# Bernstein-von Mises theorems for statistical inverse problems I: Schrödinger equation 

Received July 6, 2017 and in revised form March 22, 2018


#### Abstract

We consider the inverse problem of determining the potential $f>0$ in the partial differential equation $$
\frac{\Delta}{2} u-f u=0 \quad \text { on } \mathcal{O}, \quad u=g \text { on } \partial \mathcal{O}
$$ where $\mathcal{O}$ is a bounded $C^{\infty}$-domain in $\mathbb{R}^{d}$ and $g>0$ is a given function prescribing boundary values. The data consist of the solution $u$ corrupted by additive Gaussian noise. A nonparametric Bayesian prior for the function $f$ is devised and a Bernstein-von Mises theorem is proved which entails that the posterior distribution given the observations is approximated in a suitable function space by an infinite-dimensional Gaussian measure that has a 'minimal' covariance structure in an information-theoretic sense. As a consequence the posterior distribution performs valid and optimal frequentist statistical inference on various aspects of $f$ in the small noise limit.


Keywords. Bayesian nonlinear inverse problems, elliptic partial differential equations, inverse scattering problem, asymptotics of nonparametric Bayes procedures

## Contents

1. Introduction ..... 2
2. A statistical inverse problem for the Schrödinger equation ..... 4
3. A posterior consistency result ..... 6
3.1. Basic notation ..... 6
3.2. Wavelet bases for $L^{2}(\mathcal{O})$ ..... 7
3.3. Construction of the prior distribution ..... 9
3.4. A contraction theorem in Sobolev norms ..... 10
4. Information geometry ..... 11
4.1. Score operator and LAN expansion ..... 11
4.2. Information bound for one-dimensional subproblems ..... 13
4.3. The 'canonical' Gaussian measure $\mathcal{N}_{f_{0}}$ ..... 14
5. A Bernstein-von Mises theorem in $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$ ..... 15
5.1. Some applications to uncertainty quantification ..... 17
6. Proofs ..... 18
6.1. Proofs of Propositions 6 and 2 ..... 18
6.2. Proof of Theorem 1 ..... 21
6.3. Proof of Theorem 8 ..... 26
6.4. Proofs for Section 5.1 ..... 43
7. Appendix ..... 44
7.1. The metric entropy inequality for sub-Gaussian processes ..... 44
7.2. Some properties of Schrödinger operators ..... 44
7.3. Some properties of Hölder-type spaces ..... 48
7.4. Likelihood functions and a contraction theorem for general inverse problems ..... 49
7.5. Information lower bounds in function space ..... 51
7.6. Some facts about weak convergence of random probability measures ..... 51
References ..... 52

## 1. Introduction

Inverse problems form a vast and well studied area within applied mathematics. In the 'information age' we live in, algorithms that successfully solve these problems must be robust to the presence of statistical noise and measurement error. A principled approach to such statistical inverse problems is the Bayesian one, and it has been shown in influential work in the last decade that modern MCMC methodology can be used to construct computationally efficient Bayesian algorithms for complicated non-linear inverse problems in infinite-dimensional settings. This methodology is attractive for scientists because the Bayesian posterior distribution automatically delivers an estimate of the statistical uncertainty in the reconstruction, and hence suggests 'confidence' intervals that allow one to reject or accept scientific hypotheses. The literature on applications of Bayes procedures in inverse problems is growing rapidly and cannot be reviewed here; we only mention Andrew Stuart's survey papers [17, 50] and the contributions [13-16, 28, 35, 44, 49], where many further references can be found.

Algorithms that solve ill-posed inverse problems typically involve a regularisation step, for instance via a penalised variational procedure or a spectral cut-off. In the Bayesian approach this step is provided by the prior distribution. If the prior represents a regularisation tool rather than subjective prior beliefs about the state of nature, then a natural question arises as to whether such Bayesian algorithms deliver adequate inferential conclusions that are independent of the prior. If so, one may further ask whether Bayes solutions of inverse problems allow for recovery of an unknown parameter $f$ in a statistically optimal way. Of particular importance in this context is to understand whether 'credible regions' constructed from the posterior distribution are objectively valid, approximate 'frequentist' confidence sets. A paradigm to answer these questions is provided by performing an analysis of the Bayesian algorithm in the 'large sample' or 'small noise' limit, and under the assumption that the data are generated from a fixed 'true' function $f=f_{0}$ (instead of $f$ being drawn at random from the prior distribution). That Bayesian inference can be valid in this setting has been well studied in mathematical statistics since Laplace [34, Chapter VI]. We refer to the recent monograph [24] and also to [27, Chapter 7.3] for an account of such 'frequentist consistency results' for non-parametric Bayes procedures in standard statistical models, and to the recent paper [51] and its discussion for results concerning the particular question of non-parametric Bayesian credible regions and their frequentist properties.

For linear inverse problems the theory from Bayesian non-parametric statistics often carries over to the inverse setting, using the singular value decomposition (SVD) of the forward operator and/or conjugacy of Gaussian priors. See the papers [33], [45] and [3] and also the more recent references [31, 32, 40, 51]. From these results one can deduce objective guarantees for the Bayesian approach, and in fact Bayes point estimates (such as MAP statistics) can be shown to be closely linked to commonly used Tikhonov regularisation or penalisation methods in these inverse problems (see [16], [28]).

In the case of non-linear statistical inverse problems, which include many important examples arising with partial differential equations (PDEs), little is known about the frequentist performance of Bayesian methods. For the problem of inferring the diffusion coefficient from noisy observations of the solution of an elliptic PDE, [59] provides some contraction rate results that entail that the posterior distribution arising from certain priors concentrates on a neighbourhood of the true parameter $f_{0}$ (although the conditions and rates obtained there are far from optimal). Recently it was shown in [43] in a related parabolic problem that Bayesian inference for the coefficients of a scalar elliptic differential operator based on discrete samples of the associated diffusion Markov process can result in optimal posterior contraction rates about the true parameter. A form of weak consistency in this model had earlier been proved in [56], and a related recent contribution is [1]. While consistency and contraction rates are relevant results, they do not per se justify Bayesian 'credible sets' and related uncertainty quantification methodology. Such guarantees may be derived from more precise Bernstein-von Mises theorems, which establish that the posterior distribution is approximated by a canonical Gaussian distribution in the small noise or large sample limit. While well understood in the finite-dimensional case [36,57], the Bernstein-von Mises phenomenon is more subtle in the setting of highand infinite-dimensional statistical models (see [22]), and for non-linear inverse problems such results are currently not available; exceptions are [38] where normal approximations of posteriors in Euclidean spaces of increasing dimension are obtained, but under conditions whose suitability for inverse problems is unclear, and also the sequel [42] to the present paper.

This article constitutes an attempt to advance our understanding of the statistical performance of Bayes algorithms for non-linear inverse problems. In view of the importance of inverse problems arising in PDE models (see the various examples in the survey papers [17,50]) we study here a basic situation where one wishes to recover a coefficient of an elliptic partial differential operator from an observation of a solution to the associated PDE under some boundary conditions, corrupted by additive Gaussian noise. In fact we shall lay out the theory in a simple 'Laplacian plus potential' situation which comprises all the conceptual difficulties but permits a cleaner analytic exposition of the main ideas behind the results we obtain. We first derive minimax optimal (within log-factors) rates of posterior contraction about the unknown potential term in $L^{2}$-distance (Theorems 1 and 13). The main contribution of this article however is Theorem 8 which provides an infinite-dimensional Bernstein-von Mises theorem for the posterior distribution resulting from a carefully chosen series prior for the potential term. It is shown that the posterior measure is approximated, in a sense to be made precise, by a certain 'canonical' Gaussian measure whose covariance structure attains the statistical information lower bound
for this inference problem. As a consequence the resulting posterior parameter inferences are statistically optimal from an objective, information-theoretic and asymptotic minimax point of view.

In proving our main result we follow the program put forward in [9, 10] (see also [27, Section 7.3.4]) and understand the Bernstein-von Mises problem in infinite-dimensional statistical models as one of showing weak convergence of the (scaled and centred) posterior measure to a fixed Gaussian measure in a function space that is large enough to support the limit distribution as a tight probability measure. The topology of this function space turns out to be weaker than the standard $L^{p}$-norm that one might otherwise consider, but one can show, again following [9], that arguments from interpolation theory imply that Bayesian credible sets constructed for these weak norms are valid frequentist confidence sets whose diameter also converges to zero in the stronger $L^{1}$-distance.

In summary, our results give some theoretical support for the assertion that the Bayes approach can in principle be expected to provide efficient solutions of non-linear statistical inverse problems, and further that the associated uncertainty quantification can be objectively valid in the large sample/small noise limit. By emphasising 'in principle' we wish to point out though very clearly that our results should not be construed as giving general guarantees for Bayes solutions of arbitrary inverse problems: Even though our proofs do give a template for obtaining similar theorems in other settings, the details depend strongly on the fact that the inverse problem associated with the Schrödinger equation is in a certain sense 'globally stable' on the parameter space we consider, and that the prior is taken to be supported in that parameter space (see also Remark 9). Obtaining similar results for different inverse problems or priors requires a careful analysis of various properties of the forward operator and of its linearisation, and it is conceivable that in certain situations the 'information geometry' induced by the forward map may be such that Bernstein-von Mises theorems do not in fact hold true. A comprehensive understanding of these questions for general non-linear problems remains a formidable challenge for future research in this area.

## 2. A statistical inverse problem for the Schrödinger equation

Let $\mathcal{O}$ be a bounded $C^{\infty}$-domain in $\mathbb{R}^{d}, d \geq 2$, with boundary $\partial \mathcal{O}$. Let $\overline{\mathcal{O}}$ be the closure of $\mathcal{O}$ in $\mathbb{R}^{d}$ and let $C(\overline{\mathcal{O}})$ be the space of continuous functions on $\overline{\mathcal{O}}$. For $\Delta=\sum_{i=1}^{d} \partial^{2} / \partial x_{i}^{2}$ the standard Laplacian operator, we consider the (time-independent) Schrödinger equation

$$
\begin{equation*}
\frac{\Delta}{2} u-f u=0 \quad \text { on } \mathcal{O}, \quad u=g \quad \text { on } \partial \mathcal{O}, \tag{1}
\end{equation*}
$$

where $g: \partial \mathcal{O} \rightarrow \mathbb{R}$ is a given function prescribing boundary values, and $f \in C(\overline{\mathcal{O}})$ is a non-negative 'potential'. For $f>0$ and $g$ sufficiently regular, a unique solution $u_{f} \in C(\overline{\mathcal{O}})$ to (1) exists and has probabilistic representation in terms of the FeynmanKac formula

$$
\begin{equation*}
u_{f}(x)=u_{f, g}(x)=E^{x}\left[g\left(X_{\tau_{\mathcal{O}}}\right) e^{-\int_{0}^{\tau \mathcal{O}} f\left(X_{s}\right) d s}\right], \quad x \in \mathcal{O}, \tag{2}
\end{equation*}
$$

where ( $X_{s}: s \geq 0$ ) is a $d$-dimensional Brownian motion started at $x \in \mathcal{O}$, with exit time $\tau_{\mathcal{O}}$ from $\mathcal{O}$ satisfying

$$
\sup _{x \in \mathcal{O}} E^{x} \tau_{\mathcal{O}} \leq K(\operatorname{vol}(\mathcal{O}), d)<\infty
$$

We refer to Section 7.2 below, particularly Proposition 25, for details.
In inverse problems terminology, $f \mapsto u_{f}$ is the forward map, and the inverse problem is to recover $f$ given $u_{f}$ (and $g$ ). If $\|f\|_{\infty} \equiv \sup _{x \in \mathcal{O}}|f(x)| \leq D$ and $\inf _{x \in \partial \mathcal{O}} g(x) \geq$ $g_{\text {min }}>0$ then we can apply Jensen's inequality to (2) to obtain

$$
\begin{equation*}
u_{f}(x) \geq g_{\min } e^{-\|f\|_{\infty} E^{x} \tau_{\mathcal{O}}} \geq c>0, \quad c=c\left(g_{\min }, D, \operatorname{vol}(\mathcal{O}), d\right) \tag{3}
\end{equation*}
$$

so that given $u=u_{f}$ we can solve for $f$ simply by taking

$$
\begin{equation*}
f(x)=\frac{1}{2} \frac{\Delta u}{u}(x), \quad x \in \mathcal{O} . \tag{4}
\end{equation*}
$$

The potential $f$ models an attenuation of the solution of the Dirichlet boundary value problem for the standard Laplace equation. In physical language, $f$ describes a local 'cooling' of the equilibrium temperature distribution of the classical heat equation, when initial 'boundary temperatures' are given by $g$. In the theory that follows, $\Delta$ could be replaced by a general, strongly elliptic second order partial differential operator with known coefficients, at the expense of mostly notational changes. Recent applications of inverse problems of this kind can be found, e.g., in photo-acoustics [4] or scattering problems [5, 29].

The question we ask here is how to optimally solve this non-linear inverse problem when the observations are corrupted by statistical noise. A natural measurement model to consider is

$$
Y_{i}=u_{f}\left(x_{i}\right)+w_{i}, \quad i=1, \ldots, n, n \in \mathbb{N},
$$

where the $w_{i}$ are independent standard normal $N(0,1)$ random variables, and the $x_{i}$ are 'equally spaced' approximate lattice points in the domain $\mathcal{O}$. By standard arguments from asymptotic statistics (see [6,47] or [27, Section 1.2.3]) this discrete measurement model is asymptotically (as $n \rightarrow \infty$ ) equivalent to observing the functional equation

$$
\begin{equation*}
Y=u_{f}+\varepsilon \mathbb{W}, \quad \varepsilon=1 / \sqrt{n}, \tag{5}
\end{equation*}
$$

where $\mathbb{W}$ is a Gaussian white noise (see Section 7.4 for details) in the Hilbert space

$$
L^{2}(\mathcal{O})=\left\{h: \mathcal{O} \rightarrow \mathbb{R}:\|h\|_{L^{2}(\mathcal{O})}^{2} \equiv \int_{\mathcal{O}} h(x)^{2} d x<\infty\right\}
$$

and $\varepsilon>0$ is the noise level. We will develop the theory in this equivalent 'continuous' model as it allows for a cleaner exposition of the key ideas.

The noise process $\mathbb{W}$ belongs to the Sobolev space $H^{\eta}(\mathcal{O})$ only for $\eta<-d / 2$ and is not pointwise defined-direct recovery of $f$ from $Y$ by a simple equation such as (4) thus becomes impossible. So while the noiseless inverse problem has an essentially straightforward solution, the statistical one has not: Even though (5) is a standard non-linear Gaus-
sian regression model, the parameter space of admissible regression functions $u_{f}$ carries non-trivial non-linear constraints (since the $u_{f}$ have to be solutions of the PDE (1)), and to deal with these constraints in a statistically efficient way is a non-obvious task.

A principled approach to solve such problems is the Bayesian one, which devises a prior distribution $\Pi$ for the 'unknown' function $f$. More precisely, for $\mathcal{F} \subseteq L^{2}(\mathcal{O})$ some parameter space, we consider $f$ distributed according to $\Pi$ where $\Pi$ is some ('prior') probability measure on the trace Borel $\sigma$-field $\mathcal{B}$ of $\mathcal{F}$. Then by standard results (see Section 7.4 below), for any $u_{f} \in L^{2}(\mathcal{O})$ we can define the likelihood function $p_{f}(Y)$ describing the density of the law $P_{f}^{Y} \equiv P_{u_{f}}^{Y}$ of $Y \mid f$ in (5) for a suitable fixed dominating measure. The posterior distribution $\Pi(\cdot \mid Y)$ is the law of $f \mid Y$ obtained from Bayes' formula

$$
\begin{equation*}
\Pi(B \mid Y)=\frac{\int_{B} p_{f}(Y) d \Pi(f)}{\int_{\mathcal{F}} p_{f}(Y) d \Pi(f)} \equiv \frac{\int_{B} e^{\ell(f)} d \Pi(f)}{\int_{\mathcal{F}} e^{\ell(f)} d \Pi(f)}, \quad B \in \mathcal{B} . \tag{6}
\end{equation*}
$$

As is common in the inverse problems literature, we consider the situation where $f$ has some regularity, say a prescribed number of bounded continuous derivatives on $\mathcal{O}$. We will see how the regularity influences our ability to reconstruct $f$ from $Y, g$, and we will show that a suitable Bayesian algorithm recovers $f$ in an optimal way in various loss functions, such as $L^{2}(\mathcal{O})$ or certain Sobolev norms.

## 3. A posterior consistency result

### 3.1. Basic notation

For $\mathcal{O} \subseteq \mathbb{R}^{d}$ an open set and a multi-index $i=\left(i_{1}, \ldots, i_{d}\right), i_{j} \in \mathbb{N} \cup\{0\}$, of length $|i|=\sum_{j} i_{j}$ (not to be confused with $|\cdot|$ otherwise denoting the standard Euclidean norm on $\mathbb{R}^{d}$ ), and $D^{i}$ the associated (weak) partial differential operator, the usual Sobolev spaces are defined as

$$
\begin{equation*}
H^{\alpha}(\mathcal{O})=\left\{f \in L^{2}(\mathcal{O}): D^{i} f \in L^{2}(\mathcal{O}) \forall|i| \leq \alpha\right\}, \quad \alpha \in \mathbb{N} \tag{7}
\end{equation*}
$$

where $H^{0}(\mathcal{O})=L^{2}(\mathcal{O})$. When no confusion may arise we will sometimes omit $\mathcal{O}$ in the notation, and we take as norm on $H^{\alpha}$ the functional

$$
\begin{equation*}
\|f\|_{H^{\alpha}}=\sum_{|i| \leq \alpha}\left\|D^{i} f\right\|_{L^{2}} \tag{8}
\end{equation*}
$$

For $\alpha \notin \mathbb{N}$ one defines $H^{\alpha}(\mathcal{O})$ by interpolation (see [37, Chapter I.9] or [2, Section 7]).
We can similarly define the spaces $C^{\alpha}(\mathcal{O}), \alpha \in \mathbb{N} \cup\{0\}$, by replacing $\left(L^{2},\|\cdot\|_{L^{2}}\right)$ in (7), (8) by the space $\left(C(\mathcal{O}),\|\cdot\|_{\infty}\right)$ of bounded uniformly continuous functions on $\mathcal{O}$-for elements $f \in C^{\alpha}(\mathcal{O})$ the functions $D^{i} f, 0 \leq|i| \leq \alpha$, then all have unique continuous extensions to $\overline{\mathcal{O}}$, and we sometimes write $C^{\alpha}(\overline{\mathcal{O}})$ to make explicit that we view such functions as being defined on $\overline{\mathcal{O}}$. The symbol $C^{\infty}(\mathcal{O})$ denotes the set of all infinitely differentiable functions on $\mathcal{O}$, and $C_{0}(\mathcal{O})$ denotes those elements in $C(\overline{\mathcal{O}})$ whose boundary trace $h_{\mid \partial \mathcal{O}}$ is zero. For $\alpha \notin \mathbb{N}$ we say $f \in C^{\alpha}(\mathcal{O})$ if $D^{i} f,|i|=[\alpha]$, where $[\alpha]$ is the
integer part of $\alpha$, exists and is $\alpha-[\alpha]$-Hölder continuous. The norm on the space $C^{\alpha}(\mathcal{O})$ is then given by

$$
\begin{equation*}
\|f\|_{C^{\alpha}(\mathcal{O})} \equiv\|f\|_{C^{[\alpha]}(\mathcal{O})}+\max _{|i|=[\alpha]} \sup _{x, y \in \mathcal{O} ; x \neq y} \frac{\left|D^{i} f(x)-D^{i} f(y)\right|}{|x-y|^{\alpha-[\alpha]}} . \tag{9}
\end{equation*}
$$

We also need the Hölder-Zygmund spaces $\mathcal{C}^{\alpha}(\mathcal{O}), \alpha \geq 0$ (see [52, Section 3.4.2] for definitions). One has $\mathcal{C}^{\alpha}(\mathcal{O})=C^{\alpha}(\mathcal{O}), \alpha \notin \mathbb{N}$, (with equivalent norms) and continuous imbeddings $\mathcal{C}^{\alpha^{\prime}} \subsetneq C^{\alpha} \subsetneq \mathcal{C}^{\alpha}, \alpha \in \mathbb{N} \cup\{0\}, \alpha^{\prime}>\alpha$. Attaching a subscript $c$ to any space $S(\mathcal{O})$ will denote the subspace $\left(S_{c}(\mathcal{O}),\|\cdot\|_{S(\mathcal{O})}\right)$ consisting of functions of compact support in $\mathcal{O}$, and $\left(S_{K}(\mathcal{O}),\|\cdot\|_{S(\mathcal{O})}\right)$ will denote the subspace consisting of elements of $S(\mathcal{O})$ supported in a subset $K \subseteq \mathcal{O}$.

For $\alpha>d / 2$ and $\mathcal{O}$ a bounded $C^{\infty}$-domain, the Sobolev imbedding implies that $H^{\alpha}(\mathcal{O})$ embeds continuously into $C^{\beta}(\overline{\mathcal{O}})$ for any $0<\beta<\alpha-d / 2$ and we further have

$$
\begin{equation*}
\|f g\|_{H^{\alpha}} \leq c\|f\|_{H^{\alpha}}\|g\|_{H^{\alpha}}, \quad \alpha>d / 2, \tag{10}
\end{equation*}
$$

for some $c=c(\alpha, d, \mathcal{O})$. The above facts are classical for $\alpha \in \mathbb{N}$ (see [2]) and also hold for $\alpha \notin \mathbb{N}$ by the use of interpolation theory [37,52]. We also repeatedly use the inequalities

$$
\begin{equation*}
\|f g\|_{H^{\alpha}} \leq c\|f\|_{\mathcal{C}^{\alpha}}\|g\|_{H^{\alpha}}, \quad\|f g\|_{\mathcal{C}^{\alpha}} \leq c\|f\|_{\mathcal{C}^{\alpha}}\|g\|_{\mathcal{C}^{\alpha}}, \quad \alpha \geq 0 \tag{11}
\end{equation*}
$$

which follow from [52, Remark 1, p. 143].
On a Hilbert space $H$ we will denote by $\langle\cdot, \cdot\rangle_{H}$ the inner product generating the norm $\|\cdot\|_{H}$. For an arbitrary normed linear space $\left(X,\|\cdot\|_{X}\right)$, the topological dual space is

$$
\begin{aligned}
X^{*} & =\left(X,\|\cdot\|_{X}\right)^{*} \\
: & =\left\{L: X \rightarrow \mathbb{R}:|L(x)| \leq C\|x\|_{X} \text { for all } x \in X \text { and some } C>0\right\}
\end{aligned}
$$

which is a Banach space for the norm $\|L\|_{X^{*}} \equiv \sup _{x \in X,\|x\|_{X} \leq 1}|L(x)|$.
If $\mu$ is a probability measure on some measurable space, then $X \sim \mu$ means that $X$ is a random variable in that space drawn from distribution $\mu$, also called the law of $X, \mathcal{L}(X)=\mu$. We write $X \stackrel{\mathcal{L}}{=} Y$ if two random variables $X, Y$ have the same law, $\mathcal{L}(X)=\mathcal{L}(Y)$.

We will sometimes use the symbols $\lesssim, \gtrsim, \simeq$ to denote one- or two-sided inequalities up to multiplicative constants that may be either universal or 'fixed' in the context where the symbols appear. We also use the standard $O_{P}, o_{P}, O, o$ notation to estimate the order of magnitude of sequences of random variables and real numbers, respectively.

### 3.2. Wavelet bases for $L^{2}(\mathcal{O})$

Consider an orthonormal system of sufficiently smooth (' $S$-regular'), compactly supported Daubechies tensor wavelet basis functions

$$
\left\{\Phi_{l, r}: r \in \mathbb{Z}, l \in \mathbb{N} \cup\{-1,0\}\right\}, \quad \Phi_{l r}=2^{l d / 2} \Phi_{0, r}\left(2^{l} \cdot\right) \quad \text { for } l \geq 0
$$

of the Hilbert space $L^{2}\left(\mathbb{R}^{d}\right)$ (see $[18,39]$ and [27, Chapter 4]). [In proofs we shall sometimes use the last dilation identity also when $l=-1$, in slight abuse of notation.] We will use the fact that such a basis characterises elements of classical function spaces on $\mathbb{R}^{d}$ by the decay of the sequence norms of wavelet coefficients; for instance

$$
\begin{equation*}
\|f\|_{H^{\alpha}\left(\mathbb{R}^{d}\right)}^{2} \simeq \sum_{l, r} 2^{2 l \alpha}\left\langle f, \Phi_{l, r}\right\rangle_{L^{2}\left(\mathbb{R}^{d}\right)}^{2} \tag{12}
\end{equation*}
$$

and for some constant $C>0$ and all $\alpha \geq 0$,

$$
\begin{equation*}
f \in C^{\alpha}\left(\mathbb{R}^{d}\right) \Longrightarrow \sup _{l, r} 2^{l(\alpha+d / 2)}\left|\left\langle f, \Phi_{l, r}\right\rangle_{L^{2}\left(\mathbb{R}^{d}\right)}\right| \leq C\|f\|_{C^{\alpha}\left(\mathbb{R}^{d}\right)} \tag{13}
\end{equation*}
$$

with a converse of the last inequality holding as well when $\alpha \notin \mathbb{N}$. To be precise, the previous inequalities hold for all $\alpha \leq S$, where $S \in \mathbb{N}$ measures the 'regularity' of the wavelet basis used, in particular the $\Phi_{0, r}$ are in $C_{c}^{S}\left(\mathbb{R}^{d}\right)$. Note that $S$ can be chosen as large as desired.

For a bounded $C^{\infty}$-domain $\mathcal{O}$ in $\mathbb{R}^{d}$, one can then also construct an orthonormal wavelet basis of the Hilbert space $L^{2}(\mathcal{O})$ given by

$$
\begin{equation*}
\left\{\Phi_{l, r}^{\mathcal{O}}: r \leq N_{l}, l \in \mathbb{N} \cup\{-1,0\}\right\}, \quad N_{l} \in \mathbb{N} \tag{14}
\end{equation*}
$$

consisting of all those $\Phi_{l, r}^{\mathcal{O}}=\Phi_{l, r}$ that are compactly supported in $\mathcal{O}$, and of some boundary corrected wavelets $\Phi_{l, r}^{\mathcal{O}}=\Phi_{l, r}^{b c}$ which are orthonormalised linear combinations

$$
\Phi_{l, r}^{b c}(x)=\sum_{\left|m-m^{\prime}\right| \leq K} d_{m, m^{\prime}}^{l} \Phi_{l, m^{\prime}}(x), \quad m=m(l, r), K \in \mathbb{N}, x \in \mathcal{O}
$$

of those basic Daubechies wavelets $\Phi_{l, m^{\prime}}$ that have support both in and outside $\mathcal{O}$. We refer to [53, Theorem 2.33 and Definition 2.4] for details, but record the key properties that for all $l$ and some fixed positive constants $c_{0}, c_{1}, D$,

$$
\begin{align*}
& N_{l} \leq c_{0} 2^{l d} ; \quad \sum_{\left|m-m^{\prime}\right| \leq K}\left|d_{m, m^{\prime}}^{l}\right| \leq D  \tag{15}\\
& \operatorname{supp} \Phi_{l, r}^{b c} \subseteq\left\{x \in \mathcal{O}: \operatorname{dist}(x, \partial \mathcal{O})<c_{1} / 2^{l}\right\} ; \quad \sum_{r}\left|\Phi_{0, r}\right| \in C\left(\mathbb{R}^{d}\right)
\end{align*}
$$

the last property holding as well with $\Phi_{0, r}$ replaced by any derivative $D^{i} \Phi_{0, r},|i| \leq S$.
For the above basis any function $f \in L^{2}(\mathcal{O})$ has an orthogonal wavelet series expansion

$$
f=\sum_{l} \sum_{r=1}^{N_{l}}\left\langle f, \Phi_{l, r}^{\mathcal{O}}\right\rangle_{L^{2}(\mathcal{O})} \Phi_{l, r}^{\mathcal{O}} \quad \text { in } L^{2}(\mathcal{O})
$$

and we denote by $\Pi_{V_{J}}(f)$ the $J$-th partial sum of this series, equal to the $L^{2}(\mathcal{O})$-projection onto the linear span $V_{J}$ of $\left\{\Phi_{l, r}^{\mathcal{O}}: r \leq N_{l}, l \leq J\right\}$.

We next define Hölder-Zygmund type spaces for this wavelet basis by

$$
\begin{equation*}
f \in \mathcal{C}^{\alpha, W}(\mathcal{O}) \Longleftrightarrow \sup _{l, r} 2^{l(\alpha+d / 2)}\left|\left\langle f, \Phi_{l, r}^{\mathcal{O}}\right\rangle_{L^{2}(\mathcal{O})}\right| \equiv\|f\|_{\mathcal{C}^{\alpha, W}(\mathcal{O})}<\infty \tag{16}
\end{equation*}
$$

a definition that makes sense for all values $\alpha \in \mathbb{R}$ if $f$ is a linear functional whose action $f\left(\Phi_{l, r}^{\mathcal{O}}\right)$ on the $\left\{\Phi_{l, r}^{\mathcal{O}}\right\}$ is well defined.

The above boundary corrected wavelet basis conveniently retains the multi-scale and orthonormal basis properties in $L^{2}(\mathcal{O})$, but may not model the regularity of a function $f$ correctly near $\partial \mathcal{O}$. Describing smoothness of functions near the boundary by decay of wavelet coefficients is a delicate problem that we avoid here. For our purposes it will be sufficient that the spaces $\mathcal{C}_{c}^{\alpha}(\mathcal{O})$ and thus also $C_{c}^{\alpha}(\mathcal{O})$ are continuously embedded into $\mathcal{C}^{\alpha, W}(\mathcal{O})$ if $\alpha<S$ (see Proposition 26).

### 3.3. Construction of the prior distribution

Since $f>0$ is assumed, we seek a prior $\Pi$ for $f$ that is the law of a 'generic' nonnegative random function that possesses enough regularity so that the solution $u_{f}$ of (1) exists for every $f \sim \Pi$. We work with the $S$-regular wavelet basis (14) of $L^{2}(\mathcal{O})$ from the previous subsection, where $0<s<S$ is a fixed integer. We then take as prior distribution $\Pi=\Pi_{J}$ the law $\mathcal{L}(f)$ of the random function

$$
\begin{equation*}
f(x)=e^{\varphi(x)} \equiv \exp \left\{\sum_{l \leq J} \sum_{r=1}^{N_{l}} b_{l, r} \Phi_{l, r}^{\mathcal{O}}(x)\right\}, \quad x \in \mathcal{O}, J \in \mathbb{N}, \tag{17}
\end{equation*}
$$

where, for every $l$, the $\left(b_{l, r}: r=1, \ldots, N_{l}\right)$ are drawn independently and identically from the uniform distribution

$$
U\left(-B 2^{-l(s+d / 2)} \bar{l}^{-2}, B 2^{-l(s+d / 2)} \bar{l}^{-2}\right), \quad \bar{l}=\max (l, 1)
$$

For the prior to be fully adaptive one would have to further model $B, s, J$ by hierarchical priors, but for the results that follow we will restrict to the case where $s$ is given and $B$ an arbitrary but fixed positive constant. The truncation point $J \in \mathbb{N}$ will be chosen to increase as the noise level $\varepsilon$ decreases-thus $\Pi$ is a 'high-dimensional' prior.

The factors $2^{-l(s+d / 2)} \bar{l}^{-2}$ appearing in the weights $b_{l, r}$ imply that a function drawn from the prior lies in a ball of $C^{s}(\mathcal{O})$ almost surely: From (15) and if $|i| \leq s<S$ we deduce that

$$
\left|D^{i} \varphi\right|=\left|\sum_{l \leq J, r} b_{l, r} D^{i} \Phi_{l, r}^{\mathcal{O}}\right| \leq B \sum_{l \leq J, r} \bar{l}^{-2} 2^{l(|i|-s)}\left|D^{i} \Phi_{0, r}\right| \leq c B \sum_{l \leq J} \bar{l}^{-2} \leq C(B)
$$

for some finite constant $C(B)$, hence $\varphi$ is contained in $C^{\alpha}(\mathcal{O})$ for any $|\alpha| \leq s, s \in \mathbb{N}$. Since the exponential map is smooth on bounded sets, this further implies that

$$
\begin{equation*}
\|\varphi\|_{C^{s}(\mathcal{O})} \leq C, \quad\|f\|_{C^{s}(\mathcal{O})} \leq C \tag{18}
\end{equation*}
$$

for some finite constant $C$ that depends only on $s, B$ and the wavelet basis used.
'Besov priors' as in (17) have been proposed in inverse problem settings before in [15, 35, 45, 49]; particularly [49] studies such series priors with uniformly distributed coefficients in related PDE-type inverse problems. Other priors than (17) may be of interest, for instance those where the $b_{l, r}$ are drawn from a Gaussian or Laplace distribution. The
mathematical techniques we develop in the present paper apply in principle to such priors too, but the assumption that $\Pi$ is supported in a fixed ball of $C^{s}(\mathcal{O})$ is used crucially in many places in our proofs, and cannot easily be relaxed. Generalising our results to priors with unbounded coefficients $b_{l, r}$ remains a challenging open problem for future research.

### 3.4. A contraction theorem in Sobolev norms

Our first result states that the posterior distribution is consistent in that it concentrates around any 'true' function $f_{0}$ that generates equation (5), and we quantify the 'contraction rate' in $L^{2}$ in terms of the smoothness $s$ of $f_{0}$.

To reduce technicalities related to boundary issues we restrict to functions $\varphi_{0}=$ $\log f_{0} \in C^{s}(\mathcal{O})$ that have compact support in $\mathcal{O}$. Any such $\varphi_{0}$ is contained in $C_{c}^{s}(\mathcal{O}) \subseteq$ $\mathcal{C}^{s, W}(\mathcal{O})$ by Proposition 26, which implies by (16) that the wavelet coefficients of $\varphi_{0}$ decay like $2^{-l(s+d / 2)}$ as $l \rightarrow \infty$. We will strengthen this decay assumption slightly to

$$
\begin{equation*}
\sup _{l, r} 2^{l(s+d / 2)} \bar{l}^{2}\left|\left\langle\varphi_{0}, \Phi_{l, r}\right\rangle_{L^{2}(\mathcal{O})}\right| \leq B \tag{19}
\end{equation*}
$$

to exactly match the decay of the coefficients $b_{l, r}$ in the wavelet prior (17)-this is equivalent to assuming some additional Hölder regularity of $\varphi_{0}$ on the logarithmic scale. We will also assume $s \in \mathbb{N}$ for simplicity. Finally, and without loss of generality, we realise $g=g_{\mid \partial \mathcal{O}}$ as the boundary trace of an element of $C^{s+2}(\overline{\mathcal{O}})$. We then have the following contraction result.

Theorem 1. Let $f_{0}>0$ be such that $\varphi_{0}=\log f_{0} \in C_{c}^{s}(\mathcal{O})$ satisfies (19) for some $s>2+d / 2, s \in \mathbb{N}$. Let $\Pi=\Pi_{J}$ be as in (17) with $J \in \mathbb{N}$ such that $2^{J} \simeq \varepsilon^{-2 /(2 s+4+d)}$, and let $\Pi(\cdot \mid Y)$ be the resulting posterior distribution (6) arising from observing (5) with $g \in C^{s+2}(\overline{\mathcal{O}}), g \geq g_{\min }>0$. If $P_{f_{0}}^{Y}$ is the law of $Y=u_{f_{0}}+\varepsilon \mathbb{W}$, then for all $M$ large enough and $\gamma=s /(2 s+4)$ we have, as $\varepsilon \rightarrow 0$,

$$
\Pi\left(f:\left\|f-f_{0}\right\|_{L^{2}(\mathcal{O})}>M \varepsilon^{2 s /(2 s+4+d)} \log ^{\gamma}(1 / \varepsilon) \mid Y\right) \xrightarrow{P_{f_{0}}^{Y}} 0 .
$$

Since (18) implies that $f-f_{0}$ is bounded in $C^{s}(\mathcal{O}) \subseteq H^{s}(\mathcal{O})$, the standard interpolation inequality for Sobolev norms (see (49) below) further implies the contraction rates

$$
\begin{equation*}
\Pi\left(f:\left\|f-f_{0}\right\|_{H^{\alpha}(\mathcal{O})}>M \varepsilon^{(2 s-2 \alpha) /(2 s+4+d)} \log ^{\gamma^{\prime}}(1 / \varepsilon) \mid Y\right) \xrightarrow{P_{f_{0}}^{Y}} 0, \quad 0 \leq \alpha<s \tag{20}
\end{equation*}
$$

with $\gamma^{\prime}=(s-\alpha) /(2 s+4)$, which by the Sobolev imbedding (any $\alpha>d / 2$ ) also implies a contraction rate in the uniform norm $\|\cdot\|_{\infty}$.

Except for the log-factor, the rates obtained in Theorem 1 are optimal, as the following proposition shows. Removal of the log-factor is possible by more sophisticated prior choices, but for the main results of this article that follow, this will not be relevant.

Proposition 2. For every $s>2+d / 2, B>0,0<\alpha<s, f_{\min }>0$, any $g$ as in Theorem 1 and $E_{f}^{Y}$ the expectation operator corresponding to $P_{f}^{Y}$ we have, as $\varepsilon \rightarrow 0$,

$$
\inf _{\tilde{f}=\tilde{f}(Y, g)} \sup _{f \in C^{s}(\mathcal{O}):\|f\|_{C^{s}(\mathcal{O})} \leq B, f \geq f_{\min }>0} \varepsilon^{-(2 s-2 \alpha) /(2 s+4+d)} E_{f}^{Y}\|\tilde{f}-f\|_{H^{\alpha}(\mathcal{O})} \simeq c,
$$

where $c=c\left(\alpha, s, B, g_{\min }, f_{\min }, d, \operatorname{vol}(\mathcal{O})\right)$ is a finite positive constant.
Inspection of the proof of this proposition reveals that the minimax convergence rate $\varepsilon^{(2 s-2 \alpha) /(2 s+4+d)}$ improves by at most a power of $\log (1 / \varepsilon)$ if assumption (19) is also imposed on $f$, and also that in case $\alpha=0, L^{2}$-loss can be replaced by $L^{p}$-loss for any $1 \leq p<\infty$. Moreover the lower bound for the minimax rate can be shown to remain valid when $f$ is such that $\log f$ has compact support in $\mathcal{O}$.

## 4. Information geometry

While Theorem 1 is already quite satisfactory in that it shows that the posterior concentrates near the 'true' value, we may hope that the posterior distribution reveals more precise information about $f_{0}$. This will ultimately be formalised by proving a Bernsteinvon Mises theorem in a suitable function space, showing that the posterior distribution is approximated by a canonical Gaussian distribution with a covariance structure that is 'minimal' in an information-theoretic sense. In the language of asymptotic statistics $[30,36,57]$ the structure of this covariance is determined by the 'LAN' expansion of the log-likelihood ratio process, which we obtain now for the observation scheme (5) considered here. From it we can then construct the Gaussian measure that has to appear as the limit of the scaled and centred posterior measure.

### 4.1. Score operator and LAN expansion

We start with the following simple lemma which is a standard application of equation (110) in the Appendix.

Lemma 3. Let $\ell(f)=\ell(f, Y)=\log p_{f}(Y)$ be the log-likelihood function arising from (5), write $G(f)=u_{f}$, and assume $Y=G\left(f_{0}\right)+\varepsilon \mathbb{W}$ for some fixed $f_{0}$ such that $G\left(f_{0}\right) \in L^{2}(\mathcal{O})$. Then for any $f, g$ with $G(f), G(g) \in L^{2}(\mathcal{O})$ we have

$$
\begin{aligned}
& \ell(f)-\ell(g) \\
& =-\frac{1}{2 \varepsilon^{2}}\left(\left\|G(f)-G\left(f_{0}\right)\right\|_{L^{2}(\mathcal{O})}^{2}-\left\|G(g)-G\left(f_{0}\right)\right\|_{L^{2}(\mathcal{O})}^{2}\right)+\frac{1}{\varepsilon}\langle G(f)-G(g), \mathbb{W}\rangle_{L^{2}(\mathcal{O})} .
\end{aligned}
$$

Since $G$ is non-linear we next have to find a suitable linear approximation to $G(f)-$ $G(f+h)$ for a small perturbation $h$ of $f$. In statistical terminology this amounts to finding the 'score operator' of the model. The following proposition shows that this score operator can be understood as the solution operator for an inhomogeneous Schrödinger
equation (22) (see Section 7.2 for existence of its solutions and further details). The remainder terms of the linear approximation can be controlled in a 'weak' norm that we define now: we introduce the functional

$$
\begin{equation*}
\|f\|_{\left(H_{0}^{2}\right)^{*}}=\sup _{\varphi \in C_{0}(\mathcal{O}):\|\varphi\|_{H^{2}(\mathcal{O})} \leq 1}\left|\int_{\mathcal{O}} \varphi(x) f(x) d x\right| . \tag{21}
\end{equation*}
$$

Clearly $\|f\|_{\left(H_{0}^{2}\right)^{*}} \leq\|f\|_{L^{2}}$ but $\|\cdot\|_{\left(H_{0}^{2}\right)^{*}}$ generates a strictly weaker topology than the norm of $L^{2}(\mathcal{O})$, a fact that will be useful in our proofs.

We also write $\|h\|_{(d)}=\|h\|_{L^{2}}$ when $d<4$ but $\|h\|_{(d)}=\|h\|_{\infty}$ for $d \geq 4$. This accommodates the fact that the Sobolev space $H^{2}(\mathcal{O})$, which plays a key role for regularity estimates of solutions of the Schrödinger equation, imbeds into $C(\mathcal{O})$ only when $d<4$.

Proposition 4 (Score operator). Let $G(f)=u_{f}$ solve (1) where $f \in C^{s}(\mathcal{O}), f>0$, and $g \in C^{s+2}(\overline{\mathcal{O}})$ for some $s>0$. For $h \in C(\overline{\mathcal{O}})$, denote by $D G_{f}[h]=v_{h}$ the solution $v$ to the inhomogeneous Schrödinger equation

$$
\begin{equation*}
\frac{\Delta}{2} v-f v=h u_{f} \quad \text { on } \mathcal{O}, \quad v=0 \quad \text { on } \partial \mathcal{O} \tag{22}
\end{equation*}
$$

Let $f, h$ further be such that $f+h \geq f_{\min }>0$ and $\|f\|_{C^{s}(\mathcal{O})}+\|h\|_{C^{s}(\mathcal{O})} \leq D$ for some $D, f_{\min }>0$. Then there exists $c_{1}=c_{1}\left(\mathcal{O}, d, D, f_{\min },\|g\|_{C^{s+2}(\overline{\mathcal{O}})}\right)$ such that

$$
\left\|G(f+h)-G(f)-D G_{f}[h]\right\|_{L^{2}} \leq c_{1}\|h\|_{(d)}\|h\|_{\left(H_{0}^{2}\right)^{*}}
$$

Moreover the score operator $D G_{f}$ satisfies the estimate

$$
\begin{equation*}
\left\|D G_{f}[h]\right\|_{L^{2}} \leq c_{2}\|h\|_{\left(H_{0}^{2}\right)^{*}} \leq c_{2}\|h\|_{L^{2}}, \quad h \in C(\overline{\mathcal{O}}) \tag{23}
\end{equation*}
$$

where the constant $c_{2}$ depends only on upper bounds for $\|f\|_{C^{s}(\mathcal{O})},\|g\|_{C^{s+2}(\overline{\mathcal{O}})}$ and on $\mathcal{O}, d$.

Proof. Let us write $u=u_{f}$ and $u_{h}=u_{f+h}$ in slight abuse of notation, in this proof. Then in $\mathcal{O}$ we must have

$$
(\Delta / 2) u_{h}-(f+h) u_{h}=(\Delta / 2) u-f u
$$

or equivalently

$$
(\Delta / 2)\left(u_{h}-u\right)-(f+h)\left(u_{h}-u\right)=h u .
$$

Thus if $v_{h}$ solves (22) then

$$
(\Delta / 2)\left(u_{h}-u-v_{h}\right)-(f+h)\left(u_{h}-u-v_{h}\right)=h v_{h}
$$

on $\mathcal{O}$ and $u_{h}-u-v_{h}=0$ on $\partial \mathcal{O}$. We conclude that $w_{h}=u_{h}-u-v_{h}$ is itself a solution to the inhomogeneous Schrödinger equation

$$
\begin{equation*}
(\Delta / 2) w_{h}-(f+h) w_{h}=h v_{h}, \quad f+h \geq f_{\min }>0 \tag{24}
\end{equation*}
$$

with zero boundary condition on $\partial \mathcal{O}$. Lemma 24 , specifically (100) below, and definition (21) now imply that

$$
\left\|w_{h}\right\|_{L^{2}} \leq C\left\|h v_{h}\right\|_{\left(H_{0}^{2}\right)^{*}} \leq C\|h\|_{\left(H_{0}^{2}\right)^{*}} \sup _{\|\varphi\|_{H^{2}} \leq 1}\left\|\varphi v_{h}\right\|_{H^{2}} .
$$

Using (10) when $d<4$ gives $\left\|\varphi v_{h}\right\|_{H^{2}} \leq c\|\varphi\|_{H^{2}}\left\|v_{h}\right\|_{H^{2}}$, whereas for $d \geq 4$ we use (11) and have to replace $\left\|v_{h}\right\|_{H^{2}}$ by $\left\|v_{h}\right\|_{\mathcal{C}^{2}}$ in the last inequality. Finally, since $v_{h}$ solves (22), by the regularity estimates in Lemma 24 with $\beta=0$ and since $\left\|u_{f}\right\|_{\mathcal{C}^{2}} \leq$ $c\left(D,\|g\|_{C^{s+2}(\overline{\mathcal{O}})}\right)$ by Proposition 25 ,

$$
\left\|v_{h}\right\|_{H^{2}} \leq C^{\prime}\left\|u_{f} h\right\|_{L^{2}} \leq C^{\prime \prime}\|h\|_{L^{2}}, \quad\left\|v_{h}\right\|_{\mathcal{C}^{2}} \leq C^{\prime}\left\|u_{f} h\right\|_{\mathcal{C}^{0}} \leq C^{\prime \prime}\|h\|_{\infty}
$$

proving the first claim. Inequality (23) now follows similarly, by using the definition of $D G_{f}$ and again Lemma 24, (11) and the bound on $\left\|u_{f}\right\|_{\mathcal{C}^{2}}$.
We immediately obtain the pathwise LAN expansion for the log-likelihood ratio process.
Proposition 5. For any $f_{0}, h \in C^{s}(\overline{\mathcal{O}}), f_{0}>0, g \in C^{s+2}(\overline{\mathcal{O}}), s>0$, if $Y=$ $G\left(f_{0}\right)+\varepsilon \mathbb{W}$ then

$$
\log \frac{p_{f_{0}+\varepsilon h}}{p_{f_{0}}}(Y)=\left\langle D G_{f_{0}}[h], \mathbb{W}\right\rangle_{L^{2}}-\frac{1}{2}\left\|D G_{f_{0}}[h]\right\|_{L^{2}}^{2}+o_{P_{f_{0}}^{Y}}(1) \quad \text { as } \varepsilon \rightarrow 0,
$$

and the LAN-norm is given by

$$
\|h\|_{\mathrm{LAN}} \equiv\left\|D G_{f_{0}}[h]\right\|_{L^{2}}
$$

Proof. Combine Lemma 3 with Proposition 4 and notice that $\varepsilon^{2}\|h\|_{(d)}\|h\|_{\left(H_{0}^{2}\right)^{*}} \leq$ $\varepsilon^{2}\|h\|_{\infty}^{2}$ as well as that $\langle g, \mathbb{W}\rangle=O_{P}\left(\|g\|_{L^{2}}\right)$ by Markov's inequality.

### 4.2. Information bound for one-dimensional subproblems

To apply the general theory of statistical efficiency (see Section 7.5 in the appendix for a review) to estimate a linear functional $\Psi(f)$ of $f$ at $f_{0}$, we need to find the Riesz representer of such a functional for the LAN inner product $\langle\cdot, \cdot\rangle_{\text {LAN }} \equiv\left\langle D G_{f_{0}}[\cdot], D G_{f_{0}}[\cdot]\right\rangle_{L^{2}(\mathcal{O})}$. We restrict to functionals of the form $\Psi(f)=\langle f, \psi\rangle_{L^{2}}$ for some fixed compactly supported function $\psi$. Using the results about the $f$-Green operator $V_{f}$ for the Schrödinger operator $S_{f}$ from Section 7.2 we have

$$
\left\langle h_{1}, h_{2}\right\rangle_{\mathrm{LAN}}=\left\langle D G_{f_{0}}\left[h_{1}\right], D G_{f_{0}}\left[h_{2}\right]\right\rangle_{L^{2}}=\left\langle V_{f_{0}}\left[u_{f_{0}} h_{1}\right], V_{f_{0}}\left[u_{f_{0}} h_{2}\right]\right\rangle_{L^{2}} .
$$

For $f_{0} \in C^{s}(\overline{\mathcal{O}}), g \in C^{s+2}(\overline{\mathcal{O}}), f_{0}, g>0, s>0$, we know from Proposition 25 and (3) that $u_{f_{0}} \geq c>0$ on $\overline{\mathcal{O}}$, and if also $s>2$ then the chain rule further implies $1 / u_{f_{0}} \in C^{4}(\mathcal{O})$. As a consequence, for any $\psi \in C_{c}^{4}(\mathcal{O})$ we can define

$$
\begin{equation*}
\widetilde{\Psi}=\widetilde{\Psi}_{f_{0}}=\left[S_{f_{0}} S_{f_{0}}\left[\psi / u_{f_{0}}\right]\right] / u_{f_{0}} \tag{25}
\end{equation*}
$$

which has the pointwise representation

$$
\begin{equation*}
\widetilde{\Psi} u_{f_{0}}=\frac{1}{4} \Delta^{2}\left[\frac{\psi}{u_{f_{0}}}\right]-\frac{1}{2} \Delta\left[\frac{f_{0} \psi}{u_{f_{0}}}\right]-\frac{f_{0}}{2} \Delta\left[\frac{\psi}{u_{f_{0}}}\right]+\frac{f_{0}^{2} \psi}{u_{f_{0}}}, \tag{26}
\end{equation*}
$$

and hence is also compactly supported in $\mathcal{O}$. Then for all $h \in C(\overline{\mathcal{O}})$, Proposition 22 implies

$$
\begin{align*}
\langle\tilde{\Psi}, h\rangle_{\mathrm{LAN}} & =\left\langle V_{f_{0}}\left[u_{f_{0}} \tilde{\Psi}\right], V_{f_{0}}\left[h u_{f_{0}}\right]\right\rangle_{L^{2}}=\left\langle V_{f_{0}} V_{f_{0}} S_{f_{0}} S_{f_{0}}\left[\psi / u_{f_{0}}\right], h u_{f_{0}}\right\rangle_{L^{2}} \\
& =\langle\psi, h\rangle_{L^{2}}, \tag{27}
\end{align*}
$$

so that $\tilde{\Psi}$ is indeed the Riesz representer. From the results in Section 7.5 with $H=C^{s}(\overline{\mathcal{O}})$ and again Proposition 22 it now follows that the information lower bound for estimating $\langle f, \psi\rangle$ at $f=f_{0}$ from observations $Y$ in (5) equals

$$
\begin{equation*}
\left\|\widetilde{\Psi}_{f_{0}}\right\|_{\mathrm{LAN}}^{2}=\left\|V_{f_{0}}\left[u_{f_{0}} \widetilde{\Psi}_{f_{0}}\right]\right\|_{L^{2}}^{2}=\left\|S_{f_{0}}\left[\psi / u_{f_{0}}\right]\right\|_{L^{2}}^{2} \tag{28}
\end{equation*}
$$

### 4.3. The 'canonical' Gaussian measure $\mathcal{N}_{f_{0}}$

Collecting the Riesz representers $\widetilde{\Psi}$ for all linear functions $\left(\Psi(f)=\langle f, \psi\rangle_{L^{2}(\mathcal{O})}\right.$ : $\psi \in C_{c}^{\infty}(\mathcal{O})$ ) from Section 4.2, and for $S_{f_{0}}$ the Schrödinger operator from Section 7.2, we define the centred Gaussian process
$\left(X(\psi): \psi \in C_{c}^{\infty}(\mathcal{O})\right)$ such that $E\left[X(\psi) X\left(\psi^{\prime}\right)\right]=\left\langle S_{f_{0}}\left[\psi / u_{f_{0}}\right], S_{f_{0}}\left[\psi^{\prime} / u_{f_{0}}\right]\right\rangle_{L^{2}}$.
By Kolmogorov's consistency theorem [19, Section 12.1]) this process defines a probability measure $\mathcal{N}_{f_{0}}$ on the cylindrical $\sigma$-field of $\mathbb{R}^{C_{c}^{\infty}(\mathcal{O})}$.

In the Bernstein-von Mises theorem that follows we will want to prove weak convergence in probability of the centred and scaled posterior measure towards $\mathcal{N}_{f_{0}}$ in a suitable Banach space. The notion of weak convergence we will employ below naturally requires tightness of the limit law, and we thus first investigate when the measure $\mathcal{N}_{f_{0}}$ extends to a tight Gaussian Borel probability measure in the scale of dual spaces $\left(C_{c}^{\alpha}(\mathcal{O})\right)^{*}, \alpha>0$.

Proposition 6. Let $f_{0} \in C^{s}(\mathcal{O})$ and $g \in C^{s+2}(\overline{\mathcal{O}}), s>0$, satisfy $f_{0}, g>0$.
(A) For any $\alpha>2+d / 2$, the law $\mathcal{N}_{f_{0}}$ induced by the Gaussian process (29) defines a tight Gaussian Borel probability measure on $\left(C_{c}^{\alpha}(\mathcal{O})\right)^{*}$.
(B) If $\alpha<2+d / 2$ then for every $f_{0} \in C^{s}(\mathcal{O}), s>2+d / 2$, we have

$$
\mathcal{N}_{f_{0}}\left(x:\|x\|_{\left(C_{c}^{\alpha}(\mathcal{O})\right)^{*}}<\infty\right)=0
$$

(C) Both $(\mathrm{A})$ and $(\mathrm{B})$ remain true if $\left(C_{c}^{\alpha}(\mathcal{O})\right)^{*}$ is replaced by the space $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}=$ $\left(C_{K}^{\alpha}(\mathcal{O}),\|\cdot\|_{C^{\alpha}(\mathcal{O})}\right)^{*}$ for any compact set $K \subseteq \mathcal{O}$ with non-empty interior.

To be precise, the proof of (A) shows that the process (29) has a version that acts linearly on $C_{c}^{\infty}(\mathcal{O})$, and that the cylindrical law of that linear version on $\left(C_{c}^{\alpha}(\mathcal{O})\right)^{*}=$ $\left(C_{c}^{\infty}(\mathcal{O}),\|\cdot\|_{C^{\alpha}(\mathcal{O})}\right)^{*} \subseteq \mathbb{R}^{C_{c}^{\infty}(\mathcal{O})}$ extends to a tight Gaussian measure on the Borel $\sigma$-field
induced by $\|\cdot\|_{\left(C_{c}^{\alpha}(\mathcal{O})\right)^{*}}$ when $\alpha>2+d / 2$. In (B), the proof implies that for any version of $(X(\psi))$ and $\alpha<2+d / 2$, the norm $\sup _{\psi:\|\psi\|_{C_{c}^{\alpha}(\mathcal{O})} \leq 1}|X(\psi)|$ is $\infty$ almost surely.

For $\alpha>d / 2$ the measure $\mathcal{N}_{f_{0}}$ is tight and thus the results in Sections 4.2 and 7.5 imply that its covariance structure represents the information lower bound for estimating $f$ in $\left(C_{c}^{\alpha}(\mathcal{O})\right)^{*}$ or $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$, based on observations from (5).

One may investigate the limiting case $\alpha=2+d / 2$ further, but this requires the introduction of a 'logarithmic scale' to measure smoothness of functions, and we abstain from doing so for ease of exposition. Note that one may also show that $\mathcal{N}_{f_{0}}$ is not tight when $\alpha=2+d / 2$ (see Remark 12).

## 5. A Bernstein-von Mises theorem in $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$

We now turn to the main result of this article which gives an infinite-dimensional normal approximation of the posterior distribution arising from the prior (17), in the small noise limit $\varepsilon \rightarrow 0$. The following condition on the prior and on $f_{0}$ will be employed. It requires $f_{0}$ to be an 'interior' point of the support of the prior (17)-just as in the finitedimensional situation a Gaussian approximation of the posterior distribution cannot be expected when the 'true value' lies at the boundary of the support supp $\Pi$ of $\Pi$ (see also Remark 9).

Condition 7. Let $s \in \mathbb{N}$ satisfy $s>\max (2+d / 2, d)$. Suppose the prior $\Pi$ arises from (17) with $J \in \mathbb{N}$ chosen such that $2^{J} \simeq \varepsilon^{-2 /(2 s+4+d)}$, and is based on an $S$-regular wavelet basis with $S>2+s+d$. Suppose that $f_{0}>0$ and $\varphi_{0}=\log f_{0} \in C_{c}^{S}(\mathcal{O})$ satisfies, for some fixed $\epsilon>0$,

$$
\left|\left\langle\varphi_{0}, \Phi_{l, r}^{\mathcal{O}}\right\rangle\right| \leq(B-\epsilon) 2^{-l(s+d / 2)} \bar{l}^{-2}, \quad \bar{l}=\max (1, l)
$$

Again, the restriction to $s \in \mathbb{N}$ is just for convenience and could be removed at the expense of introducing further technicalities in the proofs.

We will interpret the (scaled and centred) posterior distribution $\Pi(\cdot \mid Y)$ of the random function $f \mid Y$ as the law of the stochastic process

$$
\begin{equation*}
\left(\varepsilon^{-1}\langle f-\bar{f}, \psi\rangle_{L^{2}(\mathcal{O})} \mid Y:\|\psi\|_{C_{K}^{\alpha}(\mathcal{O})} \leq 1\right), \quad K \subseteq \mathcal{O} \tag{30}
\end{equation*}
$$

conditional on $Y$, where $\bar{f}=\bar{f}(Y)$ is the posterior mean (since $f \mid Y$ is bounded in $L^{2}(\mathcal{O})$, the posterior mean $E^{\Pi}[f \mid Y]$ is well defined as a Bochner integral (see, e.g., [27, p. 68]). As $\varepsilon \rightarrow 0$ the conditional laws of these stochastic processes will be shown to converge weakly to the law $\mathcal{N}_{f_{0}}$ of the Gaussian process $X$ from (29), uniformly in the collection of functions $\left\{\psi:\|\psi\|_{C_{K}^{\alpha}(\mathcal{O})} \leq 1\right\}$, with $P_{f_{0}}^{Y}$-probability approaching 1, whenever $\alpha>2+3 d / 2$, and for any compact subset $K \subseteq \mathcal{O}$. Since both processes induce a linear action on $C_{K}^{\alpha}(\mathcal{O})$, this is equivalent to weak convergence in probability of the induced probability measures $\mathcal{L}\left(\varepsilon^{-1}(f-\bar{f}) \mid Y\right)$ to $\mathcal{N}_{f_{0}}$ on the dual space $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}=\left(C_{K}^{\alpha}(\mathcal{O}),\|\cdot\|_{C^{\alpha}(\mathcal{O})}\right)^{*}$.

To make these notions of convergence rigorous, recall that a sequence of probability measures $\mu_{n}$ on a metric space $(S, \rho)$ converges weakly to $\mu$, written $\mu_{n} \rightarrow^{\mathcal{L}} \mu$, if $\int_{S} F d \mu_{n} \rightarrow \int_{S} F d \mu$ for all bounded $\rho$-continuous functions $F: S \rightarrow \mathbb{R}$. Whenever $\mu$ is tight, this notion of convergence can be metrised by the bounded Lipschitz (BL) metric

$$
\begin{align*}
\beta_{S}(v, \mu) & =\sup _{F: S \rightarrow \mathbb{R},\|F\|_{\text {Lip }} \leq 1}\left|\int_{S} F d(\mu-v)\right| \\
\|F\|_{\text {Lip }} & \equiv \sup _{x \in S}|F(x)|+\sup _{x \neq y, x, y \in S} \frac{|F(x)-F(y)|}{\rho(x, y)} \tag{31}
\end{align*}
$$

(see [20, Theorem 3.28]). It is therefore natural to say that a sequence of random probability measures converges weakly in probability to some tight limiting probability measure if their BL-distance converges to zero in probability. This definition is easily seen to be independent of the metric for weak convergence that is used. We refer to Section 7.6 for some basic facts about this notion of convergence.

Theorem 8. Let prior $\Pi$ and $f_{0}$ satisfy Condition 7 for the given $s>0$, and let $\Pi(\cdot \mid Y)$ be the posterior distribution (6) resulting from observing (5) with $g \in C^{s+2}(\mathcal{O}), g \geq$ $g_{\min }>0$. Let $f \sim \Pi(\cdot \mid Y)$ conditional on $Y$, let $\bar{f}=\bar{f}(Y)=E^{\Pi}[f \mid Y]$ be the posterior mean, and let $P_{f_{0}}^{Y}$ denote the law of $Y=G\left(f_{0}\right)+\varepsilon \mathbb{W}$.

For every compact subset $K \subseteq \mathcal{O}$ any $\alpha>2+3 d / 2$, as $\varepsilon \rightarrow 0$ we have

$$
\begin{equation*}
\beta_{\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}}\left(\mathcal{L}\left(\varepsilon^{-1}(f-\bar{f}) \mid Y\right), \mathcal{N}_{f_{0}}\right) \xrightarrow{P_{f_{0}}^{Y}} 0 \tag{32}
\end{equation*}
$$

where $\mathcal{L}\left(\varepsilon^{-1}(f-\bar{f}) \mid Y\right)$ is the law on $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$ induced by the stochastic process (30), and where $\mathcal{N}_{f_{0}}$ is the tight Gaussian Borel probability measure on $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$ constructed in Proposition 6. In particular $\bar{f}$ is an efficient estimator of $f_{0}$ : as $\varepsilon \rightarrow 0$ and under $P_{f_{0}}^{Y}$,

$$
\varepsilon^{-1}\left(\bar{f}-f_{0}\right) \xrightarrow{\mathcal{L}} \mathcal{N}_{f_{0}} \quad \text { in }\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}
$$

Remark 9. From the point of view of weak convergence towards $\mathcal{N}_{f_{0}}$, the condition $\alpha>2+d / 2$ would be sufficient (and necessary) in the above theorem (see Proposition 6). The stronger condition $\alpha>2+3 d / 2$ is, however, required in Theorem 8 for the prior constructed in (17), and we wish to give some intuition for this fact: In statistical models with a parameter space of fixed finite dimension, it is well known that the Bernstein-von Mises theorem does not hold when the true parameter lies at the boundary of the support of the prior. When proving Bernstein-von Mises theorems in high and infinite dimensions, this phenomenon becomes 'quantitative'-this was already observed in [7, 10] where priors with 'heavy tails' were required to obtain sub-Gaussian approximations of posterior distributions. In the non-linear inverse problem setting here, the prior (17) strongly regularises the potential $f$ to lie in a fixed ball of $C^{s}(\mathcal{O})$. As a consequence, the 'stability estimate' induced by (4) holds globally in the support of the prior (see the proof of Lemma 14), ensuring consistency of Bayesian inversion (Theorem 1). This choice
of prior, however, also limits the directions $\psi$ along which $\varepsilon^{-1}\langle f-\bar{f}, \psi\rangle_{L^{2}(\mathcal{O})} \mid Y$ is approximately sub-Gaussian by requiring $\psi$ to be sufficiently well aligned with the support ellipsoid described by the prior (see the 'change of measure' argument in the proof, particularly in the construction of admissible 'interior' directions in Lemma 19). While $\alpha>2+3 d / 2$ appears necessary for the prior (17), it remains an interesting open question whether the condition $\alpha>2+d / 2$ can be attained by other priors.

### 5.1. Some applications to uncertainty quantification

A key consequence of a Bernstein-von Mises result is that the uncertainty quantification provided by the posterior distribution for the parameter $f$ is optimal in a frequentist, priorindependent, sense if the credible sets are constructed for the norm in which the normal approximation to the posterior holds.

To start with a simple example, suppose $\psi \in C_{c}^{\alpha}(\mathcal{O}), \alpha>2+3 d / 2$, is a fixed 'test' function ( $\psi \neq 0$ to avoid triviality) and our target of statistical inference is the scalar quantity $\Psi(f)=\langle f, \psi\rangle_{L^{2}}$. We can use the induced posterior distribution $\Pi_{\psi}(\cdot \mid Y) \equiv$ $\Pi(\cdot \mid Y) \circ \Psi^{-1}$ to construct a prescribed level $1-\beta$ credible set for $\Psi(f)$ by taking

$$
\begin{align*}
& C_{\varepsilon}=\left\{z \in \mathbb{R}:\left|z-\langle\bar{f}, \psi\rangle_{L^{2}}\right| \leq R_{\varepsilon}\right\} \\
& \quad \text { with } R_{\varepsilon}=R_{\varepsilon}(Y, \beta) \text { such that } \Pi_{\psi}\left(C_{\varepsilon} \mid Y\right)=1-\beta, \beta>0 . \tag{33}
\end{align*}
$$

The Bernstein-von Mises and continuous mapping theorems now imply that the law of $\varepsilon^{-1}(\Psi(f)-\Psi(\bar{f}))$ is approximated in the small noise limit by an $N\left(0,\left\|S_{f_{0}}\left(\psi / u_{f_{0}}\right)\right\|_{L^{2}}^{2}\right)$ distribution. Let $\Phi(t)=\operatorname{Pr}(Z \in[-t, t]), t>0$, for $Z \sim N\left(0,\left\|S_{f_{0}}\left(\psi / u_{f_{0}}\right)\right\|_{L^{2}}^{2}\right)$ with continuous inverse $\Phi^{-1}:[0,1] \rightarrow \mathbb{R}$. The following corollary implies that the posterior credible set $C_{\varepsilon}$ is an efficient frequentist confidence set for the parameter $\Psi(f)$.
Corollary 10. The frequentist coverage probability of $C_{\varepsilon}$ satisfies

$$
P_{f_{0}}^{Y}\left(\Psi\left(f_{0}\right) \in C_{\varepsilon}\right) \rightarrow 1-\beta
$$

as $\varepsilon \rightarrow 0$, and the diameter of $C_{\varepsilon}$ is of order $R_{\varepsilon}=O_{P}(\varepsilon)$, in fact as $\varepsilon \rightarrow 0$,

$$
\varepsilon^{-1} R_{\varepsilon} \xrightarrow{P_{f_{0}}^{Y}} \Phi^{-1}(1-\beta) .
$$

The above result immediately generalises to fixed non-linear but differentiable functionals $\Psi$ of $f$, by arguing just as in [9, Section 2.3.2].

For the proof of the previous corollary one would not require the full strength of Theorem 8. But the infinite-dimensional approximation is required to demonstrate that also the uncertainty quantification provided by the posterior distribution for the entire parameter $f$-by intersecting all admissible linear constraints in (33) for $\psi$ in the unit ball of $C_{K}^{\alpha}(\mathcal{O})$-can be valid. More precisely, if we choose posterior quantiles $R_{\varepsilon}=R_{\varepsilon}(Y, \beta)$ such that

$$
\begin{equation*}
C_{\varepsilon}=\left\{f \in \operatorname{supp} \Pi(\cdot \mid Y):\|f-\bar{f}\|_{\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}} \leq R_{\varepsilon}\right\}, \quad \Pi\left(C_{\varepsilon} \mid Y\right)=1-\beta \tag{34}
\end{equation*}
$$

where $\beta>0$ is some fixed significance level, and $\alpha>2+3 d / 2$ is arbitrary, then one can prove the following result.

Corollary 11. The frequentist coverage probability of $C_{\varepsilon}$ from (34) satisfies

$$
P_{f_{0}}^{Y}\left(f_{0} \in C_{\varepsilon}\right) \rightarrow 1-\beta
$$

and as $\varepsilon \rightarrow 0$ the quantile constants satisfy

$$
\varepsilon^{-1} R_{\varepsilon} \xrightarrow{P_{f_{0}}^{Y}} \text { const. }
$$

Although the credible ball (34) is constructed in the weaker topology of $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$, its diameter can be shown to converge to zero in probability at rate $\varepsilon^{2(s-d-\kappa) /(2 s+4+d)}$, for every $\kappa>0$, also in the much stronger $\|f\|_{L^{1}(\bar{K})}=\int_{\bar{K}}|f|$ norm (where the Bernsteinvon Mises theorem does not hold), for any compact strict subset $\bar{K} \subsetneq K$. This can be proved by noting that the $L^{1}(\bar{K})$-norm can be bounded via interpolation between the bounded support of $\Pi(\cdot \mid Y)$ in $C^{s}(\mathcal{O})$ and the convergence rate in $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$. See [9, 10] and [46, Section 5] for more discussion of this 'multi-scale' phenomenon.

## 6. Proofs

### 6.1. Proofs of Propositions 6 and 2

In order to prove Proposition $6(\mathrm{~A})$, let $X \sim \mathcal{N}_{f_{0}}$ with covariance metric $d\left(\psi, \psi^{\prime}\right)=$ $\left\|S_{f_{0}}\left[\left(\psi-\psi^{\prime}\right) / u_{f_{0}}\right]\right\|_{L^{2}} \lesssim\left\|\psi-\psi^{\prime}\right\|_{C^{2}}$. The metric entropy estimate

$$
\begin{equation*}
\log N\left(\{ \psi : \| \psi \| _ { C _ { K } ^ { \alpha } ( \mathcal { O } ) } \leq 1 , \eta , d ) \lesssim \operatorname { l o g } N \left(\left\{\psi:\|\psi\|_{C_{K}^{\alpha-2}(\mathcal{O})} \leq 1, c \eta,\|\cdot\| \infty\right), \quad \eta>0,\right.\right. \tag{35}
\end{equation*}
$$

for some constant $c>0$, combined with (107), $\alpha>2+d / 2$, and [27, Proposition 2.1.5 and Theorem 2.3.7] imply that $X$ defines a tight Gaussian Borel random variable in the space of bounded and uniformly $d$-continuous functions on the unit ball of $C_{K}^{\alpha}(\mathcal{O})$. That a version of $X$ exists that acts linearly on $C_{K}^{\alpha}(\mathcal{O})$, thus defining an element of $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$, follows from [27, proof of Theorem 3.7.28] and linearity in $\psi$ of the covariance in (29).

We now prove Proposition $6(\mathrm{~B})$ for $\left(C_{c}^{\alpha}(\mathcal{O})\right)^{*}$; the case of $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$ is proved in the same way by replacing $\mathcal{O}$ with a ball contained in $K$. By the continuous imbeddings $C_{c}^{\alpha^{\prime}} \subseteq C_{c}^{\alpha}, \alpha^{\prime}>\alpha$, it suffices to prove the result for $\alpha=2+d / 2-\epsilon$ for every $\epsilon>0$ small enough. Suppose to the contrary that

$$
\mathcal{N}_{f_{0}}\left(x:\|x\|_{\left(C_{c}^{\alpha}(\mathcal{O})\right)^{*}}<\infty\right)=\operatorname{Pr}\left(\sup _{h \in C_{c}^{\alpha}(\mathcal{O}):\|h\|_{C^{\alpha} \leq 1} \leq}|X(h)|<\infty\right)>0
$$

then by separability of the unit ball of $C_{c}^{\alpha}(\mathcal{O}), \alpha>2$, for the covariance metric $d$ of $\mathcal{N}_{f_{0}}$ (cf. (35) and (107)), and by [27, Proposition 2.1.12 and Theorem 2.1.20], $X \sim \mathcal{N}_{f_{0}}$ must satisfy $E\|X\|_{\left(C_{c}^{\alpha}(\mathcal{O})\right)^{*}}<\infty$, which will lead to a contradiction. If $V_{f_{0}}$ denotes the inverse Schrödinger operator from Section 7.2, and if $X \sim \mathcal{N}_{f_{0}}$, then by Proposition 22 the Gaussian process $\widetilde{X}$ defined by the action

$$
\left(\tilde{X}(\varphi)=X\left(u_{f_{0}} V_{f_{0}}[\varphi]\right), \varphi \in C_{c}^{\infty}(\mathcal{O})\right)
$$

has the law of a standard Gaussian white noise $\mathbb{W}$ on $L^{2}(\mathcal{O})$. By Proposition 22 again,

$$
\begin{aligned}
& E \sup _{h \in C_{c}^{\alpha}(\mathcal{O}):\|h\|_{C^{\alpha}} \leq 1}|X(h)|=E \sup _{h \in C_{c}^{\alpha}(\mathcal{O}):\|h\|_{C^{\alpha}} \leq 1}\left|X\left(u_{f_{0}} V_{f_{0}}\left[S_{f_{0}}\left(h / u_{f_{0}}\right)\right]\right)\right| \\
& \quad=E \sup _{h \in C_{c}^{\alpha}(\mathcal{O}):\|h\|_{C^{\alpha}} \leq 1}\left|\widetilde{X}\left(S_{f_{0}}\left(h / u_{f_{0}}\right)\right)\right|=E \sup _{h \in C_{c}^{\alpha}(\mathcal{O}):\|h\|_{C^{\alpha} \leq 1}}\left|\mathbb{W}\left(S_{f_{0}}\left(h / u_{f_{0}}\right)\right)\right| \\
& \quad \geq E \sup _{h \in C_{c}^{\alpha}(\mathcal{O}):\|h\|_{C^{\alpha}} \leq 1}\left|\left\langle\mathbb{W}, \Delta\left(h / 2 u_{f_{0}}\right)\right\rangle_{L^{2}}\right|-E \sup _{h \in C_{c}^{\alpha}(\mathcal{O}):\|h\|_{C^{\alpha}} \leq 1}\left|\left\langle\mathbb{W}, f_{0} h / u_{f_{0}}\right)\right\rangle_{L^{2}} \mid \\
& \quad \geq E \sup _{h \in C_{c}^{\alpha}(\mathcal{O}):\|h\|_{C^{\alpha}} \leq 1}\left|\left\langle\mathbb{W}, \Delta\left(h / 2 u_{f_{0}}\right)\right\rangle_{L^{2}}\right|-C \\
& \quad \geq(1 / \kappa) E \sup _{\bar{h} \in C_{c}^{\alpha}(\mathcal{O}):\|\bar{h}\|_{C^{\alpha}} \leq 1}\left|\langle\mathbb{W}, \Delta \bar{h}\rangle_{L^{2}}\right|-C
\end{aligned}
$$

for some $\kappa>0$, where we have used $f_{0}, u_{f_{0}}, 1 / u_{f_{0}} \in C^{\alpha}(\mathcal{O})$ under the maintained assumptions (Section 7.2), that the supremum of the standard white noise process $\mathbb{W}$ on balls in $C_{c}^{\alpha}(\mathcal{O})(\alpha=2+d / 2-\epsilon, \epsilon<2)$ is bounded in expectation by a fixed constant (use Proposition 27(C) and (91)), and where we have taken $h=2 \bar{h} u_{f_{0}} /\left\|\bar{h} u_{f_{0}}\right\|_{C^{\alpha}}$ and used $2\left\|\bar{h} u_{f_{0}}\right\|_{C^{\alpha}} \leq \kappa$ by (11).

We complete the proof by showing

$$
\begin{equation*}
E \sup _{h \in C_{c}^{\alpha}(\mathcal{O}):\|h\|_{C^{\alpha}} \leq 1}\left|\langle\mathbb{W}, \Delta h\rangle_{L^{2}(\mathcal{O})}\right|=\infty \tag{36}
\end{equation*}
$$

as follows: For every $j \in \mathbb{N}$ there exists a small positive constant $c_{0}^{\prime}$ and $c_{0}^{\prime} 2^{j d}=n_{j}$ Daubechies wavelets ( $\Phi_{j, r}: r=1, \ldots, n_{j}$ ) that have disjoint compact supports within $\mathcal{O}$ (Section 3.2). Let $b_{m, \text {. }}$ be a point in the discrete hypercube $\{-1,1\}^{n_{j}}$ and, for $\kappa^{\prime}$ small enough chosen below, define functions

$$
\begin{equation*}
h_{m}(x) \equiv h_{m, j}(x)=\kappa^{\prime} \sum_{r=1}^{n_{j}} b_{m, r} 2^{-j(\alpha+d / 2)} \Phi_{j, r}(x), \quad x \in \mathcal{O}, m=1, \ldots, 2^{n_{j}} \tag{37}
\end{equation*}
$$

The interior wavelets $\Phi_{j, r}$ are all orthogonal to the boundary wavelets, thus

$$
\left\|h_{m}\right\|_{\mathcal{C}^{\alpha, W}}=\sup _{l, r} 2^{l(\alpha+d / 2)}\left|\left\langle\Phi_{l, r}^{\mathcal{O}}, h_{m}\right\rangle_{L^{2}(\mathcal{O})}\right|=\kappa^{\prime},
$$

and since by the remarks after (13) the wavelet norm $\|\cdot\|_{\mathcal{C}^{\alpha, W}}$ is equivalent to $\|\cdot\|_{C^{\alpha}}$ on such $h_{m}$ 's (we can assume $\alpha \notin \mathbb{N}$ by the choice of $\epsilon$ ), the $h_{m}$ 's are all contained in $\left\{h \in C_{c}^{\alpha}(\mathcal{O}):\|h\|_{C^{\alpha}} \leq 1\right\}$ for $\kappa^{\prime}$ small enough. We conclude that the supremum in (36) is lower bounded by

$$
E \max _{m}\left|\left\langle\mathbb{W}, \Delta h_{m}\right\rangle_{L^{2}(\mathcal{O})}\right| .
$$

The Gaussian process $\left(\mathbb{W}\left(h_{m}\right): m=1, \ldots, 2^{n_{j}}\right)$ has covariance metric

$$
d^{2}\left(h_{m}, h_{m}^{\prime}\right)=\left\|\Delta\left(h_{m}-h_{m}^{\prime}\right)\right\|_{L^{2}(\mathcal{O})}^{2},
$$

and by the Varshamov-Gilbert bound [27, Example 3.1.4] there exist $\left\{b_{m,}: m=\right.$ $\left.1, \ldots, M_{j}\right\} \subseteq\{-1,1\}^{n_{j}}$ with $M_{j} \geq 3^{n_{j} / 4}$ that are $n_{j} / 8$-separated for the Hamming distance. Then for the $h_{m}$ 's corresponding to these separated $b_{m}$ 's we have

$$
\begin{equation*}
d^{2}\left(h_{m}, h_{m^{\prime}}\right)=\left(\kappa^{\prime}\right)^{2} 2^{-2 j(\alpha+d / 2)}\left\|\sum_{r}\left(b_{m, r}-b_{m^{\prime}, r}\right) \Delta \Phi_{j, r}\right\|_{L^{2}(\mathcal{O})}^{2} . \tag{38}
\end{equation*}
$$

The $\Delta \Phi_{j, r}$ all have disjoint supports and thus, when normalised by $\left\|\Delta \Phi_{j, r}\right\|_{L^{2}}=$ $2^{2 j}\left\|\Delta \Phi_{0,0}\right\|_{L^{2}}$, form an orthonormal system in $L^{2}(\mathcal{O})$. Thus the last term equals, by Parseval's identity,
$d^{2}\left(h_{m}, h_{m^{\prime}}\right)=\left(\kappa^{\prime}\right)^{2} 2^{2 j(2-\alpha-d / 2)} \sum_{r}\left(b_{m, r}-b_{m^{\prime}, r}\right)^{2} \geq c 2^{2 j(2-\alpha-d / 2)} n_{j} \geq\left(c^{\prime}\right)^{2} 2^{2 j(2-\alpha)}$,
for $m \neq m^{\prime}$ in the separated set. By Sudakov's lower bound ([27, Theorem 2.4.12] with $\varepsilon$ there chosen as $\left.\varepsilon=c^{\prime} 2^{j(2-\alpha)}\right)$ we then have

$$
E \max _{m}\left|\left\langle\mathbb{W}, \Delta h_{m}\right\rangle\right| \geq c^{\prime \prime} 2^{j(2-\alpha)} \sqrt{\log M_{j}} \geq c^{\prime \prime \prime} 2^{j(2+d / 2-\alpha)},
$$

which for $\alpha<2+d / 2$ can be made as large as desired as $j \rightarrow \infty$, completing the proof.
Remark 12. A slightly more involved version of the above argument also shows that $\mathcal{N}_{f_{0}}$ cannot be tight in $\left(C_{c}^{\alpha}(\mathcal{O})\right)^{*}$ in the boundary case $\alpha=2+d / 2$, as this would force the Gaussian process $X$ indexed by the unit ball in $C_{c}^{\alpha}(\mathcal{O})$ to be sample continuous for its covariance metric (by arguing as in [27, Proposition 2.1.7]). Sudakov's lower bound argument can then be refined (using [27, Corollary 2.4.14]) to also cover $\alpha=2+d / 2$.

We finally turn to the proof of Proposition 2: The upper bound follows from Lemma 14 with $\eta \simeq \varepsilon^{(2 s+4) /(2 s+4+d)}$ when $\alpha=0$-since one may take $v \gtrsim \bar{D}^{2}$ in that lemma, we can integrate tail probabilities to also bound the expected risk. The case of general $\alpha$ then follows from (49) below. For the lower bound we apply the general Theorem 6.3.2 of [27] and arguments akin to those used in the proof of Proposition 6(B) just given. We assume that $g=1$ on $\partial \mathcal{O}$ and that $B$ is large and $f_{\min }$ small enough (chosen below); the general result requires only minor modifications. Let $\bar{g}$ be a smooth function such that $\bar{g} \geq 1$ on $\overline{\mathcal{O}}$ and let $g_{0}$ be the smooth solution to the Poisson equation $\Delta g_{0}=\bar{g}$ on $\mathcal{O}$ with Dirichlet boundary conditions $g_{0}=1$ on $\partial \mathcal{O}$ (a unique such solution exists by [25, Theorem 6.14]). Since $\bar{g} \geq 1$ on $\mathcal{O}$ and $g \geq 1$ on $\partial \mathcal{O}$, the maximum principle for harmonic functions [25, Theorem 2.3] implies that also $g_{0} \geq 1$ on $\overline{\mathcal{O}}$. Now define

$$
u_{0}=g_{0}, \quad u_{m}=g_{0}+h_{m}, \quad m=1, \ldots, M_{j},
$$

where $h_{m}$ is as in (37) but with $\alpha$ there replaced by $s+2$, and $M_{j}$ as before (38). Arguing as after (37), we see that the $u_{m}$ are contained in a ball of $C^{s+2}(\mathcal{O})$ of radius $\left\|g_{0}\right\|_{C^{s+2}}+c \kappa^{\prime}$, $c>0$, and for $\kappa^{\prime}$ small enough both $u_{m}$ and $\Delta u_{m}=\bar{g}+\Delta h_{m}$ are positive and bounded away from zero on $\overline{\mathcal{O}}$. Setting $2^{j} \simeq \varepsilon^{-2 /(2 s+4+d)}$ and since the KL-divergence in the
statistical model (5) equals $\varepsilon^{-2}$ times the squared $L^{2}$-distance [27, (6.16)]), given $\epsilon>0$ we can choose $\kappa^{\prime}$ small enough so that

$$
\varepsilon^{-2}\left\|u_{0}-u_{m}\right\|_{L^{2}}^{2} \leq\left(\kappa^{\prime}\right)^{2} \varepsilon^{-2} \varepsilon^{2(2 s+4+d) /(2 s+4+d)} \sum_{r \leq n_{j}} 1 \leq \epsilon \log M_{j},
$$

verifying [27, (6.103)]. Observe next that the $f_{m}=\frac{1}{2} \frac{\Delta u_{m}}{u_{m}}$ are all greater than some $f_{\text {min }}>0$, lie in a fixed ball of $C^{s}(\mathcal{O})$ of radius $B$, and solve the Schrödinger equation $\Delta u_{m} / 2-f_{m} u_{m}=0$ subject to $u_{m}=g$ on $\partial \mathcal{O}$. Now the result follows since, adapting the argument after (38) to the present choice of $s$, we see that the $f_{m}$ are $H^{\alpha}(\mathcal{O})$-separated by at least
$\left\|f_{m}-f_{m}^{\prime}\right\|_{H^{\alpha}} \geq c\left\|\Delta\left(u_{m}-u_{m}^{\prime}\right)\right\|_{H^{\alpha}}-C\left\|u_{m}^{-1}-u_{m}^{-1}\right\|_{H^{\alpha}} \geq c 2^{-j(s-\alpha)} \simeq \varepsilon^{2(s-\alpha) /(2 s+4+d)}$, verifying the hypotheses of [27, Theorem 6.3.2].

### 6.2. Proof of Theorem 1

We start with a general contraction theorem that applies to priors that are supported on a fixed $s$-Hölder ball. Recall (21) for the definition of the $\left(H_{0}^{2}\right)^{*}$-norm. We apply the general approach of [23] to Bayesian contraction theorems (via Theorem 28), but, following [26], we use frequentist estimators to construct the relevant 'test functions'.

Theorem 13. Let $\Pi=\Pi_{\varepsilon}$ be a Borel probability measure supported on a measurable subset $\mathcal{F}$ of $L^{2}(\mathcal{O})$ that satisfies

$$
\mathcal{F} \subseteq\left\{\inf _{x \in \mathcal{O}} f(x) \geq f_{\min },\|f\|_{C^{s}(\mathcal{O})} \leq D\right\}, \quad D>0, f_{\min }>0, s>2+d / 2, s \in \mathbb{N}
$$

Let $\Pi(\cdot \mid Y)$ be the resulting posterior distribution (6) arising from observing (5) with $g \in C^{s+2}(\overline{\mathcal{O}}), g \geq g_{\min }>0$. Let $P_{f_{0}}^{Y}$ be the law generating $Y=u_{f_{0}}+\varepsilon \mathbb{W}$ for fixed $f_{0} \in \mathcal{F}$. Let $\eta=\eta_{\varepsilon}$ be a sequence satisfying

$$
\eta_{\varepsilon} \geq \bar{c} \varepsilon^{(2 s+4) /(2 s+4+d)}
$$

and

$$
\begin{equation*}
\Pi\left(f:\left\|f-f_{0}\right\|_{\left(H_{0}^{2}\right)^{*}}<\eta_{\varepsilon}\right) \geq e^{-C\left(\eta_{\varepsilon} / \varepsilon\right)^{2}} \tag{39}
\end{equation*}
$$

for some constants $\bar{c}, C>0$ and all $\varepsilon$ small enough. Then there exists a finite constant $M$ depending only on $\bar{c}, C, D, d, s, \mathcal{O},\|g\|_{C^{s+2}(\mathcal{O})}$ such that as $\varepsilon \rightarrow 0$,

$$
\Pi\left(f:\left\|f-f_{0}\right\|_{L^{2}}>M \eta_{\varepsilon}^{s /(s+2)} \mid Y\right) \xrightarrow{P_{f_{0}}^{Y}} 0
$$

and

$$
\Pi\left(f:\left\|f-f_{0}\right\|_{\left(H_{0}^{2}\right)^{*}}>M \eta_{\varepsilon} \mid Y\right) \xrightarrow{P_{f_{0}}^{Y}} 0 .
$$

Proof. Proposition 25 implies that for each $f \in \mathcal{F}$ a unique solution $u_{f}$ to (1) exists, and that $\left\{u_{f}: f \in \mathcal{F}\right\}$ is bounded in $H^{s+2}(\mathcal{O})$. We apply Theorem 28 below with $\mathcal{F}=\mathcal{F}_{\varepsilon}$ and its trace Borel $\sigma$-field $\mathcal{B}_{\mathcal{F}}$ of $L^{2}(\mathcal{O}), G(f)=u_{f}, \mathbb{H}=L^{2}(\mathcal{O})$ and $\eta_{\varepsilon}=\bar{\eta}_{\varepsilon} / c$ where
$c$ is the constant from Proposition 25(B), so that (39) gives
$\Pi\left(f \in \mathcal{F}:\left\|G(f)-G\left(f_{0}\right)\right\|_{\mathbb{H}}<\bar{\eta}_{\varepsilon}\right) \geq \Pi\left(f \in \mathcal{F}:\left\|f-f_{0}\right\|_{\left(H_{0}^{2}\right)^{*}}<\eta_{\varepsilon}\right) \geq e^{-C\left(\eta_{\varepsilon} / \varepsilon\right)^{2}}$,
which proves (112) in Theorem 28 with $C^{\prime}=C c^{2}$. Thus if we can construct tests $\Psi(Y)$ such that

$$
\begin{gather*}
E_{f_{0}}^{Y} \Psi(Y)+\sup _{f \in \mathcal{F}:\left\|f-f_{0}\right\|_{L^{2} \geq M} \eta_{\varepsilon}^{s /(s+2)}} E_{f}^{Y}(1-\Psi(Y)) \leq L e^{-\left(C^{\prime}+4\right)\left(\eta_{\varepsilon} / \varepsilon\right)^{2}},  \tag{40}\\
E_{f_{0}}^{Y} \Psi(Y)+\sup _{f \in \mathcal{F}:\left\|f-f_{0}\right\|_{\left(H_{0}^{2}\right)^{*} \geq} \geq M \eta_{\varepsilon}} E_{f}^{Y}(1-\Psi(Y)) \leq L e^{-\left(C^{\prime}+4\right)\left(\eta_{\varepsilon} / \varepsilon\right)^{2}} \tag{41}
\end{gather*}
$$

for $M$ large enough, respectively, then the result will follow from appropriate choices of $\eta_{\varepsilon}^{*}, d(\cdot, \cdot)$ in Theorem 28. We achieve this by first constructing an estimator for $u_{f}$ from which we obtain a plug-in test for $f$, using also (3) and the resulting identification equation (4). The estimate of $u_{f}$ is obtained as the non-parametric least squares (or maximum likelihood) estimator $\widehat{u}$ obtained from maximising the log-likelihood function (cf. (110))

$$
\begin{equation*}
\bar{\ell}(u)=\frac{1}{\varepsilon^{2}}\langle Y, u\rangle_{L^{2}(\mathcal{O})}-\frac{1}{2 \varepsilon^{2}}\|u\|_{L^{2}(\mathcal{O})}^{2} \text { over } \mathcal{U}_{g}=\left\{\|u\|_{H^{s+2}(\mathcal{O})} \leq D^{\prime}, u_{\partial \mathcal{O}}=g \text { on } \partial \mathcal{O}\right\} . \tag{42}
\end{equation*}
$$

Here $D^{\prime}$ is chosen large enough so that $\left\{u_{f}: f \in \mathcal{F}\right\} \subseteq \mathcal{U}_{g}$. By the Sobolev imbedding $H^{s+2}(\mathcal{O}) \subseteq C^{1}(\overline{\mathcal{O}})$, so by [19, Theorem 2.4.7] the set $\mathcal{U}_{g}$ is totally bounded in $C(\overline{\mathcal{O}})$. Since $\left\{u:\|u\|_{H^{s+2}(\mathcal{O})} \leq D^{\prime}\right\}$ is easily seen to be closed for the $\|\cdot\|_{\infty}$-topology and since limits $u$ of uniformly convergent sequences in $\mathcal{U}_{g}$ necessarily satisfy $u=g$ on $\partial \mathcal{O}$, we deduce that $\mathcal{U}_{g}$ is also closed and hence a compact subset of $C(\overline{\mathcal{O}})$. Next, the $L^{2}$-metric entropy of the bounded subset $\mathcal{U}_{g}$ of $H^{s+2}(\mathcal{O})$ can be shown [21, Chapter 3] to be at most

$$
\begin{equation*}
\log N\left(\mathcal{U}_{g}, \gamma,\|\cdot\|_{L^{2}(\mathcal{O})}\right) \lesssim(A / \gamma)^{d /(s+2)}, \quad \forall \gamma>0, A=A\left(D^{\prime}\right) \tag{43}
\end{equation*}
$$

Since $s+2>d / 2$ the square root of the $L^{2}$-metric entropy is $\gamma$-integrable at zero and by [27, Proposition 2.1.5 and Theorem 2.3.7] the real-valued maps $u \mapsto\langle u, \mathbb{W}\rangle_{L^{2}}$ and then also $\bar{\ell}(u)$ define Borel random variables in the space of $L^{2}$-continuous functions on $\mathcal{U}_{g}$. Thus, in view of [27, Exercise 7.2.3] a (measurable) maximiser $\widehat{u} \in \mathcal{U}_{g}$ of $\bar{\ell}(u)$ over $\mathcal{U}_{g}$ exists almost surely. We now derive its rate of convergence to $u_{f}$, initially in $L^{2}(\mathcal{O})$ distance, using 'peeling' techniques commonly used in $M$-estimation [54, 55]. For all $f \in \mathcal{F}$ and $D_{0}$ large enough, we have

$$
\begin{align*}
& P_{f}^{Y}\left(\left\|\widehat{u}-u_{f}\right\|_{L^{2}} \geq D_{0} \eta_{\varepsilon}\right)=P_{f}^{Y}\left(\bar{\ell}(\widehat{u})-\bar{\ell}\left(u_{f}\right) \geq 0,\left\|\widehat{u}-u_{f}\right\|_{L^{2}} \geq D_{0} \eta_{\varepsilon}\right) \\
& \quad=P_{f}^{Y}\left(-\frac{1}{2 \varepsilon^{2}}\left\|\widehat{u}-u_{f}\right\|_{L^{2}}^{2}+\frac{1}{\varepsilon}\left\langle\widehat{u}-u_{f}, \mathbb{W}\right\rangle_{L^{2}} \geq 0,\left\|\widehat{u}-u_{f}\right\|_{L^{2}} \geq D_{0} \eta_{\varepsilon}\right) \\
& \quad=P_{f}^{Y}\left(\frac{\left\langle\widehat{u}-u_{f}, \mathbb{W}\right\rangle_{L^{2}}}{\left\|\widehat{u}-u_{f}\right\|_{L^{2}}^{2}} \geq \frac{1}{2 \varepsilon},\left\|\widehat{u}-u_{f}\right\|_{L^{2}} \geq D_{0} \eta_{\varepsilon}\right) \\
& \quad \leq \sum_{r=0}^{R} \operatorname{Pr}\left(\sup _{u \in \mathcal{U}_{g}: 2^{r} D_{0} \eta_{\varepsilon} \leq\left\|u-u_{f}\right\|_{L^{2}} \leq 2^{r+1} D_{0} \eta_{\varepsilon}}\left|\left\langle u-u_{f}, \mathbb{W}\right\rangle\right| \geq \frac{2^{2(r+1)} D_{0}^{2} \eta_{\varepsilon}^{2}}{8 \varepsilon}\right), \tag{44}
\end{align*}
$$

where $R=R\left(D^{\prime}\right)<\infty$. Now using the metric entropy bound (92) for Gaussian processes and (43), for all $\sigma>0$ we have

$$
E \sup _{u \in \mathcal{U}_{g}:\left\|u-u_{f}\right\|_{L^{2}} \leq \sigma}\left|\left\langle u-u_{f}, \mathbb{W}\right\rangle\right| \lesssim \int_{0}^{\sigma}(A / \gamma)^{d /(2 s+4)} d \gamma \leq C^{\prime} \sigma^{1-d /(2 s+4)}
$$

For choices $\sigma=\sigma_{r} \equiv 2^{r+1} D_{0} \eta_{\varepsilon}, r \geq 0$, using $\eta_{\varepsilon} \geq \bar{c} \varepsilon^{(2 s+4) /(2 s+4+d)}$ and for $D_{0}$ large enough, we have $C^{\prime} \sigma^{1-d /(2 s+4)} \leq \sigma^{2} /(16 \varepsilon)$ and thus, by using the Borell-SudakovTsirelson inequality [27, Theorem 2.5.8], the sum in (44) can be bounded by

$$
\begin{aligned}
& \sum_{r=0}^{R} \operatorname{Pr}\left(\sup _{u \in \mathcal{U}_{g}:\left\|u-u_{f}\right\|_{L^{2}} \leq \sigma_{r}}\left|\left\langle u-u_{f}, \mathbb{W}\right\rangle\right|-E \sup _{u \in \mathcal{U}_{g}:\left\|u-u_{f}\right\|_{L^{2}} \leq \sigma_{r}}\left|\left\langle u-u_{f}, \mathbb{W}\right\rangle\right| \geq \frac{\sigma_{r}^{2}}{16 \varepsilon}\right) \\
& \leq 2 \sum_{r=0}^{R} \exp \left\{-\frac{c D_{0}^{2} 2^{2 r} \eta_{\varepsilon}^{2}}{\varepsilon^{2}}\right\} \leq L \exp \left\{-\frac{c^{\prime} D_{0}^{2} \eta_{\varepsilon}^{2}}{\varepsilon^{2}}\right\}
\end{aligned}
$$

where $c^{\prime}=c^{\prime}(R)$ and $L>0$ are fixed constants. Since both $\widehat{u}, u_{f}$ are contained in a ball of $H^{s+2}(\mathcal{O})$ of radius $D^{\prime}$, what precedes and the interpolation inequality (49) imply

$$
\begin{equation*}
P_{f}^{Y}\left(\left\|\widehat{u}-u_{f}\right\|_{H^{\alpha}} \geq c^{\prime \prime} D_{0} \eta_{\varepsilon}^{(s+2-\alpha) /(s+2)}\right) \leq L \exp \left\{-c^{\prime} D_{0}^{2} \eta_{\varepsilon}^{2} / \varepsilon^{2}\right\}, \quad 0 \leq \alpha<s+2 \tag{45}
\end{equation*}
$$

for a fixed constant $c^{\prime \prime}>0$, which by the Sobolev imbedding for $d / 2+2<\alpha<s+2$ also implies a convergence rate in $\|\cdot\|_{C^{2}(\mathcal{O})}$-norm:

$$
\begin{equation*}
P_{f}^{Y}\left(\left\|\widehat{u}-u_{f}\right\|_{C^{2}} \geq c^{\prime \prime \prime} D_{0} \eta_{\varepsilon}^{\zeta}\right) \leq L \exp \left\{-c^{\prime} D_{0}^{2} \eta_{\varepsilon}^{2} / \varepsilon^{2}\right\}, \quad \zeta, c^{\prime \prime \prime}>0 \tag{46}
\end{equation*}
$$

Now recall (4) and define

$$
\begin{equation*}
\widehat{f}=\frac{\Delta \widehat{u}}{2 \widehat{u}} 1_{A_{\varepsilon}}, \quad A_{\varepsilon}=\left\{\inf _{x \in \mathcal{O}} \widehat{u}(x) \geq c_{0},\|\Delta \widehat{u}\|_{\infty} \leq D_{2}\right\}, \quad c_{0}>0, D_{2}<\infty, \tag{47}
\end{equation*}
$$

which defines a random variable in $C(\overline{\mathcal{O}})$.
Lemma 14. For every $v>0$ there exist $c_{0}>0$ small enough and $D_{2}, D_{3}$ large enough such that for all $\bar{D} \geq D_{3}$ and $\eta_{\varepsilon}$ as in the theorem we have, for all $f \in \mathcal{F}$,

$$
P_{f}^{Y}\left(\left\{\|\widehat{f}-f\|_{L^{2}}>\bar{D} \eta_{\varepsilon}^{s /(s+2)}\right\} \cup\left\{\|\widehat{f}-f\|_{\left(H_{0}^{2}\right)^{*}}>\bar{D} \eta_{\varepsilon}\right\}\right) \leq e^{-v\left(\eta_{\varepsilon} / \varepsilon\right)^{2}}
$$

Proof. If $B$ is the event whose probability we want to bound, we can write

$$
P_{f}^{Y}(B)=P_{f}^{Y}\left(A_{\varepsilon} \cap B\right)+P_{f}^{Y}\left(A_{\varepsilon}^{c} \cap B\right)
$$

and the second probability is no greater than $P_{f}^{Y}\left(A_{\varepsilon}^{c}\right)$, which is less than $e^{-v(\eta / \varepsilon)^{2}} / 3$ for $c_{0}$ small and $D_{2}$ large enough, since $u_{f} \in C^{2}(\mathcal{O})$ is bounded away from zero (see (3)), and
using (46) for sufficiently large $D_{0}$. It remains to bound $P_{f}^{Y}\left(A_{\varepsilon} \cap B\right)$ by $2 e^{-v(\eta / \varepsilon)^{2}} / 3$. For $\widehat{f}$ as in (47) we see that (3), (4) imply on $A_{\varepsilon}$ that

$$
\begin{aligned}
\|\widehat{f}-f\|_{L^{2}} & \lesssim\left\|\widehat{u}-u_{f}\right\|_{H^{2}}, \\
\|\widehat{f}-f\|_{\left(H_{0}^{2}\right)^{*}} & \lesssim\left\|\left[\widehat{u}^{-1}-u_{f}^{-1}\right] \Delta \widehat{u}_{f}\right\|_{L^{2}}+\left\|\Delta\left(\widehat{u}-u_{f}\right) u_{f}^{-1}\right\|_{\left(H_{0}^{2}\right)^{*}} \\
& \lesssim\left\|\widehat{u}-u_{f}\right\|_{L^{2}}+\left\|\Delta\left(\widehat{u}-u_{f}\right)\right\|_{\left(H_{0}^{2}\right)^{*}} .
\end{aligned}
$$

In view of density of $C^{2}(\mathcal{O})$ in $H^{2}(\mathcal{O})$, the last term can be further bounded by

$$
\begin{aligned}
\left\|\Delta\left(\widehat{u}-u_{f}\right)\right\|_{\left(H_{0}^{2}\right)^{*}} & =\sup _{h \in C_{0} \cap C^{2}:\|h\|_{H^{2}} \leq 1}\left|\int_{\mathcal{O}} h \Delta\left(\widehat{u}-u_{f}\right)\right| \\
& =\sup _{h \in C_{0} \cap C^{2}:\|h\|_{H^{2}} \leq 1}\left|\int_{\mathcal{O}} \Delta h\left(\widehat{u}-u_{f}\right)\right| \\
& \leq \sup _{\|h\|_{H^{2}} \leq 1}\|\Delta h\|_{L^{2}}\left\|\widehat{u}-u_{f}\right\|_{L^{2}} \lesssim\left\|\widehat{u}-u_{f}\right\|_{L^{2}}
\end{aligned}
$$

where we use Green's identity [25, p. 17] twice, the fact that both $h$ and $\widehat{u}-u_{f}$ have vanishing boundary traces 0 and $g-g=0$, respectively, and the Cauchy-Schwarz inequality.

Summarising we conclude from (45) that for every $v$ we can find $\bar{D}$ and $D_{0}$ large enough so that

$$
\begin{aligned}
& P_{f}^{Y}\left(\|\widehat{f}-f\|_{L^{2}}>\bar{D} \eta_{\varepsilon}^{s /(s+2)}\right) \leq P_{f}^{Y}\left(\left\|\widehat{u}_{f}-u_{f}\right\|_{H^{2}}>d \bar{D} \eta_{\varepsilon}^{s /(s+2)}\right) \leq e^{-v\left(\eta_{\varepsilon} / \varepsilon\right)^{2}} / 3, \\
& P_{f}^{Y}\left(\|\widehat{f}-f\|_{\left(H_{0}^{2}\right)^{*}}>\bar{D} \eta_{\varepsilon}\right) \leq P_{f}^{Y}\left(\left\|\widehat{u}_{f}-u_{f}\right\|_{L^{2}}>d^{\prime} \bar{D} \eta_{\varepsilon}\right) \leq e^{-v\left(\eta_{\varepsilon} / \varepsilon\right)^{2}} / 3,
\end{aligned}
$$

which completes the proof of the lemma by the union bound.
The usual plug-in test [27, Proposition 6.2.2]) defined by

$$
\Psi(Y)=1\left\{\left\|\widehat{f}-f_{0}\right\|_{L^{2}} \geq \bar{D} \eta_{\varepsilon}^{s /(s+2)}\right\}
$$

for $\bar{D}$ a large enough constant, and likewise with $\left(H_{0}^{2}\right)^{*}, \eta_{\varepsilon}$ replacing $L^{2}(\mathcal{O}), \eta_{\varepsilon}^{s /(s+2)}$, gives

$$
E_{f_{0}}^{Y} \Psi(Y) \leq P_{f_{0}}^{Y}\left(\left\|\widehat{f}-f_{0}\right\|_{L^{2}}>\bar{D} \eta_{\varepsilon}^{s /(s+2)}\right) \leq e^{-\left(C^{\prime}+4\right)\left(\eta_{\varepsilon} / \varepsilon\right)^{2}}
$$

and also, for $M$ large enough and $f$ in the alternative (40),

$$
\begin{aligned}
E_{f}^{Y}(1 & -\Psi(Y)) \\
& =P_{f}^{Y}\left(\left\|\widehat{f}-f_{0}\right\|_{L^{2}} \leq \bar{D} \eta_{\varepsilon}^{s /(s+2)}\right) \leq P_{f}^{Y}\left(\left\|f-f_{0}\right\|_{L^{2}}-\bar{D} \eta_{\varepsilon}^{s /(s+2)} \leq\|\widehat{f}-f\|_{L^{2}}\right) \\
& \leq P_{f}^{Y}\left(\|\widehat{f}-f\|_{L^{2}} \geq(M-\bar{D}) \eta_{\varepsilon}^{s /(s+2)}\right) \leq e^{-\left(C^{\prime}+4\right)\left(\eta_{\varepsilon} / \varepsilon\right)^{2}},
\end{aligned}
$$

and likewise for the $\left(H_{0}^{2}\right)^{*}$-errors, completing the proof of the theorem.
We finally turn to the verification of (39) for priors featuring in Theorem 1.

Proposition 15. Under the conditions of Theorem 1 and given $C>0$, we can choose $\bar{L}$ large enough such that the prior (17) with $J$ such that $2^{J} \simeq \varepsilon^{-2 /(2 s+4+d)}$ satisfies (39) for

$$
\eta_{\varepsilon}=\bar{L} \sqrt{\log (1 / \varepsilon)} \cdot \varepsilon^{(2 s+4) /(2 s+4+d)}
$$

As a consequence,

$$
\eta_{\varepsilon}^{s /(s+2)} \simeq \varepsilon^{\frac{2 s}{2 s+4+d}} \log ^{\gamma}(1 / \varepsilon), \quad \gamma=s /(2 s+4)
$$

is the posterior contraction rate about $f_{0}$ in $L^{2}$, and in $\left(H_{0}^{2}\right)^{*}$ we obtain the contraction rate $\eta_{\varepsilon}$, both up to multiplicative constants.

Proof. The prior (17) defines a Borel probability measure on $V_{J}$ and thus on the class $\mathcal{F}$ from Theorem 13. As a consequence of (18) the functions $\varphi, \varphi_{0}$ and then also $e^{\varphi_{0}} \sum_{k=2}^{\infty}\left(\varphi-\varphi_{0}\right)^{k} / k$ ! are uniformly bounded in $C^{2}(\mathcal{O}) \subseteq \mathcal{C}^{2}(\mathcal{O})$. Thus, using also (11) we have

$$
\left\|f-f_{0}\right\|_{\left(H_{0}^{2}\right)^{*}} \leq \sup _{\|h\|_{H^{2}(\mathcal{O})} \leq 1}\left|\int_{\mathcal{O}} h\left(e^{\varphi}-e^{\varphi_{0}}\right)\right| \leq c_{0}\left\|\varphi-\varphi_{0}\right\|_{\left(H^{2}\right)^{*}}
$$

We can decompose $\left\|\varphi-\varphi_{0}\right\|_{\left(H^{2}\right)^{*}} \leq\left\|\Pi_{V_{J}}\left(\varphi-\varphi_{0}\right)\right\|_{\left(H^{2}\right)^{*}}+\left\|\Pi_{V_{J}}\left(\varphi_{0}\right)-\varphi_{0}\right\|_{\left(H^{2}\right)^{*}}$. For the second term we notice that by the compact support of $\varphi_{0}$ and (15), for $l>J$ large enough necessarily $\int_{\mathcal{O}} \varphi_{0} \Phi_{l, r}^{b c}=0$ for the boundary corrected wavelets, and $\Pi_{V_{J}}\left(\varphi_{0}\right)-\varphi_{0}$ then consists of a sum of interior Daubechies wavelets $\Phi_{l, r}^{\mathcal{O}}=\Phi_{l, r}$ supported compactly in $\mathcal{O}$, in particular there exists $K \subseteq \mathcal{O}$ compact such that $\Pi_{V_{J}}\left(\varphi_{0}\right)-\varphi_{0}$ is supported within $K$ for $J$ large enough. Multiplying $h \in H^{2}(\mathcal{O})$ by a function in $C_{c_{\tilde{h}}}^{\infty}(\mathcal{O})$ that equals 1 on $K$ we obtain $\widetilde{h} \in H^{2}\left(\mathbb{R}^{d}\right) \cap C_{c}(\mathcal{O})$ such that $\widetilde{h}=h$ on $K$ and $\|\widetilde{h}\|_{H^{2}\left(\mathbb{R}^{d}\right)} \leq c\|h\|_{H^{2}(\mathcal{O})}$. Then using Parseval's identity and the Cauchy-Schwarz inequality, we can estimate the dual norm by

$$
\begin{align*}
\left\|\Pi_{V_{J}}\left(\varphi_{0}\right)-\varphi_{0}\right\|_{\left(H^{2}\right)^{*}} & =\sup _{\|h\|_{H^{2}(\mathcal{O})} \leq 1}\left|\int_{\mathcal{O}} h\left(\Pi_{V_{J}}\left(\varphi_{0}\right)-\varphi_{0}\right)\right| \\
& =\sup _{\|h\|_{H^{2}(\mathcal{O})} \leq 1}\left|\int_{\mathbb{R}^{d}} \widetilde{h}\left(\Pi_{V_{J}}\left(\varphi_{0}\right)-\varphi_{0}\right)\right| \\
& =\sup _{\|h\|_{H^{2}(\mathcal{O})} \leq 1}\left|\sum_{l>J, r} 2^{-2 l}\left\langle\varphi_{0}, \Phi_{l, r}\right\rangle_{L^{2}\left(\mathbb{R}^{d}\right)} 2^{2 l}\left\langle\widetilde{h}, \Phi_{l, r}\right\rangle_{L^{2}\left(\mathbb{R}^{d}\right)}\right| \\
& \leq c^{\prime} \sup _{\|h\|_{H^{2}(\mathcal{O})} \leq 1}\|\widetilde{h}\|_{H^{2}\left(\mathbb{R}^{d}\right)} \sqrt{\sum_{l>J, r} 2^{-4 l}\left|\left\langle\Phi_{l, r}, \varphi_{0}\right\rangle\right|^{2}} \\
& \leq c^{\prime \prime} 2^{-J(s+2)} J^{-2}, \tag{48}
\end{align*}
$$

where we use (12) and (19). Now introducing 'true' coefficients

$$
\varphi_{0, l, r}=2^{-l(s+d / 2)} \bar{l}^{-2} u_{0, l, r}, \quad\left|u_{0, l, r}\right| \leq B,
$$

and if $U$ is a uniform $U(-1,1)$ random variable and $u_{l, r} \stackrel{\text { i.i.d. }}{\sim} U(-B, B)$, then using $\sum_{l \leq J} N_{l} \leq \bar{c}_{0} 2^{J d}$ and what precedes we can lower bound, for $\bar{L}$ large enough,

$$
\begin{aligned}
\Pi\left(f:\left\|f-f_{0}\right\|_{\left(H_{0}^{2}\right)^{*}}<\eta_{\varepsilon}\right) & \geq \Pi\left(\varphi:\left\|\varphi-\varphi_{0}\right\|_{\left(H^{2}\right)^{*}}<\eta_{\varepsilon} / c_{0}\right) \\
& \geq \Pi\left(\varphi:\left\|\Pi_{V_{J}}\left(\varphi-\varphi_{0}\right)\right\|_{\left(H^{2}\right)^{*}}<\eta_{\varepsilon} / c_{0}-c^{\prime \prime} J^{-2} 2^{-J(s+2)}\right) \\
& \geq \Pi\left(\varphi:\left\|\Pi_{V_{J}}\left(\varphi-\varphi_{0}\right)\right\|_{L^{2}}^{2}<c_{1} \eta_{\varepsilon}^{2}\right) \\
& =\operatorname{Pr}\left(\sum_{l \leq J, r} 2^{-l(2 s+d)} \bar{l}^{-2}\left|u_{l, r}-u_{0, l, r}\right|^{2}<c_{1} \eta_{\varepsilon}^{2}\right) \\
& \geq \operatorname{Pr}\left(\max _{l \leq J, r}\left|u_{l, r}-u_{0, l, r}\right|<c_{3} \eta_{\varepsilon}\right) \\
& =\prod_{l \leq J} \prod_{r=1}^{N_{l}} \operatorname{Pr}\left(\left|U-u_{0, l, r} / B\right|<c_{4} \eta_{\varepsilon}\right) \\
& \geq \operatorname{Pr}\left(c_{4} \eta_{\varepsilon} / 2\right)^{\bar{c}_{0} 2^{J d}} \geq e^{-c_{5} \log \left(c_{4} / \eta_{\varepsilon}\right) \varepsilon^{-2 d /(2 s+4+d)}} \geq e^{-C\left(\eta_{\varepsilon} / \varepsilon\right)^{2}}
\end{aligned}
$$

completing the proof.
Remark 16. If $\|\cdot\|$ is either the $L^{2}$ - or $\left(H_{0}^{2}\right)^{*}$-norm and $r_{\varepsilon}$ the corresponding contraction rate, then the proof of Theorem 1 via Theorem 28 actually implies that as $\varepsilon \rightarrow 0$, for some $c>0$,

$$
\Pi\left(\left\|f-f_{0}\right\|>r_{\varepsilon} \mid Y\right)=O_{P_{f_{0}}^{Y}}\left(e^{-c\left(\eta_{\varepsilon} / \varepsilon\right)^{2}}\right)
$$

### 6.3. Proof of Theorem 8

The proof is organised in a sequence of steps, and the main strategy is to prove the Bernstein-von Mises theorem via simultaneously controlling the Laplace transform of a collection of suitably regular linear functionals, an approach inspired by the papers of Ismaël Castillo and co-authors [7-11]. Of course a main challenge is to make this proof work outside of the basic LAN models considered in the references just mentioned, namely in the PDE setting considered here. However, even disregarding the different LAN structure, our proof needs to confront several new challenges when compared to the above papers: our prior has a boundary and hence using the 'perturbation of the likelihood function' approach employed in [48], [11], [10] needs some adjustments near the boundary. Moreover, our prior is supported in an $s$-regular Hölder wavelet ellipsoid that is asymptotically smaller than the ellipsoids considered in [10], which puts stronger constraints on the admissible directions one can choose when constructing perturbations (see also Remark 9).

Step I: Localisation of the posterior near $f_{0}$. We see from Proposition 15 that as $\varepsilon \rightarrow 0$,

$$
\Pi\left(\left\|f-f_{0}\right\|_{L^{2}(\mathcal{O})}>M \varepsilon^{2 s /(2 s+4+d)} \log ^{\gamma}(1 / \varepsilon) \mid Y\right) \xrightarrow{P_{f_{0}}^{Y}} 0
$$

and

$$
\Pi\left(\left\|f-f_{0}\right\|_{\left(H_{0}^{2}\right)^{*}}>M \varepsilon^{(2 s+4) /(2 s+4+d)} \sqrt{\log (1 / \varepsilon)} \mid Y\right) \xrightarrow{P_{f_{0}}^{Y}} 0
$$

From (18) we know $\|f\|_{C^{s}(\mathcal{O})} \leq C^{\prime}$ for any $f=e^{\varphi}$ and so both prior and posterior are supported in a fixed ball of $C^{s}(\mathcal{O}) \subseteq H^{s}(\mathcal{O})$. Since also $f_{0} \in H^{s}(\mathcal{O})$, we have $\left\|f-f_{0}\right\|_{H^{s}} \leq D$ for some fixed constant $D$, and recalling the standard interpolation inequality for Sobolev norms (see [37, Remark 1.9.1, p. 44])

$$
\begin{equation*}
\|f\|_{H^{\alpha}(\mathcal{O})} \lesssim\|f\|_{H^{s}(\mathcal{O})}^{\alpha / s}\|f\|_{L^{2}(\mathcal{O})}^{(s-\alpha) / s}, \quad 0<\alpha<s \tag{49}
\end{equation*}
$$

we also find, for all $M$ large enough and $\bar{\gamma}=\bar{\gamma}(\gamma, s, \alpha)>0$, that, as $\varepsilon \rightarrow 0$,

$$
\begin{equation*}
\Pi\left(\left\|f-f_{0}\right\|_{H^{\alpha}(\mathcal{O})}>M \varepsilon^{2(s-\alpha) /(2 s+4+d)} \log ^{\bar{\gamma}}(1 / \varepsilon) \mid Y\right) \xrightarrow{P_{f_{0}}^{Y}} 0, \quad 0<\alpha<s \tag{50}
\end{equation*}
$$

Since $s>2+d / 2$, we can apply the Sobolev imbedding theorem to deduce

$$
\begin{equation*}
\Pi\left(\left\|f-f_{0}\right\|_{C^{\beta}(\mathcal{O})}>M \varepsilon^{(2 s-2 \beta-d) /(2 s+4+d)} \log ^{\bar{\gamma}}(1 / \varepsilon) \mid Y\right) \xrightarrow{P_{f_{0}}^{Y}} 0,0<\beta<s-d / 2, \tag{51}
\end{equation*}
$$

which also implies a contraction rate in the uniform norm $\|\cdot\|_{\infty}$.
Since all $f \in \operatorname{supp} \Pi$ and $f_{0}$ are bounded and bounded away from zero on $\mathcal{O}$, the $\|\cdot\|_{L^{2}}$ and $\|\cdot\|_{\infty}$ contraction rates extend to $\varphi=\log f$ around $\varphi_{0}=\log f_{0}$, by simply using the estimate $\left\|\varphi-\varphi_{0}\right\| \lesssim\left\|f-f_{0}\right\|$ for these norms. The $\left(H_{0}^{2}\right)^{*}$-rates carry over too: To see this, notice that on the events in (51) with $\beta=2<s-d / 2$ we have $\left\|f-f_{0}\right\|_{C^{2}}$ $\rightarrow 0$ and then also $\left\|\left(f-f_{0}\right) / f_{0}\right\|_{C^{2}} \rightarrow 0$, so that using the Taylor expansion of the logarithm we also have

$$
\begin{align*}
\left\|\varphi-\varphi_{0}\right\|_{\left(H_{0}^{2}\right)^{*}} & =\sup _{g \in C_{0}(\mathcal{O}):\|g\|_{H^{2}} \leq 1}\left|\int_{\mathcal{O}} g \log \left(f / f_{0}\right)\right| \\
& =\sup _{g \in C_{0}(\mathcal{O}):\|g\|_{H^{2}} \leq 1}\left|\int_{\mathcal{O}} g\left(f-f_{0}\right) f_{0}^{-1} \sum_{k} \frac{(-1)^{k}}{k}\left(\frac{f-f_{0}}{f_{0}}\right)^{k-1}\right| \\
& \lesssim\left\|f-f_{0}\right\|_{\left(H_{0}^{2}\right)^{*}} \tag{52}
\end{align*}
$$

since for $\left\|\left(f-f_{0}\right) / f_{0}\right\|_{C^{2}}<1 / 2$ the series $w=f_{0}^{-1} \sum_{k}(-1)^{k} k^{-1}\left(\left(f-f_{0}\right) / f_{0}\right)^{k-1}$ converges absolutely in $C^{2}(\mathcal{O})$ and hence, after dividing and multiplying by $\|g w\|_{H^{2}} \leq$ $c\|g\|_{H^{2}}\|w\|_{C^{2}}$ via (11), the result follows on noting also that $g \in C_{0}(\mathcal{O})$ implies $g w \in C_{0}(\mathcal{O})$.

For a fixed constant $M$ to be chosen, let us now define the event

$$
\begin{equation*}
D_{\varepsilon}^{M}=\left\{f=e^{\varphi} \in \operatorname{supp} \Pi:\left\|\varphi-\varphi_{0}\right\|^{(i)} \leq \operatorname{Mr}(i, \varepsilon), i=1,2,3\right\} \tag{53}
\end{equation*}
$$

where, for some $\eta=\eta(s, \gamma)>0$ and $0<\xi / 2<s-2-d / 2$ small enough,

$$
\begin{array}{ll}
\|\cdot\|^{(1)}=\|\cdot\|_{L^{2}}, & r(1, \varepsilon)=\varepsilon^{2 s /(2 s+4+d)} \log ^{\eta}(1 / \varepsilon) \\
\|\cdot\|^{(2)}=\|\cdot\|_{\left(H_{0}^{2}\right)^{*}}, & r(2, \varepsilon)=\varepsilon^{(2 s+4) /(2 s+4+d)} \log ^{\eta}(1 / \varepsilon), \\
\|\cdot\|^{(3)}=\|\cdot\|_{\infty}, & r(3, \varepsilon)=\varepsilon^{(2 s-d-\xi) /(2 s+4+d)} \log ^{\eta}(1 / \varepsilon) .
\end{array}
$$

By increasing the constant $M$ to $\bar{M}$ we also obtain

$$
\begin{equation*}
D_{\varepsilon}^{M} \subseteq\left\{f \in \operatorname{supp} \Pi:\left\|f-f_{0}\right\|^{(i)} \leq \bar{M} r(i, \varepsilon), i=1,2,3\right\} \tag{54}
\end{equation*}
$$

which is obvious for $i=1,3$ and follows also for $i=2$ by just applying the argument in (52) with the Taylor series expansion of the exponential (instead of the logarithm) function, using the fact that $\left\|\left(f-f_{0}\right) / f_{0}\right\|_{C^{2}}$ is bounded by a fixed constant.

By Theorem 1 and what precedes we have

$$
\Pi\left(\left(D_{\varepsilon}^{M}\right)^{c} \mid Y\right) \rightarrow 0 \quad \text { in } P_{f_{0}}^{Y} \text { probability, }
$$

for all $M$ large enough. In particular if $W$ is the constant in inequality (61) below, then there exists $M_{0}$ such that the last limit holds for all $M \geq M_{0} /(8 W)$, and given such an $M_{0}$ we fix any value $M \geq M_{0}$ in what follows and just write $D_{\varepsilon}$ for $D_{\varepsilon}^{M}$ when no confusion may arise. Then, if $\Pi^{D_{\varepsilon}}(\cdot \mid Y)$ is the posterior distribution arising from the prior $\Pi$ restricted to the set $D_{\varepsilon}$, that is, from $\Pi\left(\cdot \cap D_{\varepsilon}\right) / \Pi\left(D_{\varepsilon}\right)$, standard arguments imply that

$$
\begin{equation*}
\sup _{B}\left|\Pi(B \mid Y)-\Pi^{D_{\varepsilon}}(B \mid Y)\right| \leq 2 \Pi\left(D_{\varepsilon}^{c} \mid Y\right) \xrightarrow{P_{f_{0}}^{Y}} 0 \tag{55}
\end{equation*}
$$

where the supremum extends over all measurable sets $B$ in supp $\Pi$.
Step II: Construction of the perturbation $f_{\tau}$. Fix $\gamma>0$. For arbitrary $\psi \in C_{K}^{4}(\mathcal{O})$ such that $\|\psi\|_{C^{2+d / 2+\gamma}(\mathcal{O})} \leq 1$, we now construct a suitable perturbation $f_{\tau}$ of $f \in$ supp $П$ —an asymptotic expansion of the log-likelihood ratio $\ell(f)-\ell\left(f_{\tau}\right)$ obtained in the next step will be a key element of our proof. For such $\psi$ define

$$
\begin{equation*}
\widetilde{\psi}=-\widetilde{\Psi} / f_{0} \tag{56}
\end{equation*}
$$

with $\widetilde{\Psi}$ as in (25). From the representation (26), Proposition 25(A), the definition of the $C^{\alpha}$-norms and the hypotheses on $f_{0}, g$ which imply that $u_{f_{0}}^{-1} \in C^{s+2}(\mathcal{O}), f_{0}^{-1} \in C^{s}(\mathcal{O})$, we deduce that $\tilde{\psi} \in C_{K}(\mathcal{O})$, and if $-2+d / 2 \geq 0$ then also $\|\tilde{\psi}\|_{C^{-2+d / 2+\gamma}(\mathcal{O})} \leq$ $c\|\psi\|_{C^{2+d / 2+\gamma}(\mathcal{O})}$ for some finite constant $c>0$. Thus from Proposition 26 and (16),

$$
\begin{equation*}
\|\tilde{\psi}\|_{\mathcal{C}^{-2+d / 2+\gamma, W}(\mathcal{O})} \leq C\|\psi\|_{C^{2+d / 2+\gamma}(\mathcal{O})}, \quad \text { so } \quad\left|\left\langle\tilde{\psi}, \Phi_{l, r}^{\mathcal{O}}\right\rangle_{L^{2}}\right| \leq C 2^{-l(d-2+\gamma)} \tag{57}
\end{equation*}
$$

The last estimate is true also in the case $-2+d / 2<0$, as follows from (26) and the inequalities (105), (106) in Proposition 27 below.

For $f \in \operatorname{supp} \Pi$ the perturbation will be

$$
\begin{equation*}
f_{\tau}=f \exp \{\tau\}=e^{\varphi+\tau} \tag{58}
\end{equation*}
$$

(with a slight abuse of notation when $\tau=0$ ) where $\tau$ is defined as follows: Let $J$ be the cut-off parameter of the prior, and for $L \in \mathbb{N}, L \leq J$, define $W_{L} \subseteq L^{2}(\mathcal{O})$ to be the linear span of those wavelet basis functions $\left(\Phi_{l, r}^{\mathcal{O}}: l \leq L\right)$ for which either $\left\langle\widetilde{\psi}, \Phi_{l, r}^{\mathcal{O}}\right\rangle \neq 0$ or
$\left\langle\Phi_{l, r}^{\mathcal{O}}, \varphi_{0}\right\rangle_{L^{2}} \neq 0$; in other words, $W_{L}$ is the subspace of $V_{L}$ where $\Pi_{V_{L}}\left(\varphi_{0}\right)$ and $\Pi_{V_{L}}(\widetilde{\psi})$ are supported. For $\Pi_{W_{L}}$ the corresponding $L^{2}$-projection operator set

$$
\begin{equation*}
\tau \equiv \tau(\varphi, \psi, t, L)=\Pi_{W_{L}}\left[t \varepsilon \tilde{\psi}+\delta_{\varepsilon}\left(\varphi_{0}-\varphi\right)\right]=t \varepsilon \Pi_{W_{L}}[\tilde{\psi}]+\delta_{\varepsilon} \Pi_{W_{L}}\left[\varphi_{0}-\varphi\right], t \in \mathbb{R} \tag{59}
\end{equation*}
$$

with, for $\gamma_{1}>0$ small enough chosen below (depending on $\gamma$ ),

$$
\begin{equation*}
\delta_{\varepsilon}=\varepsilon^{\frac{2 d+\gamma_{1}}{2 s+4+d}} \tag{60}
\end{equation*}
$$

[The second summand in the perturbation (59) is different from the perturbations used in previous proofs of this kind $[10,11,48]$. It will be useful to deal with the boundary of the support of the prior in Step IV below, and was used in a related context of non-parametric maximum likelihood estimation over parameter spaces with boundaries in [41].]

The projection $\Pi_{W_{L}}$ is a bounded operator for all the norms introduced after (53), that is,

$$
\begin{equation*}
\left\|\Pi_{W_{L}}\left(\varphi-\varphi_{0}\right)\right\|^{(i)} \leq W\left\|\varphi-\varphi_{0}\right\|^{(i)}, \quad i=1,2,3, \forall \varphi \in V_{J} \tag{61}
\end{equation*}
$$

for some fixed finite constant $W \geq 1$. This is clear for $i=1$ and also follows easily for $i=2,3$, by proceeding as in the corresponding proofs for wavelet bases of $L^{2}\left(\mathbb{R}^{d}\right)$ $[18,27,39]$ and making use of (15) and the compact support of $\varphi_{0}, \widetilde{\psi}$. [In case $i=2$, one initially proves that $\Pi_{W_{L}}$ defines a bounded operator on $H^{2}(\mathcal{O})$.]

Using Parseval's identity and (57) we then have

$$
\begin{equation*}
\left\|\Pi_{W_{L}} \tilde{\psi}\right\|_{L^{2}(\mathcal{O})} \leq c \sqrt{\sum_{l \leq J} \sum_{r}\left\langle\widetilde{\psi}, \Phi_{l, r}^{\mathcal{O}}\right\rangle_{L^{2}}^{2}} \leq C \sqrt{\sum_{l<J} 2^{l(4-d-2 \gamma)}} \lesssim 2^{J(2-d / 2)}=o\left(\varepsilon^{-1}\right) \tag{62}
\end{equation*}
$$

for all $\gamma>0$ and likewise, using (15) and (57), for every $i,|i|=\beta, 0 \leq \beta<s+d, \gamma>0$,

$$
\begin{align*}
\left\|D^{i} \Pi_{W_{L}} \tilde{\psi}\right\|_{\infty} & \leq \sum_{l \leq J} 2^{l(\beta+d / 2)} \sup _{x} \sum_{r}\left|\left\langle\tilde{\psi}, \Phi_{l, r}^{\mathcal{O}}\right\rangle_{L^{2}}\right|\left|\left(D^{i} \Phi_{0, r}^{\mathcal{O}}\right)\left(2^{l} x\right)\right| \\
& \leq c \sum_{l \leq J} 2^{l(2+\beta-d / 2-\gamma)} \lesssim \varepsilon^{(-4-2 \beta+d) /(2 s+4+d)}=o\left(\varepsilon^{-1}\right) \tag{63}
\end{align*}
$$

in particular $\varepsilon\left\|\Pi_{W_{L}} \widetilde{\psi}\right\|_{C^{\beta}(\mathcal{O})}=o(1)$. Now since $\varphi, \varphi_{0}$ are all bounded functions, we conclude that $\|\tau\|_{\infty} \rightarrow 0$ as $\varepsilon \rightarrow 0$, and thus also

$$
\begin{equation*}
f_{\tau} \geq c^{\prime}>0, \quad\left|f_{\tau}-f\right|=|f|\left|1-e^{\tau}\right| \leq c^{\prime \prime}|\tau| \underset{\varepsilon \rightarrow 0}{\longrightarrow} 0, \quad c^{\prime}, c^{\prime \prime}>0 \tag{64}
\end{equation*}
$$

Likewise we have

$$
\begin{align*}
\left\|f_{\tau}-f\right\|_{\left(H_{0}^{2}\right)^{*}} & =\sup _{g \in C_{0}(\mathcal{O})::\|g\|_{H^{2} \leq 1}}\left|\int_{\mathcal{O}} g f\left(1-e^{\tau}\right)\right| \\
& =\sup _{g \in C_{0}(\mathcal{O}):\|g\|_{H^{2}} \leq 1}\left|\int_{\mathcal{O}} \tau\left(g f \sum_{k=1}^{\infty} \tau^{k-1} / k!\right)\right| \lesssim\|\tau\|_{\left(H_{0}^{2}\right)^{*}} \tag{65}
\end{align*}
$$

since the above and the assumptions on $f, f_{0}, \varphi, \varphi_{0}$ imply that $f \sum_{k=1}^{\infty} \tau^{k-1} / k$ ! is bounded in $C^{2}(\mathcal{O})$ and hence by (11) the functions

$$
g f \sum_{k=1}^{\infty} \frac{\tau^{k-1}}{k!} \in C_{0}(\mathcal{O}), \quad\|g\|_{H^{2}} \leq 1
$$

are bounded in $H^{2}(\mathcal{O})$.

## Step III: Expanding the likelihood in the Laplace transform

Theorem 17. For $\psi \in C_{K}^{4}(\mathcal{O})$ define $\tilde{\Psi}$ as in (25). Let $t>0$ and $f_{\tau}$ be as in (58), (59) for any $\gamma_{1}>0$. Suppose $L=L_{\varepsilon} \uparrow \infty$ as $\varepsilon \rightarrow 0$ is a sequence of integers such that either
(i) $L=J$, or
(ii) $L \leq J$ but $\left\|\Pi_{W_{L}}(\widetilde{\Psi})-\widetilde{\Psi}\right\|_{\left(H_{0}^{2}\right)^{*}} \leq c \varepsilon^{\bar{\gamma} /(2 s+4+d)}$ for some $\bar{\gamma}>d$ and some constant $c>0$.
If $\Pi^{D_{\varepsilon}}(\cdot \mid Y)$ is the posterior distribution arising from the prior $\Pi$ restricted to the set $D_{\varepsilon}$ from (53) with $M$ chosen as before (55), and if $D G_{f_{0}}$ is the score operator from Lemma 4, then for all $t \in \mathbb{R}$,

$$
\begin{equation*}
E^{\Pi^{D_{\varepsilon}}}\left[\left.e^{\frac{t}{\varepsilon}\left\langle f-f_{0}, \psi\right\rangle_{L^{2}}} \right\rvert\, Y\right]=e^{-t\left\langle D G_{f_{0}}[\widetilde{\Psi}], \mathbb{W}\right\rangle_{L^{2}}+\frac{t^{2}}{2}\left\|D G_{f_{0}}[\widetilde{\Psi}]\right\|_{L^{2}}^{2}} \cdot \frac{\int_{D_{\varepsilon}} e^{\ell\left(f_{\tau}\right)} d \Pi(f)}{\int_{D_{\varepsilon}} e^{\ell(f)} d \Pi(f)} \cdot e^{R_{\varepsilon}} \tag{66}
\end{equation*}
$$

where $R_{\varepsilon}=o_{P_{f_{0}}^{Y}}(1)$ uniformly in $|t| \leq T$ for any $T$ and in $\psi \in \mathcal{C}(b) \equiv\left\{\psi \in C_{K}^{4}(\mathcal{O})\right.$ : $\left.\|\psi\|_{C^{2+d / 2+\gamma}(\mathcal{O})} \leq b\right\}$ for any $b, \gamma>0$, and every fixed compact subset $K \subseteq \mathcal{O}$.
Proof. From (6) and recalling $\ell(f)=\log p_{f}(Y)$ we have, for all $t \in \mathbb{R}$,

$$
\begin{equation*}
E^{\Pi^{D_{\varepsilon}}}\left[\left.e^{\frac{t}{\varepsilon}\left\langle f-f_{0}, \psi\right\rangle_{L^{2}}} \right\rvert\, Y\right]=\frac{\int_{D_{\varepsilon}} e^{(t / \varepsilon)\left\langle f-f_{0}, \psi\right\rangle_{L^{2}}+\ell(f)-\ell\left(f_{\tau}\right)+\ell\left(f_{\tau}\right)} d \Pi(f)}{\int_{D_{\varepsilon}} e^{\ell(f)} d \Pi(f)} \tag{67}
\end{equation*}
$$

The main technical work now is to obtain an asymptotic expansion of $\ell(f)-\ell\left(f_{\tau}\right)$ that is uniform in $f, \psi$. To do this, we need to start with two linearisation steps: the first takes care of the non-linearity of the forward operator $G$, and the second of the exponential nature of our perturbation $f_{\tau}$, ultimately leading to (73) below.

A: Linearisation of the $G$ operator. Using Lemma 3 and Proposition 4 to expand both $G(f)=u_{f}$ and $G\left(f_{\tau}\right)=u_{f_{\tau}}$ about $G\left(f_{0}\right)=u_{f_{0}}-$ from (63) and the definition of the prior we know that $f, f_{\tau}, f_{0}$ are all bounded in $C^{s}(\overline{\mathcal{O}})$ and bounded away from zero-we obtain the approximation, for $f \in D_{\varepsilon}$,

$$
\begin{align*}
& \ell(f)-\ell\left(f_{\tau}\right)=-\frac{1}{2 \varepsilon^{2}}\left(\left\|u_{f}-u_{f_{0}}\right\|_{L^{2}}^{2}-\left\|u_{f_{\tau}}-u_{f_{0}}\right\|_{L^{2}}^{2}\right)+\frac{1}{\varepsilon}\left\langle u_{f}-u_{f_{\tau}}, \mathbb{W}\right\rangle_{L^{2}} \\
&=-\frac{1}{2 \varepsilon^{2}}\left(\left\|D G_{f_{0}}\left[f-f_{0}\right]\right\|_{L^{2}}^{2}-\left\|D G_{f_{0}}\left[f_{\tau}-f_{0}\right]\right\|_{L^{2}}^{2}\right)+\frac{1}{\varepsilon}\left\langle D G_{f_{0}}\left[f-f_{\tau}\right], \mathbb{W}\right\rangle_{L^{2}} \\
&+R(f, \psi)+N(f, \psi) \tag{68}
\end{align*}
$$

where for some $c>0$, using also the Cauchy-Schwarz inequality and Proposition 25(B),

$$
|R(f, \psi)| \leq \frac{c}{\varepsilon^{2}}\left\|f-f_{0}\right\|_{(d)}^{2}\left\|f-f_{0}\right\|_{\left(H_{0}^{2}\right)^{*}}^{2}+\frac{c}{\varepsilon^{2}}\left\|f_{\tau}-f\right\|_{(d)}^{2}\left\|f_{\tau}-f\right\|_{\left(H_{0}^{2}\right)^{*}}^{2}
$$

and where the stochastic remainder term is given by

$$
N(f)=\varepsilon^{-1}\langle w(f, \psi), \mathbb{W}\rangle_{L^{2}}
$$

with $w(f, \psi)$ described below.
For the first term in the bound for $R(f, \psi)$ let us first assume $d<4$; then on $D_{\varepsilon}$ (see (54)) the inequality

$$
\frac{c \bar{M}^{4}}{\varepsilon^{2}} r^{2}(2, \varepsilon) r^{2}(1, \varepsilon) \lesssim \varepsilon^{-(4 s+8+2 d) /(2 s+4+d)} \varepsilon^{(8 s+8) /(2 s+4+d)} \log ^{4 \eta}(1 / \varepsilon)=o(1)
$$

holds since $s>d / 2$. Also from (64), (65) we know

$$
\left\|f_{\tau}-f\right\|_{L^{2}}^{2}\left\|f_{\tau}-f\right\|_{\left(H_{0}^{2}\right)^{*}}^{2} \lesssim\|\tau\|_{L^{2}}^{2}\|\tau\|_{\left(H_{0}^{2}\right)^{*}}^{2},
$$

which is $o\left(\varepsilon^{2}\right)$ uniformly in $f \in D_{\varepsilon}$ and $\psi \in \mathcal{C}(b)$ since

$$
\begin{equation*}
\delta_{\varepsilon}^{2}\left\|\varphi-\varphi_{0}\right\|_{L^{2}}\left\|\varphi-\varphi_{0}\right\|_{\left(H_{0}^{2}\right)^{*}}=o(\varepsilon) \tag{69}
\end{equation*}
$$

since by (62),

$$
\begin{equation*}
\left\|\Pi_{W_{L}}[\tilde{\psi}]\right\|_{L^{2}}\left\|\Pi_{W_{L}}[\tilde{\psi}]\right\|_{\left(H_{0}^{2}\right)^{*}} \lesssim\left\|\Pi_{W_{L}}[\tilde{\psi}]\right\|_{L^{2}}^{2}=o\left(\varepsilon^{-2}\right) \tag{70}
\end{equation*}
$$

and bounding the 'cross term' similarly. For $d \geq 4$ the $\|\cdot\|_{L^{2}}$-norms have to be replaced by $\|\cdot\|_{\infty}$-norms in the above estimates, which results in slightly worse convergence rates on $D_{\varepsilon}$ and the requirement $s>d$ in place of $s>d / 2$.

For the stochastic term $N(f, \psi)$ we notice that $w(f, \psi)=w_{f, \psi}-w_{f, \psi}^{\prime}$ in

$$
\begin{aligned}
G(f)-G\left(f_{\tau}\right) & =D G_{f}\left[f-f_{\tau}\right]-w_{f, \psi}^{\prime} \\
& =D G_{f_{0}}\left[f-f_{\tau}\right]+\left(D G_{f}-D G_{f_{0}}\right)\left(f-f_{\tau}\right)-w_{f, \psi}^{\prime} \\
& \equiv D G_{f_{0}}\left[f-f_{\tau}\right]-w_{f, \psi}^{\prime}+w_{f, \psi}
\end{aligned}
$$

where $w_{f, \psi}^{\prime}$ solves (just as in (24)) the inhomogeneous Schrödinger equation

$$
(\Delta / 2) w-f_{\tau} w=-\left(f-f_{\tau}\right) V_{f}\left[u_{f}\left(f_{\tau}-f\right)\right] \quad \text { on } \mathcal{O}, \quad w=0 \quad \text { on } \partial \mathcal{O}
$$

and where $V_{f}$ denotes the inverse Schrödinger operator from Section 7.2. Recalling (53) and (62), we see that both $\left\{\varphi: e^{\varphi} \in D_{\varepsilon}\right\}$ and $\left\{\varepsilon \Pi_{W_{L}}[\widetilde{\psi}]: \psi \in \mathcal{C}(b)\right\}$ are bounded subsets of the linear space $\left(V_{J},\|\cdot\|_{L^{2}}\right)$ which is isomorphic to a $c_{0} 2^{J d}$-dimensional Euclidean space (by Parseval's identity). Any such set can be covered by $(A / \eta)^{c_{0} 2^{J d}}$ balls of $L^{2}$ diameter at most $\eta$ for all $0<\eta<A$ and some $A<\infty$ [27, Proposition 4.3.34], and
moreover, using Lemma 23 and Proposition 25 one shows

$$
\begin{aligned}
& \left\|w_{f, \psi}-w_{f^{\prime}, \psi}\right\|_{L^{2}} \\
& \quad=\left\|V_{f_{0}}\left[u_{f_{0}}\left(f-f_{\tau}\right)\right]-V_{f}\left[u_{f}\left(f-f_{\tau}\right)\right]-\left(V_{f_{0}}\left[u_{f_{0}}\left(f^{\prime}-f_{\tau}^{\prime}\right)\right]-V_{f^{\prime}}\left[u_{f^{\prime}}\left(f^{\prime}-f_{\tau}^{\prime}\right)\right]\right)\right\|_{L^{2}} \\
& \quad \lesssim\left\|f-f^{\prime}\right\|_{L^{2}} \lesssim\left\|\varphi-\varphi^{\prime}\right\|_{L^{2}}
\end{aligned}
$$

and likewise $\left\|w_{f^{\prime}, \psi}-w_{f^{\prime}, \psi^{\prime}}\right\|_{L^{2}} \lesssim\left\|\Pi_{W_{L}}[\widetilde{\psi}]-\Pi_{W_{L}}\left[\widetilde{\psi^{\prime}}\right]\right\|_{L^{2}}$. Thus the set $\mathcal{W}=\left\{w_{f, \psi}\right.$ : $\left.f \in D_{\varepsilon}, \psi \in \mathcal{C}(b)\right\}$ can be covered in $L^{2}$-metric by at most $\left(A^{\prime} / \eta\right)^{c^{\prime} 2^{J d}}$ balls of radius $\eta$, and if $\sigma \geq \sup _{w \in \mathcal{W}}\|w\|_{L^{2}}$ then (92) below applies to the Gaussian process $\left\{\langle\mathbb{W}, w\rangle_{L^{2}(\mathcal{O})}: w \in \mathcal{W}\right\}$ which is sub-Gaussian for the metric $\left\|w_{f, \psi}-w_{f^{\prime}, \psi^{\prime}}\right\|_{L^{2}}$. We conclude that

$$
\begin{equation*}
E \sup _{f \in D_{\varepsilon}, \psi \in \mathcal{C}(b)}\left|\left\langle w_{f, \psi}, \mathbb{W}\right\rangle_{L^{2}(\mathcal{O})}\right| \leq C\left(\sigma+\int_{0}^{\sigma} 2^{J d / 2} \sqrt{2 \log \left(A^{\prime} / \eta\right)} d \eta\right) \tag{71}
\end{equation*}
$$

We see from Proposition 25 and Lemma 23 that

$$
\left\|w_{f, \psi}\right\|_{L^{2}}=\left\|V_{f_{0}}\left[u_{f_{0}}\left(f-f_{\tau}\right)\right]-V_{f}\left[u_{f}\left(f-f_{\tau}\right)\right]\right\|_{L^{2}} \lesssim\left\|f-f_{0}\right\|_{L^{2}}\|\tau\|_{\infty},
$$

a bound that will be seen to converge to zero at a polynomial rate in $\varepsilon$, uniformly in $w \in \mathcal{W}$. By basic calculus [27, p. 190] we can thus bound (71) up to constants by $\sup _{f \in D_{\varepsilon}, \psi \in \mathcal{C}(b)} 2^{J d / 2}\left\|f-f_{0}\right\|_{L^{2}}\|\tau\|_{\infty} \sqrt{\log (1 / \varepsilon)}$. Next we turn to suprema of the Gaussian process $\left(\left\langle\mathbb{W}, w_{f, \psi}^{\prime}\right\rangle_{L^{2}}: f \in D_{\varepsilon}, \psi \in \mathcal{C}(b)\right\}$. Using again the results from Section 7.2 we have

$$
\begin{aligned}
\left\|w_{f, \psi}^{\prime}-w_{f^{\prime}, \psi}^{\prime}\right\|_{L^{2}} & =\left\|V_{f_{\tau}}\left[\left(f-f_{\tau}\right) V_{f}\left[u_{f}\left(f_{\tau}-f\right)\right]\right]-V_{f_{\tau}^{\prime}}\left[\left(f^{\prime}-f_{\tau}^{\prime}\right) V_{f^{\prime}}\left[u_{f^{\prime}}\left(f_{\tau}^{\prime}-f^{\prime}\right)\right]\right]\right\|_{L^{2}} \\
& \lesssim\left\|f-f^{\prime}\right\|_{L^{2}} \lesssim\left\|\varphi-\varphi^{\prime}\right\|_{L^{2}},
\end{aligned}
$$

and $\left\|w_{f^{\prime}, \psi}^{\prime}-w_{f^{\prime}, \psi^{\prime}}^{\prime}\right\|_{L^{2}} \lesssim\left\|\Pi_{W_{L}}\left(\widetilde{\psi}-\widetilde{\psi}^{\prime}\right)\right\|_{L^{2}}$ as well as

$$
\left\|w_{f}^{\prime}\right\|_{L^{2}}=\left\|V_{f_{\tau}}\left[\left(f-f_{\tau}\right) V_{f}\left[\left(f_{\tau}-f\right) u_{f}\right]\right]\right\|_{L^{2}} \lesssim\|\tau\|_{\infty}\left\|f_{\tau}-f\right\|_{L^{2}} \equiv \sigma
$$

so that repeating the arguments above we obtain

$$
\begin{aligned}
& E \sup _{f \in D_{\varepsilon}, \psi \in \mathcal{C}(b)}|N(f, \psi)| \\
& \lesssim \sup _{f \in D_{\varepsilon}, \psi \in \mathcal{C}(b)} \varepsilon^{-1} 2^{J d / 2}\left(\left\|f-f_{0}\right\|_{L^{2}}+\left\|f_{\tau}-f\right\|_{L^{2}}\right)\|\tau\|_{\infty} \sqrt{\log (1 / \varepsilon)}
\end{aligned}
$$

Using the definition of $\tau$ and $s>2+d / 2+\xi / 2$, we get

$$
\begin{aligned}
& \varepsilon^{-1} 2^{J d / 2} \sqrt{\log (1 / \varepsilon)} \delta_{\varepsilon}\left\|\varphi-\varphi_{0}\right\|_{\infty}\left\|f-f_{0}\right\|_{L^{2}} \\
& \quad \lesssim \varepsilon^{\left(-2 s-4-d-d+2 d+2 s-d-\xi+2 s+\gamma_{1}\right) /(2 s+4+d)} \log ^{2 \eta+1 / 2}(1 / \varepsilon)=o(1),
\end{aligned}
$$

the same bound holds with $\left\|f-f_{0}\right\|_{L^{2}}$ replaced by $\delta_{\varepsilon}\left\|\varphi-\varphi_{0}\right\|_{L^{2}}$. Similarly using (63) with $\beta=0$ in
$2^{J d / 2} \sqrt{\log (1 / \varepsilon)}\left\|\Pi_{W_{L}} \tilde{\psi}\right\|_{\infty}\left\|f-f_{0}\right\|_{L^{2}} \lesssim \varepsilon^{(-d-4+d+2 s) /(2 s+4+d)} \log ^{\eta+1 / 2}(1 / \varepsilon)=o(1)$, and in

$$
\varepsilon 2^{J d / 2} \sqrt{\log (1 / \varepsilon)}\left\|\Pi_{W_{L}} \widetilde{\psi}\right\|_{\infty}\left\|\Pi_{W_{L}} \widetilde{\psi}\right\|_{L^{2}}=o(1)
$$

we obtain

$$
\sup _{f \in D_{\varepsilon}, \psi \in \mathcal{C}(b)}|N(f, \psi)|=o_{P_{f_{0}}^{Y}}(1)
$$

since the preceding bounds are all uniform in $f, \psi$.
In view of the above estimates for $R(f, \psi), N(f, \psi)$ the LAN expansion (68) now simplifies to

$$
\begin{align*}
\ell(f)-\ell\left(f_{\tau}\right)= & \frac{1}{\varepsilon}\left\langle D G_{f_{0}}\left[f-f_{\tau}\right], \mathbb{W}\right\rangle_{L^{2}}+\frac{1}{2 \varepsilon^{2}}\left\|D G_{f_{0}}\left[f-f_{\tau}\right]\right\|_{L^{2}}^{2} \\
& +\frac{1}{\varepsilon^{2}}\left\langle D G_{f_{0}}\left[f-f_{0}\right], D G_{f_{0}}\left(f-f_{\tau}\right)\right\rangle_{L^{2}}+Z_{n} \\
& \equiv I+I I+I I I+Z_{\varepsilon} \tag{72}
\end{align*}
$$

where $Z_{\varepsilon}=o_{P_{f_{0}}^{Y}}(1)$ uniformly in $f \in D_{\varepsilon}$ and $\psi \in \mathcal{C}(b)$, for any $\gamma>0$.
B: Linearisation of the exponential perturbation. We now linearise $f_{\tau}-f=f\left(1-e^{\tau}\right)$ in $\tau$. Since $\|\tau\|_{\infty} \rightarrow 0$ as $\varepsilon \rightarrow 0$, we assume $\|\tau\|_{\infty}<1 / 2$ in what follows. We have

$$
D G_{f_{0}}\left[f-f_{\tau}\right]-D G_{f_{0}}\left[f_{0} \tau\right]=D G_{f_{0}}\left[f-f_{\tau}-f \tau-\left(f_{0}-f\right) \tau\right] .
$$

Thus from (23) and the triangle inequality we deduce

$$
\begin{aligned}
\left\|D G_{f_{0}}\left[f-f_{\tau}\right]-D G_{f_{0}}\left[f_{0} \tau\right]\right\|_{L^{2}}^{2} & \leq c_{2}\left\|f-f_{\tau}-f \tau\right\|_{\left(H_{0}^{2}\right)^{*}}^{2}+c_{2}\left\|\left(f_{0}-f\right) \tau\right\|_{\left(H_{0}^{2}\right)^{*}}^{2} \\
& \lesssim\|f\|_{\infty}^{2}\left\|\tau^{2}\right\|_{L^{2}}^{2}+\left\|\left(f_{0}-f\right) \tau\right\|_{L^{2}}^{2} \\
& \lesssim\|\tau\|_{\infty}^{2}\left(\|\tau\|_{L^{2}}^{2}+\left\|f-f_{0}\right\|_{L^{2}}^{2},\right.
\end{aligned}
$$

which we shall use in the following estimates: First, for term II in (72) we have

$$
\left|I I-\frac{1}{2 \varepsilon^{2}}\left\|D G_{f_{0}}\left[f_{0} \tau\right]\right\|_{L^{2}}^{2}\right| \leq \frac{c^{\prime}}{\varepsilon^{2}}\|\tau\|_{\infty}^{2}\left(\|\tau\|_{L^{2}}^{2}+\left\|f-f_{0}\right\|_{L^{2}}^{2}\right)=o(1)
$$

using the definition of $\tau$ in (59), (62), (63) and $s>2$ in the bounds

$$
\begin{aligned}
\left\|\Pi_{W_{L}} \widetilde{\psi}\right\|_{\infty}^{2}\left\|f-f_{0}\right\|_{L^{2}}^{2} & \lesssim \varepsilon^{(-8+2 d+4 s) /(2 s+4+d)} \log ^{2 \eta}(1 / \varepsilon)=o(1), \\
\left\|\Pi_{W_{L}} \widetilde{\psi}\right\|_{\infty}^{2} \varepsilon^{2}\left\|\Pi_{W_{L}} \widetilde{\psi}\right\|_{L^{2}}^{2} & \lesssim \varepsilon^{2} \varepsilon^{(-16+4 d) /(2 s+4+d)}=o(1), \\
\varepsilon^{-2} \delta_{\varepsilon}\left\|\Pi_{W_{L}}\left(\varphi-\varphi_{0}\right)\right\|_{\infty}^{2}\left\|f-f_{0}\right\|_{L^{2}}^{2} & \lesssim \varepsilon^{-2} \varepsilon^{(4 d+4 s-2 d+4 s-2 \xi) /(2 s+4+d)} \log ^{4 \eta}(1 / \varepsilon) \\
& =o(1),
\end{aligned}
$$

and noting that the term involving $\delta_{\varepsilon}\left(\varphi-\varphi_{0}\right)$ in place of $f-f_{0}$ is of even smaller order of magnitude on $D_{\varepsilon}$. Similarly

$$
\begin{aligned}
\left|I I I-\frac{1}{\varepsilon^{2}}\left\langle D G_{f_{0}}\left[f-f_{0}\right], D G_{f_{0}}\left(f_{0} \tau\right)\right\rangle_{L^{2}}\right| & \lesssim \varepsilon^{-2}\left\|f-f_{0}\right\|_{\left(H_{0}^{2}\right)^{*}}\|\tau\|_{\infty}\left(\|\tau\|_{L^{2}}+\left\|f-f_{0}\right\|_{L^{2}}\right) \\
\lesssim & \varepsilon^{-2}\left\|f-f_{0}\right\|_{\left(H_{0}^{2}\right)^{*}}\left(\delta_{\varepsilon}\left\|\varphi-\varphi_{0}\right\|_{\infty}+\varepsilon\left\|\Pi_{W_{L}} \widetilde{\psi}\right\|_{\infty}\right) \\
& \times\left(\left\|f-f_{0}\right\|_{L^{2}}+\delta_{\varepsilon}\left\|\varphi-\varphi_{0}\right\|_{L^{2}}+\varepsilon\left\|\Pi_{W_{L}} \widetilde{\psi}\right\|_{L^{2}}\right)
\end{aligned}
$$

is $o(1)$ on $D_{\varepsilon}$ using $s>2$ and (63) in

$$
\begin{array}{r}
\varepsilon^{-1}\left\|f-f_{0}\right\|_{\left(H_{0}^{2}\right)}\left\|\Pi_{W_{L}} \widetilde{\psi}\right\|_{\infty}\left\|f-f_{0}\right\|_{L^{2}} \lesssim \varepsilon^{(-2 s-4-d+4 s+4+d-4) /(2 s+4+d)} \log ^{2 \eta}(1 / \varepsilon) \\
=o(1) \\
\left\|f-f_{0}\right\|_{\left(H_{0}^{2}\right)^{*}}\left\|\Pi_{W_{L}} \widetilde{\psi}\right\|_{\infty}\left\|\Pi_{W_{L}} \widetilde{\psi}\right\|_{L^{2}} \lesssim \varepsilon^{(2 s+4-8+2 d) /(2 s+4+d)} \log ^{\eta}(1 / \varepsilon)=o(1)
\end{array}
$$

and $s>2+d / 2+\xi / 2$ in

$$
\begin{aligned}
& \varepsilon^{-2} \delta_{\varepsilon}\left\|f-f_{0}\right\|_{\left(H_{0}^{2}\right)^{*}}\left\|\varphi-\varphi_{0}\right\| \infty\left\|f-f_{0}\right\|_{L^{2}} \\
& \lesssim \varepsilon^{(-4 s-8-2 d+2 d+2 s+4+2 s-d-\xi+2 s) /(2 s+4+d)} \log ^{3 \eta}(1 / \varepsilon) \\
&=\varepsilon^{(2 s-4-d-\xi) /(2 s+4+d)} \log ^{3 \eta}(1 / \varepsilon)=o(1)
\end{aligned}
$$

that $\delta_{\varepsilon}\left\|\varphi-\varphi_{0}\right\|_{L^{2}}$ is of smaller order of magnitude than $\left\|f-f_{0}\right\|_{L^{2}}$, and that a similar bound holds if in the last display $\left\|f-f_{0}\right\|_{L^{2}}$ is replaced by $\left\|\Pi_{W_{L}} \widetilde{\psi}\right\|_{L^{2}}$.

Finally, for the stochastic term $I$ we have
$\left|I-\frac{1}{\varepsilon}\left\langle D G_{f_{0}}\left[f_{0} \tau\right], \mathbb{W}\right\rangle_{L^{2}}\right| \lesssim \varepsilon^{-1}\left|\left\langle D G_{f_{0}}\left[f-f_{\tau}-f \tau\right], \mathbb{W}\right\rangle\right|+\varepsilon^{-1}\left|\left\langle D G_{f_{0}}\left[\left(f_{0}-f\right) \tau\right], \mathbb{W}\right\rangle\right|$
and we apply arguments as above (71) to the Gaussian processes $\left\{\left\langle W, w_{f, \psi}\right\rangle_{L^{2}}\right.$ : $\left.f=e^{\varphi} \in D_{\varepsilon}, \psi \in \mathcal{C}(b)\right\}$, now with $w_{f, \psi}$ equal to either $D G_{f_{0}}\left[f-f_{\tau}-f \tau\right]$ or $D G_{f_{0}}\left[\left(f_{0}-f\right) \tau\right]$ : In both cases by $L^{2}$-continuity of $D G_{f_{0}}$ (see (23)) and the properties of $f_{\tau}, \tau$ we can bound

$$
\left\|w_{f, \psi}-w_{f^{\prime}, \psi^{\prime}}\right\|_{L^{2}} \lesssim\left\|\varphi-\varphi^{\prime}\right\|_{L^{2}}+\left\|\Pi_{W_{L}}\left(\tilde{\psi}-\widetilde{\psi}^{\prime}\right)\right\|_{L^{2}}
$$

so that using (92) as after (71) gives, for some $\eta^{\prime}>0$,

$$
\begin{aligned}
& \varepsilon^{-1} E \sup _{f \in D_{\varepsilon}, \psi \in \mathcal{C}(b)}\left|\left\langle D G_{f_{0}}\left[\left(f_{0}-f\right) \tau\right], \mathbb{W}\right\rangle\right| \\
& \quad \lesssim \varepsilon^{-1} 2^{J d / 2} \sup _{f \in D_{\varepsilon}, \psi \in \mathcal{C}(b)}\left\|D G_{f_{0}}\left[\left(f_{0}-f\right) \tau\right]\right\|_{L^{2}} \sqrt{\log (1 / \varepsilon)} \\
& \quad \lesssim \varepsilon^{-1} 2^{J d / 2} \sup _{f \in D_{\varepsilon}, \psi \in \mathcal{C}(b)}\left(\delta_{\varepsilon}\left\|\varphi-\varphi_{0}\right\|_{\infty}+\varepsilon\left\|\Pi_{W_{L}} \widetilde{\psi}\right\|_{\infty}\right)\left\|f-f_{0}\right\|_{L^{2}} \sqrt{\log (1 / \varepsilon)} \\
& \quad \lesssim \log ^{\eta^{\prime}}(1 / \varepsilon)\left(\varepsilon^{(-2 s-4-2 d+2 d+2 s-d-\xi+2 s) /(2 s+4+d)}+\varepsilon^{(-d-4+d+2 s) /(2 s+4+d)}\right)=o(1)
\end{aligned}
$$

since $s>2+d / 2+\xi / 2$, and likewise

$$
\begin{aligned}
& \varepsilon^{-1} E \sup _{f \in D_{\varepsilon}, \psi \in \mathcal{C}(b)}\left|\left\langle D G_{f_{0}}\left[f-f_{\tau}-f \tau\right], \mathbb{W}\right\rangle\right| \\
& \quad \lesssim \varepsilon^{-1} 2^{J d / 2} \sup _{f \in D_{\varepsilon}, \psi \in \mathcal{C}(b)}\left\|D G_{f_{0}}\left[f-f_{\tau}-f \tau\right]\right\|_{L^{2}} \sqrt{\log (1 / \varepsilon)} \\
& \quad \leq \varepsilon^{-1} 2^{J d / 2} \sup _{f \in D_{\varepsilon}, \psi \in \mathcal{C}(b)}\|\tau\|_{\infty}\|\tau\|_{L^{2}} \sqrt{\log (1 / \varepsilon)} \\
& \quad \leq \varepsilon^{-1} \sqrt{\log (1 / \varepsilon)} 2^{J d / 2} \\
& \sup _{f \in D_{\varepsilon}, \psi \in \mathcal{C}(b)}\left(\delta_{\varepsilon}\left\|\varphi-\varphi_{0}\right\|_{\infty}+\varepsilon\left\|\Pi_{W_{L}} \widetilde{\psi}\right\|_{\infty}\right)\left(\delta_{\varepsilon}\left\|\varphi-\varphi_{0}\right\|_{L^{2}}+\varepsilon\left\|\Pi_{W_{L}} \widetilde{\psi}\right\|_{L^{2}}\right)
\end{aligned}
$$

is $o(1)$. So overall (72) (and then (68)) becomes

$$
\begin{align*}
\ell(f)-\ell\left(f_{\tau}\right)= & \frac{1}{\varepsilon}\left\langle D G_{f_{0}}\left[f_{0} \tau\right], \mathbb{W}\right\rangle_{L^{2}}+\frac{1}{2 \varepsilon^{2}}\left\|D G_{f_{0}}\left[f_{0} \tau\right]\right\|_{L^{2}}^{2} \\
& +\frac{1}{\varepsilon^{2}}\left\langle D G_{f_{0}}\left[f-f_{0}\right], D G_{f_{0}}\left(f_{0} \tau\right)\right\rangle_{L^{2}}+\bar{Z}_{\varepsilon} \tag{73}
\end{align*}
$$

where $\bar{Z}_{\varepsilon}=o_{P_{f_{0}}^{Y}}(1)$ uniformly in $f \in D_{\varepsilon}$ and $\psi \in \mathcal{C}(b)$, for any $\gamma>0$.
C: Completion of the expansion (critical terms). Now recalling (59) the right hand side of (73) can further be rewritten as

$$
\begin{align*}
t\left\langle D G_{f_{0}}\left[f_{0} \Pi_{W_{L}} \widetilde{\psi}\right], \mathbb{W}\right\rangle_{L^{2}} & +\frac{t^{2}}{2}\left\|D G_{f_{0}}\left[f_{0} \Pi_{W_{L}} \widetilde{\psi}\right]\right\|_{L^{2}}^{2} \\
& +\frac{t}{\varepsilon}\left\langle D G_{f_{0}}\left[f-f_{0}\right], D G_{f_{0}}\left(f_{0} \Pi_{W_{L}} \widetilde{\psi}\right)\right\rangle_{L^{2}}+\bar{Z}_{\varepsilon}+\bar{Z}_{\varepsilon}^{\prime} \tag{74}
\end{align*}
$$

where $\bar{Z}_{\varepsilon}^{\prime}=o_{P_{f_{0}}^{Y}}(1)$ uniformly in $f \in D_{\varepsilon}$ and $\psi \in \mathcal{C}(b)$, since

$$
\begin{aligned}
\left|\bar{Z}_{\varepsilon}^{\prime}\right| \leq & \frac{\delta_{\varepsilon}}{\varepsilon}\left|\left\langle D G_{f_{0}}\left[f_{0} \Pi_{W_{L}}\left(\varphi-\varphi_{0}\right)\right], \mathbb{W}\right\rangle_{L^{2}}\right|+\frac{\delta_{\varepsilon}^{2}}{2 \varepsilon^{2}}\left\|D G_{f_{0}}\left[f_{0} \Pi_{W_{L}}\left(\varphi-\varphi_{0}\right)\right]\right\|_{L^{2}}^{2} \\
& +\frac{\delta_{\varepsilon}}{\varepsilon^{2}}\left|\left\langle D G_{f_{0}}\left[f-f_{0}\right], D G_{f_{0}}\left(f_{0} \Pi_{W_{L}}\left(\varphi-\varphi_{0}\right)\right)\right\rangle_{L^{2}}\right| \\
= & A+B+C .
\end{aligned}
$$

For $A$ we need to bound the supremum of the Gaussian process $\left\{\left\langle\mathbb{W}, q_{\varphi}\right\rangle_{L^{2}}: e^{\varphi} \in D_{\varepsilon}\right\}$ where $q_{\varphi}=D G_{f_{0}}\left[f_{0} \Pi_{W_{L}}\left(\varphi-\varphi_{0}\right)\right]$ : The $L^{2}$-continuity of $D G_{f_{0}}$ from (23) implies $\left\|D G_{f_{0}}\left[f_{0} \Pi_{W_{L}}\left(\varphi-\varphi^{\prime}\right)\right]\right\|_{L^{2}} \lesssim\left\|f_{0}\right\|_{\infty}\left\|\varphi-\varphi^{\prime}\right\|_{L^{2}}$ and thus arguing as after (71) we obtain

$$
\begin{aligned}
& E \sup _{\varphi: e^{\varphi} \in D_{\varepsilon}} \mid\left\langle D G_{f_{0}}\left[f_{0} \Pi_{W_{L}}\left(\varphi-\varphi_{0}\right), \mathbb{W}\right\rangle_{L^{2}}\right| \\
& \lesssim \sup _{\varphi: e^{\varphi} \in D_{\varepsilon}}\left\|D G_{f_{0}}\left[f_{0} \Pi_{W_{L}}\left(\varphi-\varphi_{0}\right)\right]\right\|_{L^{2} 2^{J d / 2}}^{\log (1 / \varepsilon)},
\end{aligned}
$$

which combined with previous bounds and Proposition 4 gives

$$
\begin{aligned}
A & =O_{P_{f_{0}}^{Y}}\left(\varepsilon^{-1} \delta_{\varepsilon}\left\|\varphi-\varphi_{0}\right\|_{\left(H_{0}^{2}\right)^{*}} 2^{J d / 2} \sqrt{\log (1 / \varepsilon)}\right) \\
& =O_{P_{f_{0}}^{Y}}\left(\varepsilon^{\left(-2 s-4-d+2 d+\gamma_{1}+2 s+4-d\right) /(2 s+4+d)} \log ^{\eta+1 / 2}(1 / \varepsilon)\right)=o_{P}(1)
\end{aligned}
$$

uniformly in $f \in D_{\varepsilon}$ for every $\gamma_{1}>0$. Likewise $C=o_{P}(1)$ since for every $\gamma_{1}>0$,

$$
\begin{aligned}
& O_{P}\left(\delta_{\varepsilon} \varepsilon^{-2}\left\|f-f_{0}\right\|_{\left(H_{0}^{2}\right)^{*}}\left\|\varphi-\varphi_{0}\right\|_{\left(H_{0}^{2}\right)^{*}}\right) \\
&=O_{P}\left(\varepsilon^{\left(2 d+\gamma_{1}-4 s-8-2 d+4 s+8\right) /(2 s+4+d)} \log ^{2 \eta+1 / 2}(1 / \varepsilon)\right)
\end{aligned}
$$

and $B=o_{P}(1)$ follows too as $B$ is stochastically smaller than $C$ thanks to the extra $\delta_{\varepsilon}$ factor.

The last step is to let $L \rightarrow \infty$ in the three main terms in (74). We recall from Step II that $\widetilde{\psi} \in C_{K}(\mathcal{O})$. For $l$ large enough the boundary corrected wavelets do not intersect the support $K$ of $\widetilde{\psi}$ and the dual norm of $\left(H^{2}\right)^{*}$ is thus estimated by the dual norm of $\left(H^{2}(\mathbb{R})\right)^{*}$, just as in (48); we thus have, under assumption (i) of Theorem 17 with $L=J$, and by (11), (57),

$$
\begin{align*}
& \left\|D G_{f_{0}}\left[f_{0}\left(\Pi_{W_{J}} \tilde{\psi}-\widetilde{\psi}\right)\right]\right\|_{L^{2}}^{2} \lesssim\left\|f_{0}\right\|_{C^{2}}^{2}\left\|\Pi_{W_{J}} \tilde{\psi}-\widetilde{\psi}\right\|_{\left(H_{0}^{2}\right)^{*}}^{2} \\
& \quad=\sum_{l>J} 2^{-4 l} \sum_{r}\left\langle\tilde{\psi}, \Phi_{l, r}\right\rangle_{L^{2}\left(\mathbb{R}^{d}\right)}^{2} \leq \sum_{l>J} 2^{-l(d+2 \gamma)} \leq 2^{-J(d+2 \gamma)}=o(1) . \tag{75}
\end{align*}
$$

By (25) we see that $\left\{D G_{f_{0}}\left[f_{0} \tilde{\psi}\right]=-V_{f_{0}}\left[S_{f_{0}}\left[S_{f_{0}}\left[\psi / u_{f_{0}}\right]\right]\right]=-S_{f_{0}}\left[\psi / u_{f_{0}}\right]: \psi \in \mathcal{C}(b)\right\}$ is bounded in $C_{c}^{d / 2+\gamma}(\mathcal{O})$ and has covering numbers bounded by Proposition 27(C). Moreover by $L^{2}$-continuity of $D{\underset{\sim}{f}}_{f_{0}}$ and arguing as before (71), we find that the covering numbers of $\left\{D G_{f_{0}}\left[f_{0} \Pi_{W_{J}} \widetilde{\psi}\right]: \psi \in \mathcal{C}(b)\right\}$ are bounded by those of a ball in a $c_{0} 2^{J d}$-dimensional space. The class of differences of such functions then has covering numbers bounded by the product of the covering numbers of each class, and using (91) and also (75) to bound $\sigma$ we have, for any $\gamma>0$,

$$
\begin{aligned}
& E \sup _{\psi \in \mathcal{C}(b)}\left|\left\langle D G_{f_{0}}\left[f_{0}\left(\Pi_{W_{J}} \tilde{\psi}-\widetilde{\psi}\right)\right], \mathbb{W}\right\rangle_{L^{2}}\right| \\
& \lesssim 2^{J d / 2} \int_{0}^{\sigma} \sqrt{\log (A / \eta)} d \eta+\int_{0}^{\sigma}\left(A^{\prime} / \gamma\right)^{(d / 2) /(\gamma+d / 2)} d \gamma \\
& \lesssim 2^{J d / 2} 2^{-J(d / 2+\gamma)} \sqrt{\log (1 / \varepsilon)}+o(\sigma)=o(1)
\end{aligned}
$$

Finally, using again the previous estimate for $\left\|\Pi_{W_{J}} \tilde{\psi}-\widetilde{\psi}\right\|_{\left(H_{0}^{2}\right)^{*}}^{2}$ we obtain, for every $\gamma>0$,

$$
\begin{aligned}
\left.\frac{t}{\varepsilon} \right\rvert\,\left\langle D G_{f_{0}}\left[f-f_{0}\right], D G_{f_{0}}\left(f_{0}( \right.\right. & \left.\left.\Pi_{W_{J}} \tilde{\psi}-\widetilde{\psi}\right)\right\rangle_{L^{2}} \mid \\
& \lesssim \varepsilon^{-1}\left\|f_{0}\right\|_{C^{2}}\left\|f-f_{0}\right\|_{\left(H_{0}^{2}\right)^{*}}\left\|\Pi_{W_{J}}(\tilde{\psi}-\widetilde{\psi})\right\|_{\left(H_{0}^{2}\right)^{*}} \\
& =O\left(\varepsilon^{-1} \varepsilon^{(2 s+4+d+2 \gamma) /(2 s+4+d)} \log ^{\eta}\left(\varepsilon^{-1}\right)\right)=o(1)
\end{aligned}
$$

finishing the proof under assumption (i). Under assumption (ii) the proof of the last step proceeds analogously, with the previous estimates for $\left\|\Pi_{W_{J}} \widetilde{\psi}-\widetilde{\psi}\right\|_{\left(H_{0}^{2}\right)^{*}}^{2}$ replaced by the hypothesis featuring in Theorem 17 (ii).

To conclude, since $f_{0} \widetilde{\psi}=-\widetilde{\Psi}$, the LAN expansion (74) becomes

$$
\begin{align*}
\ell(f)-\ell\left(f_{\tau}\right)= & -t\left\langle D G_{f_{0}}[\widetilde{\Psi}], \mathbb{W}\right\rangle_{L^{2}} \\
& +\frac{t^{2}}{2}\left\|D G_{f_{0}}[\widetilde{\Psi}]\right\|_{L^{2}}^{2}-\frac{t}{\varepsilon}\left\langle D G_{f_{0}}\left[f-f_{0}\right], D G_{f_{0}}[\widetilde{\Psi}]\right\rangle_{L^{2}}+Z_{\varepsilon}^{\prime \prime} \tag{76}
\end{align*}
$$

for some $Z_{\varepsilon}^{\prime \prime}=o_{P_{f_{0}}^{Y}}(1)$ uniformly in $f \in D_{\varepsilon}$ and $\psi \in \mathcal{C}(b)$. Now using (27) we have

$$
\begin{equation*}
\left\langle D G_{f_{0}}\left[f-f_{0}\right], D G_{f_{0}}[\widetilde{\Psi}]\right\rangle_{L^{2}}=\left\langle f-f_{0}, \psi\right\rangle_{L^{2}} \tag{77}
\end{equation*}
$$

and we can insert (76) into (67) to obtain

$$
\begin{align*}
E^{\Pi^{D_{\varepsilon}}}\left[\left.e^{\frac{t}{\varepsilon}\left\langle f-f_{0}, \psi\right\rangle_{L^{2}}} \right\rvert\, Y\right]= & e^{\left.-t\left\langle D G_{f_{0}}[\widetilde{\Psi}], \mathbb{W}\right\rangle_{L^{2}}+t^{2} \| D G_{f_{0}} \tilde{\Psi}\right] \|_{L^{2}}^{2}} \\
& \cdot \frac{\int_{D_{\varepsilon}} e^{\ell\left(f_{\tau}\right)} d \Pi(f)}{\int_{D_{\varepsilon}} e^{\ell(f)} d \Pi(f)} \cdot e^{R_{\varepsilon}}, \tag{78}
\end{align*}
$$

completing the proof.
Step IV: Change of variables. We now analyse the ratio of high-dimensional integrals appearing on the r.h.s. of (78).

Proposition 18. For every $\gamma>0$ and $\gamma_{1}$ (appearing in the definition of $\delta_{\varepsilon}$ in (60)) satisfying $0<\gamma_{1}<2 \gamma$ we have, for the perturbation $f_{\tau}$ from (58) and any integer $L \leq J$, as $\varepsilon \rightarrow 0$,

$$
\begin{equation*}
\frac{\int_{D_{\varepsilon}} e^{\ell\left(f_{\tau}\right)} d \Pi(f)}{\int_{D_{\varepsilon}} e^{\ell(f)} d \Pi(f)}=(1+o(1)) \cdot \frac{\int_{D_{\varepsilon, \tau}} e^{\ell(f)} d \Pi(f)}{\int_{D_{\varepsilon}} e^{\ell(f)} d \Pi(f)}=O_{P_{f_{0}}^{Y}}(1) \tag{79}
\end{equation*}
$$

uniformly in $|t| \leq T$ for any $T$ and in $\psi \in \mathcal{C}(b)$ from Theorem 17, where

$$
D_{\varepsilon, \tau}=\left\{g=f_{\tau}=e^{\varphi+\tau}: f=e^{\varphi} \in D_{\varepsilon}\right\} \subseteq \operatorname{supp} П
$$

Proof. Write $\bar{\ell}(\varphi)=\ell(f(\varphi))$ when viewing the function $\ell(f), f=e^{\varphi}$, as a function of $\varphi$. Thus $\ell\left(f_{\tau}\right)=\bar{\ell}(\varphi+\tau)$, and so

$$
\begin{equation*}
\frac{\int_{D_{\varepsilon}} e^{\ell\left(f_{\tau}\right)} d \Pi(f)}{\int_{D_{\varepsilon}} e^{\ell(f)} d \Pi(f)}=\frac{\int_{D_{\varepsilon}} e^{\bar{\ell}(\varphi+\tau)} d \pi(\varphi)}{\int_{D_{\varepsilon}} e^{\bar{\ell}(\varphi)} d \pi(\varphi)} \tag{80}
\end{equation*}
$$

where $\pi$ is the law of $\varphi=\log f, f \sim \Pi$, and where and in what follows, by abuse of notation, the integration domain $D_{\varepsilon}$ is viewed interchangeably as a set of $f$ 's or of $\varphi$ 's. By definition of the prior $\Pi$ and (15) the $d \pi$-integrals are product integrals over hyperellipsoids

$$
\operatorname{supp} \pi=\prod_{l \leq J} I_{l}^{N_{l}} \subseteq \mathbb{R}^{\bar{c}_{0} 2^{J d}}, \quad N_{l} \leq c_{0} 2^{l d}, \bar{c}_{0}<\infty
$$

with each marginal distribution being the uniform distribution on the intervals

$$
I_{l}=\left[-B 2^{-l(s+d / 2)} \bar{l}^{-2}, B 2^{-l(s+d / 2)} \bar{l}^{-2}\right] .
$$

By definition of $\tau$ in (59) we have the orthogonal decomposition
$\varphi+\tau=\Pi_{V_{J} \backslash W_{L}} \varphi+\left(1-\delta_{\varepsilon}\right) \Pi_{W_{L}} \varphi+\delta_{\varepsilon} p(\widetilde{\psi}), \quad$ where $\quad p(\widetilde{\psi})=\frac{t \varepsilon}{\delta_{\varepsilon}} \Pi_{W_{L}}(\widetilde{\psi})+\Pi_{W_{L}} \varphi_{0}$,
thus the perturbation $\tau$ affects only the subspaces $W_{L}$ in the support of the prior, and in these subspaces it equals a convex combination of points $\varphi$ in the support with the point $p(\widetilde{\psi})$ which, as the next lemma shows, is also contained in supp $\pi$.

Lemma 19. Let $L \leq J$ and $\psi \in \mathcal{C}(b),|t| \leq T$. Then for every $\gamma>0$ and $0<\gamma_{1}<2 \gamma$ we have $p(\widetilde{\psi}) \in \operatorname{supp} \pi$ for all $\varepsilon$ small enough.

Proof. By Condition 7 we know that $\varphi_{0}$ is an interior point satisfying

$$
\left|\left\langle\varphi_{0}, \Phi_{l, r}^{\mathcal{O}}\right\rangle\right| \leq(B-\epsilon) 2^{-l(s+d / 2)} \bar{l}^{-2}
$$

for some $\epsilon>0$ and all $l, r$. Thus to show that $p(\widetilde{\psi}) \in \operatorname{supp} \pi$ it suffices to verify that

$$
\frac{t \varepsilon}{\delta_{\varepsilon}}\left|\left\langle\tilde{\psi}, \Phi_{l, r}^{\mathcal{O}}\right\rangle\right| \leq \epsilon 2^{-l(s+d / 2)} \bar{l}^{-2}
$$

for all $l \leq J, r$. Using (57) we check that for our choice of $\delta_{\varepsilon}, J$ and $\gamma>\gamma_{1} / 2$,

$$
\begin{aligned}
\max _{l \leq J, r} 2^{l(s+d / 2)} \bar{l}^{2} \frac{\varepsilon}{\delta_{\varepsilon}}\left|\left\langle\tilde{\psi}, \Phi_{l, r}^{\mathcal{O}}\right\rangle\right| & \leq \max _{l \leq J, r} 2^{l(s+2-d / 2-\gamma)} \bar{l}^{2} \varepsilon \varepsilon^{-\left(2 d+\gamma_{1}\right) /(2 s+4+d)} \\
& \lesssim \varepsilon^{\left(-2 s-4+d+2 s+4-d+2 \gamma-\gamma_{1}\right) /(2 s+4+d)} \log ^{2}(1 / \varepsilon)<\epsilon
\end{aligned}
$$

for $\varepsilon$ small enough.
We deduce from the lemma and (81) that by convexity of supp $\pi$ we must have $\varphi+\tau \in$ supp $\pi$ too. If $\pi_{\tau}$ denotes the law of $\varphi+\tau$ then obviously on $V_{J} \backslash W_{L}$ we have $\pi_{\tau}=\pi$ and for the marginal coordinates of the subspace $W_{L}$ the densities of the law $\pi_{\tau}$ with respect to the law $\pi$ equal the constant $\left(1-\delta_{\varepsilon}\right)^{-1}$ on a strict subinterval $\widetilde{I}_{l} \subseteq I_{l}$ (they are the densities of $\left(1-\delta_{\varepsilon}\right) U+\delta_{\varepsilon} p$ for $U$ a uniform random variable and $p$ a constant). The density of the product integrals is then also constant and on the support of $\pi_{\tau}$ it is given by

$$
\begin{equation*}
\frac{d \pi_{\tau}}{d \pi}=\prod_{l \leq L}\left(\frac{1}{1-\delta_{\varepsilon}}\right)^{N_{0}^{l}}=\left(\frac{1}{1-\delta_{\varepsilon}}\right)^{\bar{c} 2^{J d}}=1+o(1), \quad \bar{c}>0, \tag{82}
\end{equation*}
$$

where $N_{0}^{l} \leq N_{l}$ is the dimension of $W_{l} \backslash W_{l-1}$ and where we have used the definition of $\delta_{\varepsilon}$ from (60) in

$$
2^{J d} \delta_{\varepsilon}=\varepsilon^{\left(-2 d+2 d+\gamma_{1}\right) /(2 s+4+d)} \rightarrow 0
$$

as $\varepsilon \rightarrow 0$ for any $\gamma_{1}>0$. We can thus write the 1.h.s. of (79) as

$$
\frac{\int_{D_{\varepsilon, \tau}} e^{\ell(f(\varphi))} \frac{d \pi_{\tau}(\varphi)}{d \pi(\varphi)} d \pi(\varphi)}{\int_{D_{\varepsilon}} e^{\ell(f(\varphi))} d \pi(\varphi)}=(1+o(1)) \cdot \frac{\int_{D_{\varepsilon, \tau}} e^{\ell(f)} d \Pi(f)}{\int_{D_{\varepsilon}} e^{\ell(f)} d \Pi(f)}
$$

where

$$
D_{\varepsilon, \tau}=\left\{g=f_{\tau}=e^{\varphi+\tau}: f=e^{\varphi} \in D_{\varepsilon}\right\} \subseteq \operatorname{supp} \Pi,
$$

completing the proof of the first identity in the proposition. Renormalising by $\int_{\mathcal{F}} e^{\ell(f)} d \Pi(f)$ we find that the last ratio equals

$$
\frac{\Pi\left(D_{\varepsilon, \tau} \mid Y\right)}{\Pi\left(D_{\varepsilon} \mid Y\right)} \leq \frac{1}{\Pi\left(D_{\varepsilon} \mid Y\right)}=O_{P_{f_{0}}^{Y}}(1)
$$

using also $\Pi\left(D_{\varepsilon} \mid Y\right) \rightarrow 1$ in $P_{f_{0}}^{Y}$-probability (Step I above), completing the proof.
Step V: Convergence of finite-dimensional distributions. For any $\psi \in C_{K}^{4}(\mathcal{O})$ and $\mathbb{W}=Y-u_{f_{0}}$ define the random variables

$$
\begin{equation*}
\tilde{f}(\psi)=\langle\tilde{f}, \psi\rangle_{L^{2}}=\left\langle f_{0}, \psi\right\rangle_{L^{2}}-\varepsilon\left\langle D G_{f_{0}}[\tilde{\Psi}], \mathbb{W}\right\rangle_{L^{2}} \tag{83}
\end{equation*}
$$

which in view of Proposition 6 and the results from Section 4.2 define the random variable

$$
\begin{equation*}
\tilde{f}=\widetilde{f}(Y) \stackrel{\mathcal{L}}{=} f_{0}+\varepsilon \widetilde{X}, \quad \widetilde{X} \sim \mathcal{N}_{f_{0}} \tag{84}
\end{equation*}
$$

in $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$. Let further $\widetilde{\Pi}(\cdot \mid Y)$ be the law of $\varepsilon^{-1}(f-\tilde{f}) \mid Y$ in $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$ conditional on $Y$.

Proposition 20. For any finite vector $\left(\psi_{1}, \ldots, \psi_{k}\right), k \in \mathbb{N}$, of functions $\psi_{i} \in C_{K}^{S}(\mathcal{O})$, $S \in \mathbb{N}$ as in Condition 7, and any random variable $Z$ of law $\mu$ in $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$, let $\mu_{k}$ be the distribution of the random vector $\left(Z\left(\psi_{1}\right), \ldots, Z\left(\psi_{k}\right)\right)$ in $\mathbb{R}^{k}$. Then for $\beta=\beta_{\mathbb{R}^{k}}$ the bounded-Lipschitz distance for weak convergence of probability measures on $\mathbb{R}^{k}$ and every fixed $k \in \mathbb{N}$, we have

$$
\beta\left(\widetilde{\Pi}(\cdot \mid Y)_{k},\left(\mathcal{N}_{f_{0}}\right)_{k}\right) \xrightarrow{P_{f_{0}}^{Y}} 0 \quad \text { as } \varepsilon \rightarrow 0 .
$$

Proof. In view of (55) and since the total variation distance dominates any metric for weak convergence, it suffices to prove the result for $\Pi(\cdot \mid Y)$ replaced by $\Pi^{D_{\varepsilon}}(\cdot \mid Y)$. We first need

Lemma 21. Fix $\psi \in C_{c}^{S}(\mathcal{O})$. Let $\tau$ from (59) be the perturbation in $W_{L}$ associated to $\psi$ with L satisfying $2^{L(s+d / 2)} L^{2+\eta} \varepsilon^{2 s /(2 s+4+d)}=o(1)$, and let $D_{\varepsilon, \tau}$ be as in Proposition 18. Then, as $\varepsilon \rightarrow 0$ and in $P_{f_{0}}^{Y}$-probability,

$$
\Pi\left(\mathcal{D}_{\varepsilon, \tau} \mid Y\right) \rightarrow 1 \quad \text { and } \quad \frac{\int_{D_{\varepsilon, \tau}} e^{\ell(f)} d \Pi(f)}{\int_{D_{\varepsilon}} e^{\ell(f)} d \Pi(f)} \rightarrow 1
$$

Proof. Recall that by our choice of $M_{0}$ in the paragraph above (55) we have $\Pi\left(\mathcal{D}_{\varepsilon}^{M_{0} /(8 W)} \mid Y\right) \rightarrow 1$ in $P_{f_{0}}^{Y}$-probability. We show that $\mathcal{D}_{\varepsilon, \tau}=\mathcal{D}_{\varepsilon, M, \tau}$ contains the set $\mathcal{D}_{\varepsilon}^{M_{0} /(8 W)}$ for all $\varepsilon$ small enough so that the first limit of the lemma will follow. By definition of $\mathcal{D}_{\varepsilon, M, \tau}$ and (81) we need to show that for every $\bar{f} \in \mathcal{D}_{\varepsilon}^{M_{0} /(8 W)}$ the function $\bar{\varphi}=\log \bar{f}$ equals

$$
\varphi^{*}+\tau=\Pi_{V_{J} \backslash W_{L}} \varphi^{*}+(1-\delta) \Pi_{W_{L}} \varphi^{*}+\delta \Pi_{W_{L}}\left(\varphi_{0}\right)+\varepsilon t \Pi_{W_{L}}(\tilde{\psi})
$$

for some $\varphi^{*}$ such that $e^{\varphi^{*}} \in \mathcal{D}_{\varepsilon}^{M}$. Define

$$
\begin{equation*}
\varphi^{*}=\Pi_{V_{J} \backslash W_{L}} \bar{\varphi}+\Pi_{W_{L}}\left(\varphi_{0}\right)+\frac{\Pi_{W_{L}} \bar{\varphi}-\Pi_{W_{L}}\left(\varphi_{0}\right)-\varepsilon t \Pi_{W_{L}}(\tilde{\psi})}{1-\delta_{\varepsilon}} \tag{85}
\end{equation*}
$$

Then by (81) we have

$$
\varphi^{*}+\tau=\Pi_{V_{J} \backslash W_{L}} \bar{\varphi}+\Pi_{W_{L}} \bar{\varphi}=\bar{\varphi}
$$

so it remains to show that $e^{\varphi^{*}}$ is indeed in $\mathcal{D}_{\varepsilon}^{M}$. To achieve this we need to show that $\varphi^{*} \in \operatorname{supp} \pi$ and the $\|\cdot\|^{(i)}$-norm inequalities in (53) are satisfied. Since $\bar{\varphi} \in V_{J}$, we have

$$
\left\|\varphi^{*}-\varphi_{0}\right\|^{(i)} \leq\left\|\Pi_{W_{L}}\left(\varphi^{*}-\varphi_{0}\right)\right\|^{(i)}+\left\|\left(\operatorname{Id}-\Pi_{W_{L}}\right)\left(\bar{\varphi}-\varphi_{0}\right)\right\|^{(i)}
$$

and the second term is less than $\left[M_{0} /(8 W)+\left(M_{0} / 8\right)\right] r(i, \varepsilon) \leq M_{0} / 4$ since $e^{\bar{\varphi}} \in$ $\mathcal{D}_{\varepsilon}^{M_{0} /(8 W)}$ and using (61). For the first we use (61) once more and have, for $\delta_{\varepsilon}$ and $\varepsilon$ small enough,

$$
\begin{aligned}
\left\|\Pi_{W_{L}}\left(\varphi^{*}-\varphi_{0}\right)\right\|^{(i)} & \leq W\left(1-\delta_{\varepsilon}\right)^{-1}\left[\left\|\bar{\varphi}-\varphi_{0}\right\|^{(i)}+t \varepsilon\left\|\Pi_{W_{L}}(\tilde{\psi})\right\|^{(i)}\right] \\
& \leq\left(M_{0} / 4\right) r(i, \varepsilon)+o(r(i, \varepsilon)) \leq\left(M_{0} / 2\right) r(i, \varepsilon),
\end{aligned}
$$

using $e^{\bar{\varphi}} \in \mathcal{D}_{\varepsilon}^{M_{0} /(8 W)}$ and $\varepsilon\|\tilde{\psi}\|^{(i)}=O(\varepsilon)$ for any fixed $\psi \in C_{c}^{S}(\mathcal{O}), S>4$, by definition of $\widetilde{\psi}$. Finally, we also have $\varphi^{*} \in \operatorname{supp} \Pi$ : this is clear for $l>L$ as then $\varphi^{*}=\bar{\varphi}$, and also for $l \leq L$ since then, using Condition 7 and $\widetilde{\psi} \in L^{2}$ in view of $S>4$, we find that

$$
\begin{aligned}
& 2^{l(s+d / 2)} \bar{l}^{2}\left|\left\langle\varphi^{*}, \Phi_{l, r}^{\mathcal{O}}\right\rangle\right| \\
& \quad \leq 2^{l(s+d / 2)} \bar{l}^{2}\left|\left\langle\varphi_{0}, \Phi_{l, r}^{\mathcal{O}}\right\rangle_{L^{2}}\right|+\left(1-\delta_{\varepsilon}\right)^{-1} 2^{l(s+d / 2)} \bar{l}^{2}\left[\left|\left\langle\bar{\varphi}-\varphi_{0}, \Phi_{l, r}^{\mathcal{O}}\right\rangle_{L^{2}}\right|+\varepsilon t\left|\left\langle\tilde{\psi}, \Phi_{l, r}^{\mathcal{O}}\right\rangle_{L^{2}}\right|\right] \\
& \leq(B-\epsilon)+(4 / 3) 2^{L(s+d / 2)} L^{2}\left(\left\|\bar{\varphi}-\varphi_{0}\right\|_{L^{2}}+O(\varepsilon)\right) \leq B
\end{aligned}
$$

for $\varepsilon$ small enough, using the fact that $\bar{\varphi} \in \mathcal{D}_{\varepsilon}^{M_{0} /(8 W)}$ implies $\left\|\bar{\varphi}-\varphi_{0}\right\|_{L^{2}}=$ $O\left(\varepsilon^{2 s /(2 s+4+d)} \log ^{\eta}(1 / \varepsilon)\right)$ and the hypothesis on $L$. This proves the first claim of the lemma. For the second we can renormalise both the numerator and denominator by $\int_{\mathcal{F}} e^{\ell(f)} d \Pi(f)$ to obtain

$$
\frac{\int_{D_{\varepsilon, \tau}} e^{\ell(f)} d \Pi(f)}{\int_{D_{\varepsilon}} e^{\ell(f)} d \Pi(f)}=\frac{\Pi\left(D_{\varepsilon, \tau} \mid Y\right)}{\Pi\left(D_{\varepsilon} \mid Y\right)} \xrightarrow{P_{f_{0}}^{Y}} 1 .
$$

We are now ready to combine this lemma for the choice

$$
2^{L} \simeq \varepsilon^{-(2 s /(s+d / 2)) /(2 s+4+d)} / \log ^{\gamma_{0}}(1 / \varepsilon), \quad L \leq J
$$

$\gamma_{0}$ a large enough constant, with Theorem 17(ii) for fixed $\psi \in C_{K}^{S}(\mathcal{O})$ : Since $\psi \in C_{K}^{S}(\mathcal{O})$ we have $\widetilde{\psi} \in C_{K}^{S-4}(\mathcal{O})$, and since for large $l$ the support $K$ of $\widetilde{\psi}$ does not overlap the support of the boundary corrected wavelets, we have, as in (48),

$$
\begin{aligned}
\left\|\Pi_{W_{L}}(\widetilde{\psi})-\widetilde{\psi}\right\|_{\left(H_{0}^{2}\right)^{*}} & \lesssim 2^{-L(S-2)} \\
& \lesssim \varepsilon^{[(S-2)(2 s /(s+d / 2)] /(2 s+4+d)} \log ^{\gamma_{0}^{\prime}}(1 / \varepsilon)=O\left(\varepsilon^{\bar{\gamma} /(2 s+4+d)}\right)
\end{aligned}
$$

for some $\bar{\gamma}>d$ by assumption on $S$, $s$. Thus Theorem 17(ii) and the previous lemma imply, by definition of $\widetilde{f}$ and for $r_{\varepsilon}=o_{P_{f_{0}}^{Y}}(1)$ that

$$
E^{D_{\varepsilon}}\left[e^{t \varepsilon^{-1}\langle f-\tilde{f}, \psi\rangle} \mid Y\right]=e^{r_{\varepsilon}} e^{\frac{t^{2}}{2}\left\|D G_{f_{0}}(\widetilde{\Psi}]\right\|_{L^{2}}^{2}}, \quad \forall t \in \mathbb{R}
$$

for all $\psi \in C_{K}^{S}(\mathcal{O})$. Applying this to any linear combination $\psi=\sum_{i \leq k} a_{i} \psi_{i}, a_{i} \in \mathbb{R}$, which still defines an element of $C_{K}^{S}(\mathcal{O})$, we conclude that the convergence of the Laplace transform and the Cramer-Wold device (Section 7.6) gives the desired weak convergence in probability in the $\beta_{\mathbb{R}^{k}}$-metric.
Step VI: Tightness estimates and convergence in function space. We now prove Theorem 8 with $\tilde{f}$ from (84) in place of $\bar{f}$. That $\widetilde{f}$ can be replaced by $\bar{f}$ will constitute the last Step VII. By (55) and since the total variation distance dominates any metric for weak convergence, it suffices to prove Theorem 8 for $\Pi^{D_{\varepsilon}}(\cdot \mid Y)$ replacing $\Pi(\cdot \mid Y)$. Let $f \sim \Pi^{D_{\varepsilon}}(\cdot \mid Y)$ conditionally on $Y$ and consider the stochastic process

$$
\left(X_{1}(\psi)=\varepsilon^{-1}\left(\langle f, \psi\rangle_{L^{2}}-\tilde{f}(\psi)\right): \psi \in C_{K}^{\alpha}(\mathcal{O})\right)
$$

whose law on $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$ we denote by $\widetilde{\Pi}^{D_{\varepsilon}}(\cdot \mid Y)$. Further let $X_{2} \sim \mathcal{N}_{f_{0}}$. For $\lambda \in \mathbb{N}$ large enough to be chosen below and finite-dimensional spaces $W_{\lambda}$ with projection operator $\Pi_{W_{\lambda}}$ as in Step II, define probability measures $\widetilde{\Pi}_{\lambda}^{D_{\varepsilon}}(\cdot \mid Y), \mathcal{N}_{f_{0}, \lambda}$ as the laws of the stochastic processes

$$
P_{\lambda}\left(X_{i}\right) \equiv\left(X_{i}\left(\Pi_{W_{\lambda}} \psi\right): \psi \in C_{K}^{\alpha}(\mathcal{O})\right), \quad i=1,2
$$

respectively, which, as projections, are defined on the same probability space as the $X_{i}$ 's. Then using the triangle inequality for the metric $\beta=\beta_{\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}}$, we obtain

$$
\begin{align*}
\beta\left(\widetilde{\Pi}^{D_{\varepsilon}}(\cdot \mid Y), \mathcal{N}_{f_{0}}\right) & \leq \beta\left(\widetilde{\Pi}_{\lambda}^{D_{\varepsilon}}(\cdot \mid Y), \mathcal{N}_{f_{0}, \lambda}\right)+\beta\left(\widetilde{\Pi}^{D_{\varepsilon}}(\cdot \mid Y), \widetilde{\Pi}_{\lambda}^{D_{\varepsilon}}(\cdot \mid Y)\right)+\beta\left(\mathcal{N}_{f_{0}}, \mathcal{N}_{f_{0}, \lambda}\right) \\
& =\beta\left(\widetilde{\Pi}_{\lambda}^{D_{\varepsilon}}(\cdot \mid Y), \mathcal{N}_{f_{0}, \lambda}\right)+\sum_{i=1}^{2} \sup _{\|F\|_{\text {Lip }} \leq 1}\left|E\left[F\left(X_{i}\right)-F\left(P_{\lambda}\left(X_{i}\right)\right)\right]\right| \\
& \leq \beta_{W_{\lambda}}\left(\widetilde{\Pi}_{\lambda}^{D_{\varepsilon}}(\cdot \mid Y),\left(\mathcal{N}_{f_{0}}\right) \lambda\right)+\sum_{i} E\left\|X_{i}-P_{\lambda}\left(X_{i}\right)\right\|_{\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}} \\
& =A+B+C, \tag{86}
\end{align*}
$$

and we wish to show that the last three terms converge to zero in $P_{f_{0}}^{Y}$-probability.

For $B$, still writing $E=E^{\Pi^{D_{\varepsilon}}}[\cdot \mid Y]$, from Parseval's identity and (13), for 'interior' wavelets $\Phi_{l, r}^{\mathcal{O}}=\Phi_{l, r}$ with $l \geq \lambda$ large enough, and for some $\gamma>0$ such that $\alpha^{\prime}=$ $2+d / 2+\gamma<\alpha-d$, we have

$$
\begin{align*}
E\left\|X_{1}-P_{\lambda}\left(X_{1}\right)\right\|_{\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}} & =E \sup _{\|\psi\|_{C_{K}^{\alpha}(\mathcal{O})} \leq 1} \varepsilon^{-1}\left|\left\langle f-\tilde{f}, \psi-\Pi_{W_{\lambda}} \psi\right\rangle_{L^{2}}\right| \\
& \lesssim E \sum_{\lambda<l, r} 2^{-l(\alpha+d / 2)} \varepsilon^{-1}\left|\left\langle f-\widetilde{f}, \Phi_{l, r}^{\mathcal{O}}\right\rangle_{L^{2}}\right| \\
& =\sum_{\lambda<l, r} 2^{l\left(\alpha^{\prime}-\alpha\right)} E \varepsilon^{-1}\left|\left\langle f-\widetilde{f}, 2^{-l\left(\alpha^{\prime}+d / 2\right)} \Phi_{l, r}^{\mathcal{O}}\right\rangle_{L^{2}}\right| . \tag{87}
\end{align*}
$$

The functions $2^{-l\left(\alpha^{\prime}+d / 2\right)} \Phi_{l, r}^{\mathcal{O}}$ are all contained in the set $\mathcal{C}(b)$ from Theorem 17 with $\gamma=\alpha^{\prime}-2-d / 2$ (cf. after (13)), and that theorem combined with (79) and the results from Section 4.2 now imply that

$$
\begin{equation*}
E^{\Pi^{D_{\varepsilon}}}\left[\left.e^{\frac{t}{\varepsilon}\left\langle f-\tilde{f}, 2^{-l\left(\alpha^{\prime}+d / 2\right)} \Phi_{l, r}^{\mathcal{O}}\right\rangle_{L^{2}}} \right\rvert\, Y\right] \leq r_{\varepsilon} e^{\frac{t^{2}}{2}\left\|S_{f_{0}}\left[2^{-l\left(\alpha^{\prime}+d / 2\right)} \Phi_{l, r}^{\mathcal{O}} / u_{f_{0}}\right]\right\|_{L^{2}}^{2}} \lesssim r_{\varepsilon} e^{c t^{2}}, \quad|t| \leq T, \tag{88}
\end{equation*}
$$

for some $r_{\varepsilon}=O_{P_{f_{0}}^{Y}}(1), c>0$. Then using the inequality $E|Z| \leq E e^{Z}+E e^{-Z}$ for any random variable $Z$ we have the bound

$$
\sum_{\lambda<l \leq J, r} 2^{l\left(\alpha^{\prime}-\alpha\right)} E^{\Pi^{D_{\varepsilon}}}\left[\varepsilon^{-1}\left|\left\langle f-\tilde{f}, 2^{-l\left(\alpha^{\prime}+d / 2\right)} \Phi_{l, r}^{\mathcal{O}}\right\rangle_{L^{2}}\right| \mid Y\right] \lesssim r_{\varepsilon} \sum_{\lambda<l \leq J} 2^{l\left(\alpha^{\prime}+d-\alpha\right)},
$$

which is $o(1)$ as $\lambda \rightarrow \infty$ since $\alpha>\alpha^{\prime}+d$.
For term $C$ in (86), using the fact that $\left\langle X_{2}, \Phi_{l, r}^{\mathcal{O}}\right\rangle \sim N\left(0,\left\|S_{f_{0}}\left[\Phi_{l, r}^{\mathcal{O}} / u_{f_{0}}\right]\right\|_{L^{2}}^{2}\right)$ with $\left\|S_{f_{0}}\left[\Phi_{l, r}^{\mathcal{O}} / u_{f_{0}}\right]\right\|_{L^{2}} \lesssim 2^{2 l}$, we have

$$
\begin{align*}
& E\left\|X_{2}-P_{\lambda}\left(X_{2}\right)\right\|_{\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*} \leq E} \sup _{\|\psi\|_{C_{K}^{\alpha}(\mathcal{O})} \leq 1}\left|\left\langle X_{2}, \psi-\Pi_{W_{\lambda}} \psi\right\rangle_{L^{2}}\right| \\
&  \tag{89}\\
& \lesssim \sum_{\lambda<l, r} 2^{-l(\alpha+d / 2)} E\left|\left\langle X_{2}, \Phi_{l, r}^{\mathcal{O}}\right\rangle_{L^{2}}\right| \lesssim \sum_{\lambda<l} 2^{-l(\alpha-d / 2-2)} \underset{\lambda \rightarrow \infty}{=} o(1) .
\end{align*}
$$

To conclude the proof, let $\epsilon^{\prime}>0$ be given. By the preceding bounds we can choose $\lambda=\lambda\left(\epsilon^{\prime}\right)$ large enough so that the terms $B, C$ in (86) are each less than $\epsilon^{\prime} / 3$ (with $P_{f_{0}}^{Y}$-probability as close to 1 as desired in case $B$ ). Next, for every fixed $\lambda=\lambda(\epsilon)$, by Proposition 20 with $k=\operatorname{dim}\left(W_{\lambda}\right)$, the term $A$ can be made less than $\epsilon^{\prime} / 3$ with $P_{f_{0}}^{Y}$ probability as close to 1 as desired, for $\varepsilon$ small enough. Overall the quantity in (86) is thus less than $\epsilon^{\prime}>0$ arbitrary, with probability as close to 1 as desired, proving

$$
\beta\left(\widetilde{\Pi}^{D_{\varepsilon}}(\cdot \mid Y), \mathcal{N}_{f_{0}}\right) \xrightarrow{P_{f_{0}}^{Y}} 0
$$

STEP VII: Convergence of moments and posterior mean. Recall $\bar{f}=E^{\Pi}[f \mid Y]$. From the previous step we know that the law of $\varepsilon^{-1}(f-\widetilde{f})$ converges weakly in probability to $\mathcal{N}_{f_{0}}$ in $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$ as $\varepsilon \rightarrow 0$. By tightness of the Gaussian law $\mathcal{N}_{f_{0}}$ in $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$ and
[27, Theorem 2.1.20 and Exercise 2.1.2], we have $E\|\widetilde{f}\|_{\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}}^{4}<\infty$ and also, for some constants $C>2\left\|f_{0}\right\|_{\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}}$ and $c^{\prime}>0$,

$$
P_{f_{0}}^{Y}\left(\|\tilde{f}\|_{\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}}>C\right) \leq \operatorname{Pr}\left(\|\tilde{X}\|_{\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}}>C /(2 \varepsilon)\right) \leq 2 e^{-c^{\prime} / \varepsilon^{2}}
$$

Moreover, since $f \mid Y$ is bounded in $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$, we see from (55), Remark 16 and the Cauchy-Schwarz inequality that

$$
\begin{aligned}
& \varepsilon^{-2} E^{\Pi}\left[\|f-\tilde{f}\|_{\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}}^{2} \mid Y\right] \\
& \lesssim E^{\Pi_{D_{\varepsilon}}}\left[\varepsilon^{-2}\|f-\widetilde{f}\|_{\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}}^{2} \mid Y\right]+C^{2} \varepsilon^{-2} \Pi\left(D_{\varepsilon}^{c} \mid Y\right) \\
& \quad+\varepsilon^{-2}\|\widetilde{f}\|_{\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}}^{2} 1\left\{\|\tilde{f}\|_{\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}}>C\right\} \\
&= O_{P_{f_{0}}^{Y}}(1)+O_{P_{f_{0}}^{Y}}\left(\varepsilon^{-2} e^{-c\left(\eta_{\varepsilon} / \varepsilon^{2}\right)}+\varepsilon^{-2} e^{-c^{\prime} /\left(2 \varepsilon^{2}\right)}\right)=O_{P_{f_{0}}^{Y}}(1),
\end{aligned}
$$

where the first term can be bounded by similar arguments to those after (87) above, using also $E Z^{2} \leq 2\left(E e^{Z}+E e^{-Z}\right)$ for any random variable $Z$. We conclude that on an event of probability as close to 1 as desired, $\varepsilon^{-1}(f-\widetilde{f}) \mid Y$ has uniformly bounded second (norm-) moments. Using the Skorokhod imbedding [19, Theorem 11.7.2]) and standard uniform integrability arguments we can argue by contradiction and deduce that weak convergence implies convergence of the first moment (see also Section 7.6) and thus, on the above event,

$$
\begin{equation*}
\varepsilon^{-1}(\bar{f}-\tilde{f})=E^{\Pi}\left[\varepsilon^{-1}(f-\tilde{f}) \mid Y\right] \rightarrow E^{\mathcal{N}_{f_{0}}}(X)=0 \quad \text { in }\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*} \tag{90}
\end{equation*}
$$

so that indeed $\|\bar{f}-\widetilde{f}\|_{\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}}=o_{P_{f_{0}}^{Y}}(\varepsilon)$ and we can replace $\tilde{f}$ by $\bar{f}$ in Theorem 8 .

### 6.4. Proofs for Section 5.1

We sketch the proof of Corollary 11; the proof of Corollary 10 is the same (in fact simpler) and omitted. The proof of coverage of $C_{\varepsilon}$ in Corollary 11 follows [27, proof of Theorem 7.3.23] (see also [9, Theorem 1]), and we only indicate the necessary modifications: We first notice that the space $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$ is separable (since finite-dimensional wavelet approximations are norm dense, with the dual norm estimated as in (87)), which implies that balls in this space form uniformity classes for weak convergence towards $\mathcal{N}_{f_{0}}$. Moreover the mapping $\Phi(t)=\mathcal{N}_{f_{0}}\left(x:\|x\|_{\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}} \leq t\right)$ is strictly increasing in $t$ since any shell $\left\{x: t<\|x\|_{\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}}<t+\delta\right\}, \delta>0$, of $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$ contains an element of the reproducing kernel Hilbert space of $\mathcal{N}_{f_{0}}$ (this space equals the image of $L^{2}$ under the Schrödinger operator $S_{f_{0}}$; using arguments from Section 7.2 it is easily seen that this space contains elements of $\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}$ of any given norm $\|h\|_{\left(C_{K}^{\alpha}(\mathcal{O})\right)^{*}}, \alpha>2+d / 2$ ). Thus [27, proof of Theorem 7.3.23] applies directly to give coverage of $C_{\varepsilon}$ and also shows that $\varepsilon^{-1} R_{\varepsilon}$ converges in $P_{f_{0}}^{Y}$-probability to $\Phi^{-1}(1-\beta)$.

## 7. Appendix

### 7.1. The metric entropy inequality for sub-Gaussian processes

'Dudley's metric entropy inequality' for suprema of sub-Gaussian random processes has been used repeatedly in the proofs: A centred stochastic process $(X(s): s \in S)$ is said to be sub-Gaussian for some metric $d$ on its index set $S$ if $E e^{\lambda(X(s)-X(t))} \leq e^{\lambda^{2} d^{2}(s, t) / 2}$ for all $s, t \in S$ an $\lambda \in \mathbb{R}$. Denote by $N(S, \gamma, d)$ the $\gamma$-covering numbers of the metric space $(S, d)$. Then for some $s_{0} \in S, D$ any upper bound for the diameter of the metric space $(S, d)$ and any $\delta>0$, we have the inequalities

$$
\begin{align*}
& E \sup _{s \in S}|X(s)| \leq E\left|X\left(s_{0}\right)\right|+4 \sqrt{2} \int_{0}^{D / 2} \sqrt{2 \log N(S, \gamma, d)} d \gamma  \tag{91}\\
& E \sup _{s, t \in S: d(s, t) \leq \delta}|X(s)-X(t)| \leq(16 \sqrt{2}+2) \int_{0}^{\delta} \sqrt{2 \log N(S, \gamma, d)} d \gamma \tag{92}
\end{align*}
$$

We also note that provided the last integrals converge, the process $(X(s): s \in S)$ has a version for which the last suprema are measurable. See [27, Theorem 2.3.7] for proofs.

### 7.2. Some properties of Schrödinger operators

The PDE (1) has been extensively studied, and we review here a few facts that were used in our proofs. For $f \in C(\overline{\mathcal{O}})$, the Schrödinger operator

$$
\begin{equation*}
S_{f}[u]=\frac{\Delta}{2} u-f u \tag{93}
\end{equation*}
$$

is defined classically for all $u \in C^{2}(\mathcal{O})$, and weakly for all locally integrable functions $u$ by the action $\int_{\mathcal{O}} S_{f}[u] \varphi=\int_{\mathcal{O}} u S_{f}[\varphi]$ on all $\varphi \in C_{c}^{\infty}(\mathcal{O})$.

Under suitable conditions the Schrödinger operator will be seen to have an inverse ' $f$-Green' operator $V_{f}$ which describes the unique solutions $V_{f}[h]$ of the inhomogeneous Schrödinger equation

$$
\begin{equation*}
S_{f}[u]=(\Delta / 2) u-f u=h \quad \text { on } \mathcal{O}, \quad u=0 \quad \text { on } \partial \mathcal{O} \tag{94}
\end{equation*}
$$

The operator $V_{f}$ has a probabilistic representation by the Feynman-Kac formula

$$
\begin{equation*}
V_{f}[h](x)=-E^{x}\left[\int_{0}^{\tau_{\mathcal{O}}} h\left(X_{t}\right) e^{-\int_{0}^{t} f\left(X_{s}\right) d s} d t\right], \quad x \in \mathcal{O} \tag{95}
\end{equation*}
$$

where $X_{s}$ is a standard $d$-dimensional Brownian motion started at $x$ with exit time $\tau_{\mathcal{O}}$ from $\mathcal{O}$ (see [12, p. 88]). We also recall (e.g., see [12, Theorem 1.17])

$$
\sup _{x \in \mathcal{O}} E^{x} \tau_{\mathcal{O}} \leq K(\operatorname{vol}(\mathcal{O}), d)<\infty
$$

Some key properties of $S_{f}$ and $V_{f}$ are summarised in the following result.

Proposition 22. Let $\mathcal{O}$ be a bounded $C^{\infty}$-domain in $\mathbb{R}^{d}$. Suppose $f \in C(\overline{\mathcal{O}})$ satisfies $f \geq f_{\text {min }}>0$ on $\overline{\mathcal{O}}$. Then for any $h \in C(\overline{\mathcal{O}}), V_{f}[h] \in C_{0}(\mathcal{O})$ satisfies

$$
\begin{equation*}
S_{f}\left[V_{f}[h]\right]=h \quad \text { on } \mathcal{O} \tag{96}
\end{equation*}
$$

and is the unique solution of (94). If $h \in C^{2}(\overline{\mathcal{O}}) \cap C_{0}(\mathcal{O})$ then also

$$
\begin{equation*}
V_{f}\left[S_{f}[h]\right]=h \quad \text { on } \mathcal{O} . \tag{97}
\end{equation*}
$$

The operator $V_{f}[h]=\int_{\mathcal{O}} v(\cdot, y) h(y) d y$ admits a symmetric kernel $v(x, y)=v(y, x)$, $x, y \in \mathcal{O}$, and extends to a self-adjoint operator on $L^{2}(\mathcal{O})$.

Proof. These results follow from [12, Theorems 3.18 and 3.22], where we notice that the space $\mathbb{F}(D, q)$, defined just before Proposition 3.16 in that reference, for $f=-q$ bounded away from zero contains all bounded functions. We notice further that $V_{f}$ is defined in [12, (39)], and for $f \geq 0$ we obviously see from (95) that $\sup _{x} V_{f}[1](x) \leq$ $\sup _{x} E^{x} \tau_{\mathcal{O}}<\infty$, so that the conditions of [12, Theorems 3.18 and 3.22] are satisfied. Symmetry of the kernel $v$ follows from [12, Corollary 3.18]. Lemma 23 below implies that the operator $V_{f}$ is continuous on $L^{2}(\mathcal{O})$ and hence extends to a self-adjoint operator on that space.

We next turn to the mapping properties of $V_{f}$ and $S_{f}$. For $\mathcal{O}$ a bounded $C^{\infty}$-domain, $\partial \mathcal{O}$ is a compact smooth manifold and the spaces $H^{\beta}(\partial \mathcal{O})$ and $\mathcal{C}^{\beta}(\partial \mathcal{O})$ can be defined as usual [37, 52]. The Schrödinger operator $S_{f}$ is properly elliptic in the sense of [37, p. 110f.], and if $\operatorname{tr}[u]=u_{\mid \partial \mathcal{O}}$ is the usual boundary trace map, then $\left(S_{f}, \operatorname{tr}\right)$ can be shown to realise a topological isomorphism of $H^{\beta+2}(\mathcal{O})$ onto $H^{\beta}(\mathcal{O}) \times H^{\beta+3 / 2}(\partial \mathcal{O})$ for every $\beta \geq 0$ (see [37, Theorem II.5.4] or [52, Theorem 4.3.3]. Likewise, $\left(S_{f}, \operatorname{tr}\right)$ realises an isomorphism of the Hölder-Zygmund space $\mathcal{C}^{\beta+2}(\mathcal{O})$ onto $\mathcal{C}^{\beta}(\mathcal{O}) \times \mathcal{C}^{\beta+2}(\partial \mathcal{O})$ for all $\beta \geq 0$ (see [52, Theorem 4.3.4]). These isomorphisms are proved under the further assumption that the only smooth solution to the problem $S_{f}[u]=0$ on $\mathcal{O}$ with $u=0$ on $\partial \mathcal{O}$ equals $u=0$ identically, which is true in view of (95) and Proposition 22. From the above isomorphisms we deduce

$$
\begin{equation*}
\|u\|_{H^{\beta+2}(\mathcal{O})} \lesssim\left\|S_{f}[u]\right\|_{H^{\beta}(\mathcal{O})}+\|\operatorname{tr}[u]\|_{H^{\beta+2}(\partial \mathcal{O})} \quad \forall u \in H^{\beta+2}(\mathcal{O}), \beta \geq 0 \tag{98}
\end{equation*}
$$

using also $H^{\beta+2}(\partial \mathcal{O}) \subseteq H^{\beta+3 / 2}(\partial \mathcal{O})$, and

$$
\begin{equation*}
\|u\|_{\mathcal{C}^{\beta+2}(\mathcal{O})} \lesssim\left\|S_{f}[u]\right\|_{\mathcal{C}^{\beta}(\mathcal{O})}+\|\operatorname{tr}[u]\|_{\mathcal{C}^{\beta+2}(\partial \mathcal{O})} \quad \forall u \in \mathcal{C}^{\beta+2}(\mathcal{O}), \beta \geq 0 \tag{99}
\end{equation*}
$$

In the above references, the coefficient $f$ is assumed to be a smooth function, but basic arguments show that (98) and (99) remain valid whenever $f \in \mathcal{C}^{\beta}(\mathcal{O})$. The constants in the preceding inequalities then depend only on $\beta, d, \mathcal{O}$ and on a bound for $\|f\|_{\mathcal{C}^{\beta}(\mathcal{O})}$.

Lemma 23. For any $f$ as in Proposition 22 and some constant $c$,

$$
\left\|V_{f}[h]\right\|_{L^{p}} \leq c\|h\|_{L^{p}} \quad \forall h \in C(\overline{\mathcal{O}}), p \in\{2, \infty\}
$$

Proof. Since $V_{f}[h] \in C_{0}(\mathcal{O})$ by Proposition 22, its boundary trace vanishes and thus we immediately deduce the case $p=2$ from that proposition and (98) with $\beta=0$ and since $H^{2}(\mathcal{O})$ imbeds continuously in $L^{2}(\mathcal{O})$. The case $p=\infty$ follows directly from (95).

For the next more precise lemma we recall the dual norm (21).
Lemma 24. Let $f \in C^{s}(\overline{\mathcal{O}})$ satisfy $f>0,\|f\|_{C^{s}} \leq D$ for some $s>0$, and let $h \in$ $C(\overline{\mathcal{O}})$ be given. Suppose that $\omega$ is a solution of the inhomogeneous equation

$$
(\Delta / 2) \omega-f \omega=h \quad \text { on } \mathcal{O}, \quad \omega=0 \quad \text { on } \partial \mathcal{O}
$$

Then for some constant $C(D, d, \mathcal{O})$,

$$
\begin{equation*}
\|\omega\|_{L^{2}}=\left\|V_{f}[h]\right\|_{L^{2}} \leq C\|h\|_{\left(H_{0}^{2}\right)^{*}} \tag{100}
\end{equation*}
$$

If moreover $\beta$ is a non-negative integer and $s \geq \beta$, then also, for some constant $C^{\prime}=$ $C^{\prime}(D, \beta, d, \mathcal{O})$,

$$
\begin{equation*}
\|\omega\|_{H^{\beta+2}}=\left\|V_{f}[h]\right\|_{H^{\beta+2}} \leq C^{\prime}\|h\|_{H^{\beta}} \tag{101}
\end{equation*}
$$

and for all $h \in C^{\beta}(\mathcal{O})$ and $s>\beta$,

$$
\begin{equation*}
\|\omega\|_{\mathcal{C}^{\beta+2}}=\left\|V_{f}[h]\right\|_{\mathcal{C}^{\beta+2}} \leq C^{\prime}\|h\|_{\mathcal{C}^{\beta}} . \tag{102}
\end{equation*}
$$

Proof. We first prove (101) for $h \in H^{\beta}(\mathcal{O})$. By [25, Theorems 8.9, 8.13] and Proposition 22 the solution $\omega$ is unique, lies in $H^{\beta+2}(\mathcal{O}) \cap C_{0}(\mathcal{O})$ and can be represented as $\omega=V_{f}[h]$. Then (98) gives

$$
\|\omega\|_{H^{\beta+2}(\mathcal{O})} \lesssim\left\|S_{f}\left[V_{f}[h]\right]\right\|_{H^{\beta}(\mathcal{O})}+\left\|\operatorname{tr}\left[V_{f}[h]\right]\right\|_{H^{\beta+2}(\partial \mathcal{O})}=\|h\|_{H^{\beta}(\mathcal{O})}
$$

Replacing (98) by (99) and using [25, Theorems 6.14, 6.19] and again Proposition 22 to ensure that $\omega \in C^{\beta+2}(\mathcal{O}) \cap C_{0}(\mathcal{O})$ for $h \in C^{s}(\mathcal{O}), f \in C^{s}(\overline{\mathcal{O}})$, the same estimate follows for $\mathcal{C}^{\beta}$-norms replacing $H^{\beta}$-norms, giving (102) for such $h$. Since $C^{s}(\mathcal{O}), s>\beta$, is dense in $C^{\beta}(\mathcal{O}) \subseteq \mathcal{C}^{\beta}(\mathcal{O})$, the overall result follows from a basic approximation argument.

We finally prove (100): By Proposition 22 the operator $V_{f}$ is self-adjoint and takes values in $C_{0}(\mathcal{O})$. We thus see from (101) with $\beta=0$ that

$$
\begin{aligned}
& \|w\|_{L^{2}}=\left\|V_{f}[h]\right\|_{L^{2}}=\sup _{\varphi \in C_{c}(\mathcal{O}):\|\varphi\|_{L^{2}} \leq 1}\left|\int_{\mathcal{O}} \varphi V_{f}[h]\right| \\
& \quad=\sup _{\varphi \in C_{c}(\mathcal{O}):\|\varphi\|_{L^{2}} \leq 1}\left|\int_{\mathcal{O}} V_{f}[\varphi] h\right| \leq \sup _{\varphi \in C_{c}(\mathcal{O}):\|\varphi\|_{L^{2}} \leq 1}\left\|V_{f}[\varphi]\right\|_{H^{2}}\|h\|_{\left(H_{0}^{2}\right)^{*}} \leq C\|h\|_{\left(H_{0}^{2}\right)^{*}},
\end{aligned}
$$

completing the proof.
We turn to existence and properties of solutions to the homogeneous Schrödinger equation (1), and some basic stability estimates for the 'solution maps' of (1), (94).

Proposition 25. Let $f>0$ satisfy $\|f\|_{C^{s}(\mathcal{O})} \leq D$ for some $s, D>0$, and assume that $g \in C^{s+2}(\overline{\mathcal{O}})$.
(A) A unique solution $u_{f} \in C^{2}(\overline{\mathcal{O}})$ of (1) exists and has Feynman-Kac representation (2). Moreover, for every non-negative integer $0 \leq \beta \leq s$ we have

$$
\begin{equation*}
\left\|u_{f}\right\|_{H^{\beta+2}(\mathcal{O})} \leq D^{\prime}\|g\|_{H^{\beta+2}(\partial \mathcal{O})} \tag{103}
\end{equation*}
$$

and if $0 \leq \beta<s$ then also

$$
\begin{equation*}
\left\|u_{f}\right\|_{\mathcal{C}^{\beta+2}(\mathcal{O})} \leq D^{\prime}\|g\|_{\mathcal{C}^{\beta+2}(\partial \mathcal{O})} \tag{104}
\end{equation*}
$$

where $D^{\prime}$ depends only on $D, d, \mathcal{O}$.
(B) If $f, h \in C^{s}(\mathcal{O}), s>0$, and $u_{f}, u_{h} \in C(\overline{\mathcal{O}})$ are solutions to (1) with coefficients $f$, h, respectively, then

$$
\left\|u_{f}-u_{h}\right\|_{L^{2}} \leq c\|f-h\|_{\left(H_{0}^{2}\right)^{*}} \leq c\|f-h\|_{L^{2}}
$$

where $c>0$ depends only on upper bounds for $\|f\|_{C^{s}(\mathcal{O})},\|h\|_{C^{s}(\mathcal{O})},\|g\|_{C^{s+2}(\overline{\mathcal{O}})}$ and on $d, \mathcal{O}$.
(C) If $f, h \in C(\overline{\mathcal{O}})$ both satisfy $f, h \geq f_{\min }>0$, then for all $q \in C(\overline{\mathcal{O}})$,

$$
\left\|V_{f}(q)-V_{h}(q)\right\|_{L^{2}} \lesssim\|f-h\|_{L^{2}}\|q\|_{\infty}
$$

Proof. (A) The existence result follows from [25, Theorem 6.14]. The Feynman-Kac representation (2) is derived, e.g., in [12, Theorem 4.7]. To prove (103) we notice that [25, Theorem 8.13] and the hypotheses imply $u_{f} \in H^{\beta+2}$ so that (98) gives

$$
\left\|u_{f}\right\|_{H^{\beta+2}(\mathcal{O})} \lesssim\left\|S_{f}\left[u_{f}\right]\right\|_{H^{\beta}(\mathcal{O})}+\left\|\operatorname{tr}\left[u_{f}\right]\right\|_{H^{\beta+2}(\partial \mathcal{O})} \lesssim\|g\|_{H^{\beta+2}(\partial \mathcal{O})} .
$$

We can then prove (104) completely analogously, using [25, Theorem 6.19] to establish $u_{f} \in C^{\beta+2}(\mathcal{O})$ and then (99).
(B) We notice that $\omega=u_{f}-u_{h}$ solves the inhomogeneous equation $(\Delta / 2) \omega-f \omega=$ $(f-h) u_{h}$ on $\mathcal{O}$ with $\omega=g-g=0$ on $\partial \mathcal{O}$. By Proposition 22, Lemma 24 and (11) we thus have

$$
\begin{aligned}
\left\|u_{f}-u_{h}\right\|_{L^{2}} & =\left\|V_{f}\left[(f-h) u_{h}\right]\right\|_{L^{2}} \leq C\left\|u_{h}(f-h)\right\|_{\left(H_{0}^{2}\right)^{*}} \\
& \leq C \sup _{\|g\|_{H^{2}} \leq 1}\left\|g u_{h}\right\|_{H^{2}}\|f-h\|_{\left(H_{0}^{2}\right)^{*}} \leq c\left\|u_{h}\right\|_{\mathcal{C}^{2}}\|f-h\|_{\left(H_{0}^{2}\right)^{*}},
\end{aligned}
$$

and the result follows since $\left\|u_{h}\right\|_{\mathcal{C}^{2}} \leq$ const by part (A) with $\beta=0$.
(C) Notice that $v=V_{f}[q]-V_{h}[q]$ solves $\Delta v / 2-f v=(f-h) V_{h}[q]$ on $\mathcal{O}$ subject to zero boundary conditions. Thus $v=V_{f}\left[(f-h) V_{h}[q]\right]$ and so by Lemma 23 we must have $\|v\|_{L^{2}} \lesssim\|f-h\|_{L^{2}}\left\|V_{h}[q]\right\|_{L^{\infty}} \lesssim\|f-h\|_{L^{2}}\|q\|_{\infty}$.

### 7.3. Some properties of Hölder-type spaces

Proposition 26. We have $C_{c}^{\alpha}(\mathcal{O}) \subseteq \mathcal{C}_{c}^{\alpha}(\mathcal{O}) \subseteq \mathcal{C}^{\alpha, W}(\mathcal{O})$ for all $0<\alpha<S$, and $\|f\|_{\mathcal{C}^{\alpha, W}(\mathcal{O})} \leq c_{2}^{\prime}\|f\|_{\mathcal{C}^{\alpha}(\mathcal{O})} \leq c_{2}\|f\|_{C^{\alpha}(\mathcal{O})}$, with uniform constants $c_{2}, c_{2}^{\prime}$.
Proof. We first prove the result for $f \in C_{c}^{\alpha}(\mathcal{O})$, which has a zero extension from $\mathcal{O}$ to $\mathbb{R}^{d}$ that defines an element of $C^{\alpha}\left(\mathbb{R}^{d}\right)$ and the global Hölder norm is equal to the intrinsic one. Thus by (13), for all the interior wavelets $\Phi_{l, r}^{\mathcal{O}}=\Phi_{l, r}$ supported within $\mathcal{O}$ we have

$$
2^{l(\alpha+d / 2)}\left|\left\langle f, \Phi_{l, r}^{\mathcal{O}}\right\rangle_{L^{2}(\mathcal{O})}\right|=2^{l(\alpha+d / 2)}\left|\left\langle f, \Phi_{l, r}\right\rangle_{L^{2}\left(\mathbb{R}^{d}\right)}\right| \leq c\|f\|_{C^{\alpha}(\mathcal{O})}
$$

and for the boundary wavelets, by the support of $f$ in $\mathcal{O}$ and (15),

$$
\begin{aligned}
\sup _{l, r} 2^{l(\alpha+d / 2)}\left|\left\langle f, \Phi_{l, r}^{\mathcal{O}}\right\rangle_{L^{2}(\mathcal{O})}\right| & \leq c \sup _{l, r} \sum_{\left|m-m^{\prime}\right| \leq K}\left|d_{m, m^{\prime}}^{l}\right| 2^{l(\alpha+d / 2)}\left|\left\langle f, \Phi_{l, m^{\prime}}\right\rangle_{L^{2}\left(\mathbb{R}^{d}\right)}\right| \\
& \leq c D\|f\|_{C^{\alpha}(\mathcal{O})} .
\end{aligned}
$$

The preceding proof also applies to $f \in \mathcal{C}_{c}^{\alpha}(\mathcal{O})$ since the wavelet norm (13) is an equivalent norm on $\mathcal{C}^{\alpha}\left(\mathbb{R}^{d}\right)$ [27, Chapter 4], and since $C_{c}^{\alpha}(\mathcal{O}) \subseteq \mathcal{C}_{c}^{\alpha}(\mathcal{O})$ (see after (9)).

Some further properties are the content of the next proposition. For these we recall the usual definition of Besov spaces $B_{p q}^{\alpha}\left(\mathbb{R}^{d}\right), \alpha \in \mathbb{R}$, by tensor Littlewood-Paley decomposition and of $B_{p q}^{\alpha, W}\left(\mathbb{R}^{d}\right), \alpha \in \mathbb{R}$, by a tensor wavelet basis, equivalent to each other and to the classical definition in terms of moduli of continuity when $\alpha>0$. See [39,52,53] and also [27, Section 4.3.1] (adapted to the multi-dimensional setting) for details.

Proposition 27. (A) Let $\alpha>0$ but $\alpha-2 \ell<0$ for some $\ell \in \mathbb{N}$. Then for any $h$ in $C_{c}^{\alpha}(\mathcal{O}) \subseteq C_{c}^{\alpha}\left(\mathbb{R}^{d}\right)$ and $\Delta^{\ell}=\Delta \ldots \Delta$ ( $\ell$ times) the iterated Laplace operator, we have

$$
\begin{equation*}
\left\|\Delta^{\ell} h\right\|_{B_{\infty \infty}^{\alpha-2 \ell}\left(\mathbb{R}^{d}\right)} \leq C\|h\|_{C^{\alpha}(\mathcal{O})} \tag{105}
\end{equation*}
$$

(B) Let $\alpha<0, \rho>|\alpha|$ and $g \in C^{\rho}(\mathcal{O})$. Then, using the definition of the $\mathcal{C}^{\alpha, W}(\mathcal{O})$-norms in (16) also for negative values of $\alpha$, we have, for all $h \in C_{c}(\mathcal{O})$ and some constant $c>0$,

$$
\begin{equation*}
\|h g\|_{\mathcal{C}^{\alpha, W}(\mathcal{O})} \leq c\|h\|_{B_{\infty \infty}^{\alpha}\left(\mathbb{R}^{d}\right)}\|g\|_{C^{\rho}(\mathcal{O})} \tag{106}
\end{equation*}
$$

(C) For every $\alpha>0$ there exist finite constants $A>0$ depending on $d$, $\alpha$ such that the $\|\cdot\|_{\infty}$-metric entropy of the unit ball of $C_{c}^{\alpha}(\mathcal{O})$ can be bounded as

$$
\begin{equation*}
\log N\left(\left\{\psi:\|\psi\|_{C_{c}^{\alpha}(\mathcal{O})} \leq 1\right\}, \eta,\|\cdot\|_{\infty}\right) \leq(A / \eta)^{d / \alpha}, \quad 0<\eta<A / 2 \tag{107}
\end{equation*}
$$

Proof. For (A), we have the inequalities

$$
\left\|\Delta^{\ell} h\right\|_{B_{\infty \infty}^{\alpha-2 \ell}\left(\mathbb{R}^{d}\right)} \leq c\|h\|_{B_{\infty \infty}^{\alpha}\left(\mathbb{R}^{d}\right)} \leq c^{\prime}\|h\|_{C^{\alpha}(\mathcal{O})}
$$

where the first inequality follows from [52, Theorem 2.3.8] (or as in [27, proof of Proposition 4.3.19]), and where we have used the continuous imbedding $C_{c}^{\alpha}(\mathcal{O}) \subseteq C^{\alpha}\left(\mathbb{R}^{d}\right) \subseteq$ $B_{\infty \infty}^{\alpha}\left(\mathbb{R}^{d}\right)$ (see [52]) in the second inequality.

We now prove (B): since $h$ has compact support in $\mathcal{O}$ we can multiply $g$ by a function in $C_{c}^{\infty}(\mathcal{O})$ so that $h g=h \bar{g}$ on $\mathcal{O}$ and $\bar{g} \in C_{c}^{\rho}(\mathcal{O})$ satisfies $\|\bar{g}\|_{C^{\rho}(\mathcal{O})} \leq c\|g\|_{C^{\rho}(\mathcal{O})}$. Then [52, Theorem 2.8.2] implies

$$
\|h \bar{g}\|_{B_{\infty \infty}^{\alpha}\left(\mathbb{R}^{d}\right)} \leq c^{\prime}\|h\|_{B_{\infty \infty}^{\alpha}\left(\mathbb{R}^{d}\right)}\|\bar{g}\|_{B_{\infty \infty}^{\rho}\left(\mathbb{R}^{d}\right)} \leq c^{\prime \prime}\|h\|_{B_{\infty \infty}^{\alpha}\left(\mathbb{R}^{d}\right)}\|g\|_{C^{\rho}(\mathcal{O})}
$$

The $\|h \bar{g}\|_{B_{\infty}^{\alpha}}^{\alpha}$-norm bounds the $\|h \bar{g}\|_{B_{\infty}^{\alpha, W}}$-norm up to a constant multiple, which implies the desired decay of the wavelet coefficients of $h \bar{g}$-the bound on the $\|h g\|_{\mathcal{C}^{\alpha, W}}$-norm now follows by just repeating the estimates from the proof of Proposition 26.

Finally, the metric entropy bound in (C) is proved in [58, Theorem 2.7.1] (the domain $\mathcal{O}$ there has to be bounded and convex but we can always extend elements of $C_{c}^{\alpha}(\mathcal{O})$ by zero to a larger bounded convex domain without increasing the norm).

### 7.4. Likelihood functions and a contraction theorem for general inverse problems

Let $\mathcal{F}$ be a Polish space with Borel $\sigma$-field $\mathcal{B}_{\mathcal{F}}$ and let $\mathbb{H}$ be a real Hilbert space that is separable for the norm induced by the inner product $\langle\cdot, \cdot\rangle_{\mathbb{H}}$, with Borel $\sigma$-field $\mathcal{B}_{\mathbb{H}}$. Suppose $G: \mathcal{F} \rightarrow \mathbb{H}$ is a Borel-measurable mapping and, for $\varepsilon>0$ a scalar 'noise level', consider the formal equation in $\mathbb{H}$ given by

$$
\begin{equation*}
Y=G(f)+\varepsilon \mathbb{W} . \tag{108}
\end{equation*}
$$

Here $\mathbb{W}$ is a centred Gaussian white noise process $(\mathbb{W}(h): h \in \mathbb{H})$ with covariance $E \mathbb{W}(h) \mathbb{W}(g)=\langle h, g\rangle_{\mathbb{H}}$, defined on some probability space $(\Omega, \mathcal{A}, \mu)$ (we can take $\Omega=\mathbb{R}^{\mathbb{N}}$ with its cylindrical $\sigma$-algebra and $\mu$ the law of $\otimes_{k=1}^{\infty} N(0,1)$, see [27, Example 2.1.11]). Observing (108) then means that we observe a realisation of the Gaussian process $\left(Y(h)=\langle Y, h\rangle_{\mathbb{H}}: h \in \mathbb{H}\right)$ with marginal distributions $Y(h) \sim$ $N\left(\langle G(f), h\rangle_{\mathbb{H}},\|h\|_{\mathbb{H}}^{2}\right)$. We sometimes write $\langle\mathbb{W}, h\rangle_{\mathbb{H}}$ for the random variable $\mathbb{W}(h)$.

If ( $e_{k}: k \in \mathbb{Z}$ ) form an orthonormal basis of $\mathbb{H}$, and if $w=\left(w_{k}: k \in \mathbb{Z}\right) \in \ell_{2}$, $w_{k} \geq 0, w_{k} \downarrow 0$ as $|k| \rightarrow \infty$, then we can define the new Hilbert space

$$
\begin{equation*}
\mathbb{H}_{w}=\left\{f=\sum_{k} e_{k} f_{k}: \sum_{k} f_{k}^{2} w_{k}^{2}=\|f\|_{\mathbb{H}_{w}}^{2}<\infty\right\} \tag{109}
\end{equation*}
$$

as the completion of $\mathbb{H}$ with respect to the $\|\cdot\|_{\mathbb{H}_{w}}$-norm, where $f_{k}=\left\langle f, e_{k}\right\rangle_{\mathbb{H}}$. By definition

$$
E\|\mathbb{W}\|_{\mathbb{H}_{w}}^{2}=\sum_{k} E \mathbb{W}\left(e_{k}\right)^{2} w_{k}^{2}<\infty,
$$

so using Ulam's theorem [19, Theorem 7.1.4]) and separability of $\mathbb{H}_{w}$, the cylindrically defined law of $\mathbb{W}$ extends to a tight Gaussian probability measure on the Borel $\sigma$-field $\mathcal{B}_{\mathbb{H}_{w}}$ of $\mathbb{H}_{w}$. The equation $Y=G(f)+\varepsilon \mathbb{W}$ then makes rigorous sense in $\mathbb{H}_{w}$, with $P_{G(f)}^{Y}$ denoting the shifted Gaussian law of the random variable $Y:(\Omega, \mathcal{A}) \rightarrow$ $\left(\mathbb{H}_{w}, \mathcal{B}_{\mathbb{H}_{w}}\right)$. If we let the law $P_{0}^{Y}$ of $\varepsilon \mathbb{W}$ serve as a common dominating measure then
for $G(f)$ contained in the RKHS $\mathbb{H}$ of $P_{0}^{Y}$, the Cameron-Martin theorem (e.g., as in [27, Proposition 6.1.5]) allows us to define the log-likelihood function

$$
\begin{equation*}
\ell(f)=\log p_{f}(Y) \equiv \log \frac{d P_{G(f)}^{Y}}{d P_{0}^{Y}}(Y)=\frac{1}{\varepsilon^{2}}\langle Y, G(f)\rangle_{\mathbb{H}}-\frac{1}{2 \varepsilon^{2}}\|G(f)\|_{\mathbb{H}}^{2} . \tag{110}
\end{equation*}
$$

The mapping ( $\omega, f$ ) $\mapsto p_{f}\left(Y(\omega)\right.$ ) is jointly measurable from $\left(\Omega \times \mathcal{F}, \mathcal{A} \otimes \mathcal{B}_{\mathcal{F}}\right)$ to $\left(\mathbb{R}, \mathcal{B}_{\mathbb{R}}\right)$. [Indeed, using the $\mathcal{B}_{\mathcal{F}}-\mathcal{B}_{\mathbb{H}}$-measurability of $G$ it suffices to show joint measurability of the map $L:(\omega, h) \mapsto\langle\mathbb{W}(\omega), h\rangle_{\mathbb{H}}$ defined on $\left(\Omega \times \mathbb{H}, \mathcal{A} \otimes \mathcal{B}_{\mathbb{H}}\right)$. By what precedes, $\left(\langle\mathbb{W}(\omega), h\rangle_{\mathbb{H}}: h \in \mathbb{H}\right)$ is a centred Gaussian process, so joint measurability follows from separability of $\mathbb{H}$ and [27, Proposition 2.1.12 and Definition 2.1.2].]

If now $\Pi$ is a prior probability distribution on $\left(\mathcal{F}, \mathcal{B}_{\mathcal{F}}\right)$, then we can apply Bayes' theorem in the product space

$$
\left(\Omega \times \mathcal{F}, \mathcal{A} \otimes \mathcal{B}_{\mathcal{F}}, Q\right), \quad d Q(y, f)=p_{f}(Y(\omega)) \mu(\omega) \Pi(f)
$$

to deduce (as in [24, p. 7]) that the posterior distribution of $f|Y=f| Y(\omega)$ equals

$$
\begin{equation*}
\Pi(B \mid Y)=\frac{\int_{B} p_{f}(Y) d \Pi(f)}{\int_{\mathcal{F}} p_{f}(Y) d \Pi(f)}=\frac{\int_{B} p_{f}(Y(\omega)) d \Pi(f)}{\int_{\mathcal{F}} p_{f}(Y(\omega)) d \Pi(f)}, \quad B \in \mathcal{B}_{\mathcal{F}} \tag{111}
\end{equation*}
$$

A general contraction theorem can now be proved as in the standard direct setting [23, 24, 27], after noting that the induced 'information distance' is $\left\|G(f)-G\left(f_{0}\right)\right\|_{\mathbb{H}}$ in the measurement model (108).

Theorem 28. Let $\Pi=\Pi_{\varepsilon}$ be a sequence of prior distributions on $\left(\mathcal{F}, \mathcal{B}_{\mathcal{F}}\right)$, and let $d(\cdot, \cdot)$ be a (measurable) distance function on $\mathcal{F}$. Let $\Pi(\cdot \mid Y)$ be the posterior distribution from (111) and suppose $Y=G\left(f_{0}\right)+\varepsilon \mathbb{W}$ has law $P_{f_{0}}^{Y} \equiv P_{G\left(f_{0}\right)}^{Y}$, for some fixed $f_{0} \in \mathcal{F}$. For a sequence of numbers $\bar{\eta}_{\varepsilon} \rightarrow 0$ such that $\bar{\eta}_{\varepsilon} / \varepsilon \rightarrow \infty$ and $C^{\prime}$, L fixed constants, suppose $\Pi$ satisfies

$$
\begin{equation*}
\Pi\left(f \in \mathcal{F}:\left\|G(f)-G\left(f_{0}\right)\right\|_{\mathbb{H}}<\bar{\eta}_{\varepsilon}\right) \geq e^{-C^{\prime}\left(\bar{\eta}_{\varepsilon} / \varepsilon\right)^{2}} \tag{112}
\end{equation*}
$$

and

$$
\Pi\left(\mathcal{F} \backslash \mathcal{F}_{\varepsilon}\right) \leq L e^{-\left(C^{\prime}+4\right)\left(\bar{\eta}_{\varepsilon} / \varepsilon\right)^{2}}
$$

for a sequence of measurable sets $\mathcal{F}_{\varepsilon} \subseteq \mathcal{F}$ and numbers $\eta_{\varepsilon}^{*}$ for which we can find tests (indicator functions) $\Psi(Y)$ such that

$$
E_{f_{0}}^{Y} \Psi(Y)+\sup _{f \in \mathcal{F}_{\varepsilon}: d\left(f, f_{0}\right) \geq \eta_{\varepsilon}^{*}} E_{f}^{Y}(1-\Psi(Y)) \leq L e^{-\left(C^{\prime}+4\right)\left(\bar{\eta}_{\varepsilon} / \varepsilon\right)^{2}}
$$

Then for some $c>0$ and as $\varepsilon \rightarrow 0$, we have

$$
\begin{equation*}
\Pi\left(d\left(f, f_{0}\right) \geq \eta_{\varepsilon}^{*} \mid Y\right)=O_{P_{f_{0}}^{Y}}\left(e^{-c\left(\bar{\eta}_{\varepsilon} / \varepsilon\right)^{2}}\right)=o_{P_{f_{0}}^{Y}}(1) \tag{113}
\end{equation*}
$$

Proof. The result is proved just as [27, Theorem 7.3.5], by adapting the proof of Lemma 7.3.4 there to the present situation by replacing the $L^{2}$-norm by our $\mathbb{H}$-norm, $1 / \sqrt{n}$ there by our $\varepsilon$, and $\varepsilon_{n}$ there by our $\bar{\eta}_{\varepsilon}$, respectively. Inspection of the proof shows that also the stronger conclusion (113) is satisfied (cf. also [24, Exercise 8.7, p. 230]).

### 7.5. Information lower bounds in function space

We now recall some standard facts from efficient estimation in infinite-dimensional parameter spaces (see [57, Chapter 25] for an introduction to the general theory). Assume that for all $h$ in some linear subspace $H$ of an inner product space with Hilbert norm $\|\cdot\|_{\text {LAN }}$, the log-likelihood-ratio process of a statistical model of laws $\left\{\mathbb{P}_{f+h}^{\varepsilon}: h \in H\right\}$ on some sequence $\left(\mathcal{X}_{\varepsilon}\right)$ of measurable spaces has locally asymptotically normal (LAN) expansion

$$
\log \frac{d \mathbb{P}_{f+\varepsilon h}^{\varepsilon}}{d \mathbb{P}_{f}^{\varepsilon}}=D_{\varepsilon}(h)-\frac{1}{2}\|h\|_{\mathrm{LAN}}^{2}, \quad h \in H
$$

where, as $\varepsilon \rightarrow 0, D_{\varepsilon}(h)$ converges in distribution under $\mathbb{P}_{f}^{\varepsilon}$ to $D(h) \sim N\left(0,\|h\|_{\text {LAN }}^{2}\right)$ for every fixed $h \in H$. Next, let $\left(C,\|\cdot\|_{C}\right)$ be a Banach space and consider a continuous linear map

$$
\kappa:\left(H,\|\cdot\|_{\text {LAN }}\right) \rightarrow\left(C,\|\cdot\|_{C}\right) .
$$

Theorem 3.11.5 in [58] implies that the information lower bound for estimating the parameter $\kappa(f)$ is given by the Gaussian random variable $\mathcal{G}$ on $C$ with marginal distributions

$$
\begin{equation*}
T(\mathcal{G}) \sim N\left(0,\left\|\widetilde{\kappa}_{T}\right\|_{\mathrm{LAN}}^{2}\right), \quad T \in C^{*} \tag{114}
\end{equation*}
$$

where $\widetilde{\kappa}_{T}$ is the Riesz representer of the continuous linear map $T \circ \kappa:\left(H,\|\cdot\|_{\text {LAN }}\right) \rightarrow \mathbb{R}$. Note that $\widetilde{\kappa}_{T}$ necessarily lies in the completion $\bar{H}$ of $H$ for the $\|\cdot\|_{\text {LAN }}$-norm. In particular,

$$
\liminf _{\varepsilon \rightarrow 0} \inf _{\widehat{\kappa}: \mathcal{X} \mathcal{X}_{\varepsilon} \rightarrow C} \sup _{f} \varepsilon^{-2} E_{f}\|\widehat{\kappa}-\kappa(f)\|_{C}^{2} \geq E\|\mathcal{G}\|_{C}^{2}
$$

where the supremum extends over an $\varepsilon$-neighbourhood of $f$ in $H$. The above result holds whenever $\mathcal{G}$ is a tight Borel random variable on $C$. If $C=\mathbb{R}$ then the lower bound is simply given by $\|\widetilde{\kappa}\|_{\text {LAN }}^{2}$ where $\widetilde{\kappa}$ is the Riesz representer of the map $\kappa:\left(H,\|\cdot\|_{\text {LAN }}\right) \rightarrow \mathbb{R}$.

### 7.6. Some facts about weak convergence of random probability measures

The following result is given in [11, Lemma 2 of supplement].
Proposition 29. Let $P_{n}, P, n \in \mathbb{N}$, be random probability measures on $\mathbb{R}$. Suppose that for any real $t$, the Laplace transform $\int_{\mathbb{R}} e^{t x} d P(x)$ is finite almost surely and $\int_{\mathbb{R}} e^{t x} d P_{n}(x) \rightarrow \int_{\mathbb{R}} e^{t x} d P(x)$ in probability. Then for any metric $\beta$ for weak convergence of probability measures, $\beta\left(P_{n}, P\right) \rightarrow 0$ in probability.

As a consequence, we also deduce that by the 'Cramer-Wold device', if $P_{n}, P$ are random probability measures on $\mathbb{R}^{K}$ and $\int e^{\langle t, x\rangle} d P_{n}(x) \rightarrow \int e^{\langle t, x\rangle} d P(x)$ in probability for all $t \in \mathbb{R}^{K}$, then $\beta\left(P_{n}, P\right) \rightarrow 0$ in probability. Here $\langle\cdot, \cdot\rangle$ is the standard Euclidean inner product.

Since convergence in probability implies convergence almost surely along a subsequence, if $\beta_{S}\left(P_{n}, P\right) \rightarrow 0$ in probability on any separable metric space $S$, then any 'limiting' consequence of weak convergence such as the continuous mapping theorem,
convergence of moments or uniform convergence along classes of Borel sets or functions, also holds in probability, simply by arguing by contradiction and extracting subsequential almost sure limits. See the appendix in [9] or [11] for more details.

Acknowledgments. This research was supported by the European Research Council under ERC grant agreement UQMSI (No. 647812). I would like to thank Gabriel P. Paternain for many helpful discussions and Sven Wang for helpful remarks about Section 7.2. I am also grateful to two anonymous referees for valuable remarks and suggestions.

## References

[1] Abraham, K.: Nonparametric Bayesian posterior contraction rates for scalar diffusions with high-frequency data. Bernoulli 25, 2696-2728 (2019) Zbl 1428.62139 MR 4003562
[2] Adams, R. A., Fournier, J. J. F.: Sobolev Spaces. Elsevier/Academic Press, Amsterdam (2003) Zbl 1098.46001 MR 2424078
[3] Agapiou, S., Larsson, S., Stuart, A. M.: Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems. Stochastic Process. Appl. 123, 3828-3860 (2013) Zbl 1284.62289 MR 3084161
[4] Bal, G., Uhlmann, G.: Inverse diffusion theory of photoacoustics. Inverse Problems 26, art. 085010, 20 pp. (2010) Zbl 1197.35311 MR 2658827
[5] Bao, G., Li, P.: Inverse medium scattering problems for electromagnetic waves. SIAM J. Appl. Math. 65, 2049-2066 (2005) Zbl 1114.78006 MR 2177738
[6] Brown, L. D., Low, M. G.: Asymptotic equivalence of nonparametric regression and white noise. Ann. Statist. 24, 2384-2398 (1996) Zbl 0867.62022 MR 1425958
[7] Castillo, I.: On Bayesian supremum norm contraction rates. Ann. Statist. 42, 2058-2091 (2014) Zbl 1305.62189 MR 3262477
[8] Castillo, I.: Pólya tree posterior distributions on densities. Ann. Inst. H. Poincaré Probab. Statist. 53, 2074-2102 (2017) Zbl 1384.62156 MR 3729648
[9] Castillo, I., Nickl, R.: Nonparametric Bernstein-von Mises theorems in Gaussian white noise. Ann. Statist. 41, 1999-2028 (2013) Zbl 1285.62052 MR 3127856
[10] Castillo, I., Nickl, R.: On the Bernstein-von Mises phenomenon for nonparametric Bayes procedures. Ann. Statist. 42, 1941-1969 (2014) Zbl 1305.62190 MR 3262473
[11] Castillo, I., Rousseau, J.: A Bernstein-von Mises theorem for smooth functionals in semiparametric models. Ann. Statist. 43, 2353-2383 (2015) Zbl 1327.62302 MR 3405597
[12] Chung, K. L., Zhao, Z. X.: From Brownian Motion to Schrödinger's Equation. Springer, Berlin (1995) Zbl 0819.60068 MR 1329992
[13] Cotter, S. L., Dashti, M., Robinson, J. C., Stuart, A. M.: Bayesian inverse problems for functions and applications to fluid mechanics. Inverse Problems 25, art. 115008, 43 pp. (2009) Zbl 1228.35269 MR 2558668
[14] Cotter, S. L., Roberts, G. O., Stuart, A. M., White, D.: MCMC methods for functions: modifying old algorithms to make them faster. Statist. Sci. 28, 424-446 (2013) Zbl 1331.62132 MR 3135540
[15] Dashti, M., Harris, S., Stuart, A. M.: Besov priors for Bayesian inverse problems. Inverse Problems Imaging 6, 183-200 (2012) Zbl 1243.62032 MR 2942737
[16] Dashti, M., Law, K. J. H., Stuart, A. M., Voss, J.: MAP estimators and their consistency in Bayesian nonparametric inverse problems. Inverse Problems 29, art. 095017, 27 pp. (2013) Zbl 1281.62089 MR 3104933
[17] Dashti, M., Stuart, A. M.: The Bayesian approach to inverse problems. In: Handbook of Uncertainty Quantification, Vol. 1, R. Ghanem et al. (eds.), Springer, 311-428 (2017) MR 3839555
[18] Daubechies, I.: Ten Lectures on Wavelets. CBMS-NSF Reg. Conf. Ser. Appl. Math. 61, Amer. Math. Soc. (1992) Zbl 0776.42018 MR 1162107
[19] Dudley, R. M.: Real Analysis and Probability. Cambridge Univ. Press, Cambridge (2002) Zbl 1023.60001 MR 1932358
[20] Dudley, R. M.: Uniform Central Limit Theorems. 2nd ed., Cambridge Univ. Press, New York (2014) Zbl 1317.60030 MR 3445285
[21] Edmunds, D. E., Triebel, H.: Function Spaces, Entropy Numbers, Differential Operators. Cambridge Univ. Press, Cambridge (1996) Zbl 0865.46020 MR 1410258
[22] Freedman, D.: On the Bernstein-von Mises theorem with infinite-dimensional parameters. Ann. Statist. 27, 1119-1140 (1999) Zbl 0957.62002 MR 1740119
[23] Ghosal, S., Ghosh, J. K., van der Vaart, A. W.: Convergence rates of posterior distributions. Ann. Statist. 28, 500-531 (2000) Zbl 1105.62315 MR 1790007
[24] Ghosal, S., van der Vaart, A. W.: Fundamentals of Nonparametric Bayesian Inference. Cambridge Univ. Press, New York (2017) Zbl 1376.62004 MR 3587782
[25] Gilbarg, D., Trudinger, N. S.: Elliptic Partial Differential Equations of Second Order. Springer, Berlin (2001) Zbl 1042.35002 MR 1814364
[26] Giné, E., Nickl, R.: Rates of contraction for posterior distributions in $L^{r}$-metrics, $1 \leq r \leq \infty$. Ann. Statist. 39, 2883-2911 (2011) Zbl 1246.62095 MR 3012395
[27] Giné, E., Nickl, R.: Mathematical Foundations of Infinite-Dimensional Statistical Models. Cambridge Univ. Press, New York (2016) Zbl 1358.62014 MR 3588285
[28] Helin, T., Burger, M.: Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems. Inverse Problems 31, art. 085009, 22 pp. (2015) Zbl 1325.62058 MR 3377106
[29] Hohage, T., Weidling, F.: Verification of a variational source condition for acoustic inverse medium scattering problems. Inverse Problems 31, art. 075006, 14 pp . (2015) Zbl 1321.35162 MR 3366755
[30] Ibragimov, I. A., Khasminski, R. Z.: Statistical Estimation. Asymptotic Theory. Springer, New York (1981) Zbl 0467.62026 MR 0620321
[31] Kekkonen, H., Lassas, M., Siltanen, S.: Posterior consistency and convergence rates for Bayesian inversion with hypoelliptic operators. Inverse Problems 32, art. $085005,31 \mathrm{pp}$. (2016) Zbl 1390.35422 MR 3535664
[32] Knapik, B., Salomond, J.-B.: A general approach to posterior contraction in nonparametric inverse problems. Bernoulli 24, 2091-2121 (2018) Zbl 1419.62089 MR 3757524
[33] Knapik, B., van der Vaart, A. W., van Zanten, J. H.: Bayesian inverse problems with Gaussian priors. Ann. Statist. 39, 2626-2657 (2011) Zbl 1232.62079 MR 2906881
[34] de Laplace, P.-S. M.: Théorie analytiques des probabilités. Courcier, Paris (1812)
[35] Lassas, M., Saksman, E., Siltanen, S.: Discretization-invariant Bayesian inversion and Besov space priors. Inverse Problems Imaging 3, 87-122 (2009) Zbl 1191.62046 MR 2558305
[36] Le Cam, L.: Asymptotic Methods in Statistical Decision Theory. Springer, New York (1986) MR 0856411
[37] Lions, J.-L., Magenes, E.: Non-Homogeneous Boundary Value Problems and Applications. Vol. I. Springer, New York (1972) Zbl 0223.35039 MR 0350177
[38] Lu, Y.: On the Bernstein-von Mises theorem for high dimensional nonlinear Bayesian inverse problems. arXiv:1706.00289 (2017)
[39] Meyer, Y.: Wavelets and Operators. Cambridge Univ. Press, Cambridge (1992) Zbl 0776.42019 MR 1228209
[40] Monard, F., Nickl, R., Paternain, G. P.: Efficient Bayesian nonparametric inference for X-ray transforms. Ann. Statist. 47, 1113-1147 (2019) Zbl 1417.62060 MR 3909962
[41] Nickl, R.: Donsker-type theorems for nonparametric maximum likelihood estimators. Probab. Theory Related Fields 138, 411-449 (2007) Zbl 1113.60028 MR 2299714
[42] Nickl, R., Söhl, J.: Bernstein-von Mises theorems for statistical inverse problems II: compound Poisson processes. Electron. J. Statist. 13, 3513-3571 (2019) Zbl 1429.62168 MR 4013745
[43] Nickl, R., Söhl, J.: Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. Ann. Statist. 45, 1664-1693 (2017) Zbl 1411.62087 MR 3670192
[44] Papaspiliopoulos, O., Pokern, Y., Roberts, G. O., Stuart, A. M.: Nonparametric estimation of diffusions: a differential equations approach. Biometrika 99, 511-531 (2012) Zbl 06085152 MR 2966767
[45] Ray, K.: Bayesian inverse problems with non-conjugate priors. Electron. J. Statist. 7, 25162549 (2013) Zbl 1294.62107 MR 3117105
[46] Ray, K.: Adaptive Bernstein-von Mises theorems in Gaussian white noise. Ann. Statist. 45, 2511-2536 (2017) Zbl 1384.62158 MR 3737900
[47] Reiß, M.: Asymptotic equivalence for nonparametric regression with multivariate and random design. Ann. Statist. 36, 1957-1982 (2008) Zbl 1142.62023 MR 2435461
[48] Rivoirard, V., Rousseau, J.: Bernstein-von Mises theorem for linear functionals of the density. Ann. Statist. 40, 1489-1523 (2012) Zbl 1257.62036 MR 3015033
[49] Schwab, C., Stuart, A. M.: Sparse deterministic approximation of Bayesian inverse problems. Inverse Problems 28, art. 045003, 32 pp. (2012) Zbl 1236.62014 MR 2903278
[50] Stuart, A. M.: Inverse problems: a Bayesian perspective. Acta Numer. 19, 451-559 (2010) Zbl 1242.65142 MR 2652785
[51] Szabó, B., van der Vaart, A. W., van Zanten, H.: Frequentist coverage of adaptive nonparametric Bayesian credible sets. Ann. Statist. 43, 1391-1428 (2015) (with discussion) Zbl 1317.62040 MR 3357861
[52] Triebel, H.: Theory of Function Spaces. Monogr. Math. 78, Birkhäuser, Basel (1983) Zbl 0546.46027 MR 0781540
[53] Triebel, H.: Function Spaces and Wavelets on Domains. EMS Tracts in Math. 7, Eur. Math. Soc., Zürich (2008) Zbl 1158.46002 MR 2455724
[54] van de Geer, S. A.: Hellinger-consistency of certain nonparametric maximum likelihood estimators. Ann. Statist. 21, 14-44 (1993) Zbl 0779.62033 MR 1212164
[55] van de Geer, S. A.: Applications of Empirical Process Theory. Cambridge Univ. Press, Cambridge (2000) Zbl 0953.62049 MR 1739079
[56] van der Meulen, F., van Zanten, H.: Consistent nonparametric Bayesian inference for discretely observed scalar diffusions. Bernoulli 19, 44-63 (2013) Zbl 1259.62070 MR 3019485
[57] van der Vaart, A. W.: Asymptotic Statistics. Cambridge Univ. Press, Cambridge (1998) Zbl 0910.62001 MR 1652247
[58] van der Vaart, A. W., Wellner, J. A.: Weak Convergence and Empirical Processes. Springer Ser. Statistics, Springer, New York (1996) Zbl 0862.60002 MR 1385671
[59] Vollmer, S. J.: Posterior consistency for Bayesian inverse problems through stability and regression results. Inverse Problems 29, art. 125011, 32 pp. (2013) Zbl 1284.35064 MR 3141858

