Posterior contraction and credible regions for level sets

Wei Li

Department of Mathematics,
Syracuse University,
Syracuse, NY 13244-1150, U.S.A.
e-mail: wli169@syr.edu

Subhashis Ghosal

Department of Statistics,
North Carolina State University,
Raleigh, NC 27695-8203, U.S.A.
e-mail: sghosal@stat.ncsu.edu

Abstract: For a given function f on a multivariate domain, the level sets, given by $\{x : f(x) = c\}$ for different values of c, provide important geometrical insights about the underlying function of interest. The distance on level sets of two functions may be measured by the Hausdorff metric or a metric based on the Lebesgue measure of a discrepancy, both of which can be linked with the L^∞-distance on the underlying functions. In a Bayesian framework, we derive posterior contraction rates and optimal sized credible sets with assured frequentist coverage for level sets in some nonparametric settings by extending some univariate L^∞-posterior contraction rates to the corresponding multivariate settings. For the multivariate Gaussian white noise model, adaptive Hausdorff and Lebesgue contraction rates for levels sets of the signal function and its mixed order partial derivatives are derived using a wavelet series prior on the function. Assuming a known smoothness level of the signal function, an optimal sized credible region for a level set with assured frequentist coverage is derived based on a multidimensional trigonometric series prior. For the nonparametric regression problem, adaptive rates for level sets of the function and its mixed partial derivatives are obtained using a multidimensional wavelet series prior. When the smoothness level is given, optimal sized credible regions with assured frequentist coverage are obtained using a finite random series prior based on tensor products of B-splines. We also derive Hausdorff and Lebesgue contraction rates of a multivariate density function under a known smoothness setting.

Keywords and phrases: Level sets, posterior contraction, credible region, coverage.

Received September 2020.

1. Introduction

For a given constant c, the c-level set for a smooth function $f : \mathbb{R}^d \to \mathbb{R}$ is defined as the set $\{x \in \mathbb{R}^d : f(x) = c\}$. In the literature, by level sets, some authors mean the sets $\{x \in \mathbb{R}^d : f(x) \geq c\}$. To avoid a possible confusion, it would perhaps be appropriate to term the sets $\{x \in \mathbb{R}^d : f(x) = c\}$ as level

curves when $d = 2$, but owing to the fact that in higher dimensions, these are surfaces and hypersurfaces, we shall use the term "level set" for $\{x \in \mathbb{R}^d : f(x) = c\}$. Estimation of level sets is helpful in understanding the geometry of the function surface, in the problems like clustering (Cuevas et al. [18], Rinaldo and Wasserman [40]), support estimation of the density function (Cuevas and Fraiman [17], Biau et al. [5]) and binary classification (Mammen and Tsybakov [33]).

The most common approach to the estimation of level sets is the so-called "plug-in" approach, when, in computing a level set, the argument f is replaced by some nonparametric estimate $\hat{f}$. Under some suitable metrics, convergence rates for density level set estimates by the plug-in method were studied in Tsybakov [46], and later in Baíllo et al. [1], Cadre [7], Singh and Nowak [43] and Rigollet and Vert [39]. Asymptotic normality of some measure of the symmetric difference between the level set and its plug-in estimator was studied in Mason and Polonik [34]. Cavalier [12] studied nonparametric regression level sets and Cuevas et al. [19] studied level sets of a general smooth function. Besides the plug-in approach to estimation, a direct approach called "excess mass approach" was proposed and studied in Polonik [36] and Polonik and Wang [37] for density and regression functions respectively. More recent papers studying statistical inference for level sets include Jankowski and Stanberry [28], Mammen and Polonik [32] and Chen et al. [13]. The constructions of confidence regions proposed in the first two papers both relied on the estimates of the sup-norm distance between f and $\hat{f}$, while the third paper made use of the estimate of variation in the Hausdorff metric directly. Inferential problems of similar types have also been studied in econometrics literature where the objects of interests are identified as sets Chernozhukov et al. [14, 15].

All the papers from the literature mentioned above followed the frequentist approach and the inference procedures about level sets mainly rely on the bootstrap procedure. To the best of our knowledge, Gayraud and Rousseau [20] is the only paper on level set estimation using the Bayesian approach, where the authors studied the contraction rate for a level set $\{f = c\}$ in density estimation with respect to the Lebesgue distance, that is, the Lebesgue measure of the symmetric difference of $\{f \geq c\}$ with its true value $\{f_0 \geq c\}$. The present paper has two primary goals. The first is to study contraction rates for the level set $\{f = c\}$ of a function f of interest in terms of both Hausdorff and Lebesgue distance in several nonparametric settings such as for the signal function or its mixed partial derivatives in a multivariate signal with a Gaussian white noise model, for the multivariate regression function or a mixed partial derivative of it in a nonparametric regression setting, and for the density function in a multivariate density estimation problem, and check if such rates can be automatically adapted to the smoothness of the underlying function. Second, to validate Bayesian credible regions for level sets from the frequentist perspective by obtaining adequate frequentist coverage of such a region while maintaining its optimal size. The primary technique is to link the rates and the sizes of the credible regions respectively to the L^∞-contraction rate of the underlying function and a posterior quantile of the L^∞-distance between the function of

interest and its Bayes estimate.

The general theory of posterior contraction rates in Bayesian nonparametric models was developed in Ghosal et al. [23] and Ghosal and van der Vaart [21], respectively for independent and identically distributed (i.i.d.) observations and general observations. In this general theory, the existence of tests of exponentially decaying errors plays a fundamental role. Shen and Wasserman [42] also studied the posterior contraction rates for i.i.d. observations but under somewhat stronger conditions. For a thorough account of the development of the theory of posterior contraction, see the recent monograph Ghosal and van der Vaart [22]. These works primarily focus on consistency and contraction rates in the Hellinger metric, or metrics on the underlying parameters which can be related to the Hellinger (or average root-squared-Hellinger distance in case of non-i.i.d. observations) relatively easily.

A challenge with our goals is that the contraction rate in terms of the L^∞-distance is harder to obtain using the general theory of posterior contraction, because certain exponentially powerful tests as required by the theory can exist on much smaller portions of the parameter space, and the role of the prior becomes a lot more important. The resulting rates may turn out to be suboptimal for some commonly used natural priors. This led to the investigation of optimal rates through more direct analyses of the properties of the posterior distribution. For the one-dimensional signal with Gaussian white noise model, Giné and Nickl [24] obtained the optimal rate using conjugacy with Gaussian priors on wavelet coefficients, while Hoffmann et al. [27] obtained adaptive L^∞-contraction rates using a spike-and-slab prior on these coefficients. Castillo [9] introduced a method based on a Bernstein-von Mises theorem to study sup-norm contraction rates and this method leads to minimax contraction rates in (univariate) density estimation with a wavelet series prior on log-density, and the signal with Gaussian white noise model using non-conjugate priors. Yoo and Ghosal [48] established optimal L^∞-rates for the nonparametric multivariate regression function and its mixed partial derivative using tensor products of B-splines prior. Yoo et al. [50] obtained adaptive posterior L^∞-contraction rates in the same setup using a spike-and-slab prior. Shen and Ghosal [41] studied both L^2-and L^∞-rates for density and its derivatives and obtained minimax rate for the L^2-distance, but suboptimal for the L^∞-distance. Castillo and Nickl [10] obtained exact asymptotic L^∞-coverage of an optimal-sized Bayesian credible set in the signal with a Gaussian white noise model through a Bernstein-von Mises theorem in a larger space. For the multivariate regression problem, Yoo and Ghosal [48] showed that a suitably inflated Bayesian credible L^∞-ball around the posterior mean has high frequentist coverage and the optimal size. A more recent paper Castillo and van der Pas [11] derived optimal L^∞-rates for the univariate hazard rate function. Nickl and Ray [35] studied d-dimensional diffusions and obtained the optimal L^∞-rates for the drift vector field when $d \leq 4$ and sub-optimal L^∞-rates for $d \geq 5$.

In this paper, we address the questions of posterior contraction and coverage of credible sets for level sets of functions appearing in nonparametric modeling. We address these problems by linking the posterior contraction rates and

credible sets for level sets with those on the underlying function in terms of the L^∞-distance. More specifically, three nonparametric models are considered — signal with Gaussian white noise, nonparametric Gaussian regression, and multivariate smooth density on d-dimensional hypercube $[0, 1]^d$. For the first and the third setup, we slightly extend known L^∞-contraction results of respectively Hoffmann et al. [27] and Castillo [9] to the higher dimensional settings. Unfortunately, we find that the approach of Castillo [9] does not seem to give the optimal L^∞-rate for a multivariate density function, although it does improve upon the rate using a finite random series prior based on tensor products of B-splines; see Proposition 6.2 of Shen and Ghosal [41]. For the nonparametric Gaussian regression problem, the L^∞-posterior contraction results on the function and its mixed partial derivatives are directly usable from Yoo et al. [50]. The L^∞-contraction rates for the signal with Gaussian white noise and nonparametric Gaussian regression also automatically adapt to the smoothness of the underlying function, leading to adaptive posterior contraction rates for the level sets.

The paper is organized as follows. In the next section, we present notations, preliminaries, and a brief tutorial on wavelets on the domain $[0, 1]^d$. In Section 3, we present results on posterior contraction rates for level sets for all three models. The coverage of credible regions is addressed in Section 4. Computational algorithm and simulation results are presented in Section 5. The proofs and other related results are given in Section 6. Unless otherwise stated, throughout the paper it is implicitly assumed that $d \geq 2$.

2. Preliminaries

2.1. Notations and definitions

Let $\mathbb{N} = \{1, 2, .., \}$, $\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$ and $\mathbb{Z}$ be the set of integers. Given two real sequences a_n and b_n, $a_n = O(b_n)$ or $a_n \lesssim b_n$ means that a_n/b_n is bounded, while $a_n = o(b_n)$ or $a_n \ll b_n$ means that $a_n/b_n \to 0$. Also $a_n \asymp b_n$ means that both $a_n = O(b_n)$ and $b_n = O(a_n)$. For a sequence of random elements Z_n, $Z_n = O_p(a_n)$ means that $\mathrm{P}(|Z_n| \leq C_n a_n) \to 1$ for every $C_n \to \infty$.

Consider $f : U \mapsto \mathbb{R}$ defined on some bounded set $U \subset \mathbb{R}^d$. The L^∞-norm (sup-norm) is denoted by $\|f\|_\infty = \sup_{x \in U} |f(x)|$. The L^2-norm (with respect to Lebesgue measure) is denoted by $\| \cdot \|_2$ with the associated inner product $\langle \cdot, \cdot \rangle$. Any integration without specification of the underlying measure is implicitly taken with respect to the Lebesgue measure. For $g : U \mapsto \mathbb{R}^1$ on some bounded set $U \subset \mathbb{R}^d$, let ∇g be the gradient of g, which is a $d \times 1$ vector of functions. For a d-dimensional multi-index $r = (r_1, \ldots, r_d) \in \mathbb{N}_0^d$, let D^r be the partial derivative operator $\partial^{|r|}/\partial x_1^{r_1} \cdots \partial x_d^{r_d}$ where $|r| = \sum_{l=1}^d r_l$.

The Hölder Space $\mathcal{H}^\alpha([0, 1]^d)$ of order $\alpha > 0$ consists of functions $f : [0, 1]^d \mapsto \mathbb{R}$ such that $\|f\|_{\alpha,\infty} < \infty$, where $\| \cdot \|_{\alpha,\infty}$ is the Hölder norm

$$\|f\|_{\alpha,\infty} = \max_{r:|r| \leq \lfloor \alpha \rfloor} \sup_x |D^r f(x)| + \max_{r:|r| = \lfloor \alpha \rfloor} \sup_{x,y:x \neq y} \frac{|D^r f(x) - D^r f(y)|}{\|x - y\|^{\alpha - \lfloor \alpha \rfloor}},$$

where supremum is taken over the support of f and $\lfloor \alpha \rfloor$ is the largest integer strictly smaller than α.

Given two sets A and B in a Euclidean Space, let $d(A|B) := \sup_{x \in A} \inf_{y \in B} \|x - y\|$. If $A = \{x\}$, write $d(x|B)$ for $d(\{x\}|B) := \inf\{\|x-y\| : y \in B\}$. The Hausdorff distance between A and B is defined as

$$\text{Haus}(A, B) = \max\{d(A|B), d(B|A)\}. \tag{1}$$

2.2. Level sets

Let $f : \mathbb{R}^d \mapsto \mathbb{R}$ and consider some value c that belongs to the range of f. Then the level set at c is given by

$$\mathcal{L}_c := \mathcal{L}_c(f) = \{x : f(x) = c\}. \tag{2}$$

We shall assume the following conditions.

(A1) There exist $\epsilon > 0$ and $\delta_1 > 0$, such that for any $\tilde{c} \in [c - \epsilon, c + \epsilon]$, and x satisfying $|f(x) - \tilde{c}| \le \delta_1$, it holds that $d(x|\{f = \tilde{c}\}) \le A_1|f(x) - \tilde{c}|^{\nu_1}$ for some $A_1 > 0, \nu_1 > 0$.

Assumption (A1) precludes functions arbitrarily flat around the level c. In particular, the condition implies that $\mathcal{L}_c$ does not include any stationary points of the function f (i.e., points at which ∇f equals zero) or any flat part of f at the level c. Hence points of local extrema are excluded. To see this, consider $t_n = c + \epsilon_n$, for some small $\epsilon_n > 0$ such that $\epsilon_n < \delta_1$. Clearly, for $x \in \mathcal{L}_c$, $|f(x) - t_n| = \epsilon_n < \delta_1$, thus $d(x|\{f = t_n\}) \le A_1|f(x) - t_n|^{\nu_1}$, giving $\inf\{\|x - x_n\| : f(x_n) = t_n\} \le A_1 \epsilon_n^{\nu_1}$. Therefore, there exists some x_n such that $f(x_n) = t_n$ and $\|x-x_n\| \le A_1 \epsilon_n^{\nu_1}$. Hence there exists some sequence $x_n \to x$ and $f(x_n) > c$. Similarly, it is straightforward to see that there also exists some sequence $x_n \to x$ and $f(x_n) < c$.

The following lemma gives a sufficient condition for (A1).

Lemma 2.1. *If there exists some $\epsilon_1 > 0$ and $c_0 > 0$ such that $\inf\{\|\nabla f(x)\| : |f(x) - c| \le \epsilon_1\} > c_0$, then Condition (A1) hold with $\nu_1 = 1$.*

This lemma essentially ensures that the set $\{x : f(x) = c\}$ is a lower-dimensional manifold. When $d = 2$, the level set becomes a curve and could be called a "level curve". The following two lemmas are slight generalizations of Theorem 1 and Theorem 2 of Cuevas et al. [19].

Lemma 2.2. *Let f be a continuous function satisfying Condition (A1). Then there exists a $\delta > 0$ such that if g is another function with $\|f - g\|_\infty < \delta$, then*

$$\text{Haus}(\mathcal{L}_c(f), \mathcal{L}_c(g)) \le C_1 \|f - g\|_\infty^{\nu_1}, \tag{3}$$

for some C_1, which may be taken as $6A_1$.

The order in the bound above is sharp in the situation where the level set is a regular curve ($\nu_1 = 1$). To see this, let f and g respectively be two-dimensional centered normal densities with covariance matrix identity and $(1 + \delta)$-times identity. Then for a given c, the level sets for both f and g are circles whose radii differ by the order of δ as $\delta \to 0$. Hence their Hausdorff distance is of the order of δ. Now, as both f and g are radial functions, their L^∞-distance can be seen to be also $O(\delta)$, as $f(0) - g(0)$ is of the order of δ and the maximum of $g - f$ is also easily seen to be of the order of δ. Thus for small $\|f - g\|_\infty$, the inequality may not be improved except for the constant. This implies that the optimal rate for L^∞-estimation is also the optimal rate for level set estimation in the Hausdorff metric in the present situation. The conclusion remains valid even if the domain is restricted to a disc with a sufficiently large radius.

We also consider another metric given by the Lebesgue measure of symmetric difference between the regions enclosed by the level sets, namely, $\lambda\big(\{f_1 \geq c\} \triangle \{f_2 \geq c\}\big)$, where λ stands for the Lebesgue measure. We shall call this the Lebesgue metric and denote it by $\mathrm{Leb}(\mathcal{L}_c(f_1), \mathcal{L}_c(f_2))$. The following assumption will be needed to study distances in terms of the Lebesgue metric, where ν_2 plays a role similar to ν_1 in Assumption (A1).

(A1$'$) For all sufficiently small $\epsilon > 0$, we have $\lambda\{c - \epsilon < f < c + \epsilon\} \leq A_2\epsilon^{\nu_2}$ for some $A_2 > 0, \nu_2 > 0$.

Lemma 2.3. *Let f be a continuous function satisfying (A1$'$) and $1 \leq p < \infty$. Then there exists $\delta > 0$ such that $\|f - g\|_\infty < \delta$ for a function g implies that*

$$\mathrm{Leb}\big(\mathcal{L}_c(f), \mathcal{L}_c(g)\big) \leq C_2\|f - g\|_\infty^{\nu_2}, \tag{4}$$

for some constant C_2, which may be taken as A_2.

Note that the bivariate normal density function satisfies (A1$'$) with $\nu_2 = 1$, and in this regular case, it can be argued similarly as above that the inequality (4) is sharp.

2.3. Wavelets on $[0, 1]^d$

There is a huge literature on both theory and applications of wavelets; see Härdle et al. [26], Giné and Nickl [25] and the references therein. Suppose that ϕ, ψ are the scaling function and wavelet of a Daubechies wavelet basis of $L^2(\mathbb{R})$. For $j \in \mathbb{N}_0$, $k \in \mathbb{Z}$, let $\phi_k = \phi(\cdot - k)$, $\psi_k = \psi(\cdot - k)$, $\phi_{j,k} = 2^{j/2}\phi_k(2^j\cdot)$ and $\psi_{j,k} = 2^{j/2}\psi_k(2^j\cdot)$. A wavelet basis on $[0, 1]^d$ (called CDV-wavelet basis, named after Cohen, Daubechies and Vial) can be constructed from ϕ, ψ starting from some sufficiently large fixed resolution level J_0; see Cohen et al. [16]. We write this basis as

$$\{\phi_k, \psi_{j,k'} : 0 \leq k \leq 2^{J_0} - 1, 0 \leq k' \leq 2^j - 1, j \geq J_0\}.$$

Let $\mathcal{K}(j) = \{0, \ldots, 2^j - 1\}^d$ and $\mathcal{I}$ is the set of sequence $i = (i_1, \ldots, i_d)$ of zeros and ones excluding $i = \{0, \ldots, 0\}$. Both ϕ and ψ can be constructed to be of

sufficiently high regularity, say being q-times $(q > \alpha)$ continuously differentiable with their derivatives up to q-th order uniformly bounded. A wavelet series for $f \in L^2([0,1]^d)$ can be written as

$$f = \sum_{k \in \mathcal{K}(J_0)} \langle f, \Phi_k \rangle \, \Phi_k + \sum_{j=J_0}^{\infty} \sum_{i \in \mathcal{I}, k \in \mathcal{K}(j)} \langle f, \Psi_{j,k}^i \rangle \, \Psi_{j,k}^i,$$

where

$$\Phi_k(x) := \Phi_{J_0,k}(x) = \phi_{J_0,k_1}(x_1) \cdots \phi_{J_0,k_d}(x_d),$$
$$\Psi_{j,k}^i(x) := \psi_{j,k_1}^{i_1}(x_1) \cdots \psi_{j,k_d}^{i_d}(x_d),$$

$\psi_{j,k}^0(\cdot) := \phi_{j,k}(\cdot)$ and $\psi_{j,k}^1(\cdot) := \psi_{j,k}(\cdot)$. It should be pointed out that the summation over $i \in \mathcal{I}$ may be omitted throughout the proof with the understanding that $|\mathcal{I}| \leq 2^d$ and thus inclusion or exclusion of this summation does not affect the asymptotic analysis. Therefore, up to renumbering indices, it is more convenient to write

$$f = \sum_{j=0}^{\infty} \sum_{k \in \mathcal{K}(j)} \langle f, \Psi_{j,k} \rangle \, \Psi_{j,k} = \sum_{(j,k) \in \Lambda} \langle f, \Psi_{j,k} \rangle \, \Psi_{j,k},$$

where the level $j = 0$ corresponds to the father wavelet, $\Lambda := \{(j,k) : j \in \mathbb{N}_0, k \in \mathcal{K}(j)\}$ and $|\mathcal{K}(j)| = O(2^{jd})$.

In the proof, we shall use the following properties of this wavelet basis.

(i) $\left\| \sum_{k \in \mathcal{K}(j)} |D^r \Psi_{j,k}| \right\|_{\infty} \lesssim 2^{j(d/2+|r|)}$.

(ii) The wavelets bases characterize the so-called Besov Space $\mathcal{B}_{\infty,\infty}^{\alpha}([0,1]^d)$ with $\alpha < q$, in that for any f in the Besov Spaces $\mathcal{B}_{\infty,\infty}^{\alpha}([0,1]^d)$,

$$\|f\|_{\infty,\infty,\alpha} := \sup_{j \in \mathbb{N}_0, k \in \mathcal{K}(j)} 2^{j(\alpha+d/2)} |\langle f, \Psi_{j,k} \rangle| < \infty.$$

In fact, $\mathcal{B}_{\infty,\infty}^{\alpha}([0,1]^d)$ coincides with $\mathcal{H}^{\alpha}([0,1]^d)$ when $\alpha > 0$ is non-integer and that $\mathcal{H}^{\alpha}([0,1]^d) \subset \mathcal{B}_{\infty,\infty}^{\alpha}([0,1]^d)$ when α is an integer.

(iii) $\Psi_{j,k}$ has support $S_{j,k}$ with area (or volume) at most 2^{-jd}, and $\|\Psi_{j,k}\|_{\infty} \lesssim 2^{jd/2}$, $\int |\Psi_{j,k}| \lesssim 2^{-jd/2}$.

(iv) For fixed level j, given a fixed $\Psi_{j,k}$ with its support $S_{j,k}$, the number of wavelets of the level $j' < j$ with support intersecting $S_{j,k}$ is bounded by a universal constant (independent of j, j', k); the number of wavelets of the level $j' > j$ with support intersecting $S_{j,k}$ is bounded by $2^{(j'-j)d}$ times a universal constant.

3. Posterior contraction rates for level sets

Given a prior Π on the parameters describing f, let $\Pi(\cdot|\mathbb{D}_n)$ denote the posterior distribution, where n is the sample size, $\mathbb{D}_n$ are the observations, and $\mathrm{E}^{\Pi}(\cdot|\mathbb{D}_n)$

is the expectation under the law $\Pi(\cdot|\mathbb{D}_n)$. When the context of the prior is clear, we may drop the superscript Π in $\mathrm{E}^\Pi(\cdot|\mathbb{D}_n)$. If the observations $\mathbb{D}_n$ follow (the n-fold product of) the true law P_0, let E_0 denote the expectation under the true law. We denote the true function (signal function, regression function or density function) by f_0.

3.1. Signal with Gaussian white noise on $[0, 1]^d$

Consider the multivariate white noise model which is defined through the stochastic differential equation $dY_n(t) = f(t)dt + n^{-1/2}dW(t)$ for $t \in [0, 1]^d$, where dW is defined through a multivariate stochastic integral with respect to independent standard Brownian motions $(W_1(t_1), \ldots, W_d(t_d))$ and that

$$\int g(t)dW(t) \sim \mathrm{N}(0, \|g\|^2) \text{ for any } g \in L^2([0, 1]^d).$$

With above choice of the basis, the equivalent sequence space model is given by

$$Y_{j,k} = \theta_{j,k} + \frac{1}{\sqrt{n}}\varepsilon_{j,k}, \quad k \in \{0, \ldots, 2^j - 1\}^d, j \geq 0,$$

where the parameters are given by $\theta_{j,k} := \langle f, \Psi_{j,k} \rangle$ and $\varepsilon_{j,k}$ are all i.i.d. N(0,1). As in Hoffmann et al. [27], we put a spike-and-slab prior on $\theta_{j,k}$. Let J_n be a deterministic increasing function of n to be defined in the following theorem, $\theta_{j,k}$'s drawn independently as

$$\theta_{j,k} \sim \begin{cases} (1 - w_j)\delta_0 + w_j g(\cdot), & \text{if } j \leq J_n, \\ \delta_0, & \text{if } j > J_n; \end{cases}$$

here δ_0 is the point mass at zero and g is a bounded density function of $\mathbb{R}$ which satisfies

$$\inf_{x \in [-L_0, L_0]} g(x) > 0, \tag{5}$$

for some suitable $L_0 > 0$.

Let f_0 stand for the true signal function and let P_0 be the true distribution of the observed signal. A multivariate version of Theorem 3.1 of Hoffmann et al. [27] implies the following theorem.

Theorem 3.1. *Consider a spike and slab prior defined above, where $2^{J_n} \asymp n^{1/d}$ and for some constants $s > 0, \tau > 1/2$, $n^{-s} \leq w_j \leq \min\{2^{-jd(1+\tau)}, 1/2\}$. For every $\alpha > 0$, $L_0 - 1 \geq M > 0$, $K_n \to \infty$, for any f_0 with $\|f_0\|_{\alpha,\infty} \leq M$,*

(i) if f_0 satisfies Assumption (A1), then

$$\Pi\Big(\mathrm{Haus}(\mathcal{L}_c(f), \mathcal{L}_c(f_0)) > K_n\big(n/\log n\big)^{-\nu_1\alpha/(d+2\alpha)}\big|Y\Big) \xrightarrow{\mathrm{P}_0} 0;$$

(ii) if f_0 satisfies Assumption (A1'), then

$$\Pi\Big(\mathrm{Leb}\big(\mathcal{L}_c(f), \mathcal{L}_c(f_0)\big) > K_n\big(n/\log n\big)^{-\nu_2\alpha/(d+2\alpha)}\big|Y\Big) \xrightarrow{\mathrm{P}_0} 0.$$

Remark 3.1. The corresponding rate for the level sets of the derivatives of the function $D^r f$ is given by $(n/\log n)^{\nu_l(|r|-\alpha)/(d+2\alpha)}$ where $|r| < \alpha$, $l = 1, 2$, for the case (i) and (ii) respectively.

The rate above is the optimal L^∞-rate for α-smooth functions in d-dimension. In view of the sharpness of (3) when $\nu_1 = 1$ (or (4) when $\nu_2 = 1$) discussed earlier, this rate is also optimal for estimating a level set in terms of the Hausdorff metric and Lebesgue metric. Since the prior does not depend on the smoothness level α of the underlying true function, the posterior contraction rate is adaptive.

3.2. Nonparametric Gaussian regression

Consider the Gaussian nonparametric regression problem with observations (Y_i, X_i), where $Y_i = f(X_i) + \varepsilon_i$, $\varepsilon_i \overset{\mathrm{i.i.d.}}{\sim} N(0, \sigma^2)$, $i = 1, \ldots, n$, and the covariates take values in $[0, 1]^d$. We write $Y = (Y_1, \ldots, Y_n)^T$, $X = (X_1^T, \ldots, X_n^T)^T$, $F = (f(X_1), \ldots, f(X_n))^T$ and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$, so that $Y = F + \varepsilon$.

Our study allows for both fixed and random designs. For the fixed design case, X is considered fixed and each Y_i given X_i is independently distributed. For the random design case, each X_i is further assumed to be independent and identically distributed (i.i.d.). If the covariates are random, they are assumed to follow the uniform distribution on $[0, 1]^d$. If the covariates are deterministic, let $\mathbb{G}_n$ stand for the empirical distribution function of $X_1, \ldots, X_n$. We assume that

$$\sup_{x \in [0,1]^d} |\mathbb{G}_n(x) - U(x)| = O(n^{-1}), \tag{6}$$

where U stands for the cumulative distribution function of the uniform distribution on $[0, 1]^d$. More generally, if the relation holds with U replaced by a cumulative distribution function G having a positive and continuous density, then the case can be reduced to (6) by a transformation on x, but then the interpretation of the regression function will change too. However, the level sets are not affected by a transformation of x, so we can assume that the covariates are uniformly distributed.

As in Yoo et al. [50], the priors on $\theta_{j,k} = \langle f, \Psi_{j,k} \rangle$ and σ^2 are mutually independent with each other. Specifically, $\theta_{j,k}$'s are drawn independently as

$$\theta_{j,k} = \begin{cases} g(\cdot), & \text{if } j = 0, \\ (1 - w_j)\delta_0 + w_j g(\cdot), & \text{if } 0 < j \le J_n, \\ \delta_0, & \text{if } j > J_n; \end{cases}$$

here, the density function g also satisfies Condition (5). The prior for σ is taken to be some positive and continuous density on $(0, \infty)$.

Assume that the true error distribution is sub-Gaussian with mean zero and variance σ_0^2. In view of Theorem 4.2 of Yoo et al. [50], the following theorem and remark readily follow.

Theorem 3.2. *Consider a spike-and-slab prior defined above, where $2^{J_n} \asymp (n/\log n)^{1/(2d)}$ and $n^{-s} \leq w_j \leq \min\{2^{-jd(1+\tau)}, 1/2\}$ for some constants $s > 0, \tau > 1/2$. For every $\alpha > d/2$, $L_0 - 1/2 \geq M > 0$, $K_n \to \infty$, for any f_0 with $\|f_0\|_{\alpha,\infty} \leq M$,*

(i) if f_0 satisfies Assumption (A1), then

$$\Pi\Big(\mathrm{Haus}(\mathcal{L}_c(f), \mathcal{L}_c(f_0)) > K_n\big(n/\log n\big)^{-\nu_1\alpha/(d+2\alpha)}\big|\mathbb{D}_n\Big) \xrightarrow{\mathrm{P}_0} 0;$$

(ii) if f_0 satisfies Assumption (A1$'$), then

$$\Pi\Big(\mathrm{Leb}\big(\mathcal{L}_c(f), \mathcal{L}_c(f_0)\big) > K_n\big(n/\log n\big)^{-\nu_2\alpha/(d+2\alpha)}\big|\mathbb{D}_n\Big) \xrightarrow{\mathrm{P}_0} 0.$$

Remark 3.2. The corresponding rate for the level sets of the derivatives of the function $D^r f$ is given by $(n/\log n)^{\nu_l(|r|-\alpha)/(d+2\alpha)}$ where $|r| < \alpha$, $l = 1, 2$, for the case (i) and (ii) respectively.

As the prior is free of the smoothness level α, the posterior of the level set of the regression function contracts adaptively at the optimal rate $(n/\log n)^{-\alpha/(2\alpha+d)}$ for the L^∞-metric, which is also the optimal rate for level sets in the regular situation in view of the sharpness of (3) and (4).

3.3. Density estimation

Suppose we have $X_i \overset{\mathrm{i.i.d.}}{\sim} \mathrm{P}_0$, whose density function f_0 is supported on $[0,1]^d$. Following Castillo [9], we put a prior through an exponentiation of a wavelet prior, generalized to d-dimension. More specifically, let $f(x) = \exp\{T(x) - c(T)\}$, where

$$T(x) = \sum_{j=0}^{J_n} \sum_{k \in \mathcal{K}(j)} \sigma_{j,k}\alpha_{j,k}\Psi_{j,k}, \qquad c(T) = \log \int_0^1 \exp\{T(x)\}dx,$$

and $\alpha_{j,k}$ are drawn independently from some fixed density φ if $j \leq J_n$ and are set to 0 if $j > J_n$. There are two types of density functions for φ under consideration. The first is the log-Lipschitz case, that is, $\log(\varphi)$ is a Lipschitz function. For this class, we assume that

$$\sigma_{j,k} \equiv \sigma_j \geq 2^{-j(\alpha+d/2)}. \tag{7}$$

The second type is to take φ to be the standard normal density function. In this case, we assume that for some r such that $0 < r < \alpha - d/4$,

$$\sigma_{j,k} \equiv \sigma_j = 2^{-j(r+d/2)}. \tag{8}$$

Theorem 3.3. *Suppose a positive density function $f_0 \in \mathcal{H}^\alpha([0,1]^d)$ for some $\alpha > d > 1$. Using the priors described above with $2^{J_n} \asymp (n/\log n)^{1/(2\alpha+d)}$. For any $K_n \to \infty$,*

    (i) if f_0 satisfies Assumption (A1), then

$$\Pi\Big(\mathrm{Haus}(\mathcal{L}_c(f), \mathcal{L}_c(f_0)) > K_n \epsilon_n^{\nu_1} \Big| \mathbb{D}_n\Big) \xrightarrow{\mathrm{P}_0} 0;$$

    (ii) if f_0 satisfies Assumption (A1′), then

$$\Pi\Big(\mathrm{Leb}\big(\mathcal{L}_c(f), \mathcal{L}_c(f_0)\big) > K_n \epsilon_n^{\nu_2} \Big| \mathbb{D}_n\Big) \xrightarrow{\mathrm{P}_0} 0,$$

where $\epsilon_n = (n/\log n)^{-(\alpha-(d/2)+\alpha/(\alpha+1))/(2\alpha+d)}$.

As the prior relies on the knowledge of the smoothness level α, the above rate is not adaptive. It is useful to point out also, the suboptimal rate arises from the suboptimal L^∞-rate obtained for the multivariate density function when $d > 1$; see the discussion following the proof of Propositions 6.2. It is also useful to compare the obtained rate with those obtained by Gayraud and Rousseau [20] and Polonik [36]. The former used the weaker Lebesgue measure distance and concluded that for $d = 2$, the rate is $\max\{n^{-\alpha/(3\alpha+3)}, n^{-1/4}\}$, which is always weaker than the rate in Theorem 3.3 for smoothness $\alpha \geq 2$. The later paper used different conditions which include a metric entropy condition for the underlying class of candidate sets, and a different metric. But when the true density is bounded and bounded away from zero, the metric behaves like the Lebesgue measure metric. Under a comparable situation which involves at least the existence of a gradient, the rate is at the best $n^{-1/3}$, which is always weaker than ours for smoothness $\alpha \geq 3$. Thus even though our obtained rate for density is weaker than the anticipated optimal rate $(n/\log n)^{-\alpha/(2\alpha+2)}$, we improve upon the available results when the function is at least moderately smooth. Unlike ours, their rates are capped and do not improve with smoothness. Further, our rate is also with respect to the Hausdorff distance, which is typically stronger than the Lebesgue measure metric in a bounded domain, if measured from a curve with bounded length (that is, not too wiggly).

4. Optimal credible regions with assured coverage

In this section, we provide credible regions for level sets with sufficient coverage for the signal with a Gaussian white noise model using the trigonometric series prior, and the Gaussian nonparametric regression using B-splines series priors assuming that the targeted smoothness level of the true function is known. The issue of adaptive size and coverage is a lot more complex. It is known that even for a one-dimensional signal in a white noise model with the L^2-distance, it is impossible to obtain coverage with adaptive size sets by any method, Bayesian or not; see for instance Li [30], Baraud [2], Cai and Low [8]. Size of a credible region can adapt to the underlying complexity only if certain parts of the parameter

space are removed from consideration by using restrictions like "self-similarity", "polished tail" or "excessive bias restriction"; see Szabó et al. [45], Belitser and Nurushev [4], Belitser [3], Sniekers and van der Vaart [44] and Ray [38] for results on adaptive Bayesian coverage in certain conjugacy settings with the size measured by L^2-type distances. At present, it is unclear whether L^∞-adaptive credible sets are possible to obtain, and if possible, which part of the parameter space is to be removed to maintain coverage with adaptive L^∞-size, as the notion of a bias-variance decomposition for the L^2-setting is not easily extended to the L^∞-setting. Thus we study coverage of credible sets only under the known smoothness setting for the signal with a Gaussian white noise model and the nonparametric Gaussian regression model using conjugate priors, respectively based on multivariate trigonometric series and tensor products of B-splines. The conjugacy allows certain explicit calculations necessary for lower-bounding the size of the posterior spread, which is the key to obtaining frequentist coverage. The approach will induce credible regions for the level sets from those on the function constructed using posterior quantiles of the L^∞-spread of the function around its center. The approach is similar to that of Yoo and Ghosal [49] and Li and Ghosal [31] respectively for the mode and the filaments of the regression function.

4.1. Signal with Gaussian white noise model

We start with the Gaussian white noise model. Let $\tilde{\mathcal{L}}$ be the induced level set of $\tilde{f}$, where $\tilde{f}$ is the posterior mean function. For some $0 < \gamma < 1/2$, let $R_{n,\gamma}$ denotes the $(1 - \gamma)$-quantile of the posterior distribution of $\|f - \tilde{f}\|_\infty$. Let

$$\mathbb{C}_\gamma := \{f : \|f - \tilde{f}\|_\infty \le \rho R_{n,\gamma}\}, \tag{9}$$

where ρ is some positive constant sufficiently large. The following theorem provides a credible region with sufficiently high frequentist coverage in the signal with Gaussian white noise model using trigonometric series prior.

Let $\{e_i : i = 1, 2, \ldots,\}$ be an orthonormal basis of $L^2([0,1]^d)$. Then the equivalent sequence space model is $Y_i = \theta_i + n^{-1/2}\varepsilon_i$, $Y_i := \int e_i(t)dY(t)$, $\theta_i := \langle e_i, f\rangle = \int e_i(t)f(t)dt$ and $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0,1)$, $i = 1, 2, \ldots$.

To obtain the coverage of credible regions, conjugacy will be essential. The prior will be based on conjugate normal distributions on the coefficients of an orthogonal basis expansion. While tensor products of CDV wavelets can again be used, we illustrate the results using the basis of tensor products of trigonometric functions. The details of the proofs for the posterior contractions rates for these two priors are different.

For $j = (j_1, \ldots, j_d) \in \{0, 1, \ldots\}^d$ and $x \in [0,1]^d$, we denote the trigonometric basis functions by

$$\phi_{j,k}(x) = \prod_{l=1}^{d} \sqrt{2}[(1 - k_l)\cos(2\pi j_l x_l) + k_l \sin(2\pi j_l x_l)],$$

where $k \in \mathcal{K}(j) := \{k \in \{0,1\}^d : k_l = 0 \text{ when } j_l = 0, l = 1, \ldots, d\}$. The collection of $\{\phi_{j,k} : (j,k) \in \{0,1,\ldots,\}^d \times \mathcal{K}(j)\}$ forms a basis for $L^2([0,1]^d)$. Expand the function f_0 in the wavelet series as

$$f_0(x) = \sum_{j_1}^{\infty} \cdots \sum_{j_d}^{\infty} \sum_{k \in \mathcal{K}(j)} \phi_{j,k}\theta_{0,j,k},$$

for some collection of coefficients $\{\theta_{0,j,k} : j \in \{0,1,\ldots\}^d$ and $k \in \mathcal{K}(j)\}$. For the sake of brevity, we may suppress the ranges of the indices of summation and just write the expansion as $f = \sum_j \sum_k \phi_{j,k}\theta_{j,k}$. The equivalent sequence model is $Y_{j,k} = \theta_{j,k} + n^{-1/2}\varepsilon_{j,k}$. If the prior is that $\theta_{j,k} \stackrel{\text{ind}}{\sim} \mathrm{N}(0,\mu_j)$, the posterior distribution of $\theta_{j,k}$ is Gaussian with mean $\tilde{\theta}_{j,k} = n\mu_j Y_{j,k}/(n\mu_j + 1)$ and variance $\mu_j/(n\mu_j + 1)$.

While the posterior contraction rate and the convergence of the Bayes estimator can be obtained as before, we only present results on uncertainty quantification. The theorem below holds for $d = 2$ or 3. We do not prove it for $d > 3$ as the terms quickly become very cumbersome and the argument becomes tedious. However, by inspecting the proofs, we conjecture that it holds also for $d > 3$.

Theorem 4.1. *Suppose $d = 2$ or 3. Let $\mu_j = \prod_{l=1}^{d} \mu_{j_l}$, where $\mu_{j_l} = j_l^{-(1+2\alpha/d)}$ if $j_l \neq 0$ but $\mu_{j_l} = 1$ if $j_l = 0$. Assume that the true signal f_0 satisfies the condition (A1) and its Fourier coefficients $\{\theta_{0,j,k} : j \in \{0,1,\ldots\}^d$ and $k \in \mathcal{K}(j)\}$ satisfy the smoothness condition*

$$\sum_{j_1=0}^{\infty} \cdots \sum_{j_d=0}^{\infty} \sum_{k \in \mathcal{K}(j)} \sup_{\substack{s_1,\ldots,s_d \geq 0 \\ \sum_{l=1}^{d} s_l = \alpha}} \prod_{l=1}^{d} j_l^{s_l}|\theta_{0,j,k}| \leq L, \tag{10}$$

for some positive constant L and $\alpha > d$. Then the following assertions hold:

- (i) *for the credible set $\mathcal{C}_\gamma = \{\mathcal{L}_c(f) : f \in \mathbb{C}_\gamma\}$ for $\mathcal{L}_c$, with $\mathbb{C}_\gamma$ as defined in (9), both the credibility and its coverage probability for $\mathcal{L}_c(f_0)$ tend to 1;*
- (ii) *for the credible set $\bar{\mathcal{C}}_\gamma := \{\mathcal{L}_c : \mathrm{Haus}(\mathcal{L}_c, \tilde{\mathcal{L}}_c) \leq C_1\rho^{\nu_1}R_{n,\gamma}^{\nu_1}\}$ where C_1 is as in (3), we have that $\Pi(\mathcal{C}_\gamma \subset \bar{\mathcal{C}}_\gamma | \mathbb{D}_n) \to 1$ in P_0-probability;*
- (iii) $\mathrm{diam}_{\mathrm{Haus}}(\mathcal{C}_\gamma) = O_{P_0}(n^{-\nu_1\alpha/(2\alpha+d)}(\log n)^{\nu_1 d/2})$.

If f_0 satisfies (A1′), then

- (iv) *For the credible set $\tilde{\mathcal{C}}_\gamma := \{\mathcal{L}_c(f) : \mathrm{Leb}(\mathcal{L}_c(f), \mathcal{L}_c(\tilde{f})) \leq C_2\rho^{\nu_2}R_{n,\gamma}^{\nu_2}\}$, both the credibility and its coverage tend to one, where C_2 is as in (4);*
- (v) $\mathrm{diam}_{\mathrm{Leb}}(\tilde{\mathcal{C}}_\gamma) = O_{P_0}(n^{-\nu_2\alpha/(2\alpha+d)}(\log n)^{\nu_2 d/2})$.

Remark 4.1. Note that the smoothness condition (10) is stronger than conditions imposed commonly on Fourier coefficients to quantify their decay using square summability. This becomes necessary as such conditions characterize only L^2-Sobolev-type smoothness, not Hölder-type. The condition (10) implies L^∞-type-bounds and is slightly stronger than a typical Hölder condition.

Remark 4.2. We note that the size of the credible region matches with the optimal rate up to a logarithmic factor in the regular situations.

Remark 4.3. An alternative use of a posterior distribution in Bayesian analysis is to quantify the uncertainty that the level set will be contained in $\mathcal{C}'_\gamma = \{x : |\tilde{f}(x) - c| < \rho R_{n,\gamma}\}$. This is attractive because of its simplicity, as the set is just a band in $\mathbb{R}^d$ instead of a subset of the space of curves on $\mathbb{R}^d$. To examine its credibility, note that

$$
\begin{aligned}
\Pi(\mathcal{L}_c \subset \mathcal{C}'_\gamma | \mathbb{D}_n) &= \Pi(f : \{x : f(x) = c\} \subset \{x : |\tilde{f}(x) - c| < \rho R_{n,\gamma}\} | \mathbb{D}_n) \\
&= \Pi(f : \{|\tilde{f}(x) - f(x)| < \rho R_{n,\gamma}, \forall x : f(x) = c\} | \mathbb{D}_n) \\
&\geq \Pi(f : \|\tilde{f} - f\|_\infty < \rho R_{n,\gamma} | \mathbb{D}_n),
\end{aligned} \tag{11}
$$

which tends to one by the proof of Theorem 4.1. Similarly,

$$
\mathrm{P}_0(\mathcal{L}_c(f_0) \subset \mathcal{C}'_\gamma) \geq \mathrm{P}_0(\|\tilde{f} - f_0\|_\infty < \rho R_{n,\gamma}) \to 1. \tag{12}
$$

Also observe that if the gradient is nonzero throughout on the level curve, then the spread of the band is of the order

$$
\rho R_{n,\gamma} = O_{P_0}(n^{-\alpha/(2\alpha+d)}(\log n)^{d/2}).
$$

The advantage of this set $\mathcal{C}'_\gamma$ is that the posterior mean and $R_{n,\gamma}$ can be obtained from posterior sampling. If one computes the set $\{x : |\tilde{f}(x) - c| < \rho R_{n,\gamma}\}$ numerically, it will serve as a confidence set. Here, instead of relying on bootstrap to get the quantiles, the cut-off $R_{n,\gamma}$ is obtained from the posterior. This credible set has to be interpreted as a possible region of controlled size that contains the true level set.

4.2. Nonparametric Gaussian regression

For credible sets with coverage in the Gaussian nonparametric regression problem, we shall use B-splines functions. Choose a fixed order q and a sequence of knots $0 = t_0 < \cdots < t_{N+1} = 1$. Denote the B-spline functions of order q by B_i and form the tensor products $B_{i_1}(x_1) \cdots B_{i_d}(x_d)$. Let $b_{J_1,\ldots,J_d}(x) = (B_{j_1,\ldots,j_d}(x) = \prod_{k=1}^d B_{j_k}(x_k) : 1 \leq j_k \leq J_k)^T$ to be a column vector of tensor product of B-splines functions; here N_k denotes the number of interior points and $J_k = q_k + N_k$ denotes the number of basis functions used for each coordinate. If desired, N_k, q_k and the knot locations $t_{k,0}, \ldots, t_{k,N_k+1}$ can be chosen differently for different k. Let $\delta_{k,\ell} = t_{k,\ell+1} - t_{k,\ell}$. We assume the quasi-uniformity of the knots, in that $\max\{\delta_{k,\ell} : 0 \leq \ell \leq N_k\}/\min\{\delta_{k,\ell} : 0 \leq \ell \leq N_k\} \leq C$ for some $C > 0$. To ensure quality of the approximation, we fix $q_k > \alpha$, and let $N_k = N_k(n)$ and $J_k = J_k(n) := N_k(n) + q_k$.

As in Yoo and Ghosal [48], we put a prior on the regression function through the representation $f = b^T \theta$, $\theta | \sigma^2 \sim \mathrm{N}(\theta_0, \sigma^2 \Lambda_0)$, where $c_1 I \leq \Lambda_0 \leq c_2 I$, for some constants $0 < c_1 \leq c_2 < \infty$ and I is the $\prod J_k \times \prod J_k$ identity matrix. Using B

to denote $(b_{J_1,\ldots,J_d}(X_1),\ldots,b_{J_1,\ldots,J_d}(X_n))^T$, the model becomes $Y|(X,\theta,\sigma^2) \sim$ N$(B\theta,\sigma^2 I_n)$. It follows then

$$\theta|(\mathbb{D}_n,\sigma^2) \sim \text{N}\left((\Lambda_0^{-1} + B^T B)^{-1}(B^T Y + \Lambda_0^{-1}\theta_0), \sigma^2(B^T B + \Lambda_0^{-1})^{-1}\right).$$

Writing

$$b_J^{(r)}(x) = \left(\frac{\partial^{r_1}}{\partial x_1^{r_1}}B_{j_1,q_1}(x_1)\cdots\frac{\partial^{r_d}}{\partial x_d^{r_d}}B_{j_d,q_d}(x_d) : 1 \le j_k \le J_k, k = 1,\ldots,d\right)^T,$$

(13)

the posterior distribution for f and its partial derivatives are obtained as

$$\Pi(D^r f|\mathbb{D}_n,\sigma^2) \sim \text{GP}(A_r Y + C_r\theta_0, \sigma^2\Sigma_r),$$ (14)

where $A_r(x) = b_J^{(r)}(x)^T (B^T B + \Lambda_0^{-1})^{-1}B^T$, $C_r(x) = b_J^{(r)}(x)^T (B^T B + \Lambda_0^{-1})^{-1}\Lambda_0^{-1}$ and $\Sigma_r(x,y) = b_J^{(r)}(x)^T (B^T B + \Lambda_0^{-1})^{-1}b_J^{(r)}(y)$ and GP denotes a Gaussian process. To handle σ^2, we can either put a conjugate inverse-gamma prior $\sigma^2 \sim \text{IG}(a/2, b/2)$ with shape parameter $a/2 > 2$ and rate parameter $b/2 > 0$ or plug-in an estimate for σ^2. The theoretical study is similar in both cases, using the consistency of the marginal maximum likelihood estimator of σ, so for the ease of exposition, we consider the second approach only. The plug-in posterior is given by

$$\Pi(D^r f|\mathbb{D}_n,\sigma^2)|_{\sigma^2=\hat\sigma_n^2} \sim \text{GP}(A_r Y + C_r\theta_0, \hat\sigma_n^2\Sigma_r),$$ (15)

where $\hat\sigma_n^2 = n^{-1}(Y - B\theta_0)^T (B\Lambda_0 B^T + I_n)^{-1}(Y - B\theta_0)$.

We assume that $Y_i = f_0(X_i) + \varepsilon_i$, where ε_i are i.i.d. sub-Gaussian with mean 0 and variance σ_0^2 for $i = 1,\ldots,n$. Notice that under P$_0$, $x \mapsto A_r(x)\varepsilon/\sigma_0$ is a mean zero process with a sub-Gaussian tail. As in the adaptive case, the result below extends to both fixed and random design, except that Condition (6) can be replaced by $\sup_x |\mathbb{G}_n(x) - G(x)| = o(\prod_{k=1}^d J_k^{-1})$, where G can be taken as the uniform distribution for the fixed design, the sampling distribution of X_i for the random design.

The following result shows the coverage of a Bayesian credible set.

Theorem 4.2. *If f_0 satisfy Assumption (A1), $J_k \asymp (n/\log n)^{1/(2\alpha+d)}$, $k = 1,\ldots,d$, then with $\mathbb{C}_\gamma := \{f : \|f - \tilde f\|_\infty \le \rho R_{n,\gamma}\}$, for some $\rho > 0$ sufficiently large, the following assertions hold:*

(i) *the credibility of $\mathcal{C}_\gamma := \{\mathcal{L}_c(f) : f \in \mathbb{C}_\gamma\}$ and its coverage probability for $\mathcal{L}_c(f_0)$ both tend to 1;*

(ii) *$\Pi(\mathcal{C}_\gamma \subset \bar{\mathcal{C}}_\gamma|\mathbb{D}_n) \to 1$ in P$_0$-probability, where $\bar{\mathcal{C}}_\gamma := \{\mathcal{L}_c : \text{Haus}(\mathcal{L}_c,\tilde{\mathcal{L}}) \le C_1\rho^{\nu_1}R_{n,\gamma}^{\nu_1}\}$ and C_1 is as in (3),;*

(iii) *$\text{diam}_{\text{Haus}}(\mathcal{C}_\gamma) = O_{P_0}((n/\log n)^{-\nu_1\alpha/(2\alpha+d)})$.*

If f_0 satisfies Assumption (A1'), then

(iv) *the credibility and coverage of $\tilde{\mathcal{C}}_\gamma := \{\mathcal{L}_c(f) : \text{Leb}(\mathcal{L}_c(f),\mathcal{L}_c(\tilde f)) \le C_2\rho^{\nu_2}R_{n,\gamma}^{\nu_2}\}$ both tend to one, where C_2 is as in (4);*

(v) $\mathrm{diam}_{\mathrm{Leb}}(\tilde{\mathcal{C}}_\gamma) = O_{P_0}((n/\log n)^{-\nu_2\alpha/(2\alpha+d)})$.

Remark 4.4. Both Remark 4.2 and Remark 4.3 for the Gaussian white noise model also hold for nonparametric regression. Furthermore, the same method can be used to construct credible region for $\mathcal{L}_c(D^r f)$ by replacing f by $D^r f$ in the Theorem 4.2.

5. Computation and simulations

5.1. Algorithm

We give a description of the two generic algorithms we use for finding level sets.

Algorithm 1 (Fixed Point Algorithm). Set $\epsilon > 0$, $\tau > 0$, $\bar{a} > 0$ and select a collection of starting points $\{x_1, ..., x_n\}$, compute $f(x_i)$ and keep only those points x_i for which $f(x_i) > \tau$. Now for each x_i, let $x_i^{(1)} = x_i$; iterate through the following steps from $t = 1$:

(1) evaluate $\nabla f(x_i^{(t)})$;
(2) compute $(f(x_i^{(t)}) - c)\nabla f(x_i^{(t)})$;
(3) update $x_i^{(t+1)} = -\bar{a}(f(x_i^{(t)}) - c)\nabla f(x_i^{(t)}) + x_i^{(t)}$;
(4) stop if $\|x_i^{(t+1)} - x_i^{(t)}\| < \epsilon$ or $\|\nabla f(x_i^{(t+1)})\| < \epsilon$.

The second algorithm is based on the idea of simulated annealing and posterior sampling from the unnormalized density $\exp[-(f(x) - c)^2/\bar{a}]$, where $\bar{a}$ is a small positive tuning parameter.

Algorithm 2 (Simulated-annealing Based Algorithm). Set $\tau > 0$, some small positive $\bar{a}$ and burn-in time $B > 0$ sufficiently large. Select a collection of starting points $\{x_1, ..., x_n\}$, compute $f(x_i)$ and keep only those points for which $f(x_i) > \tau$. Now for each x_i, let $x_i^{(1)} = x_i$; iterate the following step from $t = 1, \ldots, B$:

- Update $x_i^{(t+1)} \leftarrow x_i^{(t)}$ by a slice sampler that gives rise to the invariant distribution whose density is proportional to $\exp\{-(f(x) - c)^2/\bar{a}\}$.

At last, collect all points $\{x_i^{(B+1)}\}$.

5.2. Simulation results

We work with the nonparametric Gaussian regression setting. In the simulation, we consider the following function

$$f(x_1, x_2) = 1 + \left(g\left(\sqrt{x_1^2 + x_2^2}\right)\right)^{1+\cos^2(\tan^{-1}(x_2/x_1))},$$

where g is the normal density function with mean 0.5 and standard deviation 0.3. We generate i.i.d. data X_i uniformly on $[0, 1] \times [0, 1]$ and i.i.d. ε_i from
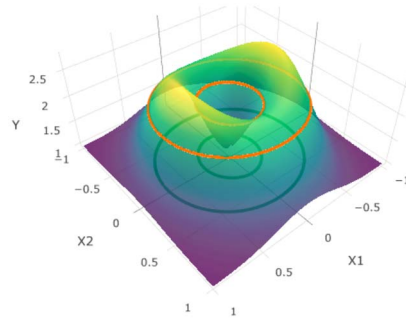
FIG 1. *The function f and its level curves.*

the normal distribution with mean 0 and standard deviation 0.1 and then set $Y_i = f(X_{i,1}, X_{i,2}) + \varepsilon_i, i = 1, \ldots, n$. The level set of interest is $\{x : f(x) = 2.1\}$. The true level set consists of two circles, as given in Figure 1.

The sample size is 2000. We use fifth-order B-splines functions, that is $q_1 = q_2 = 5$. One can choose the pair (J_1, J_2) by their posterior mode by maximizing (in the logarithmic scale)

$$\log \Pi(J_1, J_2 | \mathbb{D}_n, \hat{\sigma}^2) = -2n \log \hat{\sigma} - \log(\det(B\Lambda_0 B^T) + I_n) + \text{const.}$$

We give the results using a simulated-annealing based algorithm. We choose $\tau = 0.5$ and $\bar{a} = 10^{-5}$. The tuning parameters J_1 and J_2 are both chosen 9 throughout the pilot experiments.

We also experimented with different choices of J_1 and J_2. Figure 2 shows that the posterior mean under different smoothing levels. When $J_1 = J_2 = 7$, just slightly smaller than the 9, it can be seen that the posterior mean completely fails to approximate the inner circle. The approximation seems quite reasonable with a larger J.

Figure 3 gives uncertainty quantification for two cases $J_1 = J_2 = 7$ and $J_2 = J_2 = 9$. We choose $\gamma = 0.1$ and $\rho = 1.2$. This choice of ρ should give sufficiently large credibility but not too high. This can be done by some pilot simulation using the posterior samples. To evaluate the $(1 - \gamma)$-quantile $R_{n,\gamma}$, we first draw 200 posterior samples of θ to compute their posterior mean $\tilde{\theta}$. We compute $\|b_{J_1,J_2}(x)^T(\theta - \tilde{\theta})\|_\infty$ by searching on a crude grid and pick the maximum point on the grid. Then starting from this maximum point, we apply the gradient ascent method to check if nearby points can achieve greater (absolute) value. We keep the largest value as the supremum. The $(1 - \gamma)$-empirical quantile over all these suprema gives our estimate of $R_{n,\gamma}$. Finally, we draw 100 posterior level sets (level sets induced from the posterior samples of f) and keep those that fall in the set $\mathcal{C}_\gamma$. The curves of different colors correspond to these level sets. It is obvious that undersmoothing tends to deliver better and more robust coverage.

Now we assess the credibility and coverage performance over 100 iterations. Some experimentation suggests that the choice $\rho = 1.2$ works well, giving 97.13%
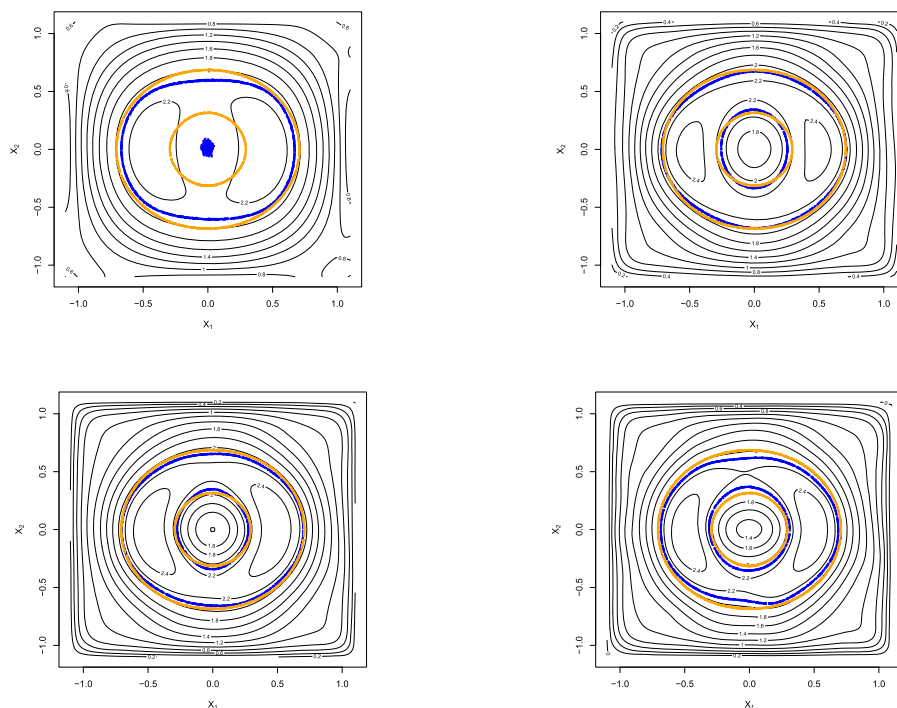
FIG 2. *Nonparametric regression. Effects of the smoothing parameters J_1 and J_2. The orange circle is the truth, the blue curve is estimated level curve induced by the posterior mean. Top left: $J_1 = J_2 = 7$; top right: $J_1 = J_2 = 9$; bottom left: $J_1 = J_2 = 11$; bottom right: $J_1 = J_2 = 15$.*

credibility averaging over all iterations. To evaluate the coverage performance, we compute the Hausdorff distance between the truth $\mathcal{L}_c^*$ and $\tilde{\mathcal{L}}_c$. Simulation shows that $\mathcal{L}_c^*$ belongs to $\bar{\mathcal{C}}_\gamma$ respectively $88\%, 90\%, 94\%, 96\%, 98\%$ time when C takes values respectively $0.21, 0.22, 0.23, 0.24, 0.25$. Inspecting the proof of Lemma 2.1, the constant A_1 may be estimated as $1/\inf\{\|\nabla \tilde{f}(x)\| : x \in \{\tilde{f} = c\}\}$. When we compute with $C_1 = 6A_1$ in this way, the frequentist coverage turns out to be 100%, as the theory predicts.

6. Proofs

Proof of Lemma 2.1. Let x be such that $|f(x) - \tilde{c}| < \delta_1$, where we take $\tilde{c}$ between $c - \epsilon$ and $c + \epsilon$. Consider $\tilde{y}$ to be the projection of x onto $\{f = \tilde{c}\}$ (just take one if not unique). We assume $x \neq \tilde{y}$ (otherwise the bound is trivially true). Therefore $\tilde{y} \in \{f = \tilde{c}\}$ and $d(x|\{f = \tilde{c}\}) = \|\tilde{y} - x\|$. Since the gradient $\nabla f(\tilde{y})$ is normal to the level curve $\{f = \tilde{c}\}$ at $\tilde{y}$ and x lies on the normal line, it holds that $x = \tilde{y} + \tilde{t}\nabla f(\tilde{y})$ for some $\tilde{t} \neq 0$. Without loss of generality, take δ_1 to be a
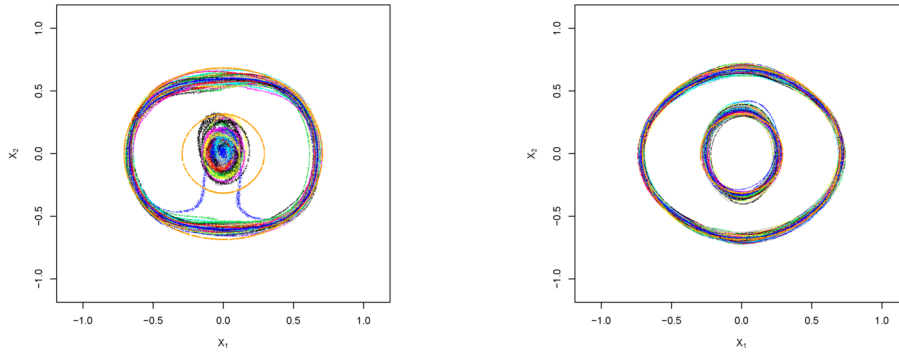
FIG 3. *Nonparametric regression. Uncertainty Quantification: the orange circle is the truth, while the remaining colored curves are the level sets (from the posterior samples) that belong to $\mathcal{C}_\gamma$. Left: $J_1 = J_2 = 7$. Right: $J_1 = J_2 = 9$.*

small value. By Taylor's expansion,

$$\left| f(x) - f(\tilde{y}) \right| = \left| (x - \tilde{y})^T \nabla f(\tilde{y}) + o(\|x - \tilde{y}\|) \right| = \|x - \tilde{y}\| \| \nabla f(\tilde{y}) + o(1) \|.$$

Therefore, $\|x - \tilde{y}\| < A |f(x) - f(\tilde{y})|$ for some A, because $\|\nabla f(\tilde{y})\| > c_0 > 0$. $\quad\square$

Proof of Theorem 3.1. The proof follows from the L_∞-contraction rate for a multivariate Gaussian white noise model derived in Proposition 6.1 below in conjugation with (3) and (4). $\quad\square$

Proposition 6.1. *Consider the spike-and-slab priors defined in Section 3.1, where $2^{J_n} \asymp n^{1/d}$ and for some constants $s > 0$, $\tau > 1/2$, $n^{-s} \le w_j \le \min\{2^{-jd(1+\tau)}, 1/2\}$. Then for every $\alpha > 0$, r such that $|r| < \alpha$, $L_0 - 1 \ge M > 0$, $K_n \to \infty$, we have that*

$$\sup_{f_0 : \|f_0\|_{\alpha,\infty} \le M} \Pi\Big(\|D^r f - D^r f_0\|_\infty > K_n \big(n/\log n\big)^{(|r|-\alpha)/(d+2\alpha)} \big| Y \Big) \xrightarrow{\mathrm{P}_0} 0.$$

Proof. We only sketch the proof. Let $S = \{(j,k) : \theta_{j,k} \ne 0\}$, $\theta_{0,j,k} = \langle f_0, \Psi_{j,k} \rangle$, and

$$\mathcal{J}_n(\gamma) = \{(j,k) : |\theta_{0,j,k}| > \gamma \sqrt{(\log n)/n}\}, \gamma > 0.$$

Consider the following events

$$E_1 = \{ \sup_{(j,k) \in \mathcal{J}_n(\underline{\gamma})} |\theta_{j,k} - \theta_{0,j,k}| \le \bar{\gamma} \sqrt{\log n/n} \},$$

$$E_2 = \{ S^c \cap \mathcal{J}_n(\bar{\gamma}) = \varnothing \},$$

$$E_3 = \{ S \cap \mathcal{J}_n^c(\underline{\gamma}) = \varnothing \},$$

for some $\bar{\gamma}$ sufficiently large and $\underline{\gamma}$ sufficiently small. Assume for now that $\mathrm{E}_0[\Pi(E_1^c \cap E_2 \cap E_3 | Y)] \to 0$, $\mathrm{E}_0[\Pi(E_2^c | Y)] \to 0$ and $\mathrm{E}_0[\Pi(E_3^c | Y)] \to 0$. Then it suffices to show that

$$\mathrm{E}_0[\Pi(\{ \|f - f\|_\infty > \epsilon_{n,r,\alpha} \} \cap E_2 \cap E_3 | Y)] \to 0,$$

where $\epsilon_{n,r,\alpha} := (\log n/n)^{(\alpha-|r|)/(2\alpha+d)}$. In view that $\mathrm{E}_0[\Pi(E_1^c \cap E_2 \cap E_3|Y)] \to 0$, then it suffices to show that on the events $E_1 \cap E_2 \cap E_3$, $\|D^r f - D^r f\|_\infty \le \epsilon_{n,r,\alpha}$. To this goal, one can find some suitable choice of $J_n(\alpha) \asymp (n/\log n)^{1/(2\alpha+d)}$ such that

$$\mathcal{J}_n(\underline{\gamma}) \subset \bar{\mathcal{J}}_n(\alpha) := \{(j,k) : j \le J_n(\alpha), k \in \mathcal{K}(j)\}.$$

Note that

$$\|D^r f - D^r f_0\|_\infty \le \Big\| \sum_{(j,k) \in \Lambda} |\theta_{j,k} - \theta_{0,j,k}| |D^r \Psi_{j,k}| \Big\|_\infty$$

$$\le \sum_{j \in \mathbb{N}} \max_{k \in \mathcal{K}(j)} |\theta_{j,k} - \theta_{0,j,k}| \sum_{k \in \mathcal{K}(j)} \|D^r \Psi_{k,j}\|$$

$$\le \sum_{j \in \mathbb{N}} \max_{k \in \mathcal{K}(j)} 2^{(|r|+d/2)j} |\theta_{j,k} - \theta_{0,j,k}|.$$

The sum-of-max over (j,k) can be taken over three different regions $\bar{\mathcal{J}}_n(\alpha) \cap \mathcal{J}_n(\underline{\gamma})$, $\bar{\mathcal{J}}_n(\alpha) \cap (\mathcal{J}_n(\underline{\gamma}))^c$ and $(\bar{\mathcal{J}}_n(\alpha))^c$. On the first region, since $\mathcal{J}_n(\underline{\gamma}) \subset \bar{\mathcal{J}}_n(\alpha)$ and $J_n(\alpha) < J_n$, it can be bounded on E_1 as

$$\Big(\max_{(j,k) \in \mathcal{J}_n(\underline{\gamma})} |\theta_{j,k} - \theta_{0,j,k}| \Big) \sum_{j \le J_n} 2^{(|r|+d/2)j} \lesssim \sqrt{(\log n)/n} \times 2^{J_n(|r|+d/2)} \lesssim \epsilon_{n,r,\alpha}.$$

On the second region, since $|\theta_{0,j,k}| < \underline{\gamma}\sqrt{(\log n)/n}$, and on E_3, $\theta_{j,k} = 0$, together implying that $2^{(|r|+d/2)j}|\theta_{j,k} - \theta_{0,j,k}|$ is bounded by $2^{(|r|+d/2)j}\sqrt{(\log n)/n}$ up to some multiplicative constant. The sum-of-max over this region is again upper bounded by $\epsilon_{n,r,\alpha}$. On the third region, on E_3, noting that $\theta_{j,k} = 0$, the term can be dominated by

$$\sum_{j > J_n(\alpha)} \max_{k \in \mathcal{K}(j)} 2^{(|r|+d/2)j} |\theta_{0,j,k}| \lesssim \sum_{j > J_n(\alpha)} 2^{(|r|+d/2)j} 2^{-j(\alpha+d/2)} \lesssim 2^{J_n(|r|-\alpha)},$$

which is further bounded by $\epsilon_{n,r,\alpha}$.

Finally it can be shown by following the proof of Lemma 1 of Hoffmann et al. [27] that indeed $\mathrm{E}_0[\Pi(E_2^c|Y)] \to 0$, $\mathrm{E}_0[\Pi(E_3^c|Y)] \to 0$ and $\mathrm{E}_0[\Pi(E_1^c \cap E_2 \cap E_3|Y)] \to 0$, noting that the dimension d does not alter most of the argument. □

Proof of Theorem 3.3. The proof follows from L^∞-contraction rate for multivariate density model derived in Proposition 6.2 below in conjugation with (3) and (4). □

Proposition 6.2. *Given a positive density function $f_0 \in \mathcal{H}^\alpha([0,1]^d)$ for some $\alpha > d > 1$; assume that the Gaussian prior under the condition (8) or the log-Lipschitz prior under the condition (7). Then for $2^{J_n} \asymp (n/\log n)^{1/(2\alpha+d)}$ and any $K_n \to \infty$, we have*

$$\Pi\Big(\|f - f_0\|_\infty > K_n(n/\log n)^{-(\alpha-d/2+\alpha/(\alpha+1))/(2\alpha+d)}\Big) \xrightarrow{\mathrm{P}_0} 0.$$

Proof. The proof mainly follows the argument in Castillo [9]. For clarity, we sketch the main ideas of the proof, and present the key results that contribute to the final rates for the multivariate case ($d > 1$). We only provide proofs for key steps and those lemmas that are illustrative of the subtleties for the case $d > 1$, whereas other proofs are referred to the original paper.

Given two density functions, f_1 and f_2, let $h(f_1, f_2)$ denote their Hellinger distance. Some key intermediate results can be shown to hold (the Lemma 4 and 8 of Castillo [9]), that is, for any $K_n \to \infty$,

$$\Pi\Big(h(f, f_0) > K_n \epsilon_{n,\alpha} | \mathbb{D}_n \Big) \xrightarrow{P_0} 0,$$

$$\Pi\Big(\|f - f_0\|_2 > K_n \epsilon_{n,\alpha} | \mathbb{D}_n \Big) \xrightarrow{P_0} 0,$$

$$\Pi\Big(\|f - f_0\|_\infty > K_n \zeta_{n,\alpha} | \mathbb{D}_n \Big) \xrightarrow{P_0} 0,$$

where $\epsilon_{n,\alpha} := (\log n)^{\eta_1} n^{-\alpha/(2\alpha+d)}$ for $\eta_1 = \alpha/(2\alpha+d)$ and the intermediate L^∞-rate $\zeta_{n,\alpha} = 2^{J_n d/2} \epsilon_{n,\alpha} = n^{-(2\alpha-d)/(2(2\alpha+d))}(\log n)^{\eta_2}$ for $\eta_2 = (2\alpha - d)/(2(2\alpha + 2))$. Let D_n be a measurable set given by

$$D_n := \{f : h(f, f_0) \leq \epsilon_{n,\alpha}, \|f - f_0\|_2 \leq \epsilon_{n,\alpha}, \|f - f_0\|_\infty \leq \zeta_{n,\alpha}\}.$$

Denote by Π^{D_n} the normalization of Π to D_n, that is $d\Pi^{D_n} = \mathbb{1}_{D_n} d\Pi(\cdot)/\Pi(D_n)$. Therefore,

$$\mathrm{E}_0 \Pi(D_n | \mathbb{D}_n) = 1 + o(1). \tag{16}$$

Let $g_0 = \log f_0$ and $g = \log f = T - c(T)$. Since $\|f - f_0\|_\infty = \|e^{g_0}(e^{g-g_0} - 1)\|_\infty \lesssim \|g - g_0\|_\infty$ using the bound $|e^x - 1| \lesssim |x|$ for bounded x and the fact that f_0 is bounded, it suffices to bound $\|g - g_0\|_\infty$. By Markov's inequality, for any $K_n \to \infty$ and $c_n > 0$,

$$\mathrm{E}_0\Big(\Pi(g : \|g - g_0\|_\infty > c_n | \mathbb{D}_n)\Big) \leq c_n^{-1} \mathrm{E}_0\Big(\mathrm{E}^{\Pi^{D_n}}(\|g - g_0\|_\infty | \mathbb{D}_n)\Pi(D_n | \mathbb{D}_n)\Big)$$
$$+ o(1). \tag{17}$$

We then bound

$$\mathrm{E}^{\Pi^{D_n}}(\|g - g_0\|_\infty | \mathbb{D}_n) \leq \int \|g^{J_n} - g_0\|_\infty d\Pi^{D_n}(f | \mathbb{D}_n)$$
$$+ \int \|g^{J_n^c}\| d\Pi^{D_n}(f | \mathbb{D}_n) + \|g_0^{J_n^c}\|_\infty, \tag{18}$$

where g^{J_n} is the L^2-projection of g up to level J_n. The second term is 0 by the definition of T, and the constant function is orthogonal to higher level wavelets. For the third term, let $g_{0,j,k}$ denote the wavelet coefficient of g_0, we bound

$$\|g_0^{J_n^c}\|_\infty = \Big\| \sum_{j > J_n} \sum_{k \in \mathcal{K}(j)} g_{0,j,k} \Psi_{j,k} \Big\|_\infty$$

$$\leq \sum_{j>J_n} \left[\max_k |g_{0,j,k}| \left\| \sum_k |\Psi_{j,k}| \right\|_\infty \right]$$

$$\leq \sum_{j>J_n} 2^{-j(\alpha+d/2)} 2^{jd/2},$$

which is bounded by $2^{-J_n\alpha} \lesssim \epsilon_{n,\alpha} = (n/\log n)^{-\alpha/(2\alpha+d)}$.

To estimate the first term on the right side of (18), we introduce a few more notations. Let the log-likelihood be denoted by $\ell_n(f) = \sum_{i=1}^n \log f(X_i)$, $W_n(f) = n^{-1/2} \sum_{i=1}^n (f(X_i) - P_0 f)$, where P_0 denotes integration with respect to the density f_0. Let $\zeta_{j,k} := \Psi_{j,k}/f_0$ and let $\mathcal{A}_{j,k}$ denote the L^2-projection of $\zeta_{j,k}$ up to the level j ($j \leq J_n$), that is,

$$\mathcal{A}_{j,k} \equiv \sum_{\lambda \leq J_n} \sum_{u \in \mathcal{K}(j)} \langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle \Psi_{\lambda,u}.$$

Consider

$$\Gamma^{J_n} := g_0^{J_n} + \frac{1}{\sqrt{n}} \sum_{j \leq J_n} \sum_{k \in \mathcal{K}(j)} W_n(\mathcal{A}_{j,k}) \Psi_{j,k}.$$

Now

$$\int \|g^{J_n} - g_0\|_\infty d\Pi^{D_n}(f|\mathbb{D}_n) \leq \int \|g^{J_n} - \Gamma^{J_n}\|_\infty d\Pi^{D_n}(f|\mathbb{D}_n) + \|\Gamma^{J_n} - g_0^{J_n}\|_\infty.$$

By Lemma 6.4 and Lemma 6.7, the second term on the right hand side of the above expression can be bounded by $\epsilon_{n,\alpha}$, so we shall focus on the first term. Note that

$$\|g^{J_n} - \Gamma^{J_n}\|_\infty = \| \sum_{j \leq J_n} \sum_{k \in \mathcal{K}(j)} \langle g^{J_n} - \Gamma^{J_n}, \Psi_{j,k} \rangle \Psi_{j,k} \|_\infty$$

$$\leq \sum_{j \leq J_n} \max_{k \in \mathcal{K}(j)} |\langle g^{J_n} - \Gamma^{J_n}, \Psi_{j,k} \rangle| \left\| \sum_{k \in \mathcal{K}(j)} |\Psi_{j,k}| \right\|_\infty$$

$$\lesssim \frac{1}{\sqrt{n}} \sum_{j \leq J_n} 2^{jd/2} \max_k \sqrt{n} |g_{j,k} - g_{0,j,k} - \frac{1}{\sqrt{n}} W_n(\mathcal{A}_{j,k})|.$$

Therefore, for any $t > 0$, using Jensen's inequality and $e^{|x|} \leq e^x + e^{-x}$, the term $\mathrm{E}^{\Pi^{D_n}}(\|g^{J_n} - \Gamma^{J_n}\|_\infty |\mathbb{D}_n)$ can be bounded by

$$\frac{1}{t} \mathrm{E}^{\Pi^{D_n}} \left\{ t \frac{1}{\sqrt{n}} \sum_{j \leq J_n} 2^{jd/2} \max_k \sqrt{n} |g_{j,k} - g_{0,j,k} - \frac{1}{\sqrt{n}} W_n(\mathcal{A}_{j,k})| \Big| \mathbb{D}_n \right\}$$

$$= \frac{1}{t} \frac{1}{\sqrt{n}} \sum_{j \leq J_n} 2^{jd/2} \mathrm{E}^{\Pi^{D_n}} \left\{ \log \left(e^{\max_k t\sqrt{n}|g_{j,k}-g_{0,j,k}-\frac{1}{\sqrt{n}}W_n(\mathcal{A}_{j,k})|} \right) \Big| \mathbb{D}_n \right\}$$

$$\leq \frac{1}{\sqrt{n}} \sum_{j \leq J_n} 2^{jd/2} \frac{1}{t} \log \left\{ \mathrm{E}^{\Pi^{D_n}} \left(e^{\max_k t\sqrt{n}|g_{j,k}-g_{0,j,k}-\frac{1}{\sqrt{n}}W_n(\mathcal{A}_{j,k})|} \Big| \mathbb{D}_n \right) \right\}$$

$$\leq \frac{1}{\sqrt{n}} \sum_{j \leq J_n} 2^{jd/2} \Big\{ \frac{1}{t} \log \Big(\sum_k \mathcal{M}_{j,k}(t) \Big) \Big\} + \frac{1}{\sqrt{n}} \sum_{j \leq J_n} 2^{jd/2} \Big\{ \frac{1}{t} \log \Big(\sum_k \mathcal{M}_{j,k}(-t) \Big) \Big\},$$

where $\mathcal{M}_{j,k}(t) := e^{-tW_n(\mathcal{A}_{j,k})} \mathrm{E}^{\Pi^{D_n}}[e^{t\sqrt{n}(g_{j,k}-g_{0,j,k})}|\mathbb{D}_n]$. Note that

$$g - g_0 = \frac{f - f_0}{f_0} + \rho\Big(\frac{f - f_0}{f_0}\Big),$$

where $\rho(x) := \log(1+x) - x$. As argued in p. 2074 of Castillo [9], $g_{j,k} - g_{0,j,k}$ can be written as

$$g_{j,k} - g_{0,j,k} = \langle f - f_0, \zeta_{j,k} - \mathcal{A}_{j,k} \rangle + \langle f - f_0, \mathcal{A}_{j,k} \rangle + \int \rho\Big(\frac{f - f_0}{f_0}\Big) \Psi_{j,k}.$$

Hence $\mathcal{M}_{j,k}(t)$ is given by

$$e^{-tW_n(\mathcal{A}_{j,k})} \mathrm{E}^{\Pi^{D_n}} \Big[e^{t\sqrt{n}\langle f - f_0, \zeta_{j,k} - \mathcal{A}_{j,k} \rangle} e^{t\sqrt{n}\langle f - f_0, \mathcal{A}_{j,k} \rangle} e^{t\sqrt{n}(\int \rho(\frac{f-f_0}{f_0})\Psi_{j,k})} \Big| \mathbb{D}_n \Big].$$

Note that, since $|\rho(x)| \leq x^2$ for x small, and $\|f - f_0\|_\infty \lesssim \zeta_{n,\alpha}$ and by the assumption $\alpha > d$, on D_n,

$$\sqrt{n} \Big| \int \rho\Big(\frac{f - f_0}{f_0}\Big) \Psi_{j,k} \Big| \lesssim \sqrt{n} \|\Psi_{j,k}\|_\infty \int \Big(\frac{f - f_0}{f_0}\Big)^2$$
$$\lesssim \sqrt{n} 2^{jd/2} \|f - f_0\|_2^2$$
$$\lesssim \sqrt{n} 2^{jd/2} \epsilon_{n,\alpha}^2. \tag{19}$$

By Lemma 6.5, maximizing the upper bound (by choosing $j = \lceil J_n \alpha/(\alpha+1) \rceil$) on D_n,

$$\sqrt{n} |\langle f - f_0, \zeta_{j,k} - \mathcal{A}_{j,k} \rangle| \lesssim \sqrt{n} 2^{-J_n \alpha/(\alpha+1)} \|f - f_0\|_2$$
$$\lesssim \sqrt{n} 2^{-J_n \alpha/(\alpha+1)} \epsilon_{n,\alpha}. \tag{20}$$

In view of the bounds in the above two displays, and by Lemma 6.4 and Lemma 6.6 (setting $\gamma_n = \mathcal{A}_{j,k}$), we have, for any $t > 0$,

$$\mathrm{E}^{\Pi^{D_n}}(\|g^{J_n} - \Gamma^{J_n}\|_\infty | \mathbb{D}_n)$$
$$\lesssim \frac{1}{\sqrt{n}} \sum_{j \leq J_n} 2^{jd/2} \frac{1}{t} \log \Big(\sum_k e^{Ct^2} \frac{\int e^{\ell_n(f_t) - \ell_n(f_0)} d\Pi^{D_n}(f)}{\int e^{\ell_n(f) - \ell_n(f_0)} d\Pi^{D_n}(f)} \Big) \tag{21}$$
$$+ \frac{1}{\sqrt{n}} \sum_{j \leq J_n} 2^{jd/2} (\sqrt{n} 2^{jd/2} \epsilon_{n,\alpha}^2) \tag{22}$$
$$+ \frac{1}{\sqrt{n}} \sum_{j \leq J_n} 2^{jd/2} (\sqrt{n} 2^{-J_n \alpha/(\alpha+1)} \epsilon_{n,\alpha}). \tag{23}$$

The second term (22) on the right hand side of above expression is bounded by $\rho_{n,2} := (n/\log n)^{(d-2\alpha)/(2\alpha+d)} (\log n)^{2\eta_1}$, and hence $\rho_{n,2} \lesssim \epsilon_{n,\alpha}$ as long as $\alpha > d$.

A simplification shows that the third term (23) is bounded by

$$\rho_{n,1} = (n/\log n)^{-(\alpha - d/2 + \alpha/(\alpha+1))/(2\alpha+d)}.$$

As will be clear later, this term turns out to be the dominating factor for rates in the multivariate case $(d > 1)$.

All that remains is to bound the first term (21) on the right hand side of above expression. One can write the ratio inside the parentheses as

$$\frac{\int \mathbb{1}_{D^n}(f)e^{\ell_n(f_t)-\ell_n(f_0)}d\Pi(f)}{\Pi(D_n|\mathbb{D}_n)\int e^{\ell_n(f)-\ell_n(f_0)}d\Pi(f)}. \tag{24}$$

Recall that $f = \exp(T(x) - c(T))$, where $T = \sum_{\lambda \leq J_n}\sum_u \theta_{\lambda,u}\Psi_{\lambda,u}$, $\theta_{\lambda,u} := \sigma_{\lambda,u}\alpha_{\lambda,u}$, and

$$f_t = \exp\left\{T - \frac{t}{\sqrt{n}}\mathcal{A}_{j,k} - c(T - \frac{t\mathcal{A}_{j,k}}{\sqrt{n}})\right\}.$$

Let $\tilde{T} := T - tn^{-1/2}\mathcal{A}_{j,k} = \sum_{\lambda \leq J_n}\sum_u (\theta_{\lambda,u} - tn^{-1/2}\langle \mathcal{A}_{j,k}, \Psi_{\lambda,u}\rangle)\Psi_{\lambda,u}$. The prior on f effectively is the prior on the coefficients $\theta = \{\theta_{\lambda,u}\}$ and therefore the numerator of (24) is the integration over the law of θ, whose density is denoted by p_θ. We then proceed to apply change of variables in this integration, by changing $\theta_{\lambda,u} \to \theta^*_{\lambda,u} := \theta_{\lambda,u} - tn^{-1/2}\langle \mathcal{A}_{j,k}, \Psi_{\lambda,u}\rangle$. Therefore, the numerator can be written as

$$\int \mathbb{1}_{D'_n}(f_{\theta^*})e^{\ell_n(f_{\theta^*})-\ell_n(f_0)}\left[\prod_{\lambda \leq J_n, u \in \mathcal{K}(\lambda)}\frac{\varphi(\theta_{\lambda,u}/\sigma_\lambda)}{\varphi\left(\frac{\theta_{\lambda,u}-t\langle \mathcal{A}_{j,k}, \Psi_{\lambda,u}\rangle/\sqrt{n}}{\sigma_\lambda}\right)}\right]p_\theta(\theta^*)d(\theta^*),$$

for some suitably transformed set D'_n. By Lemma 6.8, for the Gaussian prior case, the term in the square bracket is bounded by $\exp\{C|t| + Ct^2\}$ for some constant $C > 0$. Therefore, (24) can be bounded by $\exp\{C|t| + Ct^2\}/\Pi(D_n|\mathbb{D}_n)$. Now the term on the right hand side of (21) can be bounded as

$$\frac{1}{\sqrt{n}}\sum_{j \leq J_n} 2^{jd/2}\frac{1}{t}\log\left(\sum_k e^{C|t|+Ct^2}\Pi(D_n|\mathbb{D}_n)^{-1}\right)$$

$$\leq \frac{1}{\sqrt{n}}\sum_{j \leq J_n} 2^{jd/2}\frac{1}{t}\log\left(2^{jd}e^{C|t|+Ct^2}\right) + \frac{1}{\sqrt{n}}\sum_{j \leq J_n} 2^{jd/2}\frac{1}{t}\log\frac{1}{\Pi(D_n|\mathbb{D}_n)},$$

which, by setting $t = \sqrt{j}$, can be further bounded by $\epsilon_{n,\alpha} + \epsilon_{n,\alpha}\log(1/\Pi(D_n|\mathbb{D}_n))$. Combining all the above results, we obtain

$$E^{\Pi^{D_n}}(\|g^{J_n} - \Gamma^{J_n}\|_\infty|\mathbb{D}_n) \lesssim \epsilon_{n,\alpha} + \epsilon_{n,\alpha}\log(1/\Pi(D_n|\mathbb{D}_n)) + \rho_{n,1} + \rho_{n,2}.$$

We choose $c_n = K_n\epsilon^*_{n,\alpha}$, where $K_n \to \infty$ is a given sequence, and $\epsilon^*_{n,\alpha} := \max(\epsilon_{n,\alpha}, \rho_{n,1}, \rho_{n,2})$ in (17). By (16), we obtain

$$E_0\left(\Pi(g : \|g - g_0\|_\infty > K_n\epsilon^*_{n,\alpha}|\mathbb{D}_n)\right) \to 0.$$

By the assumption $\alpha > d > 1$, $\rho_{n,2} \lesssim \epsilon_{n,\alpha} \lesssim \rho_{n,1}$, thus $\epsilon^*_{n,\alpha} = \rho_{n,1}$, completing the proof for the Gaussian prior. $\qquad\square$

Remark 6.1. For $d = 1$, the optimal rate $\epsilon_{n,\alpha} := (n/\log n)^{-\alpha/(2\alpha+d)}$ was obtained. For $d > 1$, only a suboptimal rate $\rho_{n,1}$ can be obtained, which may be due to some possibly crude bounds used in Lemma 6.5. It is not immediately clear whether the limitation is due to the techniques of the proof or due to the specification of the prior.

Note also that $\rho_{n,1} \lesssim \zeta_{n,\alpha}$, and therefore, $\rho_{n,1}$ is an improvement over the intermediate L^∞-rate $\zeta_{n,\alpha}$. In fact, when $d = 1$, $\alpha > 1$, $\rho_{n,1} \lesssim \epsilon_{n,\alpha}$, which recovers the univariate result in Castillo [9], i.e., $\epsilon_{n,\alpha}^* = \epsilon_{n,\alpha}$. A similar argument goes through for the log-Lipschitz prior. For the most practically interesting case for level sets $d = 2$, the suboptimality disappears with increasing smoothness as $\alpha \to \infty$.

There are two directions that are worth of exploring in the future. One is to obtain improved results for lower smoothness situation possibly $\alpha > d/2$. Another one is to obtain optima sup-norm rates for $d > 1$. There are some very recent results in these directions in other related nonparametric estimation problems; see for instance Castillo and van der Pas [11] for survival analysis and Nickl and Ray [35] for diffusions. The first paper deals with univariate functions, and its technique may be used to obtain results for lower smoothness situation. The second paper develops some techniques which makes it possible to obtain optimal sup-norm rates for $d \leq 4$ in the multidimensional diffusion problems. It remains an open problem if similar techniques can be used for improving rates in the multivariate density estimation.

Lemma 6.3. *Suppose that $f_0 \in \mathcal{H}^\alpha([0,1]^d)$ for some $\alpha > 1$, and is bounded away from 0 and ∞. Let $\zeta_{j,k} = \Psi_{j,k}/f_0$. Then for $k \in \mathcal{K}(j)$, $u \in \mathcal{K}(\lambda)$, it holds that*

$$| \langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle | \lesssim \begin{cases} 2^{(\lambda-j)d/2}, & \text{if } j \geq \lambda, \\ 2^{(j-\lambda)(\alpha+d/2)}, & \text{if } j < \lambda. \end{cases}$$

Proof. If $j \geq \lambda$, then

$$| \langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle | \leq \int \left| \frac{\Psi_{j,k}\Psi_{\lambda,u}}{f_0} \right| \lesssim \int |\Psi_{j,k}\Psi_{\lambda,u}| \leq \|\Psi_{\lambda,u}\|_\infty \int |\Psi_{j,k}| \lesssim 2^{(\lambda-j)d/2},$$

where the last step is due to the property (iii) of the wavelets bases. When $j < \lambda$, a symmetric bound can be obtained, but a better bound is possible. To see this,

$$\begin{aligned} | \langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle | &\leq 2^{-\lambda(\alpha+d/2)} \max_{u \in \mathcal{K}(\lambda)} 2^{\lambda(\alpha+d/2)} | \langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle | \\ &\leq 2^{-\lambda(\alpha+d/2)} \|\zeta_{j,k}\|_{\infty,\infty,\alpha} \\ &\lesssim 2^{(j-\lambda)(\alpha+d/2)}, \end{aligned}$$

in view of Lemma 5 of Castillo [9] which implies that $\zeta_{j,k} \in \mathcal{B}_{\infty,\infty}^\alpha([0,1]^d)$ with norm $\|\zeta_{j,k}\|_{\infty,\infty,\alpha} \lesssim 2^{j(\alpha+d/2)}$. $\square$

Lemma 6.4. *Let f_0 be a density function bounded away from 0 and ∞ and let P_0 stand for the integration with respect to the density f_0. Then for any $j \leq J_n$ and $k \in \mathcal{K}(j)$, we have*

$$\|\mathcal{A}_{j,k}\|_\infty \lesssim 2^{J_n d/2}, \quad \|\mathcal{A}_{j,k}\|_2 = O(1), \quad \|P_0 \mathcal{A}_{j,k}\|_\infty = O(1).$$

Proof. These bounds are not explicitly stated in Castillo [9] but yet are needed for the conditions of Lemma 6.6 and 6.7. We give a proof here for completeness. Write

$$
\begin{aligned}
\mathcal{A}_{j,k} &= \sum_{\lambda \leq J_n} \sum_{u \in \mathcal{K}(\lambda)} \langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle \Psi_{\lambda,u} \\
&= \sum_{\lambda \leq j} \sum_{u \in \mathcal{K}(\lambda)} \langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle \Psi_{\lambda,u} + \sum_{j < \lambda \leq J_n} \sum_{u \in \mathcal{K}(\lambda)} \langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle \Psi_{\lambda,u}.
\end{aligned}
$$

When $\lambda \leq j$, for any fixed λ, a fixed number of $\{\Psi_{\lambda,u}\}$ have support intersecting with the support of $\zeta_{j,k}$. By Lemma (6.3), $|\langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle| \lesssim 1$, and the property (iii) of the wavelets, the first term on the right hand side can be bounded as

$$\Big\| \sum_{\lambda \leq j} \sum_{u \in \mathcal{K}(\lambda)} \langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle \Psi_{\lambda,u} \Big\|_\infty \lesssim \sum_{\lambda \leq j} \|\Psi_{\lambda,u}\|_\infty \leq \sum_{\lambda \leq j} 2^{\lambda d/2} \lesssim 2^{J_n d/2}.$$

Since $\lambda > j$, by Lemma (6.3) again, the second term can be bounded as

$$
\begin{aligned}
\Big\| \sum_{j < \lambda \leq J_n} \sum_{u \in \mathcal{K}(\lambda)} \langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle \Psi_{\lambda,u} \Big\|_\infty &\lesssim \sum_{j < \lambda \leq J_n} \max_{u \in \mathcal{K}(\lambda)} |\langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle| \Big\| \sum_{u \in \mathcal{K}(\lambda)} |\Psi_{\lambda,u}| \Big\|_\infty \\
&\lesssim \sum_{j < \lambda \leq J_n} 2^{(j-\lambda)(\alpha+d/2)} 2^{\lambda d/2},
\end{aligned}
$$

which is bounded by a constant multiple of $2^{jd/2} \leq 2^{J_n d/2}$. Similarly,

$$
\begin{aligned}
\|\mathcal{A}_{j,k}\|_2^2 &= \sum_{\lambda \leq J_n} \sum_{u \in \mathcal{K}(\lambda)} |\langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle|^2 \\
&= \sum_{\lambda \leq j} \sum_{u \in \mathcal{K}(\lambda)} |\langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle|^2 + \sum_{j < \lambda \leq J_n} \sum_{u \in \mathcal{K}(\lambda)} |\langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle|^2.
\end{aligned}
$$

The first term on the right-hand side is bounded by $\sum_{\lambda \leq j} 2^{(\lambda-j)d}$, which is $O(1)$. The second term is bounded by $\sum_{j < \lambda \leq J_n} 2^{(j-\lambda)2\alpha} \lesssim 2^{-2\alpha} = O(1)$.

For the last statement,

$$
\begin{aligned}
P_0 \mathcal{A}_{j,k} &= \sum_{\lambda \leq J_n} \sum_{u \in \mathcal{K}(\lambda)} \langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle P_0 \Psi_{\lambda,u} \\
&= \sum_{\lambda \leq j} \sum_{u \in \mathcal{K}(\lambda)} \langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle P_0 \Psi_{\lambda,u} + \sum_{j < \lambda \leq J_n} \sum_{u \in \mathcal{K}(\lambda)} \langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle P_0 \Psi_{\lambda,u}.
\end{aligned}
$$

Note that $P_0 \Psi_{\lambda,u} \lesssim \int |\Psi_{\lambda,u}| \lesssim 2^{-\lambda d/2}$, by similar argument, we can obtain $\|P_0 \mathcal{A}_{j,k}\|_\infty = O(1)$. $\qquad\square$

Lemma 6.5. *Let $f_0 \in \mathcal{H}^\alpha([0,1]^d)$ for some $\alpha > 1$, and be bounded away from 0 and ∞. Then for any $j \le J_n$ and $k \in \mathcal{K}(j)$, any density f bounded away from 0 and ∞, the following holds*

$$\|\mathcal{A}_{j,k} - \zeta_{j,k}\|_\infty \lesssim 2^{j(\alpha+d/2)}2^{-\alpha J_n},$$

$$\left| \int_{[0,1]^d} (\mathcal{A}_{j,k} - \zeta_{j,k})(f - f_0) \right| \lesssim (2^{(j-J_n)\alpha} \wedge 2^{-j})\|f - f_0\|_2.$$

Proof. To show the first statement, write

$$\|\mathcal{A}_{j,k} - \zeta_{j,k}\|_\infty = \left\| \sum_{\lambda > J_n} \sum_{u \in \mathcal{K}(\lambda)} \langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle \Psi_{\lambda,u} \right\|_\infty$$

$$\le \sum_{\lambda > J_n} \max_{u \in \mathcal{K}(\lambda)} |\langle \zeta_{j,k}, \Psi_{\lambda,u} \rangle| \left\| \sum_{u \in \mathcal{K}(\lambda)} |\Psi_{\lambda,u}| \right\|_\infty$$

$$\lesssim \sum_{\lambda > J_n} 2^{(j-\lambda)(\alpha+d/2)}2^{\lambda d/2}$$

$$\lesssim 2^{j(\alpha+d/2)}2^{-\alpha J_n},$$

where the third line follows from Lemma 6.3 and the property (i) of the wavelets.

To show the second statement, recall first the support of $\Psi_{j,k}$ is denoted by $S_{j,k}$ whose volume $\mathrm{vol}(S_{j,k}) \lesssim 2^{-jd}$ (since along each coordinate, the length of support is of order 2^{-j}). For any fixed j, k, $\mathcal{A}_{j,k} - \zeta_{j,k}$ is a linear combination of high frequency wavelets ($\lambda > J_n \ge j$). Observe first that for any $\lambda > J_n$, and admissible u, each coordinate of $\Psi_{\lambda,u}$ has a support of length of order $\mathrm{vol}(S_{\lambda,u})^{1/d}$, which is smaller than that of each coordinate of $\zeta_{j,k}$ (whose length is of the order of $\mathrm{vol}(S_{j,k})^{1/d}$), that is, $\mathrm{vol}(S_{\lambda,u})^{1/d} \lesssim R \times \mathrm{vol}(S_{j,k})^{1/d}$ for some constant R. Therefore, along each coordinate, supports of all $\Psi_{\lambda,u}$'s which intersect with that of $\zeta_{j,k}$ along the same coordinate are contained in an interval of length at most $(2R + 1)\mathrm{vol}(S_{j,k})^{1/d}$. Hence, the supports of $\Psi_{\lambda,u}$'s which intersect with the support of $\zeta_{j,k}$ are contained in a hyper-rectangle of volume at most $(2R+1)^d\mathrm{vol}(S_{j,k})$. This implies that the volume of the support of $\zeta_{j,k} - \mathcal{A}_{j,k}$ is bounded by a constant multiple of $\mathrm{vol}(S_{j,k})$.

Let $\Delta_{j,k}$ denote the support of $\zeta_{j,k} - \mathcal{A}_{j,k}$. Now bounding $\mathcal{A}_{j,k} - \zeta_{j,k}$ by its supremum and applying Cauchy-Schwarz inequality,

$$\left| \int_{[0,1]^d} (\mathcal{A}_{j,k} - \zeta_{j,k})(f - f_0)\mathbb{1}_{\Delta_{j,k}} \right| \lesssim \|\mathcal{A}_{j,k} - \zeta_{j,k}\|_\infty \sqrt{\mathrm{vol}(\Delta_{j,k})}\|f - f_0\|_2$$

$$\lesssim 2^{j(\alpha+d/2)}2^{-\alpha J_n}2^{-jd/2}\|f - f_0\|_2$$

$$= 2^{(j-J_n)\alpha}\|f - f_0\|_2.$$

To obtain the bound for the other side, let $D_{j,k} := \Psi_{j,k}/[\bar{f}_0]_{j,k}$, where $[\bar{f}_0]_{j,k} := (\mathrm{vol}(S_{j,k}))^{-1}\int_{S_{j,k}} f_0$, which is bounded below from 0 (by the assumption on f_0). Since $j \le J_n$, by the property of L^2-projection, $\|\zeta_{j,k} - \mathcal{A}_{j,k}\|_2 \le$

$\|\zeta_{j,k} - D_{j,k}\|_2$. Since $D_{j,k}$ has the same support as that of $\Psi_{j,k}$,

$$\|\zeta_{j,k} - \mathcal{A}_{j,k}\|_2^2 \le \int \mathbb{1}_{S_{j,k}}(D_{j,k} - \zeta_{j,k})^2 \le \|D_{j,k} - \zeta_{j,k}\|_\infty^2 \mathrm{vol}(S_{j,k}).$$

To bound the sup-norm,

$$
\begin{aligned}
\|D_{j,k} - \zeta_{j,k}\|_\infty &\lesssim \|\Psi_{j,k}(f_0 - [\bar{f}_0]_{j,k})\|_\infty \\
&\le \|\Psi_{j,k}\|_\infty \Big(\sup_{x \in S_{j,k}} |f_0(x) - [\bar{f}_0]_{j,k}| \Big) \\
&= \|\Psi_{j,k}\|_\infty \Big(\sup_{x \in S_{j,k}} \Big| \int_{S_{j,k}} \frac{f_0(x)}{\mathrm{vol}(S_{j,k})} du - \frac{1}{\mathrm{vol}(S_{j,k})} \int_{S_{j,k}} f_0(u) du \Big| \Big) \\
&\le \|\Psi_{j,k}\|_\infty \frac{1}{\mathrm{vol}(S_{j,k})} \Big(\sup_{x \in S_{j,k}} \int_{S_{j,k}} |f_0(x) - f_0(u)| du \Big) \\
&\lesssim \|\Psi_{j,k}\|_\infty \frac{1}{\mathrm{vol}(S_{j,k})} \Big(\sup_{x \in S_{j,k}} \int_{S_{j,k}} \|x - u\| du \Big) \\
&\lesssim \|\Psi_{j,k}\|_\infty \frac{1}{\mathrm{vol}(S_{j,k})} \Big(\mathrm{vol}(S_{j,k}) \mathrm{vol}(S_{j,k})^{1/d} \Big) \\
&\lesssim 2^{jd/2} 2^{-j},
\end{aligned}
$$

where the fifth line holds because $f_0 \in \mathcal{H}^\alpha([0,1]^d)$ with $\alpha > 1$; and in the second from the last line above, the first factor in the parentheses $\mathrm{vol}(S_{j,k})$ is due to the volume from the support and the second factor $\mathrm{vol}(S_{j,k})^{1/d}$ is due to the maximum length of support along each coordinate (or the diameter of the support). Combing the above results,

$$\left| \int_{[0,1]^d} (\mathcal{A}_{j,k} - \zeta_{j,k})(f - f_0) \right| \lesssim \|\zeta_{j,k} - \mathcal{A}_{j,k}\|_2 \|f - f_0\|_2 \lesssim 2^{-j} \|f - f_0\|_2. \qquad \square$$

Lemma 6.6. *Let f_0 be a density function bounded away from 0 and ∞. Let a_n be a sequence of real numbers with $na_n^2 > 1$, any $n \ge 1$. Let $\{\Pi_n\}$ be a collection of priors on densities restricted to the set $\{f : h(f, f_0) \le a_n\}$. Let $\{\gamma_n\}$ be an arbitrary sequence in $L^\infty([0,1]^d)$. Set $\tilde{\gamma}_n := \gamma_n - P_0\gamma_n$, where P_0 denotes integration with respect to the density f_0. Suppose that for some $m > 0$ and all $n > 1$,*

$$P_0 \tilde{\gamma}_n^2 \le m, \qquad \|\tilde{\gamma}_n\|_\infty \le (4 a_n \log(n+1))^{-1}.$$

Then there exists $C > 0$ depending on m and $\|f_0\|_\infty$ only such that for any $n \ge 1$ and any $|t| \le \log n$,

$$\mathrm{E}^\Pi(e^{t\sqrt{n}\langle f - f_0, \gamma_n \rangle} | \mathbb{D}_n) \le \exp\{Ct^2 + tW_n(\gamma_n)\} \frac{\int e^{\ell_n(f_t) - \ell_n(f_0)} d\Pi_n(f)}{\int e^{\ell_n(f) - \ell_n(f_0)} d\Pi_n(f)}, \qquad (25)$$

where f_t is defined through the expression

$$\log f_t = \log f - t \frac{\tilde{\gamma}_n}{\sqrt{n}} - c\Big(\log f - \frac{t\tilde{\gamma}_n}{\sqrt{n}}\Big).$$

In particular, the result holds for $\gamma_n = \mathcal{A}_{j,k}$ for any j such that $2^j \leq 2^{J_n} \asymp (n/\log n)^{1/(2\alpha+d)}$.

Proof. See the proof of Lemma 3 of Castillo [9]. $\qquad\square$

Lemma 6.7. *Suppose that f_0 is bounded away from 0 and ∞. Let $g_0 = \log f_0$. For $\mathcal{A}_{j,k}$ any element of $L^\infty([0,1]^d)$ such that there exists constants c_1, c_2, for any $j, k \in \mathcal{K}(j)$ with $2^j \leq 2^{J_n} \asymp (n/\log n)^{1/(2\alpha+d)}$, any n sufficiently large,*

$$\|\mathcal{A}_{j,k}\|_\infty \leq c_1 \sqrt{n/\log n}, \qquad \|\mathcal{A}_{j,k}\|_2 \leq c_2.$$

Then for any n sufficiently large,

$$\mathrm{E}_0 \|\Gamma^{J_n} - g_0^{J_n}\|_\infty \lesssim \epsilon_{n,\alpha}.$$

Proof. See the proof of Lemma 7 of Castillo [9], with some necessary adaption using the properties of tensor product wavelet basis. $\qquad\square$

Lemma 6.8. *For $\theta = \{\theta_{\lambda,u} : \lambda \leq J_n, u \in \mathcal{K}(j)\}$ such that $f_\theta \in D_n$, where $f_\theta = \exp(T(x) - c(T))$, $T = \sum_{\lambda \leq J_n} \sum_u \theta_{\lambda,u} \Psi_{\lambda,u}$, and $2^{J_n} \asymp (n/\log n)^{1/(2\alpha+d)}$. For the Gaussian prior under the condition (8), the following holds for some constant $C > 0$*

$$\prod_{\lambda \leq J_n, u \in \mathcal{K}(\lambda)} \frac{\varphi(\theta_{\lambda,u}/\sigma_\lambda)}{\varphi\left(\frac{\theta_{\lambda,u} - t\langle \mathcal{A}_{j,k}, \Psi_{\lambda,u}\rangle/\sqrt{n}}{\sigma_\lambda}\right)} \leq \exp\{C|t| + Ct^2\}.$$

For the log-Lipschitz prior under the condition (7), the upper bound is $e^{C|t|}$ for some $C > 0$.

Proof. See the argument in pages 2078–2080 of Castillo [9], with some necessary adaption using the properties of tensor product wavelet basis and Lemma 6.3. $\qquad\square$

The following result, needed to prove Theorem 4.1, is also of independent interest.

Proposition 6.9. *In the setup of Theorem 4.1, the following assertions hold, for $d = 2, 3$:*

(a) $\mathrm{E}(\|f - \tilde{f}\|_\infty | \mathbb{D}_n) \lesssim n^{-\alpha/(2\alpha+d)}(\log n)^{d/2}$;
(b) $\mathrm{E}_0 \|\tilde{f} - \mathrm{E}_0 \tilde{f}\|_\infty \lesssim n^{-\alpha/(2\alpha+d)}(\log n)^{d/2}$;
(c) $\|\mathrm{E}_0 \tilde{f} - f_0\|_\infty \lesssim n^{-\alpha/(2\alpha+d)}$,

and hence the posterior for f contracts at f_0 at the rate $n^{-\alpha/(2\alpha+d)}(\log n)^{d/2}$ with respect to the supremum distance.

Proof. To avoid the complication of notation, we first demonstrate the proof for the case $d = 2$ and then sketch the proof for the dominating terms for the case $d = 3$ after that.

Let $Z = (f - \tilde{f}) = \sum_j \sum_k \phi_{j,k}(\theta_{j,k} - \tilde{\theta}_{j,k})$. The posterior distribution of Z given the data is a centered Gaussian process that does not depend on $\mathbb{D}_n$. To

ease the notation, we may just use E to denote the expectation with respect to the posterior distribution, i.e, we write $E(\|Z\|_\infty)$ for $E(\|Z\|_\infty | \mathbb{D}_n)$. Hence (a) is the same as showing $E(\|Z\|_\infty) \lesssim n^{-\alpha/(2\alpha+2)} \log n$.

Proof for $d = 2$:

Note that $E(Z^2(x)) = \sum_j \sum_k \mu_j/(n\mu_j + 1)\phi_{j,k}^2(x)$. By the uniform boundedness of the basis functions, $EZ^2(x)$ is bounded by

$$\frac{1}{n} + \sum_{j_1=1}^\infty \frac{1}{n + j_1^{\alpha+1}} + \sum_{j_2=1}^\infty \frac{1}{n + j_2^{\alpha+1}} + \sum_{j_1=1}^\infty \sum_{j_2=1}^\infty \frac{1}{n + (j_1 j_2)^{\alpha+1}}.$$

Let $n_\alpha = n^{1/(\alpha+1)}$. The second term can be bounded by

$$\sum_{j_1=1}^{n_\alpha} \frac{1}{n} + \sum_{j_1 > n_\alpha}^\infty j_2^{-(1+\alpha)} \lesssim n^{-\alpha/(1+\alpha)}.$$

The third term can be bounded in a similar way. For the fourth term, we split in two cases — one is where $j_1 j_2 \leq n_\alpha$ and the other $j_1 j_2 > n_\alpha$. For the first case,

$$\sum_{j_1 j_2 \leq n_\alpha} \frac{1}{n + (j_1 j_2)^{\alpha+1}} \lesssim \frac{1}{n} \#\{(j_1, j_2) : j_1 j_2 \leq n_\alpha, j_1, j_2 \in \mathbb{N}\}.$$

For any fixed j_1 so that $1 \leq j_1 \leq n_\alpha$, the number of j_2 such that $j_2 \leq n_\alpha/j_1$ is bounded by n_α/j_1. Therefore, $n^{-1}\#\{(j_1, j_2) : j_1 j_2 \leq n_\alpha, j_1, j_2 \in \mathbb{N}\} \leq n^{-1} n_\alpha \sum_{j_1=1}^{n_\alpha} j_1^{-1} \lesssim n^{-\alpha/(1+\alpha)} \log n$. For the second case, we can have two scenarios: (i) when $j_1 \geq n_\alpha, j_2 \geq 1$,

$$\sum_{j_1 j_2 > n_\alpha} \frac{1}{n + (j_1 j_2)^{\alpha+1}} \lesssim \sum_{j_1 \geq n_\alpha}^\infty j_1^{-(1+\alpha)} \sum_{j_2=1}^\infty j_2^{-(1+\alpha)} = n^{-\alpha/(1+\alpha)},$$

while for (ii) $j_1 < n_\alpha, j_2 \geq n_\alpha/j_1$,

$$\sum_{j_1=1}^{n_\alpha} \sum_{j_2 > n_\alpha/j_1}^\infty \frac{1}{(j_1 j_2)^{1+\alpha}} = \sum_{j_1=1}^{n_\alpha} j_1^{-(1+\alpha)} \sum_{j_2 > n_\alpha/j_1}^\infty j_2^{-(1+\alpha)}$$

$$= \sum_{j_1=1}^{n_\alpha} j_1^{-(1+\alpha)} \left(\frac{n_\alpha}{j_1}\right)^{-\alpha}$$

$$= n^{-\alpha/(1+\alpha)} \sum_{j_1=1}^{n_\alpha} j_1^{-1},$$

and hence is $O(n^{-\alpha/(1+\alpha)} \log n)$. Thus $\|EZ^2(\cdot)\|_\infty \lesssim n^{-\alpha/(1+\alpha)} \log n$.

We next show that $E|Z(x) - Z(y)|^2$ is bounded by some power of n multiplied by $\|x - y\|^2$. Note that

$$E|Z(x) - Z(y)|^2 = \sum_{j_1=0}^\infty \sum_{j_2=0}^\infty \sum_{k \in \mathcal{K}(j)} \frac{1}{n + (\mu_{j_1} \mu_{j_2})^{-1}} \left|\phi_{j_1,j_2,k}(x) - \phi_{j_1,j_2,k}(y)\right|^2.$$

Clearly, when $j_1 = j_2 = 0$, the summand is bounded by n^{-1}. Consider the summation over $j_1 = 0, j_2 \in \mathbb{N}$. Using the fact that $|\phi_{0,j_2,k}(x) - \phi_{0,j_2,k}(y)| \lesssim |x_2 - y_2|^2 j_2^2$, the summation is bounded by

$$\sum_{j_2=1}^{\infty} \frac{1}{n + j_2^{1+\alpha}} |\phi_{0,j_2,k}(x) - \phi_{0,j_2,k}(y)|^2 \lesssim \sum_{j_2=1}^{\infty} \frac{j_2^2}{n + j_2^{1+\alpha}} |x_2 - y_2|^2 \lesssim n^{(2-\alpha)/(1+\alpha)},$$

because

$$\sum_{j_2=1}^{\infty} \frac{j_2^2}{n + j_2^{1+\alpha}} \lesssim \Big(\sum_{j_2=1}^{n_\alpha} \frac{j_2^2}{n} + \sum_{j_2 > n_\alpha}^{\infty} \frac{j_2^2}{j_2^{1+\alpha}} \Big) \lesssim \Big(n^{(2-\alpha)/(1+\alpha)} + n_\alpha^{2-\alpha} \Big) \lesssim n^{(2-\alpha)/(1+\alpha)}.$$

Notice that for the second inequality above to hold, α is required to be greater than 2. A bound of the same order can be obtained by the summation over $j_1 \in \mathbb{N}, j_2 = 0$. Consider the summation over $j_1, j_2 \in \mathbb{N}$. Using the bound $|\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i| \leq \sum_{i=1}^{n} |a_i - b_i|$ for $|a_i|, |b_i| \leq 1$, we have that

$$\left| \phi_{j_1,k_1}(x_1) \phi_{j_2,k_2}(x_2) - \phi_{j_1,k_1}(y_1) \phi_{j_2,k_2}(y_2) \right| = 2 \sum_{i=1}^{2} \left| \frac{1}{\sqrt{2}} \phi_{j_i,k_i}(x_i) - \frac{1}{\sqrt{2}} \phi_{j_i,k_i}(y_i) \right|$$

$$\lesssim \sum_{i=1}^{2} j_i |x_i - y_i|.$$

Therefore,

$$|\phi_{j_1,j_2,k}(x) - \phi_{j_1,j_2,k}(y)|^2 \lesssim \Big(\sum_{i=1}^{2} j_i |x_i - y_i| \Big)^2 \leq (j_1^2 + j_2^2) \|x - y\|^2$$

by the Cauchy-Schwarz inequality. Hence

$$\sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} \sum_{k} \frac{1}{n + (\mu_{j_1} \mu_{j_2})^{-1}} |\phi_{j_1,j_2,k}(x) - \phi_{j_1,j_2,k}(y)|^2$$

$$\lesssim \|x - y\|^2 \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} \frac{j_1^2 + j_2^2}{n + (j_1 j_2)^{1+\alpha}}.$$

As before, we consider the two cases for the summation. One is the summation over $j_1 j_2 \leq n_\alpha$ and the other is over $j_1 j_2 > n_\alpha$. For the first case,

$$\sum_{j_1 j_2 \leq n_\alpha} \frac{j_1^2 + j_2^2}{n + (j_1 j_2)^{\alpha+1}} \leq \frac{1}{n} \sum_{j_1 j_2 \leq n_\alpha} (j_1^2 + j_2^2) \leq \frac{1}{n} \sum_{j_1}^{n_\alpha} \sum_{j_2 < (n_\alpha/j_1)} (j_1^2 + j_2^2).$$

By a straightforward calculation, it is bounded by $n^{(2-\alpha)/(1+\alpha)}$. For the second case, we again consider two scenarios (i) $j_1 \geq n_\alpha, j_2 \geq 1$ and (ii) $j_2 < n_\alpha, j_2 \geq n_\alpha/j_1$. Proceeding similarly as before, we can get the bound $n^{(2-\alpha)/(1+\alpha)}$. In summary, we obtain that

$$\mathrm{E}|Z(x) - Z(y)|^2 \lesssim n^{(2-\alpha)/(1+\alpha)} \|x - y\|^2.$$

Now, by Lemma A.11 of Yoo and Ghosal [48] with $\delta_n = n^{-p}$ for $p > 0$ sufficiently large, we obtain $\mathrm{E}\|Z\|_\infty^2 \lesssim \log n \times \|\mathrm{E}(Z^2(\cdot))\|_\infty$. By the bound $n^{-\alpha/(1+\alpha)} \log n$ for $\|\mathrm{E}(Z^2(\cdot))\|_\infty$ obtained earlier in the proof, these imply

$$\mathrm{E}\|Z\|_\infty \lesssim n^{-\alpha/(2\alpha+2)} \log n.$$

To prove (b), let $V = \tilde{f} - \mathrm{E}_0\tilde{f}$. Consider a mean-zero Gaussian process given by

$$V = \sum_{j_1=0}^\infty \sum_{j_2=0}^\infty \sum_{k\in\mathcal{K}(j)} \frac{\sqrt{n}\varepsilon_{j,k}}{n + (\mu_{j_1j_2})^{-1}} \phi_{j,k}.$$

Following the same sequence of argument used for Z is used, we can establish that

$$\|\mathrm{E}_0(V^2(\cdot))\|_\infty \lesssim n^{-\alpha/(1+\alpha)} \log n, \tag{26}$$
$$\mathrm{E}_0|V(x) - V(y)|^2 \lesssim n^{(2-\alpha)/(1+\alpha)} \|x - y\|^2.$$

Therefore, $\mathrm{E}_0\|V\|_\infty \lesssim n^{-\alpha/(2\alpha+2)} \log n$.

To prove (c), note that

$$\mathrm{E}_0(\tilde{f}) - f_0 = \sum_{j_1=0}^\infty \sum_{j_2=0}^\infty \sum_{k\in\mathcal{K}(j)} \left(\frac{n}{n + \mu_{j_2}^{-1}} - 1\right)\phi_{j,k}\theta_{0,j,k}.$$

The summand of $j_1 = j_2 = 0$ is bounded by n^{-1}. Considering the summation over $j_1 = 0, j_2 \in \mathbb{N}$,

$$\left\|\sum_{j_2\geq 1}\sum_{k\in\mathcal{K}(j)} \left(\frac{n}{n + \mu_j^{-1}} - 1\right)\phi_{j,k}\theta_{0,j,k}\right\|_\infty$$
$$\lesssim \left\|\sum_{j_2\geq 1}^{n_\alpha}\sum_{k\in\mathcal{K}(0,j_2)} \left(\frac{j_2^{1+\alpha}}{n + j_2^{1+\alpha}}\right)\phi_{j,k}\theta_{0,j,k}\right\|_\infty$$
$$+ \left\|\sum_{j_2>n_\alpha}^\infty\sum_{k\in\mathcal{K}(0,j_2)} \left(\frac{n}{n + j_2^{1+\alpha}} - 1\right)\phi_{j,k}\theta_{0,j,k}\right\|_\infty,$$

where the first term on the right hand side can be bounded as

$$\left\|\sum_{j_2\geq 1}^{n_\alpha}\sum_{k\in\mathcal{K}(0,j_2)} (j_2/n)j_2^\alpha \phi_{j,k}\theta_{0,j,k}\right\|_\infty \leq (n_\alpha/n)\sum_{j_2\geq 1}^{n_\alpha}\sum_{k\in\mathcal{K}(0,j_2)} j_2^\alpha|\theta_{0,j,k}|\|\phi_{j,k}\|_\infty,$$

which is bounded by a constant multiple of $n_\alpha/n = n^{-\alpha/(1+\alpha)}$ by the uniform boundedness of the basis functions and the assumption on the smoothness of f_0 (take $i_1 = 0, i_2 = \alpha$). Since for $j_2 > n_\alpha$, $\sum_{j_2>n_\alpha}\sum_{k\in\mathcal{K}(0,j_2)}|\theta_{0,j,k}| \lesssim n_\alpha^{-\alpha}$, the second term of the right hand side of above display can be bounded as $\sum_{j_2>n_\alpha}\sum_{k\in\mathcal{K}(0,j_2)}|\theta_{0,j,k}| \lesssim n_\alpha^{-\alpha} = n^{-\alpha/(\alpha+1)}$. Therefore

$$\left\|\sum_{j_2\geq 1}\sum_{k\in\mathcal{K}(0,j_2)} \left(\frac{n}{n + \mu_{j_2}^{-1}} - 1\right)\phi_{j,k}\theta_{0,j,k}\right\|_\infty \lesssim n^{-\alpha/(\alpha+1)}.$$

Finally, consider the sum over $j_1, j_2 \in \mathbb{N}$. We consider two scenarios (i) $j_1 j_2 > n_\alpha = n^{1/(\alpha+1)}$ and (ii) $j_1 j_2 \leq n_\alpha$. For the first scenario, by the uniform boundedness of basis functions and the assumption on the smoothness of f_0 (taking $s_1 = s_2 = \alpha/2$), it holds that $\sum_{j_1 j_2 > n_\alpha} \sum_{k \in \mathcal{K}(j)} |\theta_{0,j,k}| \lesssim n_\alpha^{-\alpha/2}$. It is easy to see that

$$\left\| \sum_{j_1 j_2 > n_\alpha} \sum_{k \in \mathcal{K}(j)} \left(\frac{n}{n + \mu_j^{-1}} - 1 \right) \phi_{j,k} \theta_{0,j,k} \right\|_\infty$$

is of the order $n_\alpha^{-\alpha/2} = n^{-\alpha/(2(1+\alpha))}$. In the second scenario, we bound the term as

$$\left\| \sum_{j_1 j_2 \leq n_\alpha} \sum_{k \in \mathcal{K}(j)} \frac{(j_1 j_2)^{1+\alpha/2}}{n + (j_1 j_2)^{1+\alpha}} (j_1 j_2)^{\alpha/2} \phi_{j,k} \theta_{0,j,k} \right\|_\infty$$

$$\lesssim \frac{n_\alpha^{1+\alpha/2}}{n} \left\| \sum_{j_1 j_2 \leq n_\alpha} \sum_{k \in \mathcal{K}(j)} (j_1 j_2)^{\alpha/2} \phi_{j,k} \theta_{0,j,k} \right\|_\infty,$$

which is of the order $n^{-\alpha/(2+2\alpha)}$ by the uniform boundedness of basis functions and the assumption on the smoothness of f_0. In summary, $\|E_0 \tilde{f} - f_0\|_\infty \lesssim n^{-\alpha/(2+2\alpha)}$ as claimed. This completes our proof for the case $d = 2$.

Proof for $d = 3$:

Now we have that $\mathbb{E}Z^2(x)$ is bounded by

$$\frac{1}{n} + \sum_{j_1=1}^\infty \frac{1}{n + j_1^{1+2\alpha/3}} + \sum_{j_2=1}^\infty \frac{1}{n + j_2^{1+2\alpha/3}} + \sum_{j_3=1}^\infty \frac{1}{n + j_3^{1+2\alpha/3}}$$

$$+ \sum_{j_1=1}^\infty \sum_{j_2=1}^\infty \frac{1}{n + (j_1 j_2)^{1+2\alpha/3}} + + \sum_{j_2=1}^\infty \sum_{j_3=1}^\infty \frac{1}{n + (j_2 j_3)^{1+2\alpha/3}}$$

$$+ \sum_{j_1=1}^\infty \sum_{j_3=1}^\infty \frac{1}{n + (j_1 j_3)^{1+2\alpha/3}} + \sum_{j_1=1}^\infty \sum_{j_2=1}^\infty \sum_{j_3=1}^\infty \frac{1}{n + (j_1 j_2 j_3)^{1+2\alpha/3}}.$$

We bound the dominating term $\sum_{j_1=1}^\infty \sum_{j_2=1}^\infty \sum_{j_3=1}^\infty [n + (j_1 j_2 j_3)^{1+2\alpha/3}]^{-1}$. Other terms can be bounded similarly. Let $n_\alpha = n^{3/(2\alpha+3)}$. We consider two cases.

The first case is $j_1 j_2 j_3 \leq n_\alpha$. The upper bound is given by

$$\frac{1}{n} \sum_{j_1 j_2 j_3 \leq n_\alpha} 1 \lesssim \frac{1}{n} (\log n)^2 n_\alpha,$$

by the so-called Piltz divisor problem (Bordellès [6]). Therefore it is bounded by $n^{-2\alpha/(2\alpha+3)} (\log n)^2$.

The second case is $j_1 j_2 j_3 > n_\alpha$. Due to symmetry of the indices, we only need to consider the three sub-cases: (i) $j_1 > n_\alpha, j_2 \geq 1, j_3 \geq 1$, (ii) $j_1 < n_\alpha, j_2 > n_\alpha/j_1, j_3 \geq 1$, (iii) $j_1 < n_\alpha, j_1 j_2 < n_\alpha/j_1, j_3 > n_\alpha/j_1 j_2$.

For (i):

$$\sum_{j_1>n_\alpha}\sum_{j_2=1}\sum_{j_3=1}\frac{1}{(j_1j_2j_3)^{1+2\alpha/3}}\lesssim\sum_{j_1>n_\alpha}\frac{1}{j_1^{1+2\alpha/3}},$$

which can be bounded by $(n_\alpha)^{-2\alpha/3}=n^{-2\alpha/(2\alpha+3)}$.

For (ii): $\sum_{j_1<n_\alpha}\sum_{j_2>n_\alpha/j_1}\sum_{j_3=1}(j_1j_2j_3)^{-(1+2\alpha/3)}$ can be bounded as

$$\sum_{j_1<n_\alpha}\sum_{j_2>n_\alpha/j_1}\frac{1}{(j_1j_2)^{1+2\alpha/3}}=\sum_{j_1<n_\alpha}j_1^{-(1+2\alpha/3)}\sum_{j_2>n_\alpha/j_1}j_2^{-(1+2\alpha/3)}$$

$$\lesssim\sum_{j_1<n_\alpha}j_1^{-(1+2\alpha/3)}\left(\frac{n_\alpha}{j_1}\right)^{-2\alpha/3},$$

which can be bounded by $(n_\alpha)^{-2\alpha/3}\sum_{j_1=1}^{n_\alpha}j_1^{-1}\lesssim(\log n)n^{-2\alpha/(2\alpha+3)}$

For (iii): $\sum_{j_1<n_\alpha}\sum_{j_2<n_\alpha/j_1}\sum_{j_3>n_\alpha/j_1j_2}(j_1j_2j_3)^{-(1+2\alpha/3)}$ can be bounded as

$$\sum_{j_1<n_\alpha}j_1^{-(1+2\alpha/3)}\sum_{j_2<n_\alpha/j_1}j_2^{-(1+2\alpha/3)}\sum_{j_3>n_\alpha/j_1j_2}j_3^{-(1+2\alpha/3)}$$

$$\lesssim n_\alpha^{-2\alpha/3}\sum_{j_1<n_\alpha}j_1^{-1}\sum_{j_2<n_\alpha/j_1}j_2^{-1}$$

$$<n_\alpha^{-2\alpha/3}\sum_{j_1<n_\alpha}j_1^{-1}\sum_{j_2<n_\alpha}j_2^{-1}$$

$$\lesssim n_\alpha^{-2\alpha/3}(\log n)^2,$$

which is $n^{-2\alpha/(2\alpha+3)}(\log n)^2$.

Combining the assertions above, $\|\mathrm{E}Z^2(\cdot)\|_\infty\lesssim n^{-2\alpha/(2\alpha+3)}(\log n)^2$.

We next turn to show that $E|Z(x)-Z(y)|^2$ is bounded by some power of n multiplied by $\|x-y\|^2$. We can split the term as in the $d=2$ case and shall demonstrate here bounding the dominating term

$$\sum_{j_1=1}^\infty\sum_{j_2=1}^\infty\sum_{j_3=1}^\infty\frac{j_1^2+j_2^2+j_3^2}{n+(j_1j_2j_3)^{1+2\alpha/3}}.$$

The first case to consider is when $j_1j_2j_3\leq n_\alpha$. The above term can be bounded as

$$\frac{1}{n}\sum_{j_1j_2j_3\leq n_\alpha}(j_1^2+j_2^2+j_3^2)=\frac{1}{n}\sum_{j_1j_2\leq n_\alpha}\sum_{j_3<n_\alpha/j_1j_2}(j_1^2+j_2^2+j_3^2)$$

$$\lesssim\frac{1}{n}\sum_{j_1\leq n_\alpha}\sum_{j_2\leq n_\alpha/j_1}\left(j_1^2+j_2^2+\left(\frac{n_\alpha}{j_1j_2}\right)^2\right)$$

$$\lesssim\frac{1}{n}\sum_{j_1\leq n_\alpha}\left(j_1^2+\left(\frac{n_\alpha}{j_1}\right)^3+n_\alpha j_1\right),$$

which can be bounded by a constant multiple of n_α^3/n.

The second case to consider is when $j_1 j_2 j_3 > n_\alpha$. Again we only need to consider the three sub-cases: (i) $j_1 > n_\alpha, j_2 \geq 1, j_3 \geq 1$, (ii) $j_1 < n_\alpha, j_2 > n_\alpha/j_1, j_3 \geq 1$, (iii) $j_1 < n_\alpha, j_1 j_2 < n_\alpha/j_1, j_3 > n_\alpha/j_1 j_2$.

For (i): Since $\alpha > d$ is assumed, $\sum_{j_2=1} j_2^{1-2\alpha/d} = O(1)$. We can bound

$$\sum_{j_1 > n_\alpha}^{\infty} \sum_{j_2=1}^{\infty} \sum_{j_3=1}^{\infty} \frac{j_1^2 + j_2^2 + j_3^2}{n + (j_1 j_2 j_3)^{1+2\alpha/3}} \lesssim \sum_{j_1 > n_\alpha} j_1^{1-2\alpha/d} + \sum_{j_1 > n_\alpha} j_1^{-(1+2\alpha/d)}$$
$$\lesssim n_\alpha^{2-2\alpha/d}.$$

For (ii): we can bound

$$\sum_{j_1 < n_\alpha} \sum_{j_2 > n_\alpha/j_1}^{\infty} \sum_{j_3=1}^{\infty} \frac{j_1^2 + j_2^2 + j_3^2}{n + (j_1 j_2 j_3)^{1+2\alpha/3}}$$
$$\lesssim \sum_{j_1 < n_\alpha} \sum_{j_2 > n_\alpha/j_1}^{\infty} \sum_{j_3=1}^{\infty} \frac{j_1^2}{n + (j_1 j_2 j_3)^{1+2\alpha/3}}$$
$$+ \sum_{j_1 < n_\alpha} \sum_{j_2 > n_\alpha/j_1}^{\infty} \sum_{j_3=1}^{\infty} \frac{j_2^2}{n + (j_1 j_2 j_3)^{1+2\alpha/3}}$$
$$+ \sum_{j_1 < n_\alpha} \sum_{j_2 > n_\alpha/j_1}^{\infty} \sum_{j_3=1}^{\infty} \frac{j_3^2}{n + (j_1 j_2 j_3)^{1+2\alpha/3}}.$$

The first term on the right hand side can be bounded as

$$\sum_{j_1 < n_\alpha} \sum_{j_2 > n_\alpha/j_1}^{\infty} j_1^{1-2\alpha/3} j_2^{-(1+2\alpha/3)} \lesssim \sum_{j_1=1}^{n_\alpha} j_1^{1-2\alpha/3} \left(\frac{n_\alpha}{j_1} \right)^{-2\alpha/3},$$

which is bounded by $n_\alpha^{-2\alpha/3} \sum_{j_1=1}^{n_\alpha} j_1 \lesssim n_\alpha^{2-2\alpha/3}$. Similarly, we can show the second and the third term are bounded by $n_\alpha^{2-2\alpha/3}$, with details omitted.

For (iii): as in the case (ii), we bound three terms and shall show for bounding the first one, while the remaining two terms have similar bounds but of smaller order:

$$\sum_{j_1 < n_\alpha} \sum_{j_2 < n_\alpha/j_1} \sum_{j_3 > n_\alpha/j_1 j_2} \frac{j_1^2}{n + (j_1 j_2 j_3)^{1+2\alpha/3}}$$
$$\lesssim \sum_{j_1 < n_\alpha} j_1^{-\left(\frac{2\alpha}{3}-1\right)} \sum_{j_2 < n_\alpha/j_1} j_2^{-\left(\frac{2\alpha}{3}+1\right)} \sum_{j_3 > n_\alpha/j_1 j_2} j_3^{-\left(1+\frac{2\alpha}{3}\right)}$$
$$\lesssim n_\alpha^{-2\alpha/3} \sum_{j_1 < n_\alpha} j_1 \sum_{j_2 < n_\alpha/j_1} j_2^{-1}$$
$$\lesssim n_\alpha^{-2\alpha/3} \sum_{j_1 < n_\alpha} j_1 \sum_{j_2 < n_\alpha} j_2^{-1},$$

$$\lesssim n_\alpha^{2-2\alpha/3} \log n,$$

which is bounded by $n^{9/(2\alpha+3)}$. Therefore, by the same argument to that for the case $d = 2$, now $\mathrm{E}\|Z\|_\infty \lesssim ((\log n)\|\mathrm{E}Z^2(\cdot)\|_\infty)^{1/2}$, which is bounded by a constant multiple of $(\log n)^{3/2} n^{-\alpha/(2\alpha+3)}$ when $d = 3$.

By a similar argument used for bounding Z, we can prove that, for the case $d = 3$, $\|\mathrm{E}_0(V^2(\cdot))\|_\infty \lesssim (\log n)^2 n^{-2\alpha/(3+2\alpha)}$ and also $\mathrm{E}_0(V(x) - V(y))^2 \lesssim (\log n)^2 n^{(6-2\alpha)/(2\alpha+3)}\|x - y\|^2$. Therefore, $\mathrm{E}_0\|V\|_\infty \lesssim (\log n)^{3/2} n^{-\alpha/(2\alpha+3)}$.

Finally, we bound $\|\mathrm{E}_0\tilde{f} - f_0\|_\infty$. It turns out the dominating term is given by the following

$$\sum_{j_1=1} \sum_{j_2=1} \sum_{j_3=1} \Big(\frac{n}{n + (j_1 j_2 j_3)^{1+2\alpha/3}} - 1\Big)\phi_{j,k}\theta_{0,j,k}.$$

This term is bounded by

$$\sum_{j_1 j_2 j_3 \leq n_\alpha} \Big(\frac{(j_1 j_2 j_3)^{1+2\alpha/3}}{n + (j_1 j_2 j_3)^{1+2\alpha/3}}\Big)\phi_{j,k}\theta_{0,j,k}$$
$$+ \sum_{j_1 j_2 j_3 > n_\alpha} \Big(\frac{n}{n + (j_1 j_2 j_3)^{1+2\alpha/3}} - 1\Big)\phi_{j,k}\theta_{0,j,k}.$$

By assumption $\sum_{j_1} \sum_{j_2} \sum_{j_3} (j_1 j_2 j_3)^{\alpha/3}|\theta_{0,j,k}| \lesssim L$ and the uniform boundedness of the basis functions, the first term on the right hand side of the above expression is bounded by

$$\frac{n_\alpha^{1+\alpha/3}}{n}\Big|\sum_{j_1} \sum_{j_2} \sum_{j_3}(j_1 j_2 j_3)^{\alpha/3}|\theta_{0,j,k}|\Big| \lesssim n^{-\alpha/(2\alpha+3)}.$$

Noting by the same assumption, $n_\alpha^{\alpha/3} \sum_{j_1 j_2 j_3 > n_\alpha}(j_1 j_2 j_3)^{\alpha/3}|\theta_{0,j,k}| \lesssim L$. Hence, the second term is bounded by

$$\Big\|\sum_{j_1 j_2 j_3 > n_\alpha} |\phi_{j,k}||\theta_{0,j,k}|\Big\|_\infty \lesssim n_\alpha^{-\alpha/3} = n^{-\alpha/(2\alpha+3)}.$$

Therefore, $\|\mathrm{E}_0\tilde{f} - f_0\|_\infty \lesssim n^{-\alpha/(2\alpha+3)}$. $\qquad\square$

Proof of Theorem 4.1. We show the proof for $d = 2$ and sketch the proof for $d = 3$ after that. As in the proof of Proposition 6.9, let $Z = (f - \tilde{f})$. The proof of this theorem consists of two parts. In Part (1) we show that $\mathrm{E}\|Z\|_\infty \gtrsim (\log n)n^{-\alpha/(2(1+\alpha))}$. In Part (2) we prove the stated assertions.

Proof for $d = 2$:

Proof of Part (1): Let $1 \leq l_1, l_2, l'_1, l'_2 \leq n_\alpha^\tau$ for some $0 < \tau < 1$, where $n_\alpha = n^{1/(1+\alpha)}$ and $(l_1, l_2) \neq (l'_1, l'_2)$. Without loss of generality, we take $l_1 > l'_1$. Consider

$$\mathrm{E}\left(\Big|Z\Big(\frac{l_1}{n_\alpha}, \frac{l_2}{n_\alpha}\Big) - Z\Big(\frac{l'_1}{n_\alpha}, \frac{l'_2}{n_\alpha}\Big)\Big|^2\right)$$

$$= \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} \sum_{k \in \mathcal{K}(j)} \frac{1}{n + (\mu_{j_1}\mu_{j_2})^{-1}} \left| \phi_{j_1,j_2,k}\left(\frac{l_1}{n_\alpha}, \frac{l_2}{n_\alpha}\right) - \phi_{j_1,j_2,k}\left(\frac{l_1'}{n_\alpha}, \frac{l_2'}{n_\alpha}\right) \right|^2.$$

It is clearly lower bounded by

$$\sum_{j_1 j_2 \leq n_\alpha} \frac{1}{n + (j_1 j_2)^{1+\alpha}} \left| 4\sin\left(\frac{2\pi l_1 j_1}{n_\alpha}\right)\sin\left(\frac{2\pi l_2 j_2}{n_\alpha}\right) - 4\sin\left(\frac{2\pi l_1' j_1}{n_\alpha}\right)\sin\left(\frac{2\pi l_2' j_2}{n_\alpha}\right) \right|^2. \tag{27}$$

Fix j_1 such that $1 \leq j_1 \leq n_\alpha/2$, thus $j_2 \leq n_\alpha/j_1$. Since $\sin(2\pi j_1 l_1/n_\alpha) = 0$ if and only if $2j_1 l_1/n_\alpha = m$ for some $m \in \mathbb{Z}$. Let

$$\mathcal{J}_1 := \{j_1 : 1 \leq j_1 \leq n_\alpha/2, \frac{j_1}{n_\alpha} \neq \frac{m}{2l_1}, m \in \{1, 2, \ldots, \lfloor l_1 \rfloor\}\}.$$

Note that $\sin(x) = \frac{b}{a}\sin(y), a \neq 0$ if and only if $x = (-1)^m \sin^{-1}(\frac{b}{a}\sin(y)) + m\pi$ for some $m \in \mathbb{Z}$. Let $a = \sin(2\pi j_1 l_1/n_\alpha), b = \sin(2\pi j_1 l_1'/n_\alpha)$. Fix $j_1 \in \mathcal{J}_1$ so that $a \neq 0$. Pick some $\epsilon > 0$ arbitrarily small. Note that

$$\left| a\sin\left(\frac{2\pi l_2 j_2}{n_\alpha}\right) - b\sin\left(\frac{2\pi l_2' j_2}{n_\alpha}\right) \right| > \epsilon \tag{28}$$

whenever the following inequality hold:

$$\left| 2\pi j_2 \frac{l_2}{n_\alpha} - (-1)^m \sin^{-1}\left(\frac{b}{a}\sin\left(2\pi j_2 \frac{l_2'}{n_\alpha}\right)\right) - m\pi \right| > \delta \text{ for some } m \in \mathbb{Z}, \tag{29}$$

for some small positive constant δ. Let $s := j_2/n_\alpha \in (0, 1/j_1]$, as $j_2 \leq n_\alpha/j_1$. The expression (28) holds when s is outside the neighborhood of the zeros of $2\pi s l_2 = (-1)^m \sin^{-1}\left(\frac{b}{a}\sin\left(2\pi s l_2'\right)\right) - m\pi$ for an integer m. Since $s \in (0, 1/j_1]$, l_2, l_2' are fixed, there are at most finitely many such zeros. Thus when s is not in a set which is a union of finitely many arbitrary small intervals, (28) holds. Note that the set is not changing with n_α for fixed l_1, l_1', l_2, l_2'. The complement set has positive Lebesgue measure. Therefore, for any $j_1 \in \mathcal{J}_1$, the fraction of $\{j_2 : 1 \leq j_2 \leq n_\alpha/j_1\}$ such that (28) holds is bounded away from zero. Hence the expression in (27) is at least a constant multiple of

$$\frac{1}{2n} \sum_{j_1 \in \mathcal{J}_1} \sum_{j_2 \leq n_\alpha/j_1} 1 \gtrsim \frac{\epsilon^2}{2n} \sum_{j_1 \in \mathcal{J}_1} n_\alpha/j_1 \gtrsim \frac{n_\alpha}{n} \sum_{j_1 \in \mathcal{J}_1} \frac{1}{j_1}.$$

Note also that $\sum_{j_1 \in \mathcal{J}_1} 1/j_1 = \sum_{j_1=1}^{n_\alpha/2} 1/j_1 - \sum_{j_1 \in \mathcal{J}_1^c} 1/j_1$. To bound the second term

$$\sum_{j_1 \in \mathcal{J}_1^c} \frac{1}{j_1} \leq \sum_{m=1}^{\lfloor l_1 \rfloor} \frac{2l_1}{mn_\alpha} \leq \frac{2l_1}{n_\alpha}\left(\sum_{m=1}^{n_\alpha^\tau} \frac{1}{m}\right) \lesssim \sum_{m=1}^{n_\alpha^\tau} \frac{1}{m}.$$

Therefore, $\sum_{j_1 \in \mathcal{J}_1} 1/j_1 \geq \sum_{j_1=1}^{n_\alpha/2} 1/j_1 - \sum_{m=1}^{n_\alpha^\tau} 1/m \gtrsim \log n$. Summarizing, the expression in (27) is at least a multiple of $(\log n)n^{-\alpha/(1+\alpha)}$.

Let $U_{l_1,l_2} = \sqrt{2}c^{-1/2}(\log n)^{-1/2}n^{\alpha/(2(1+\alpha))}Z(l_1/n_\alpha, l_2/n_\alpha)$. Then

$$\mathrm{E}(U_{l_1,l_2} - U_{l'_1,l'_2})^2 \geq 2 = \mathrm{E}(V_{l_1,l_2} - V_{l'_1,l'_2})^2,$$

where V_{l_1,l_2} are i.i.d. standard normal random variables. Note that, for a sequence of i.i.d normal random variables $W_1,\ldots,W_n \sim \mathrm{N}(0,\sigma^2)$, for some universal constants C_1 and $C_2 > 0$,

$$C_1\sigma\sqrt{\log n} \leq \mathrm{E}\max_{1\leq j\leq n} W_j \leq \mathrm{E}\max_{1\leq j\leq n} |W_j| \leq C_2\sigma\sqrt{\log n};$$

see (3.14) of Ledoux and Talagrand [29]. Now by Slepian's lemma (cf. Corollary 3.14 of Ledoux and Talagrand [29] or Proposition A.2.6 of Van der Vaart and Wellner [47])

$$\mathrm{E}(\max_{l_1,l_2} U_{l_1,l_2}) \gtrsim \mathrm{E}(\max_{l_1,l_2} V_{l_1,l_2}) \gtrsim \sqrt{\log(n_\alpha^2)} \asymp \sqrt{\log n}.$$

Since $\mathrm{E}(\max_{l_1,l_2} Z(l_1/n_\alpha, l_2/n_\alpha)) \gtrsim (\log n)^{1/2}n^{-\alpha/(2(1+\alpha))}\mathrm{E}(\max_{l_1,l_2} U_{l_1,l_2})$, we obtain

$$\mathrm{E}\|Z\|_\infty \geq \mathrm{E}\Big(\max_{l_1,l_2} Z\Big(\frac{l_1}{n_\alpha}, \frac{l_2}{n_\alpha}\Big)\Big) \gtrsim n^{-\alpha/(2(1+\alpha))}\log n.$$

Proof of Part (2): Note that since for $0 < \gamma < 1/2$, $R_{n,\gamma}$ is greater than the posterior median of $\|Z\|_\infty$. By the estimate (26) and Borell's inequality (cf. Proposition A.2.1 of Van der Vaart and Wellner [47]), the posterior mean of $\|Z\|_\infty$ and the posterior median of $\|Z\|_\infty$ are of the same order. As $(1-\gamma)$-quantile of $\|Z\|_\infty$ is larger than its median, by Part (1), it follows that

$$R_{n,\gamma} \gtrsim \mathrm{E}\|Z\|_\infty \gtrsim n^{-\alpha/(2(1+\alpha))}\log n.$$

Let $\epsilon_n = n^{-\alpha/(2(1+\alpha))}\log n$. By Borell's inequality (see Proposition A.2.1 of Van der Vaart and Wellner [47]),

$$\begin{aligned}
\Pi(\mathcal{L}_c \notin \mathcal{C}_\gamma|\mathbb{D}_n) &= \Pi(f \notin \mathbb{C}_\gamma|\mathbb{D}_n)\\
&= \mathrm{P}(\|Z\|_\infty - \mathrm{E}\|Z\|_\infty > \rho R_{n,\gamma} - \mathrm{E}\|Z\|_\infty)\\
&\leq \exp[-C^2\epsilon_n^2/c_n],
\end{aligned}$$

where $c_n = \|\mathrm{E}Z^2(\cdot)\|_\infty$ is bounded by $(\log n)n^{-\alpha/(1+\alpha)}$ in view of (26), and C is a constant, whose positivity can be ensured if ρ is chosen sufficiently large. Therefore, the above posterior probability tends to zero.

Finally,

$$\begin{aligned}
\mathrm{P}_0(\mathcal{L}_c(f_0) \in \mathcal{C}_\gamma) &= \mathrm{P}_0(\|\tilde{f} - f_0\|_\infty \leq \rho R_{n,\gamma})\\
&= \mathrm{P}_0(\|\mathrm{E}_0\tilde{f} - f_0 + \tilde{f} - \mathrm{E}_0\tilde{f}\|_\infty \leq \rho R_{n,\gamma}) \to 1,
\end{aligned}$$

by the Parts (b) and (c) of Proposition 6.9, establishing the coverage of $\mathcal{C}_\gamma$.

Observe that $\mathcal{C}_\gamma \subset \bar{\mathcal{C}}_\gamma$, because any $\mathcal{L}_c \in \mathcal{C}_\gamma$ is induced by some f such that $\|f - \tilde{f}\|_\infty \leq \rho R_{n,\gamma}$. In view of Part (a) of Proposition 6.9, by the third assertion

of Proposition A.2.1 of Van der Vaart and Wellner [47], $R_{n,\gamma}$ is bounded by $\sqrt{8\mathrm{E}(\|Z\|_\infty^2|\mathbb{D}_n)\log(2/\gamma)}$, which is of the order $n^{-\alpha/(2(1+\alpha))}\log n$. By Lemma 2.2, it then establishes the claim about the size of the credible set.

Proof for $d = 3$:

The key step is to prove that $\mathrm{E}\|Z\|_\infty \gtrsim n^{-\alpha/(2\alpha+3)}(\log n)^{3/2}$. The rest of the proof follows similarly. In order to lower bound $\mathrm{E}\|Z\|_\infty$, let $1 \le l_1, l_2, l_3, l_1', l_2', l_3' \le n_\alpha^\tau$ for some $0 < \tau < 1$, where $n_\alpha = n^{3/(3+2\alpha)}$ and $(l_1, l_2, l_3) \ne (l_1', l_2', l_3')$. Note that

$$\mathrm{E}\left(\left|Z\left(\frac{l_1}{n_\alpha}, \frac{l_2}{n_\alpha}, \frac{l_3}{n_\alpha}\right) - Z\left(\frac{l_1'}{n_\alpha}, \frac{l_2'}{n_\alpha}, \frac{l_3'}{n_\alpha}\right)\right|^2\right)$$

is lower bounded by

$$\sum_{j_1 j_2 j_3 \le n_\alpha} \frac{1}{n + (j_1 j_2 j_3)^{1+2\alpha/3}}\left|8\sin\left(\frac{2\pi l_1 j_1}{n_\alpha}\right)\sin\left(\frac{2\pi l_2 j_2}{n_\alpha}\right)\sin\left(\frac{2\pi l_3 j_3}{n_\alpha}\right)\right.$$
$$\left. - 8\sin\left(\frac{2\pi l_1' j_1}{n_\alpha}\right)\sin\left(\frac{2\pi l_2' j_2}{n_\alpha}\right)\sin\left(\frac{2\pi l_3' j_3}{n_\alpha}\right)\right|^2. \tag{30}$$

Choose $j_1 < n_\alpha^{\bar\tau}, j_2 < n_\alpha/j_1, j_3 < n_\alpha/j_1 j_2$, where $0 < \bar\tau < 1$. Let

$$\mathcal{J}_1 := \{j_1 : 1 \le j_1 \le n_\alpha, \frac{j_1}{n_\alpha} \ne \frac{m}{2l_1}, m \in \{1, 2, \ldots, \lfloor 2l_1\rfloor\}\},$$

$$\mathcal{J}_2(j_1) := \{j_2 : 1 \le j_2 \le \frac{n_\alpha}{j_1}, \frac{j_2}{n_\alpha} \ne \frac{m}{2l_1}, m \in \{1, 2, \ldots, \lfloor 2l_2/j_1\rfloor\}\}.$$

Let $a = \sin(2\pi j_1 l_1/n_\alpha)\sin(2\pi j_2 l_2/n_\alpha), b = \sin(2\pi j_1 l_1'/n_\alpha)\sin(2\pi j_2 l_2'/n_\alpha)$. Fix $j_1 \in \mathcal{J}_1, j_2 \in \mathcal{J}_2(j_1)$ so that $a \ne 0$. Pick some $\epsilon > 0$ arbitrarily small. Note that

$$\left|a\sin\left(\frac{2\pi l_3 j_3}{n_\alpha}\right) - b\sin\left(\frac{2\pi l_3' j_3}{n_\alpha}\right)\right| > \epsilon, \tag{31}$$

whenever the following inequality hold:

$$\left|2\pi j_3\frac{l_3}{n_\alpha} - (-1)^m\sin^{-1}\left(\frac{b}{a}\sin\left(2\pi j_3\frac{l_3'}{n_\alpha}\right)\right) - m\pi\right| > \delta \text{ for some } m \in \mathbb{Z}, \tag{32}$$

for some small positive constant δ. Let $s := j_3/n_\alpha \in (0, 1/(j_1 j_2)]$, as $j_3 \le n_\alpha/(j_1 j_2)$. Arguing as in the case $d = 2$, for any $j_1 \in \mathcal{J}_1$ and $j_2 \in \mathcal{J}_2(j_1)$, the fraction of $\{j_3 : 1 \le j_3 \le n_\alpha/j_1 j_2\}$ such that (31) holds is bounded away from zero. Therefore, the expression in (30) is at least a constant multiple of $\epsilon^2 n_\alpha/(2n)\sum_{j_1 \mathcal{J}_1}\sum_{j_2\in\mathcal{J}_2(j_1)} 1/(j_1 j_2)$. Note also that $\sum_{j_2\in\mathcal{J}_2(j_1)} 1/j_2 = \sum_{j_2=1}^{n_\alpha/j_1} 1/j_2 - \sum_{j_2\in\mathcal{J}_2^c(j_1)} 1/j_2$ where the second term can be bounded as

$$\sum_{j_2\in\mathcal{J}_2^c(j_1)} \frac{1}{j_2} \le \frac{2l_2}{n_\alpha}\left(\sum_{m=1}^{n_\alpha^\tau/j_1} \frac{1}{m}\right) \lesssim \sum_{m=1}^{n_\alpha^\tau/j_1} \frac{1}{m}.$$

Therefore, $\sum_{j_2 \in \mathcal{J}_2^c(j_1)} 1/j_2 \gtrsim \sum_{j_2=n_\alpha^\tau/j_1}^{n_\alpha/j_1} 1/j_2 \gtrsim \log(n_\alpha/j_1) \gtrsim \log n$. Similarly, $\sum_{j_1 \mathcal{J}_1} 1/j_1 \gtrsim \log n$. Combining above results, $(30) \gtrsim n^{-2\alpha/(2\alpha+3)}(\log n)^2$. By the arguments used for $d = 2$, we have $\mathrm{E}\|Z\|_\infty \gtrsim n^{-\alpha/(2\alpha+3)}(\log n)^{3/2}$.   □

Proof of Theorem 4.2. We take $J_k \asymp J_n^* = (n/\log n)^{1/(2\alpha+d)}, k = 1, \ldots, d$. For $0 < \gamma < 1/2$, by argument in the proof of Theorem of 5.3 of Yoo and Ghosal [48], one can establish that

$$R_{n,\gamma} \asymp \mathrm{E}\Big(\|f - \tilde{f}\|_\infty | \mathbb{D}_n, \sigma^2\Big) \asymp \sqrt{\frac{\log n}{n}}(J_n^*)^{d/2} = (\log n/n)^{\alpha/(2\alpha+d)}$$

in probability.

Recall that from (15), $\Pi(f|\mathbb{D}_n, \hat{\sigma}^2) \sim \mathrm{GP}(\tilde{f}, \hat{\sigma}^2\Sigma)$. Arguing as in the proof of Theorem 4.1,

$$\Pi(\mathcal{L}_c \notin \mathcal{C}_\gamma | \mathbb{D}_n, \hat{\sigma}^2) \leq \exp[-C^2 R_{n,\gamma}^2/c_n],$$

where $c_n = \sup_x \mathrm{var}((f - \tilde{f})(x)|\mathbb{D}_n, \hat{\sigma}^2)$ is bounded by a constant multiple of

$$\sup_x \Sigma(x, x) \lesssim \sup_x \|b(x)\|^2 \|(B^T B + \Lambda_0^{-1})^{-1}\|_{(2,2)} \lesssim (J_n^*)^d/n.$$

Therefore, the above posterior probability is bounded by $e^{-c \log n}$ for some $c > 0$, and hence tends to zero. In addition,

$$\mathrm{P}_0(\mathcal{L}_c(f_0) \in \mathcal{C}_\gamma) = \mathrm{P}_0(\|\tilde{f} - f_0\|_\infty \leq \rho R_{n,\gamma}) \to 1,$$

by the third assertion of Theorem 5.3 of Yoo and Ghosal [48] establishing the coverage of $\mathcal{C}_\gamma$. To see $\mathcal{C}_\gamma \subset \bar{\mathcal{C}}_\gamma$, as in the last theorem, note that any $\mathcal{L}_c \in \mathcal{C}_\gamma$ is induced by some f such that $\|f - \tilde{f}\|_\infty \leq \rho R_{n,\gamma}$. In view of Lemma 2.2 and Theorem 4.4 of Yoo and Ghosal [48] and its proof, the assertion about the size of the credible region immediately follows.   □

References

[1] Amparo Baíllo, Juan A Cuesta-Albertos, and Antonio Cuevas. Convergence rates in nonparametric estimation of level sets. *Statistics & Probability Letters*, 53(1):27–35, 2001. MR1843338

[2] Yannick Baraud. Confidence balls in Gaussian regression. *The Annals of Statistics*, 32(2):528–551, 2004. MR2060168

[3] Eduard Belitser. On coverage and local radial rates of credible sets. *The Annals of Statistics*, 45(3):1124–1151, 2017. MR3662450

[4] Eduard Belitser and Nurzhan Nurushev. Needles and straw in a haystack: robust confidence for possibly sparse sequences. *Bernoulli*, 26:191–225, 2020. MR4036032

[5] Gérard Biau, Benoît Cadre, and Bruno Pelletier. Exact rates in density support estimation. *Journal of Multivariate Analysis*, 99(10):2185–2207, 2008. MR2463383

[6] Olivier Bordellès. An inequality for the class number. *Journal of Inequalities in Pure and Applied Mathematics*, 7(3), 2006. MR2257286

[7] Benoît Cadre. Kernel estimation of density level sets. *Journal of Multivariate Analysis*, 97(4):999–1023, 2006. MR2256570

[8] T Tony Cai and Mark G Low. An adaptation theory for nonparametric confidence intervals. *The Annals of statistics*, 32(5):1805–1840, 2004. MR2102494

[9] Ismaël Castillo. On Bayesian supremum norm contraction rates. *The Annals of Statistics*, 42(5):2058–2091, 2014. MR3262477

[10] Ismaël Castillo and Richard Nickl. On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *The Annals of Statistics*, 42(5):1941–1969, 2014. MR3262473

[11] Ismaël Castillo and Stéphanie van der Pas. Multiscale Bayesian survival analysis. https://arxiv.org/pdf/2005.02889.pdf, 2020.

[12] Laurent Cavalier. Nonparametric estimation of regression level sets. *Statistics, A Journal of Theoretical and Applied Statistics*, 29(2):131–160, 1997. MR1484386

[13] Yen-Chi Chen, Christopher R Genovese, and Larry Wasserman. Density level sets: Asymptotics, inference, and visualization. *Journal of the American Statistical Association*, 112(520):1–13, 2017. MR3750891

[14] Victor Chernozhukov, Han Hong, and Elie Tamer. Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, 75(5):1243–1284, 2007. MR2347346

[15] Victor Chernozhukov, Emre Kocatulum, and Konrad Menzel. Inference on sets in finance. *Quantitative Economics*, 6(2):309–358, 2015. MR3388152

[16] A Cohen, I Daubechies, and P Vial. Wavelets on the interval and fast wavelet transform. *Applied and Computational Harmonic Analysis*, 1:58–81, 1993. MR1256527

[17] Antonio Cuevas and Ricardo Fraiman. A plug-in approach to support estimation. *The Annals of Statistics*, 25(6):2300–2312, 1997. MR1604449

[18] Antonio Cuevas, Manuel Febrero, and Ricardo Fraiman. Estimating the number of clusters. *Canadian Journal of Statistics*, 28(2):367–382, 2000. MR1792055

[19] Antonio Cuevas, Wenceslao González-Manteiga, and Alberto Rodríguez-Casal. Plug-in estimation of general level sets. *Australian & New Zealand Journal of Statistics*, 48(1):7–19, 2006. MR2234775

[20] Ghislaine Gayraud and Judith Rousseau. Rates of convergence for a Bayesian level set estimation. *Scandinavian Journal of Statistics*, 32(4):639–660, 2005. MR2232347

[21] Subhashis Ghosal and Aad van der Vaart. Convergence rates of posterior distributions for noniid observations. *The Annals of Statistics*, 35(1):192–223, 2007. MR2332274

[22] Subhashis Ghosal and Aad van der Vaart. *Fundamentals of Nonparametric Bayesian Inference*, volume 44. Cambridge University Press, 2017. MR3587782

[23] Subhashis Ghosal, Jayanta K Ghosh, and Aad W van der Vaart. Conver-

gence rates of posterior distributions. *The Annals of Statistics*, 28(2):500–531, 2000. MR1790007

[24] Evarist Giné and Richard Nickl. Rates of contraction for posterior distributions in L^r-metrics, $1 \leq r \leq \infty$. *The Annals of Statistics*, 39(6):2883–2911, 2011. MR3012395

[25] Evarist Giné and Richard Nickl. *Mathematical Foundations of Infinite-dimensional Statistical Models*, volume 40. Cambridge University Press, 2015. MR3588285

[26] Wolfgang Härdle, Gerard Kerkyacharian, Dominique Picard, and Alexander Tsybakov. *Wavelets, Approximation, and Statistical Applications*, volume 129. Springer Science & Business Media, 2012. MR1618204

[27] Marc Hoffmann, Judith Rousseau, and Johannes Schmidt-Hieber. On adaptive posterior concentration rates. *The Annals of Statistics*, 43(5):2259–2295, 2015. MR3396985

[28] Hanna Jankowski and Larissa Stanberry. Confidence regions in level set estimation. http://www.math.yorku.ca/~hkj/Research/level.pdf, 2012.

[29] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: isoperimetry and processes*. Springer Science & Business Media, 2013. MR2814399

[30] Ker-Chau Li. Honest confidence regions for nonparametric regression. *The Annals of Statistics*, 17(3):1001–1008, 1989. MR1015135

[31] Wei Li and Subhashis Ghosal. Posterior contraction and credible sets for filaments of regression functions. *Electronic Journal of Statistics*, 14(1):1707–1743, 2020. MR4083733

[32] Enno Mammen and Wolfgang Polonik. Confidence regions for level sets. *Journal of Multivariate Analysis*, 122:202–214, 2013. MR3189318

[33] Enno Mammen and Alexandre B Tsybakov. Smooth discrimination analysis. *The Annals of Statistics*, 27(6):1808–1829, 1999. MR1765618

[34] David M Mason and Wolfgang Polonik. Asymptotic normality of plug-in level set estimates. *The Annals of Applied Probability*, 19(3):1108–1142, 2009. MR2537201

[35] Richard Nickl and Kolyan Ray. Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. *The Annals of Statistics*, 48(3):1383–1408, 2020. MR4124327

[36] Wolfgang Polonik. Measuring mass concentrations and estimating density contour clusters-an excess mass approach. *The Annals of Statistics*, 23(3):855–881, 1995. MR1345204

[37] Wolfgang Polonik and Zailong Wang. Estimation of regression contour clusters—an application of the excess mass approach to regression. *Journal of Multivariate Analysis*, 94(2):227–249, 2005. MR2167913

[38] Kolyan Ray. Adaptive Bernstein–von Mises theorems in Gaussian white noise. *The Annals of Statistics*, 45(6):2511–2536, 2017. MR3737900

[39] Philippe Rigollet and Régis Vert. Optimal rates for plug-in estimators of density level sets. *Bernoulli*, 15(4):1154–1178, 2009. MR2597587

[40] Alessandro Rinaldo and Larry Wasserman. Generalized density clustering. *The Annals of Statistics*, 38(5):2678–2722, 2010. MR2722453

[41] Weining Shen and Subhashis Ghosal. Posterior contraction rates of density derivative estimation. *Sankhyā, Series A*, 79(2):336–354, 2017. MR3707424

[42] Xiaotong Shen and Larry Wasserman. Rates of convergence of posterior distributions. *The Annals of Statistics*, 29(3):687–714, 2001. MR1865337

[43] Scott C. Singh, A. and R. Nowak. Adaptive Hausdorff estimation of density level sets. *The Annals of Statistics*, 37(5B):2769–2782, 2009. MR2541446

[44] Suzanne Sniekers and Aad van der Vaart. Adaptive Bayesian credible sets in regression with a Gaussian process prior. *Electronic Journal of Statistics*, 9(2):2475–2527, 2015. MR3425364

[45] Botond Szabó, AW van der Vaart, and JH van Zanten. Frequentist coverage of adaptive nonparametric Bayesian credible sets. *The Annals of Statistics*, 43(4):1391–1428, 2015. MR3357861

[46] Alexandre B Tsybakov. On nonparametric estimation of density level sets. *The Annals of Statistics*, 25(3):948–969, 1997. MR1447735

[47] Aad W Van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes*. Springer, New York, 1996. MR1385671

[48] William Weimin Yoo and Subhashis Ghosal. Supremum norm posterior contraction and credible sets for nonparametric multivariate regression. *The Annals of Statistics*, 44(3):1069–1102, 2016. MR3485954

[49] William Weimin Yoo and Subhashis Ghosal. Bayesian mode and maximum estimation and accelerated rates of contraction. *Bernoulli*, 25(3):2330–2358, 2019. MR3961250

[50] William Weimin Yoo, Judith Rousseau, and Vincent Rivoirard. Adaptive supremum norm posterior contraction: Wavelet spike-and-slab and anisotropic Besov spaces. https://arxiv.org/abs/1708.01909, 2018.