

Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem

Matteo Giordano¹ and Richard Nickl¹

Centre for Mathematical Sciences, University of Cambridge, Wilberforce Rd, Cambridge, CB3 0WA, United Kingdom

Received 16 October 2019, revised 4 February 2020

Accepted for publication 5 March 2020

Published 20 August 2020



Abstract

For \mathcal{O} a bounded domain in \mathbb{R}^d and a given smooth function $g : \mathcal{O} \rightarrow \mathbb{R}$, we consider the statistical nonlinear inverse problem of recovering the conductivity $f > 0$ in the divergence form equation $\nabla \cdot (f \nabla u) = g$ on \mathcal{O} , $u = 0$ on $\partial\mathcal{O}$, from N discrete noisy point evaluations of the solution $u = u_f$ on \mathcal{O} . We study the statistical performance of Bayesian nonparametric procedures based on a flexible class of Gaussian (or hierarchical Gaussian) process priors, whose implementation is feasible by MCMC methods. We show that, as the number N of measurements increases, the resulting posterior distributions concentrate around the true parameter generating the data, and derive a convergence rate $N^{-\lambda}$, $\lambda > 0$, for the reconstruction error of the associated posterior means, in $L^2(\mathcal{O})$ -distance.

Keywords: inverse problems, Bayesian inference, Gaussian prior, frequentist consistency

1. Introduction

Statistical inverse problems arise naturally in many applications in physics, imaging, tomography, and generally in engineering and throughout the sciences. A prototypical example involves a domain $\mathcal{O} \subset \mathbb{R}^d$, some function $f : \mathcal{O} \rightarrow \mathbb{R}$ of interest, and indirect measurements $G(f)$ of f , where G is a given solution (or ‘forward’) operator of some partial differential equation (PDE) governed by the unknown coefficient f . A natural statistical observational model postulates data

$$Y_i = G(f)(X_i) + \sigma W_i, \quad i = 1, \dots, N, \quad (1)$$

¹ Author to whom any correspondence should be addressed.



Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

where the X_i 's are design points at which the PDE solution $G(f)$ is measured, and where the W_i 's are standard Gaussian noise variables scaled by a noise level $\sigma > 0$. The aim is then to infer f from the data $(Y_i, X_i)_{i=1}^N$. The study of problems of this type has a long history in applied mathematics, see the monographs (Engl *et al* 1996, Kaltenbacher *et al* 2008), although explicit *statistical* noise models have been considered only more recently (Bissantz *et al* 2004, Bissantz *et al* 2007, Hohage and Pricop 2008, Kaipio and Somersalo 2004). Recent survey articles on the subject are (Arridge *et al* 2019, Benning and Burger 2018) where many more references can be found.

For many of the most natural PDEs—such as the divergence form elliptic equation (2) considered below—the resulting maps G are *non-linear* in f , and this poses various challenges: among other things, the negative log-likelihood function associated to the model (1), which equals the least squares criterion (see (10) below for details), is then possibly *non-convex*, and commonly used statistical algorithms (such as maximum likelihood estimators, Tikhonov regularisers or MAP estimates) defined as optimisers in f of likelihood-based objective functions can not reliably be computed by standard convex optimisation techniques. While iterative optimisation methods (such as Landweber iteration) may overcome such challenges (Kaltenbacher *et al* 2008, Hanke *et al* 1995, Kaltenbacher *et al* 2009, Qi-nian 2000), an attractive alternative methodology arises from the Bayesian approach to inverse problems advocated in an influential paper by Stuart (Stuart 2010): one starts from a *Gaussian process prior* Π for the parameter f or in fact, as is often necessary, for a suitable vector-space valued re-parameterisation F of f . One then uses Bayes' theorem to infer the best posterior guess for f given data $(Y_i, X_i)_{i=1}^N$. Posterior distributions and their expected values can be approximately computed via Markov chain Monte Carlo (MCMC) methods (see, for example, Beskos *et al* 2017, Conrad *et al* 2016, Cotter *et al* 2013 and references therein) as soon as the forward map $G(\cdot)$ can be evaluated numerically, avoiding optimisation algorithms as well as the use of (potentially tedious, or non-existent) inversion formulas for G^{-1} ; see subsection 2.3.1 below for more discussion. The Bayesian approach has been particularly popular in application areas as it does not only deliver an estimator for the unknown parameter f but simultaneously provides uncertainty quantification methodology for the recovery algorithm via the probability distribution of $f|(Y_i, X_i)_{i=1}^N$ (see, for example, Dashti and Stuart 2016). Conceptually related is the area of 'probabilistic numerics' (Briol *et al* 2019) in the noise-less case $\sigma = 0$, with key ideas dating back to work by Diaconis (1988).

As successful as this approach may have proved to be in algorithmic practice, for the case when the forward map G is non-linear we currently only have a limited understanding of the statistical validity of such Bayesian inversion methods. By validity we mean here *statistical guarantees* for convergence of natural Bayesian estimators such as the posterior mean $\bar{f} = E^\Pi[f|(Y_i, X_i)_{i=1}^N]$ towards the ground truth f_0 generating the data. Without such guarantees, the interpretation of posterior based inferences remains vague: the randomness of the prior may have propagated into the posterior in a way that does not 'wash out' even when very informative data is available (e.g., small noise variance and/or large sample size N), rendering Bayesian methods potentially ambiguous for the purposes of valid statistical inference and uncertainty quantification.

In the present article we attempt to advance our understanding of this problem area in the context of the following basic but representative example for a non-linear inverse problem: let g be a given smooth 'source' function, and let $f : \mathcal{O} \rightarrow \mathbb{R}$ be a an unknown conductivity

parameter determining solutions $u = u_f$ of the PDE

$$\begin{cases} \nabla \cdot (f \nabla u) = g & \text{on } \mathcal{O}, \\ u = 0 & \text{on } \partial \mathcal{O}, \end{cases} \quad (2)$$

where we denote by $\nabla \cdot$ the divergence and by ∇ the gradient operator, respectively. Under mild regularity conditions on f , and assuming that $f \geq K_{\min} > 0$ on \mathcal{O} , standard elliptic theory implies that (2) has a unique classical C^2 -solution $G(f) \equiv u_f$. Identification of f from an observed solution u_f of this PDE has been considered in a large number of articles both in the applied mathematics and statistics communities—we mention here (Stuart 2010, Beskos *et al* 2017, Dashti and Stuart 2016, Briol *et al* 2019, Alessandrini 1986, Bonito *et al* 2017, Dashti and Stuart 2011, Falk 1983, Hoffmann and Sprekels 1985, Ito and Kunisch 1994, Knowles 2001, Kohn and Lowe 1988, Kravaris and Seinfeld 1985, Nickl *et al* 2020, Richter 1981, Schwab and Stuart 2012, Vollmer 2013) and the many references therein.

The main contributions of this article are as follows: we show that posterior means arising from a large class of Gaussian (or conditionally Gaussian) process priors for f provide statistically consistent recovery (with explicit polynomial convergence rates as the number N of measurements increases) of the unknown parameter f in (2) from data in (1). While we employ the theory of posterior contraction from Bayesian non-parametric statistics (Ghosal and van der Vaart 2017, van der Vaart and van Zanten 2008, van der Vaart and van Zanten 2009), the non-linear nature of the problem at hand leads to substantial additional challenges arising from the fact that (a) the Hellinger distance induced by the statistical experiment is not naturally compatible with relevant distances on the actual parameter f and that (b) the ‘push-forward’ prior induced on the information-theoretically relevant regression functions $G(f)$ is non-explicit (in particular, non-Gaussian) due to the non-linearity of the map G . Our proofs apply recent ideas from Monard *et al* (2020) to the present elliptic situation. In the first step we show that the posterior distributions arising from the priors considered (optimally) solve the PDE-constrained regression problem of inferring $G(f)$ from data (1). Such results can then be combined with a suitable ‘stability estimate’ for the inverse map G^{-1} to show that, for large sample size N , the posterior distributions concentrate around the true parameter generating the data at a convergence rate $N^{-\lambda}$ for some $\lambda > 0$. We ultimately deduce the same rate of consistency for the posterior mean from quantitative uniform integrability arguments.

The first results we obtain apply to a large class of ‘rescaled’ Gaussian process priors similar to those considered in Monard *et al* (2020), addressing the need for additional *a priori* regularisation of the posterior distribution in order to tame non-linear effects of the ‘forward map’. This rescaling of the Gaussian process depends on sample size N . From a non-asymptotic point of view this just reflects an adjustment of the covariance operator of the prior, but following Diaconis (1988) one may wonder whether a ‘fully Bayesian’ solution of this non-linear inverse problem, based on a prior that does *not* depend on N , is also possible. We show indeed that a hierarchical prior that randomises a finite truncation point in the Karhunen–Loève-type series expansion of the Gaussian base prior will also result in consistent recovery of the conductivity parameter f in equation (2) from data (1), at least if f is smooth enough.

Let us finally discuss some related literature on statistical guarantees for Bayesian inversion: to the best of our knowledge, the only previous paper concerned with (frequentist) consistency of Bayesian inversion in the elliptic PDE (2) is by Vollmer (2013). The proofs in Vollmer (2013) share a similar general idea in that they rely on a preliminary treatment of the associated regression problem for $G(f)$, which is then combined with a suitable stability estimate for G^{-1} . However, the convergence rates obtained in Vollmer (2013) are only implicitly given and sub-optimal, also (unlike ours) for ‘prediction risk’ in the PDE-constrained regression problem.

Moreover, when specialised to the concrete non-linear elliptic problem (2) considered here, the results in section 4 in Vollmer (2013) only hold for priors with bounded C^β -norms, such as ‘uniform wavelet type priors’, similar to the ones used in Nickl (2018), Nickl and Söhl (2017), and Nickl and Söhl (2019) for different non-linear inverse problems. In contrast, our results hold for the more practical Gaussian process priors which are commonly used in applications, and which permit the use of tailor-made MCMC methodology—such as the pCN algorithm discussed in subsection 2.3.1—for computation.

The results obtained in Nickl *et al* (2020) for the maximum *a posteriori* (MAP) estimates associated to the priors studied here are closely related to our findings in several ways. Ultimately the proof methods in Nickl *et al* (2020) are, however, based on variational methods and hence entirely different from the Bayesian ideas underlying our results. Moreover, the MAP estimates in Nickl *et al* (2020) are difficult to compute due to the lack of convexity of the forward map, whereas posterior means arising from Gaussian process priors admit explicit computational guarantees, see Hairer *et al* (2014) and also subsection 2.3.1 for more details.

It is further of interest to compare our results to those recently obtained in Abraham and Nickl (2019), where the statistical version of the *Caldéron problem* is studied. There the ‘Dirichlet-to-Neumann map’ of solutions to the PDE (2) is observed, corrupted by appropriate Gaussian matrix noise. In this case, as only boundary measurements of u_f at $\partial\mathcal{O}$ are available, the statistical convergence rates are only of order $\log^{-\gamma}(N)$ for some $\gamma > 0$ (as $N \rightarrow \infty$), whereas our results show that when interior measurements of u_f are available throughout \mathcal{O} , the recovery rates improve to $N^{-\lambda}$ for some $\lambda > 0$.

There is of course a large literature on consistency of Bayesian *linear* inverse problems with Gaussian priors, we only mention Agapiou *et al* (2013), Kekkonen *et al* (2016), Knapik *et al* (2011), Monard *et al* (2019), and Ray (2013) and references therein. The non-linear case considered here is fundamentally more challenging and cannot be treated by the techniques from these papers—however, some of the general theory we develop in the appendix provides novel proof methods also for the linear setting.

This paper is structured as follows. Section 2 contains all the main results for the inverse problem arising with the PDE model (2). The proofs, which also include some theory for general non-linear inverse problems that is of independent interest, are given in section 3 and appendix A. Finally, appendix B provides additional details on some facts used throughout the paper.

2. Main results

2.1. A statistical inverse problem with elliptic PDEs

2.1.1. Main notation. Throughout the paper, $\mathcal{O} \subset \mathbb{R}^d$, $d \in \mathbb{N}$, is a given nonempty open and bounded set with smooth boundary $\partial\mathcal{O}$ and closure $\bar{\mathcal{O}}$.

The spaces of continuous functions defined on \mathcal{O} and $\bar{\mathcal{O}}$ are respectively denoted $C(\mathcal{O})$ and $C(\bar{\mathcal{O}})$, and endowed with the supremum norm $\|\cdot\|_\infty$. For positive integers $\beta \in \mathbb{N}$, $C^\beta(\mathcal{O})$ is the space of β -times differentiable functions with uniformly continuous derivatives; for non-integer $\beta > 0$, $C^\beta(\mathcal{O})$ is defined as

$$C^\beta(\mathcal{O}) = \left\{ f \in C^{\lfloor \beta \rfloor}(\mathcal{O}) : \forall |i| = \lfloor \beta \rfloor, \sup_{x,y \in \mathcal{O}, x \neq y} \frac{|D^i f(x) - D^i f(y)|}{|x - y|^{\beta - \lfloor \beta \rfloor}} < \infty \right\},$$

where $\lfloor \beta \rfloor$ denotes the largest integer less than or equal to β , and for any multi-index $i = (i_1, \dots, i_d)$, D^i is the i th partial differential operator. $C^\beta(\mathcal{O})$ is normed by

$$\|f\|_{C^\beta(\mathcal{O})} = \sum_{|i| \leq \lfloor \beta \rfloor} \sup_{x \in \mathcal{O}} |D^i f(x)| + \sum_{|i| = \lfloor \beta \rfloor} \sup_{x, y \in \mathcal{O}, x \neq y} \frac{|D^i f(x) - D^i f(y)|}{|x - y|^{\beta - \lfloor \beta \rfloor}},$$

where the second summand is removed for integer β . We denote by $C^\infty(\mathcal{O}) = \cap_\beta C^\beta(\mathcal{O})$ the set of smooth functions, and by $C_c^\infty(\mathcal{O})$ the subspace of elements in $C^\infty(\mathcal{O})$ with compact support contained in \mathcal{O} .

Denote by $L^2(\mathcal{O})$ the Hilbert space of square integrable functions on \mathcal{O} , equipped with its usual inner product $\langle \cdot, \cdot \rangle_{L^2(\mathcal{O})}$. For integer $\alpha \geq 0$, the order- α Sobolev space on \mathcal{O} is the separable Hilbert space

$$H^\alpha(\mathcal{O}) = \{f \in L^2(\mathcal{O}) : \forall |i| \leq \alpha, \exists D^i f \in L^2(\mathcal{O})\},$$

$$\langle f, g \rangle_{H^\alpha(\mathcal{O})} = \sum_{|i| \leq \alpha} \langle D^i f, D^i g \rangle_{L^2(\mathcal{O})}.$$

For non-integer $\alpha \geq 0$, $H^\alpha(\mathcal{O})$ can be defined by interpolation, for example, Lions and Magenes (1972). For any $\alpha \geq 0$, $H_c^\alpha(\mathcal{O})$ will denote the completion of $C_c^\infty(\mathcal{O})$ with respect to the norm $\|\cdot\|_{H^\alpha(\mathcal{O})}$. Finally, if K is a nonempty compact subset of \mathcal{O} , we denote by $H_K^\alpha(\mathcal{O})$ the closed subspace of functions in $H^\alpha(\mathcal{O})$ with support contained in K . Whenever there is no risk of confusion, we will omit the reference to the underlying domain \mathcal{O} .

Throughout, we use the symbols \lesssim and \gtrsim for inequalities holding up to a universal constant. Also, for two real sequences (a_N) and (b_N) , we say that $a_N \simeq b_N$ if both $a_N \lesssim b_N$ and $b_N \lesssim a_N$ for all N large enough. For a sequence of random variables Z_N we write $Z_N = O_{\text{Pr}}(a_N)$ if for all $\varepsilon > 0$ there exists $M_\varepsilon < \infty$ such that for all N large enough, $\Pr(|Z_N| \geq M_\varepsilon a_N) < \varepsilon$. Finally, we will denote by $\mathcal{L}(Z)$ the law of a random variable Z .

2.1.2. Parameter spaces and link functions. Let $g \in C^\infty(\mathcal{O})$ be an arbitrary source function, which will be regarded as fixed throughout. For $f \in C^\beta(\mathcal{O})$, $\beta > 1$, consider the boundary value problem

$$\begin{cases} \nabla \cdot (f \nabla u) = g & \text{on } \mathcal{O}, \\ u = 0 & \text{on } \partial\mathcal{O}. \end{cases} \quad (3)$$

If we assume that $f \geq K_{\min} > 0$ on \mathcal{O} , then standard elliptic theory (e.g. Gilbarg and Trudinger (1998)) implies that (3) has a classical solution $G(f) \equiv u_f \in C(\bar{\mathcal{O}}) \cap C^{1+\beta}(\mathcal{O})$.

We consider the following parameter space for f : for integer $\alpha > 1 + d/2$, $K_{\min} \in (0, 1)$, and denoting by $n = n(x)$ the outward pointing normal at $x \in \partial\mathcal{O}$, let

$$\mathcal{F}_{\alpha, K_{\min}} = \left\{ f \in H^\alpha(\mathcal{O}) : \inf_{x \in \mathcal{O}} f(x) > K_{\min}, f|_{\partial\mathcal{O}} = 1, \frac{\partial^j f}{\partial n^j}|_{\partial\mathcal{O}} = 0 \text{ for } 1 \leq j \leq \alpha - 1 \right\}. \quad (4)$$

Our approach will be to place a prior probability measure on the unknown conductivity f and base our inference on the posterior distribution of f given noisy observations of $G(f)$, via Bayes' theorem. It is of interest to use *Gaussian process priors*. Such probability measures are naturally supported in linear spaces (in our case $H_c^\alpha(\mathcal{O})$) and we now introduce a bijective

re-parametrisation so that the prior for f is supported in the relevant parameter space $\mathcal{F}_{\alpha, K_{\min}}$. We follow the approach of using regular link functions Φ as in Nickl *et al* (2020).

Condition 1. For given $K_{\min} > 0$, let $\Phi : \mathbb{R} \rightarrow (K_{\min}, \infty)$ be a smooth, strictly increasing bijective function such that $\Phi(0) = 1$, $\Phi'(t) > 0$, $t \in \mathbb{R}$, and assume that all derivatives of Φ are bounded on \mathbb{R} .

For some of the results to follow it will prove convenient to slightly strengthen the previous condition.

Condition 2. Let Φ be as in condition 1, and assume furthermore that Φ' is nondecreasing and that $\liminf_{t \rightarrow -\infty} \Phi'(t)t^a > 0$ for some $a > 0$.

For $a = 2$, an example of such a link function is given in example 24 below. Note however that the choice of $\Phi = \exp$ is not permitted in either condition.

Given any link function Φ satisfying condition 1, one can show (cf. Nickl *et al* (2020), section 3.1) that the set $\mathcal{F}_{\alpha, K_{\min}}$ in (4) can be realised as the family of composition maps

$$\mathcal{F}_{\alpha, K_{\min}} = \{\Phi \circ F : F \in H_c^\alpha(\mathcal{O})\}, \quad \alpha \in \mathbb{N}.$$

We then regard the solution map associated to (3) as one defined on H_c^α via

$$\mathcal{G} : H_c^\alpha(\mathcal{O}) \rightarrow L^2(\mathcal{O}), \quad F \mapsto \mathcal{G}(F) := G(\Phi \circ F), \quad (5)$$

where $G(\Phi \circ F)$ is the solution to (3) now with $f = \Phi \circ F \in \mathcal{F}_{\alpha, K_{\min}}$. In the results to follow, we will implicitly assume a link function Φ to be given and fixed, and understand the re-parametrised solution map \mathcal{G} as being defined as in (5) for such choice of Φ .

2.1.3. Measurement model. Define the uniform distribution on \mathcal{O} by $\mu = dx/\text{vol}(\mathcal{O})$, where dx is the Lebesgue measure and $\text{vol}(\mathcal{O}) = \int_{\mathcal{O}} dx$, and consider random design variables

$$(X_i)_{i=1}^N \stackrel{\text{iid}}{\sim} \mu, \quad N \in \mathbb{N}. \quad (6)$$

For unknown $f \in \mathcal{F}_{\alpha, K_{\min}}$, we model the statistical errors under which we observe the corresponding measurements $\{G(f)(X_i)\}_{i=1}^N$ by i.i.d. Gaussian random variables $W_i \sim N(0, 1)$, all independent of the X_i 's. Using the re-parameterisation $f = \Phi \circ F$ via a given link function from the previous subsection, the observation scheme is then

$$Y_i = \mathcal{G}(F)(X_i) + \sigma W_i, \quad i = 1, \dots, N, \quad (7)$$

where $\sigma > 0$ is the noise amplitude. We will often use the shorthand notation $Y^{(N)} = (Y_i)_{i=1}^N$, with analogous definitions for $X^{(N)}$ and $W^{(N)}$. The random vectors (Y_i, X_i) on $\mathbb{R} \times \mathcal{O}$ are then i.i.d with laws denoted as P_F^i . Writing dy for the Lebesgue measure on \mathbb{R} , it follows that P_F^i has Radon–Nikodym density

$$p_F(y, x) := \frac{dP_F^i}{dy \times d\mu}(y, x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[y - \mathcal{G}(F)(x)]^2/(2\sigma^2)}, \quad y \in \mathbb{R}, x \in \mathcal{O}. \quad (8)$$

We will write $P_F^N = \otimes_{i=1}^N P_F^i$ for the joint law of $(Y^{(N)}, X^{(N)})$ on $\mathbb{R}^N \times \mathcal{O}^N$, with E_F^i , E_F^N the expectation operators corresponding to the laws P_F^i , P_F^N respectively. In the sequel we sometimes use the notation P_f^N instead of P_F^N when convenient.

2.1.4. The Bayesian approach. In the Bayesian approach one models the parameter $F \in H_c^\alpha(\mathcal{O})$ by a Borel probability measure Π supported in the Banach space $C(\mathcal{O})$. Since the map $(F, (y, x)) \mapsto p_F(y, x)$ can be shown to be jointly measurable, the posterior distribution $\Pi(\cdot | Y^{(N)}, X^{(N)})$ of $F | (Y^{(N)}, X^{(N)})$ arising from data in model (7) equals, by Bayes' formula (p 7, Ghosal and van der Vaart 2017),

$$\Pi(B | Y^{(N)}, X^{(N)}) = \frac{\int_B e^{\ell^{(N)}(F)} d\Pi(F)}{\int_{C(\mathcal{O})} e^{\ell^{(N)}(F)} d\Pi(F)} \quad \text{any Borel set } B \subseteq C(\mathcal{O}), \quad (9)$$

where

$$\ell^{(N)}(F) = -\frac{1}{2\sigma^2} \sum_{i=1}^N [Y_i - \mathcal{G}(F)(X_i)]^2 \quad (10)$$

is (up to an additive constant) the joint log-likelihood function.

2.2. Statistical convergence rates

In this section we will show that the posterior distribution arising from certain priors concentrates near any sufficiently regular ground truth F_0 (or, equivalently, f_0), and provide a bound on the rate of this contraction, assuming the observation $(Y^{(N)}, X^{(N)})$ to be generated through model (7) of law $P_{F_0}^N$. We will regard $\sigma > 0$ as a fixed and known constant; in practice it may be replaced by the estimated sample variance of the Y_i 's.

The priors we will consider are built around a Gaussian process base prior Π' , but to deal with the non-linearity of the inverse problem, some additional regularisation will be required. We first show how this can be done by an N -dependent 'rescaling' step as suggested in Monard *et al* (2020). We then further show that a randomised truncation of a Karhunen–Loeve-type series expansion of the base prior also leads to a consistent, 'fully Bayesian' solution of this inverse problem.

2.2.1. Results with re-scaled Gaussian priors. We will freely use terminology from the basic theory of Gaussian processes and measures, see, for example, Giné and Nickl (2016), chapter 2 for details.

Condition 3. Let $\alpha > 1 + d/2$, $\beta \geq 1$, and let \mathcal{H} be a Hilbert space continuously imbedded into $H_c^\alpha(\mathcal{O})$. Let Π' be a centred Gaussian Borel probability measure on the Banach space $C(\mathcal{O})$ that is supported on a separable measurable linear subspace of $C^\beta(\mathcal{O})$, and assume that the reproducing-kernel Hilbert space (RKHS) of Π' equals \mathcal{H} .

As a basic example of a Gaussian base prior Π' satisfying condition 3, consider a Whittle–Matérn process $M = \{M(x), x \in \mathcal{O}\}$ indexed by \mathcal{O} and of regularity α (cf. example 25 below for full details). We will assume that it is known that $F_0 \in H^\alpha(\mathcal{O})$ is supported inside a given compact subset K of the domain \mathcal{O} , and fix any smooth cut-off function $\chi \in C_c^\infty(\mathcal{O})$ such that $\chi = 1$ on K . Then, $\Pi' = \mathcal{L}(\chi M)$ is supported on the separable linear subspace $C^{\beta'}(\mathcal{O})$ of $C^\beta(\mathcal{O})$ for any $\beta < \beta' < \alpha - d/2$, and its RKHS $\mathcal{H} = \{\chi F, F \in H^\alpha(\mathcal{O})\}$ is continuously imbedded into $H_c^\alpha(\mathcal{O})$ (and contains $H_K^\alpha(\mathcal{O})$). The condition $F_0 \in \mathcal{H}$ that is employed in the following theorems then amounts to the standard assumption that $F_0 \in H^\alpha(\mathcal{O})$ be supported in a strict subset K of \mathcal{O} .

To proceed, if Π' is as above and $F' \sim \Pi'$, we consider the 're-scaled' prior

$$\Pi_N = \mathcal{L}(F_N), \quad F_N = \frac{1}{N^{d/(4\alpha+4+2d)}} F', \quad (11)$$

Then Π_N again defines a centred Gaussian prior on $C(\mathcal{O})$, and a basic calculation (e.g., exercise 2.6.5 in Giné and Nickl (2016)) shows that its RKHS \mathcal{H}_N is still given by \mathcal{H} but now with norm

$$\|F\|_{\mathcal{H}_N} = N^{d/(4\alpha+4+2d)} \|F\|_{\mathcal{H}} \quad \forall F \in \mathcal{H}. \quad (12)$$

Our first result shows that the posterior contracts towards F_0 in ‘prediction’-risk at rate $N^{-(\alpha+1)/(2\alpha+2+d)}$ and that, moreover, the posterior draws possess a bound on their C^β -norm with overwhelming frequentist probability.

Theorem 4. *For fixed integer $\alpha > \beta + d/2$, $\beta \geq 1$, consider the Gaussian prior Π_N in (11) with base prior $F' \sim \Pi'$ satisfying condition 3 for RKHS \mathcal{H} . Let $\Pi_N(\cdot|Y^{(N)}, X^{(N)})$ be the resulting posterior distribution arising from observations $(Y^{(N)}, X^{(N)})$ in (7), set $\delta_N = N^{-(\alpha+1)/(2\alpha+2+d)}$, and assume $F_0 \in \mathcal{H}$.*

Then for any $D > 0$ there exists $L > 0$ large enough (depending on $\sigma, F_0, D, \alpha, \beta$, as well as on \mathcal{O}, d, g) such that, as $N \rightarrow \infty$,

$$\Pi_N(F : \|\mathcal{G}(F) - \mathcal{G}(F_0)\|_{L^2} > L\delta_N | Y^{(N)}, X^{(N)}) = O_{P_{F_0}^N}(e^{-DN\delta_N^2}), \quad (13)$$

and for sufficiently large $M > 0$ (depending on σ, D, α, β)

$$\Pi_N(F : \|F\|_{C^\beta} > M | Y^{(N)}, X^{(N)}) = O_{P_{F_0}^N}(e^{-DN\delta_N^2}). \quad (14)$$

Following ideas in Monard *et al* (2020), we can combine (13) with the regularisation property (14) and a suitable stability estimate for G^{-1} to show that the posterior contracts about f_0 also in L^2 -risk. We shall employ the stability estimate proved in Nickl *et al* (2020), lemma 24, which requires the source function g in the base PDE (3) to be strictly positive, a natural condition ensuring injectivity of the map $f \mapsto G(f)$, see Richter (1981). Denote the push-forward posterior on the conductivities f by

$$\tilde{\Pi}_N(\cdot|Y^{(N)}, X^{(N)}) := \mathcal{L}(f), \quad f = \Phi \circ F : F \sim \Pi_N(\cdot|Y^{(N)}, X^{(N)}). \quad (15)$$

Theorem 5. *Let $\Pi_N(\cdot|Y^{(N)}, X^{(N)})$, δ_N and F_0 be as in theorem 4 for integer $\beta > 1$. Let $f_0 = \Phi \circ F_0$ and assume in addition that $\inf_{x \in \mathcal{O}} g(x) \geq g_{\min} > 0$. Then for any $D > 0$ there exists $L > 0$ large enough (depending on $\sigma, f_0, D, \alpha, \beta, \mathcal{O}, g_{\min}, d$) such that, as $N \rightarrow \infty$,*

$$\tilde{\Pi}_N(f : \|f - f_0\|_{L^2} > LN^{-\lambda} | Y^{(N)}, X^{(N)}) = O_{P_{f_0}^N}(e^{-DN\delta_N^2}), \quad \lambda = \frac{(\alpha+1)(\beta-1)}{(2\alpha+2+d)(\beta+1)}.$$

We note that as the smoothness α of f_0 increases, we can employ priors of higher regularity α, β . In particular, if $F_0 \in C^\infty = \cap_{\alpha>0} H^\alpha$, we can let the above rate $N^{-\lambda}$ be as closed as desired to the ‘parametric’ rate $N^{-1/2}$.

We conclude this section showing that the posterior mean $E^\Pi[F|Y^{(N)}, X^{(N)}]$ of $\Pi_N(\cdot|Y^{(N)}, X^{(N)})$ converges to F_0 at the rate $N^{-\lambda}$ from theorem 5. We formulate this result at the level of the vector space valued parameter F (instead of for conductivities f), as the most commonly used MCMC algorithms (such as pCN, see subsection 2.3.1) target the posterior distribution of F .

Theorem 6. *Under the hypotheses of theorem 5, let $\bar{F}_N = E^\Pi[F|Y^{(N)}, X^{(N)}]$ be the (Bochner-)mean of $\Pi_N(\cdot|Y^{(N)}, X^{(N)})$. Then, as $N \rightarrow \infty$,*

$$P_{F_0}^N(\|\bar{F}_N - F_0\|_{L^2} > N^{-\lambda}) \rightarrow 0. \quad (16)$$

The same result holds for the implied conductivities, that is, for $\|\Phi \circ \bar{F}_N - f_0\|_{L^2}$ replacing $\|\bar{F}_N - F_0\|_{L^2}$, since composition with Φ is Lipschitz.

2.2.2. Extension to high-dimensional Gaussian sieve priors. It is often convenient, for instance for computational reasons as discussed in subsection 2.3.1, to employ ‘sieve’-priors that are concentrated on a finite-dimensional approximation of the parameter space supporting the prior. For example a truncated Karhunen–Loeve-type series expansion (or some other discretisation) of the Gaussian base prior Π' is frequently used (Dashti and Stuart 2011, Hairer *et al* 2014). The theorems of the previous subsection remain valid if the approximation spaces are appropriately chosen.

Let us illustrate this by considering a Gaussian series prior based on an orthonormal basis $\{\Psi_{\ell r}, \ell \geq -1, r \in \mathbb{Z}^d\}$ of $L^2(\mathbb{R}^d)$ composed of sufficiently regular, compactly supported Daubechies wavelets (see chapter 4 in Giné and Nickl (2016) for details). We assume that $F_0 \in H_K^\alpha(\mathcal{O})$ for some $K \subset \mathcal{O}$, and denote by \mathcal{R}_ℓ the set of indices r for which the support of $\Psi_{\ell r}$ intersects K . Fix any compact $K' \subset \mathcal{O}$ such that $K \subsetneq K'$, and a cut-off function $\chi \in C_c^\infty(\mathcal{O})$ such that $\chi = 1$ on K' . For any real $\alpha > 1 + d/2$, consider the prior Π'_J arising as the law of the Gaussian random sum

$$\Pi'_J = \mathcal{L}(\chi F), \quad F = \sum_{\ell \leq J, r \in \mathcal{R}_\ell} 2^{-\ell\alpha} F_{\ell r} \Psi_{\ell r}, \quad F_{\ell r} \stackrel{\text{iid}}{\sim} N(0, 1), \quad (17)$$

where $J = J_N \rightarrow \infty$ is a (deterministic) truncation point to be chosen. Then Π'_J defines a centred Gaussian prior that is supported on the finite-dimensional space

$$\mathcal{H}_J = \text{span}\{\chi \Psi_{\ell r}, \ell \leq J, r \in \mathcal{R}_\ell\} \subset C(\mathcal{O}). \quad (18)$$

Proposition 7. Consider a prior Π_N as in (11) where now $F' \sim \Pi'_J$ and $J = J_N \in \mathbb{N}$ is such that $2^J \simeq N^{1/(2\alpha+2+d)}$. Let $\Pi_N(\cdot | Y^{(N)}, X^{(N)})$ be the resulting posterior distribution arising from observations $(Y^{(N)}, X^{(N)})$ in (7), and assume $F_0 \in H_K^\alpha(\mathcal{O})$. Then the conclusions of theorems 4–6 remain valid (under the respective hypotheses on α, β, g).

A similar result could be proved for more general Gaussian priors (not of wavelet type), but we refrain from giving these extensions here.

2.2.3. Randomly truncated Gaussian series priors. In this section we show that instead of rescaling Gaussian base priors Π', Π'_J in an N -dependent way to attain extra regularisation, one may also randomise the dimensionality parameter J in (17) by a hyper-prior with suitable tail behaviour. While this is computationally somewhat more expensive (by necessitating a hierarchical sampling method, see subsection 2.3.1), it gives a possibly more principled approach to (‘fully’) Bayesian regularisation in our inverse problem. The theorem below will show that such a procedure is consistent in the frequentist sense, at least for smooth enough F_0 .

For the wavelet basis and cut-off function χ introduced before (17), we consider again a random (conditionally Gaussian) sum

$$\Pi = \mathcal{L}(\chi F), \quad F = \sum_{\ell \leq J, r \in \mathcal{R}_\ell} 2^{-\ell\alpha} F_{\ell r} \Psi_{\ell r}, \quad F_{\ell r} \stackrel{\text{iid}}{\sim} N(0, 1) \quad (19)$$

where now J is a random truncation level, independent of the random coefficients $F_{\ell r}$, satisfying the following inequalities:

$$\Pr(J > j) = e^{-2^{jd} \log 2^{jd}} \forall j \geq 1; \quad \Pr(J = j) \gtrsim e^{-2^{jd} \log 2^{jd}}, \quad j \rightarrow \infty. \quad (20)$$

When $d = 1$, a (log-)Poisson random variable satisfies these tail conditions, and for $d > 1$ such a random variable J can be easily constructed too—see example 28 below.

Our first result in this section shows that the posterior arising from the truncated series prior in (19) achieves (up to a log-factor) the same contraction rate in L^2 -prediction risk as the one obtained in theorem 4. Moreover, as is expected in light of the results in van der Vaart and van Zanten (2009) and Ray (2013), the posterior adapts to the unknown regularity α_0 of F_0 when it exceeds the base smoothness level α .

Theorem 8. *For any $\alpha > 1 + d/2$, let Π be the random series prior in (19), and let $\Pi(\cdot|Y^{(N)}, X^{(N)})$ be the resulting posterior distribution arising from observations $(Y^{(N)}, X^{(N)})$ in (7). Then, for each $\alpha_0 \geq \alpha$ and any $F_0 \in H_K^{\alpha_0}(\mathcal{O})$, we have that for any $D > 0$ there exists $L > 0$ large enough (depending on $\sigma, F_0, D, \alpha, \mathcal{O}, d, g$) such that, as $N \rightarrow \infty$,*

$$\Pi(F : \|\mathcal{G}(F) - \mathcal{G}(F_0)\|_{L^2} > L\xi_N | Y^{(N)}, X^{(N)}) = O_{P_{F_0}^N}(e^{-DN\xi_N^2}),$$

where $\xi_N = N^{-(\alpha_0+1)/(2\alpha_0+2+d)} \log N$. Moreover, for \mathcal{H}_J the finite-dimensional subspaces in (18) and $J_N \in \mathbb{N}$ such that $2^{J_N} \simeq N^{1/(2\alpha_0+2+d)}$, we also have that for sufficiently large $M > 0$ (depending on D, α)

$$\Pi(F : F \in \mathcal{H}_{J_N}, \|F\|_{H^\alpha} \leq M2^{J_N\alpha}N\xi_N^2 | Y^{(N)}, X^{(N)}) = 1 - O_{P_{F_0}^N}(e^{-DN\xi_N^2}). \quad (21)$$

We can now exploit the previous result along with the finite-dimensional support of the posterior and again the stability estimate from Nickl *et al* (2020) to obtain the following consistency theorem for $F_0 \in H^{\alpha_0}$ if α_0 is large enough (with a precise bound $\alpha_0 \geq \alpha^*$ given in the proof of lemma 12).

Theorem 9. *Let the link function Φ in the definition (5) of \mathcal{G} satisfy condition 2. Let $\Pi(\cdot|Y^{(N)}, X^{(N)})$, ξ_N be as in theorem 8, assume in addition $g \geq g_{\min} > 0$ on \mathcal{O} , and let $\tilde{\Pi}(\cdot|Y^{(N)}, X^{(N)})$ be the posterior distribution of f as in (15). Then for $f_0 = \Phi \circ F_0$ with $F_0 \in H_K^{\alpha_0}(\mathcal{O})$ for α_0 large enough (depending on α, d, a) and for any $D > 0$ there exists $L > 0$ large enough (depending on $\sigma, f_0, D, \alpha, \mathcal{O}, g_{\min}, d$) such that, as $N \rightarrow \infty$,*

$$\tilde{\Pi}(f : \|f - f_0\|_{L^2} > LN^{-\rho} | Y^{(N)}, X^{(N)}) = O_{P_{f_0}^N}(e^{-DN\xi_N^2}), \quad \rho = \frac{(\alpha_0 + 1)(\alpha - 1)}{(2\alpha_0 + 2 + d)(\alpha + 1)}.$$

Just as before, for $f_0 \in C^\infty$ the above rate can be made as close as desired to $N^{-1/2}$ by choosing α large enough. Moreover, the last contraction theorem also translates into a convergence result for the posterior mean of F .

Theorem 10. *Under the hypotheses of theorem 9, let $\bar{F}_N = E^\Pi[F|Y^{(N)}, X^{(N)}]$ be the mean of $\Pi(\cdot|Y^{(N)}, X^{(N)})$. Then, as $N \rightarrow \infty$,*

$$P_{F_0}^N(\|\bar{F}_N - F_0\|_{L^2} > N^{-\rho}) \rightarrow 0. \quad (22)$$

We note that the proof of the last two theorems crucially takes advantage of the ‘non-symmetric’ and ‘non-exponential’ nature of the stability estimate from Nickl *et al* (2020), and may not hold in other non-linear inverse problems where such an estimate may not be available (e.g., as in Monard *et al* (2020), Abraham and Nickl (2020) or also in the Schrödinger equation setting studied in Nickl *et al* (2020), Nickl (2018)).

Let us conclude this section by noting that hierarchical priors such as the one studied here are usually devised to ‘adapt to unknown’ smoothness α_0 of F_0 , see van der Vaart and van Zanten (2009) and Ray (2013). Note that while our posterior distribution is adaptive to α_0

in the ‘prediction risk’ setting of theorem 8, the rate $N^{-\rho}$ obtained in theorems 9 and 10 for the inverse problem *does* depend on the minimal smoothness α , and is therefore *not adaptive*. Nevertheless, this hierarchical prior gives an example of a fully Bayesian, consistent solution of our inverse problem.

2.3. Concluding discussion

2.3.1. Posterior computation. As mentioned in the introduction, in the context of the elliptic inverse problem considered in the present paper, posterior distributions arising from Gaussian process priors such as those above can be computed by MCMC algorithms, see Beskos *et al* (2017), Conrad *et al* (2016), Cotter *et al* (2013), and computational guarantees can be obtained as well: for Gaussian priors, Hairer *et al* (2014) establish non-asymptotic sampling bounds for the ‘preconditioned Crank–Nicholson (pCN)’ algorithm, which hold even in the absence of log-concavity of the likelihood function, and which imply bounds on the approximation error for the computation of the posterior mean. The algorithm can be implemented as long as it is possible to evaluate the forward map $F \mapsto \mathcal{G}(F)(x)$ at $x \in \mathcal{O}$, which in our context can be done by using standard numerical methods to solve the elliptic PDE (3). In practice, these algorithms often employ a finite-dimensional approximation of the parameter space (see subsection 2.2.2).

In order to sample from the posterior distribution arising from the more complex hierarchical prior (19), MCMC methods based on fixed Gaussian priors (such as the pCN algorithm) can be employed within a suitable Gibbs-sampling scheme that exploits the conditionally Gaussian structure of the prior. The algorithm would then alternate, for given J , an MCMC step targeting the marginal posterior distribution of $F|(Y^{(N)}, X^{(N)}, J)$, followed by, given the actual sample of F , a second MCMC run with objective the marginal posterior of $J|(Y^{(N)}, X^{(N)}, F)$. A related approach to hierarchical inversion is empirical Bayesian estimation. In the present setting this would entail first estimating the truncation level J from the data, via an estimator $\hat{J} = \hat{J}(Y^{(N)}, X^{(N)})$ (e.g., the marginal maximum likelihood estimator), and then performing inference based on the fixed finite-dimensional prior $\Pi_{\hat{J}}$ (defined as in (19) with J replaced by \hat{J}). See Knapik *et al* (2015) where this is studied in a diagonal linear inverse problem.

2.3.2. Open problems: towards optimal convergence rates. The convergence rates obtained in this article demonstrate the frequentist consistency of a Bayesian (Gaussian process) inversion method in the elliptic inverse problem (2) with data (1) in the large sample limit $N \rightarrow \infty$. While the rates approach the optimal rate $N^{-1/2}$ for very smooth models ($\alpha \rightarrow \infty$), the question of optimality for fixed α remains an interesting avenue for future research. We note that for the ‘PDE-constrained regression’ problem of recovering $\mathcal{G}(F_0)$ in ‘prediction’ loss, the rate $\delta_N = N^{-(\alpha+1)/(2\alpha+2+d)}$ obtained in theorems 4 and 8 can be shown to be minimax optimal (as in Nickl *et al* (2020), theorem 10). But for the recovery rates for f obtained in theorems 6 and 10, no matching lower bounds are currently known. Related to this issue, in Nickl *et al* (2020) faster (but still possibly suboptimal) rates are obtained for the modes of our posterior distributions (MAP estimates, which are not obviously computable in polynomial time), and one may loosely speculate here about computational hardness barriers in our non-linear inverse problem. These issues pose formidable challenges for future research and are beyond the scope of the present paper.

3. Proofs

We assume without loss of generality that $\text{vol}(\mathcal{O}) = 1$. In the proof, we will repeatedly exploit properties of the (re-parametrised) solution map \mathcal{G} defined in (5), which was studied in detail

in Nickl *et al* (2020). Specifically, in the proof of theorem 9 in Nickl *et al* (2020) it is shown that, for all $\alpha > 1 + d/2$ and any $F_1, F_2 \in H_c^\alpha(\mathcal{O})$,

$$\|\mathcal{G}(F_1) - \mathcal{G}(F_2)\|_{L^2(\mathcal{O})} \lesssim (1 + \|F_1\|_{C^1(\mathcal{O})}^4 \vee \|F_2\|_{C^1(\mathcal{O})}^4) \|F_1 - F_2\|_{(H^1(\mathcal{O}))^*}, \quad (23)$$

where we denote by X^* the topological dual Banach space of a normed linear space X . Secondly, we have (lemma 20 in Nickl *et al* (2020)) for some constant $c > 0$ (only depending on d , \mathcal{O} and K_{\min}),

$$\sup_{F \in H_c^\alpha} \|\mathcal{G}(F)\|_\infty \leq c \|g\|_\infty < \infty. \quad (24)$$

Therefore the inverse problem (7) falls in the general framework considered in appendix A below (with $\beta = \kappa = 1$, $\gamma = 4$ in (A2) and $S = c \|g\|_\infty$ in (A3)); in particular theorems 4 and 8 then follow as particular cases of the general contraction rate results derived in theorems 14 and 19, respectively. It thus remains to derive theorems 5 and 6 from theorem 4, and theorems 9 and 10 from theorem 8, respectively.

To do so we recall here another key result from Nickl *et al* (2020), namely their stability estimate lemma 24: for $\alpha > 2 + d/2$, if $G(f)$ denotes the solution of the PDE (3) with g satisfying $\inf_{x \in \mathcal{O}} g(x) \geq g_{\min} > 0$, then for fixed $f_0 \in \mathcal{F}_{\alpha, K_{\min}}$ and all $f \in \mathcal{F}_{\alpha, K_{\min}}$

$$\|f - f_0\|_{L^2(\mathcal{O})} \lesssim \|f\|_{C^1(\mathcal{O})} \|G(f) - G(f_0)\|_{H^2(\mathcal{O})}, \quad (25)$$

with multiplicative constant independent of f .

3.1. Proofs for section 2.2.1

Proof of theorem 5. The conclusions of theorem 4 can readily be translated for the push-forward posterior $\tilde{\Pi}_N(\cdot | Y^{(N)}, X^{(N)})$ from (15). In particular, (13) implies, for $f_0 = \Phi \circ F_0$, as $N \rightarrow \infty$,

$$\tilde{\Pi}_N(f : \|G(f) - G(f_0)\|_{L^2} > L\delta_N | Y^{(N)}, X^{(N)}) = O_{P_{f_0}^N}(e^{-DN\delta_N^2}); \quad (26)$$

and using lemma 29 in Nickl *et al* (2020) and (14) we obtain for sufficiently large $M' > 0$

$$\begin{aligned} \tilde{\Pi}_N(f : \|f\|_{C^\beta} > M' | Y^{(N)}, X^{(N)}) &\leq \Pi_N(F : \|F\|_{C^\beta} > M | Y^{(N)}, X^{(N)}) \\ &= O_{P_{F_0}^N}(e^{-DN\delta_N^2}). \end{aligned} \quad (27)$$

From the previous bounds we now obtain the following result. □

Lemma 11. For $\Pi_N(\cdot | Y^{(N)}, X^{(N)})$, δ_N and F_0 as in theorem 4, let $\tilde{\Pi}_N(\cdot | Y^{(N)}, X^{(N)})$ be the push-forward posterior distribution from (15). Then, for $f_0 = \Phi \circ F_0$ and any $D > 0$ there exists $L > 0$ large enough such that, as $N \rightarrow \infty$,

$$\tilde{\Pi}_N(f : \|G(f) - G(f_0)\|_{H^2} > L\delta_N^{(\beta-1)/(\beta+1)} | Y^{(N)}, X^{(N)}) = O_{P_{F_0}^N}(e^{-DN\delta_N^2}).$$

Proof. Using the continuous imbedding of $C^\beta \subset H^\beta$, $\beta \in \mathbb{N}$, and (27), for some $M' > 0$ as $N \rightarrow \infty$,

$$\tilde{\Pi}_N(f : \|f\|_{H^\beta} > M' | Y^{(N)}, X^{(N)}) = O_{P_{F_0}^N}(e^{-DN\delta_N^2}).$$

Now if $f \in H^\beta$ with $\|f\|_{H^\beta} \leq M'$, lemma 23 in Nickl *et al* (2020) implies $G(f), G(f_0) \in H^{\beta+1}$, with

$$\|G(f_0)\|_{H^{\beta+1}} \lesssim 1 + \|f_0\|_{H^\beta}^{\beta(\beta+1)} < \infty, \quad \|G(f)\|_{H^{\beta+1}} \lesssim 1 + \|f\|_{H^\beta}^{\beta(\beta+1)} < M'' < \infty;$$

and by the usual interpolation inequality for Sobolev spaces (Lions and Magenes 1972),

$$\begin{aligned} \|G(f) - G(f_0)\|_{H^2} &\lesssim \|G(f) - G(f_0)\|_{L^2}^{(\beta-1)/(\beta+1)} \|G(f) - G(f_0)\|_{H^{\beta+1}}^{2/(\beta+1)} \\ &\lesssim \|G(f) - G(f_0)\|_{L^2}^{(\beta-1)/(\beta+1)}. \end{aligned}$$

Thus, by what precedes and (26), for sufficiently large $L > 0$

$$\begin{aligned} \tilde{\Pi}_N(f : \|G(f) - G(f_0)\|_{H^2} > L\delta_N^{(\beta-1)/(\beta+1)} | Y^{(N)}, X^{(N)}) \\ \leq \tilde{\Pi}_N(f : \|G(f) - G(f_0)\|_{L^2} > L'\delta_N | Y^{(N)}, X^{(N)}) \\ + \tilde{\Pi}_N(f : \|f\|_{H^\beta} > M'' | Y^{(N)}, X^{(N)}) = O_{P_{F_0}^N}(\mathrm{e}^{-DN\delta_N^2}), \end{aligned}$$

as $N \rightarrow \infty$. □

To prove theorem 5 we use (25), (27) and lemma 11 to the effect that for any $D > 0$ we can find $L, M > 0$ large enough such that, as $N \rightarrow \infty$,

$$\begin{aligned} \tilde{\Pi}_N(f : \|f - f_0\|_{L^2} > L\delta_N^{\frac{\beta-1}{\beta+1}} | Y^{(N)}, X^{(N)}) \\ \leq \tilde{\Pi}_N(f : \|G(f) - G(f_0)\|_{H^2} > L'\delta_N^{\frac{\beta-1}{\beta+1}} | Y^{(N)}, X^{(N)}) \\ + \tilde{\Pi}_N(f : \|f\|_{C^\beta} > M | Y^{(N)}, X^{(N)}) = O_{P_{F_0}^N}(\mathrm{e}^{-DN\delta_N^2}). \end{aligned}$$

Proof of theorem 6. The proof largely follows ideas of Monard *et al* (2020) but requires a slightly more involved, iterative uniform integrability argument to also control the probability of events $\{F : \|F\|_{C^\beta} > M\}$ on whose complements we can subsequently exploit regularity properties of the inverse link function Φ^{-1} .

Using Jensen's inequality, it is enough to show, as $N \rightarrow \infty$,

$$P_{F_0}^N(E^\Pi[\|F - F_0\|_{L^2}^2 | Y^{(N)}, X^{(N)}] > N^{-\lambda}) \rightarrow 0.$$

For $M > 0$ sufficiently large to be chosen, we decompose

$$\begin{aligned} E^\Pi[\|F - F_0\|_{L^2} | Y^{(N)}, X^{(N)}] &= E^\Pi[\|F - F_0\|_{L^2} 1_{\|F\|_{C^\beta} \leq M} | Y^{(N)}, X^{(N)}] \\ &\quad + E^\Pi[\|F - F_0\|_{L^2} 1_{\|F\|_{C^\beta} > M} | Y^{(N)}, X^{(N)}]. \end{aligned} \quad (28)$$

Using the Cauchy–Schwarz inequality we can upper bound the expectation in the second summand by

$$\sqrt{E^\Pi[\|F - F_0\|_{L^2}^2 | Y^{(N)}, X^{(N)}]} \sqrt{\Pi_N(F : \|F\|_{C^\beta} > M | Y^{(N)}, X^{(N)})}.$$

In view of (14), for all $D > 0$ we can choose $M > 0$ large enough to obtain

$$P_{F_0}^N \left(E^\Pi [\|F - F_0\|_{L^2}^2 | Y^{(N)}, X^{(N)}] \Pi_N(F : \|F\|_{C^\beta} > M | Y^{(N)}, X^{(N)}) > N^{-2\lambda} \right) \\ \leq P_{F_0}^N \left(E^\Pi [\|F - F_0\|_{L^2}^2 | Y^{(N)}, X^{(N)}] e^{-DN\delta_N^2} > N^{-2\lambda} \right) + o(1).$$

To bound the probability in the last line, let \mathcal{B}_N be the sets defined in (A4) below, note that lemmas 16 and 23 below jointly imply that $\Pi_N(\mathcal{B}_N) \geq a e^{-AN\delta_N^2}$ for some $a, A > 0$. Also, let $\nu(\cdot) = \Pi_N(\cdot \cap \mathcal{B}_N) / \Pi_N(\mathcal{B}_N)$, and let \mathcal{C}_N be the event from (A10), for which lemma 7.3.2 in Giné and Nickl (2016) implies that $P_{F_0}^N(\mathcal{C}_N) \rightarrow 1$ as $N \rightarrow \infty$. Then

$$P_{F_0}^N \left(E^\Pi [\|F - F_0\|_{L^2}^2 | Y^{(N)}, X^{(N)}] e^{-DN\delta_N^2} > N^{-2\lambda} \right) \\ \leq P_{F_0}^N \left(\frac{\int_{C(\mathcal{O})} \|F - F_0\|_{L^2}^2 \prod_{i=1}^N p_F / p_{F_0}(Y_i, X_i) d\Pi_N(F)}{\Pi(\mathcal{B}_N) \int_{\mathcal{B}_N} \prod_{i=1}^N p_F / p_{F_0}(Y_i, X_i) d\nu(F)} e^{-DN\delta_N^2} > N^{-2\lambda}, \mathcal{C}_N \right) + o(1) \\ \leq P_{F_0}^N \left(\int_{C(\mathcal{O})} \|F - F_0\|_{L^2}^2 \prod_{i=1}^N p_F / p_{F_0}(Y_i, X_i) d\Pi_N(F) > N^{-2\lambda} a e^{(D-A-2)N\delta_N^2} \right) + o(1)$$

which is upper bounded, using Markov's inequality and Fubini's theorem, by

$$\frac{1}{a} e^{-(D-A-2)N\delta_N^2} N^{2\lambda} \int_{C(\mathcal{O})} \|F - F_0\|_{L^2}^2 E_{F_0}^N \left(\prod_{i=1}^N \frac{p_F}{p_{F_0}}(Y_i, X_i) \right) d\Pi_N(F).$$

Taking $D > A + 2$ (and M large enough in (28)), using the fact that $E_{F_0}^N \left(\prod_{i=1}^N p_F / p_{F_0}(Y_i, X_i) \right) = 1$, and that $E^{\Pi_N} \|F\|_{L^2} < \infty$ (by Fernique's theorem, e.g., Giné and Nickl (2016), exercise 2.1.5), we then conclude

$$P_{F_0}^N \left(E^\Pi [\|F - F_0\|_{L^2}^2 1_{\|F\|_{C^\beta} > M} | Y^{(N)}, X^{(N)}] > N^{-\lambda} \right) \rightarrow 0, \quad N \rightarrow \infty. \quad (29)$$

To handle the first term in (28), let $f = \Phi \circ F$ and $f_0 = \Phi \circ F_0$. Then for all $x \in \mathcal{O}$, by the mean value and inverse function theorems,

$$|F(x) - F_0(x)| = |\Phi^{-1} \circ f(x) - \Phi^{-1} \circ f_0(x)| = \frac{1}{|\Phi'(\Phi^{-1}(\eta))|} |f(x) - f_0(x)|$$

for some η lying between $f(x)$ and $f_0(x)$. If $\|F\|_{C^\beta} \leq M$ then, as Φ is strictly increasing, necessarily $f(x) = \Phi(F(x)) \in [\Phi(-M), \Phi(M)]$ for all $x \in \mathcal{O}$. Similarly, the range of f_0 is contained in the compact interval $[\Phi(-M), \Phi(M)]$ for $M \geq \|F_0\|_\infty$, so that

$$|\Phi^{-1} \circ f(x) - \Phi^{-1} \circ f_0(x)| \leq \frac{1}{\min_{z \in [-M, M]} \Phi'(z)} |f(x) - f_0(x)| \lesssim |f(x) - f_0(x)|$$

for a multiplicative constant not depending on $x \in \mathcal{O}$. It follows

$$\|F - F_0\|_{L^2} 1_{\|F\|_{C^\beta} \leq M} \lesssim \|f - f_0\|_{L^2} 1_{\|F\|_{C^\beta} \leq M},$$

and

$$E^\Pi [\|F - F_0\|_{L^2} 1_{\|F\|_{C^\beta} \leq M} | Y^{(N)}, X^{(N)}] \lesssim E^{\tilde{\Pi}} [\|f - f_0\|_{L^2} | Y^{(N)}, X^{(N)}].$$

Noting that for each $L > 0$ the last expectation is upper bounded by

$$\begin{aligned} & LN^{-\lambda} + E^{\tilde{\Pi}} \left[\|f - f_0\|_{L^2} 1_{\|f - f_0\|_{L^2} > LN^{-\lambda}} | Y^{(N)}, X^{(N)} \right] \\ & \leq LN^{-\lambda} + \sqrt{E^{\tilde{\Pi}}[\|f - f_0\|_{L^2}^2 | Y^{(N)}, X^{(N)}]} \sqrt{\tilde{\Pi}_N(f : \|f - f_0\|_{L^2} > LN^{-\lambda} | Y^{(N)}, X^{(N)})}, \end{aligned}$$

we can repeat the above argument, with the event $\{F : \|F\|_{C^\beta} > M\}$ replaced by the event $\{f : \|f - f_0\|_{L^2} > LN^{-\lambda}\}$, to deduce from theorem 5 that for $D > A + 2$ there exists $L > 0$ large enough such that

$$\begin{aligned} & P_{F_0}^N \left(E^{\tilde{\Pi}}[\|f - f_0\|_{L^2}^2 | Y^{(N)}, X^{(N)}] \tilde{\Pi}_N(f : \|f - f_0\|_{L^2} > LN^{-\lambda} | Y^{(N)}, X^{(N)}) > N^{-\lambda} \right) \\ & \lesssim e^{-(D-A-2)N\delta_N^2} N^{2\lambda} \end{aligned}$$

which combined with (29) and the definition of δ_N concludes the proof. \square

3.2. Sieve prior proofs

The proof only requires minor modification from the proofs of section 2.2.1. We only discuss here the main points: one first applies the L^2 -prediction risk theorem 14 with a sieve prior. In the proof of the small ball lemma 16 one uses the following observations: the projection $P_{\mathcal{H}_J}(F_0) \in \mathcal{H}_J$ of $F_0 \in H_K^\alpha$ defined in (B4) satisfies by (B6)

$$\|F_0 - P_{\mathcal{H}_J}(F_0)\|_{(H^1(\mathcal{O}))^*} \lesssim 2^{-J(\alpha+1)},$$

hence choosing J such that $2^J \simeq N^{1/(2\alpha+2+d)}$, and noting also that $\|P_{\mathcal{H}_J}(F_0)\|_{C^1} \leq \|F_0\|_{C^1} < \infty$ for all J by standard properties of wavelet bases, it follows from (23) that

$$\|\mathcal{G}(F_0) - \mathcal{G}(P_{\mathcal{H}_J}(F_0))\|_{L^2} \lesssim \|F_0 - P_{\mathcal{H}_J}(F_0)\|_{(H^1)^*} \lesssim N^{-(\alpha+1)/(2\alpha+2+d)} = \delta_N.$$

Therefore, by the triangle inequality,

$$\Pi_N(F : \|\mathcal{G}(F) - \mathcal{G}(F_0)\|_{L^2} \geq q\delta_N) \geq \Pi_N(F : \|\mathcal{G}(F) - \mathcal{G}(P_{\mathcal{H}_N}(F_0))\|_{L^2} \geq q'\delta_N).$$

The rest of the proof of lemma 16 then carries over (with $P_{\mathcal{H}_J}(F_0)$ replacing F_0), upon noting that (B3) and a Sobolev imbedding imply

$$\sup_{J \in \mathbb{N}} E^{\Pi_J} \|F\|_{C^1}^2 < \infty, \text{ as well as } \|F\|_{H^\alpha} \leq c \|F\|_{\mathcal{H}_J} \text{ for all } F \in \mathcal{H}_J$$

for some constant $c > 0$ independent of J . Moreover, the last two properties are sufficient to prove an analogue of lemma 17 as well, so that theorem 14 indeed applies to the sieve prior. The proof from here onwards is identical to the ones of theorems 4–6 for the unsieved case, using also that what precedes implies that $\sup_J E^{\Pi_J} \|F\|_{L^2}^2 < \infty$, relevant in the proof of convergence of the posterior mean.

3.3. Proofs for section 2.2.3

Inspection of the proofs for rescaled priors implies that theorems 9 and 10 can be deduced from theorem 8 if we can show that posterior draws lie in a α -Sobolev ball of fixed radius with sufficiently high frequentist probability. This is the content of the next result.

Lemma 12. *Under the hypotheses of theorem 9, there exists $\alpha^* > 0$ (depending on α, d and a) such that for each $F_0 \in H_K^{\alpha_0}(\mathcal{O})$, $\alpha_0 > \alpha^*$, and any $D > 0$ we can find $M > 0$ large enough such that, as $N \rightarrow \infty$,*

$$\Pi(F : \|F\|_{H^\alpha} \leq M | Y^{(N)}, X^{(N)}) = 1 - O_{P_{F_0}^N}(e^{-DN\xi_N^2}).$$

Proof. Theorem 8 implies that for all $D > 0$ and sufficiently large $L, M > 0$, if $J_N \in \mathbb{N} : 2^{J_N} \simeq N^{1/(2\alpha_0+2+d)}$ and denoting by

$$\mathcal{A}_N = \{F \in \mathcal{H}_{J_N} : \|F\|_{H^\alpha} \leq M2^{J_N\alpha}\sqrt{N}\xi_N, \|\mathcal{G}(F) - \mathcal{G}(F_0)\|_{L^2} \leq L\xi_N\},$$

then as $N \rightarrow \infty$

$$\Pi(\mathcal{A}_N | Y^{(N)}, X^{(N)}) = 1 - O_{P_{F_0}^N}(e^{-DN\xi_N^2}). \quad (30)$$

Next, note that if $F \in \mathcal{H}_{J_N}$, then by standard properties of wavelet bases (cf. (63)), $\|F\|_{H^\alpha} \lesssim 2^{J_N\alpha}\|F\|_{L^2}$ for all N large enough. Thus, for $P_{\mathcal{H}_{J_N}}(F_0)$ the projection of F_0 onto \mathcal{H}_{J_N} defined in (B4),

$$\|F\|_{H^\alpha} \leq \|F - P_{\mathcal{H}_{J_N}}(F_0)\|_{H^\alpha} + \|P_{\mathcal{H}_{J_N}}(F_0)\|_{H^\alpha} \leq 2^{J_N\alpha}\|F - F_0\|_{L^2} + \|F_0\|_{H^\alpha},$$

and a Sobolev imbedding further gives $\|F\|_{L^\infty} \leq M'2^{J_N\alpha}\sqrt{N}\xi_N$, for some $M' > 0$. Now letting $f = \Phi \circ F$ and $f_0 = \Phi \circ F_0$, by similar argument as in the proof of theorem 6 combined with monotonicity of Φ' , we see that for all N large enough

$$\|F - F_0\|_{L^2} \leq \frac{1}{\Phi'(-M'2^{J_N\alpha}\sqrt{N}\xi_N)}\|f - f_0\|_{L^2}.$$

Then, using the assumption on the left tail of Φ in condition 2, and the stability estimate (25),

$$\|F - F_0\|_{L^2} \lesssim (2^{J_N\alpha}\sqrt{N}\xi_N)^a \|f\|_{H^\alpha} \|G(f) - G(f_0)\|_{H^2}.$$

Finally, by the interpolation inequality for Sobolev spaces (Lions and Magenes 1972) and lemma 23 in Nickl *et al* (2020),

$$\begin{aligned} \|G(f) - G(f_0)\|_{H^2} &\lesssim \|G(f) - G(f_0)\|_{L^2}^{(\alpha-1)/(\alpha+1)} \|G(f) - G(f_0)\|_{H^{\alpha+1}}^{2/(\alpha+1)} \\ &\lesssim \xi_N^{(\alpha-1)/(\alpha+1)} (\|G(f)\|_{H^{\alpha+1}} + \|G(f_0)\|_{H^{\alpha+1}})^{2/(\alpha+1)} \\ &\lesssim \xi_N^{(\alpha-1)/(\alpha+1)} (1 + \|f\|_{H^\alpha}^{a^2+\alpha})^{2/(\alpha+1)}, \end{aligned}$$

so that, in conclusion, for each $F \in \mathcal{A}_N$ and sufficiently large N ,

$$\|F\|_{H^\alpha} \lesssim 1 + 2^{J_N\alpha} (2^{J_N\alpha}\sqrt{N}\xi_N)^a \|f\|_{H^\alpha} \xi_N^{\frac{\alpha-1}{\alpha+1}} (1 + \|f\|_{H^\alpha}^{a^2+\alpha})^{\frac{2}{\alpha+1}}.$$

The last term is bounded, using lemma 29 in Nickl *et al* (2020), by a multiple of

$$\xi_N^{\frac{\alpha-1}{\alpha+1}} 2^{J_N \alpha} (2^{J_N \alpha} \sqrt{N} \xi_N)^{2\alpha^2+2\alpha+a} = N^{-\frac{(\alpha_0+1)(\alpha-1)}{(2\alpha_0+2+d)(\alpha+1)}} N^{\frac{2\alpha^3+(2+d)\alpha^2+(1+a+d)\alpha+ad/2}{2\alpha_0+2+d}}$$

the last identity holding up to a log factor. Hence, if

$$\alpha_0 > \alpha^* := \frac{[2\alpha^3 + (2+d)\alpha^2 + (1+a+d)\alpha + ad/2](\alpha+1)}{(\alpha-1)}$$

then we conclude overall that $\|F\|_{H^\alpha} \lesssim 1 + o(1)$ as $N \rightarrow \infty$ for all $F \in \mathcal{A}_N$, proving the claim in view of (30). \square

Replacing β by α in the conclusion of lemma 11, the proof of theorem 9 now proceeds as in the proof of theorem 5 without further modification. Likewise, theorem 10 can be shown following the same argument as in the proof of theorem 6, noting that for Π the random series prior in (19), it also holds that $E^\Pi \|F\|_{L^2}^2 < \infty$.

Acknowledgments

The authors are grateful to three anonymous referees for critical remarks and suggestions, as well as to Sven Wang for helpful discussions. We further acknowledge support by the European Research Council under ERC Grant agreement No. 647812 (UQMSI).

Appendix A. Results for general inverse problems

Let $\mathcal{O} \subset \mathbb{R}^d$, $d \in \mathbb{N}$, be a nonempty open and bounded set with smooth boundary, and assume that \mathcal{D} is a nonempty and bounded measurable subset of \mathbb{R}^p , $p \geq 1$. Let $\mathcal{F} \subseteq L^2(\mathcal{O})$ be endowed with the trace Borel σ -field of $L^2(\mathcal{O})$, and consider a Borel-measurable ‘forward mapping’

$$\mathcal{G}: \mathcal{F} \rightarrow L^2(\mathcal{D}), \quad F \mapsto \mathcal{G}(F).$$

For $F \in \mathcal{F}$, we are given noisy discrete measurements of $\mathcal{G}(F)$ over a grid of points drawn uniformly at random on \mathcal{D} ,

$$Y_i = \mathcal{G}(F)(X_i) + \sigma W_i, \quad i = 1, \dots, N, \quad X_i \stackrel{\text{iid}}{\sim} \mu, \quad W_i \stackrel{\text{iid}}{\sim} N(0, 1), \quad (\text{A1})$$

for some $\sigma > 0$. Above μ denotes the uniform (probability) distribution on \mathcal{D} and the design variables $(X_i)_{i=1}^N$ are independent of the noise vector $(W_i)_{i=1}^N$. We assume without loss of generality that $\text{vol}(\mathcal{D}) = 1$, so that $\mu = dx$, the Lebesgue measure on \mathcal{D} .

We take the noise amplitude $\sigma > 0$ in (A1) to be fixed and known, and work under the assumption that the forward map \mathcal{G} satisfies the following local Lipschitz condition: for given $\beta, \gamma, \kappa \geq 0$, and all $F_1, F_2 \in C^\beta(\mathcal{O}) \cap \mathcal{F}$,

$$\|\mathcal{G}(F_1) - \mathcal{G}(F_2)\|_{L^2(\mathcal{D})} \lesssim (1 + \|F_1\|_{C^\beta(\mathcal{O})}^\gamma \vee \|F_2\|_{C^\beta(\mathcal{O})}^\gamma) \|F_1 - F_2\|_{(H^\kappa(\mathcal{O}))^*} \quad (\text{A2})$$

where we recall that X^* denotes the topological dual Banach space of a normed linear space X . Additionally, we will require \mathcal{G} to be uniformly bounded on its domain:

$$S := \sup_{F \in \mathcal{F}} \|\mathcal{G}(F)\|_{L^\infty(\mathcal{D})} < \infty. \quad (\text{A3})$$

As observed in (23), the elliptic inverse problem considered in this paper falls in this general framework, which also encompasses other examples of nonlinear inverse problems such as those involving the Schrödinger equation considered in Nickl *et al* (2020) and Nickl (2018), for which the results in this section would apply as well. It also includes many linear inverse problems such as the classical Radon transform, see Nickl *et al* (2020).

A.1. General contraction rates in Hellinger distance

Using the same notation as in section 2.1.2, and given a sequence of Borel prior probability measures Π_N on \mathcal{F} , we write $\Pi_N(\cdot|Y^{(N)}, X^{(N)})$ for the posterior distribution of $F|Y^{(N)}, X^{(N)}$ (arising as after (9) and (10)). We also continue to use the notation p_F for the densities from (8) now in the general observation model (A1) (and implicitly assume that the map $(F, (y, x)) \mapsto p_F(y, x)$ is jointly measurable to ensure existence of the posterior distribution). Below we formulate a general contraction theorem in the Hellinger distance that forms the basis of the proofs of the main results. It closely follows the general theory in Ghosal and van der Vaart (2017) and its adaptation to the inverse problem setting in Monard *et al* (2020)—we include a proof for conciseness and convenience of the reader.

Define the Hellinger distance $h(\cdot, \cdot)$ on the set of probabilities density functions on $\mathbb{R} \times \mathcal{D}$ (with respect to the product measure $\mathrm{d}y \times \mathrm{d}x$) by

$$h^2(p_1, p_2) := \int_{\mathbb{R} \times \mathcal{D}} \left[\sqrt{p_1(y, x)} - \sqrt{p_2(y, x)} \right]^2 \mathrm{d}y \mathrm{d}x.$$

For any set \mathcal{A} of such densities, let $N(\eta; \mathcal{A}, h)$, $\eta > 0$, be the minimal number of Hellinger balls of radius η needed to cover \mathcal{A} .

Theorem 13. *Let Π_N be a sequence of prior Borel probability measures on \mathcal{F} , and let $\Pi_N(\cdot|Y^{(N)}, X^{(N)})$ be the resulting posterior distribution arising from observations $(Y^{(N)}, X^{(N)})$ in model (A1). Assume that for some fixed $F_0 \in \mathcal{F}$, and a sequence $\delta_N > 0$ such that $\delta_N \rightarrow 0$ and $\sqrt{N}\delta_N \rightarrow \infty$ as $N \rightarrow \infty$, the sets*

$$\mathcal{B}_N := \left\{ F : E_{F_0}^1 \left[\log \frac{p_{F_0}(Y_1, X_1)}{p_F(Y_1, X_1)} \right] \leq \delta_N^2, E_{F_0}^1 \left[\log \frac{p_{F_0}(Y_1, X_1)}{p_F(Y_1, X_1)} \right]^2 \leq \delta_N^2 \right\}, \quad (\text{A4})$$

satisfy for all N large enough

$$\Pi_N(\mathcal{B}_N) \geq a e^{-AN\delta_N^2}, \quad \text{some } a, A > 0. \quad (\text{A5})$$

Further assume that there exists a sequence of Borel sets $\mathcal{A}_N \subset \mathcal{F}$ for which

$$\Pi_N(\mathcal{A}_N^c) \lesssim e^{-BN\delta_N^2}, \quad \text{some } B > A + 2 \quad (\text{A6})$$

for all N large enough, as well as

$$\log N(\delta_N; \mathcal{A}_N, h) \leq CN\delta_N^2, \quad \text{some } C > 0. \quad (\text{A7})$$

Then, for sufficiently large $L = L(B, C) > 4$ such that $L^2 > 12(B \vee C)$, and all $0 < D < B - A - 2$, as $N \rightarrow \infty$,

$$\Pi_N(F \in \mathcal{A}_N : h(p_F, p_{F_0}) \leq L\delta_N | Y^{(N)}, X^{(N)}) = 1 - O_{P_{F_0}^N}(e^{-DN\delta_N^2}). \quad (\text{A8})$$

Proof. We start noting that by theorem 7.1.4 in Giné and Nickl (2016), for each $L > 4$ satisfying $L^2 > 12(B \vee C)$ we can find tests (random indicator functions) $\Psi_N = \Psi_N(Y^{(N)}, X^{(N)})$ such that as $N \rightarrow \infty$

$$E_{F_0}^N \Psi_N \rightarrow 0, \quad \sup_{F \in \mathcal{A}_N : h(p_F, p_{F_0}) \geq L\delta_N} E_F^N (1 - \Psi_N) \leq e^{-BN\delta_N^2}. \quad (\text{A9})$$

Next, denote the set whose posterior probability we want to lower bound by

$$\tilde{\mathcal{A}}_N = \{F \in \mathcal{A}_N : h(p_F, p_{F_0}) \leq L\delta_N\},$$

and, using the first display in (A9), decompose the probability of interest as

$$\begin{aligned} P_{F_0}^N \left(\Pi_N(\tilde{\mathcal{A}}_N^c | Y^{(N)}, X^{(N)}) \geq e^{-DN\delta_N^2} \right) &= P_{F_0}^N \left(\Pi_N(\tilde{\mathcal{A}}_N^c | Y^{(N)}, X^{(N)}) \geq e^{-DN\delta_N^2}, \Psi_N = 0 \right) \\ &\quad + P_{F_0}^N \left(\Pi_N(\tilde{\mathcal{A}}_N^c | Y^{(N)}, X^{(N)}) \geq e^{-DN\delta_N^2}, \Psi_N = 1 \right) \\ &= P_{F_0}^N \left(\Pi_N(\tilde{\mathcal{A}}_N^c | Y^{(N)}, X^{(N)}) \geq e^{-DN\delta_N^2}, \Psi_N = 0 \right) + o(1). \end{aligned}$$

Next, let $\nu(\cdot) = \Pi_N(\cdot \cap \mathcal{B}_N) / \Pi_N(\mathcal{B}_N)$ be the restricted normalised prior on \mathcal{B}_N , and define the event

$$\mathcal{C}_N = \left\{ \int_{\mathcal{B}_N} \prod_{i=1}^N \frac{p_F}{p_{F_0}}(Y_i, X_i) d\nu(F) \geq e^{-2N\delta_N^2} \right\}, \quad (\text{A10})$$

for which lemma 7.3.2 in Giné and Nickl (2016) implies that $P_{F_0}^N(\mathcal{C}_N) \rightarrow 1$ as $N \rightarrow \infty$. We then further decompose

$$\begin{aligned} P_{F_0}^N \left(\Pi_N(\tilde{\mathcal{A}}_N^c | Y^{(N)}, X^{(N)}) \geq e^{-DN\delta_N^2}, \Psi_N = 0 \right) \\ = P_{F_0}^N \left(\Pi_N(\tilde{\mathcal{A}}_N^c | Y^{(N)}, X^{(N)}) \geq e^{-DN\delta_N^2}, \Psi_N = 0, \mathcal{C}_N \right) + o(1) \end{aligned}$$

and in view of condition (A5) and the above definition of \mathcal{C}_N , we see that

$$\begin{aligned} P_{F_0}^N \left(\Pi_N(\tilde{\mathcal{A}}_N^c | Y^{(N)}, X^{(N)}) \geq e^{-DN\delta_N^2}, \Psi_N = 0, \mathcal{C}_N \right) \\ = P_{F_0}^N \left(\frac{\int_{\tilde{\mathcal{A}}_N^c} \prod_{i=1}^N p_F / p_{F_0}(Y_i, X_i) d\Pi_N(F)}{\int_{\mathcal{F}} \prod_{i=1}^N p_F / p_{F_0}(Y_i, X_i) d\Pi_N(F)} \geq e^{-DN\delta_N^2}, \Psi_N = 0, \mathcal{C}_N \right) \\ \leq P_{F_0}^N \left(\frac{\int_{\tilde{\mathcal{A}}_N^c} (1 - \Psi_N) \prod_{i=1}^N p_F / p_{F_0}(Y_i, X_i) d\Pi_N(F)}{\int_{\mathcal{B}_N} \prod_{i=1}^N p_F / p_{F_0}(Y_i, X_i) d\nu(F)} \geq \Pi_N(\mathcal{B}_N) e^{-DN\delta_N^2}, \mathcal{C}_N \right) \\ \leq P_{F_0}^N \left(\int_{\tilde{\mathcal{A}}_N^c} (1 - \Psi_N) \prod_{i=1}^N \frac{p_F}{p_{F_0}}(Y_i, X_i) d\Pi_N(F) \geq a e^{-(A+D+2)N\delta_N^2} \right). \end{aligned}$$

We conclude applying Markov's inequality and Fubini's theorem, jointly with the fact that for all $F \in \mathcal{F}$

$$E_{F_0}^N \left[\left(1 - \Psi_N\right) \prod_{i=1}^N \frac{p_F}{p_{F_0}}(Y_i, X_i) \right] = E_{F_0}^N \left[\left(1 - \Psi_N\right) \prod_{i=1}^N \frac{dP_F^1}{dP_{F_0}^1}(Y_i, X_i) \right] = E_F^N[1 - \Psi_N],$$

to upper bound the last probability by

$$\begin{aligned} & \frac{1}{a} e^{(A+D+2)N\delta_N^2} \left(\int_{\mathcal{A}_N^c} E_F^N[1 - \Psi_N] d\Pi_N(F) + \int_{\{F \in \mathcal{A}_N : h(p_{F_0}, p_F) > L\delta_N\}} E_F^N[1 - \Psi_N] d\Pi_N(F) \right. \\ & \quad \left. + \int_{\{F \in \mathcal{A}_N^c : h(p_{F_0}, p_F) > L\delta_N\}} E_F^N[1 - \Psi_N] d\Pi_N(F) \right) \\ & \leq \frac{1}{a} e^{(A+D+2)N\delta_N^2} \left(2\Pi_N(\mathcal{A}_N^c) + \int_{\{F \in \mathcal{A}_N : h(p_{F_0}, p_F) > L\delta_N\}} E_F^N[1 - \Psi_N] d\Pi_N(F) \right) \\ & \lesssim e^{-(B-A-D-2)N\delta_N^2} = o(1) \end{aligned}$$

as $N \rightarrow \infty$ since $B > A + D + 2$, having used the excess mass condition (A6) and the second display in (A9). \square

A.2. Contraction rates for rescaled Gaussian priors

While the previous result assumed a general sequence of priors, we now derive explicit contraction rates in L^2 -prediction risk for the specific choices of priors considered in section 2.2. We start with the 're-scaled' priors of section 2.2.1.

Theorem 14. *Let the forward map \mathcal{G} satisfy (A2) and (A3) for given $\beta, \gamma, \kappa \geq 0$ and $S > 0$. For integer $\alpha > \beta + d/2$, consider a Gaussian prior Π_N constructed as in (11) with scaling $N^{d/(4\alpha+4\kappa+2d)}$ and with base prior $F' \sim \Pi'$ satisfying condition 3 with RKHS \mathcal{H} . Let $\Pi_N(\cdot | Y^{(N)}, X^{(N)})$ be the resulting posterior arising from observations $(Y^{(N)}, X^{(N)})$ in (A1), assume $F_0 \in \mathcal{H}$ and set $\delta_N = N^{-(\alpha+\kappa)/(2\alpha+2\kappa+d)}$.*

Then for any $D > 0$ there exists $L > 0$ large enough (depending on σ, F_0, D, α , and $\beta, \gamma, \kappa, S, d$) such that, as $N \rightarrow \infty$,

$$\Pi_N(F : \|\mathcal{G}(F) - \mathcal{G}(F_0)\|_{L^2(\mathcal{D})} > L\delta_N | Y^{(N)}, X^{(N)}) = O_{P_{F_0}^N}(e^{-DN\delta_N^2}), \quad (\text{A11})$$

and for sufficiently large $M > 0$ (depending on $\sigma, D, \alpha, \beta, \gamma, \kappa, d$)

$$\Pi_N(F : \|F\|_{C^\beta} > M | Y^{(N)}, X^{(N)}) = O_{P_{F_0}^N}(e^{-DN\delta_N^2}). \quad (\text{A12})$$

Remark 15. Inspection of the proof shows that if $\kappa = 0$ in (A12), then the RKHS \mathcal{H} in condition 3 can be assumed to be continuously imbedded in $H^\alpha(\mathcal{O})$ only instead of $H_c^\alpha(\mathcal{O})$. The same remark in fact applies for $\kappa < 1/2$.

Proof. In view of the boundedness assumption (A3) on \mathcal{G} , we have by lemma 23 below that for some $q > 0$ (depending on σ, S)

$$E_{F_0}^1 \left[\log \frac{p_{F_0}(Y_1, X_1)}{p_F(Y_1, X_1)} \right] \vee E_{F_0}^1 \left[\log \frac{p_{F_0}(Y_1, X_1)}{p_F(Y_1, X_1)} \right]^2 \leq q \|\mathcal{G}(F_0) - \mathcal{G}(F)\|_{L^2(\mathcal{D})}^2.$$

Hence, for \mathcal{B}_N the sets from (A4) we have $\{F : \|\mathcal{G}(F_0) - \mathcal{G}(F)\|_{L^2(\mathcal{D})} \leq \delta_N/q\} \subseteq \mathcal{B}_N$, which in turn implies the small ball condition (A5) since by lemma 16 below (premultiplying if needed δ_N by a sufficiently large but fixed constant):

$$\Pi_N(F : \|\mathcal{G}(F) - \mathcal{G}(F_0)\|_{L^2(\mathcal{D})} \leq \delta_N/q) \gtrsim e^{-AN\delta_N^2}$$

for some $A > 0$ and all N large enough. Next, for all $D > 0$ and any $B > A + D + 2$, we can choose sets \mathcal{A}_N as in lemmas 17 and 18 and verify the excess mass condition (A6) as well as the complexity bound (A7). Note that $\|F\|_{C^\beta} \leq M$ for all $F \in \mathcal{A}_N$. We then conclude by theorem 13 that for some $L' > 0$ large enough

$$\Pi_N(F \in \mathcal{A}_N : h(p_F, p_{F_0}) \leq L'\delta_N |Y^{(N)}, X^{(N)}) = 1 - O_{P_{F_0}^N}(e^{-DN\delta_N^2})$$

yielding the claim for some appropriate $L > 0$ using the first inequality in (A26). \square

The following key lemma shows that the (non-Gaussian) prior induced on the regression functions $\mathcal{G}(F)$ assigns sufficient mass to a L^2 -neighbourhood of $\mathcal{G}(F_0)$.

Lemma 16. *Let Π_N, F_0 and δ_N be as in theorem 14. Then, for all sufficiently large $q > 0$ there exists $A > 0$ (depending on $q, F_0, \alpha, \beta, \gamma, \kappa, d$) such that*

$$\Pi_N(F : \|\mathcal{G}(F) - \mathcal{G}(F_0)\|_{L^2(\mathcal{D})} \leq q\delta_N) \gtrsim e^{-AN\delta_N^2} \quad (\text{A13})$$

for all N large enough.

Proof. Using (A2) and noting that $\|F_0\|_{C^\beta} < \infty$ for $F_0 \in \mathcal{H}$ by a Sobolev imbedding, for any fixed constant $M > 1 \vee \|F_0\|_{C^\beta}$,

$$\begin{aligned} \Pi_N(F : \|\mathcal{G}(F) - \mathcal{G}(F_0)\|_{L^2(\mathcal{D})} \leq q\delta_N) &\geq \Pi_N(F : \|F - F_0\|_{(H^\kappa)^*} \leq q'\delta_N, \|F - F_0\|_{C^\beta} \leq M) \\ &= \Pi_N(F : F - F_0 \in C_1 \cap C_2), \end{aligned}$$

having defined

$$C_1 := \{F : \|F\|_{(H^\kappa)^*} \leq q'\delta_N\}, \quad C_2 := \{F : \|F\|_{C^\beta} \leq M\},$$

whose intersection is a symmetric set in the ambient space $C(\mathcal{O})$. Then, since $F_0 \in \mathcal{H}$, recalling that the RKHS \mathcal{H}_N of Π_N coincides with \mathcal{H} with RKHS norm $\|\cdot\|_{\mathcal{H}_N}$ given in (12), now with scaling $N^{d/(4\alpha+4\kappa+2d)} = \sqrt{N}\delta_N$, we can use corollary 2.6.18 in Giné and Nickl (2016) to lower bound the last probability by

$$\begin{aligned} e^{-\|F_0\|_{\mathcal{H}_N}^2/2} \Pi_N(C_1 \cap C_2) &= e^{-\frac{1}{2}N\delta_N^2\|F_0\|_{\mathcal{H}}^2} \Pi_N(C_1 \cap C_2) \\ &\geq e^{-\frac{1}{2}N\delta_N^2\|F_0\|_{\mathcal{H}}^2} (\Pi_N(C_1) - \Pi_N(C_2^c)). \end{aligned}$$

We proceed finding a lower bound for the prior probability of C_1 , which, by construction of Π_N , satisfies

$$\Pi_N(F \in C_1) = \Pi'(F' : \|F'\|_{(H^\kappa)^*} \leq q' \sqrt{N} \delta_N^2).$$

For any integer $\alpha > 0$ and any $\kappa \geq 0$, letting $B_c^\alpha(r) := \{F \in H_c^\alpha, \|F\|_{H^\alpha} \leq r\}$, $r > 0$, we have the metric entropy estimate:

$$\log N(\eta; B_c^\alpha(r), \|\cdot\|_{(H^\kappa)^*}) \lesssim (r/\eta)^{d/(\alpha+\kappa)} \quad \forall \eta > 0, \quad (\text{A14})$$

see the proof of lemma 19 in Nickl *et al* (2020) for the case $\kappa \geq 1/2$, and theorem 4.10.3 in Triebel (1978) for $\kappa < 1/2$ (in the latter case, we note in fact that the estimate holds true also for balls in the whole space H^α). Hence, since \mathcal{H} is continuously imbedded into H_c^α , letting $B_{\mathcal{H}}(1)$ be the unit ball of \mathcal{H} , we have $B_{\mathcal{H}}(1) \subseteq B_c^\alpha(r)$ for some $r > 0$, implying that for all $\eta > 0$

$$\log N(\eta; B_{\mathcal{H}}(1), \|\cdot\|_{(H^\kappa)^*}) \leq \log N(\eta; B_c^\alpha(r), \|\cdot\|_{(H^\kappa)^*}) \lesssim \eta^{-d/(\alpha+\kappa)}.$$

Then, for all N large enough, the small ball estimate in theorem 1.2 in Li and Linde (1999) yields

$$\begin{aligned} -\log \Pi'(F' : \|F'\|_{(H^\kappa)^*} \leq q' \sqrt{N} \delta_N^2) &\lesssim (\sqrt{N} \delta_N^2)^{-2 \frac{d}{\alpha+\kappa} (2-d/(\alpha+\kappa))^{-1}} \\ &= [N^{-(\alpha+\kappa-d/2)/(2\alpha+2\kappa+d)}]^{-\frac{2d}{2\alpha+2\kappa-d}} \\ &= N \delta_N^2, \end{aligned} \quad (\text{A15})$$

implying $\Pi_N(C_1) \geq e^{-q'' N \delta_N^2}$, for some $q'' > 0$. Note that q'' can be made as small as desired by taking q in (A13) large enough.

We conclude obtaining a suitable upper bound for $\Pi_N(C_2^c)$. In particular, by construction of Π_N , recalling $N^{d/(2\alpha+2\kappa+d)} = N \delta_N^2$,

$$\Pi_N(C_2) = \Pr(\|F'\|_{C^\beta} \leq MN \delta_N^2), \quad F' \sim \Pi'.$$

By condition 3, F' defines a centred Gaussian Borel random element in a separable measurable subspace \mathcal{C} of C^β , and by the Hahn–Banach theorem and the separability of \mathcal{C} , $\|F'\|_{C^\beta}$ can then be represented as a countable supremum

$$\|F'\|_{C^\beta} = \sup_{T \in \mathcal{T}} |T(F')|$$

of actions of bounded linear functionals $\mathcal{T} = (T_m)_{m \in \mathbb{N}} \subset (C^\beta)^*$. It follows that the collection $\{T_m(F')\}_{m \in \mathbb{N}}$ is a centred Gaussian process with almost surely finite supremum, so that by Fernique's theorem Giné and Nickl (2016), theorem 2.1.20:

$$E\|F'\|_{C^\beta} = E \sup_{m \in \mathbb{N}} |T_m(F')| < \infty; \quad \tau^2 := \sup_{m \in \mathbb{N}} E|T_m(F')|^2 < \infty.$$

We then apply the Borell–Sudakov–Tirelson inequality Giné and Nickl (2016), theorem 2.5.8, to obtain for all N large enough,

$$\Pr\left(\|F'\|_{C^\beta} \geq M \sqrt{N} \delta_N\right) \leq \Pr\left(\|F'\|_{C^\beta} \geq E\|F'\|_{C^\beta} + M \sqrt{N} \delta_N/2\right) \leq e^{-\frac{1}{8}(M/\tau)^2 N \delta_N^2}.$$

Given our initial choice of M , the proof is then concluded taking q in lemma 16 sufficiently large so that $q'' < (M/\tau)^2/8$. \square

We now construct suitable approximating sets for which we check the excess mass condition (A6).

Lemma 17. *Let Π_N and δ_N be as in theorem 14. Define for any $M, Q > 0$*

$$\mathcal{A}_N = \{F : F = F_1 + F_2 : \|F_1\|_{(H^\kappa)^*} \leq Q\delta_N, \|F_2\|_{\mathcal{H}} \leq M, \|F\|_{C^\beta} \leq M\}. \quad (\text{A16})$$

Then for any $B > 0$ and for sufficiently large M, Q (both depending on $B, \alpha, \beta, \gamma, \kappa, d$), for all N large enough,

$$\Pi_N(\mathcal{A}_N^c) \leq 2e^{-BN\delta_N^2}. \quad (\text{A17})$$

Proof. We note that the last inequality at the end of the proof of the previous lemma implies that for $M \gtrsim \sqrt{B}$ and all N large enough, $\Pi_N(F : \|F\|_{C^\beta} \leq M) \geq 1 - e^{-BN\delta_N^2}$. Thus, the claim will follow if we can derive a similar lower bound for

$$\begin{aligned} \Pi_N(F : F = F_1 + F_2 : \|F_1\|_{(H^\kappa)^*} \leq Q\delta_N, \|F_2\|_{\mathcal{H}} \leq M) \\ \geq \Pi'(F' : F' = F'_1 + F'_2, \|F'_1\|_{(H^\kappa)^*} \leq Q\sqrt{N}\delta_N^2, \|F'_2\|_{\mathcal{H}} \leq M\sqrt{N}\delta_N), \end{aligned}$$

having used that $N^{d/(4\alpha+4\kappa+d)} = \sqrt{N}\delta_N$. Using theorem 1.2 in Li and Linde (1999) as before (A15), we deduce that for some $q > 0$

$$-\log \Pi'(F' : \|F'\|_{(H^\kappa)^*} \leq Q\sqrt{N}\delta_N^2) \leq q(Q\sqrt{N}\delta_N^2)^{-\frac{2d}{2\alpha+2\kappa-d}}$$

so that taking any $Q > (B/q)^{-(2\alpha+2\kappa-d)/(2d)}$ implies

$$-\log \Pi'(F' : \|F'\|_{(H^\kappa)^*} \leq Q\sqrt{N}\delta_N^2) \leq B(\sqrt{N}\delta_N^2)^{-\frac{2d}{2\alpha+2\kappa-d}} = BN\delta_N^2. \quad (\text{A18})$$

Next, denote

$$M_N = -2\Phi^{-1}(e^{-BN\delta_N^2})$$

where Φ is the standard normal cumulative distribution function. Then by standard inequalities for Φ^{-1} we have $M_N \simeq \sqrt{BN}\delta_N$ as $N \rightarrow \infty$, so that taking $M \gtrsim \sqrt{B}$

$$\begin{aligned} \Pi'(F' : F' = F'_1 + F'_2, \|F'_1\|_{(H^\kappa)^*} \leq Q\sqrt{N}\delta_N^2, \|F'_2\|_{\mathcal{H}} \leq M\sqrt{N}\delta_N) \\ \geq \Pi'(F' : F' = F'_1 + F'_2, \|F'_1\|_{(H^\kappa)^*} \leq Q\sqrt{N}\delta_N^2, \|F'_2\|_{\mathcal{H}} \leq M_N). \end{aligned}$$

By the isoperimetric inequality for Gaussian processes Giné and Nickl (2016), theorem 2.6.12, the last probability is then lower bounded, using (A18), by

$$\begin{aligned} \Phi(\Phi^{-1}[\Pi'(F' : \|F'\|_{(H^\kappa)^*} \leq Q\sqrt{N}\delta_N^2)] + M_N) &\geq \Phi(\Phi^{-1}[e^{-BN\delta_N^2}] + M_N) \\ &= 1 - e^{-BN\delta_N^2}, \end{aligned}$$

concluding the proof. □

We conclude with the verification of the complexity bound (A7) for the sets \mathcal{A}_N .

Lemma 18. *Let \mathcal{A}_N be as in lemma 17 for some fixed $M, Q > 0$. Then,*

$$\log N(\delta_N; \mathcal{A}_N, h) \leq CN\delta_N^2,$$

for some constant $C > 0$ (depending on $\sigma, M, Q, \alpha, \beta, \gamma, \kappa, d, S$) and all N large enough.

Proof. If $F \in \mathcal{A}_N$, then $F = F_1 + F_2$ with $\|F_1\|_{(H^\kappa)^*} \leq Q\delta_N$ and $\|F_2\|_{H^\alpha} \leq M'$, the latter inequality following from the continuous imbedding of \mathcal{H} into H_c^α . Thus, recalling the metric entropy estimate (A14), if

$$\{H_1, \dots, H_P\} \subset B_c^\alpha(M'), \quad P \leq e^{-q\delta_N^{-d/(\alpha+\kappa)}} = e^{-qN\delta_N^2}, \quad q > 0,$$

is a δ_N -net with respect to $\|\cdot\|_{(H^\kappa)^*}$, we can find H_i such that $\|F_2 - H_i\|_{(H^\kappa)^*} \leq \delta_N$. Then, using the second inequality in (A26) below and the local Lipschitz estimate (A2),

$$h(p_F, H_i) \lesssim \|\mathcal{G}(F) - \mathcal{G}(H_i)\|_{L^2(\mathcal{D})} \lesssim (1 + \|F\|_{C^\beta}^\gamma \vee \|H_i\|_{C^\beta}^\gamma) \|F - H_i\|_{(H^\kappa)^*}.$$

Recalling that if $F \in \mathcal{A}_N$ then also $\|F\|_{C^\beta} \leq M$, and using the Sobolev imbedding of H^α into C^β to bound $\|H_i\|_{C^\beta}$, we then obtain

$$h(p_F, H_i) \lesssim \|F - H_i\|_{(H^\kappa)^*} \lesssim \|F - F_2\|_{(H^\kappa)^*} + \|F_2 - H_i\|_{(H^\kappa)^*} \lesssim \delta_N.$$

It follows that $\{H_1, \dots, H_P\}$ also forms a $q'\delta_N$ -net for \mathcal{A}_N in the Hellinger distance for some $q' > 0$, so that

$$\log N(\delta_N; \mathcal{A}_N, h) \leq \log N(\delta_N/q'; B_c^\alpha(M), \|\cdot\|_{(H^\kappa)^*}) \lesssim N\delta_N^2.$$

□

A.3. Contraction rates for hierarchical Gaussian series priors

We now derive contraction rates in L^2 -prediction risk in the inverse problem (A1), for the truncated Gaussian random series priors introduced in section 2.2.3. The proof again proceeds by an application of theorem 13.

Theorem 19. *Let the forward map \mathcal{G} satisfy (A2) and (A3) for given $\beta, \gamma, \kappa \geq 0$ and $S > 0$. For any $\alpha > \beta + d/2$, let Π be the random series prior in (19), and let $\Pi(\cdot|Y^{(N)}, X^{(N)})$ be the resulting posterior distribution arising from observations $(Y^{(N)}, X^{(N)})$ in (A1). Then, for each $\alpha_0 \geq \alpha$, any $F_0 \in H_K^{\alpha_0}(\mathcal{O})$ and any $D > 0$ there exists $L > 0$ large enough (depending on $\sigma, F_0, D, \alpha, \beta, \gamma, \kappa, S, d$) such that, as $N \rightarrow \infty$,*

$$\Pi(F : \|\mathcal{G}(F) - \mathcal{G}(F_0)\|_{L^2(\mathcal{D})} > L\xi_N | Y^{(N)}, X^{(N)}) = O_{P_{F_0}^N}(e^{-DN\xi_N^2}), \quad (\text{A19})$$

where $\xi_N = N^{-(\alpha_0+\kappa)/(2\alpha_0+2\kappa+d)} \log N$. Moreover, for \mathcal{H}_J the finite-dimensional subspaces from (18) and $J_N \in \mathbb{N}$ such that $2^{J_N} \simeq N^{1/(2\alpha_0+2\kappa+d)}$, we also have that for sufficiently large $M > 0$ (depending on D, α, β, d):

$$\Pi(F : F \in \mathcal{H}_{J_N}, \|F\|_{H^\alpha} \leq M 2^{J_N \alpha} N \xi_N^2 |Y^{(N)}, X^{(N)}) = O_{P_{F_0}^N}(e^{-DN \xi_N^2}). \quad (\text{A20})$$

We begin deriving a suitable small ball estimate in the L^2 -prediction risk.

Lemma 20. *Let Π , F_0 and ξ_N be as in theorem 19. Then, for all sufficiently large $q > 0$ there exists $A > 0$ (depending on $q, F_0, \alpha, \beta, \gamma, \kappa, d$) such that*

$$\Pi(F : \|\mathcal{G}(F) - \mathcal{G}(F_0)\|_{L^2(\mathcal{D})} \leq q \xi_N) \gtrsim e^{-AN \xi_N^2} \quad (\text{A21})$$

for all N large enough.

Proof. For each $j \in \mathbb{N}$, denote by Π_j the Gaussian probability measure on the finite dimensional subspace \mathcal{H}_j in (18) defined as after (19) with the series truncated at j . For $J_N \in \mathbb{N} : 2^{J_N} \simeq N^{1/(2\alpha_0+2\kappa+d)}$, note

$$2^{J_N d} \log 2^{J_N d} \simeq N^{d/(2\alpha_0+2\kappa+d)} \log N = N \xi_N^2, \quad (\text{A22})$$

so that, recalling the properties (20) of the random truncation level J , for some $s > 0$,

$$\Pr(K = J_N) \gtrsim e^{-2^{J_N d} \log 2^{J_N d}} \geq e^{-sN \xi_N^2}$$

for all N large enough. It follows

$$\begin{aligned} \Pi(F : \|\mathcal{G}(F) - \mathcal{G}(F_0)\|_{L^2} \leq q \xi_N) &\geq \Pi_{J_N}(F : \|\mathcal{G}(F) - \mathcal{G}(F_0)\|_{L^2} \leq q \xi_N) \Pr(K = J_N) \\ &\gtrsim \Pi_{J_N}(F : \|\mathcal{G}(F) - \mathcal{G}(F_0)\|_{L^2} \leq q \xi_N) e^{-sN \xi_N^2}. \end{aligned}$$

Next, let

$$P_{\mathcal{H}_{J_N}}(F_0) = \chi \sum_{\ell \leq J_N, r \in \mathcal{R}_\ell} \langle F_0, \Psi_{\ell r} \rangle_{L^2} \Psi_{\ell r}$$

be the ‘projection’ of F_0 onto \mathcal{H}_{J_N} . Since $F_0 \in H_K^{\alpha_0} \subset C^\beta$ by a Sobolev imbedding, it follows using (A2) and standard approximation properties of wavelets (B6),

$$\|\mathcal{G}(F_0) - \mathcal{G}(P_{\mathcal{H}_{J_N}}(F_0))\|_{L^2(\mathcal{D})} \lesssim \|F_0 - P_{\mathcal{H}_{J_N}}(F_0)\|_{(H^\kappa)^*} \lesssim 2^{-J_N(\alpha_0+\kappa)} = N^{-\frac{\alpha_0+\kappa}{2\alpha_0+2\kappa+d}},$$

which implies by the triangle inequality that

$$\begin{aligned} \Pi_{J_N}(F : \|\mathcal{G}(F) - \mathcal{G}(F_0)\|_{L^2} \leq q \xi_N) \\ &\geq \Pi_{J_N}(F : \|\mathcal{G}(F) - \mathcal{G}(P_{\mathcal{H}_{J_N}}(F_0))\|_{L^2} \leq q \xi_N - \|\mathcal{G}(F_0) - \mathcal{G}(P_{\mathcal{H}_{J_N}}(F_0))\|_{L^2}) \\ &\geq \Pi_{J_N}(F : \|\mathcal{G}(F) - \mathcal{G}(P_{\mathcal{H}_{J_N}}(F_0))\|_{L^2} \leq q' \xi_N). \end{aligned}$$

Using again that H^α imbeds continuously into C^β as well as (A2) and (B5), we can lower bound the last probability by

$$\begin{aligned} \Pi_{J_N}(F : \|\mathcal{G}(F) - \mathcal{G}(P_{\mathcal{H}_{J_N}}(F_0))\|_{L^2(\mathcal{D})} \leq q' \xi_N, \|F - P_{\mathcal{H}_{J_N}}(F_0)\|_{H^\alpha(\mathcal{O})} \leq \xi_N) \\ &\geq \Pi_{J_N}(F : \|F - P_{\mathcal{H}_{J_N}}(F_0)\|_{(H^\kappa(\mathcal{O}))^*} \leq q'' \xi_N, \|F - P_{\mathcal{H}_{J_N}}(F_0)\|_{H^\alpha(\mathcal{O})} \leq \xi_N) \\ &\geq \Pi_{J_N}(F : \|F - P_{\mathcal{H}_{J_N}}(F_0)\|_{H^\alpha(\mathcal{O})} \leq q''' \xi_N), \end{aligned}$$

which, by corollary 2.6.18 in Giné and Nickl (2016) and in view of (B2) is further lower bounded by

$$e^{-\frac{1}{2}\|P_{\mathcal{H}_{J_N}}(F_0)\|_{\mathcal{H}_{J_N}}^2} \Pi_{J_N}(F : \|F\|_{H^\alpha} \leq q''' \xi_N) \geq e^{-s' \|F_0\|_{H^{\alpha_0}}^2} \Pi_{J_N}(F : \|F\|_{H^\alpha} \leq q''' \xi_N).$$

Now since $f \mapsto \chi f$, $\chi \in C^\infty(\mathcal{O})$, is continuous on $H^\alpha(\mathcal{O})$,

$$\begin{aligned} \Pi_{J_N}(F : \|F\|_{H^\alpha} \leq q''' \xi_N) &= \Pr \left(\left\| \chi \sum_{\ell \leq J_N, r \in \mathcal{R}_\ell} 2^{-\ell\alpha} F_{\ell r} \Psi_{\ell r} \right\|_{H^\alpha} \leq q''' \xi_N \right) \\ &\geq \Pr \left(\sum_{m=1}^{\dim(\mathcal{H}_{J_N})} Z_m^2 \leq t \xi_N^2 \right) \end{aligned}$$

for some $t > 0$, where $Z_m \stackrel{\text{iid}}{\sim} N(0, 1)$, and where we have used the wavelet characterisation of the $H^\alpha(\mathbb{R}^d)$ norm. To conclude, note that the last probability is greater than

$$\begin{aligned} \Pr \left(\sqrt{\dim(\mathcal{H}_{J_N})} \max_{m \leq \dim(\mathcal{H}_{J_N})} |Z_m| \leq \sqrt{t} \xi_N \right) &\geq \Pr \left(\max_{m \leq \dim(\mathcal{H}_{J_N})} |Z_m| \leq t' N^{-\frac{\alpha_0 + \kappa}{2\alpha_0 + 2\kappa + d}} N^{-\frac{d/2}{2\alpha_0 + 2\kappa + d}} \right) \\ &= \prod_{m \leq \dim(\mathcal{H}_{J_N})} \Pr \left(|Z_m| \leq t' N^{-\frac{\alpha_0 + \kappa + d/2}{2\alpha_0 + 2\kappa + d}} \right). \end{aligned}$$

Finally, a standard calculation shows that $\Pr(|Z_1| \leq t) \gtrsim t$ if $t \rightarrow 0$, and hence the last product is lower bounded, for large N , by

$$\left(t' N^{-\frac{\alpha_0 + \kappa + d/2}{2\alpha_0 + 2\kappa + d}} \right)^{\dim(\mathcal{H}_{J_N})} = e^{\dim(\mathcal{H}_{J_N}) \log \left(t' N^{-\frac{\alpha_0 + \kappa + d/2}{2\alpha_0 + 2\kappa + d}} \right)} \geq e^{-t'' 2^{J_N d} \log N} = e^{-t''' N \xi_N^2}.$$

□

In the following lemma we construct suitable approximating sets, for which we check the excess mass condition (A6) and the complexity bound (A7) required in theorem 13.

Lemma 21. *Let Π , ξ_N and J_N be as in theorem 8, and let \mathcal{H}_{J_N} be the finite dimensional subspace defined in (18) with $J = J_N$. Define for each $M > 0$*

$$\mathcal{A}_N = \{F \in \mathcal{H}_{J_N}, \|F\|_{H^\alpha} \leq M 2^{J_N \alpha} N \xi_N^2\}. \quad (\text{A23})$$

Then, for any $B > 0$ there exists $M > 0$ large enough (depending on B, α, β, d) such that, for sufficiently large N

$$\Pi(\mathcal{A}_N^c) \leq 2e^{-BN \xi_N^2}. \quad (\text{A24})$$

Moreover, for each fixed $M > 0$ and all N large enough

$$\log N(\xi_N; \mathcal{A}_N, h) \leq CN\xi_N^2 \quad (\text{A25})$$

for some $C > 0$ (depending on $\sigma, \alpha, \beta, \gamma, \kappa, S, d$).

Proof. Letting $Z_m \stackrel{\text{iid}}{\sim} N(0, 1)$, noting $\|F\|_{H^\alpha}^2 \leq 2^{2J_N\alpha} \sum_{\ell \leq J_N, r \in \mathcal{R}_\ell} F_{\ell r}^2$ for all $F \in \mathcal{H}_{J_N}$ (cf. (B2)) and using (A22) and (20), we have for sufficiently large N

$$\begin{aligned} \Pi(\mathcal{A}_N^c) &\leq \Pr(J > J_N) + \Pr\left(\sum_{\ell \leq J \wedge J_N, r \in \mathcal{R}_\ell} F_{\ell r}^2 \leq MN\xi_N^2\right) \\ &\leq e^{-2^{J_N d} \log 2^{J_N d}} + \Pr\left(\sum_{m \leq \dim(\mathcal{H}_{J_N})} Z_m^2 > MN\xi_N^2\right) \\ &\leq e^{-BN\xi_N^2} + \Pr\left(\sum_{m \leq \dim(\mathcal{H}_{J_N})} (Z_m^2 - 1) > \bar{M}N\xi_N^2\right) \end{aligned}$$

for any constant $0 < \bar{M} < M^2 - 1$, since $\dim(\mathcal{H}_{J_N}) \lesssim 2^{J_N d} \simeq N^{d/(2\alpha+2+d)} = o(N\xi_N^2)$. The bound (A24) then follows applying theorem 3.1.9 in Giné and Nickl (2016) to upper bound the last probability, for any $B >$ and for sufficiently large M and \bar{M} , by

$$e^{-\frac{\bar{M}^2(N\xi_N^2)^2}{4\dim(\mathcal{H}_{J_N}) + \bar{M}N\xi_N^2}} \leq e^{-BN\xi_N^2}.$$

We proceed with the derivation of (A25). By choice of J_N , if $F \in \mathcal{A}_N$ then $\|F\|_{H^\alpha}^2 \lesssim N^{(2\alpha)/(2\alpha+2\kappa+d)} N\xi_N^2$. Hence, by the second inequality in (A26), using (A2) and the Sobolev imbedding of H^α into C^β , if $F_1, F_2 \in \mathcal{A}_N$ then

$$\begin{aligned} h(p_{F_1}, p_{F_2}) &\lesssim \|\mathcal{G}(F_1) - \mathcal{G}(F_2)\|_{L^2(\mathcal{D})} \\ &\lesssim \left(1 + \left(N^{\frac{\alpha}{2\alpha+2\kappa+d}} \sqrt{N}\xi_N\right)^\gamma\right) \|F_1 - F_2\|_{(H^\kappa)^*} \\ &\lesssim N^{\frac{\alpha\gamma}{2\alpha+2\kappa+d}} (\sqrt{N}\xi_N)^\gamma \sqrt{\sum_{\ell \leq J_N, r \in \mathcal{R}_\ell} (F_{1,\ell r} - F_{2,\ell r})^2}. \end{aligned}$$

Therefore, using the standard metric entropy estimate for balls $B_{\mathbb{R}^p}(r)$, $r > 0$, in Euclidean spaces (Giné and Nickl (2016), proposition 4.3.34), we see that for N large enough

$$\begin{aligned} \log N(\xi_N; \mathcal{A}_N, h) &\lesssim \log N\left(\xi_N N^{\frac{-\alpha\gamma}{2\alpha+2\kappa+d}} (\sqrt{N}\xi_N)^{-\gamma}; B_{\mathbb{R}^{\dim(\mathcal{H}_{J_N})}}(M\sqrt{N}\xi_N), \|\cdot\|_{\mathbb{R}^{\dim(\mathcal{H}_{J_N})}}\right) \\ &\leq \dim(\mathcal{H}_{J_N}) \log \frac{3M\sqrt{N}\xi_N}{\xi_N N^{\frac{\alpha\gamma}{2\alpha+2\kappa+d}} (\sqrt{N}\xi_N)^{-\gamma}} \lesssim N\xi_N^2. \end{aligned}$$

□

A.4. Information theoretic inequalities

In the following lemma due to Birgé (2004) we exploit the boundedness assumption (A3) on \mathcal{G} to show the equivalence between the Hellinger distance appearing in the conclusion of theorem 13 and the L^2 -distance on the ‘regression functions’ $\mathcal{G}(F)$.

Lemma 22. *Let the forward map \mathcal{G} satisfy (A3) for some $S > 0$. Then, for all $F_1, F_2 \in \mathcal{F}$*

$$\frac{1 - e^{-S^2/(2\sigma^2)}}{4S^2} \|\mathcal{G}(F_1) - \mathcal{G}(F_2)\|_{L^2(\mathcal{D})}^2 \leq h^2(p_{F_1}, p_{F_2}) \leq \frac{1}{4\sigma^2} \|\mathcal{G}(F_1) - \mathcal{G}(F_2)\|_{L^2(\mathcal{D})}^2. \quad (\text{A26})$$

Proof. Note $h^2(p_{F_1}, p_{F_2}) = 2 - 2\rho(p_{F_1}, p_{F_2})$, where

$$\rho(p_{F_1}, p_{F_2}) := \int_{\mathbb{R} \times \mathcal{D}} \sqrt{p_{F_1}(y, x) p_{F_2}(y, x)} dy dx$$

is the Hellinger affinity. Using the expression of the likelihood in (8) (with \mathcal{D} instead of \mathcal{O}), the right hand side is seen to be equal to

$$\begin{aligned} & \int_{\mathbb{R} \times \mathcal{D}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\{[y - \mathcal{G}(F_1)(x)]^2 - [y - \mathcal{G}(F_2)(x)]^2\}/(4\sigma^2)} dy dx \\ &= \int_{\mathcal{D}} e^{-\{[\mathcal{G}(F_1)(x)]^2 + [\mathcal{G}(F_2)(x)]^2\}/(4\sigma^2)} \\ & \quad \times \left[\int_{\mathbb{R}} \frac{e^{-y^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}} e^{y[\mathcal{G}(F_1)(x) + \mathcal{G}(F_2)(x)]/(2\sigma^2)} dy \right] dx \\ &= \int_{\mathcal{D}} e^{-\{[\mathcal{G}(F_1)(x)]^2 + [\mathcal{G}(F_2)(x)]^2\}/(4\sigma^2)} e^{[\mathcal{G}(F_1)(x) + \mathcal{G}(F_2)(x)]^2/(8\sigma^2)} dx \end{aligned}$$

having used that the moment generating function of $Z \sim N(0, \sigma^2)$ satisfies $E e^{tZ} = e^{\sigma^2 t^2/2}$, $t \in \mathbb{R}$. Thus, the latter integral equals

$$\int_{\mathcal{D}} e^{-\{[\mathcal{G}(F_1)(x)]^2 + [\mathcal{G}(F_2)(x)]^2 - 2\mathcal{G}(F_2)(x)\mathcal{G}(F_1)(x)\}/(8\sigma^2)} dx = E^\mu e^{-\{\mathcal{G}(F_1)(X) - \mathcal{G}(F_2)(X)\}^2/(8\sigma^2)}.$$

To derive the second inequality in (A26), we use Jensen’s inequality to lower bound the expectation in the last line by

$$e^{-E^\mu \{\mathcal{G}(F_1)(X) - \mathcal{G}(F_2)(X)\}^2/(8\sigma^2)} = e^{-\|\mathcal{G}(F_1) - \mathcal{G}(F_2)\|_{L^2(\mathcal{D})}^2/(8\sigma^2)}.$$

Hence

$$h^2(p_{F_1}, p_{F_2}) \leq 2 \left[1 - e^{-\|\mathcal{G}(F_1) - \mathcal{G}(F_2)\|_{L^2(\mathcal{D})}^2/(8\sigma^2)} \right],$$

whereby the claim follows using the basic inequality $1 - e^{-z/c} \leq z/c$, for all $c, z > 0$.

To deduce the first inequality we follow the proof of proposition 1 in Birgé (2004): note that for all $0 \leq z_1 < z_2$

$$e^{-z_1} \leq \frac{z_1}{z_2} e^{-z_2} + \left(1 - \frac{z_1}{z_2}\right) = \frac{e^{-z_2} - 1}{z_2} z_1 + 1.$$

Then taking $z_1 = \{\mathcal{G}(F_1)(X) - \mathcal{G}(F_2)(X)\}^2 / (8\sigma^2)$ and $z_2 = S^2 / (2\sigma^2)$,

$$E^\mu e^{-\{\mathcal{G}(F_1)(X) - \mathcal{G}(F_2)(X)\}^2 / (8\sigma^2)} \leq \frac{e^{-S^2 / (2\sigma^2)} - 1}{4S^2} \|\mathcal{G}(F_1) - \mathcal{G}(F_2)\|_{L^2(\mathcal{D})}^2 + 1$$

which in turn yields the result. \square

The next lemma bounds the Kullback–Leibler divergences appearing in (A4) in terms of the L^2 -prediction risk.

Lemma 23. *Let the observation Y_i in (A1) be generated by some fixed $F_0 \in \mathcal{F}$. Then, for each $F \in \mathcal{F}$,*

$$E_{F_0}^1 \left[\log \frac{p_{F_0}(Y_1, X_1)}{p_F(Y_1, X_1)} \right] = \frac{1}{2\sigma^2} \|\mathcal{G}(F_0) - \mathcal{G}(F)\|_{L^2(\mathcal{D})}^2,$$

and

$$E_{F_0}^1 \left[\log \frac{p_{F_0}(Y_1, X_1)}{p_F(Y_1, X_1)} \right]^2 \leq \frac{2(S^2 + \sigma^2)}{\sigma^4} \|\mathcal{G}(F_0) - \mathcal{G}(F)\|_{L^2(\mathcal{D})}^2.$$

Proof. If $Y_1 = \mathcal{G}(F_0)(X_1) + \sigma W_1$, then

$$\begin{aligned} & \log \frac{p_{F_0}(Y_1, X_1)}{p_F(Y_1, X_1)} \\ &= -\frac{1}{2\sigma^2} \{[\mathcal{G}(F_0)(X_1) + \sigma W_1 - \mathcal{G}(F_0)(X_1)]^2 - [\mathcal{G}(F_0)(X_1) + \sigma W_1 - \mathcal{G}(F)(X_1)]^2\} \\ &= \frac{1}{2\sigma^2} \{\mathcal{G}(F_0)(X_1) - \mathcal{G}(F)(X_1)\}^2 + \frac{1}{\sigma} W_1 \{\mathcal{G}(F_0)(X_1) - \mathcal{G}(F)(X_1)\}. \end{aligned}$$

Hence, since $EW_1 = 0$ and $X_1 \sim \mu$,

$$\begin{aligned} E_{F_0}^1 \left[\log \frac{p_{F_0}(Y_1, X_1)}{p_F(Y_1, X_1)} \right] &= E^\mu \left[\frac{1}{2\sigma^2} \{\mathcal{G}(F_0)(X_1) - \mathcal{G}(F)(X_1)\}^2 \right] \\ &= \frac{1}{2\sigma^2} \|\mathcal{G}(F_0) - \mathcal{G}(F)\|_{L^2(\mathcal{D})}^2. \end{aligned}$$

On the other hand,

$$\begin{aligned}
\left[\log \frac{p_{F_0}(Y_1, X_1)}{p_F(Y_1, X_1)} \right]^2 &= \left[\frac{1}{2\sigma^2} \{ \mathcal{G}(F_0)(X_1) - \mathcal{G}(F)(X_1) \}^2 + \frac{1}{\sigma} W_1 \{ \mathcal{G}(F_0)(X_1) - \mathcal{G}(F)(X_1) \} \right]^2 \\
&\leq 2 \left[\frac{1}{2\sigma^2} \{ \mathcal{G}(F_0)(X_1) - \mathcal{G}(F)(X_1) \}^2 \right]^2 + 2 \left[\frac{1}{\sigma} W_1 \{ \mathcal{G}(F_0)(X_1) - \mathcal{G}(F)(X_1) \} \right]^2 \\
&\leq \frac{2S^2}{\sigma^4} \{ \mathcal{G}(F_0)(X_1) - \mathcal{G}(F)(X_1) \}^2 + \frac{2}{\sigma^2} W_1^2 \{ \mathcal{G}(F_0)(X_1) - \mathcal{G}(F)(X_1) \}^2,
\end{aligned}$$

whence the second claim follows since $EW_1^2 = 1$. \square

Appendix B. Additional background material

In this final appendix we collect some standard materials used in the proofs for convenience of the reader.

Example 24. Take

$$\phi : \mathbb{R} \rightarrow (0, \infty), \quad \phi(t) = \frac{1}{1-t} 1_{\{t < 0\}} + (1+t) 1_{\{t \geq 0\}},$$

and let $\psi : \mathbb{R} \rightarrow [0, \infty)$ be a smooth compactly supported function such that $\int_{\mathbb{R}} \psi(t) dt = 1$. Define for any $K_{\min} \in (0, 1)$

$$\Phi(t) = K_{\min} + \frac{1 - K_{\min}}{\psi * \phi(0)} \psi * \phi(t), \quad t \in \mathbb{R}. \quad (\text{B1})$$

Then it is elementary to check that Φ is a regular link function that satisfies condition 2 (with $a = 2$).

Example 25. For any real $\alpha > d/2$, the Whittle–Matérn process with index set \mathcal{O} and regularity $\alpha - d/2 > 0$ (cf. example 11.8 in Ghosal and van der Vaart (2017)) is the stationary centred Gaussian process $M = \{M(x), x \in \mathcal{O}\}$ with covariance kernel

$$K(x, y) = \int_{\mathbb{R}^d} e^{-i\langle x-y, \xi \rangle_{\mathbb{R}^d}} \mu(d\xi), \quad \mu(d\xi) = (1 + \|\xi\|_{\mathbb{R}^d}^2)^{-\alpha} d\xi, \quad x, y \in \mathcal{O}.$$

From the results in chapter 11 in (Ghosal and van der Vaart 2017) we see that the RKHS of $(M(x) : x \in \mathcal{O})$ equals the set of restrictions to \mathcal{O} of elements in the Sobolev space $H^\alpha(\mathbb{R}^d)$, which equals, with equivalent norms, the space $H^\alpha(\mathcal{O})$ (since \mathcal{O} has a smooth boundary). Moreover, lemma I.4 in Ghosal and van der Vaart (2017) shows that M has a version with paths belonging almost surely to $C^{\beta'}$ for all $\beta' < \alpha - d/2$. Let now $K \subset \mathcal{O}$ be a nonempty compact set, and let M be a $C^{\beta'}$ -smooth version of a Whittle–Matérn process on \mathcal{O} with RKHS $H^\alpha(\mathcal{O})$. Taking $F' = \chi M$ implies (cf. exercise 2.6.5 in Giné and Nickl (2016)) that $\Pi' = \mathcal{L}(F')$ defines a centred Gaussian probability measure supported on $C^{\beta'}$, whose RKHS is given by

$$\mathcal{H} = \{\chi F, F \in H^\alpha(\mathcal{O})\},$$

and the RKHS norm satisfies that for all $F \in H^\alpha(\mathcal{O})$ there exists $F^* \in H^\alpha(\mathcal{O})$ such that $\chi F = \chi F^*$ and

$$\|\chi F\|_{\mathcal{H}} = \|F^*\|_{H^\alpha(\mathcal{O})}.$$

Thus if $F' = \chi F$ is an arbitrary element of \mathcal{H} , then

$$\|F'\|_{H^\alpha} = \|\chi F^*\|_{H^\alpha} \lesssim \|F^*\|_{H^\alpha} = \|F'\|_{\mathcal{H}},$$

which shows that \mathcal{H} is continuously embedded into $H_c^\alpha(\mathcal{O})$.

Remark 26. Let $\{\Psi_{\ell r}, \ell \geq -1, r \in \mathbb{Z}^d\}$ be an orthonormal basis of $L^2(\mathbb{R}^d)$ composed of S -regular and compactly supported Daubechies wavelets (see chapter 4 in Giné and Nickl (2016) for construction and properties). For each $0 \leq \alpha \leq S$, we have

$$H^\alpha(\mathbb{R}^d) = \left\{ F \in L^2(\mathbb{R}^d) : \sum_{\ell, r} 2^{2\ell\alpha} \langle F, \Psi_{\ell r} \rangle_{L^2(\mathbb{R}^d)}^2 < \infty \right\},$$

and the square root of the latter series defines an equivalent norm to $\|\cdot\|_{H^\alpha(\mathbb{R}^d)}$. Note that $S > 0$ can be taken arbitrarily large.

For any $\alpha \geq 0$ the Gaussian random series

$$\bar{F}_j = \sum_{\ell \leq j, r \in \mathcal{R}_\ell} F_{\ell r} 2^{-\ell\alpha} \Psi_{\ell r}, \quad F_{\ell r} \stackrel{\text{iid}}{\sim} N(0, 1)$$

defines a centred Gaussian probability measure supported on the finite-dimensional space $\bar{\mathcal{H}}_j$ spanned by the $\{\Psi_{\ell r}, \ell \leq j, r \in \mathcal{R}_\ell\}$, and its RKHS equals $\bar{\mathcal{H}}_j$ endowed with norm

$$\|\bar{H}_j\|_{\bar{\mathcal{H}}_j}^2 = \sum_{\ell \leq j, r \in \mathcal{R}_\ell} 2^{2\ell\alpha} H_{\ell r}^2 = \|\bar{H}_j\|_{H^\alpha(\mathbb{R}^d)}^2 \quad \forall \bar{H}_j \in \bar{\mathcal{H}}_j$$

(cf. example 2.6.15 in Giné and Nickl (2016)). Basic wavelet theory implies $\dim(\bar{\mathcal{H}}_j) \lesssim 2^{jd}$.

If we now fix compact $K' \subset \mathcal{O}$ such that $K \subsetneq K'$, and consider a cut-off function $\chi \in C_c^\infty(\mathcal{O})$ such that $\chi = 1$ on K' , then multiplication by χ is a bounded linear operator $\chi : H^s(\mathbb{R}^d) \rightarrow H_c^s(\mathcal{O})$. It follows that the random function

$$F_j = \chi(\bar{F}_j) = \sum_{\ell \leq j, r \in \mathcal{R}_\ell} F_{\ell r} 2^{-\ell\alpha} \chi \Psi_{\ell r}, \quad F_{\ell r} \stackrel{\text{iid}}{\sim} N(0, 1)$$

defines, according to exercise 2.6.5 in Giné and Nickl (2016), a centred Gaussian probability measure $\Pi_j = \mathcal{L}(F_j)$ supported on the finite dimensional subspace \mathcal{H}_j from (18), with RKHS norm satisfying

$$\left\| \chi \left(\sum_{\ell \leq j, r \in \mathcal{R}_\ell} H_{\ell r} \Psi_{\ell r} \right) \right\|_{\mathcal{H}_j} \leq \left\| \sum_{\ell \leq j, r \in \mathcal{R}_\ell} H_{\ell r} \Psi_{\ell r} \right\|_{\bar{\mathcal{H}}_j} = \sqrt{\sum_{\ell \leq j, r \in \mathcal{R}_\ell} 2^{2\ell\alpha} H_{\ell r}^2}. \quad (\text{B2})$$

Arguing as in the previous remark one shows further that for some constant $c > 0$,

$$\|H_j\|_{H^\alpha(\mathcal{O})} \leq c \|H_j\|_{\mathcal{H}_j} \quad \forall H_j \in \mathcal{H}_j. \quad (\text{B3})$$

Remark 27. Using the notation of the previous remark, for fixed $F_0 \in H_K^\alpha(\mathcal{O})$, consider the finite-dimensional approximations

$$P_{\mathcal{H}_j}(F_0) = \sum_{\ell \leq j, r \in \mathcal{R}_\ell} \langle F_0, \Psi_{\ell r} \rangle_{L^2} \chi \Psi_{\ell r} \in \mathcal{H}_j, \quad j \in \mathbb{N}. \quad (\text{B4})$$

Then in view of (B2), we readily check that for each $j \geq 1$

$$\|P_{\mathcal{H}_j}(F_0)\|_{\mathcal{H}_j} \leq \sqrt{\sum_{\ell \leq j, r \in \mathcal{R}_\ell} 2^{2\ell\alpha} \langle F_0, \Psi_{\ell r} \rangle_{L^2}^2} \leq \|F_0\|_{H^\alpha(\mathcal{O})} < \infty. \quad (\text{B5})$$

Also, for each $\kappa \geq 0$, and any $G \in H^\kappa(\mathcal{O})$, we see that (implicitly extending to 0 on $\mathbb{R}^d \setminus \mathcal{O}$ functions that are compactly supported inside \mathcal{O})

$$\langle F_0 - P_{\mathcal{H}_j}(F_0), G \rangle_{L^2(\mathcal{O})} = \langle F_0 - P_{\mathcal{H}_j}(F_0), \chi' G \rangle_{L^2(\mathbb{R}^d)},$$

where $\chi' \in C_c^\infty(\mathcal{O})$, with $\chi' = 1$ on $\text{supp}(\chi)$. We also note that, in view of the localisation properties of Daubechies wavelets, for some $J_{\min} \in \mathbb{N}$ large enough, if $\ell \geq J_{\min}$ and the support of $\Psi_{\ell r}$ intersects K , then necessarily $\text{supp}(\Psi_{\ell r}) \subseteq K'$, so that

$$\chi \Psi_{\ell r} = \Psi_{\ell r} \quad \forall \ell \geq J_{\min}, \quad r \in \mathcal{R}_\ell.$$

Therefore, for $j \geq J_{\min}$, by Parseval's identity and the Cauchy–Schwarz inequality

$$\begin{aligned} \langle F_0 - P_{\mathcal{H}_j}(F_0), \chi' G \rangle_{L^2(\mathbb{R}^d)} &= \sum_{\ell' > j, r' \in \mathcal{R}_{\ell'}} 2^{\ell'\alpha} \langle F_0, \Psi_{\ell' r'} \rangle_{L^2(\mathbb{R}^d)} 2^{\ell'\kappa} \langle \chi' G, \Psi_{\ell' r'} \rangle_{L^2(\mathbb{R}^d)} 2^{-\ell'(\alpha+\kappa)} \\ &\leq 2^{-j(\alpha+\kappa)} \sqrt{\sum_{\ell' > j, r' \in \mathcal{R}_{\ell'}} 2^{2\ell'\alpha} \langle F_0, \Psi_{\ell' r'} \rangle_{L^2(\mathbb{R}^d)}^2} \\ &\quad \times \sqrt{\sum_{\ell' > j, r' \in \mathcal{R}_{\ell'}} 2^{2\ell'\kappa} \langle \chi' G, \Psi_{\ell' r'} \rangle_{L^2(\mathbb{R}^d)}^2} \\ &\leq 2^{-j(\alpha+\kappa)} \|F_0\|_{H^\alpha(\mathcal{O})} \|\chi' G\|_{H^\kappa(\mathbb{R}^d)}. \end{aligned}$$

It follows by duality that for all j large enough

$$\|F_0 - P_{\mathcal{H}_j}(F_0)\|_{(H^\kappa(\mathcal{O}))^*} \lesssim 2^{-j(\alpha+\kappa)} \|F_0\|_{H^\alpha(\mathcal{O})}. \quad (\text{B6})$$

We conclude remarking that

$$\|F\|_{H^\alpha(\mathcal{O})} \lesssim 2^{j\alpha} \|F\|_{L^2(\mathcal{O})}, \quad \forall F \in \mathcal{H}_j, \quad j \geq J_{\min}. \quad (\text{B7})$$

Indeed, let $j \geq J_{\min}$, and fix $F \in \mathcal{H}_j$; then

$$F = P_{\mathcal{H}_{J_{\min}}}(F) + (F - P_{\mathcal{H}_{J_{\min}}}(F)) = \sum_{\ell \leq J_{\min}, r \in \mathcal{R}_\ell} F_{\ell r} \chi \Psi_{\ell r} + \sum_{J_{\min} < \ell \leq j, r \in \mathcal{R}_\ell} F_{\ell r} \Psi_{\ell r}.$$

But as $\mathcal{H}_{J_{\min}}$ is a fixed finite dimensional subspace, then we have $\|P_{\mathcal{H}_{J_{\min}}}(F)\|_{H^\kappa(\mathcal{O})} \lesssim \|P_{\mathcal{H}_{J_{\min}}}(F)\|_{L^2(\mathcal{O})} \leq \|F\|_{L^2(\mathcal{O})}$ for some fixed multiplicative constant only depending on J_{\min} . On the other hand, we also have

$$\begin{aligned}\|F - P_{\mathcal{H}_{J_{\min}}}(F)\|_{H^\alpha(\mathcal{O})}^2 &= \sum_{J_{\min} < \ell \leq j, r \in \mathcal{R}_\ell} 2^{2\ell\alpha} F_{\ell r}^2 \\ &\leq 2^{2j\alpha} \|F - P_{\mathcal{H}_{J_{\min}}}(F)\|_{L^2(\mathcal{O})}^2 \leq 2^{2j\alpha} \|F\|_{L^2(\mathcal{O})}^2,\end{aligned}$$

yielding (B7).

Example 28. Consider the integer-valued random variable

$$J = \lfloor \log_2(\phi^{-1}(T)^{1/d}) \rfloor + 1, \quad T \sim \text{Exp}(1),$$

where $\phi(x) = x \log x, x \geq 1$. Then for any $j \geq 1$

$$\Pr(J > j) = \Pr(\phi^{-1}(T) \geq 2^{jd}) = \Pr(T \geq 2^{jd} \log 2^{jd}) = e^{-2^{jd} \log 2^{jd}}.$$

On the other hand, since $e^{-2^{jd}(1-2^{-d}) \log 2^{(j-1)d}} \rightarrow 0$ as $j \rightarrow \infty$,

$$\begin{aligned}\Pr(J = j) &= \Pr(2^{(j-1)d} \leq \phi^{-1}(T) < 2^{jd}) = e^{-2^{(j-1)d} \log 2^{(j-1)d}} + 1 - e^{-2^{jd} \log 2^{jd}} - 1 \\ &\geq e^{-2^{(j-1)d} \log 2^{(j-1)d}} (1 - e^{-2^{jd}(1-2^{-d}) \log 2^{(j-1)d}}) \gtrsim e^{-2^{jd} \log 2^{jd}}.\end{aligned}$$

ORCID iDs

Matteo Giordano  <https://orcid.org/0000-0003-1581-1706>

Richard Nickl  <https://orcid.org/0000-0002-0889-5422>

References

- Abraham K and Nickl R 2020 On statistical Calderón problems *Math. Stat. Learn.* accepted. arXiv:[1906.03486](https://arxiv.org/abs/1906.03486)
- Agapiou S, Larsson S and Stuart A M 2013 Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems *Stochastic Process. Appl.* **123** 3828–860
- Alessandrini G 1986 An identification problem for an elliptic equation in two variables *Ann. Mat. Pura Appl.* **145** 265–95
- Arridge S, Maass P, Öktem O and Schönlieb C-B 2019 Solving inverse problems using data-driven models *Acta Numer.* **28** 1–174
- Benning M and Burger M 2018 Modern regularization methods for inverse problems *Acta Numer.* **27** 1–111
- Beskos A, Girolami M, Lan S, Farrell P and Stuart A 2017 Geometric MCMC for infinite-dimensional inverse problems *J. Comput. Phys.* **335** 327–51
- Birgé L 2004 Model selection for gaussian regression with random design *Bernoulli* **10** 1039–051
- Bissantz N, Hohage T and Munk A 2004 Consistency and rates of convergence of nonlinear Tikhonov regularization with random noise *Inverse Problems* **20** 1773–789
- Bissantz N, Hohage T, Munk A and Ruymgaart F 2007 Convergence rates of general regularization methods for statistical inverse problems and applications *SIAM J. Numer. Anal.* **45** 2610–636
- Bonito A, Cohen A, DeVore R, Petrova G and Welper G 2017 Diffusion coefficients estimation for elliptic partial differential equations *SIAM J. Math. Anal.* **49** 1570–592
- Briol F-X, Oates C J, Girolami M, Osborne M A and Sejdinovic D 2019 Probabilistic integration: a role in statistical computation? *Stat. Sci.* **34** 1–22
- Conrad P R, Marzouk Y M, Pillai N S and Smith A 2016 Accelerating asymptotically exact MCMC for computationally intensive models via local approximations *J. Am. Stat. Assoc.* **111** 1591–607
- Cotter S, Roberts G, Stuart A and White D 2013 MCMC methods for functions: modifying old algorithms to make them faster *Stat. Sci.* **28** 424–46

- Dashti M and Stuart A M 2011 Uncertainty quantification and weak approximation of an elliptic inverse problem *SIAM J. Numer. Anal.* **49** 2524–542
- Dashti M and Stuart A M 2016 The Bayesian approach to inverse problems *Handbook of Uncertainty Quantification* eds Ghanem R, Higdon D and Owhadi H (Berlin: Springer)
- Diaconis P 1988 Bayesian numerical analysis *Statistical Decision Theory and Related Topics IV* eds Berger J and Gupta S (New York: Springer) pp 163–75
- Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer Academic Publishers Group)
- Falk R S 1983 Error estimates for the numerical identification of a variable coefficient *Math. Comput.* **40** 537–46
- Ghosal S and van der Vaart A W 2017 *Fundamentals of Nonparametric Bayesian Inference* (New York: Cambridge University Press)
- Gilbarg D and Trudinger N S 1998 *Elliptic Partial Differential Equations of Second Order* (Berlin: Springer)
- Giné E and Nickl R 2016 *Mathematical Foundations of Infinite-Dimensional Statistical Models* (New York: Cambridge University Press)
- Hairer M, Stuart A and Vollmer S 2014 Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions *Ann. Appl. Probab.* **24** 2455–490
- Hanke M, Neubauer A and Scherzer O 1995 A convergence analysis of the Landweber iteration for nonlinear ill-posed problems *Numer. Math.* **72** 21–7
- Hoffmann K H and Sprekels J 1985 On the identification of coefficients of elliptic problems by asymptotic regularization *Numer. Funct. Anal. Optim.* **7** 157–77
- Hohage T and Pricop M 2008 Nonlinear Tikhonov regularization in Hilbert scales for inverse boundary value problems with random noise *Inverse Probl. Imag.* **2** 271–90
- Ito K and Kunisch K 1994 On the injectivity and linearization of the coefficient-to-solution mapping for elliptic boundary value problems *J. Math. Anal. Appl.* **188** 1040–066
- Kaipio J and Somersalo E 2004 *Statistical and Computational Inverse Problems* (Applied Mathematical Sciences vol 160) (New York: Springer)
- Kaltenbacher B, Neubauer A and Scherzer O 2008 *Iterative Regularization Methods for Nonlinear Ill-Posed Problems* (Radon Series on Computational and Applied Mathematics) (Berlin, Boston: De Gruyter)
- Kaltenbacher B, Schöpfer F and Schuster T 2009 Iterative methods for nonlinear ill-posed problems in Banach spaces: convergence and applications to parameter identification problems *Inverse Problems* **25**
- Kekkonen H, Lassas M and Siltanen S 2016 Posterior consistency and convergence rates for Bayesian inversion with hypoelliptic operators *Inverse Problems* **32** 085005
- Knapik B, van der Vaart A W and van Zanten J H 2011 Bayesian inverse problems with Gaussian priors *Ann. Stat.* **39** 2626–657
- Knapik B, Szabò B, van der Vaart A W and van Zanten H 2015 Bayes procedures for adaptive inference in inverse problems for the white noise model *Probab. Theor. Relat. Field* **164** 771–13
- Knowles I 2001 Parameter identification for elliptic problems *J. Comput. Appl. Math.* **131** 175–94
- Kohn R V and Lowe B D 1988 A variational method for parameter identification *ESAIM Math. Model. Numer. Anal.* **22** 119–58
- Kravaris C and Seinfeld J 1985 Identification of parameters in distributed parameter systems by regularization *SIAM J. Control Optim.* **23**
- Li W V and Linde W 1999 Approximation, metric entropy and small ball estimates for Gaussian measures *Ann. Probab.* **27** 1556–578
- Lions J L and Magenes E 1972 *Non-Homogeneous Boundary Value Problems and Applications* (Grundlehren der mathematischen Wissenschaften) 1st edn (Berlin: Springer)
- Monard F, Nickl R and Paternain G P 2020 Consistent inversion of noisy non-abelian x-ray transforms *Comm. Pure Appl. Math.* accepted. arXiv:1905.00860
- Monard F, Nickl R and Paternain G P 2019 Efficient nonparametric Bayesian inference for x-ray transforms *Ann. Stat.* **47** 1113–147
- Nickl R and Söhl J 2017 Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions *Ann. Stat.* **45** 1664–693
- Nickl R and Söhl J 2019 Bernstein–von Mises theorems for statistical inverse problems II: compound Poisson processes *Electron. J. Stat.* **13** 3513–571

- Nickl R, van de Geer S and Wang S 2020 Convergence rates for penalised least squares estimators in PDE-constrained regression problems *SIAM/ASA J. Uncertain. Quantification* **8** 374–413
- Nickl R 2018 Bernstein-von Mises theorems for statistical inverse problems I: Schrödinger equation *J. Eur. Math. Soc.* accepted. arXiv:[1707.01764](https://arxiv.org/abs/1707.01764)
- Qi-nian J 2000 Error estimates of some newton-type methods for solving nonlinear inverse problems in Hilbert scales *Inverse Problems* **16** 187–97
- Ray K 2013 Bayesian inverse problems with non-conjugate priors *Electron. J. Stat.* **7** 2516–549
- Richter G R 1981 An inverse problem for the steady state diffusion equation *SIAM J. Appl. Math.* **41** 210–21
- Schwab C and Stuart A M 2012 Sparse deterministic approximation of Bayesian inverse problems *Inverse Problems* **28** 045003
- Stuart A M 2010 Inverse problems: a Bayesian perspective *Acta Numer.* **19** 451–59
- Triebel H 1978 *Interpolation Theory, Function Spaces, Differential Operators* (Berlin: VEB Deutscher Verlag der Wissenschaften)
- van der Vaart A W and van Zanten J H 2008 Rates of contraction of posterior distributions based on Gaussian process priors *Ann. Stat.* **36** 1435–463
- van der Vaart A W and van Zanten J H 2009 Adaptive Bayesian estimation using a Gaussian random field with inverse gamma bandwidth *Ann. Stat.* **37** 2655–675
- Vollmer S J 2013 Posterior consistency for Bayesian inverse problems through stability and regression results *Inverse Problems* **29** 125011