Bayesian inverse problems with heterogeneous variance

Natalia Bochkina\textsuperscript{1,2} | Jenovah Rodrigues\textsuperscript{1,2}

\textsuperscript{1}School of Mathematics, University Of Edinburgh, Edinburgh, UK
\textsuperscript{2}Maxwell Institute, Edinburgh, UK

Abstract
We consider inverse problems in Hilbert spaces under correlated Gaussian noise, and use a Bayesian approach to find their regularized solution. We focus on mildly ill-posed inverse problems with fractional noise, using a novel wavelet-based vaguelette–vaguelette approach. It allows us to apply sequence space methods without assuming that all operators are simultaneously diagonalizable. The results are proved for more general bases and covariance operators. Our primary aim is to study posterior contraction rate in such inverse problems over Sobolev classes and compare it to the derived minimax rate. Secondly, we study effect of plugging in a consistent estimator of variances in sequence space on the posterior contraction rate. This result is applied to the problem with error in forward operator. Thirdly, we show that empirical Bayes posterior distribution with a plugged-in maximum marginal likelihood estimator of the prior scale contracts at the optimal rate, adaptively, in the minimax sense.

KEYWORDS
Bayesian model, contraction rate, empirical Bayes, error in operator, fractional Brownian motion, inverse problems, plug-in estimator

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2023 The Authors. Scandinavian Journal of Statistics published by John Wiley & Sons Ltd on behalf of The Board of the Foundation of the Scandinavian Journal of Statistics.

wileyonlinelibrary.com/journal/sjos

Scand J Statist. 2023;50:1116–1151.
1 | INTRODUCTION

1.1 | Bayesian linear inverse problem

Consider the following probability model for $Y$ which are noisy indirect observations of an unknown function $\mu$:

$$Y = K\mu + \epsilon W,$$

where $\mu \in H_0$, a separable Hilbert space and a known, injective, continuous, linear operator $K$ maps $\mu$ into another separable Hilbert space, $H_1$. Here a scalar $\epsilon$ represents the level of noise, and $W$ is a random Gaussian process in $H_1$. See Knapik et al. (2011) for a statistical interpretation of this model. We consider the setting where $W$ is not necessarily a white noise. This occurs for differential operators (Agapiou et al., 2013) and econometric problems (Florens & Simoni, 2016). Also, model (1) can be viewed as a continuous experiment used to study asymptotically equivalent discrete models (Johnstone & Silverman, 1997; Schmidt-Hieber, 2014).

We consider ill-posed problems when the solution, even of a noise free problem (with $\epsilon = 0$), does not depend continuously on observations. This happens, for instance when the eigenvalues of operator $KTK$ decay to 0 (Cavalier, 2008). Typically, most methods for solving linear ill-posed inverse problems involve regularizing the solution space, by constraining the set of solutions using some a priori information such as a small norm, sparsity or smoothness, normally leading to a unique solution in a noise free case. For further details, see Engl et al. (1996). Most regularized solutions can be interpreted as a Bayesian estimator where the regularization is reflected as the prior information. For a more detailed discussion of the correspondence between the penalized likelihood and Bayesian approaches, see Bochkina (2013).

However, the Bayesian perspective brings more than merely a different characterization of a familiar numerical solution. Formulating a statistical inverse problem as one of inference in a Bayesian model has great appeal, notably for what this brings in terms of coherence, the interpretability of regularization penalties, the integration of all uncertainties, and the principled way in which the set-up can be elaborated to encompass broader features of the context, such as measurement error, indirect observation, etc. The Bayesian formulation comes close to the way that most scientists intuitively regard the inferential task, and in principle allows the free use of subject knowledge in probabilistic model building (Auranen et al., 2005; Cotter et al., 2009; Kaipio et al., 1999; Röver et al., 2007; Voutilainen & Kaipio, 2009, etc.). For an interesting philosophical view on inverse problems, falsification, and the role of Bayesian argument, see Tarantola (2006). Various Bayesian methods to solve practical inverse problems have been proposed by Efendiev et al. (2010), Cotter et al. (2009), Dashti et al. (2012), among many others.

1.2 | Posterior consistency and contraction rate

The solution to an inverse problem in the presence of noise is usually analyzed by taking the limit of the noise $\epsilon \to 0$. In a Bayesian approach, the solution is a probability measure $\mathbb{P}(% \mu \in \cdot | Y)$ over a set of functions $\mu$ which depends on observations $Y$, making it a random
probability measure over a set of functions. Ghosal et al. (2000) proposed to study *contraction rate* of the posterior distribution which is defined as the smallest $r_\varepsilon = r_\varepsilon(\varepsilon)$ such that for every $M \to \infty$,

$$
\mathbb{P}(\mu : ||\mu - \mu_0|| \geq Mr_\varepsilon|Y) \xrightarrow{\mathbb{P}_{\mu_0}} 0, \text{ as } \varepsilon \to 0,
$$

uniformly over the true solution $\mu_0$ in a relevant functional class (e.g. a Sobolev class). Here $\mathbb{P}_{\mu_0}$ is the true distribution of $Y$ data under model (1) with $\mu = \mu_0$.

In case of the inverse problem under the white noise model, the nonadaptive rate of contraction of the posterior distribution in a sequence space with Gaussian prior was studied by Knapik et al. (2011). More generally, Ray (2013) studied general adaptive priors, including wavelets, with suboptimal rates. When the covariance operator of the noise is not identity, this problem was studied by Agapiou et al. (2013) and Florens and Simoni (2016), with the particular types of covariance operators motivated by the respective areas of application. Motivated by inverse problems arising in PDEs, Agapiou et al. (2013) considered a case where the covariance operators do not necessarily commute. In such a challenging setting, to obtain conditions for the contraction of the posterior distribution, the authors assumed the unknown function to be continuous, and expressed the smoothness of the unknown function in terms of the prior covariance operator; their contraction rates were slower than the minimax optimal ones in the case of white noise (Knapik et al., 2011), although in many cases the exponent in the rate could be arbitrarily close to the optimal exponent. Florens and Simoni (2016) investigated contraction of the posterior distribution in a challenging case of nontrivial covariance operators motivated by inverse problems arising in econometrics; to overcome the challenges, the authors assumed the covariance operator of the noise is trace class and true functions have monotonically decreasing coefficients in some basis (resulting in a subclass of Sobolev spaces) where they showed that the posterior contracts at the minimax rate, up to a log factor.

In the adaptive setting, where the posterior distribution adapts to the unknown smoothness of $\mu_0$, a direct problem in a sequence space with nondecreasing known variances (which is equivalent to an ill-posed inverse problem with white noise) was considered by Belitser (2017), under a restricted bias condition using a sieve prior. Agapiou and Mathé (2018) proposed to choose a data-dependent prior mean as a way of making their posterior distribution adaptive and to contract at the minimax optimal rate; such an approach is common in optimization but in its current form may be less intuitive to a Bayesian statistician. Other ways to achieve adaptation are to use an empirical Bayes and a full Bayesian estimator of, either a prior hyperparameter under a white noise error model with a self-similarity assumption (Knapik et al., 2016), the truncation number under the white noise model (Johannes et al., 2020), or a plug-in estimator of the prior scale parameter for direct models under white noise and Hölder spaces (Szabó et al., 2013), that lead to the posterior contraction rate that achieves the minimax rate, up to a constant. Florens and Simoni (2016) proposed an estimator of the prior scale parameter in inverse problems with heterogeneous noise, which resulted in a suboptimal rate of contraction. In a different setting, Knapik and Salomond (2018) showed that adaptive hierarchical mixture models lead to optimal posterior contraction rates. We emphasize that for inverse problems, only the approach of Johannes et al. (2020) leads to the posterior distribution contracting at the optimal rate (under white noise) without a restriction to a subset of functions.
1.3 | Our contribution

In this paper, we focus on linear inverse problems with Gaussian noise where it is possible to reduce model (1) to the sequence space using a Riesz basis of $H_0$ so that its image under the forward operator is a basis of $H_1$, for which there exists a biorthogonal basis. This generalizes a standard assumption of reduction to a sequence space where the covariance operator $V$ of the Gaussian noise $W$ and operator $KK^T$ are simultaneously diagonalizable. We consider covariance operators that do not have to belong to the trace class (e.g. white noise or the generalized derivative of fractional Brownian motion [fBM], called fractional noise), nor do we constrain our unknown function of interest to be continuous or to have monotonically decreasing coefficients in some basis. Motivation for this approach comes from wavelet-based approaches to inverse problems (Abramovich & Silverman, 1998), that apply to homogeneous operators $K$, and representation of fractional noise in terms of biorthogonal wavelet bases (Meyer et al., 1999), leading to a novel approach we refer to as vaguelette-vaguelette that we then generalize to non-wavelet bases. Our first contribution is the derivation of the posterior contraction rate for mildly ill-posed problems under fractional noise using the vaguelette-vaguelette approach, and identifying nonadaptive priors leading to optimal posterior contraction rate, in the minimax sense, for the true function belonging to a Sobolev space. The results are derived in a more general setting.

Our second contribution is to investigate the case where the variances in the sequence space may be unknown, and how using their plug-in estimator affects the posterior contraction rate. We illustrate this on an example with repeated observations. We also study the problem of error in operator and identify conditions when the rate of posterior contraction of $\mu$ is not affected.

Our third contribution is the study of the empirical Bayes posterior with a plugged-in maximum marginal likelihood estimator of the prior scale under an appropriate Gaussian prior, in the sequence space. We show that the corresponding empirical Bayes posterior distribution contracts at the optimal rate (in the minimax sense), adaptively over a set of Sobolev classes, under some conditions on prior smoothness. Also, we show that under an additional self-similarity condition, the maximum marginal likelihood estimator of the prior scale concentrates around its oracle value.

The paper is organized as follows. In Section 2 we introduce the Bayesian formulation of an inverse problem, fBM and its generalized derivative, mildly ill-posed inverse problems and problems with error in operator. In Section 3 we review a wavelet-based formulation of inverse problems. In Section 4 we consider mildly ill-posed inverse problems under fractional noise, formulate the vaguelette-vaguelette approach and study the posterior contractions rates for the problems discussed in Section 2. We also study properties of an empirical Bayes posterior distribution for this problem that adapts to unknown smoothness of $\mu_0$. In Section 5 we formulate this problem for more general bases that lead to a sequence space representation of (1) with arbitrary sequences of coefficients of ill-posedness, prior and noise variances, and study the posterior contractions rates for the problems discussed in Section 2. In Section 6 these general results are applied to the case where all sequences (of coefficients of ill-posedness, prior and noise variances) decay polynomially, which includes mildly ill-posed problems with fractional noise. We also study properties of an adaptive empirical Bayes posterior distribution in this setting. The effects of the choice of the prior parameters and sample size on the posterior contraction rate is illustrated on simulated (synthetic) data, with both fixed and estimated prior scale, for mildly ill-posed inverse problems (Section 7). We conclude with a discussion. The proofs are given in Appendix S1.
1.4 Notation

\( a_n \geq b_n \) denotes the existence of \( c, C : 0 < c \leq a_n / b_n \leq C < \infty \) for all \( n \geq 1 \). Here \( \| \cdot \| \) denotes the Euclidean vector norm if the argument is a vector, and the \( L^2 \) norm if the argument is a function. For an operator \( K : L^2 \to L^2 \), where \( L^2 \) is viewed as a Hilbert space, define operator \( K^T \) as follows: for any \( f, g \in L^2 \), \( \langle K^T f, g \rangle = \langle f, Kg \rangle \).

2 BAYESIAN INVERSE PROBLEM WITH HETEROGENEOUS GAUSSIAN NOISE

2.1 Bayesian inverse problem

Note that model (1) in Hilbert space \( L^2[0, 1] \) can be viewed as the following model

\[
d\tilde{Y}_t = (K\mu)(t)dt + \epsilon d\tilde{W}_t, \quad t \in [0, 1],
\]

with \( Y = (d\tilde{Y}_t / dt, t \in [0, 1]) \) and \( W = (d\tilde{W}_t / dt, t \in [0, 1]) \) where the derivatives are generalized, that is, in both cases the models can only be used to compute scalar products with appropriate test functions (Knapik et al., 2011). It has been shown that model (3) is asymptotically equivalent, in Le Cam distance, to a finite dimensional model with growing sample size \( n \); if the errors are independent then such a finite dimensional model is equivalent to (3) with \( \tilde{W}_t \) being Brownian motion and \( \epsilon = n^{-1/2} \) (the white noise model), while if the errors are correlated it is equivalent to (3) with \( \tilde{W}_t \) being fBM with parameter \( H \) and \( \epsilon = n^{-(1-H)} \) for \( H > 1/4 \) under an assumption on correlation errors, for details on the direct problem with \( K = I \) see Schmidt-Hieber (2014).

We assume that \( W_t \) is a Gaussian process with covariance operator \( V : H_1 \to H_1 \) such that for any bounded function \( f, g \in L^2[0, 1] \), \( \langle Vf, g \rangle = E \left[ \int f(s)g(t)W_sdW_t ds dt \right] \). If \( W_t \) is a derivative of a Brownian motion, that is, (1) is a white noise model, then \( V = I \). Note that \( V \), just like the identity element of \( H_1 \), does not have to be of trace class.

We consider the prior distribution of \( \mu \) to be a centred Gaussian process

\[
\mu \sim N(0, \Lambda),
\]

with covariance operator \( \Lambda : H_0 \to H_0 \) belonging to a trace class (\( \text{trace}(\Lambda) < \infty \)). Here we also consider \( H_0 = L^2[0, 1] \).

Then, the posterior distribution of \( \mu \) is given by

\[
\mu|Y \sim N \left( (K^TV^{-1}K + \epsilon^2\Lambda^{-1})^{-1}K^TV^{-1}Y, (\epsilon^{-2}K^TV^{-1}K + \Lambda^{-1})^{-1} \right).
\]

Similar results were presented in Agapiou et al. (2013) and Florens and Simoni (2016). The posterior distribution is proper if \( \text{trace} \left( (\epsilon^{-2}K^TV^{-1}K + \Lambda^{-1})^{-1} \right) < \infty \).

We assume that the true distribution of \( Y_t, t \in [0, 1] \) is given by model (1) with \( \mu = \mu_0 \) that we denote by \( P_{\mu_0} \); sometimes we write \( P_{\mu_0,Y} \) if we need to be specific about the covariance operator of the noise. The expectation with respect to this distribution is denoted by \( \mathbb{E}_{\mu_0} \).
Our aim is to determine the contraction rate of the posterior distribution of $\mu$ given $Y$ around the true value $\mu_0$ as the error $\epsilon \to 0$. Next we introduce our motivating example, fBM noise, and mildly ill-posed inverse problems.

2.2 | fBM and fractional noise

In this section we focus on a particular case where $W_t$ in (1) is a generalized derivative of fBM, which we call fractional noise, defined below.

Definition 1. Let $B = (B_t)_{t \in [0,1]}$ be a centered Gaussian process. $B$ is called a fBm with Hurst exponent $H \in (0,1)$ if it has the following covariance structure:

$$E(B_tB_s) = \frac{1}{2} (|t|^{2H} + |s|^{2H} - |t-s|^{2H}).$$

fBm is a self-similar process with stationary increments. When $H = 1/2$, it coincides with the standard Brownian motion. Sample paths of fBM are Hölder continuous of any order strictly less than $H$. See Picard (2011) for further details.

Series representations of fBM in various orthonormal bases and spanning frames of $L^2[0,1]$ are available. These bases can be normalized harmonics $\sin(\alpha_k t)$ and $1 - \cos(\theta_k t)$ with specific frequencies $\alpha_k$ and $\theta_k$. Dzhaparidze and van Zanten (2005) also extended it to the case of multidimensional support, and more involved bases in Schmidt-Hieber (2014) and Kühn and Linde (2002). A general series representation is given in Gilsing and Sottinen (2003), as well as other bases such as shifted Legendre polynomials and modified hypergeometric functions. A good introduction to fBM is given by Picard (2011). Efficient simulation of fBM is discussed in Ndaoud (2018), who gives a series representation of a fBM with Hurst exponent $H$ in terms of a trigonometric series, which implies that its generalized derivative (fractional noise) can be decomposed in a series with functions $(1, \cos(\pi kt), \sin(\pi kt), \ k \in \mathbb{N})$.

Meyer et al. (1999) showed that one-dimensional fBM with Hurst exponent $H \geq 1/4$ can be written as a wavelet series which we discuss in detail in Section 3, alongside a wavelet-based representation of inverse problem (1).

2.3 | Mildly ill-posed inverse problems

In this paper we focus on mildly ill-posed inverse problems. An inverse problem (1) is called mildly ill-posed if the eigenvalues $\{k_i^2\}$ of operator $KTK$, arranged in the decreasing order, decrease polynomially.

Assumption 1. There exist $p \geq 0, C_1 \geq 1$ such that the eigenvalues $(k_i^2)$ satisfy $C_1^{-1}i^{-p} \leq k_i \leq C_1i^{-p}$ for all $i = 1, 2, \ldots$.

Examples of such inverse problems are, for instance, Volterra operator (Section 7) and Radon operator used in tomography. They also arise in econometrics (Florens & Simoni, 2016).

2.4 | Error in operator

In this paper we also study the problem when operator $K$ is observed with error, in particular how this affects the posterior contraction rate of $\mu$. Hoffmann and Reiss (2008) studied the
inverse problem (1) with white noise (i.e. $V = I$) where the operator is observed with errors, namely
\[ \hat{K} = K + \delta Z, \]
where $Z$ is the operator white noise, independent of observational noise $W$ and $\delta$ is the observation error. We will extend this approach to a model with dependent errors (Section 5.4.3), and in particular to the case of fractional noise (Section 4.4).

3 | INVERSE PROBLEMS AND FRACTIONAL NOISE VIA WAVELETS

Donoho (1995) proposed two wavelet-based approaches to linear inverse problems with operator $K$ based on an orthonormal wavelet basis $(\psi_{jk})$, called vaguelette-wavelet and wavelet-vaguelette which we introduce below.

3.1 | Wavelets and Riesz basis

A one-dimensional, orthonormal, compactly supported wavelet basis is given by \{ $\phi$, $\psi_{j,k}$, $j = 0, 1, \ldots$, $k = 0, 1, \ldots, 2^j - 1$ \} with $\psi_{j,k}(x) = 2^{j/2}\psi(2^jx - k)$, where $\phi$ is the scaling function (father wavelet) and $\psi$ is the wavelet function (mother wavelet), both having a compact support. To simplify the notation, it is common to denote $\psi_{-1,0} = \phi$. Wavelets are said to be of regularity $r$ if they have $r$ continuous derivatives and vanishing moments up to order $r$, i.e. $\int x^k \psi(x)dx = 0$ for all $k = 1, 2, \ldots, r$. We denote an orthonormal wavelet basis on $[0, 1]^d$ for a fixed $d$ by $(\psi_v, v \in \mathcal{Y})$, for a countable $\mathcal{Y}$. In the one-dimensional setting, with the wavelet and the scaling functions supported on $[0, 1]$, $v = (j, k)$ and $\mathcal{Y} = \{(j, k) : j = 0, 1, 2, \ldots, k = 0, 1, \ldots, 2^j - 1\}$. See Vidakovic (1999) for the compactly supported wavelets and their statistical applications, including higher dimensional wavelets.

A wavelet approach to inverse problems involves generalization of an orthonormal basis to a Riesz basis that we define below.

**Definition 2.** Sequence of functions $(v_v)$ forms a Riesz basis in $L^2[0, 1]$ if $(v_v)$ spans $L^2[0, 1]$ and there exist $0 < A \leq B < \infty$ such that for any square summable sequence $(c_v)$,
\[ A \sum_{v \in \mathcal{Y}} c_v^2 \leq \| \sum_{v \in \mathcal{Y}} c_v v_v \|_2^2 \leq B \sum_{v \in \mathcal{Y}} c_v^2. \]

Typically a Riesz basis is an image of an orthonormal basis under some bounded invertible operator. Donoho (1995) showed that this also applies to what he called weakly invertible operators for a given orthonormal basis. An orthonormal basis is trivially a Riesz basis with $A = B = 1$ due to Parseval’s identity.

We will also rely on the concept of biorthogonal bases which provides an infinite dimensional generalization of the singular value decomposition of matrices.

**Definition 3.** Given a set of functions $(v_v)_{v \in \mathcal{Y}}$ spanning $L^2[0, 1]$, a biorthogonal basis $(w_v)_{v \in \mathcal{Y}}$ to $(v_v)_{v \in \mathcal{Y}}$ satisfies
\[ \langle v_v, w_v \rangle = 1 \text{ for all } v \in \mathcal{Y} \text{ and } \langle v_{v_1}, w_{v_2} \rangle = 0 \text{ for all } v_1 \neq v_2, \quad v_1, v_2 \in \mathcal{Y}. \]

For more details on biorthogonal bases, including construction, see Vidakovic (1999).
### 3.2 Sobolev classes in terms of wavelets

Note that the Sobolev norm can be written equivalently in terms of wavelet coefficients for wavelets with regularity \( r > \beta \),

\[
S^\beta(A) = \left\{ f(x) = \sum_k a_k \phi(x - k) + \sum_{j=0}^\infty \sum_k b_{jk} \psi_{jk} : ||a|| + \left[ \sum_{j=0}^\infty 2^{2j\beta} ||b_j||^2 \right]^{1/2} \leq A \right\},
\]

(8)

Johnstone and Silverman (1997) and Cohen et al. (2000) show the equivalence of the Sobolev norm and the above sequence norm for biorthogonal compactly supported wavelet bases with regularity \( r > \beta \).

### 3.3 Wavelet representation of fractional noise

Meyer et al. (1999) proved that a fractional derivative, or antiderivative \( Z_{d+1/2}(t) = D^{-d}Z(t) \), of white noise \( Z(t) \) has the same distribution as the generalized derivative of the fBM with the Hurst exponent \( H = d + 1/2 \) for \( d < 1/2 \), and the same distribution as the following wavelet random series:

\[
W_t = \sum_{k\in D} \tilde{z}_k \phi_{\Delta}^{(d)}(t - k) + \sum_{v\in \mathbb{Y}} \sigma_v \tilde{z}_v \psi_{\psi}^{(d)}(t), \tag{9}
\]

with iid \( z_v \sim N(0, 1), \sigma_v = C2^{-j(H-1/2)}, v = (j, k) \). Here \( \tilde{z}_k = \sum_{\ell=0}^\infty \gamma_{\ell}^{(d)} \xi_{k-\ell} \) is a fractional ARIMA(0, d, 0) process bounded with high probability, where the coefficients \( \gamma_{\ell}^{(d)} \) are defined by

\[
(1 - x)^{-d} = \sum_{k=0}^\infty \gamma_{k}^{(d)} x^k, \quad \text{for } |x| < 1. \tag{10}
\]

with property \( |\gamma_{k}^{(d)}| = O(k^{-1-d}) \) for large \( k \). Here the scaling function \( \phi_{\Delta}^{(d)} \) and the wavelet function \( \psi^{(d)} \) are constructed from wavelet and scaling functions \( \psi \) and \( \phi \) in terms of their Fourier transform:

\[
\phi_{\Delta}^{(d)}(\xi) = \left( \frac{e^{i\xi}}{1-e^{i\xi}} \right)^d \hat{\phi}(\xi) \quad \text{and} \quad \psi^{(d)}(\xi) = e^{i\xi d} \hat{\phi}(\xi) \]

where \( \hat{f}(\xi) = \int e^{-it\xi} f(t)dt \), so that all of their integer moments vanish (see Meyer et al., 1999, equation (3.1), for the full set of conditions and Proposition 2 for the statement). This gives a Riesz wavelet basis \((\phi_{\Delta}^{(d)}(\cdot - m), \psi_{\psi}^{(d)})\), with \( \psi_{\psi}^{(d)} \) being dilations and translations of a single function \( \psi^{(d)} \), which is biorthogonal to the wavelet Riesz basis \((\phi_{\Delta}^{(d)}(\cdot - m), \psi_{\psi}^{(d)})\). Ayache and Taqqu (2003) showed this under weaker conditions on wavelets \( \psi \), including compactly supported Daubechies wavelets of regularity \( r \geq 6 \), as well as alternative wavelet representations of fractional noise (precise conditions on wavelets \( \psi \) are given by the authors on p. 456). Abry and Sellan (1996) constructed the corresponding filters that allow fast practical implementation of wavelet transform and its reverse using the cascade algorithm. To simplify the notation, we denote \( Y = Y_\phi \cup Y_\psi \) with \( v \in Y_\phi = \{-1,k\}, k \in D \), and \( \psi_{\psi}^{(d)} = \phi_{\Delta}^{(d)}(\cdot - k) \) for \( v = (-1, k) \). For compactly supported wavelets, \( D \) and \( Y_\phi \) are finite, and \( Y_\psi = \{(j, k) : j = 0, 1, \ldots, k \in K_j, |K_j| \leq C_{\psi} 2^j \} \).
3.4 A wavelet-vaguelette approach

Wavelets transformed by some fixed operator $A$ are called vaguelettes. In this section we describe the wavelet-vaguelette approach, and in the next we discuss the vaguelette-wavelet approach.

A wavelet-vaguelette approach applies to operators $K$ such that wavelets $\psi_v$ are in their image, and involves the numerical evaluation of $K^{-1}\psi_v$ (Donoho, 1995). Then, we can write $(K\mu_0)(t) = \sum_{\varpi \in \Upsilon} c_{\varpi} \psi_v(t)$ and hence

$$
\mu_0(t) = \sum_{\varpi \in \Upsilon} c_{\varpi} \tilde{\beta}_v^{-1} (K^{-1}\psi_v)(t) \tilde{\beta}_v = \sum_{\varpi \in \Upsilon} \mu_{0,\varpi} \tilde{v}_v(t),
$$

with $\mu_{0,\varpi} = c_{\varpi} / \tilde{\beta}_v$ and normalized vaguelettes $\tilde{v}_v(t) = (K^{-1}\psi_v)(t) \tilde{\beta}_v$. Here $\tilde{\beta}_v$ are normalization factors so that $||\tilde{v}_v|| = 1$. This approach is applicable if the normalized vaguelettes $\tilde{v}_v(t)$ span $L^2[0,1]$ and form a Riesz basis.

If the noise $W$ can be written as

$$
W_t = \sum_{\varpi \in \Upsilon} \sigma_{\varpi} z_{\varpi} \psi_v(t), \quad (11)
$$

for some coefficients $\sigma_{\varpi}$ and iid $z_{\varpi} \sim N(0,1)$, then the model for the wavelet transform of the observations $y_v = \langle \psi_v, Y \rangle$, under (1), is given by $y_v | \mu_v \sim N(\mu_v \tilde{\beta}_v, \sigma_v^2 \tilde{\beta}_v^2)$, independently for $v \in \Upsilon$.

Therefore, under the above assumptions, using the wavelet-vaguelette approach reduces problem (1) to a sequence space model. For fractional noise, as far as we are aware, representation (11) does not hold.

3.5 A vaguelette-wavelet approach

A vaguelette-wavelet approach works if $\psi$ is in the image of $K_T$, and the normalized image of these functions under operator $K$ forms an orthonormal Riesz basis in $L^2[0,1]$ (Donoho, 1995). Formally these assumptions are formulated as follows.

**Definition 4.** Given a linear operator $K$ and a wavelet basis $\psi_v \in \Upsilon$ assume that there exist scalars $\beta_v$, $v \in \Upsilon$ such that the functions $v_v = K\psi_v / \beta_v$ form a Riesz basis, and there exists a biorthogonal basis $(w_v)_{v \in \Upsilon}$ for $(v_v)_{v \in \Upsilon}$, defined by (6) and (7), respectively. Such functions $(v_v)$ are called vaguelettes.

Abramovich and Silverman (1998) give examples of linear operators that satisfy this condition, such as homogeneous operators $K$, where for all $t_0$ and $s > 0$, $K[f(s(t - t_0))] = s^{-p}K(f(s(t - t_0))]$ for some constant $p$, called an index of an operator. Examples of homogeneous operators include integration, differentiation (including fractional integration and differentiation) and other operators such as Radon transform in an appropriate coordinate system (Kolaczyk, 1996). For such operators, the normalizing constants are $\beta_{v,\varpi} = C_{p} 2^{-pl}$, and the corresponding vaguelettes are dilations and translations of a single function but they are not necessarily mutually orthogonal. This property also holds for some convolution operators (Donoho, 1995).

In this approach, $\mu = \sum_{v \in \Upsilon} \mu_v \psi_v$, and hence $K\mu = \sum_{v \in \Upsilon} \mu_v \beta_v \psi_v$. If $(w_v)$ is a biorthogonal basis for $(v_v)$, i.e. $\int v_v w_v = 0$ for all $v \in \Upsilon$, then the sequence space model can be written as

$$
y_v \sim N(\mu_v \beta_v, \sigma_v^2), \quad v \in \Upsilon,
$$
where \( y_o = \int Yw_o \) and \( \sigma_y^2 \) is such that \( \epsilon^2 \sigma_y^2 = \text{Var}(y_o) = \int V(s, t)w_o(t)w_o(s)dsdt \). \( y_o \)'s are independent if \( \text{Cov}(y_o, y'_o) = \int V(s, t)w_o(t)w_o(s)dsdt = 0 \) for all \( o \neq o' \). Note that for fractional noise this approach applies with fractional wavelets which are not necessarily an image of an orthogonal wavelet basis under \( K \) (Section 3.3).

In the next section we extend these approaches to use representation of fractional noise in terms of vaguelettes \( \psi^{(d)} \) (Section 3.3) and give results on the posterior contraction rates for mildly ill-posed inverse problems, under fractional noise, using a wavelet approach with \( \mu_0 \) in a Sobolev class, which in turn will rely on general results stated in Section 5.

4 | POSTERIOR CONTRACTION RATE UNDER FRACTIONAL NOISE

In this section we study the posterior contraction rate of mildly ill-posed inverse problems introduced in Section 2.3 with fractional noise decomposed in a wavelet series (Section 3.3), including the cases of error in operator, unknown Hurst exponent \( H \) and adaptation to unknown smoothness of \( \mu_0 \). First, we introduce a new approach, called vaguelette-vaguelette, to account for the vaguelette type of wavelet used in the decomposition of a fractional noise.

4.1 | Vaguelette-vaguelette approach

Following the series decomposition of fractional noise in terms of biorthogonal wavelets \( \left( \psi^{(d)}_o \right) \) and \( \left( \psi^{(-d)}_o \right) \) with \( d = H - 1/2 \) (Section 3.3), we need to adapt the wavelet-vaguelette approach to the case of wavelet \( \psi \) being replaced by \( \psi^{(d)} \), which is a particular kind of vaguelette determined by the covariance operator of noise. We call this a vaguelette-vaguelette approach. Similarly to the wavelet-vaguelette approach, the vaguelette-vaguelette approach applies to operators \( K \) such that wavelets \( \psi^{(d)}_o \) are in their image, which holds for example if range\((K) \subseteq \text{range}(D^{-d}) \) where \( D^{-d} \) is a differentiation/integration operator (Section 3.3).

Then, we can write \( (K\mu)(t) = \sum_{o \in \Upsilon} c_o \psi^{(d)}_o(t) \) with \( c_o = \langle \psi^{(-d)}_o, K\mu \rangle \) and hence
\[
\mu(t) = \sum_{o \in \Upsilon} c_o k^{-1}_o \left( K^{-1} \psi^{(d)}_o \right)(t) k_o = \sum_{o \in \Upsilon} \mu o \tilde{\psi}^{(d)}_o(t),
\]
with \( \mu_o = c_o / k_o \) and normalized vaguelettes
\[
\tilde{\psi}^{(d)}_o / k_o = K^{-1} \psi^{(d)}_o = K^{-1} D^{-d} \psi_o,
\]
with normalization factors \( k_o \) so that \( \left\| \tilde{\psi}^{(d)}_o \right\| = 1 \).

This approach is applicable if the normalized vaguelettes \( \left( \tilde{\psi}^{(d)}_o \right) \) span \( L^2[0, 1] \) and form a Riesz basis. Under the former assumption, any \( \mu(t) \in L^2[0, 1] \) can be decomposed into this vaguelette basis, and the latter assumption allows us to study the posterior contraction rate of \( \mu \) in terms of its coefficients \( \mu_o \).

Following the wavelet-vaguelette and vaguelette-wavelet approaches, we call this inverse problem mildly ill-posed if for some \( p > 0 \)
\[
\kappa_o = \left\| K^{-1} \phi^{(d)}_{\Delta o} \right\| \lesssim 1, \; o \in \Upsilon_{\phi}, \text{ and } \kappa_o = \left\| K^{-1} \psi^{(d)}_o \right\| \lesssim 2^{-jp}, \; o = (j, k) \in \Upsilon_{\psi}.
\]
(12)
Recall that this holds for homogeneous operators $K$ (Section 3.5).

Then, using the wavelet representation of a fractional noise (Section 3.3), the sequence space model for the wavelet transform of observations $y_\nu = \langle \psi_\nu, Y \rangle$ under model (1) with a fractional noise is given by

\[ y_\nu \sim N \left( \mu_\nu \kappa_\nu, \epsilon^2 \sigma_\nu^2 \right) \text{, } \nu \in \mathcal{Y}_\psi \text{ independently} \, , \tag{13} \]

with $\sigma_\nu \propto 2^{-j(H-1/2)}$ for large $j$ and $\epsilon = n^{-H-1}$, and independently of $y_{-1,k} = \langle \phi_{\Delta,k}^{(-d)} \rangle, k \in D$. Note that the coefficient $y_{-1,k}$ is nonzero only if $(-k, -k + 1) \cap \supp \left( \phi_{\Delta}^{(-d)} \right) \neq \emptyset$, which implies that $D$ is a finite set since $\supp \left( \phi_{\Delta}^{(-d)} \right)$ is finite. Hence, the scaling coefficients $y_D = (y_{-1,k}, k \in D)$ satisfy

\[ y_D \sim N \left( K_D \mu_D, \epsilon^2 V_D \right) , \]

where $\mu_D = (\mu_{-1,k}, k \in D), K_D = \text{diag} \left( \kappa_{-1,k}, k \in D \right)$ and the covariance matrix $V_D$ has variances $\sigma_{-d}^2$ on the diagonal and $V_{D,k,m} = \sigma_{-d}^2 |_{k-m}|$ where $\sigma_{-d}^2$ and $\rho_{m}^{(-d)}$ are defined in terms of the sequence $\gamma_{\nu}^{(-d)}$ (see (10)) by

\[ \sigma_{-d}^2 = \sum_{\ell=0}^{\infty} \left[ \gamma_{\nu}^{(-d)} \right]^2 < \infty \text{ and } \rho_{m}^{(-d)} = \sigma_{-d}^2 \sum_{\ell=0}^{\infty} \gamma_{\nu}^{(-d)} \gamma_{\nu+m}^{(-d)} . \]

Note that the support of Daubechies wavelets is of length at most $2r$ where $r$ is the regularity of the wavelets, hence, $D$ has at most $2r$ elements. Recall that $\mathcal{Y}_\psi = \{(-1,k), k \in D\}$.

We consider a prior distribution on $\mu$ defined in terms of random series $\mu = \sum_{\nu \in \mathcal{Y}} \mu_\nu \tilde{\nu}^{(-d)}(t)$ with $\mu_\nu \sim N(0, \lambda_\nu)$ independently for $\nu \in \mathcal{Y}$. We take $\lambda_\nu = \tau 2^{-2\alpha j}$ for the wavelet coefficients $\nu = (j,k) \in \mathcal{Y}_\psi$, for some $\tau, \alpha > 0$. Denote the prior variances for the scaling coefficients as a diagonal matrix $\Lambda_D = \text{diag}(\lambda_{-1,k}, k \in D)$. In the context of wavelets, it is common to take a non-informative prior for scaling coefficients, that is, $\lambda_{-1,k} = +\infty$, or its proper approximation with large prior variances. For $\lambda_{-1,k} = +\infty, k \in D$, we have $\mu_D|y_D \sim N(K_D^{-1}y_D, \epsilon^2 K_D^{-1}V_D K_D^{-1})$, which holds approximately if $\lambda_{-1,k}, k \in D$ are large but finite. If $\lambda_\nu = \tau 2^{-2\alpha j}$ for $\nu = (j,k) \in \mathcal{Y}_\psi$ and $\lambda_\nu$ is finite for $\nu \in \mathcal{Y}_\psi$, then a priori, with high probability, $\mu \in S^{2\alpha}$ for all $\alpha' < \alpha$.

For such a prior distribution, the corresponding posterior distribution is

\[ \mu_\nu|y_\nu \sim N \left( \frac{y_\nu \kappa_\nu \lambda_\nu}{\lambda_\nu \kappa_\nu^2 + \epsilon^2 \sigma_\nu^2}, \frac{\sigma_\nu^2 \lambda_\nu}{\epsilon^{-2} \lambda_\nu \kappa_\nu^2 + \sigma_\nu^2} \right) \text{, independently for } \nu \in \mathcal{Y}_\psi \, , \]

and for the scaling coefficients, independently of $\mu_\nu|y_\nu$ for $\nu \in \mathcal{Y}_\psi$, we have

\[ \mu_D|y_D \sim N \left( (K_D^T V_D^{-1} K_D^T + \epsilon^2 \Lambda_D^{-1})^{-1} K_D^T V_D^{-1} y_D, \left( \epsilon^{-2} K_D^T V_D^{-1} K_D^T + \Lambda_D^{-1} \right)^{-1} \right) . \]

### 4.2 Minimax rate of an inverse problem with fractional noise

In this section we derive the minimax rate of convergence, in $L^2[0,1]$ norm, of estimators of $\mu$ under model (1) given the true model $P_{\mu_0}, W_t$ is a generalized derivative of fBM and the inverse problem is mildly ill-posed, that is, it satisfies assumption (12). Ghosal et al. (2000) proposed to
use it as a benchmark for posterior contraction rates. The minimax rate of convergence is defined in Section 5.3. We assume that the true function \( \mu_0 \) belongs to a generalized Sobolev class \( S^\theta(A) \) defined by (8) with wavelets \( \psi_v \) replaced by vaguelettes \( \hat{v}(d) \).

Johnstone and Silverman (1997) showed that the minimax rate of estimation in \( L^2 \) norm in the direct problem with fBM over Sobolev spaces \( S^\theta \) is \( n\gamma^{(1-H)/(2\beta+2-2H)} \), which is a particular subset of Besov spaces studied in that paper (\( S^\beta = B^\beta_{2,2} \)). Now we state the minimax rate for a mildly ill-posed inverse problem with fractional noise.

**Proposition 1.** The minimax rate of convergence in the mildly ill-posed inverse problem (1) under a fractional noise with Hurst exponent \( H \in (1/4, 1) \), under condition (12), over functions from a Sobolev class \( S^\theta \) defined by (8) with vaguelettes \( \hat{v}(d) \), is

\[
r^\alpha = \epsilon^{\gamma_{2 \beta+2-2H}} = n^{-\gamma H/(2\beta+2-2H)}. \tag{14}\]

For \( p = 0 \) this rate coincides with the rate for the direct problem (Johnstone & Silverman, 1997), and for \( H = 1/2 \) it coincides with the minimax rate under white noise. This result follows from Proposition 6 where the minimax rate is defined in a more general setting which applies here since the Sobolev norm (8) can be written equivalently as

\[
\mu_0 \in S^\theta(A) = \left\{ f = \sum_{n \in \mathbb{D}} \hat{v}_n f_n : \sum_{n \in \mathbb{D}} \hat{v}_n^2 + \sum_{i=1}^\infty i^{2\beta} f_i^2 \leq A^2 \right\}, \tag{15}\]

due to the set \( D \) being finite, (shown by using the bijective mapping of the set \( \{(j,k), j = 0, 1, 2, \ldots; k = 0, 1, \ldots, 2^j - 1\} \) into \( \mathbb{N} : (j,k) \to 2^j + k \) such that \( 2^j \leq i = 2^j + k < 2 \cdot 2^j \).

If a posterior contraction rate \( r_\epsilon \) in a mildly ill-posed inverse problem, over functions from a Sobolev class with a fractional noise, matches the rate stated in Proposition 1 up to a constant, i.e. there exist finite \( C > 0 \) and \( \epsilon_0 \) independent of \( \epsilon \) such that \( 0 < C^{-1} \leq r_\epsilon/r^\alpha \leq C < \infty \) for all \( \epsilon \leq \epsilon_0 \), then we will say that the posterior distribution contracts at the optimal rate in the minimax sense.

### 4.3 Posterior contraction rate for fractional noise

Now we study the posterior contraction rate in the setting of Section 4.1 under the vaguelette-vaguelette approach for a mildly ill-posed inverse problem with fractional noise. We consider a centred Gaussian prior on \( \mu_\nu \) with \( \lambda_\nu = \tau 2^{-2\alpha j} \) for wavelet coefficients \( \nu \in Y_\psi \), and finite \( \lambda_\nu \geq 1 \) for the scaling coefficients \( \nu \in Y_\phi \). We derive conditions on this prior, namely on its parameters \( \tau \) and \( \alpha \), so that the corresponding posterior contraction rate is optimal, in the minimax sense. First we state results with known \( \beta \) and operator \( K \), extending it to the cases when operator \( K \) is observed with error and when smoothness \( \beta \) is unknown. All the stated results follow from the general results on posterior concentration rates, which apply to more general Gaussian processes \( W \) and operators \( K \) (Section 5), with particular cases considered in Section 6.

**Proposition 2.** Consider the inverse problem (1) under fractional noise with Hurst exponent \( H \in (1/4, 1) \), with the vaguelette-vaguelette approach described in Section 4.1 under condition (12), for \( \mu_0 \in S^\theta(A) \) defined by (8). We consider a centred Gaussian prior on \( \mu_\nu \) with \( \lambda_\nu = \tau 2^{-2\alpha j} \) for wavelet coefficients, and finite \( \lambda_\nu \geq 1 \) for the scaling coefficients.
Then, the contraction rate of the posterior distribution of $\mu$ uniformly over $\mu_0 \in S^\beta(A)$ is given by
\[
\rho_c \asymp n^{-\alpha(1-H)/(p+1-H+\alpha)} + n^{-\beta(1-H)/(p+1-H+\alpha)} + \tau n^{-\alpha(1-H)/(p+1-H+\alpha)} \\
+ \tau^{-1} n^{-\min(2(p+1-H+\alpha),\beta)(1-H)/(p+1-H+\alpha)}.
\]

Therefore, the vaguelette-vaguelette approach leads to the posterior contraction rate of the same order as if they were an orthonormal basis. This result follows from Theorems 1 and 3.

**Corollary 1.** In the setting of Proposition 2, assume $\alpha \geq \beta / 2 - p - 1 + H$. Then, for $\tau^2 = \left[ n^{2(1-H)} \right]^{(a-\beta)/(1-H+\beta+p)}$, the posterior distribution contracts at the minimax convergence rate uniformly over $S^\beta(A)$.

If $\alpha < \beta / 2 - p - 1 + H$, the contraction rate is suboptimal for any choice of $\tau_c$.

It is possible to extend this result to derive the posterior contraction rate for wavelet-based estimation in inverse problems over Besov spaces. However it is known that for some Besov spaces linear estimators are suboptimal, therefore the considered Bayesian model with a Gaussian prior is not appropriate for a general estimation over Besov spaces (see Agapiou et al., 2021 for a Bayesian model with Besov prior). Therefore here we restrict ourselves to Sobolev spaces.

### 4.4 Error in operator

Now we consider the case when the operator $K$ may be observed with error: $\hat{K} = K + \delta B$, where the operator $B$ has the standard operator normal distribution, and study its effect on the posterior contraction rate of $\mu$. This problem is discussed in more detail in Section 5.4.3.

**Proposition 3.** Consider the inverse problem (1) under fractional noise with Hurst exponent $H \in (1/4, 1)$, with the vaguelette-vaguelette approach described in Section 4.1 under condition (12), for $\mu_0 \in S^\beta(A)$ defined by (8). We consider a centred Gaussian prior on $\mu_0$ with $\lambda_o = \tau 2^{-2a j}$ for wavelet coefficients, and finite $\lambda_o \geq 1$ for the scaling coefficients. Suppose $\kappa_v$ are unknown but we observe $\widehat{k}_v = \langle \hat{K}, \psi_{0}^{(-d)} \rangle$, $v \in Y_N$ where $Y_\phi \subset Y_N \subset Y$ of size $|Y_N| = N$ and $\hat{K}$ satisfies (5).

Then, the rate of contraction of the posterior distribution of $\mu$ given $y$ with plugged in $(\widehat{k}_v, v \in Y_N)$ is not affected by the error in operator if
\[
\delta = o \left( \left[ \tau_c n^{2(1-H)} \right]^{-p/(2p+2-2H+2\alpha)} \left[ \log(\tau_c n^{2(1-H)}) \right]^{-1/2} \right).
\]

In particular, for non-adaptive models (i.e. where $\beta$ is known), the posterior contraction rate of $\mu$ is optimal, in the minimax sense, for $\alpha \geq \beta / 2 - p - 1 + H$ and $\tau_c = \left[ n^{2(1-H)} \right]^{(a-\beta)/(1-H+\beta+p)}$ if
\[
\delta = o \left( n^{-(1-H)p/(p+1-H+\beta)} [\log n]^{-1/2} \right).
\]

The proof is a straightforward application of Lemma 16.

### 4.5 Empirical Bayes posterior distribution of $\mu$

As we have seen in Section 4.3, for the posterior distribution to contract at the optimal rate of convergence (in the minimax sense), prior parameters $\alpha$ or $\tau$ need to be chosen appropriately (Corollary 1), and their choice depends on smoothness $\beta$ of the true function which in practice
may be unknown. To adapt to the unknown smoothness of $\mu_0$, we estimate the scale $\tau$ of the prior covariance operator. In particular, we consider the maximum marginal likelihood estimator of $\tau$ based on the marginal distribution of $Y$ given $\tau$, and study the contraction rate of the posterior distribution of $\mu$ with this estimator as the plugged in value of $\tau$, which is usually called an empirical Bayes posterior distribution.

The marginal distribution of data $y_\nu$ for $\nu \in \mathcal{Y}$ given $\tau$ is

$$y_\nu | \tau \sim N \left( 0, \kappa_0^2 \lambda_0 + \epsilon^2 \sigma_\nu^2 \right),$$

independently for $\nu \in \mathcal{Y}$, and $\hat{\tau}$ is the value maximizing this marginal likelihood of $\tau$. See Section 6.8 for further details.

**Proposition 4.** Consider the inverse problem (1) under fractional noise with Hurst exponent $H \in \left( \frac{1}{4}, 1 \right)$, with the vaguelette-vaguelette approach described in Section 4.1 under condition (12), for $\mu_0 \in S^\beta(A)$ defined by (8). We consider a centered Gaussian prior on $\lambda_0$ with $\lambda_0 = \tau 2^{-2j}$ for wavelet coefficients, and finite $\lambda_0 \geq 1$ for the scaling coefficients.

Then, the maximum of marginal likelihood estimator (MMLE) of $\tau$ satisfies, almost surely for small $\epsilon$,

$$\hat{\tau} \leq \begin{cases} 
  e^{-2(\alpha-\beta)/(1+p-H+\beta)}(1 + o_p(1)), & \text{if } \alpha + 1/2 \geq \beta, \\
  e^{1/(3/2+p-H+\alpha)}(1 + o_p(1)), & \text{if } \alpha + 1/2 < \beta.
\end{cases} \quad (16)$$

Therefore, the posterior distribution with the plugged in estimator $\hat{\tau}$ is consistent. The contraction rate is optimal in the minimax sense uniformly over $S^\beta(A)$ for $0 < \beta \leq B_0 < \infty$ as long as

$$\alpha \geq \max(B_0 - 1/2, B_0/2 - p - 1 + H). \quad (17)$$

Note that for $\alpha + 1/2 \geq \beta$, under the assumptions of Proposition 4 and Assumption 9, the MMLE almost surely coincides with the oracle value of $\tau$ which is of the same order as the optimal value given in Corollary 1. This implies that the corresponding posterior distribution of $\mu$ with the plugged in value of $\tau$ contracts at the optimal rate, in the minimax sense. This proposition is a straightforward consequence of Theorem 4 with $\gamma = 1/2 - H$.

### 5 | BAYESIAN INVERSE PROBLEM WITH CORRELATED GAUSSIAN NOISE

Following the setup for the fractional noise and wavelet based approaches to inverse problems discussed in Section 4, we can generalize this approach to other types of correlated noise and other bases.

#### 5.1 | Assumptions

Recall that in model (1), $\mu \in H_0 = L^2[0, 1]$, and a linear operator $K$ maps $\mu$ into another separable Hilbert space, $H_1 = L^2[0, 1]$, which are both Hilbert spaces with scalar product $\langle f, g \rangle = \int_0^1 f(\lambda)g(\lambda) d\lambda$.

Now we generalize the vaguelette-vaguelette assumption to a more general set of biorthogonal bases.
Assumption 2. Assume that there exists a countable set of functions \((e_{1,v})_{v \in \mathcal{Y}}\) such that

1. functions \((e_{1,v})\) is a basis of \(H_1\) such that there exists a biorthogonal basis \((\tilde{e}_{1,v})\),
2. functions \((e_{0,v})\) form a Riesz basis of \(H_0\) where \(e_{0,v} = \kappa_v K^{-1} e_{1,v}\) with \(\kappa_v = 1/\|K^{-1} e_{1,v}\|\).

We assume that functions \((e_{1,v})\) and \((\tilde{e}_{1,v})\) are also normalized, that is, \(\|e_{1,v}\| = \|\tilde{e}_{1,v}\| = 1\). Here the set of indices \(\mathcal{Y}\) is countable, that is, it is isomorphic to the set of natural numbers \(\mathbb{N}\). This assumption describes all three considered approaches: wavelet-vaguelette with orthonormal wavelet basis \((e_{1,v})\), vaguelette-wavelet with orthonormal wavelet basis \((e_{0,v})\), and the vaguelette-vaguelette approach introduced in Section 4.1. This assumption holds also when \((e_{1,v})\) form an orthonormal eigenbasis of \(K^T K\) which implies that \((e_{0,v})\) form an orthonormal eigenbasis of \(K K^T\) and \((\kappa_v^2)\) are the corresponding eigenvalues.

Under this assumption, we can write

\[
\mu = \sum_v \mu_v e_{0,v}, \quad K \mu = \sum_v \mu_v K_1 e_{1,v}, \tag{18}
\]

where \(\mu_v = \langle K \mu, \tilde{e}_{1,v} \rangle / \kappa_v\). Note that here \(\mu_v\) may not be equal to \(\langle \mu, e_{0,v} \rangle\).

Assume that noise \(W\) in (1) is such that coefficients \(\xi_v = \langle \tilde{e}_{1,v}, W \rangle\) satisfy

\[
\xi_v \sim N(0, \sigma_v^2) \text{ independently for } v \in \mathcal{Y}_I, \quad \xi_D \sim N(0, V_D), \tag{19}
\]

and \(\xi_D\) is independent of \(\xi_v\) for all \(v \in \mathcal{Y}_I \subseteq \mathcal{Y}\), for a finite set \(D = \mathcal{Y} \setminus \mathcal{Y}_I\), some sequence \((\sigma_v)\) and an invertible covariance matrix \(V_D\). Note that we do not require \(\sum_{v \in \mathcal{Y}} \sigma_v^2\) to be finite. Normality and zero mean of \(\xi_v, v \in \mathcal{Y}\), follow as long as \(W\) is a centred Gaussian process.

Remark 1. Note that this approach is related to the model of Agapiou et al. (2013) where given operator \(K\) and Assumption 2, covariance operator of the noise and the prior covariance operator are assumed to be of the following type

\[
V(t, s) = \sum_{v \in \mathcal{Y}_I} \sigma_v^2 e_{1,v}(t)e_{1,v}(s) + \sum_{v \in D} \sum_{v' \in D'} V_{D,v',v} e_{1,v}(t)e_{1,v'}(s), \quad \sigma_v \in \mathbb{R}^Y, \quad V_D \in \mathbb{R}^{D \times D}, \tag{20}
\]

\[
\Lambda(t, s) = \sum_{v \in \mathcal{Y}_I} \lambda_v \kappa_v^2 (K^{-1} e_{1,v}(t))(K^{-1} e_{1,v}(s)), \quad \lambda_v > 0 \text{ and } \sqrt{\lambda_v} \in \ell^2(\mathcal{Y}), \tag{21}
\]

where \(V_D\) is positive definite and \(D \subset \mathcal{Y}\) is finite. In the notation of Agapiou et al. (2013), \(C_0 = \Lambda, C_1 = V\) and \(A = K\). In Section 6 we compare our posterior contraction rate to the rate given in Agapiou et al. (2013) for mildly ill-posed inverse problems.

### 5.2 Sequence space formulation

Under Assumption 2 and the noise in model (1) satisfying (19), model (1) can be written in terms of \(y_v = \langle Y, \tilde{e}_{1,v} \rangle\) as

\[
y_v \sim N(\kappa_v \mu_v, e^2 \sigma_v^2) \text{ independently for } v \in \mathcal{Y}_I = \mathcal{Y} \setminus D, \quad \text{independently of } y_D \sim N(\kappa_D \mu_D, e^2 V_D), \tag{22}
\]

where \(\kappa_D = \text{diag}(\kappa_v, v \in D)\) and \(\mu_D = (\mu_v, v \in D)\).
We also assume that the prior process for \( \mu \) is Gaussian and under decomposition (18) it satisfies
\[
\mu_0 \sim N(0, \lambda_0) \text{ independently for } \nu \in \gamma.
\] (23)

For the prior distribution of \( \mu \) to be proper, we assume \( \sum_{\nu \in \gamma} \lambda_0 < \infty \). Denote \( \Lambda_D = \text{diag}(\lambda_0, \nu \in D) \).

The corresponding posterior distribution \( \mu_0 | y_0 \) for each \( \nu \in \gamma \) is
\[
\mu_0 | y_0 \sim N \left( \frac{y_0 \kappa_0 \lambda_0}{\lambda_0 \kappa_0^2 + \epsilon^2 \sigma_0^2}, \frac{\sigma_0^2 \lambda_0}{\epsilon^{-2} \lambda_0 \kappa_0^2 + \sigma_0^2} \right), \quad \nu \in \gamma_I,
\] (24)

independently, and independently of
\[
\mu_D | y_D \sim N \left( \left( K_D^T V_D^{-1} - K_D + \epsilon^2 \Lambda_D^{-1} \right)^{-1} K_D^T V_D^{-1} y_D, \left( \epsilon^{-2} K_D^T V_D^{-1} K_D + \Lambda_D^{-1} \right)^{-1} \right).
\] (25)

Then, if estimates \( \hat{\mu}_0 \) for \( \nu \in \gamma \) are available, we obtain the estimate of function \( \mu \): \( \hat{\mu} = \sum_{\nu \in \gamma} \hat{\mu}_0 e_0, \nu \). Similarly, a posterior distribution on \((\mu_0, \nu \in \gamma)\) induces a posterior distribution on function \( \mu \).

In most of the paper, we consider this sequence model under the assumption that \((e_0, \nu)\) and \((\hat{e}_1, \nu)\) are known. In the case \((e_0, \nu)\) form an orthonormal eigenbasis of \(K K^T\), Koltchinskii and Lounici (2017) propose a way to estimate the corresponding eigenvectors of the covariance matrix in the corresponding finite dimensional model which we discuss in Section 6.5.2.

### 5.3 Minimax rate of convergence

In this section we define the minimax rate of convergence of estimators of \( \mu \) under model (1). This rate is usually used as a benchmark for posterior contraction rates; Ghosal et al., 2000). We consider the following smoothness classes for the true unknown function \( \mu_0 \).

**Assumption 3.** Assume that \( \mu_0 \) belongs to a smoothness class \( Q((a_\nu), A) \), with \( A > 0 \) and a nondecreasing sequence \((a_\nu)_{\nu \in \gamma} \), \( a_\nu \geq 1 \):
\[
Q((a_\nu), A) = \left\{ f = \sum_{\nu \in \gamma} e_0, \nu f_\nu : \sum_{\nu} a_\nu^2 f_\nu^2 \leq A^2 \right\}.
\] (26)

Since \( \gamma \) is isomorphic to \( \mathbb{N} \), the sequence \((a_\nu)\) is understood to be nondecreasing for some map \( \gamma \rightarrow \mathbb{N} \). For example, a Sobolev class is defined with \( e_0, \nu \) being the Fourier basis, \( \gamma = \mathbb{Z} \) and \( a_i = |i|^{\beta} \), or in terms of wavelets or vaguelettes as discussed in Section 3.2.

Define the risk of estimator \( \hat{\mu} \) of a true function \( \mu_0 \) in \( L^2 \) norm over a set of functions \( Q \) by
\[
R(\hat{\mu}, Q) = \sup_{\mu_0 \in Q} \mathbb{E}_{\mu_0} ||\hat{\mu} - \mu_0||^2.
\] (27)

**Definition 5.** \( r_c(Q) \) is the minimax rate of convergence of estimators of \( \mu \) under model (1) over class \( Q \) if \( \exists 0 < c \leq C < \infty \) that depend only on \( Q, K, \) and \( V \) such that \( c \leq \inf_{\hat{\mu}} |r_c(Q)|^{-2} R(\hat{\mu}, Q) \leq C \), where \( R(\hat{\mu}, Q) \) is defined by (27).

The minimax rate of convergence of estimating \( \mu \) in \( L^2 \) norm under model (1) in sequence space under Assumption 3 can be derived for given sequences \((\sigma_\nu), (\kappa_\nu)\) and \((a_\nu), \nu \in \gamma \), using theorem 3 in Belitser and Levit (1995) which is stated in Appendix S1 (Lemma 19).
The finite-dimensional dependent part of the model in sequence space (22) adds a term of order $\epsilon^2$ to the minimax rate as long as its covariance matrix $V_D$ has finite positive eigenvalues.

Remark 2. If the covariance matrix $V_D$ is such that $||V_D||, ||V_D^{-1}|| \in (0, \infty)$, then the part of the model (22) with dependent observations $y_D$ is stochastically dominated by a iid model with larger variances $y \sim N(\kappa_0 \mu_0, \epsilon^2 ||V_D||^2)$, $v \in D$, and stochastically dominates an iid model with smaller variances $\epsilon^2 ||V_D^{-1}||^{-2}$. Since the set $D$ is finite, in both cases these terms affect the minimax rate up to an additional $C\epsilon^2$.

The following corollary, for when a fast rate can be obtained, is a straightforward consequence of Lemma 19.

Corollary 2. Under conditions of Lemma 19 with $\tilde{\sigma}_v = \sigma_v \kappa_v^{-1}$,

1. if $\sum_{u \in Y} \kappa_u^{-2} \sigma_u^2 < \infty$ then the minimax rate of convergence is $r_\epsilon(Q) = C\epsilon$ for some $C > 0$;
2. if $\sum_{u \in Y} \kappa_u^{-2} \sigma_u^2 \asymp \log N$ for large $N$ and $Y_N$ is the set of the first $N$ elements of $Y$ (order so that sequence $(a_u)$ is nondecreasing), then the minimax rate of convergence is $r_\epsilon^2(Q) = C\epsilon^2 \log(1/\epsilon)$.

5.4 Posterior contraction rates and their optimality

Now we study conditions on the prior distribution so that the corresponding posterior distribution $\mu|Y$ contracts to the true value $\mu_0$ at the optimal rate, in the minimax sense.

5.4.1 Rate of contraction of posterior distribution

Below we present a general result that can be applied for any $a_u, \lambda_u, \kappa_u$ and $\sigma_u$ satisfying stated conditions.

Theorem 1. Consider the inverse problem (1) formulated in the sequence space under Assumption 2 with noise $W$ satisfying (19) and prior distribution (23). Assume that the true function $\mu_0$ satisfies Assumption 3. Assume also that $\min_{\nu \in D} \lambda_\nu \geq 1$ and $\|K_D^{-1}V_DK_D^{-1}\| < \infty$.

Then, for every $M \to \infty$,

$$P(\mu : ||\mu - \mu_0|| \geq Mr_\epsilon|Y) \xrightarrow{P_{\mu_0}} 0, \quad \text{as} \ \epsilon \to 0,$$

uniformly over $\mu_0$ in $Q((a_u), A)$ where $r_\epsilon$ is given by

$$r_\epsilon = \epsilon + \left[ \epsilon^2 \sum_{\nu \notin Y_\epsilon} \sigma_\nu^2 \kappa_\nu^{-2} + A^2 \sup_{\nu \in Y_\epsilon} \sigma_\nu^{-2} + \sum_{\nu \in Y_\epsilon} \lambda_\nu + \epsilon^4 \max_{\nu \in Y_\epsilon} \left[ \frac{\sigma_\nu^2}{\sigma_0^2 \kappa_\nu^2 \lambda_\nu} \right]^{2} \right]^{1/2},$$

and $Y_\epsilon = \{\nu \in Y : \sigma_\nu^2 \kappa_\nu^{-2} > \epsilon^{-2} \lambda_\nu\}$.

The first two terms in the square brackets represent the variance and squared bias, respectively, and the remaining terms, which involve prior parameters $\lambda_\nu$, represent the bias arising from the prior and the saturation term.
Remark 3. The posterior distribution can contract at rate $\epsilon$ if $\sigma^2$ is such that Assumption 8 holds and $\sum_{\alpha \in \Upsilon} \sigma^2_{\alpha} \kappa^{-2} \leq C < \infty$ which is the minimax rate (Corollary 2), under the appropriate choice of prior parameters ($\lambda_\alpha$).

Remark 4. There is a phenomenon known as saturation (Agapiou & Mathé, 2018) that constrains the posterior contraction rate for an undersmoothing prior. In the above theorem, this is described by the term $\max_{\alpha \not\in \Upsilon} \left[ \frac{\sigma^2_{\alpha}}{a_i \kappa^2 \lambda_i} \right]$. It will be illustrated in the example below.

Example 1. Motivated by an example in Agapiou et al. (2013) with covariance operator of the type $(\text{KKT})^{-1} + M_x^{-1}$ where $M_x$ is a nonlinear operator, we consider a particular linear case with $V = (\text{KKT})^a$, $a \geq 0$. Here $(e_{1,i})$ are the eigenbasis of $\text{KKT}$ and $V$, with eigenvalues $\kappa_i^2$ and $\sigma^2 = \sigma_i^2 a_i^2$, respectively. In this case,

$$r_\epsilon = \left[ \epsilon^2 \sum_{i \leq i_c} \kappa_i^{-2+2a_i} + a_i^{-2} + \sum_{i > i_c} \lambda_i + \epsilon^4 \max_{i \leq i_c} \left[ \frac{1}{a_i \kappa_i^{2(1-a)} \lambda_i} \right] \right]^{1/2},$$

where $i_c = \max \left\{ i : \kappa_i^{-2(1-a)} \leq \epsilon^{-2} \lambda_i \right\}$. In particular, if $a = 1$, that is, if $V = \text{KKT}$, then the posterior contraction rate coincides with the rate of the direct problem. This corresponds to the model of the type $Y = K(\mu + \epsilon W)$ where $W$ is white noise, that is, when the error occurs before operator $K$ is applied.

In the next section we assume that the variances $\sigma^2_{\alpha}$ are unknown, and their plug-in estimator is available.

5.4.2 Rate of contraction of posterior distribution with unknown variances $\sigma^2_{\alpha}$

When the $\sigma^2_{\alpha}$ are unknown, their plug-in estimator is used to conduct inference about $\mu$. We investigate how this affects the contraction rate of the posterior distribution of $\mu$. Theorem 1 will be used to address this question.

Suppose we have a plug-in estimator $\hat{\sigma}^2_{\alpha}$ of the error variances $\sigma^2_{\alpha}$, $\alpha \in \Upsilon$ under the model (22). We consider the case where $\hat{\sigma}^2_{\alpha}$ are consistent estimators of $\sigma^2_{\alpha}$ for all $\alpha$ and are independent of $Y$ used to estimate $(\mu_\alpha)$. Plugging in an estimator can be thought of as having an informative prior distribution on $\sigma^2_{\alpha}$ that is a point mass at $\hat{\sigma}^2_{\alpha}$. If $\sigma^2_{\alpha}$ are the eigenvalues of $V$, Koltchinskii and Lounici (2017) proposed a way to estimate the eigenfunctions of $V$ which we apply for a particular form of $V$ under repeated observations in Section 6.5.2.

These assumptions are satisfied, for instance, when variances are estimated from a different study; when the data is split into two independent parts: one is used to estimate $(\sigma_\alpha)$ and the other one to estimate $(\mu_\alpha)$; or when there are repeated observations with Gaussian error, (hence the sample mean and the sample variance are independent), convolution operators (Cavalier & Hengartner, 2005) and statistical inference in econometric problems with instruments (Florens & Simoni, 2016). Another application is an additive error in the operator (Hoffmann & Reiss, 2008) that we consider below.

We make the following assumption of consistency of $\hat{\sigma}^2_{\alpha}$ that holds relative to $\sigma^2_{\alpha}$.
**Assumption 4.** Assume that the estimated variances \( \{ \hat{\sigma}^2_v \}_{v \in \mathcal{Y}_N} \) for some \( \mathcal{Y}_N \subset \mathcal{Y} \) of size \( |\mathcal{Y}_N| = N \) are independent of \( (y_v, v \in \mathcal{Y}) \), and there exists a constant \( c_0 \) such that \( c_0 \epsilon \sigma \leq 1/2 \) and

\[
P \left( \left| \frac{\hat{\sigma}^2_v}{\sigma^2_v} - 1 \right| \leq c_0 \epsilon \sigma, \ i = 1, 2, \ldots, N \right) \rightarrow 1, \ \text{as} \ \epsilon \sigma \rightarrow 0 \ \text{and} \ N \rightarrow \infty.
\]

Note that the rate \( \epsilon \sigma \) may depend on \( N \) but it does not need to be known.

**Theorem 2.** Consider the inverse problem (1) formulated in the sequence space under Assumption 2, with noise \( W \) satisfying (19) and prior distribution (23) with \( \lambda_v = 0 \) for \( v \not\in \mathcal{Y}_N \). We assume that \( \min_{v \in \mathcal{D}} \lambda_v \geq 1 \) and \( \| K_D^{-1} V_D K_D^{-1} \| < \infty \). Assume also that the true function \( \mu_0 \) satisfies Assumption 3.

Suppose that \( \sigma_v \) are unknown, but we observe \( (\hat{\sigma}_v, v \in \mathcal{Y}_N) \) satisfying Assumption 4.

Then, the posterior distribution of \( \mu \) given \( Y \), with plugged in \( (\hat{\sigma}_v) \) instead of \( \sigma_v \), is such that for every \( M \rightarrow \infty \),

\[
\sup_{\mu_0 \in \mathcal{P}_M(A)} \mathbb{P} \left( \{ \mu : \| \mu - \mu_0 \| \geq M r_{\epsilon} |Y, \hat{Y}, N) \right) \xrightarrow{\mathbb{P}} 0, \ \text{as} \ \epsilon \rightarrow 0, \epsilon \sigma \rightarrow 0 \ \text{and} \ N \rightarrow \infty,
\]

where \( r_{\epsilon} \) is given by

\[
r_{\epsilon}^2 = \epsilon^2 \sum_{v \in \mathcal{Y}_N \& \ \epsilon \not\in \mathcal{Y}_N} \sigma_v^2 \lambda_v^{-2} + \sum_{v \in \mathcal{Y}_N \& \ \epsilon \in \mathcal{Y}_N} \lambda_v + \epsilon^2 \sum_{v \in \mathcal{Y}_N \& \ \epsilon \not\in \mathcal{Y}_N} \sigma_v^{-2} \lambda_v^{-2} \\
+ \epsilon^4 \max_{v \in \mathcal{Y}_N \& \ \epsilon \not\in \mathcal{Y}_N} \left[ \frac{\sigma_v^2 a_v^{-1}}{\lambda_v^{-2}} \right]^2 + \sup_{v \in \mathcal{Y}_N \& \ \epsilon \in \mathcal{Y}_N} a_v^{-2},
\]

where \( \mathcal{Y}_{\epsilon, \pm} = \{ v : \sigma_v^2 / (\lambda_v \kappa_v^2) > (1 \pm c_0 \epsilon \sigma)^{-1} \epsilon^{-2} \} \).

In addition to the terms in the rate with known \( V \) in Theorem 1 where \( \mathcal{Y}_\epsilon \) replaced by either \( \mathcal{Y}_{\epsilon, \pm} \) or \( \overline{\mathcal{Y}}_N \), there is an additional term (third term in the rate above) that arises from the posterior variance.

We apply this theorem to the case of repeated observations in Section 5.4.4, where we estimate \( \sigma_v \) and the power for polynomially decreasing \( \sigma_v^2 \) (Section 6.6). We also discuss the case where the difference between \( \delta^2_v \) and \( \sigma_v^2 \) is simultaneously bounded for all \( v \in \mathcal{Y} \) (Section E.1 in Appendix S1).

Next we discuss the problem of error in the forward operator \( K \), that can be reformulated as a problem of unknown variances.

### 5.4.3 Error in operator

Now we consider the case forward operator \( K \) is observed with error, under model (5) introduced in Section 2.4. Under the settings of Section 5.1, this problem can be written in the sequence space as (22) with

\[
\hat{k}_v = \kappa_v + \delta \xi_v, \ \ \xi_v \sim N(0, 1) \ \text{iid for} \ v \in \mathcal{Y}_N, \ \ (28)
\]

and independently of \( y_v \), for some set \( \mathcal{Y}_N \) such that \( D \subset \mathcal{Y}_N \subset \mathcal{Y} \) of size \( |\mathcal{Y}_N| = N \).

This observed value of \( \hat{k}_v \) can be plugged in into the posterior distribution (24) and (25) for \( v \in \mathcal{Y}_N \), and we can study whether the contraction rate of the plug-in posterior distribution is affected by plugging in \( \hat{k}_v \). Note that it is common to assume that the level of ill-posedness of \( K \) is known as well as the error level \( \delta \) (Hoffmann & Reiss, 2008).
Theorem 3. Consider the inverse problem (1) formulated in the sequence space under Assumption 2, with noise \( W \) satisfying (19). We assume that \( \min_{v \in D} \lambda_v \geq 1 \) and \( \|K^{-1}V K^{-1}\| < \infty \). Assume also that the true function \( \mu_0 \) satisfies Assumption 3.

Suppose that \( \kappa_0 \) tend to 0 and are unknown, but we observe \( (\hat{k}_v, v \in \mathcal{V} \) from the model (28) where \( \mathcal{V} \subseteq \{v \in \mathcal{V} : \kappa_v \geq b^{-1}\delta\} \) and \( \delta = \delta \sqrt{2 \log(B)} \) for some \( b \in (0, 0.5] \) such that \( B \to +\infty, |\mathcal{V}|/B \leq 1 \) and \( b^{-1}\delta \to 0 \) as \( \delta, e \to 0 \).

Take prior distribution (23) with \( \lambda_0 = 0 \) for \( v \notin \mathcal{V} \).

Then, the posterior distribution of \( \mu \) given \( Y \), with plugged in \( \hat{v}_v \) instead of \( \kappa_v \) for \( v \in \mathcal{V} \), is such that for every \( M \to \infty \),

\[
\sup_{\mu_0 \in \mathcal{P}(\mathcal{V}(a_\delta), A)} \mathbb{P}(\mu : ||\mu - \mu_0|| \geq M r_\varepsilon | Y, \hat{v}) \xrightarrow{P_{m,K}} 0, \text{ as } \varepsilon \to 0, \delta \to 0,
\]

where \( r_\varepsilon \) is given by

\[
r_\varepsilon^2 = e^2 \sum_{v \in \mathcal{V}} \sigma_v^2 \kappa_v^{-2} + A^2 \sup_{v \in \mathcal{V} \cup \mathcal{Y}(1)} a_v^{-2} \sum_{v \in \mathcal{V} \cap \mathcal{Y}_N} \lambda_v
\]

\[
+ A^2 e^4 \sup_{v \in \mathcal{V}} a_v^{-2} \sigma_v^4 \lambda_v^{-2}[\kappa_v^{-4} + 2 \delta^2 e^{-4} \sup_{v \in \mathcal{V} \cap \mathcal{Y}_N} a_v^{-2} \kappa_v^{-2} \lambda_v^{-2} \sigma_v^{-4} + A^2 \delta^2 \sup_{v \in \mathcal{V} \cap \mathcal{Y}_N} a_v^{-2} \kappa_v^{-2},
\]

where \( \mathcal{Y}_{\varepsilon, \pm} = \{v \in \mathcal{V} : \sigma_v^2 /[\lambda_v \kappa_v^{-4}] > (1 + b)^2 e^{-2}\} \).

Note that the first four terms in the squared rate are the same as in the posterior contraction rate without plugged in \( \hat{k}_v \), but over a slightly larger sets of indices. The last term includes the squared bias of \( K\mu \) for smaller indices and the penultimate term is a variant of the saturation term for \( K\mu \).

For an ill-posed problem where \( \kappa_0 \) tends to 0, the size of the set \( \mathcal{V} \) tends to infinity. As often only an asymptotic behavior of \( \kappa_0 \) is known, for example, \( \kappa_0 \asymp 2^b \) as assumed in Hoffmann and Reiss (2008), \( \mathcal{V} \) can be defined as \( \{C2^b \geq b^{-1}\delta\} \) for large enough \( C \). Reasonable choices of \( b \) and \( B \) are \( b = 0.5 \) and \( \log B \asymp \log N \), as long as the required conditions hold. We consider an application of this result to fractional noise in Section 4.4 and to geometric sequences in Section 6.7.

5.4.4 | Repeated observations

Now we suppose that we have \( m \) independent replicates of the original model (1) with \( \varepsilon \) replaced by \( \varepsilon_0 \), and hence, under the assumptions of Section 5.1, we can write the corresponding model (22) as

\[
Y_{v,j} \sim N(\kappa_0 \mu_v, \varepsilon_0^2 \sigma_v^2), \quad v \in \mathcal{V}, \quad Y_{v,j} \sim N(K \mu_D, \varepsilon_0^2 V_D), \quad j = 1, \ldots, m,
\]

independently. Consequently, for each \( v \in \mathcal{V} \), the sample mean \( \overline{Y}_v \) and the sample variance \( s_v^2 \) are

\[
\overline{Y}_v := \frac{1}{m} \sum_{j=1}^m Y_{v,j} \sim N(\kappa_0 \mu_v, \varepsilon_0^2 \sigma_v^2 / m),
\]

\[
s_v^2 := \frac{1}{\varepsilon_0^2 (m - 1)} \sum_{j=1}^m (Y_{v,j} - \overline{Y}_v)^2 \sim \frac{\sigma_v^2}{m - 1} \chi^2_{m-1},
\]
independently for all \( v \in Y_I \). Also, independently from \( \overline{Y}_o \) and \( s_0^2 \) for \( v \in Y_I \),

\[
\overline{Y}_D := \frac{1}{m} \sum_{j=1}^{m} Y_{Dj} \sim N \left(K_D \mu_D, \epsilon_0^2 V_D/m \right),
\]

\[
S_0^2 := \frac{1}{\epsilon_0^2 (m-1)} \sum_{j=1}^{m} \left(Y_{Dj} - \overline{Y}_D\right) \left(Y_{Dj} - \overline{Y}_D\right)^T \sim \frac{1}{m-1} \text{ Wishart}_D(m-1, V_D).
\]

where a \( D \)-dimensional Wishart distribution of a random positive definite matrix, parameterized by a positive definite matrix \( \Sigma \) and shape parameter \( a > 0 \), has density \( f_D(S; a, V) = |S|^{(a-D-1)/2} e^{-\frac{1}{2} \text{trace}(S^T V)} / 2^{-aD/2} |V|^{-a/2} / \Gamma_D(a/2) \) where \( \Gamma_D \) is a multivariate Gamma function.

Now we study when simultaneous asymptotic consistency of the following estimator of \((\sigma_i^2)_{i \in Y} \) holds for large \( m \) with high probability, given a set \( Y_N \) such that \( D \subset Y_N \subset Y \), \( |Y_N| = N \) and

\[
\hat{\sigma}^2 = s_0^2 I(v \in Y_N).
\]  

**Proposition 5.** Assume that we have \( m \) independent observations of inverse problem (1) that can be formulated in the sequence space under Assumption 2, with noise \( W \) satisfying (19), with repeated observations (29), with prior distribution (23), and the true function \( \mu_0 \) satisfying Assumptions 3.

Consider the estimator of \((\sigma_i^2) \) defined by (32) with \( N \) satisfying \( \log N = o(m) \). Then, for every \( M_0 \to \infty \),

\[
P \left( \mu : ||\mu - \mu_0|| \geq M_0 r_{\text{plugin}}|Y, (\hat{\sigma}_0^2) \right) \xrightarrow{P_{\mu_0|Y}} 0, \ \text{as} \ \epsilon \to 0,
\]

uniformly over \( \mu_0 \in Q((a_0), A) \), where \( r_{\text{plugin}} \) is the rate given in Theorem 2 with \( \epsilon^2 = \epsilon_0^2 / m \).

If also \( Y_N \subseteq Y \setminus Y_\epsilon \) and \( Y_{\epsilon, \pm} \approx Y_\epsilon \) (the latter holds as long as \( \epsilon \to 0 \) and \( m \to \infty \)), then \( r_{\text{plugin}} = r_\epsilon \) stated in Theorem 1, that is, it is not affected by the plug-in estimator.

The condition on \( N, \log N = o(m) \), is required to have simultaneous consistency for \( N \) estimators given their precision \( m \), and the other condition \( Y_N \subseteq Y \setminus Y_\epsilon \) ensures the plugin rate of posterior contraction of \( \mu \) coincides with the contraction rate for known \( (\sigma_i^0) \).

In the next section we illustrate how the results of this section apply to mildly ill-posed problems with correlated Gaussian noise.

## 6 MILDLY ILL-POSED INVERSE PROBLEMS WITH HETEROGENEOUS GAUSSIAN NOISE

In this section we apply the general results with known and unknown covariance operators of Gaussian noise to mildly ill-posed inverse problems, under the following assumptions.

### 6.1 Assumptions

In this section, we assume that the setting of Section 5.1 applies, leading to the sequence space problem described in Section 5.2, under mildly ill-posed inverse problems and other sequences decaying polynomially. We set \( Y = N \), taking an appropriate mapping, such that sequence \( a_i \) in the definition of the smoothness class \( Q((a_i), A) \) is nonincreasing.

**Assumption 5.** Assume that \((\sigma_i^2) \) satisfy \( C_2^{-1} i^\gamma \leq \sigma_i \leq C_2 i^\gamma \) for some \( \gamma \in \mathbb{R} \) and \( C_2 \geq 1 \).
Assumption 6. Assume that \( (a_i) \) satisfy \( C_3^{-1} i^\beta \leq a_i \leq C_3 i^\beta \) for some \( \beta > 0 \) and \( C_3 \geq 1 \).

This assumption corresponds to a generalized Sobolev class. If \( (e_{0,i}) \) is the Fourier basis and \( C_3 = 1 \), this is the definition of a standard Sobolev class. If \( (e_{0,i}) \) are wavelets then this also defines the standard Sobolev class for the mapping \( (j, k) \) to \( i = 2^j + k \) and for \( C_3 = 2^\beta \), given sufficiently regular wavelets.

Assumption 7. Eigenvalues \( (\lambda_i) \) satisfy \( \lambda_i = \tau \epsilon^{2 - i^{-1 - 2\alpha}} \) for some \( \alpha > 0 \) and \( \tau \epsilon > 0 \), such that \( e^{-2\tau \epsilon^2} \to \infty \) as \( \epsilon \to 0 \). Denote \( \tau = \tau \epsilon^2 \).

The latter assumption implies that a priori we assume \( \mu \in Q((i^\alpha'), R) = S^\alpha(R) \) almost surely, for any \( \alpha' < \alpha \) and large enough \( R \), for a fixed \( \epsilon \).

Assumption 8. \( ||V_D||, ||V_D^{-1}|| \in (0, \infty) \) and there exist constants \( C_{D,1}, C_{D,2} > 0 \) independent of \( \epsilon \) such that

\[
\left\|(K_D^T V_D^{-1} K_D + \epsilon^2 A_D^{-1})^{-1} A_D^{-1}\right\| \leq C_{D,1}, \quad \text{trace}\left((K_D^T V_D^{-1} K_D + \epsilon^2 A_D^{-1})^{-1}\right) \leq C_{D,2}.
\]

By Lemma 21, this assumption ensures that the terms in the posterior contraction rate that correspond to the dependent observations are of order \( \epsilon \).

6.2 Minimax rate of convergence

The minimax rate for model (1) with the considered parameter sequences is stated in the following proposition.

Proposition 6. Consider the inverse problem (1) formulated in the sequence space under Assumption 2 with noise \( \mathcal{W} \) following (19). Assume that Assumption 5 holds, that matrix \( V_D \) has finite positive eigenvalues and that the true function \( \mu_0 \) satisfies Assumptions 3 and 6.

Then, the minimax rate of convergence of estimating \( \mu \) is given by

\[
r_\epsilon^* = \begin{cases} 
\epsilon^{\frac{2\beta}{1+2(p+\gamma)}} & \text{if } \gamma > -p - 1/2, \\
\epsilon & \text{if } \gamma = -p - 1/2, \\
\epsilon \left(\log |\epsilon|\right)^{1/2} & \text{if } -p/2 - p - 1/2 < \gamma < -p - 1/2.
\end{cases}
\]

It follows from Lemma 19 together with Remark 2.

This result implies that for a heterogeneous variance the degree of ill-posedness for (1) changes, in particular to \( \tilde{p} = p + \gamma \) if \( p + \gamma > -1/2 \). For \( p + \gamma = 0 \), the rate coincides with the minimax rate of the direct problem. If \( p + \gamma \leq -1/2 \), the problem becomes self-regularized, that is, the rate of convergence \( \epsilon \) can be achieved (up to a log factor in the case \( p + \gamma = -1/2 \)), provided \( \gamma + p + 1/2 + \beta/2 > 0 \). According to Belitser and Levit (1995), under this constraint the minimax rate of any estimator coincides with the minimax rate of a linear estimator. Since we consider Gaussian errors and Gaussian priors, such constraint is appropriate.

6.3 Posterior contraction rates

Having found the minimax rates, we can now discuss the contraction rates achieved by the posterior distribution under the considered Bayesian model. Note these rates also apply when the problem is self-regularized, that is, \( p + \gamma \leq -1/2 \).
Theorem 4. Consider the inverse problem (1) formulated in the sequence space under Assumption 2 with noise \( W \) satisfying (19) and Assumption 5, with prior distribution (23) under Assumption 7, and the true function \( \mu \) satisfying Assumptions 3 and 6. Let Assumption 8 hold.

Then, for every \( M \to \infty \), \( E_{\mu_0} (\mu : ||\mu - \mu_0|| \geq M \rho_e \mid Y) \to 0 \) as \( \epsilon \to 0 \) uniformly over \( \mu_0 \) in \( S^\beta(A) \) where

\[
\tau_e^r := \begin{cases} 
(e^{2} \tau_e^{-2})^{\frac{\beta}{1+2p+2p_{\gamma}+2p_{\gamma}}} + \tau_e (e^{2} \tau_e^{-2})^{\frac{\beta}{1+2p+2p_{\gamma}}} , & \text{if } \gamma > -p - 1/2, \\
(e^{2} \tau_e^{-2})^{\frac{\beta}{1+2p+2p_{\gamma}}} + \epsilon \log(e^{-1} \tau_e) / 2 , & \text{if } \gamma = -p - 1/2, \\
(e^{2} \tau_e^{-2})^{\frac{\beta}{1+2p+2p_{\gamma}}} + \epsilon , & \text{if } -p - 1/2 - \alpha < \gamma < -p - 1/2.
\end{cases}
\]

The assumption of monotonicity of \( \sigma_i^2 /[\lambda_i \kappa_i^2] \approx i^{1+2\gamma+2\alpha+2p} \) for large \( i \) from Theorem 1 is reflected in the condition \(-p - 1/2 - \alpha < \gamma \). Recall that parameters \( p \) and \( \gamma \) are assumed known and given by the problem, as well as the smoothness parameter \( \beta \). Parameters of the prior \( \alpha \) and \( \tau_e \) can be chosen in some cases so that the posterior contracts at the optimal rate given in Proposition 6.

Corollary 3. Let assumptions of Theorem 3 hold. The rate of contraction of the posterior given in Theorem 3 matches the minimax rate of convergence, for the following \( \alpha \) and \( \tau_e = \tau^{1/2} \):

1. \( \tau_e = \text{const} \in (0, \infty) \) and

\[
\begin{cases} 
\alpha = \beta , & \text{if } \gamma > -\frac{1+2p}{2}, \\
\alpha \leq \beta , & \text{if } \gamma = -\frac{1+2p}{2}, \\
\alpha \leq \beta , & \text{if } -\frac{1+2p}{2} - \alpha < \gamma < -\frac{1+2p}{2}.
\end{cases}
\]

2. for \( \tau_e \) depending on \( \epsilon \):

\[
\begin{cases} 
\alpha \geq \beta / 2 - (1/2 + p + \gamma) , & \tau_e = Ce^{\frac{\beta (p + \gamma)}{1+2p+2p_{\gamma}}}, & \text{if } \gamma > -\frac{1+2p}{2}, \\
en^{-\beta} \geq \tau_e \geq Ce^{\min(1/2, 1-\frac{1+2p+2p_{\gamma}}{2p_{\gamma}})} [\log(e^{-1})]^{-0.5 \max(1, 2, \alpha / \beta)} , & \text{if } \gamma = -\frac{1+2p}{2}, \\
\tau_e \geq Ce^{\min(1/2, 1-\frac{1+2p+2p_{\gamma}}{2p_{\gamma}})} \epsilon^{2p_{\gamma}} , & \text{if } -\frac{1+2p}{2} - \alpha < \gamma < -\frac{1+2p}{2}.
\end{cases}
\]

Observe that when \( \gamma > -p - 1/2 \), the rates obtained are similar to the white noise case, albeit with a different degree of ill-posedness \( \tilde{p} = p + \gamma \). When \( p + \gamma < 0 \), the fastest rate of contraction coincides with the minimax rate of convergence of the direct problem, that is, the model self-regularizes, and it can achieved when we undersmooth a priori.

If \( \alpha < \beta / 2 - (1/2 + p + \gamma) \) and \( \gamma > -p - 1/2 \), that is, if we undersmooth too much a priori, then the minimax rate cannot be achieved for any \( \tau \). When \( \gamma = 0 \), this coincides with the findings of Knapik et al. (2011). This is known as saturation (Agapiou & Mathé, 2018). However, in the self-regularizing case \(-p - 1/2 - \alpha < \gamma \) the optimal rate can be achieved if the appropriate prior scaling is used.

Remark 5. Note that the case \( p + \gamma + 1/2 > 0 \) for a given \( \alpha \geq \beta / 2 - (1/2 + p + \gamma) \), where \( \alpha > 0 \), the value of \( \tau_e \) that leads to the minimax rate is such that the cutoff level \( i_e \sim \{\epsilon^{-2}\}^{\frac{1+2p+2p_{\gamma}}{1+2p+2p_{\gamma}}} \) is independent of \( \alpha \) and is the cutoff level corresponding to the minimax optimal projection estimator (projecting on the first \( i_e \) components).
Florens and Simoni (2016) consider the case of our setup with \( p + \gamma > 0 \) and \( |\mu_{0,i}| \approx i^{-p-1/2} \). In Agapiou et al. (2013), the rate of contraction in this setting is given only for \( \beta > \alpha + 1/2 \) (in our notation) by

\[
e^{\frac{\beta(p+\gamma+2r+1)}{2(1+2r)+12p+2+2\beta(p+\gamma+2r+1)}} , \quad \forall \delta > 0.
\]

(where \( \gamma_A = \beta/(\alpha + 1/2) \), \( \xi_A = p/(2\alpha + 1) \), \( s_{0,A} = 1/(2\alpha + 1) \), \( \Delta_A = (p + \gamma)/(\alpha + 1/2) + 1 \), with subscript \( A \) referring to parameters in Agapiou et al., 2013). In particular, the authors’ assumption \( \Delta_A > 2s_{0,A} \) is equivalent to assumption \( p + \gamma + \alpha + 1/2 > 1 \) in our notation which is stronger than our assumption \( p + \gamma + \alpha + 1/2 > 0 \). In fact, under the latter assumption, it is not possible to achieve the minimax optimal contraction rate for \( p + \gamma > -1/2 \) with constant \( \tau_c \), which recall is achieved when \( \alpha = \beta \). Also, the authors refer to the case \( p + \gamma < 0 \) as self-regularizing, with no regularization being necessary which we do not find in case \( p + \gamma \in (-1/2, 0) \); also their rate in this case is not faster unless \( \delta(1/2 + \alpha) = -(1/2 + p + \gamma) - 0.5\beta \wedge (p + \gamma + 2\alpha + 1) \) which is possible only if \( \alpha < -1.5(p + \gamma + 1/2) - 1/4 \) and \( \beta < -2(1/2 + p + \gamma) \).

**Example 2.** Consider case \( V = (KK^T)^p \). For mildly ill-posed inverse problems where \( \kappa_i \approx i^{-p} \), where \( p > 0 \), we have \( \gamma = -pa \) and \( p + \gamma = p(1 - a) \). Take the prior with \( \lambda_i = \tau_c^2 i^{-2\alpha - 1} \) and the parameters as specified in Corollary 3, so that the posterior contraction rate coincides with the minimax rate. Then, the rate is \( e^{\sqrt{\log 1/\epsilon}} \) for \( a \in (0, 1 + 0.5/p) \) the rate is \( [e^2]^{\beta/(1+2\beta+2p(1-a))} \) and for \( a > 1 + 0.5/p \) the rate is \( e \). Note that the model \( Y = K(\mu + \epsilon W) \) with white noise \( W \) corresponds to \( a = 1 \), where the rate of posterior contraction coincides with the minimax rate of the direct problem.

### 6.4 Contraction rate with plugged-in \((\sigma_i)\)

In this section, we investigate how a plug-in estimator of \((\sigma_i)\) affects the rate of contraction for geometric sequences.

**Proposition 7.** Consider the inverse problem (1) formulated in the sequence space under Assumption 2 with noise \( W \) satisfying (19) and Assumption 5 with \( \gamma < 0 \). Consider the prior distribution (23) under Assumption 7 and the true function \( \mu_0 \) satisfying Assumptions 3 and 6. Let Assumption 8 hold.

Suppose that \((\sigma_i)\) are unknown but their estimators \((\hat{\sigma}_i)\) are available.

1. Assume that estimator \((\hat{\sigma}_i)\) satisfies Assumption 10 (absolute bound) in Appendix S1.
   (a) If \( e_{\sigma} < C [e^2 \tau_c^{-2}]^{-\gamma/(\alpha+1/2+p+\gamma)} \) then the rate of contraction is not affected by using a plug-in estimator of \((\sigma_i)\), that is, it coincides with the rate given in Theorem 3, up to a constant.
   (b) If \( e_{\sigma} \geq C [e^2 \tau_c^{-2}]^{-\gamma/(\alpha+1/2+p+\gamma)} \), then the contraction rate of the posterior distribution is given by

\[
r_{\text{plugin}}^2 = e^2 (\log e_{\sigma}^{-1})^{[1+2(\gamma+\gamma+\gamma+1)]} + \tau_c^2 [e_{\sigma}^{-1} e^{-2 \tau_c^{-2}} - 2a/(1+2\alpha+2p)] + [e_{\sigma}^{-1} e^{-2 \tau_c^{-2}}]^{-2/\gamma} + e_{\sigma}^{1/2} \tau_c^{-4} e_{\sigma}^{-2} [e_{\sigma}^{-1}(1+2\alpha+2p)]^{-1/\gamma}.
\]

2. Assume that estimator \((\hat{\sigma}_i)\) satisfies Assumption 4 (relative bound). The contraction rate of the posterior distribution is given by
\[ r^2_{\text{plugin}} = e^{2\min(N, i_e)}[\log \min(N, i_e)]^{(2p+2\gamma+1)} \] 
\[ + \min(N, i_e)^{-2\beta} + e^{4} \tau_e^2 \min(N, i_e)^{(2(2p+2\gamma+2\alpha+1)-\beta)} \] 
\[ + \tau_e^{-2} \max(N^{-2\alpha-2p-2\gamma}, i_e^{-2\alpha-2p-2\gamma}) I(i_e < N) \] 
\[ I(i_e \leq N) \] 
with \( i_{e, \pm} = i_e = C[\tau_e e^{-2}]^{1/(2\gamma+2p+2\alpha+1)} \). If \( N \geq C_{\ell_e} \) for some positive \( C > 0 \), then the rate is the same as if \( \sigma_i \)'s were known; if \( N / \ell_e \rightarrow 0 \) then the rate is affected by the plug-in.

In the case the bound is absolute, if the error \( \epsilon_\sigma \) of estimating \( \sigma_i^2 \) is small enough, then the rate of estimation of \( \mu \) is not affected. In the case of relative bound, as long as \( N \geq \ell_e \), the rate of estimation is not affected.

### 6.5 Repeated observations

#### 6.5.1 Posterior contraction rate with plugged in sample variances

Now we assume that we have \( m \) repeated observations from model (1) as described in Section 5.4.4, and that we plug in \( \hat{\sigma}_i^2 = s_i^2 \) for \( i = 1, \ldots, N \). Then, under Assumptions 5, 7, 6, and 8, Proposition 7 states that as long as \( \log N = o(m) \) and \( N \geq C[\tau_e e^{-2}]^{1/(2\gamma+2p+2\alpha+1)} \), the posterior rate of contraction is the same as if we used the true values of \( \sigma_i^2 \). Such \( N \) exists if \( \log(\tau_e e^{-2}) = o(m) \), which is generally a mild constraint.

We have also applied Theorem 23 to the model with repeated observations where \( \hat{\sigma}_i \) are estimated with an absolute error (Assumption 10) given in Bochkina and Rodrigues; however, in that setting the plug-in effect can lead to a suboptimal rate for some \( \beta \).

We will discuss the choice of \( N \) in this setting in practice using the empirical Bayes approach.

#### 6.5.2 Estimation of the eigenfunctions of \( V \)

Here we assume that \( e_{t,i} = \phi_{ti} \), that is, the eigenfunctions of operator \( V_i \), and discuss their estimation using repeated observations. We apply the results of Koltchinskii and Lounici (2017) to evaluate the number of eigenfunctions of \( V \) that are possible to estimate consistently, and hence verify whether the posterior distribution can achieve the optimal rate in the minimax sense for repeated observations where the eigenfunctions are unknown. Their conditions assume that the eigenvalues are decreasing, and rely on \( \text{trace}(V) \) being finite, so for \( \sigma_i^2 \approx i^{2\gamma} \) these conditions hold only if \( \gamma \geq -1/2 \) which is the case we consider here.

**Lemma 2.** Recall that \( V = \sum_{i=m}^{\infty} \sigma_i^2 \phi_i \phi_i^T \) and denote the projection matrices by \( P_i = \phi_i \phi_i^T \). In the repeated observations model (29), denote the sample estimator of the covariance matrix by \( \hat{V} \) and the corresponding projection matrices by \( \hat{P}_i \).

Assume that \( \sigma_i^2 \approx i^{2\gamma} \) for some \( \gamma < -1/2 \). Then, \( \hat{P}_r \) are consistent estimators of \( P_r \) simultaneously for \( r = 1, \ldots, N \) with high probability if \( N = o \left( m^{1/(1/2-4\gamma)} \right) \).

Note that for a fBM with Hurst exponent \( H, \gamma = 1/2 - H \in (-1/2, 1/4) \) hence this theorem does not apply to it (see below for a discussion of the associated known series expansions). In other settings, such as a nonparametric regression with a finite-dimensional approximation of \( V \) it may be possible to estimate the eigenfunctions for other values of \( \gamma \) but this problem is beyond the scope of this paper.
Recall that for the posterior distribution to contract at the optimal rate, we need \( N \geq i_c = (e^{-2\tau_c^2})^{1/(2p+2\gamma+2a+1)} \). For \( \alpha \geq \max(\beta - 1/2, \beta/2 - (1/2 + p + \gamma)), p + \gamma + 1/2 + \alpha > 0, \gamma < -1/2 \) and \( \tau_c \) given in Corollary 3, the condition of the lemma holds if

\[
m \gg [e^{-2}]^{(1/2-4\gamma)/(2p+2\gamma+2\alpha+1)},
\]

that is, under this condition estimators of the first \( N \geq i_c \) eigenvectors are consistent and the corresponding posterior contraction rate is optimal.

### 6.6 Estimated \( \gamma \)

Assume that for large \( i, \sigma_i = Cl_i(1 + o(1)) \). Therefore, we can consider \( \sigma_i = Cl_i \) for all \( i \geq i_0 \) for some \( i_0 \geq 1 \). Suppose \( \gamma \) and \( C \) are estimated independently of \( (Y_i)_{i=1}^N \), and that with high probability \( \hat{C}/C - 1 \in [-c_{\max}, c_{\max}] \) and \( \hat{\gamma} - \gamma \in [-z_{\max}, z_{\max}] \) for some \( c_{\max}, z_{\max} > 0 \).

This implies that, with the same high probability, for all \( i = i_0, \ldots, N \),

\[
|\hat{\gamma}^2 - \gamma^2| = \max\left[c_{\max}^2N^{-2z_{\max}} - 1, 1 - c_{\max}^{-1}N^{-2z_{\max}}\right] = : \varepsilon_\sigma.
\]

If \( c_{\max} \to 1 \) and \( z_{\max} \log N \to 0 \) then the upper bound tends to 0 for all \( \gamma \in \mathbb{R} \), that is, Assumption 4 holds, as long as we have an alternative consistent estimator of \( \sigma_i \) for \( i < i_0 \) or if \( i_0 = 1 \). Such bound holds also for the absolute bound (Assumption 10 in Appendix S1); however, we found that the conditions for the corresponding plug in posterior distribution to contract at the optimal rate are stronger, so we do not use it here.

If \( c_{\max} \) and \( z_{\max} \) are known, then choosing \( N \geq c_{\max} \hat{C}[\tau_c e^{-2}]^{1/(2[\hat{\gamma} - z_{\max}] + 2p + 2\alpha + 1)} \) implies that the rate of contraction of the posterior of \( \mu \) is not affected, with high probability. Now we investigate what the expressions for \( c_{\max} \) and \( z_{\max} \) are under repeated observations.

**Example 3.** In the setting of repeated observations and \( \sigma_i^2 = Cl_i^2 \) for \( i \geq i_0 \), we can use \( \hat{C} = \hat{C}_{i_0} \) and for some subset \( I_N \subseteq \{i_0, i_0 + 1, \ldots, N\} \),

\[
\hat{\gamma} = \frac{1}{2|I_N|} \sum_{i \in I_N} \log\left(\frac{s_i^2}{s_{i_0}^2}\right) / \log(i/i_0).
\]

Now we derive confidence intervals for \( C \) and \( \gamma \) when \( m \) is large, and hence the expression for relative bound \( \varepsilon_\sigma \). The ratio \( \frac{s_i^2}{[i/i_0]^{2\gamma} s_{i_0}^2} \) has distribution \( F(m, m) \) hence, approximately for large \( m \),

\[
\log\left(\frac{s_i^2}{s_{i_0}^2}\right) - 2\gamma \log(i/i_0) \sim N(0, 1/m),
\]

implying that

\[
\hat{\gamma} - \gamma \sim N\left(0, \frac{1}{4m|I_N|} \sum_{i \in I_N} \frac{1}{[\log(i/i_0)]^2}\right).
\]

This suggests that averaging over a small number of \( i \) close to \( N \) (or even a single \( i = N \)) leads to an estimator with the smallest variance of order \( [m \log N]^{-1} \), e.g. \( I_N = \{N\} \) or \( I_N = \{N - k : N\} \) for a small \( k \), for instance \( k = 5 \), provides a more robust estimator.

For the case \( I_N = \{N\} \),

\[
\hat{\gamma} = \frac{1}{2} \log\left(\frac{s_N^2}{s_{i_0}^2}\right),
\]

(35)
which asymptotically follows distribution $N \left( \gamma, \left[ 4m(\log(N/i_0))^2 \right]^{-1} \right)$, and hence a $(1 - \alpha)100\%$ confidence interval for $\gamma$ is

$$
\left[ \hat{\gamma} - \frac{z_{\alpha/2}}{2\sqrt{m\log(N/i_0)}}, \hat{\gamma} + \frac{z_{\alpha/2}}{2\sqrt{m\log(N/i_0)}} \right].
$$

Here $z_{\alpha}$ satisfies $1 - \Phi(z_{\alpha}) = \alpha$, where $\Phi(z) = P(N(0, 1) < z)$.

For large $m$, the distribution of $m\hat{\sigma}^2_i / \sigma^2_i = m\hat{C}/C$, $\chi^2_m$, can be approximated by $N(m/2, 2m)$, and hence for known $\gamma$ with probability $1 - \alpha$, a $(1 - \alpha)100\%$ confidence interval for $C$ is defined by $\sqrt{m/2}((\hat{C}/C) - 1) \in [-z_{\alpha/2}, z_{\alpha/2}]$. Using the above $(1 - \alpha)100\%$ confidence interval for $\gamma$ based on (34) gives the following asymptotic $(1 - 2\alpha)100\%$ confidence interval for $C$ when $m$ is large:

$$
\frac{s_i^2 - 2\gamma - z_{\alpha/2}/(\sqrt{m\log(N/i_0)})}{1 + z_{\alpha/2}/\sqrt{m/2}} \leq C \leq \frac{s_i^2 - 2\gamma + z_{\alpha/2}/(\sqrt{m\log(N/i_0)})}{1 - z_{\alpha/2}/\sqrt{m/2}}.
$$

Note that if $i_0 = 1$, then a $(1 - \alpha)100\%$ confidence interval for $C$ is $\sqrt{m/2}((\hat{C}/C) - 1) \in [-z_{\alpha/2}, z_{\alpha/2}]$.

In particular, this implies that with probability $1 - 2\alpha$, $|\hat{C}/C - 1| \leq z_{\alpha/2}m^{-1/2}$

$$
\left[ \sqrt{2} + \log(i_0) / \log(N/i_0) \right] =: c_{\text{max}} \text{ and } z_{\alpha/2} = \frac{\sqrt{m\log(N/i_0)}}{m}\text{ and hence, for large } m, \text{ with probability } 1 - 2\alpha, \text{ the relative upper bound for } m \text{ is approximately } e_{\gamma} = m^{-1/2}z_{\alpha/2}[2 + \sqrt{2}].
$$

### 6.7 Error in operator

Now we consider the case when the operator $K$ may be observed with error, and study its effect on the posterior contraction rate for mildly ill-posed inverse problems (see Section 5.4.3 for the setup).

Given white noise, Theorem 3 with $b = 0.5$ and $\log B \propto \log N$ implies that for $\mu_0 \in Q ((\hat{\theta}), \hat{A})$, the posterior distribution of $\mu$ given data $y_1, \ldots, y_N$ and observed $\hat{k}_1, \ldots, \hat{k}_N$ with $N$ satisfying $N|\log N|^{1/(2p)} \lesssim \delta^{-1/p}$ contracts at the optimal rate for $\alpha$ and $\tau$ specified in Corollary 3 if $N \geq i_{\epsilon} \asymp e^{-2/(1+2\beta+2p)}$. There exists $N \geq \sqrt{\epsilon}$ satisfying both conditions if

$$
\delta \leq \epsilon^{2p/(1+2\beta+2p)} |\log \epsilon|^{-1/2},
$$

for small $\epsilon$. This holds, for instance, if $\delta \leq \epsilon$.

Another problem is to determine the rate of contraction of the posterior distribution with noisy $\hat{k}_i$ for given values of $\epsilon$ and $\delta$ and compare it to the minimax rate for linear estimators in one dimension over sequence space which is $r_\epsilon = \max(\epsilon^{2p/(3p+2\beta+1)}, \delta, \delta^{1/p})$ over $Q ((\hat{\theta}), \hat{A})$ with the given bases (Theorem 2.1 and illustration 2.2 in Johannes & Schwarz, 2013). For mildly ill-posed problems and white noise, Theorem 3 with $\tau_\epsilon$ and $\alpha$ specified below in Corollary 3 gives the rate

$$
r_\epsilon \asymp \epsilon N^{p+1/2}_\epsilon + \epsilon N^{-\alpha}_\epsilon + \epsilon^2 \epsilon^{-2} N^{2p+2a+1-\beta}_\epsilon + N^{-\beta}_\epsilon \delta e^{-2} \epsilon^2 N^{p-2a-1-\beta}_\epsilon + \delta N^{(p-\beta)}_\epsilon,
$$

where $N_\epsilon = \min(N, i_{\epsilon})$ and $i_{\epsilon} \asymp \epsilon^{-2/(2p+2\beta+1)}$. Assume that $2p + 2\alpha + 1 - \beta \geq 0$ which is necessary to obtain the optimal contraction rate with known $K$ with $\gamma = 0$. Taking $N = \min(\delta^{-1/\beta}, \delta^{-1/p} |\log \delta|^{-1/(2p)})$ where the log factor is due to the constraint
Empirical Bayes posterior distribution of \( \mu \)

As we have seen in Section 6.3, the posterior distribution of \( \mu \) contracts at the optimal rate (in the minimax sense) for a particular set of the prior parameters \( \alpha \) and \( \tau \) (Corollary 3). In this section we study the contraction rate of the posterior distribution of \( \mu \) with a plugged in estimator of the prior scale \( \tau \) that does not rely on knowing \( \beta \), true smoothness of \( \mu_0 \).

Consider the prior Gaussian distribution with \( \lambda_i = \tau \lambda_{0,i} \), with empirical Bayes posterior of \( \mu \) using the MMLE of \( \tau \) and fixed \( \lambda_{0,i} \):

\[
\hat{\tau} = \arg \max_{\tau > 0} p(\mathbf{y}|\tau) = \arg \min_{\tau > 0} \sum_{i=1}^{\infty} \left[ \frac{y_i^2}{\kappa_i^{-1} \lambda_{0,i}^2 \tau + e^2 \sigma_i^2} + \log(\kappa_i^2 \lambda_{0,i} \sigma_i^{-2} \tau e^{-2} + 1) \right],
\]

where \( p(\mathbf{y}|\tau) = \int p(\mathbf{y}|\mu) dP(\mu|\tau) \) is the marginal density of \( \mathbf{y} = (y_1, y_2, \ldots) \) with respect to measure \( \prod_{i=1}^{\infty} N(0, e^2 \sigma_i^2) \). Recall that \( \tau_c = \tau^{1/2} \).

The following assumption is necessary to show convergence of \( \hat{\tau} \) in probability (but not for optimality of the posterior).

**Assumption 9.** For \( \mu_0 \in S^{\beta}(A) \), assume that for any \( \beta' \geq \beta \) there exists \( N_0 \geq 1 \) and \( c_0 > 0 \) such that for any \( N \geq N_0, N^{-2(\beta'-\beta)} \sum_{i=1}^{N} \mu_0^2 i^{2\beta'} \geq c_0. \)
Now we prove that the posterior distribution of $\mu$ with plugged in value of $\hat{\tau}$ defined by (36) contracts at the optimal rate adaptively, uniformly over $\mu_0 \in \mathcal{S}^\beta(A)$ with $0 < \beta \leq B_0 < \infty$, in the minimax sense, under some conditions on prior smoothness $\alpha$.

**Theorem 5.** Consider the inverse problem (1) formulated in the sequence space under Assumption 2 with noise $W$ satisfying (19) and Assumption 5, with prior distribution (23) under Assumption 7 with $\tau = \tau_e^*$, and the true function $\mu_0$ satisfying Assumptions 3 and 6. Assume that $1/2 + \beta + p + \gamma > 0$, and $1/2 + \alpha + p + \gamma > 0$.

1. If $\mu_0 \neq 0$, then there exist $\tau_e$ and $\bar{\tau}_e$ such that $P(\tau_e < \hat{\tau}_e < \bar{\tau}_e) \to 1$ as $\epsilon \to 0$. The posterior distribution of $\mu$ with plugged in $\hat{\tau}_e$ is consistent.

2. If $\mu_0 \neq 0$ and $\alpha \geq \max(\beta - 0.5, \beta/2 - 1/2 - p - \gamma)$, then the posterior contraction rate with plugged in $\hat{\tau}_e$ is optimal in the minimax sense.

3. If $\mu_0 \neq 0$, $\alpha \geq \beta - 0.5$ and Assumption 9 holds, then

   (a) $\tau_e$ and $\bar{\tau}_e$ are of the same order,

   (b) $\tau_e/\tau_e^* \to 1$ in probability as $\epsilon \to 0$ where $\tau_e^* = \arg\max_{\tau_e > 0} \mathbb{E}_{\mu_0} \log p(y|\tau_e) \asymp \gamma(2(\alpha - \beta)/(1 + 2\beta/2p + 2\gamma))$.

4. Finally, if $\mu_0 = 0$, then $\hat{\tau}_e = \Theta_P(\epsilon)$ and the plug-in contraction rate is $r_e = \epsilon$.

We show in the proof that for $\mu_0 \neq 0$,

$$
\tau_e^2 \geq e^{\frac{1}{1 + 4p + 2\gamma}},
$$

$$
\tau_e^2 \lesssim \begin{cases}
   e^{-4/(1 + 2p + 2\gamma + 2\beta)}(1 + o_P(1)), & \text{if } \alpha + 1/2 \geq \beta, \\
   e^{1/(1 + p + \gamma + \alpha)}(1 + o_P(1)), & \text{if } \alpha + 1/2 < \beta
\end{cases}
$$

**Corollary 4.** Under the conditions of Theorem 4, the posterior distribution of $\mu$ with the plugged in estimator $\hat{\tau}_e$ is consistent. The contraction rate is optimal in the minimax sense adaptively over $\mathcal{S}^\beta(A)$ for $0 < \beta \leq B_0 < \infty$ as long as

$$
\alpha \geq \max(B_0 - 1/2, B_0/2 - 1/2 - p - \gamma), \quad \text{and } \alpha > (- p - \gamma - 1/2)_+.
$$

The theorem states that if the prior smoothness parameter $\alpha$ is chosen large enough, then using the MMLE $\hat{\tau}$ allows us to achieve the optimal rate of contraction. If $p + \gamma = 0$ and $\mu_0 \in H^\beta(A)$, the results coincide with those of Szabó et al. (2013). We have weakened the condition for convergence of $\hat{\tau}_e$ in probability and show that it converges to the “oracle” value of $\tau_r$. The additional condition $\alpha \geq \beta/2 - 1/2 - p - \gamma$ arises only if $p + \gamma < 0$ and comes from the saturation term.

Now we consider the case of repeated observations where $\sigma_i^2$ are estimated, and we discuss the choice of $N$ in an adaptive setting, where we plug in the MMLE $\hat{\tau}$. Recall that the rate is not affected by the plugged in estimator $(\hat{\sigma}_i^2)$ if $\log(N) = o(m)$ and $N \geq i_{\epsilon,0}$ (see Proposition 5).

**Remark 6.** For a mildly ill-posed inverse problem with geometric sequence $(\sigma_i^2)$ satisfying Assumption 5, for the repeated observations and estimated $\sigma^2$, the rate of contraction of the posterior distribution of $\mu$ with the plugged in $(\hat{\sigma}_i^2)_i = 1$ is not affected if $\log(1/\epsilon) = o(m)$, which is generally a mild constraint. In particular, we can choose $N \geq C[\hat{\epsilon}_e e^{1/2}]^{1/(1 + 2p + 2\gamma + 2\alpha)}$ for some constant $C > 0$, which leads to a consistent posterior distribution due to the lower bound on $\hat{\epsilon}_e$ given by (37) (with high probability), and for $\mu_0$ satisfying Assumption 9 the posterior contraction rate is optimal.
7 | SIMULATION

7.1 | Simulation setup

We illustrate the theoretical results obtained in Section 6 on indirect observations from model (1) corrupted by the Volterra operator (Halmos, 1974) with dependent Gaussian noise and conjugate prior (4). Here \( H_1 = H_2 = L^2[0, 1] \).

The Volterra operator, \( K : L^2[0, 1] \to L^2[0, 1] \) is defined by
\[
K\mu(x) := \int_0^x \mu(s) ds, \quad \text{and} \quad K^T\mu(x) := \int_x^1 \mu(s) ds.
\]
The eigenvalues of \( KK^T \) and the orthonormal eigenbasis for the range of \( K \) are
\[
\kappa_i := \left[ \left( i - \frac{1}{2} \right)^2 \pi^2 \right]^{-\frac{1}{2}}, \quad \text{and} \quad \phi_i(x) := \sqrt{2} \sin \left( \left( i - \frac{1}{2} \right) \pi x \right) \quad \text{for every} \quad i \in \mathbb{N},
\]
where \( \kappa_i \propto i^{-p} \) with \( p = 1 \). The corresponding orthonormal eigenbasis of \( K^T K \) is \( e_i(x) = \sqrt{2} \cos \left( \left( i - \frac{1}{2} \right) \pi x \right) \).

We will estimate the following approximation of \( \mu_0(x) \): 
\[
\mu_0^N(x) := \sum_{i=1}^N \mu_{0,i} e_i(x),
\]
where \( N \) is the truncation parameter and is large to ensure good approximation: \( N \geq \max(50, e^{-2/(1+2p)}) \). We consider a particular function with \( \mu_{0,i} := i^{-3/2} \sin(i) \) which belongs to \( S^\beta \) with \( \beta = 1 \) (this can be shown using Dirichlet’s test, see e.g. Voxman & Goetschel, 1981).

We consider Gaussian noise \( W \) with truncated covariance operator \( V^N = \sum_{i=1}^N \sigma_i^2 \phi_i \phi_i^T \) and \( \sigma_i = i^\gamma \). Realizations of the data are simulated as follows: \( Y_i \sim N(\mu_{0,i} \kappa_i, \epsilon^2 \sigma_i^2) \), independently for \( i = 1, \ldots, N \).

We consider a centered Gaussian prior distribution with \( \lambda_i = \tau^\epsilon i^{-1-2\alpha} \) and different values of \( \alpha \) and \( \tau \), including the MMLE \( \hat{\tau} \).

The corresponding posterior distribution is
\[
\hat{\mu}_N | Y \sim N \left( \sum_{i=1}^N \frac{Y_i \kappa_i \lambda_i e_i}{\lambda_i \kappa_i^2 + \epsilon^2 \sigma_i^2}, \sum_{i=1}^N \frac{\epsilon^2 \sigma_i^2 \lambda_i e_i^2}{\lambda_i \kappa_i^2 + \epsilon^2 \sigma_i^2} \right).
\]
The (truncated) true function \( \mu_0^N(x) \), along with its noisy realizations \( Y(x) \) for two different noise levels are shown in Figure 1. We can see that with the noise level \( \epsilon = 10^{-2} \), the observed function \( Y \) is very noisy compared to the signal \( Y_0 = K\mu_0 \): the range of observations is approximately \([-44, 43]\) whereas the range of the signal is \([0, 0.67]\). For a smaller noise level \( \epsilon = 10^{-4} \), the range of observations is \([0.2, 0.8]\) which is comparable to the range of \( Y_0 \).

A key property we want to study here is the variability of the posterior distribution around the true value of the function and posterior coverage, that is, the posterior probability that the true function lies in the support of the posterior. A common way to investigate the posterior support in Bayesian nonparametric models is to plot a large number of draws from the posterior distribution, where the “center” of the support is displayed using the posterior mean. We plot 100 draws as plotting a larger number in the considered examples does not much change the posterior support. Our main interest is to see how the coverage of the true function by the posterior distribution and its contraction is affected by the choice of prior smoothness \( \alpha \), both for a fixed prior scale \( \tau \) (nonadaptive case) and for the empirical Bayes \( \hat{\tau} \), as well as by the variance parameters \( \epsilon \) and \( \gamma \).
7.2 Nonadaptive posterior distribution of \( \mu \)

Firstly we study how posterior variability and the coverage of the true function varies with prior smoothness \( \alpha \) as we fix prior scale \( \tau = 1 \). In this section we also fix \( \epsilon = 10^{-2}, \gamma = 0.5, N = 2000, \tau_{\epsilon} = 1 \) and consider how different values of a priori smoothness \( \alpha \) affect behavior of the posterior distribution. We consider the following values of \( \alpha = (0.5, 0.75, 1, 2, 3, 5) \). Draws from the posterior distributions of \( \mu_{\mathbf{N}}(x) \) corresponding to these values of \( \alpha \) are given in Figure 2, for the same realization of \( Y(x) \). Individual draws from the corresponding posterior distribution with a priori smoothness \( \alpha \) are plotted in figure 5 in Appendix S1.

For \( \alpha \leq \beta = 1 \), the variability of the posterior is large so that the true function lies inside the credible band. For larger values of \( \alpha \), posterior variability around the posterior mean is much smaller, but the bias of the posterior mean increases, so the true function does not lie inside the credible band. The value of \( \alpha \) (among the considered values) that gives the posterior with the smallest uncertainty while containing the true function appears to be 1, which is equal to \( \beta \) (the smoothness of \( \mu_{0}(x) \)), as predicted by theory (Corollary 3 with constant \( \tau \)).

7.3 Empirical Bayes posterior distribution of \( \mu \)

In this section we fix \( \alpha > \beta \) and apply the Empirical Bayes estimator of \( \tau_{\epsilon} \) defined by (36), to check if the corresponding empirical Bayes posterior provides reasonable coverage of the true function, and how it is affected by the noise level \( \epsilon \) and different values of \( \gamma \).

In Figures 3 and 4 we can see that with decreasing noise level \( \epsilon \) the posterior distribution contracts to the mean, with much smaller bias than for a fixed \( \tau \)—the behavior predicted by the theory. Interestingly, for larger \( \alpha \) (\( \alpha = 5 \) in Figure 4), the bias decreases to 0 slower than for smaller \( \alpha \) (\( \alpha = 1 \) in Figure 3) however the variability of the posterior decreases faster.

Boxplots of values of \( \hat{\tau} \) over 100 simulations for different values of \( \alpha \) and different values of \( \gamma \) are given in figure 6 in Appendix S1. In each case, the sampling distribution of \( \hat{\tau} \) concentrates, and the values appear to increase exponentially as functions of \( \alpha \). This can be explained by our theory.
FIGURE 2  Plots of $\mu_0^N(x)$ (black line) along with the posterior mean (red line), and 100 draws from the posterior (blue dashes) for $\alpha = (0.5, 0.75, 1, 2, 3, 5)$ respectively, with $\epsilon = 10^{-2}, \gamma = 0.5$ and $N = 2000$ in all cases.

FIGURE 3  One hundred draws from empirical Bayes posterior with $\mu_0^N$ (black line), $\epsilon = 10^{-4}$ (left), $\epsilon = 10^{-5}$ (middle) and $\epsilon = 10^{-8}$ (right), $\alpha = 1, \gamma = 0.5$. 
For self-similar functions and $\alpha > \beta - 1/2$, $\hat{\tau}$ is close to the oracle value $\tau^*$ (Theorem 5) which increases exponentially in $\alpha$ (Corollary 3). The values of $\hat{\tau}$ do not vary much for the considered different values of $\gamma$ but they do vary with $\alpha$, as expected (Corollary 3).

We also plotted values of prior SDs $\sqrt{\lambda_i}$ with plugged in $\hat{\tau}$ for different values of $\alpha$ and $\gamma$ (figure 7 in Appendix S1). Note that the eigenvalues do not much differ for the considered values of $\gamma$, as the only effect is through $\hat{\tau}$. For each $\gamma$, the values of $\hat{\tau}$ are such that the values of $\lambda_i = \hat{\tau}i^{-2\alpha-1}$ are the same at some index $i$ (around $i = 50$). This is expected due to $\lambda_i$ corresponding to the optimal cutoff $i_c$ being independent of $\alpha$.

Therefore, the empirical Bayes posterior adapts well to the unknown function and contracts to the true value of the function as the noise level $\epsilon$ vanishes as stated by theory (Theorem 5). Choosing $\alpha$ larger than $\beta$ and using the empirical Bayes estimate of $\tau$ does lead to the contraction of the posterior distribution of $\mu$ and good coverage of $\mu_0$. This holds for various values of $\gamma$, including the case of an ill-posed inverse problem ($p + \gamma > 0$), a partially regularized model ($p + \gamma = 0$) and a case where the contraction rate is $\epsilon$ ($p + \gamma < -1/2$). The value of $\gamma$ does not have a strong effect on the posterior concentration and on the behavior of the $\hat{\tau}$ and the prior eigenvalues $\lambda_i$.

8 | DISCUSSION

We have considered the inverse problem with Gaussian errors in Hilbert spaces where the covariance operator is not constant but is decomposable in a biorthogonal basis which is an image of a Riesz basis, and both bases span the corresponding Hilbert spaces. We showed that this leads to a sequence space formulation of the inverse problem, and studied its posterior contraction rate, in particular its optimality in the minimax sense, possibly under the error in the forward operator and in the covariance operator. We focused on mildly ill-posed inverse problems with
fractional noise, where we also studied optimality of adaptive empirical Bayes posterior distribution. We also identified a setting where the posterior distribution can contract at a faster rate, effectively leading to self-regularization of the inverse problem, which was also discussed in Johannes et al. (2020).

We extend the general theorem to study the effect of using a plug-in estimator of the variances in sequence space on the posterior contraction rate. We discuss in which cases the rate of contraction of the posterior distribution is not affected whether the covariance operator is known exactly or observed with error. We consider two types of consistency of the estimator: in terms of the absolute bound on the difference between estimated and true values, and in terms of a relative bound. We study its effect on mildly ill-posed inverse problems, and consider in detail the case of repeated observations. We find that the relative consistency conditions lead to weaker effects of the plug-in on the posterior contraction rate. We have also applied these results to study an error in operator, for example, when the eigenfunctions are known but the eigenvalues are estimated, for a fractional noise.

We consider in detail a particular case of the covariance operator whose coefficients in sequence space decrease to 0 at a polynomial rate and illustrate it on the case where the noise is fBM using fractional wavelets as the bases. We also derive minimax rates of convergence of estimators of the unknown signal under this model, and show the choice of the prior parameters that leads to this contraction rate of the posterior. We studied the empirical Bayes approach that leads to the corresponding posterior contracting at the optimal rate, under some assumptions on prior smoothness. An alternative approach to adapt in inverse problems uniformly over $S^{\alpha}(A)$ for a range of $\alpha$ is a sieve prior proposed by Johannes et al. (2020). Our simulation results confirm the theoretical conclusions.

One can argue that it is not realistic to assume the knowledge of the covariance operator, and it needs to be estimated in practice. We discuss when estimating of (discretized) eigenfunctions is possible ($\gamma < -1/2$) for repeated observations, and when the number of estimated eigenfunctions is sufficient to achieve the optimal posterior contraction rate of $\mu$.

Another interesting question is estimating the Hurst exponent for fractional noise. While it is possible to estimate $H$, similarly to estimating $\gamma$, using asymptotic expression for coefficients for large indices, fractional wavelets that are used to decompose fractional noise depend on $H$. Therefore, studying the posterior contraction rate with estimated $H$ and using biorthogonal bases that depend on $H$ is a challenging question.

An open question for future research is a joint Bayesian model of the signal and the variance function when the latter is unknown, and its asymptotic behavior, and the behavior of the full Bayesian model when the scale parameter is estimated.

ACKNOWLEDGMENTS

Jenovah Rodrigues was supported by The Maxwell Institute Graduate School in Analysis and its Applications, a Centre for Doctoral Training funded by the UK Engineering and Physical Sciences Research Council (grant EP/L016508/01), the Scottish Funding Council, Heriot-Watt University, and the University of Edinburgh.

FUNDING INFORMATION

Jenovah Rodrigues was funded for his PhD by the MIGSAA Doctoral Training Centre, EPSRC grant EP/L016508/01.
REFERENCES

Abramovich, F., & Silverman, B. W. (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, 85, 115–129.

Abry, P., & Sellan, F. (1996). The wavelet-based synthesis for fractional Brownian motion proposed by F. Sellan and Y. Meyer: Remarks and fast implementation. *Applied and Computational Harmonic Analysis*, 3, 377–383.

Agapiou, S., Dashti, M., & Helin, T. (2021). Rates of contraction of posterior distributions based on p-exponential priors. *Bernoulli*, 27, 1616–1642.

Agapiou, S., Larsson, S., & Stuart, A. M. (2013). Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems. *Stochastic Processes and Their Applications*, 123, 3828–3860.

Agapiou, S., & Mathé, P. (2018). *Posterior contraction in Bayesian inverse problems under Gaussian priors*. In *New trends in parameter identification for mathematical models* (pp. 1–29). Springer.

Auranen, T., Nummenmaa, A., Hämäläinen, M. S., Jääskeläinen, I. P., Lampinen, J., Vehtari, A., & Sams, M. (2005). Bayesian analysis of the neuromagnetic inverse problem with $\ell_p$-norm priors. *NeuroImage*, 26, 870–884.

Ayache, A., & Taqqu, M. S. (2003). Rate optimality of wavelet series approximations of fractional Brownian motion. *The Journal of Fourier Analysis and Applications*, 9, 451–471.

Belitser, E. (2017). On coverage and local radial rates of credible sets. *The Annals of Statistics*, 45, 1124–1151.

Belitser, E. N., & Levit, B. Y. (1995). On minimax filtering over ellipsoids. *Mathematical Methods of Statistics*, 3, 259–273.

Bochkina, N. (2013). Consistency of the posterior distribution in generalized linear inverse problems. *Inverse Problems*, 29, 095010.

Cavalier, L. (2008). Nonparametric statistical inverse problems. *Inverse Problems*, 24, 034004.

Cavalier, L., & Hengartner, N. W. (2005). Adaptive estimation for inverse problems with noisy operators. *Inverse problems*, 21, 1345.

Cohen, A., Dahmen, W., & DeVore, R. (2000). Multiscale decompositions in bounded domains. *Transactions of the American Mathematical Society*, 352, 3651–3685.

Cotter, S. L., Dashti, M., Robinson, J. C., & Stuart, A. M. (2009). Bayesian inverse problems for functions and applications to fluid mechanics. *Inverse Problems*, 25, 115008.

Dashti, M., Harris, S., & Stuart, A. (2012). Besov priors for Bayesian inverse problems. *Inverse Problems and Imaging*, 6, 183–200.

Donoho, D. L. (1995). Nonlinear solution of linear inverse problems by Wavelet–Vaguelette decomposition. *Applied and Computational Harmonic Analysis*, 2, 101–126.

Dzhaparidze, K., & van Zanten, H. (2005). Optimality of an explicit series expansion of the fractional Brownian sheet. *Statistics & Probability Letters*, 71, 295–301.

Efendiev, Y., Datta-Gupta, A., Hwang, K., Ma, X., & Mallick, B. (2010). *Bayesian partition models for subsurface characterization*. In *Large-scale inverse problems and quantification of uncertainty* (pp. 107–122). John Wiley & Sons, Ltd.

Engl, H. W., Hanke, M., & Neubauer, A. (1996). *Regularization of inverse problems*. Springer.

Florens, J.-P., & Simoni, A. (2016). Regularizing priors for linear inverse problems. *Econometric Theory*, 32, 71–121.

Ghosal, S., Ghosh, J. K., & van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *The Annals of Statistics*, 28, 500–531.

Gilsing, H., & Sottinen, T. (2003). Power series expansions for fractional Brownian motions. *Theory of Stochastic Processes*, 9, 38–49.

Halmos, P. R. (1974). *A Hilbert space problem book* (Vol. 19, 680108946). Springer-Verlag.

Hoffmann, M., & Reiss, M. (2008). Nonlinear estimation for linear inverse problems with error in the operator. *The Annals of Statistics*, 36, 310–336.

Johannes, J., Simoni, A., & Schenk, R. (2020). Adaptive Bayesian estimation in indirect Gaussian sequence space models. *Annals of Economics and Statistics*, 137, 83–116.
Johnstone, I. M., & Silverman, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59, 319–351.

Kaipio, J. P., Kolehmainen, V., Vauhkonen, M., & Somersalo, E. (1999). Inverse problems with structural prior information. *Inverse Problems*, 15, 713–729.

Knapik, B., & Salomond, J.-B. (2018). A general approach to posterior contraction in nonparametric inverse problems. *Bernoulli*, 24, 2091–2121.

Knapik, B. T., Szabo, B. T., van der Vaart, A. W., & van Zanten, J. H. (2016). Bayes procedures for adaptive inference in inverse problems for the white noise model. *Probability Theory and Related Fields*, 164, 771–813.

Knapik, B. T., van der Vaart, A. W., & van Zanten, J. H. (2011). Bayesian inverse problems with Gaussian priors. *The Annals of Statistics*, 39, 2626–2657.

Kolaczek, Y. D. (1996). An application of wavelet shrinkage to tomography. In *Wavelets in medicine and biology* (pp. 77–92). Routledge.

Koltchinskii, V., & Lounici, K. (2017). Normal approximation and concentration of spectral projectors of sample covariance. *The Annals of Statistics*, 45, 121–157.

Kühn, T., & Linde, W. (2002). Optimal series representation of fractional Brownian sheets. *Bernoulli*, 8, 669–696.

Meyer, Y., Sellan, F., & Taqqu, M. S. (1999). Wavelets, generalized white noise and fractional integration: The synthesis of fractional Brownian motion. *The Journal of Fourier Analysis and Applications*, 5, 465–494.

Ndaoud, M. (2018). Harmonic analysis meets stationarity: A general framework for series expansions of special Gaussian processes. *The Journal of Fourier Analysis and Applications*, 9.

Picard, J. (2011). *Representation formulae for the fractional Brownian motion*. In C. Donati-Martin, A. Lejay, & A. Rouault (Eds.), *Séminaire de probabilités XLIII Lecture Notes in Mathematics* (pp. 3–70). Springer.

Ray, K. (2013). Bayesian inverse problems with non-conjugate priors. *Electronic Journal of Statistics*, 7, 2516–2549.

Röver, C., Meyer, R., Guidi, G. M., Viceré, A., & Christensen, N. (2007). Coherent Bayesian analysis of inspiral signals. *Classical and Quantum Gravity*, 24, S607–S615.

Schmidt-Hieber, J. (2014). Asymptotic equivalence for regression under fractional noise. *The Annals of Statistics*, 42, 2557–2585.

Szabó, B. T., Vaart, A. W. v. d., & Zanten, J. H. v. (2013). Empirical Bayes scaling of Gaussian priors in the white noise model. *Electronic Journal of Statistics*, 7, 991–1018.

Tarantola, A. (2006). Popper, Bayes and the inverse problem. *Nature Physics*, 2, 492–494.

Vidakovic, B. (1999). *Statistical modeling by wavelets*. John Wiley & Sons.

Voutilainen, A., & Kaipio, J. (2009). Model reduction and pollution source identification from remote sensing data. *Inverse Problems and Imaging*, 3, 711–730.

Voxman, W. L., & Goetschel, R. (1981). *Advanced calculus: An introduction to modern analysis Pure and Applied Mathematics* (Vol. 63). CRC Press.

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.

---

**How to cite this article:** Bochkina, N., & Rodrigues, J. (2023). Bayesian inverse problems with heterogeneous variance. *Scandinavian Journal of Statistics*, 50(3), 1116–1151. [https://doi.org/10.1111/sjos.12622](https://doi.org/10.1111/sjos.12622)