Posterior contraction rates
for support boundary recovery

Markus Reiß
Institute of Mathematics
Humboldt-Universität zu Berlin
mreiss@math.hu-berlin.de

Johannes Schmidt-Hieber
Mathematical Institute
Leiden University
schmidtieberaj@math.leidenuniv.nl

Abstract

Given a sample of a Poisson point process with intensity \( \lambda_f(x,y) = n \mathbf{1}(f(x) \leq y) \), we study recovery of the boundary function \( f \) from a nonparametric Bayes perspective. Because of the irregularity of this model, the analysis is non-standard. We establish a general result for the posterior contraction rate with respect to the \( L^1 \)-norm based on entropy and one-sided small probability bounds. From this, specific posterior contraction results are derived for Gaussian process priors and priors based on random wavelet series.

MSC 2000 subject classification: 62C10; 62G05; 60G55
Key words: Frequentist Bayesian analysis, posterior contraction, Poisson point process, boundary detection, one-sided entropy, Gaussian prior, wavelet prior.

1 Introduction

We consider a support boundary detection model, where a Poisson point process (PPP) \( N \) on \([0,1] \times \mathbb{R}\) is observed with intensity

\[
\lambda(x,y) = \lambda_f(x,y) = n \mathbf{1}(f(x) \leq y).
\]

The statistical task is to recover the unobserved lower boundary \( f : [0,1] \to \mathbb{R} \) of the support of \( \lambda \), see the simulated data set in Figure 1. This boundary detection model can be seen as a
Figure 1: Simulated dataset (blue) and true boundary function (black).

continuous analogue of the nonparametric regression model with discrete equidistant design and exponential errors, that is, we observe \( Y_{i,n} = f(i/n) + \epsilon_{i,n}, \ i = 1, \ldots, n, \) and \( (\epsilon_{i,n}) \) are i.i.d. exponential random variables, cf. [15, 9]. Due to the one-sided error distribution, this model, with \( f \) in a parametric class, is not Hellinger differentiable and therefore irregular. Most of the nonparametric models that have been analysed from a frequentist Bayes point of view are asymptotically equivalent to a Gaussian shift experiment. Yet Poisson experiments form another important class of limit experiments [12], whose statistical structure is very different. The laws are not mutually absolutely continuous leading to a peculiar version of the Bayes formula and one-sided entropy conditions, subsequently. Moreover, the Hellinger distance is governed by the \( L^1 \)-distance between the boundary functions in contrast to the \( L^2 \)-theory in Gaussian shift models.

The goal of this article is to study posterior contraction for the support boundary detection model. We consider the \( L^1 \)-distance as loss function, which is linked to the information geometry of the model. Posterior contraction for the Hellinger loss is well-studied and can be reduced to conditions on the entropy of the parameter space and the small ball probability of the prior, cf. [4, 5]. We derive a modification of this result which is applicable for the support boundary detection model. Related to that, we show the following surprising result: If the posterior is restricted to functions that lie below the true function, then posterior contraction follows already from the behaviour of the one-sided small ball prior probability. In this case no bound on the entropy is necessary. On the contrary, for functions which lie above the true function, we essentially only require the entropy bound.

Given the general contraction result, we apply this to concrete classes of priors. In a first step, we study Gaussian priors and derive an analogue of the result in [19] for the support boundary detection model. We then study posterior contraction for random wavelet series priors with independent but not necessarily Gaussian random coefficients. For these
priors we derive a result on small ball probabilities, which is of independent interest. The corresponding contraction rates only match with the minimax estimation rates for one smoothness index. Below this critical smoothness the contraction rates can be improved if more heavy-tailed distributions on the wavelet coefficients are used. We also prove that truncated random wavelet series priors achieve the adaptive rates up to logarithmic factors. The companion paper [17] studies compound Poisson process priors for support boundary recovery. The focus of that article is on Bernstein-von Mises type theorems for function classes with increasing parameter dimension and frequentist coverage of credible sets.

Bayesian methods for irregular or boundary detection problems have attracted considerable attention especially because the MLE approach is often inefficient. [2] compares Bayes estimators with the MLE in a parametric model that is irregular. In [1] a Bernstein-von Mises theorem is derived for parameters which are on the boundary of the parameter space. The limit distribution consists in this case of Gaussian and exponentially distributed components. [10] considers posterior contraction around \( \vartheta \) given i.i.d. observations from a class of nonparametric densities of the form \( \eta(x - \vartheta) \) with \( \eta(y) = 0 \) for \( y < 0 \) and \( \eta(y) > 0 \) for \( y \geq 0 \). This can be viewed as a semiparametric, irregular model, where the nuisance parameter is the unknown distribution of the noise. For nonparametric models, [14] considers Bayesian methods for Poisson point processes, but does not cover boundary detection. In [13] a nonparametric Bayes approach is studied for detecting the boundary of an object in an image, assuming different distributions of the response variable inside and outside the object. This boundary detection model is regular and the likelihood ratios are always well-defined. The underlying information geometry is induced by the \( L^1 \)-norm, similar to our PPP model, but there is no different treatment necessary for the posterior on functions below or above the true function.

The paper is structured as follows. In Section 2 we derive a general result relating posterior contraction to entropy and small ball estimates. This result is then used in Section 3 to derive a criterion for posterior contraction under Gaussian priors. Section 4 studies wavelet expansion priors. Technicalities and proofs are deferred to an appendix.

Notation. We write \((x)_+ = \max(x, 0)\) and denote the indicator function of a set \(A\) by \(\mathbf{1}_A = \mathbf{1}(\cdot \in A)\). For \(p \in [1, \infty]\), \(\| \cdot \|_p\) denotes the \(L^p[0, 1]\)-norm. Inequalities for \(L^1\)-functions are assumed to hold almost everywhere. Let \([\beta]\) denote the largest integer strictly smaller than \(\beta > 0\). The \(\beta\)-Hölder norm is \(\|f\|_{C^\beta} := \sum_{j=1}^{[\beta]} \|f^{(j)}\|_\infty + \sup_{x \neq y} |f^{(\lfloor \beta \rfloor)}(x) - f^{(\lfloor \beta \rfloor)}(y)|/|x - y|^\beta - \lfloor \beta \rfloor|\). We denote by \(C^\beta(R)\) the class of functions \(f\) on \([0, 1]\) with \(\|f\|_{C^\beta} \leq R\). We further write \(N = \sum_i \delta_{(X_i, Y_i)}\) for a random point measure on \([0, 1] \times \mathbb{R}\) and often identify \(N\) with its support points \((X_i, Y_i)\). For two positive sequences \((a_n)_n, (b_n)_n\) we write \(a_n \lesssim b_n\) if there is a constant \(C\) such that \(a_n \leq Cb_n\) for all \(n\). If \(a_n \lesssim b_n\) and \(b_n \lesssim a_n\) then we write \(a_n \asymp b_n\).
2 General results on posterior contraction rates

2.1 Likelihood and Bayes formula

Before stating the main result on posterior contraction, we study the likelihood in the support boundary detection model. From that we derive expressions for the information distances and a specific form of the Bayes formula.

Denote by $P_f = P^n_f$ the distribution of a PPP with intensity measure $\Lambda_f(B) = \int_B \lambda_f$ for Borel sets $B$ in $[0, 1] \times \mathbb{R}$ with Lebesgue density $\lambda_f(x, y) = n1(f(x) \leq y)$, where $f$ is some function in $L^1([0, 1])$. The likelihood ratio $dP_f/dP_g$ is only defined for $g \leq f$, otherwise $P_g$ does not dominate $P_f$. The fact that the observation laws are not necessarily mutually absolutely continuous is a distinctive feature of support estimation problems and will play a major role in the analysis. Recall that for the Poisson point process $N$ its support points in $[0, 1] \times \mathbb{R}$ are denoted by $(X_i, Y_i)_{i \geq 1}$.

2.1 Lemma. For $g \leq f$ and $f, g \in L^1([0, 1])$, the likelihood ratio has the explicit form

$$\frac{dP_f}{dP_g} = \exp \left( n \int_0^1 (f - g)(x) \, dx \right) \cdot 1(\forall i : f(X_i) \leq Y_i). \quad (2.1)$$

The information geometry of the model is driven by the $L^1([0, 1])$-norm. Indeed, the Hellinger affinity is $\rho(P_f, P_g) = \int \sqrt{dP_f dP_g} = \exp(-\frac{n}{2} \|f - g\|_1)$. This implies for the squared Hellinger distance

$$H^2(P_f, P_g) = 2 - 2\rho(P_f, P_g) = 2 - 2 \exp \left( - \frac{n}{2} \|f - g\|_1 \right) \leq n \|f - g\|_1, \quad \forall f, g \in L^1([0, 1]).$$

Similarly, the Kullback-Leibler divergence satisfies $KL(P_f, P_g) = n \|f - g\|_1$ if $g \leq f$ and $KL(P_f, P_g) = \infty$ otherwise.

Since the likelihood requires the support boundaries to be in $L^1([0, 1])$, we consider as priors distributions $\Pi$ of stochastic processes $(X_t)_{t \in [0, 1]}$ on a Polish space $(\Theta, d)$ equipped with its Borel $\sigma$-algebra, which embeds continuously into $L^1([0, 1])$. We aim for a Bayes formula of the form

$$\Pi(B|N) = \frac{\int_B \frac{dP_f}{dP_{f_0}}(N) \, d\Pi(f)}{\int_{\Theta} \frac{dP_f}{dP_{f_0}}(N) \, d\Pi(f)}. \quad (2.2)$$

Since in the boundary detection model the likelihood ratio does not exist in general, the formula has to be modified. The next result provides a Bayes formula under the frequentist assumption that the data are generated under $P_{f_0}$.

4
2.2 Lemma. For $f_0 \in L^1([0,1])$, a prior $\Pi$ on the Polish space $\Theta$ and a Borel set $B \subset \Theta$, we have an explicit Bayes formula under the law $P_{f_0}$:

$$
\Pi(B|N) = \frac{\int_B e^{-n \int f(X_i) \leq Y_i} d\Pi(f)}{\int_{\Theta} e^{-n \int f(X_i) \leq Y_i} d\Pi(f)} = \frac{\int_B e^{-n \int (f_0 - f)_+} dP_{f_0}(N) d\Pi(f)}{\int_{\Theta} e^{-n \int (f_0 - f)_+} dP_{f_0}(N) d\Pi(f)} \text{ P}_{f_0}\text{-a.s.}
$$

The right-hand side is well-defined since $dP_{f_0} / dP_{f_0}$ exists. Compared to (2.2), the likelihood ratios are reweighted in the Bayes formula by a factor $e^{-n \int (f_0 - f)_+}$. In particular, for $f \leq f_0$ the integrands are equal to the deterministic values $e^{-n \int (f_0 - f)_+}$.

2.2 Main results

We start by stating the main theorem, which reduces posterior contraction to conditions on the entropy and small ball probabilities. The result is an analogue of the general contraction theorems in [4, 5]. Denote by $N(\varepsilon, \mathcal{F}, d)$ the $\varepsilon$-covering number of a metric space $\mathcal{F}$ with respect to the metric $d$.

2.3 Theorem. If for some $\Theta_n \subset \Theta$, some rate $\varepsilon_n \to 0$ and constants $C, C', C'' \geq 1, A > 0$

(i) $N(\varepsilon_n, \Theta_n, \| \cdot \|_{\infty}) \leq C'' e^{C' n \varepsilon_n}$;

(ii) $\Pi(f : \| f - f_0 \|_1 \leq A \varepsilon_n, f \leq f_0) \geq e^{-C n \varepsilon_n}$;

(iii) $\Pi(\Theta_n^c) \leq C'' e^{-(C+A+1)n \varepsilon_n}$,

then there exists a positive constant $M$ such that

$$
E_{f_0} \Pi(f : \| f - f_0 \|_1 \geq M \varepsilon_n | N) \leq 3C'' e^{-n \varepsilon_n}.
$$

Condition (i) can be relaxed to any of the conditions of Corollary 2.6

In condition (ii) we need a lower bound on the one-sided small ball probabilities. Applying triangle inequality and $\| \cdot \|_1 \leq \| \cdot \|_{\infty}$, a stronger version of (ii), which is often easier to verify, is given by

$$
(ii)' : \Pi(f : \| f + A \varepsilon_n / 2 - f_0 \|_{\infty} \leq A \varepsilon_n / 2) \geq e^{-C n \varepsilon_n}. \quad (2.3)
$$

The proof of the theorem is deferred to the appendix, yet main intermediate results are presented here. It will be convenient to establish posterior contraction for $\int (f_0 - f)_+$ and $\int (f - f_0)_+$ separately. Surprisingly, for posterior contraction with respect to $\int (f_0 - f)_+$ we only need the small ball estimate of the prior probability, but no bound on the entropy. In contrast, posterior contraction for $\int (f - f_0)_+$ only requires that (i) and (iii) of Theorem 2.3 hold.
2.4 Proposition. If for some constants $C > 0, A \geq 1$

$$\Pi\left(f : \int (f_0 - f) \leq A\varepsilon_n, f \leq f_0 \right) \geq e^{-Cn\varepsilon_n},$$

then

$$E_{f_0}\left[\Pi\left(f : \int (f_0 - f)_{+} \geq (1 + A + C)\varepsilon_n \left| N \right.\right)\right] \leq e^{-n\varepsilon_n}.$$ 

The one-sided small ball probability can be viewed as a prior mass condition on a Kullback-Leibler ball in view of $\{f : \text{KL}(P_{f_0}, P_f) \leq A\varepsilon_n n\} = \{f : \int (f_0 - f) \leq A\varepsilon_n, f \leq f_0 \}$. To establish posterior contraction with respect to the loss $\int (f - f_0)_{+}$, we need to understand the testing theory in the boundary detection model, which is non-standard due to the lack of absolute continuity in general. The Neyman-Pearson test $\varphi = 1(dP_g/dP_{f\wedge g} \geq dP_f/dP_{f\wedge g})$ behaves well for testing $f$ against $g$:

$$E_f\varphi + E_g[1 - \varphi] = \int \left(\frac{dP_f}{dP_{f\wedge g}} \wedge \frac{dP_g}{dP_{f\wedge g}}\right) dP_{f\wedge g} \leq \rho(P_f, P_g) = e^{-\frac{n}{2}\|f - g\|_1}.$$ 

Robustness with respect to the Hellinger distance, however, in the sense that for some $\alpha, \beta > 0$, and all $n$

$$E_f\varphi + \sup_{h : \|h - g\|_1 \leq \alpha\|f - g\|_1} E_h[1 - \varphi] \leq e^{-\beta n\|f - g\|_1}$$

holds, is violated: if $f \leq g$, we have $\varphi = 1(\forall i : g(X_i) \leq Y_i)$ and thus $E_h[1 - \varphi] = 1 - e^{-n\int (f - h)_{+}}$, which for general $h$ is much larger than $e^{-\beta n\|f - g\|_1}$. Under the additional assumption $h \geq g$, however, the type II error vanishes completely and we find for $f \leq g$

$$E_f\varphi \leq e^{-\frac{n}{2}\|f - g\|_1} \quad \text{and} \quad \sup_{h \geq g} E_h[1 - \varphi] = 0.$$ 

To control the posterior, it is therefore natural to use one-sided bracketing entropy. Consider a subset $\mathcal{F}$ of $L^1([0, 1])$. The one-sided bracketing number $N_1(\delta, \mathcal{F})$ is the smallest number $M$ of functions $\ell_1, \ldots, \ell_M \in L^1([0, 1])$ such that for any $f \in \mathcal{F}$ there exists $j \in \{1, \ldots, M\}$ with $\ell_j \leq f$ and $\int (f - \ell_j) \leq \delta$. For some function $f_0$ and integer $n$ consider the separation quantity

$$S_{f_0}(n, \mathcal{F}, f_0) = \inf_{(\ell_j)_{j \in J}} \sum_{j \in J} e^{-n\int (\ell_j - f_0)_{+} \in [0, \infty]},$$

where the infimum is taken over (not necessarily finite) subsets $J$ of the integers and functions $(\ell_j)_{j \in J} \subset L^1([0, 1])$ such that for any $f \in \mathcal{F}$ there exists $j \in J$ with $\ell_j \leq f$. In both definitions the functions $\ell_j$ are not required to be in $\mathcal{F}$.

In view of the next result, the quantity $S_{f_0}$, which can be seen as a weighted covering number, is the natural complexity measure for $\Theta$. 

6
2.5 Proposition. If $\Pi(f : f \leq f_0) > 0$, then for any Borel set $B \subseteq \Theta$

$$E_{f_0}[\Pi(f \in B|N)] \leq S_i(n, B, f_0).$$

Notice that the right-hand side does not depend on the prior. Weighted covering numbers might be small even for non-compact parameter spaces and have been used before in nonparametric Bayes theory, cf. [8], Section 4. For many specific problems, covering or bracketing numbers are sufficient and we can further upper bound the right-hand side in Proposition 2.5 using $-\int (\ell_j - f_0)_+ \leq -\int (f - f_0)_+ + \int (f - \ell_j)$ which implies that for any $\Theta_n \subseteq \Theta,$

$$S_i(n, \{f \in \Theta_n : \int (f - f_0)_+ \geq \epsilon\}, f_0) \leq e^{-n\epsilon/2}N(\epsilon/2, \Theta_n) \leq e^{-n\epsilon/2}N(\epsilon/4, \Theta_n, \|\cdot\|_\infty).$$

2.6 Corollary. Work under the assumptions of Proposition 2.5. If $C \geq 1$, then

$$E_{f_0}[\Pi(f \in \Theta_n : \int (f - f_0)_+ \geq 4C\epsilon_n|N)] \leq C''e^{-n\epsilon_n}$$

holds under any of the following conditions:

(i) $S_i(n, \{f \in \Theta_n : \int (f - f_0)_+ \geq 4C\epsilon_n\}, f_0) \leq C''e^{-n\epsilon_n};$

(ii) $N(\epsilon_n, \Theta_n) \leq C''e^{Cn\epsilon_n};$

(iii) $N(C\epsilon_n, \Theta_n, \|\cdot\|_\infty) \leq C''e^{Cn\epsilon_n}.$

We can avoid the entropy condition if we control instead the risk of an estimator. Indeed, $\inf_\varphi E_{\theta_0}\varphi + \sup_{\vartheta \in \Theta : \ell(\vartheta, \vartheta_0) \geq 2\epsilon} E_{\theta_0}[1 - \varphi] \leq 2 \inf_\varphi \sup_{\vartheta \in \Theta} P_\theta(\ell(\vartheta, \vartheta) \geq \epsilon)$ which follows by studying the test $\varphi = 1(\ell(\vartheta, \vartheta) \geq \epsilon)$ given an estimator $\vartheta.$ If the nonparametric MLE for $f$ exists, we have a particularly simple relation in the support boundary detection model between posterior contraction of $\int (f - f_0)_+$ and the excess probability of the MLE.

2.7 Lemma. Assume that $\Theta_n \subseteq \Theta$ contains $f_0$ and is closed under maxima, that is, if $f, g \in \Theta_n$ then $f \vee g \in \Theta_n.$ If the maximum likelihood estimator $\hat{f}^\text{MLE}$, based on the parameter space $\Theta_n$, exists, then

$$E_{f_0}[\Pi(f \in \Theta_n : \int (f - f_0)_+ > \epsilon_n|N)] \leq P_{f_0}(\int (\hat{f}^\text{MLE} - f_0)_+ > \epsilon_n). \quad (2.4)$$

As in the proof of Proposition 2.5 the upper bound is independent of the prior. It is well-known that posterior contraction with rate $\epsilon_n$ implies existence of a frequentist estimator with rate of convergence $\epsilon_n$, cf. Theorem 2.5 in [1]. Inequality (2.4) shows that also the other direction may hold, namely that convergence of an estimator implies posterior contraction with the same rate. Regarding the assumptions, a sufficient condition for the existence of the MLE is that $\Theta$ is closed under arbitrary maxima: $f_i \in \Theta, i \in I \Rightarrow \bigvee_{i \in I} f_i \in \Theta$, see the discussion in [18]. Examples of function spaces which are closed under the maximum are Hölder balls, monotone functions and convex functions.
3 Gaussian process priors

A common choice for nonparametric Bayes methods is to pick the distribution of a Gaussian process as prior probability measure. Given a Gaussian process prior $\Pi$, the seminal work in [19] relates posterior contraction to the small ball prior probability and approximation properties in the reproducing kernel Hilbert space (RKHS) generated by $\Pi$. The following result adapts Theorem 2.1 in [19] to our setting.

3.1 Theorem. Consider as prior the distribution of a Gaussian process $X$ with sample paths in the space $(C[0,1],\|\cdot\|_\infty)$. Write $\|\cdot\|_H$ for the RKHS-norm induced by the covariance operator of $X$. If $\varepsilon_n \geq n^{-1}$ and for all $n$

$$\inf_{h: \|h+2\varepsilon_n-f_0\|_\infty \leq \varepsilon_n} \|h\|_H^2 - \log \mathbb{P}(\|X\|_\infty \leq \varepsilon_n) \leq n\varepsilon_n,$$

(3.1)

then there exists a constant $M$ such that for all $n$

$$E_{f_0}[\Pi(f : \|f-f_0\|_1 \geq M\varepsilon_n|N)] \leq 3e^{-n\varepsilon_n}.$$

Condition (3.1) is slightly different compared to (1.2) and (1.3) in [19]. As a bound we have $n\varepsilon_n$ instead of $n\varepsilon_n^2$ and in the RKHS part there is an extra term $2\varepsilon_n$ which accounts for the one-sided prior mass condition in Theorem 2.3.

As an example let us study the Brownian motion prior with a random starting value. The prior is the law of the process $(X_0 + W_t)_{t \in [0,1]}$ with a Brownian motion $W$ and an independent standard normal random variable $X_0$. Let $f_0 \in C^\beta(R)$. Arguing as in [19], Section 4.1, we find for the corresponding RKHS norm $\|h\|_H^2 = \|h'\|_2^2 + h(0)^2$ and $\inf_{h: \|h+2\varepsilon_n-f_0\|_\infty \leq \varepsilon_n} \|h\|_H^2 \lesssim \varepsilon_n^{2-2/\beta}$ as well as for the small ball probabilities $\Pi(f : \|f\|_\infty \leq \varepsilon_n) \geq \mathbb{P}(|X_0| \leq \varepsilon_n/2)\mathbb{P}(\|W\|_\infty \leq \varepsilon_n/2) \gtrsim \varepsilon_n e^{-C/\varepsilon_n^2}$. The $L^1$-contraction rate is therefore

$$\begin{cases} n^{-\frac{\beta}{2+\beta}}, & \text{for } \beta \leq 1/2, \\ n^{-\frac{1}{2}}, & \text{for } \beta \geq 1/2. \end{cases}$$

This coincides with the minimax rate $n^{-\beta/(\beta+1)}$ if $\beta = 1/2$. For $\beta > 1/2$, we do not gain anymore in the contraction rate by imposing more smoothness on the signal. For $\beta < 1/2$ the rate is slower than the minimax rate. This behaviour of the posterior for Gaussian priors is well-known in nonparametric Bayes theory.
4 Wavelet expansion priors

Series expansions provide another natural way to construct priors on function spaces. We study process priors \((X_t)_{t \in [0,1]}\) which admit an expansion in a wavelet basis \((\psi_{jk})\):

\[
X_t = \sum_{j,k} d_{j,k} \xi_{j,k} \psi_{j,k}(t) \quad \text{in} \quad L^2[0,1].
\] (4.1)

Here, \(d_{j,k}\) are real numbers and \(\xi_{j,k}\) are i.i.d. random variables with Lebesgue density \(f_\xi\). As a prior on the function \(f\) this means that each wavelet coefficient of \(f\) is drawn independently from the distribution of \(d_{j,k} \xi_{j,k}\).

For convenience, we restrict ourselves in this section to \(s\)-regular, boundary corrected and compactly supported wavelet bases \((\psi_{jk})\) in \(L^2(\mathbb{R})\) as constructed in Section 4 of [3].

Wavelet expansion priors have been studied in different nonparametric models with uniform random variables \(\xi_{j,k}\), cf. [6, 16]. Moreover, [20] derives bounds on the small ball probabilities of Gaussian processes of the form (4.1). Below, we derive posterior contraction rates for a class of distributions \(\xi_{j,k}\).

To start with, we prove the following general lower bound on small ball probabilities, which is of independent interest.

4.1 Lemma. Assume (4.1) with a symmetric and unimodal density \(f_\xi\) and \(|d_{j,k}| \asymp 2^{-\frac{j}{2}(2\alpha+1)}\) for some \(\alpha > 0\). Suppose further that there are constants \(L\) and \(\delta > 0\) such that

\[
\mathbb{E}[|\xi_{j,k}|^{(1+\delta)/\alpha}] \leq L.
\]

Then for all \(\beta \in (0, s], R > 0\) there exists a constant \(D > 0\) such that

\[
\inf_{h \in C^s(R)} \mathbb{P}(\|X - h\|_{\infty} \leq \varepsilon) \geq f_\xi(D\varepsilon^{-(\alpha-\beta)/\beta}) D\varepsilon^{-1/(\alpha \wedge \beta)} \quad \text{for all} \quad 0 < \varepsilon \leq 1.
\]

For \(\beta \geq \alpha\) the lower bound has the form \(C^{-\varepsilon^{-1/\alpha}}\) with \(C = f_\xi(D)^{-1}\). For \(\beta < \alpha\) the lower bound depends on the tails of the distribution: heavier tails lead to larger lower bounds on the small ball probabilities and in consequence to better contraction rates. The fastest contraction rate that can be obtained using the small ball estimate in Lemma 4.1 and Theorem 2.3 is \(\varepsilon_n = n^{-\alpha/(1+\alpha)}\), which is the solution of the equation \(\varepsilon_n^{-1/\alpha} = n\varepsilon_n\).

4.2 Theorem. Consider the process in (4.1) as prior with a symmetric and unimodal density \(f_\xi\) and \(|d_{j,k}| \asymp 2^{-\frac{j}{2}(2\alpha+1)}\) for some \(\alpha > 1\). Suppose \(f_\xi(x) \leq \gamma^{-1} e^{-\gamma|x|}\) for some \(\gamma > 0\) and all \(x \in \mathbb{R}\) and fix \(\beta \in (0, s], R > 0\). For any sequence \(\varepsilon_n \to 0\), satisfying

\[
\frac{1+(\alpha \wedge \beta)}{n\varepsilon_n^\alpha \beta} \asymp -\log f_\xi(D\varepsilon_n^{-(\alpha-\beta)/\beta}),
\]


there exist positive constants $M$ and $c$ such that for all $n$
\[
\sup_{f_0 \in C^\beta(R)} E_{f_0} \left[ \mathbb{P}\left( \|f - f_0\|_1 \geq M \varepsilon_n | N \right) \right] \leq e^{-cn \varepsilon_n}.
\]

The assumption $\alpha > 1$ is imposed in order to bound the bracketing entropy in Theorem 2.3. One of the consequences of Theorem 4.2 is that the posterior contracts faster in the regime $\beta < \alpha$ if heavier-tailed distributions are used. This is illustrated by two specific examples.

4.3 Example.

(a) If $\xi_{j,k} \sim \mathcal{N}(0,1)$, we find for a sufficiently large constant $C$
\[
\inf_{h \in C^\beta(R)} \mathbb{P}\left( \|X - h\|_\infty \leq \varepsilon \right) \geq \exp\left( -C \varepsilon^{-(1+2\alpha-2\beta \vee 1)} \right).
\]

For $\alpha = 1/2$, the bound becomes $\exp(-C \varepsilon^{-2(1+\beta \vee 1)})$, which is the same as for the Brownian motion prior. The resulting posterior contraction rate is
\[
\varepsilon_n = n^{-\frac{\beta \wedge \alpha}{1+\alpha+\beta}}.
\]

(b) If $\xi_{j,k}$ follows a Laplace (double-exponential) distribution, we obtain
\[
\inf_{h \in C^\beta(R)} \mathbb{P}\left( \|X - h\|_\infty \leq \varepsilon \right) \geq \exp\left( -C \varepsilon^{-(1+\alpha-\beta \vee 1)} \right).
\]

The posterior contraction rate becomes $\varepsilon_n = n^{-\frac{\beta \wedge \alpha}{1+\alpha}}$ which improves the rate in (a) for the case $\beta < \alpha$.

We can also obtain a fully adaptive result (up to log $n$ factors) using a random truncation of the wavelet expansion prior. The prior can be realized via a hierarchical construction. In a first step, we draw the maximal resolution level $J$ from a distribution satisfying $P(J = j) \propto 2^{-j}$. Given $J$, generate
\[
X_t = \sum_{j \leq J, k} \xi_{j,k} \psi_{j,k}(t) \quad (4.2)
\]
with $\psi_{j,k}$ as in (4.1) and $(\xi_{j,k})_{j,k}$ an i.i.d. sequence of random variables with positive and continuous Lebesgue density $f_\xi$. In this prior the regularization is induced by the truncation of the wavelet series and compared with (4.1) we can set $d_{j,k} = 1$.

4.4 Lemma. Consider the random truncation prior (4.2). For $\beta \in (0,s]$, $R > 0$ there exists a constant $D > 0$ such that
\[
\inf_{h \in C^\beta(R)} \mathbb{P}\left( \|X - h\|_\infty \leq \varepsilon \right) \geq \varepsilon D \varepsilon^{-1/\beta} \quad \text{for all} \quad 0 < \varepsilon \leq 1.
\]
4.5 Theorem. Consider the random truncation prior \([4,2]\). Suppose \(f_\varepsilon(x) \leq \gamma^{-1}e^{-\gamma|x|}\) for some \(\gamma > 0\) and all \(x \in \mathbb{R}\) and fix \(\beta \in (0, s], R > 0\). Then there exist constants \(M\) and \(c\) such that for all \(n\)
\[
\sup_{f_\varepsilon \in C^0(R)} E_{f_\varepsilon} \left[ \Pi \left( f : \|f - f_\varepsilon\|_1 \geq M\varepsilon_n | N \right) \right] \leq e^{-cn\varepsilon_n}
\]
with
\[
\varepsilon_n = \left( \frac{\log n}{n} \right)^{\frac{\beta}{\beta + 1}}.
\]

A key ingredient of the proof is a connection to Besov spaces which allows us to obtain entropy bounds.

5 Appendix

Proof of Lemma 2.1. The general change of measure formula for two Poisson point processes (PPPs) on \(\mathcal{X}\) with finite intensity measures \(\Lambda_1 \ll \Lambda_2\) is given by
\[
\frac{dQ_{\Lambda_1}}{dQ_{\Lambda_2}}(N) = \exp \left( \int_{\mathcal{X}} \log \frac{d\Lambda_1}{d\Lambda_2}(x) dN(x) - \Lambda_1(\mathcal{X}) + \Lambda_2(\mathcal{X}) \right),
\]
where \(\log 0 := -\infty, \exp(-\infty) := 0\). Notice that \(P_f\) and \(P_g\) have infinite intensity. We therefore apply the following decomposition first. For two functions \(f, g\) with finite intensity and we can decompose it in two independent PPPs with distribution \(P_{f,vg}\) and \(P_{g,vg}\), that is, \(P_f = P_{f,vg} \otimes P_{g,vg}\). Similarly, \(P_g = P_{g,vF} \otimes P_{f,vg}\). The PPPs \(P_{f,vg}\) and \(P_{g,vg}\) have finite intensity and we can apply \(\mathcal{X} = [0,1] \times \mathbb{R}, \Lambda_1(A) = \int_A \lambda_{f,vg}(x,y)dxdy\) and \(\Lambda_2(A) = \int_A \lambda_{g,vg}(x,y)dxdy\).

With \(N_{f,vg} := \sum_{x \in \mathcal{X} : f(x,y) \neq 0} Y_i \delta_{(X_i,Y_i)}\), this gives
\[
\frac{dP_f}{dP_g}(N) = \frac{dP_{f,vg}}{dP_{g,vg}}(N_{f,vg})
\]
\[
= \exp \left( \sum_{x \in \mathcal{X} : f(x,y) \neq 0} \frac{n1(Y_i \geq f(X_i))}{n1(Y_i \geq g(X_i))} - n \int (f \vee g - f) + n \int (f \vee g - g) \right)
\]
\[
= e^{n \int (f-g)1(\forall i: Y_i \geq f(X_i))}.
\]

Proof of Lemma 2.2. Let \(f_0 \in L^1([0,1])\) be fixed. Consider a PPP with a strictly positive intensity \(\lambda^\ast : [0,1] \times \mathbb{R} \to (0, \infty)\) satisfying
\[
\lambda^\ast(x,y) = n \text{ for } x \in [0,1], y \geq f_0(x) \text{ and } \Lambda^\ast \{ y < f_0(x) \} = \int_0^1 \int_{f_0(x)}^{f_0(x)} \lambda^\ast(x,y)dydx < \infty,
\]
and denote by $P^*$ its distribution. For any $f \in L^1([0,1])$ we can decompose the process into two independent PPP with intensities $\lambda^*_{f,1}(x,y) = \lambda^*(x,y)1(y < f \lor f_0(x))$ and $\lambda_f \wedge f_0(x,y) = n1(y \geq f \lor f_0(x))$. If $P^*_{f,1}$ and $P^*_{f \lor f_0}$ denote the corresponding laws, then we can decompose $P^* = P^*_{f,1} \otimes P^*_{f \lor f_0}$. Similarly, we can decompose the distribution $P_f$ via $P_f = P^*_{f,1} \otimes P^*_{f \lor f_0}$, where $P^*_{f,1}$ denotes the PPP distribution with intensity $\lambda^*_{f,1}(x,y) = n1(f(x) \leq y \leq f_0(x))$. Finally, for a PPP $N$ we write $N = N_{< f \lor f_0} + N_{\geq f \lor f_0}$ with $N_{< f \lor f_0} = \sum_{i:Y_i < f \lor f_0(X_i)} \delta(X_i,Y_i)$ and $N_{\geq f \lor f_0} = \sum_{i:Y_i \geq f \lor f_0(X_i)} \delta(X_i,Y_i)$. Then, using (5.1),

$$
\frac{dP_f}{dP_\infty}(N) = \frac{dP^*_{f,1}}{dP^*_{f,1}}(N_{< f \lor f_0})
= e^{-n\int(f_0-f)_+ + \lambda^*(\{y < f_0(x)\}) + n\int(f-f_0)_+ + \lambda^*_{f,1}(X_i,Y_i)}
= e^n\int(f-f_0)_+ + \lambda^*(\{y < f_0(x)\}) 1(\forall i : f(X_i) \leq Y_i)
$$

where products over empty index sets are set to one. Now, notice that under $P_0$ we have $Y_i \geq f_0(X_i)$ and thus $\lambda^*(X_i,Y_i) = n$ a.s. such that

$$
\Pi(B|N) = \frac{\int_B \frac{dP_f}{dP_\infty}(N)d\Pi(f)}{\int \frac{dP_f}{dP_\infty}(N)d\Pi(f)} = \frac{\int_B e^{n\int(f-f_0)_+} 1(\forall i : f(X_i) \leq Y_i)d\Pi(f)}{\int e^{n\int(f-f_0)_+} 1(\forall i : f(X_i) \leq Y_i)d\Pi(f)} P_0\text{-a.s.}
$$

Under $P_0$ we have $1(\forall i : f(X_i) \leq Y_i) = 1(\forall i : f \lor f_0(X_i) \leq Y_i)$ a.s. and (2.1) yields

$$
H(f) := e^n\int(f-f_0)_+ 1(\forall i : f \lor f_0(X_i) \leq Y_i) = e^{-n\int(f_0-f)_+} \frac{dP_{f \lor f_0}}{dP_f}(N),
$$

which completes the proof.

**Proof of Proposition 2.4.** Consider $H(f)$ from (5.2). By Lemma 2.2 under $P_0$

$$
\Pi(B|N) = \frac{\int_B H(f)d\Pi(f)}{\int H(f)d\Pi(f)} \leq e^{A\varepsilon_n} \frac{\int_B H(f)d\Pi(f)}{\Pi(f : \|f - f_0\|_1 \leq A\varepsilon_n \& f \leq f_0)},
$$

where we used $1(\forall i : f(X_i) \leq Y_i) = 1 P_0\text{-a.s.}$ for all $f \leq f_0$. With $B := \{f : \int(f_0-f)_+ \geq (1 + A + C)\varepsilon_n\}$ and $E_{f_0}[H(f)] = e^{-n\int(f_0-f)_+} \leq e^{-(1+A+C)\varepsilon_n}$ for $f \in B$, we obtain $E_{f_0}[\Pi(B|N)] \leq e^{-n\varepsilon_n}$.

**Proof of Proposition 2.5.** For functions $(\ell_j)_{j \in J}$, eligible in the definition of $S_\varepsilon(n, B, f_0)$, consider the test $\varphi_n = 1(3 \exists i \forall j : \ell_j(X_i) \leq Y_i)$. This test satisfies under the hypothesis $f_0$

$$
P_{f_0}(\varphi_n = 1) \leq \sum_{j \in J} P_{f_0}(\forall i : \ell_j(X_i) \leq Y_i) = \sum_{j \in J} e^{-n\int(\ell_j-f_0)_+}.
$$
By assumption and $\sigma$-continuity of $\Pi$, there exist $R > 0$ and $\delta > 0$ such that $\Pi(f : \int f \geq -R, f \leq f_0) \geq \delta$. Thus, we use formula (5.3) and bound the posterior by

$$\Pi(B|N) \leq \varphi_n + \left[ \int_B H(f)(1 - \varphi_n) d\Pi(f) \right] \leq \varphi_n + \delta^{-1} e^{nR+n} \int f_0 \int_B H(f)(1 - \varphi_n) d\Pi(f).$$

Since for $f \in B$ there is an $\ell_j \leq f$, we infer

$$H(f)(1 - \varphi_n) = e^{n(f - f_0)} 1(\forall i : f(X_i) \leq Y_i) 1(\forall j : \ell_j(X_i) > Y_i) = 0.$$

Therefore,

$$E_{f_0}[\Pi(B|N)] \leq E_{f_0}[\varphi_n] \leq \sum_{j \in J} e^{-n f(\ell_j - f_0)} +$$

and the claim follows by taking the infimum over all possible $(\ell_j)$.

**Proof of Theorem 2.3.** By Proposition 2.4 and Proposition 2.5 it remains to show $E_{f_0}[\Pi(\Theta_n^c|N)] \leq C'' e^{-n\varepsilon_n}$. By (5.3), $E_{f_0}[H(f)] \leq 1$ and condition (i) and (iii),

$$E_{f_0}[\Pi(\Theta_n^c|N)] \leq \frac{e^{nA_n} \Pi(\Theta_n^c)}{\Pi(f : \|f - f_0\|_1 \leq A\varepsilon_n, f \leq f_0)} \leq C'' e^{-n\varepsilon_n}$$

follows, which is the claim.

**Proof of Lemma 2.7.** The key observation is that we can restrict the posterior to $\{ f \leq \hat{f}_{MLE} \}$ because otherwise the likelihood is zero. To see this, note that $\forall i : f(X_i) \leq Y_i$ implies $f \leq \hat{f}_{MLE}$ because otherwise $f \lor \hat{f}_{MLE} \in \Theta_n$ would have a larger likelihood than $\hat{f}_{MLE}$. Observe that then $\int (f - f_0)_+ \leq \int (\hat{f}_{MLE} - f_0)_+$ such that

$$E_{f_0}\left[ \Pi\left( f \in \Theta_n : \int (f - f_0)_+ > \varepsilon_n | N \right) \right] \leq E_{f_0}\left[ \Pi\left( f \in \Theta_n : \int (\hat{f}_{MLE} - f_0)_+ > \varepsilon_n | N \right) \right]$$

$$= P_{f_0}\left( \int (\hat{f}_{MLE} - f_0)_+ > \varepsilon_n \right),$$

where the last equality holds because $\{ \int (\hat{f}_{MLE} - f_0)_+ > \varepsilon_n \}$ is independent of $f$.

## 6 Proofs for Section 3

We state Theorem 2.1 of [19] in a slightly more general form.
6.1 Theorem (Theorem 2.1 of [19]). Let \( X \) be a Borel-measurable, zero-mean Gaussian random element in the Banach space \((B, \| \cdot \|)\) with RKHS \((H, \| \cdot \|_{H})\) and let \( f \) be contained in the closure of \( H \) in \( B \). For any \( \varepsilon_{n} > 0 \) and \( \gamma_{n} \geq 1 \), satisfying
\[
\inf_{h: \| h-f \| \leq \varepsilon_{n}} \| h \|_{H}^{2} - \log P(\| X \| \leq \varepsilon_{n}) \leq \gamma_{n}
\]
and for any \( C_{*} \geq 1 \), there exists a Borel set \( B_{n} \subset B \) such that
\[
\log N(3\varepsilon_{n}, B_{n}, \| \cdot \|) \leq 6C_{*}\gamma_{n}, \quad P(X \notin B_{n}) \leq e^{-C_{*}\gamma_{n}}, \quad \text{and} \quad P(\| X - f \| \leq 2\varepsilon_{n}) \geq e^{-\gamma_{n}}.
\]
Proof. Replace \( n\varepsilon_{n}^{2} \) in the proof of Theorem 2.1 of [19] by \( \gamma_{n} \); in particular \( M_{n} := -2\Phi^{-1}(e^{-C_{*}\gamma_{n}}) \). For the final argument of the proof observe that \( e^{-C_{*}\gamma_{n}} < 1/2 \) due to \( C_{*}\gamma_{n} \geq 1 \).

Proof of Theorem 3.1. We apply Theorem 6.1 with \((B, \| \cdot \|) = (C[0,1], \| \cdot \|_{\infty})\), \( \gamma_{n} = n\varepsilon_{n} \) and \( C_{*} = 6 \). This shows that there exists \( \Theta_{n} \) such that \( \log N(3\varepsilon_{n}, \Theta_{n}, \| \cdot \|_{\infty}) \leq 36n\varepsilon_{n}, \Pi(\Theta_{n}^{\varepsilon}) \leq e^{-6\varepsilon_{n}} \) and \( \Pi(f : \| f + 2\varepsilon_{n} - f_{0} \|_{\infty} \leq 2\varepsilon_{n}) \geq e^{-n\varepsilon_{n}} \). Use (2.3) and (iii) in Corollary 2.6. Thus, the assumptions of Theorem 2.3 are satisfied with \( A = 4, C = C'' = 1, C' = 36 \).

7 Proofs for Section 4

Proof of Lemma 4.1. Write \( h = \sum_{j,k} h_{j,k}\psi_{j,k} \). Since \( \psi \) is a compactly supported wavelet, \( \| X - h \|_{\infty} \leq C \sum_{j} 2^{j/2} \max_{k} |d_{j,k}\xi_{j,k} - h_{j,k}| \) for a sufficiently large constant \( C \). By assumption \( \psi \) is moreover \( s \)-regular and \( h \in C^{\beta}(R) \) with \( \beta \leq s \). Using Theorem 4.4 in [3], we can find constants \( 0 < q < Q < \infty \) such that \( q2^{-\frac{1}{q}(2\alpha+1)} \leq |d_{j,k}| \leq Q2^{-\frac{1}{q}(2\alpha+1)} \) and \( |h_{j,k}| \leq Q2^{-\frac{1}{q}(2\beta+1)} \) and obtain for any \( J \),
\[
\| X - h \|_{\infty} \leq CQ \left( \sum_{j\leq J} 2^{-j\alpha} \max_{k} |\xi_{j,k} - h_{j,k}/d_{j,k}| + \sum_{j>J} 2^{-j\alpha} \max_{k} |\xi_{j,k}| + \sum_{j>J} 2^{-j\beta} \right).
\]
Let \( J_{*} \) be the smallest integer such that
\[
CQ \left( 2^{-J_{*}\alpha} \sum_{r=0}^{J_{*}} 2^{-ar} + L^{(1+\delta)}2^{-J_{*}\alpha} \sum_{r\geq 1} 2^{-r\alpha/(2+\delta)} + 2^{-J_{*}\beta} \sum_{r\geq 0} 2^{-r\beta} \right) \leq \varepsilon,
\]
which yields \( 2^{J_{*}} \asymp e^{-1/(a \wedge \delta)} \) as \( \varepsilon \to 0 \). Introduce the events \( G_{\leq} = \{ |\xi_{j,k} - h_{j,k}/d_{j,k}| \leq 2(j-J_{*})2\alpha \} \) and \( G_{>} = \{ |\xi_{j,k}| \leq L^{(1+\delta)}2(j-J_{*})\alpha/(1+\delta/2) \} \) for \( j > J_{*} \) then, thanks to the choice of \( J_{*} \), on \( G_{\leq} \cap G_{>} \) we have \( \sup_{t \in [0,1]} |X_{t} - h(t)| \leq \varepsilon \). Thus,
\[
\mathbb{P}(\| X - h \|_{\infty} \leq \varepsilon) \tag{7.1}
\]
for a sufficiently large constant $R' \geq 1$. Since the random variables $\xi_{j,k}$ are symmetric and have a unimodal density, we have $f_\xi(x) \leq 1/2$ for $x \geq 1$ as well as $\mathbb{P}(|\xi_{j,k} - h_{j,k}/d_{j,k}| \leq 2(j-J_*) \alpha/2) \geq 2(j-J_*) \alpha/2$. On the $j$-th product level there are at most $A2^j$ wavelet coefficients with $A$ some positive constant. The first product in (7.1) can therefore be bounded from below by

$$\prod_{j \leq J_*} f_\xi(R'2^{J_*(\alpha-\beta)_+} A2^j (2(j-J_*) \alpha/2)) \geq f_\xi(R'2^{J_*(\alpha-\beta)_+} A2^j \prod_{k=1}^{J_*} 2^{-2\alpha k^2 - k} A2^{J_*}$$

for a sufficiently large constant $K$. To find a lower bound of the second product in (7.1), observe that

$$\mathbb{P}(| \xi_{j,k} | \leq L^{\alpha/(1+\delta)} 2^{(j-J_*) \alpha/(1+\delta/2)}) \geq 1 - \mathbb{P}(2^{(j-J_*) \alpha/(1+\delta/2)}) \geq 1 - 2^{(J_*-j)(1+\delta)/(1+\delta/2)}.$$

For any fixed $j > J_*$, using $(1 + \delta)/(1 + \delta/2) = 1 + \delta/(2 + \delta)$ and the elementary inequality $1 - y \geq e^{-2y}$ for $0 \leq y \leq 1/2$,

$$\prod_k \mathbb{P}(| \xi_{j,k} | \leq L2^{(j-J_*) \alpha/(1+\delta/2)}) \geq \left(1 - 2^{(J_*-j)(1+\delta)/(1+\delta/2)}\right)^{A2^j} \geq \exp \left(- A2^{J_*+1} 2^{(J_*-j)\delta/(2+\delta)}\right).$$

This implies that the product $\prod_{j > J_*} \mathbb{P}(| \xi_{j,k} | \leq L^{\alpha/(1+\delta)} 2^{(j-J_*) \alpha/(1+\delta/2)})$ can be bounded from below by

$$\prod_{j > J_*} \exp \left(- A2^{J_*+1} 2^{(J_*-j)\delta/(2+\delta)}\right) = \exp \left(- A2^{J_*+1} \sum_{k \geq 1} 2^{-k\delta/(2+\delta)}\right) \geq \exp(-R''2^{J_*})$$

for a sufficiently large constant $R''$. Recall that $2^{J_*} \asymp e^{-1/\alpha \wedge \beta}$. Because of $f_\ell(x) \leq 1/2$ for $|x| \geq 1$ we have $K^{-2^{J_*}} \exp(-R''2^{J_*}) \geq f_\xi(K' e^{-(\alpha-\beta)_+}) K^{2^{J_*}}$ for a sufficiently large constant $K'$. The result follows therefore from (7.1), (7.2), and (7.3). \hfill \Box

7.1 Lemma. Suppose that for some $\gamma > 0$, $f_\xi(x) \leq \gamma^{-1} e^{-\gamma |x|}$ for all $x \in \mathbb{R}$. Let $\xi_1, \ldots, \xi_m \sim f_\xi$, independently. Then,

$$\mathbb{P}\left(\frac{1}{m} \sum_{j=1}^{m} | \xi_j | \geq 2\gamma^{-1} (t + \log(4/\gamma^2))\right) \leq e^{-tm}.$$
Proof. Set \( A(\gamma, t) := 2\gamma^{-1}(t + \log(4/\gamma^2)) \) and denote by \( f_\xi \) the density of \( |\xi| \). By assumption, \( f_\xi(x) \leq 2\gamma^{-1}e^{-\gamma x} \) holds for all \( x \geq 0 \) such that \( E[e^{\gamma|\xi|/2}] \leq 4/\gamma^2 \). Therefore, we deduce by Markov’s inequality

\[
P\left( \sum_{j=1}^{m} |\xi_j| \geq A(\gamma, t)m \right) = P\left( \exp\left( \frac{\gamma}{2} \sum_{j=1}^{m} |\xi_j| \right) \geq \exp\left( \frac{\gamma}{2} A(\gamma, t)m \right) \right) 
\leq \exp\left( m \log \left( \frac{4}{\gamma^2} \right) - \frac{\gamma}{2} A(\gamma, t)m \right) = e^{-tm}.
\]

\( \square \)

Proof of Theorem 4.2. It is enough to prove the result for all \( n \geq n_0 \) with \( n_0 \) a fixed integer. We verify the conditions (i) – (iii) of Theorem 2.3.

(i) To check the first condition, pick \( J_n \) such that \( \varepsilon_n^{1/\alpha} \leq 2J_n < 2\varepsilon_n^{-1/\alpha} \). For a constant \( K \), which will be chosen later to be large enough, define

\[
\Theta_n = \left\{ g = \sum_{j,k} \vartheta_{j,k} \psi_{j,k} : \sum_{j \leq J_n} 2^{s(2\alpha + 1)} \sum_k |\vartheta_{j,k}| \leq K2^{J_n}, \max_{j > J_n} 2^{\frac{s}{2}(2\alpha - 1)} \sum_k |\vartheta_{j,k}| \leq K \right\}.
\]

Denote by

\[
B_{p,q}^s(M) := \left\{ g = \sum_{j,k} \vartheta_{j,k} \psi_{j,k} : \left( \sum_j 2^{j(s+\frac{1}{2}-\frac{1}{p})} \left( \sum_k |\vartheta_{j,k}|^p \right)^{\frac{1}{p}} \right)^{\frac{1}{q}} \leq M \right\}
\]

the Besov \( B_{p,q}^s \)-ball with radius \( M \) and apply the usual modifications for \( p = \infty \) and \( q = \infty \). To bound the bracketing entropy of \( \Theta_n \), observe that \( \Theta_n \subseteq B_{1,1}^{\alpha+1}(K2^{J_n}) + B_{1,\infty}^\alpha(K) \) where the sum is the elementwise addition. By Theorem 4.3.36 in [7], there exists a constant \( C \) such that

\[
\log \mathcal{N}(\delta, B_{p,q}^s(M), \| \cdot \|_\infty) \leq C(M/\delta)^{1/s} \text{ if } s > 1/p.
\]

Since \( 2J_n \approx \varepsilon_n^{-1/\alpha} \) and \( \alpha > 1 \), this gives

\[
\log \mathcal{N}(C\varepsilon_n, \Theta_n, \| \cdot \|_\infty) \leq \log \mathcal{N}(C\varepsilon_n/2, B_{1,1}^{\alpha+1}(K2^{J_n}), \| \cdot \|_\infty) + \log \mathcal{N}(C\varepsilon_n/2, B_{1,\infty}^\alpha(K), \| \cdot \|_\infty)
\leq (C\varepsilon_n)^{-1/\alpha}.
\]

Notice that \( \varepsilon_n \gtrsim n^{-\alpha/(1+\alpha)} \). Making the constant \( C \) big enough, we therefore obtain

\[
\mathcal{N}(C\varepsilon_n, \Theta_n, \| \cdot \|_\infty) \leq e^{Cn\varepsilon_n}.
\]

(ii) We apply Lemma 4.1 to (2.3). Since \( \varepsilon_n \to 0 \), \( f_0 - \varepsilon_n \in C^\beta(R + 1) \) for sufficiently large \( n \) and

\[
\Pi \left( f : \| f + \varepsilon_n - f_0 \|_\infty \leq \varepsilon_n \right) \geq f_\xi \left( D\varepsilon_n^{-\alpha/\beta} / \beta \right) D\varepsilon_n^{-1/\alpha,\beta} \geq e^{-Cn\varepsilon_n}.
\]

(iii) We bound \( \Pi(\Theta^c_n) \). Recall that \( X = \sum_{j,k} d_{j,k} \xi_{j,k} \psi_{j,k} \) and \( |d_{j,k}| \leq Q2^{-\frac{s}{2}(2\alpha+1)} \) for all \( j, k \). Thus,

\[
\Pi(\Theta^c_n) \leq P \left( \sum_{j \leq J_n} 2^{\frac{s}{2}(2\alpha+1)} \sum_k |d_{j,k}| \xi_{j,k} \geq K2^{J_n} \right) + P \left( \max_{j > J_n} 2^{\frac{s}{2}(2\alpha-1)} \sum_k |d_{j,k}| \xi_{j,k} \geq K \right)
\]

16
\[ \leq \mathbb{P}\left( Q \sum_{j \leq J_n} \sum_k |\xi_{j,k}| \geq 2^j K \right) + \sum_{j > J_n} \mathbb{P}\left( Q 2^{-j} \sum_k |\xi_{j,k}| \geq K \right). \]

On the \( j \)-th resolution level there are of the order of \( 2^j \) wavelet coefficients. Recall that \( 2^{J_n} \asymp \varepsilon_n^{-1/\alpha} \) and \( \varepsilon_n \gtrsim n^{-\alpha/(1+\alpha)} \). Lemma 7.1 shows that for any constant \( c \), \( \Pi(\Theta^*_n) \leq e^{-cn\varepsilon_n} \) for sufficiently large \( K \).

The assertion follows by Theorem 2.3.

**Proof of Lemma 4.4.** Since \( \psi \) is \( s \)-regular and \( \beta \leq s \), \( |h_{j,k}| \lesssim 2^{-\frac{1}{2}(2\alpha+1)} \). As \( \psi \) has compact support, there exists a constant \( C \) such that \( \|X - h\|_\infty \leq C(\sum_{j \leq J} \max_k |\xi_{j,k} - h_{j,k}| + 2^{-J\beta}) \). Let \( J^* \) be the smallest integer such that \( C(\sum_{j \leq J^*} \varepsilon/(2JC) + 2^{-J^*}) \leq \varepsilon \). Notice that \( 2^{J^*} \asymp \varepsilon^{-1/\beta} \) as \( \varepsilon \to 0 \) and

\[ \mathbb{P}(\|X - h\|_\infty \leq \varepsilon) \geq \mathbb{P}(J = J^*) \prod_{j \leq J^*} \mathbb{P}(|\xi_{j,k} - h_{j,k}| \leq \varepsilon/(2JC)). \]

Let \( c := \inf_{x:|x| \leq (2JC)^{-1} + \max_{j \leq J^*} |h_{j,k}|} f_\varepsilon(x) \). By the assumptions on \( f_\varepsilon \), we can conclude that \( c > 0 \). Together with \( \mathbb{P}(J = J^*) \asymp 2^{-J^*} \) and the fact that on the \( j \)-th resolution level the number of wavelet coefficients is bounded by \( A2^j \) for some \( A > 0 \), this shows that

\[ \mathbb{P}(\|X - h\|_\infty \leq \varepsilon) \gtrsim 2^{-J^*} \left( \frac{c\varepsilon}{J^*} \right)^{A2^{J^*}}. \]

The result follows from \( \varepsilon \asymp 2^{J^*} \).

**Proof of Theorem 4.5.** We verify the conditions (i) – (iii) of Theorem 2.3.

(i): To check the first condition, pick \( J_n \) such that \( \varepsilon_n^{-1/\beta} \leq 2^{J_n} < 2\varepsilon_n^{-1/\beta} \). For a constant \( K \) that will be chosen later to be large enough, define

\[ \Theta_n = \left\{ g = \sum_{j \leq J_n} \vartheta_{j,k} \psi_{j,k} : J_n \leq J_n, \sum_{j \leq J_n} |\vartheta_{j,k}| \leq K2^{J_n} \right\}. \]

As the wavelet has compact support, we have for \( g = \sum_{j \leq J_n} \vartheta_{j,k} \psi_{j,k} \) and \( h = \sum_{j \leq J_n} \vartheta'_{j,k} \psi_{j,k} \), the sup-norm bound \( \|g - h\|_\infty \leq C'2^{J_n/2} \sum_{j \leq J_n} |\vartheta_{j,k} - \vartheta'_{j,k}| \) for some positive constant \( C' \). From [7], Proposition 4.3.36,

\[ \mathcal{N}(C\varepsilon_n, \Theta_n, \|\cdot\|_\infty) \leq \left( \frac{C'2^{3J_n/2}}{C\varepsilon_n} \right)^{2^{J_n}}. \]

Since \( \varepsilon_n^{-1/\beta} \leq 2^{J_n} < 2\varepsilon_n^{-1/\beta} \) and \( \varepsilon_n = (\log n/n)^{-\beta/(1+\beta)} \), we therefore obtain \( \mathcal{N}(C\varepsilon_n, \Theta_n, \|\cdot\|_\infty) \leq e^{Cn\varepsilon_n/2} \) provided \( C \) is chosen sufficiently large. Now, (i) follows from Corollary 2.6.
(ii): Since $\varepsilon_n \to 0$, $f_0 - \varepsilon_n \in C^\beta(R + 1)$ for sufficiently large $n$. The result follows from applying Lemma 4.4 to (2.3) and

$$
\Pi(f : \|f + \varepsilon_n - f_0\|_\infty \leq \varepsilon_n) \geq \varepsilon_n^{D\varepsilon_n^{-1/\beta}} \geq e^{-Cn\varepsilon_n},
$$

for sufficiently large $C$.

(iii): Observe that $\Pi(\Theta_n^c) \leq \mathbb{P}(J > J_n) + \mathbb{P}(\sum_{j \leq J_n, k} |\xi_{j,k}| \geq K2^{J_n})$. The sum $\sum_{j \leq J_n, k}$ is over $A2^{J_n}$ wavelet coefficients. Recall that $2^{J_n} \asymp \varepsilon_n^{-1/\alpha}$ and $\varepsilon_n = (\log n/n)^{\beta/(1+\beta)}$. Lemma 7.1 shows now that for any constant $c$ we obtain $\Pi(\Theta_n^c) \leq e^{-cn\varepsilon_n}$ for sufficiently large $K$. The assertion follows from Theorem 2.3.

References

[1] Bochkina, N. A., and Green, P. J. The Bernstein - von Mises theorem and nonregular models. Ann. Statist. 42, 5 (2014), 1850–1878.

[2] Chernozhukov, V., and Hong, H. Likelihood estimation and inference in a class of nonregular econometric models. Econometrica 72, 5 (2004), 1445–1480.

[3] Cohen, A., Daubechies, I., and Vial, P. Wavelets on the interval and fast wavelet transforms. Appl. Comput. Harmon. Anal. 1, 1 (1993), 54–81.

[4] Ghosal, S., Ghosh, J. K., and van der Vaart, A. W. Convergence rates of posterior distributions. Ann. Statist. 28, 2 (2000), 500–531.

[5] Ghosal, S., and van der Vaart, A. Convergence rates of posterior distributions for noniid observations. Ann. Statist. 35, 1 (2007), 192–223.

[6] Giné, E., and Nickl, R. Rates on contraction for posterior distributions in $L^r$-metrics, $1 \leq r \leq \infty$. Ann. Statist. 39, 6 (2011), 2883–2911.

[7] Giné, E., and Nickl, R. Mathematical Foundations of Infinite-Dimensional Statistical Models. Cambridge University Press, Cambridge, 2016.

[8] Hoffmann, M., Rousseau, J., and Schmidt-Hieber, J. On adaptive posterior concentration rates. Ann. Statist. 43, 5 (2015), 2259–2295.

[9] Jirak, M., Meister, A., and Reiss, M. Adaptive function estimation in nonparametric regression with one-sided errors. Ann. Statist. 42, 5 (2014), 1970–2002.

[10] Kleijn, B., and Knapik, B. Semiparametric posterior limits under local asymptotic exponentiality. ArXiv e-prints (2012).
[11] **Kutoyants, Y.** *Statistical Inference for Spatial Poisson Processes.* Springer, 1998.

[12] **Le Cam, L., and Yang, G. L.** *Asymptotics in statistics.* Springer Series in Statistics. Springer-Verlag, New York, 1990.

[13] **Li, M., and Ghosal, S.** Bayesian detection of image boundaries. *Ann. Statist.* 45, 5 (2017), 2190–2217.

[14] **Lo, A. Y.** Bayesian nonparametric statistical inference for Poisson point processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 59, 1 (1982), 55–66.

[15] **Meister, A., and Reiss, M.** Asymptotic equivalence for nonparametric regression with non-regular errors. *Probab. Theory Related Fields* 155, 1-2 (2013), 201–229.

[16] **Ray, K.** Bayesian inverse problems with non-conjugate priors. *Electron. J. Stat.* 7 (2013), 2516–2549.

[17] **Reiss, M., and Schmidt-Hieber, J.** Nonparametric Bayesian analysis of the compound Poisson prior for support boundary recovery. *ArXiv e-prints* (2017).

[18] **Reiss, M., and Selk, L.** Efficient estimation of functionals in nonparametric boundary models. *Bernoulli* 23, 2 (2017), 1022–1055.

[19] **van der Vaart, A. W., and van Zanten, H.** Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.* 36, 3 (2008), 1435–1463.

[20] **Wang, Y.** Small ball problem via wavelets for Gaussian processes. *Statist. Probab. Lett.* 32, 2 (1997), 133 – 139.