ADAPTIVE BAYESIAN ESTIMATION VIA BLOCK PRIOR

By CHAO GAO AND HARRISON H. ZHOU

Yale University

A novel block prior is proposed for adaptive Bayesian estimation. The prior does not depend on the smoothness of the function and the sample size. It puts sufficient prior mass near the true signal and automatically concentrates on its effective dimension. A rate-optimal posterior contraction is obtained in a general framework, which includes density estimation, white noise model, Gaussian sequence model, Gaussian regression and spectral density estimation.

1. Introduction. Bayesian nonparametric estimation is attracting more and more attention in a wide range of applications. We consider a fundamental question in Bayesian nonparametric estimation: Is it possible to construct a prior such that the posterior contracts to the truth with the exact optimal rate and at the same time is adaptive regardless of the unknown smoothness? We provide a positive answer to this question by designing a block prior on coefficients of orthogonal series expansion of the function.

The block prior is inspired by the blockwise estimation approach to achieve adaptation in the frequentist literature. For example, Efroimovich (1986) proposed blockwise shrinkage estimation to achieve adaptive density estimation. Subsequent works in more general settings include Hall, Kerkyacharian and Picard (1998), Cai (1999), Cai and Zhou (2009) and references therein. However, the block prior considered in this work is more than a “Bayesian version” of the frequentist results on adaptive estimation. We illustrate this point by two simple facts. First, the result of posterior contraction is stronger than the minimax convergence rate of a point estimator, since rate-optimal posterior contraction with exponential tail probability implies rate-optimal point estimation. Second, the proposed block prior does not depend on the sample size $n$. We do not need to artificially estimate every coefficient by 0 after a pre-determined frequency, which is different from some frequentist approaches. For example, the blockwise James-Stein estimator is used for adaptive estimation in Gaussian sequence model, but the procedure is only applied to the first $n$ parameters, and estimate the rest of parameters by 0, otherwise the mean squared error would be unbounded.

AMS 2000 subject classifications: 62G07, 62G20

Keywords and phrases: Bayesian nonparametrics, adaptive estimation, block prior
The asymptotic analysis of Bayesian nonparametric estimation was pioneered by Barron (1988) and Barron, Schervish and Wasserman (1999). They proved as long as the prior satisfies a Kullback-Leibler property, and there exists a sieve such that it has finite metric entropy and receives most of the prior mass, then the posterior distribution is consistent in Hellinger distance. Ghosal, Ghosh and van der Vaart (2000) extended this method and proved posterior rate of convergence. The Kullback-Leibler property of the prior is to lower bound the denominator of the posterior probability, and the construction of a sieve with proper metric entropy bound is for the testing of the true distribution and its alternatives so that the numerator can be upper bounded. This line of development follows the early works of Le Cam (1973) and Schwartz (1965) on testing theory and Bayesian estimation. Alternative methods that do not explicitly use the testing theory for Bayesian consistency and rate of convergence are referred to Walker (2004) and Zhang (2006) and references therein.

Adaptive Bayesian estimators over Sobolev balls or Hölder balls are considered in the literature. There are two main approaches in these works. The first one is to put a hyper-prior on the smoothness index $\alpha$. As is shown in Scricciolo (2006) and Ghosal, Lember and van der Vaart (2008), minimax rate can be achieved, but the set of $\alpha$ is restricted to be countable or even finite. The second approach is to put a prior on $k$, where $k$ is the number of basis functions for approximation, or the model dimension. This is called sieve prior in Shen and Wasserman (2001). Examples of using sieve prior includes Kruijer and van der Vaart (2008) and Rivoirard and Rousseau (2012). Their procedures are adaptive over all $\alpha$, but the rates have extra logarithmic terms. Other recent works in Bayesian adaptive estimation include van der Vaart and van Zanten (2009), de Jonge and van Zanten (2010), Kruijer, Rousseau and van der Vaart (2010), Rousseau (2010) and Shen, Tokdar and Ghosal (2013).

Compared to the approaches in the literature, the proposed block prior is adaptive over a continuum of smoothness, and its posterior contraction is exactly rate-optimal. This settles down a standing issue in Bayesian nonparametrics whether the logarithmic term in the contraction rate is intrinsically necessary or not. The framework for the applications of the block prior is very general. It includes density estimation, white noise, Gaussian sequence, regression and spectral density estimation.

We obtain adaptive Bayesian estimation under a Sobolev ball assumption. Assume that $f$ is a function on the unit interval $[0, 1]$. Let $\{\phi_j\}$ be the trigonometric orthogonal basis of $L^2[0, 1]$, and define $\theta_j = \int f \phi_j$ for each $j$. 
The Sobolev ball is specified as

\[ E_\alpha(Q) = \left\{ f \in L^2[0,1] : \sum_{j=1}^{\infty} j^{2\alpha} \theta_j^2 \leq Q^2, \text{ with } \theta_j = \int f \phi_j \text{ for each } j \right\}. \]

Under a general framework, we construct a prior \( \Pi \), which satisfies the Kullback-Leibler (KL) property and it automatically concentrates on the effective dimension of the signal \( f_0 \), then as a consequence the minimax posterior contraction rate is obtained, i.e.,

\[ P_{f_0}^{(\alpha)} \left( \|f - f_0\| > M n^{-\frac{\alpha}{2\alpha + \tau}} |X^n| \right) \rightarrow 0, \]

where the loss function \( \| \cdot \| \) is the \( l^2 \)-norm. The posterior tail probability is exponentially small as discussed in Section 4.

The key idea behind the adaptive block prior is as follows. We want to design a prior that puts sufficient mass near the true function with a certain smoothness level \( \alpha \). On the other hand, we need the prior to automatically concentrates on the effective dimension of the true signal \( f_0 \), so that suitable tests can be established according to the spirit conveyed by Barron, Schervish and Wasserman (1999) and Ghosal, Ghosh and van der Vaart (2000). The intuitive way of mixing the smoothness \( \alpha \) or mixing the model dimension \( k \) results in either an extra logarithmic term in the posterior contraction rate as a price of adaptation, or the inability to adaptive to an uncountable set of \( \alpha \). Our strategy is to mix the whole block of neighboring Fourier coefficients as a way to learn the smoothness, by borrowing information from the neighbors. This is called “information pooling” in Cai (2008). Cai (2008) argued that “information pooling” is necessary to achieve adaptation in the frequentist context.

The paper is organized as follows. In Section 2, we first introduce a preliminary block prior \( \bar{\Pi} \), which satisfies the Kullback-Leibler property and concentrates on the effective dimension of the truth, then present the key result of this paper, adaptive rate-optimal posterior contraction for a slightly modified prior \( \Pi \) under a general framework. As applications of the main results, we study adaptive Bayesian estimation of various models, including density estimation, white noise, Gaussian sequence model, regression and spectral density estimation in Section 3. Section 4 discussed the posterior tail probability bound and an extension of the theory to Besov balls. The main body of the proofs are presented in Section 5 and Section 6. Some auxiliary results are proved in the supplementary material.
1.1. Notations. Throughout the paper, $\mathbb{P}$ and $\mathbb{E}$ are generic probability and expectation operators, which are used whenever the distribution is clear in the context. Small and big case letters denote constants which may vary from line to line. We won’t pay attention to the values of constants which do not affect the result, unless otherwise specified. Notice these constants may or may not be universal, which we shall make clear in the context. The function $f$ and its Fourier coefficients $\theta = \{\theta_j\}$ are used interchangeably. We say $f$ is distributed by $\Pi$ if the corresponding $\theta \sim \Pi$. In the same way, the function space and the parameter space of $f$ and $\theta$ will not be distinguished.

The norm $\|\cdot\|$ denotes both the $l^2$-norm of $f$ and the $l^2$-norm of $\theta$. For two probabilities $P_1$ and $P_2$ with densities $p_1$ and $p_2$, we use the following divergences throughout the paper,

\[
D(P_1, P_2) = P_1 \log \frac{p_1}{p_2},
\]

\[
V(P_1, P_2) = P_1 \left( \log \frac{p_1}{p_2} - D(P_1, P_2) \right)^2,
\]

\[
H(P_1, P_2) = \left( \int (\sqrt{p_1} - \sqrt{p_2})^2 \right)^{1/2}.
\]

We use $\theta_j$ and $\theta_{0j}$ to indicate the $j$-th entries of vectors $\theta = \{\theta_j\}$ and $\theta_0 = \{\theta_{0j}\}$ respectively. The bold notation $\theta_k$ represents the vector $\{\theta_{jk}\}_{j \in B_k}$ for the $k$-th block. The rate $\epsilon_n$ is always the minimax rate $\epsilon_n^2 = n^{-\frac{\alpha}{2\alpha + 1}}$.

2. Main Results. In this section, we first give some necessary backgrounds of Bayes nonparametric estimation, then introduce a block prior and the result of adaptive posterior contraction.

2.1. Background. Suppose we have data $X^n \sim P_{f_0}^{(n)}$, and the distribution $P_{f_0}^{(n)}$ has density $p_{f_0}^{(n)}$ with respect to a dominating measure. The posterior distribution for a prior $\Pi$ is defined to be

\[
\Pi(A|X^n) = \frac{\int_A p_{f_0}^{(n)}(X^n)d\Pi(f)}{\int p_{f_0}^{(n)}(X^n)d\Pi(f)}, \text{ where } X^n \sim P_{f_0}^{(n)}.
\]

We need to bound the expectation of $\Pi\left(d(f, f_0) > M\epsilon_n|X^n\right)$ in this paper. To bound this quantity, it is sufficient to upper bound the numerator and lower bound the denominator. Following Barron, Schervish and Wasserman...
(1999) and Ghosal, Ghosh and van der Vaart (2000), this involves three steps:

1. Show the prior \( \Pi \) puts sufficient mass near the truth, i.e., we need
   \[
   \Pi(K_n) \geq \exp\left(-C_1 n \epsilon_n^2\right),
   \]
   where \( K_n = \left\{ P(\theta_n, f) \leq n \epsilon_n^2, V(P(\theta_n, f)) \leq n \epsilon_n^2 \right\} \).

2. Choose an appropriate set \( F_n \), and show the prior is essentially supported on \( F_n \) in the sense that
   \[
   \Pi(F_n^c) \leq \exp\left(-C_2 n \epsilon_n^2\right).
   \]
   This controls the complexity of the prior.

3. Construct a testing function \( \phi_n \) for the following testing problem
   \[
   H_0 : f = f_0 \quad \text{vs.} \quad H_1 : f \in \text{supp}(\Pi) \cap F_n \quad \text{and} \quad d(f, f_0) > M \epsilon_n.
   \]
   The testing error needs to be well controlled in the sense that
   \[
   P_f(\phi_n) \vee \sup_{f \in H_1} P_f(1 - \phi_n) \leq \exp\left(-C_3 n \epsilon_n^2\right).
   \]

Note that the constants \( C_1, C_2, \text{and} C_3 \) are different in these three steps above. Step 1 lower bounds the prior concentration near the truth, which leads to a lower bound for the denominator \( \int \frac{p_f^{(n)}}{p_{f_0}^{(n)}} (X^n) d\Pi(f) \). It is originated from Schwartz (1965). Step 2 and Step 3 are mainly for upper bounding the numerator \( \int_A \frac{p_f^{(n)}}{p_{f_0}^{(n)}} (X^n) d\Pi(f) \). The testing idea in Step 3 is initialized by Le Cam (1973) and Schwartz (1965). Step 2 goes back to Barron (1988), who proposes the idea to choose an appropriate \( F_n \) to regularize the alternative hypothesis in the test, otherwise the testing function for Step 3 may never exist (see Le Cam (1973) and Barron (1989)).

2.2. The Block Prior \( \tilde{\Pi} \). Given a sequence \( \theta = (\theta_1, \theta_2, \ldots) \) in the Hilbert space \( l^2 \). Define the blocks to be \( B_k = \{l_k, \ldots, l_{k+1} - 1\} \), and \( \{1, 2, 3, \ldots\} = \cup_{k=0}^{\infty} B_k \). Define the block size of the \( k \)-th block to be \( n_k = l_{k+1} - l_k = |B_k| \).

Remember the notation \( \theta_k \) represents the vector \( \{\theta_j\}_{j \in B_k} \). The block prior \( \tilde{\Pi} \) on the function \( f \) is induced by a distribution on its Fourier sequence \( \{\theta_j\} \).

For each \( k \), let \( g_k \) be a one-dimensional density function on \( \mathbb{R}^+ \).

We describe \( \tilde{\Pi} \) as follows.

\[
A_k \sim g_k \quad \text{independently for each} \ k,
\]
\[ \theta_k|A_k \sim N(0, A_k I_{n_k}) \] independently for each \( k \),
where \( I_{n_k} \) is the \( n_k \times n_k \) identity matrix. In this work, we specify \( l_k \) to be \( l_k = [e^k] \). The sequence of densities \( \{g_k\} \) is used to mix the scale parameter \( A_k \) for each block, and we call them mixing densities. Our theory covers a class of mixing densities. The mixing density class \( \mathcal{G} \) contains all \( \{g_k\} \) satisfying the following properties:

1. There exists \( c_1 > 0 \) such that, for any \( k \) and \( t \in [e^{-k^2}, e^{-k}] \),
   \[
   g_k(t) \geq \exp(-c_1 e^k).
   \]

2. There exists \( c_2 > 0 \), such that for any \( k \),
   \[
   \int_0^\infty t g_k(t) dt \leq 4 \exp(-c_2 k^2).
   \]

3. There exists \( c_3 > 0 \), such that for any \( k \),
   \[
   \int_{e^{-k^2}}^{\infty} g_k(t) dt \leq \exp(-c_3 e^k).
   \]

For a function \( f_0 \in E_\alpha(Q) \), define the set

\[
\mathcal{F}_n = \mathcal{F}_n(\beta) = \left\{ \theta : \sum_{j > (n\beta^{-1})^{1/2\alpha+1}} (\theta_j - \theta_0j)^2 \leq \epsilon_n^2 \right\}.
\]

We have the following theorem characterizing the property of \( \bar{\Pi} \).

**Theorem 2.1.** For the block prior \( \bar{\Pi} \) with mixing densities \( \{g_k\} \in \mathcal{G} \), let \( f_0 \in E_\alpha(Q) \) for some \( \alpha, Q > 0 \), then there exists a constant \( C > 0 \) such that

\[
\bar{\Pi} \left\{ \sum_{j=1}^\infty (\theta_j - \theta_0j)^2 \leq \epsilon_n^2 \right\} \geq \exp\left(-Cn\epsilon_n^2\right),
\]

and

\[
\bar{\Pi}\left(\mathcal{F}_n^c\right) \leq 2 \exp\left(- (C + 4)n\epsilon_n^2\right),
\]

for sufficiently large \( n \) whenever \( \beta \leq \left(\frac{c_3}{2(C+4)}\right)^{2\alpha+1} \), with \( c_3 \) defined in (2.3).
Remark 2.1. The theorem presents two properties of the block prior $\bar{\Pi}$. Property (2.5) says the prior gives sufficient mass near the true signal $f_0$. This is also recognized as the K-L condition once the Kullback-Leibler divergence is upper bounded by the $l^2$-norm in the support of the prior. Property (2.6) says the prior concentrates on the effective dimension of the true signal $f_0$ automatically. In Bayesian nonparametric theory, a testing argument is needed to prove posterior contraction rate. Such test can be established on a sieve receiving most of the prior mass. In (2.6), the set $F_n$ can be used as such a sieve.

Remark 2.2. When the smoothness $\alpha$ is known, a well-known prior $\Pi_\alpha = \bigotimes_{j=1}^{\infty} N(0, j^{-2\alpha-1})$ is used in the literature. It can be shown that this prior satisfies (2.5). The block prior $\bar{\Pi}$ satisfies (2.5) and (2.6), and it does not depend on the smoothness $\alpha$. Thus it is fully adaptive.

We claim that the mixing density class $G$ is not empty by presenting an example (Figure 1).

\begin{equation}
\begin{aligned}
g_k(t) &= \begin{cases} 
    e^{k^2} \left( \exp(-e^k) - T_k \right) t + T_k, & 0 \leq t \leq e^{-k^2}; \\
    \exp(-e^k), & e^{-k^2} < t \leq e^{-k}; \\
    0, & t > e^{-k}.
\end{cases}
\end{aligned}
\end{equation}

The value of $T_k$ is specified as

\begin{equation}
T_k = 2e^{k^2} - 2\exp(-e^k + k^2 - k) + \exp(-e^k).
\end{equation}

The following proposition is proved in the supplementary material.

Proposition 2.1. The densities $\{g_k\}$ defined in (2.7) satisfies (2.1), (2.2) and (2.3). Thus, $G$ is not empty.

2.3. Adaptive Posterior Contraction of the Modified Block Prior $\Pi$. In order to prove posterior contraction rate, it is essential to construct a suitable test. A preliminary test is first constructed in a local neighborhood. Then a global test is established by combining all the local tests when the metric entropy is well controlled. We say the distance $d$ satisfies the testing property with respect to the prior $\Pi$ and the truth $f_0$ if and only if there exists some constants $L > 0$ and $\xi \in (0, 1/2)$, such that for any $f_1 \in \text{supp}(\Pi)$ satisfying $d(f_0, f_1) > \epsilon_n$, we have

\begin{equation}
P_{f_0}^{(n)} \phi_n \leq \exp \left( -Ln^2(f_0, f_1) \right),
\end{equation}
Fig 1. The plot of the mixing density function $A_k \sim g_k$ defined in (2.7).

(2.10) $\sup_{\{f \in \text{supp}(\Pi) : d(f, f_1) \leq \xi d(f_0, f_1)\}} P_f^{(n)}(1 - \phi_n) \leq \exp\left(-Lnd^2(f_0, f_1)\right),$

for some testing function $\phi_n$. Then, a global test can be constructed for $H_0 : f = f_0$ against $H_1 = \{f \in \mathcal{F}_n \cap \text{supp}(\Pi) : d(f, f_0) > M\epsilon_n\}$ as long as $d(f_1, f_2) \asymp ||f_1 - f_2||$ for any $f_1$ and $f_2$. The equivalence of $d$ and $||\cdot||$ may not be true for $d$ being Hellinger distance or total variation. We thus consider a modification of the block prior $\bar{\Pi}$, denoted as $\Pi$, so that $d$ and $||\cdot||$ are equivalent in the support of the modified block prior $\Pi$. Define

$$\Pi(A) = \frac{\bar{\Pi}(D \cap A)}{\bar{\Pi}(D)},$$

where the constraint set $D$ needs to be designed case by case such that

$$D(P_{f_1}^{(n)}, P_{f_2}^{(n)}) \leq bn||f_1 - f_2||^2, \quad V(P_{f_1}^{(n)}, P_{f_2}^{(n)}) \leq bn||f_1 - f_2||^2,$$

$$b^{-1}d(f_1, f_2) \leq ||f_1 - f_2|| \leq bd(f_1, f_2),$$

for some constant $b > 1$. We give a specific choice of $D$ for each model considered in this paper. Another crucial property of $D$ we need is that $\Pi$
inherits properties (2.5) and (2.6) from $\Pi$. It is obvious that (2.6) is still true for $\Pi$ as long as $\Pi(D) > 0$. Therefore, one only needs to check (2.5), which is usually not hard as we will see in all the examples in Section 3. A general theorem covers all examples in Section 3 is stated as follows.

**Theorem 2.2.** For the block prior $\Pi$ with mixing densities $\{g_k\} \in \mathcal{G}$, define $\Pi(A) = \frac{\Pi(D \cap A)}{\Pi(D)}$ with the constraint set $D$ satisfying the properties above. Let the distance $d$ satisfy the testing property (2.9) and (2.10). Assume that, for any $f_0 \in E_{\alpha}(Q) \cap D$ with $\alpha \in (\alpha^*, \infty)$ and $Q \in (0, Q^*)$, the prior $\Pi$ inherits properties (2.5) and (2.6) from $\Pi$ for some $C > 0$. Then, for any such $f_0$, there exists $M > 0$, such that

$$P_{f_0}^{(n)} \left( d(f, f_0) > M n^{-\frac{\alpha}{2n+1}} \left| X^n \right) \right) \rightarrow 0.$$ 

**Remark 2.3.** We note that the range $\alpha \in (\alpha^*, \infty)$ and $Q \in (0, Q^*)$ is the adaptive region for the prior $\Pi$. It is determined by the constraint set $D$ and by whether properties (2.5) and (2.6) can be inherited from $\Pi$ to $\Pi$. In some examples such as the white noise model, the modification by $D$ is not needed, so that we have $\Pi = \Pi$. This will result in $\alpha^* = 0$ and $Q^* = \infty$, and thus the prior may adapt to all Sobolev balls. In the regression and the density estimation models, $\alpha^*$ needs to be larger than $1/2$, and $Q^*$ can be chosen arbitrarily large by properly picking the corresponding $D$. For the spectral density estimation, we need $\alpha^* > 3/2$. See Section 3 for details.

### 3. Applications.

Given the experiment $\left( (X^{(n)}, A^{(n)}, P_f^{(n)}) : f \in E_\alpha(Q) \right)$, and observation $X^n \sim P_{f_0}^{(n)}$, we estimate the function $f_0$ by an adaptive Bayesian procedure. The goal is to achieve the minimax posterior contraction rate without knowing the smoothness $\alpha$. In this section, we consider the following examples:

1. **Density Estimation.** The observations $X_1, ..., X_n$ are i.i.d. distributed according to the density

   $$p_f(t) = \frac{e^{f(t)}}{\int e^{f(t)} dt},$$

   for some function $f$ in a Sobolev ball.

2. **White Noise.** The observation $Y_t^{(n)}$ is from the following process

   $$dY_t^{(n)} = f(t) dt + \frac{1}{\sqrt{n}} dW_t,$$
where $W_t$ is the standard Wiener process.

3. **Gaussian Sequence.** We have independent observations

$$X_i = \theta_i + n^{-1/2}Z_i, \quad i \in \mathbb{N},$$

where $\{\theta_i\}$ are Fourier coefficients of $f$, and $\{Z_i\}$ are i.i.d. standard Gaussian variables.

4. **Gaussian Regression.** The design is uniform $X \sim U[0,1]$. Given $X$, $Y|X \sim N(f(X),1)$. The observations are i.i.d. pairs $(X_1,Y_1),\ldots,(X_n,Y_n)$.

5. **Spectral Density.** The observations are stationary Gaussian time series $X_1,\ldots,X_n$ with mean 0 and auto-covariance $\eta_{h}(g) = \int_{-\pi}^{\pi} e^{ih\lambda}g(\lambda)d\lambda$.

The spectral density $g$ is modeled by $g = \exp \left( f \right)$ for some symmetric $f$ in a Sobolev ball.

The above models have similar frequentist estimation procedures, which is due to the deep fact that they are asymptotically equivalent to each other under minor regularity assumptions. References for asymptotic equivalence theory include Brown and Low (1996), Nussbaum (1996), Brown et al. (2002), Brown et al. (2004) and Golubev, Nussbaum and Zhou (2010).

3.1. **Density Estimation.** Let $P_f^{(n)}$ be the product measure

$$P_f^{(n)} = \bigotimes_{i=1}^{n} P_f.$$

The data is i.i.d. $X^n = (X_1,\ldots,X_n) \sim \bigotimes_{i=1}^{n} P_{f_0}$. Let $P_f$ be dominated by Lebesgue measure $\mu$, and it has density function

$$p_f(t) = \frac{e^{f(t)}}{\int_{0}^{1} e^{f(t)} \mu(dt)}.$$

Consider the Fourier expansion $f = \sum_j \theta_j \phi_j$, and the density $p_f$ can be written in the form of infinite dimensional exponential family

$$p_f(t) = \exp \left( \sum_j \theta_j \phi_j(t) - \psi(\theta) \right),$$

where

$$\psi(\theta) = \int_{0}^{1} e^{\sum_j \theta_j \phi_j(t)} \mu(dt).$$

Notice the first Fourier base function is $\phi_1(t) = 1$. It is easy to see that different $\theta_1$’s correspond to the same $p_f$. For identifiability, we set $\theta_1 = 0,$
so that we have \(\int f(t) \mu(dt) = \sum_{j \geq 2} \theta_j \int \phi_j(t) dt = 0\). We use the modified block prior \(\Pi(A) = \frac{\Pi(D \cap A)}{\Pi(D)}\) with the constraint set

\[
D = \left\{ \theta : \sum_{j=1}^{\infty} |\theta_j| < B \right\},
\]

for some constant \(B > 0\). The next lemma shows that the modified block prior \(\Pi\) inherits properties (2.5) and (2.6) from \(\bar{\Pi}\).

**Lemma 3.1.** For \(\alpha^* > 1/2\), define the constant

\[
\gamma = \left( \sum_{j=1}^{\infty} j^{-2\alpha^*} \right)^{1/2} < \infty.
\]

For any \(f_0 \in E_\alpha(Q)\), with \(\alpha \geq \alpha^*\) and \(3\gamma Q \leq B\), there is a constant \(C > 0\), such that

\[
\Pi \left\{ \sum_{j=1}^{\infty} (\theta_{0j} - \theta_j)^2 \leq \epsilon_n^2 \right\} \geq \exp \left( - C n \epsilon_n^2 \right),
\]

and

\[
\Pi \left( \mathcal{F}_n^c \right) \leq 2 \exp \left( - (C + 4)n \epsilon_n^2 \right).
\]

For density estimation, it is natural to use Hellinger distance as the testing distance \(d\). The next lemma establishes equivalence among various distances and divergences under \(D\) defined in (3.1).

**Lemma 3.2.** On the set \(D\), there exists a constant \(b > 1\), such that

\[
D(P_{f_1}, P_{f_2}) \leq b ||\theta_1 - \theta_2||^2, \quad V(P_{f_1}, P_{f_2}) \leq b ||\theta_1 - \theta_2||^2,
\]

\[
b^{-1} H(P_{f_1}, P_{f_2}) \leq ||\theta_1 - \theta_2|| \leq b H(P_{f_1}, P_{f_2}).
\]

We will prove the above two lemmas in the supplementary material. The main result of posterior contraction for density estimation is stated as follows.

**Theorem 3.1.** Let \(\alpha^* > 1/2\) be fixed, and \(\gamma\) is the associated constant defined in (3.2). For any \(f_0 \in E_\alpha(Q)\), with \(\alpha \geq \alpha^*\) and \(B \geq 3\gamma Q\), there is a constant \(M > 0\), such that

\[
P_{f_0}^n \Pi \left( H(P_f, P_{f_0}) > M \epsilon_n |X_1, ..., X_n \right) \to 0.
\]
Proof. Let \( d(f_1, f_2) = H(P_{f_1}, P_{f_2}) \). According to the testing theory in Le Cam (1973) and Ghosal, Ghosh and van der Vaart (2000), the distance \( d \) satisfies testing property (2.9) and (2.10). Since \( \alpha \geq \alpha^* \) and \( 3 \gamma Q \leq B \) implies \( E_\alpha(Q) \subset D \), the conclusion is directly following from Theorem 2.2.

Remark 3.1. The prior \( \Pi \) depends on the value of \( B \), which determines the range of adaptation. For any \( \alpha^* > 1/2 \) and \( Q^* > 0 \), we can choose \( B \) satisfying \( B \geq 3 \gamma Q \) (\( \gamma \) depends on \( \alpha^* \)), such that the prior \( \Pi \) is adaptive for all \( E_\alpha(Q) \) with \( \alpha \geq \alpha^* \) and \( Q \leq Q^* \).

3.2. White Noise. We let \( P_f^{(n)} \) be the distribution of the following process
\[
dY_t^{(n)} = f(t)dt + \frac{1}{\sqrt{n}}dW_t, \quad t \in [0, 1],
\]
where \( W_t \) is the standard Wiener process and the signal has Fourier expansion \( f = \sum_j \theta_j \phi_j \). This model is the simplest and most studied nonparametric model. It is equivalent to the Gaussian sequence model. Since the log density ratio has form
\[
\log \frac{P_{f_0}^{(n)}}{P_f^{(n)}} = n \int (f_0(t) - f(t))dY_t^{(n)} - \frac{n}{2} ||f_0||^2 + \frac{n}{2} ||f||^2,
\]
it is easy to calculate the divergence \( D \) and \( V \). We have
\[
D(P_{f_0}^{(n)}, P_f^{(n)}) = \frac{1}{2} n ||f - f_0||^2, \quad V(P_{f_0}^{(n)}, P_f^{(n)}) = n ||f - f_0||^2.
\]
In the white noise model, it is natural to use the \( l_2 \) norm as the testing distance \( d \). The following lemma is from Lemma 5 in Ghosal and van der Vaart (2007).

Lemma 3.3. Let \( \phi_n = \left\{ 2 \int (f_1(t) - f_0(t))dY_t^{(t)} > ||f_1||^2 - ||f_0||^2 \right\} \). Then we have
\[
P_{f_0}^{(n)} \phi_n \leq 1 - \Phi(\sqrt{n}||f_1 - f_0||/2)
\]
\[
\sup_{\{f : ||f - f_0|| \leq ||f_1 - f_0||/4\}} P_f^{(n)}(1 - \phi_n) \leq 1 - \Phi(\sqrt{n}||f_1 - f_0||/4),
\]
where \( \Phi \) is the standard Gaussian cumulative distribution function.
By the property of Gaussian tail (see Lemma 1 of Mason and Zhou (2012)), we have

\[ 1 - \Phi(\sqrt{nL||f_1 - f_0||}) \leq \frac{1}{\sqrt{nL||f_1 - f_0||}} e^{-\frac{1}{2}L^2n||f_1 - f_0||^2} \leq e^{-\frac{1}{2}L^2n||f_1 - f_0||^2}, \]

provided \( \sqrt{nL||f_1 - f_0||} > 1 \), which is true because we only need to test those \( f_1 \) with \( ||f_1 - f_0|| > M\epsilon_n \), and we have \( \sqrt{n\epsilon_n} \rightarrow \infty \). Therefore, in the white noise model, the distance satisfying (2.9) and (2.10) is the \( l^2 \) norm.

Considering that the divergence \( D(P_{f_0}^{(n)}, P_f^{(n)}) \) and \( V(P_{f_0}^{(n)}, P_f^{(n)}) \) are also \( l^2 \) norm, we reach the following conclusion.

**Theorem 3.2.** In the white noise model, for any \( f_0 \in E_\alpha(Q) \), with some \( \alpha > 0 \) and \( Q > 0 \), there exists a constant \( M > 0 \), such that

\[ P_{f_0}^{(n)} \bar{\Pi}\left(||f - f_0|| > M\epsilon_n|Y_t^{(n)}\right) \rightarrow 0. \]

Hence, this is a case that we have adaptation for all Sobolev balls.

3.3. **Gaussian Sequence.** The Gaussian sequence model is equivalent to the while noise model. We present this case just for illustration of the theory. Given \( f = \sum_j \theta_j \phi_j \), the model \( P_f^{(n)} \) is in a product form

\[ (3.3) \quad P_f^{(n)} = \bigotimes_{i=1}^\infty P_{\theta_i}^{(n)} = \bigotimes_{i=1}^\infty N(\theta_i, n^{-1}). \]

Thus, the observations are independent Gaussian variables in the form

\[ X_i = \theta_i + n^{-1/2}Z_i, \quad i \in \mathbb{N}, \]

where \( \{Z_i\} \) are i.i.d. standard Gaussian variables. The divergence in this case is easy to calculate.

\[
D(P_{f_0}^{(n)}, P_f^{(n)}) = \sum_{i=1}^\infty D(N(\theta_{0i}, n^{-1}), N(\theta_i, n^{-1}))
\]

\[ = \frac{n}{2}||\theta_0 - \theta||^2, \]

\[
V(P_{f_0}^{(n)}, P_f^{(n)}) = P_{f_0}^{(n)} \left( \log \frac{P_{f_0}^{(n)}}{P_f^{(n)}} - \frac{n}{2}||\theta_0 - \theta||^2 \right)^2
\]

\[ = P_{f_0}^{(n)} \left( -\frac{n}{2}||X - \theta_0||^2 + \frac{n}{2}||X - \theta||^2 - \frac{n}{2}||\theta_0 - \theta||^2 \right)^2
\]

\[ = n||\theta_0 - \theta||^2. \]
and they are exactly the $l^2$ norm. Define

$$\phi_n(X) = \begin{cases} \|X - \theta_1\|^2 < \|X - \theta_0\|^2 \\ \|X\|^2 > \|\theta_1\|^2 - \|\theta_0\|^2 \end{cases} = \begin{cases} X^T (\theta_1 - \theta_0) > \|\theta_1\|^2 - \|\theta_0\|^2 \end{cases}.$$

We observe this is exactly the same test in the white noise model, and thus Lemma 3.3 applies here. Therefore,

$$P^{(n)}_{f_0} \phi_n \leq e^{-\frac{1}{2}n\|\theta_0 - \theta_1\|^2},$$

$$\sup_{\{\theta : \|\theta - \theta_1\| \leq \|\theta_1 - \theta_0\|/4\}} P^{(n)}_f (1 - \phi_n) \leq e^{-\frac{1}{4}n\|\theta_0 - \theta_1\|^2}.$$

The $d$ satisfying the testing property (2.9) and (2.10) can be chosen as the $l^2$ norm. We thus reach the following conclusion.

**Theorem 3.3.** In the Gaussian sequence model, for any $f_0 \in E_\alpha(Q)$, with some $\alpha > 0$ and $Q > 0$, there exists a constant $M > 0$, such that

$$P^{(n)}_{f_0} \Pi \left( \|\theta - \theta_0\| > M\epsilon_n |X_1, X_2, ... \right) \rightarrow 0.$$

We have adaptation for all Sobolev balls.

**3.4. Gaussian Regression.** We consider uniform random design instead of fixed design, because the random design allows simple connection between various divergences and the $l^2$ distance. The model $P^{(n)}_f$ gives i.i.d. observations $(X_1, Y_1), ..., (X_n, Y_n)$ with distribution

$$X \sim U[0, 1], \ Y | X \sim N(f(X), 1).$$

The theory is easily extended to general random design with $X \sim q$ for some density $q$ on $[0, 1]$ bounded from above and below. We choose the uniform design for simplicity of presentation. The function has Fourier expansion $f = \sum_j \theta_j \phi_j$ so that we can apply the modified block prior on $f$. Let $P_f$ be the distribution of a single observation, and we need to calculate $D(P_{f_0}, P_f)$ and $V(P_{f_0}, P_f)$. Let $\phi$ be the standard normal density, and we have

$$D(P_{f_0}, P_f) = \int_0^1 \int \phi(y - f_0(x)) \log \frac{\phi(y - f_0(x))}{\phi(y - f(x))} dy dx$$

$$= \frac{1}{2} \int_0^1 (f_0(x) - f(x))^2 dx$$

$$= \frac{1}{2} \|f - f_0\|^2.$$
and

\[
V(P_{f_0}, P_f) = \int_0^1 \int \phi(y - f_0(x)) \left( \log \frac{\phi(y - f_0(x))}{\phi(y - f(x))} - \frac{1}{2} \| f - f_0 \|^2 \right)^2 dy dx \\
\leq \int_0^1 \int \phi(y - f_0(x)) \left( \log \frac{\phi(y - f_0(x))}{\phi(y - f(x))} \right)^2 dy dx \\
\leq \left( 1 + \frac{1}{2} (\| f \|^2_{\infty} + \| f_0 \|^2_{\infty}) \right) \| f - f_0 \|^2.
\]

As what we have done in the density estimation case, we use the modified block prior \( \Pi(A) = \frac{n(A \cap D)}{n(D)} \) with the constraint set \( D = \{ \sum_{j=1}^{\infty} |\theta_j| < B \} \).

According to Lemma 3.1, the prior \( \Pi \) inherits properties (2.5) and (2.6) from \( \tilde{\Pi} \). Moreover, for \( f \) and \( f_0 \in D \),

\[
V(P_{f_0}, P_f) \leq \left( 1 + 2B^2 \right) \| f - f_0 \|^2.
\]

Next, we deal with the testing procedure. We use likelihood ratio test as in the white noise and Gaussian sequence model cases, and the error is bounded in the following lemma.

**Lemma 3.4.** There exists a constant \( L > 0 \), such that for any \( f_0, f_1 \in D \) satisfying \( \sqrt{n} \| f_1 - f_0 \| > 1 \), there exists a testing function \( \phi_n \) with error probability bounded as

\[
P^{(n)}_{f_0} \phi_n \leq e^{-Ln \| f_0 - f_1 \|^2},
\]

\[
\sup_{\{f \in \text{supp}(\Pi) : \| f - f_0 \|^2 \leq \frac{1}{12} \| f_1 - f_0 \|^2\}} P^{(n)}_f (1 - \phi_n) \leq e^{-Ln \| f_0 - f_1 \|^2}.
\]

The lemma will be proved in later sections. It says \( l^2 \) norm satisfies the testing property (2.9) and (2.10). Using Theorem 2.2, we reach the following conclusion.

**Theorem 3.4.** Let \( \alpha^* > 1/2 \) and \( \gamma \) be the constant defined in (3.2). In the Gaussian regression model with uniform random design, suppose the data is generated by the regression function \( f_0 \in E_{\alpha}(Q) \), with \( \alpha \geq \alpha^* \) and \( 3\gamma Q \leq B \), there exists a constant \( M > 0 \), such that

\[
P^{(n)}_{f_0} \Pi \left( \| f - f_0 \| > M\epsilon_n |X_1, \ldots, X_n, Y_1, \ldots, Y_n\right) \longrightarrow 0.
\]
Remark 3.2. The prior $\Pi$ depends on the value of $B$, which determines the range of adaptation. For any $\alpha^* > 1/2$ and $Q^* > 0$, we can choose $B$ satisfying $B \geq 3\gamma Q^*$ ($\gamma$ depends on $\alpha^*$), such that the prior $\Pi$ is adaptive for all $E_\alpha(Q)$ with $\alpha \geq \alpha^*$ and $Q \leq Q^*$.

3.5. Spectral Density Estimation. Suppose the probability $P^{(n)}_f$ generates stationary Gaussian time series data $X_1,\ldots,X_n$ with mean 0 and spectral density $g = e^f$, with $f(t) = f(-t)$. We assume the spectral density to be a function on $[-\pi,\pi]$. The auto-covariance is $\eta_n = \int_{-\pi}^{\pi} e^{iht}g(t)dt$. Thus, the observation $(X_1,\ldots,X_n)$ follows $P^{(n)}_f = N\left(0,\Gamma_n(g)\right)$, where the covariance matrix is

$$
\begin{pmatrix}
\eta_0 & \eta_1 & \cdots & \eta_{n-1} \\
\eta_1 & \eta_0 & \cdots & \eta_{n-2} \\
\vdots & \vdots & \ddots & \vdots \\
\eta_{n-1} & \eta_{n-2} & \cdots & \eta_0
\end{pmatrix}.
$$

We model the exponent of the spectral density by

$$
f(t) = \sum_{j=0}^{\infty} \theta_j \cos(jt).
$$

According to Parseval’s identity,

$$
2\pi||g||^2 = ||\eta||^2, \quad 2\pi||f||^2 = ||\theta||^2.
$$

We use the modified block prior $\Pi(A) = \frac{\Pi(D \cap A)}{\Pi(D)}$ with the constraint set

$$
D = \left\{ \sum_{j=0}^{\infty} j|\theta_j| < B \right\},
$$

The constraint set (3.4) is stronger than (3.1). Thus, in order that the modified prior $\Pi$ inherits properties (2.5) and (2.6) from the block prior $\Pi$, we need $\alpha > 3/2$. The following lemma will be proved in the supplementary material.

Lemma 3.5. For an arbitrary $\alpha^* > 3/2$, and the constant $\gamma$ defined as

$$
\gamma = \sum_{j=1}^{\infty} j^{2-2\alpha^*}.
$$
For any $f_0 \in E_{\alpha}(Q)$, with $\alpha \geq \alpha^*$ and $3\gamma Q \leq B$, there is a constant $C > 0$, such that

$$\Pi \left\{ \sum_{j=1}^{\infty} (\theta_{0j} - \theta_j)^2 \leq \epsilon_n^2 \right\} \geq \exp \left( - Cn\epsilon_n^2 \right),$$

and

$$\Pi \left( \mathcal{F}_n^c \right) \leq 2 \exp \left( - (C + 4)n\epsilon_n^2 \right).$$

The following lemma, comparing the $l^2$ norm with $D(P_{f_0}^{(n)}, P_{f_1}^{(n)})$ and $V(P_{f_0}^{(n)}, P_{f_1}^{(n)})$, will be proved in the supplementary material.

**Lemma 3.6.** For any $f_0, f_1 \in D$, we have

$$D(P_{f_0}^{(n)}, P_{f_1}^{(n)}) \leq bn\|f_0 - f_1\|^2,$$

$$V(P_{f_0}^{(n)}, P_{f_1}^{(n)}) \leq bn\|f_0 - f_1\|^2,$$

where $b > 1$ is a constant only depending on $\Pi$.

The testing distance satisfying the testing properties (2.9) and (2.10) is the $l^2$-norm.

**Lemma 3.7.** There exists constants $L > 0$ and $0 < \xi < 1/2$, such that for any $f_0, f_1 \in D$ with $\|f_0 - f_1\|^2 \geq \epsilon_n^2$, there exists a testing function $\phi_n$ such that

$$P_{f_0}^{(n)} \phi_n \leq \exp \left( - Ln\|f_0 - f_1\|^2 \right),$$

$$\sup_{\{f \in \text{supp}(\Pi) : \|f - f_1\| \leq \xi \|f_1 - f_0\|\}} P_{f_1}^{(n)}(1 - \phi_n) \leq \exp \left( - Ln\|f_0 - f_1\|^2 \right).$$

The lemma will be proved in later sections. We state the main result of posterior contraction of spectral density estimation as follows.

**Theorem 3.5.** In the spectral density estimation problem, let $(X_1, ..., X_n) \sim P_{f_0}^{(n)}$. For any $f_0 \in E_{\alpha}(Q)$ with $\alpha$ and $Q$ satisfying Lemma 3.5, there is a constant $M > 0$, such that

$$P_{f_0}^{(n)} \Pi \left( \|f - f_0\| > M\epsilon_n |X_1, ..., X_n| \right) \rightarrow 0.$$

**Remark 3.3.** The prior $\Pi$ depends on the value of $B$, which determines the range of adaptation. For any $\alpha^* > 3/2$ and $Q^* > 0$, we can choose $B$ satisfying $B \geq 3\gamma Q^*$ ($\gamma$ depends on $\alpha^*$), such that the prior $\Pi$ is adaptive for all $E_{\alpha}(Q)$ with $\alpha \geq \alpha^*$ and $Q \leq Q^*$. Notice the definition of $\gamma$ in (3.5) is different from that in (3.2).
4. Discussion.

4.1. Exponential Tail of the Posterior. The conclusion of the main posterior contraction result in Theorem 2.2 does not specify a decaying rate of the posterior tail. In fact, by scrutinizing the its proof, it has the following polynomial tail

\[ P_{f_0}^{(n)} \Pi \left( \| \theta - \theta_0 \| > M \epsilon_n | X^n \right) \leq \frac{C'}{n \epsilon_n^2}. \]

However, to obtain a point estimator such as posterior mean with the same rate of convergence as \( \epsilon_n \), faster posterior tail probability is needed (see, for example, Ghosal, Ghosh and van der Vaart (2000) and Shen and Wasserman (2001)). In this section, we show that this polynomial tail can be improved to exponential tail in all the examples we consider in Section 3. The critical step is the following lemma, which improves Lemma 5.6 in the proof of the general result of Theorem 2.2.

**Lemma 4.1.** For all statistical models we consider in Section 3 and the corresponding modified block prior \( \Pi \), let \( C \) be the constant with which \( \Pi \) satisfies (2.5) and (2.6). Define

(4.1) \[ \mathcal{H}_n = \left\{ \int \frac{p_{f_0}^{(n)}}{p_{f_0}^{(n)}} (X^n) d\Pi(f) \geq \exp \left( - (C + b + 1)n \epsilon_n^2 \right) \right\}. \]

Then we have \( P_{f_0}^{(n)} (\mathcal{H}_n^c) \leq \exp \left( - \tilde{C} n \epsilon_n^2 \right) \) for \( f_0 \in E_\alpha(Q) \cap D \) and some \( \tilde{C} > 0 \).

From Lemma 4.1, we have the following improved result for posterior contraction.

**Theorem 4.1.** The conclusions of Theorem 3.1, Theorem 3.2, Theorem 3.3, Theorem 3.4 and Theorem 3.5 can be strengthened as

\[ P_{f_0}^{(n)} \Pi \left( \| \theta - \theta_0 \| > M \epsilon_n | X^n \right) \leq \exp \left( - C' n \epsilon_n^2 \right), \]

under their corresponding settings.

As a consequence, the posterior mean serves as a rate-optimal point estimator.
Corollary 4.1. Under the setting of Theorem 3.1, Theorem 3.2, Theorem 3.3, Theorem 3.4 and Theorem 3.5, we have

\[ P_{f_0}^{(n)} \|E_{\Pi}(\theta|X^n) - \theta_0\|^2 \leq M'\epsilon_n^2, \]

for some constant \(M' > 0\).

The proofs of Lemma 4.1, Theorem 4.1 and Corollary 4.1 are presented in the supplementary material.

4.2. Extension to Besov Balls. Besov balls provide a more flexible collection of functions than Sobolev balls. They are related to wavelet bases. The block prior we propose in this paper naturally takes advantage of the multi-resolution structure of Besov balls. Given a sequence \(\{\theta_j\}\), define \(\theta_k = \{\theta_{2^k+l}\}_{l=0}^{2^k-1}\) for \(k = 0, 1, 2, \ldots\). We can view the signals on each resolution level \(\theta_k\) as a natural block with size \(n_k = 2^k\). The Besov ball is defined as

\[ B_{\alpha p,q}(Q) = \left\{ \theta : \sum_k 2^{skq}||\theta_k||_p^q \leq Q^q \right\}, \]

where \(s = \alpha + \frac{1}{2} - \frac{1}{p}\) and \(||\cdot||_p\) is the vector \(l^p\)-norm. We consider the non-sparse case where the parameters are restricted by

\[(\alpha, p, q, Q) \in (0, \infty) \times [2, \infty] \times [1, \infty] \times (0, \infty).\]

Under such restriction, the block prior is suitable for estimating the signal in \(B_{\alpha p,q}(Q)\). We describe the prior \(\bar{\Pi}\) as follows.

\[ A_k \sim g_k \text{ independently for each } k, \]

\[ \theta_k|A_k \sim N(0, A_k I_{n_k}) \text{ independently for each } k, \]

where \(I_{n_k}\) is the \(2^k \times 2^k\) identity matrix. The mixing densities \(\{g_k\}\) are defined through (2.7) and (2.8) with the constant \(e\) replaced by 2. It is clear that the new mixing densities \(\{g_k\}\) satisfies (2.1), (2.2) and (2.3) with every \(e\) replaced by 2. Define the new sieve

\[ F_n = \left\{ \sum_{k>(2\alpha+1)^{-1}\log_2(n^{\beta-1})} ||\theta_k - \theta_{0k}||^2 \leq \epsilon_n^2 \right\}. \]

We state the property of the block prior \(\bar{\Pi}\) targeting at Besov balls below.
**Theorem 4.2.** For the block prior $\tilde{\Pi}$ defined above, let $\theta_0 \in B_{p,q}^\alpha(Q)$ with $(\alpha,p,q,Q)$ satisfying (4.2), then there exists a constant $C > 0$ such that

\[
\tilde{\Pi} \left\{ \sum_{j=1}^{\infty} (\theta_j - \theta_{0j})^2 \leq \frac{\xi_n^2}{2} \right\} \geq 2^{-Cn\xi_n^2},
\]

and

\[
\tilde{\Pi}(\mathcal{F}_n^c) \leq 2^{1-(C+4)n\xi_n^2},
\]

for sufficiently large $n$ whenever $\beta < \left( \frac{c_3}{2(C+4)} \right)^{2\alpha+1}$, with $c_3$ defined in (2.3) where $e$ is replaced by $2$.

We apply the prior to the Gaussian sequence model. For other models, some slightly extra works are needed.

**Theorem 4.3.** For the Gaussian sequence model (3.3) with any $\theta_0 \in B_{p,q}^\alpha(Q)$, where $(\alpha,p,q,Q)$ satisfies (4.2), then there exists $M > 0$, such that

\[
P_{\theta_0}(\|\theta - \theta_0\| > M\epsilon_n \mid X_1, X_2, \ldots) \rightarrow 0.
\]

Thus, the prior is adaptive for all Besov balls satisfying (4.2).

We prove the results of the extension in the supplementary material.

5. Proof of Main Results.

5.1. Proof of Theorem 2.1. We first outline the proof and list some preparatory lemmas, and then state the proof in details. We introduce the notation $\tilde{\Pi}^A$ to be defined as

\[
\tilde{\Pi}^A = \bigotimes_{k=1}^{\infty} N(0, A_k I_{n_k}).
\]

Given a scale sequence $A = \{A_k\}$, the random function $f = \sum_j \theta_j \phi_j$ is distributed by $\tilde{\Pi}^A$ if for each block $B_k$, $\theta_k = \{\theta_j\}_{j \in B_k} \sim N(0, A_k I_{n_k})$. Then, $\tilde{\Pi}^A$ is a Gaussian process for a given $A$, and the block prior is a mixture of Gaussian process with $A$ distributed by the mixing densities $\{g_k\} \in \mathcal{G}$.
Since $\bar{\Pi}$ itself is not a Gaussian process, the result for the $l^2$ small ball probability asymptotics for Gaussian process cannot be applied directly. Our strategy is to pick a collection $V_\alpha$, and by conditioning, we have

\begin{equation}
\bar{\Pi}\left(\cdot\right) \geq \mathbb{P}(V_\alpha)\mathbb{E}\left(\bar{\Pi}^A\left(\cdot\right) \bigg| A \in V_\alpha\right).
\end{equation}

Then as long as for each $A \in V_\alpha$, there is constants $C_1, C_2 > 0$ independent of $A$, such that

\begin{equation}
\bar{\Pi}^A\left\{\sum_{j=1}^{\infty} (\theta_j - \theta_{0j})^2 \leq \epsilon_n^2\right\} \geq \exp\left(-C_1n\epsilon_n^2\right),
\end{equation}

and

\begin{equation}
\mathbb{P}(V_\alpha) \geq \exp\left(-C_2n\epsilon_n^2\right),
\end{equation}

then the property (2.5) is a direct consequence with $C = C_1 + C_2$. Thus, picking such $V_\alpha$ is important. Generally speaking, for each $A \in V_\alpha$, we need $\bar{\Pi}^A$ to behave just like a Gaussian prior designed for estimating $f_0 \in E_\alpha(Q)$ when $\alpha$ is known.

The distribution $\bar{\Pi}^A$ may be hard to deal with. Our strategy is to use the following simple comparison result so that we can study a simpler distribution instead. The lemma will be proved in the supplementary material.

**Lemma 5.1.** For standard i.i.d. Gaussian sequence $\{Z_j\}$ and sequences $\{a_j\}, \{b_j\}$ and $\{c_j\}$, suppose there is a constant $R > 0$ such that

\[ R^{-1}a_j \leq b_j \leq Ra_j, \quad \text{for all } j, \]

then we have

\[ \mathbb{P}\left(\sum_j b_j(Z_j-c_j)^2 \leq R^{-1}\epsilon^2\right) \leq \mathbb{P}\left(\sum_j a_j(Z_j-c_j)^2 \leq \epsilon^2\right) \leq \mathbb{P}\left(\sum_j b_j(Z_j-c_j)^2 \leq R\epsilon^2\right). \]

Define $J_\alpha$ to be the smallest integer such that $J_\alpha \geq (8Q^2)^{\frac{1}{2}}n^{\frac{1}{2\alpha+1}}$. Let $K$ to be the smallest integer such that $e^K > J_\alpha$, and define $J = [e^K]$. Inspired by the comparison lemma, we define

\begin{equation}
V_\alpha = V_{\alpha,R} = \left\{A : R^{-1} \leq \min_{1 \leq k \leq K} \frac{A_k}{A_{\alpha,k}} \leq \max_{1 \leq k \leq K} \frac{A_k}{A_{\alpha,k}} \leq R\right\},
\end{equation}
with
\[ A_{\alpha,k} = \frac{l_k^{-2\alpha} - l_{k+1}^{-2\alpha}}{2\alpha(l_{k+1} - l_k)}, \quad \text{for } k = 1, 2, \ldots, K. \]

Define the truncated Gaussian process,
\[ \Pi_{\alpha} = \bigotimes_{k=1}^{K} N(0, A_{\alpha,k} I_{n_k}). \]

A random function \( f = \sum_j \theta_j \phi_j \) is distributed by \( \Pi_{\alpha} \) if \( \theta_k \sim N(0, A_{\alpha,k} I_{n_k}) \) for each \( k = 1, \ldots, K \) and \( \theta_k = 0 \) for \( k > K \). The comparison lemma implies that we can control \( \Pi_{\alpha} \) for each \( A \in V_{\alpha} \) by the truncated Gaussian process \( \Pi_{\alpha} \). Additionally, the small ball probability of \( \Pi_{\alpha} \) can be established. The argument is separated in the following lemmas, which will be proved in later sections.

**Lemma 5.2.** For any \( \alpha > 0 \), and \( f_0 \in E_{\alpha}(Q) \), there exists \( C_3 > 0 \), such that
\[ \Pi_{\alpha} \left\{ \sum_{j=1}^{\infty} (\theta_j - \theta_{0j})^2 \leq \epsilon_n^2 \right\} \geq \exp \left( -C_3 n \epsilon_n^2 \right). \]

**Lemma 5.3.** For each \( k \), let \( A_k \sim g_k \), with \( \{g_k\} \in \mathcal{G} \). we have
\[ \mathbb{P}(V_{\alpha}) \geq \exp \left( -C_2 n \epsilon_n^2 \right). \]

**Lemma 5.4.** For \( J \) defined above, and \( f_0 \in E_{\alpha}(Q) \), we have
\[ \Pi \left\{ \sum_{j>J} (\theta_j - \theta_{0j})^2 \leq \epsilon_n^2 \right\} \geq \frac{1}{2}, \]

for sufficiently large \( n \).

**Proof of (2.5) in Theorem 2.1.** We first introduce the truncated version of \( \Pi_{\alpha} \) to be
\[ \Pi_{\alpha} = \bigotimes_{k=1}^{K} N(0, A_{k} I_{n_k}). \]
By Lemma 5.4, we have
\[ \tilde{\Pi} \left\{ \sum_{j=1}^{\infty} (\theta_j - \theta_{0j})^2 \leq \frac{\epsilon_n^2}{2} \right\} \geq \tilde{\Pi} \left\{ \sum_{j=1}^{J} (\theta_j - \theta_{0j})^2 \leq \frac{\epsilon_n^2}{2} \right\} \]
\[ = \tilde{\Pi} \left\{ \sum_{j=1}^{J} (\theta_j - \theta_{0j})^2 \leq \frac{\epsilon_n^2}{2} \right\} \tilde{\Pi} \left\{ \sum_{j>J} (\theta_j - \theta_{0j})^2 \leq \frac{\epsilon_n^2}{2} \right\} \]
\[ \geq \frac{1}{2} \tilde{\Pi} \left\{ \sum_{j=1}^{J} (\theta_j - \theta_{0j})^2 \leq \frac{\epsilon_n^2}{2} \right\}, \]
where we have used independence between different blocks in the above equality. In the spirit of (5.2), we have
\[ (5.7) \]
\[ \tilde{\Pi} \left\{ \sum_{j=1}^{J} (\theta_j - \theta_{0j})^2 \leq \frac{\epsilon_n^2}{2} \right\} \geq \mathbb{P}(V_\alpha) \mathbb{E} \left( \Pi_K^A \left\{ \sum_{j=1}^{\infty} (\theta_j - \theta_{0j})^2 \leq \frac{\epsilon_n^2}{2} \right\} | A \in V_\alpha \right). \]
By Lemma 5.1, for each $A \in V_\alpha$,
\[ \Pi_K^A \left\{ \sum_{j=1}^{\infty} (\theta_j - \theta_{0j})^2 \leq \frac{\epsilon_n^2}{2} \right\} \geq \tilde{\Pi}^{A_\alpha} \left\{ \sum_{j=1}^{\infty} (\theta_j - \theta_{0j})^2 \leq \frac{\epsilon_n^2}{2R} \right\}. \]
By Lemma 5.2, we have
\[ \tilde{\Pi}^{A_\alpha} \left\{ \sum_{j=1}^{\infty} (\theta_j - \theta_{0j})^2 \leq \frac{\epsilon_n^2}{2R} \right\} \geq \exp \left( - C' n \epsilon_n^2 \right). \]
Combining what we have derived and Lemma 5.3, (2.5) is proved. \[ \blacksquare \]

**Proof of (2.6) in Theorem 2.1.** We fix the constant $C$ in (2.5), and we are going to prove (2.6) with the same $C$. Remember the sieve $\mathcal{F}_n$ is defined by (2.4). Define the set
\[ \mathcal{A}_n = \left\{ A_k \leq e^{-k^2} \text{ for all } k > \frac{1}{2\alpha + 1} \log(n\beta^{-1}) \right\}. \]
Then,
\[ \tilde{\Pi}(\mathcal{F}_n^c) \leq \sup_{A \in \mathcal{A}_n} \tilde{\Pi}^A(\mathcal{F}_n^c) + \mathbb{P}(\mathcal{A}_n^c). \]
Condition (2.3) implies
\[
\mathbb{P}(A^c_n) \leq \sum_{k>(2\alpha+1)^{-1}\log(n\beta^{-1})} \mathbb{P}(A_k > e^{-k^2})
\]
\[
\leq \sum_{k>(2\alpha+1)^{-1}\log(n\beta^{-1})} \exp \left( -c_3 e^k \right)
\]
\[
\leq \exp \left( -\frac{1}{2} c_3 n^{2\alpha+1} \beta^{-\frac{1}{2\alpha+1}} \right)
\]
\[
\leq \exp \left( - (C + 4) n\epsilon^2 \right).
\]
The last inequality is because \( \beta \leq \left( \frac{c_3}{2(C+4)} \right)^{2\alpha+1} \). We bound \( \bar{\Pi}^A(F_n^c) \) for each \( A \in \mathcal{A}_n \). By Anderson’s lemma,
\[
\bar{\Pi}^A(F_n^c) = \bar{\Pi}^A \left\{ \sum_{j>(n\beta^{-1})^{2\alpha+1}} (\theta_j - \theta_0)^2 > \epsilon_n^2 \right\}
\]
\[
\leq \bar{\Pi}^A \left\{ \sum_{j>(n\beta^{-1})^{2\alpha+1}} \theta_j^2 \geq \epsilon_n^2 \right\}
\]
\[
\leq \bar{\Pi}^A \left\{ \sum_{k>(2\alpha+1)^{-1}\log(n\beta^{-1})} ||\theta_k||^2 \geq \epsilon_n^2 \right\}
\]
\[
\leq \sum_{k>(2\alpha+1)^{-1}\log(n\beta^{-1})} \bar{\Pi}^A \left\{ \sum_{k>(2\alpha+1)^{-1}\log(n\beta^{-1})} ||\theta_k||^2 \geq a_k \epsilon_n^2 \right\},
\]
where \( \sum_k a_k \leq 1 \). We choose \( a_k = ak^{-2} \). Define \( \chi_d^2 \) to be the chi-square random variable with degree of freedom \( d \).
\[
\sum_{k>(2\alpha+1)^{-1}\log(n\beta^{-1})} \bar{\Pi}^A \left\{ ||\theta_k||^2 \geq a_k \epsilon_n^2 \right\}
\]
\[
= \sum_{k>(2\alpha+1)^{-1}\log(n\beta^{-1})} \mathbb{P} \left\{ a_k^{-1} A_k \chi_{n_k}^2 \geq \epsilon_n^2 \right\}
\]
\[
= \sum_{k>(2\alpha+1)^{-1}\log(n\beta^{-1})} \mathbb{P} \left\{ \epsilon_n^{-2} C' e^k a_k^{-1} A_k \chi_{n_k}^2 \geq C' e^k \right\}
\]
\[
\leq \sum_{k>(2\alpha+1)^{-1}\log(n\beta^{-1})} \exp \left( - C' e^k \right) \left( 1 - 2\epsilon_n^{-2} C' e^k a_k^{-1} A_k \right)^{-\frac{n_k}{2}},
\]
where we can choose $C'$ sufficiently large. On the set $A_k$, for $n$ sufficiently large,

\[ A_k \leq e^{-k^2} \leq \frac{1}{4C'} a_k e^{-k^2} \epsilon_n^2, \quad \text{for all } k > \frac{1}{2\alpha + 1} \log(n\beta^{-1}). \]

Therefore,

\[
\sum_{k > (2\alpha + 1)^{-1} \log(n\beta^{-1})} \exp \left( -C' e^k \right) \left( 1 - 2\epsilon_n^2 C' e^k a_k^{-1} A_k \right)^{-\frac{n}{2}kA_k} \\
\leq \sum_{k > (2\alpha + 1)^{-1} \log(n\beta^{-1})} \exp \left( -C' e^k \right) (\sqrt{2})^{nk} \\
\leq \sum_{k > (2\alpha + 1)^{-1} \log(n\beta^{-1})} \exp \left( - (C' - \frac{1}{2} \log 2) e^k \right) \\
\leq \exp \left( - \frac{1}{2} \left( C' - \frac{1}{2} \log 2 \right) \beta^{-\frac{1}{2\alpha + 1} \log 2} \right) \\
\leq \exp \left( - (C + 4) n\epsilon^2_n \right),
\]

with sufficiently large $C'$ and $n$. Hence,

\[
\sup_{A \in A_n} \Pi^A(F_n^c) \leq \exp \left( - (C + 4) n\epsilon^2_n \right),
\]

and we have

\[
\Pi(F_n^c) \leq 2 \exp \left( - (C + 4) n\epsilon^2_n \right).
\]

Thus the proof is complete. $\blacksquare$

5.2. Proof of Theorem 2.2. Before stating the proof of Theorem 2.2, we need to establish a testing result. It will be proved in later sections.

**Lemma 5.5.** Let $d$ be a distance satisfying the testing property (2.9) and (2.10). Suppose that there is $b > 0$ such that for all $f_1, f_2 \in D$,

\[
b^{-1}d(f_1, f_2) \leq ||f_1 - f_2|| \leq bd(f_1, f_2).
\]

Then for any sufficiently large $M > 0$, there exists a testing function $\phi_n$, such that

\[
P_{f_0}^{(n)}(\phi_n) \leq 2 \exp \left( - \frac{1}{2} L M^2 n\epsilon^2_n \right),
\]

\[
\sup_{\{f \in F_n \cap \supp(\Pi) : d(f, f_0) > M\epsilon_n\}} P_f^{(n)}(1 - \phi_n) \leq \exp \left( - L^2 n\epsilon^2_n \right).
\]
The following result is Lemma 10 in Ghosal and van der Vaart (2007). It lower bounds the denominator of the posterior distribution in probability.

**Lemma 5.6.** Consider $\mathcal{H}_n$ defined in (4.1), as long as

\[
\Pi \left\{ D(P_f^{(n)}, P_f^{(n)}) \leq b n \epsilon_n^2, V(P_f^{(n)}, P_f^{(n)}) \leq b n \epsilon_n^2 \right\} \geq \exp \left( -C n \epsilon_n^2 \right),
\]

we have $P_f^{(n)}(\mathcal{H}_n^c) \leq \frac{1}{C^2 n \epsilon_n^2}$ for some $\bar{C} > 0$.

**Proof of Theorem 2.2.** Notice the prior $\Pi$ inherits the properties (2.5) and (2.6) from $\bar{\Pi}$. Since both $D(P_f^{(n)}, P_f^{(n)})$ and $V(P_f^{(n)}, P_f^{(n)})$ are upper bounded by $b n ||\theta_0 - \theta||^2$, we have

\[
\Pi \left\{ D(P_f^{(n)}, P_f^{(n)}) \leq b n \epsilon_n^2, V(P_f^{(n)}, P_f^{(n)}) \leq b n \epsilon_n^2 \right\}
\geq \Pi \left\{ \sum_{j=1}^{\infty} (\theta_j - \theta_0)^2 \leq \epsilon_n^2 \right\}
\geq \exp \left( -C n \epsilon_n^2 \right),
\]

for the constant $C$ with which $\Pi$ satisfies (2.5) and (2.6). By Lemma 5.6, the K-L property of prior implies $P_f^{(n)}(\mathcal{H}_n^c) \leq \frac{1}{C^2 n \epsilon_n^2}$. Let $\mathcal{F}_n$ be the sieve defined in (2.4) and we have

\[
\Pi(\mathcal{F}_n^c) \leq 2 \exp \left( -(C + 4) n \epsilon_n^2 \right).
\]

Letting $\phi_n$ be the testing function in Lemma 5.5, we have

\[
P_f^{(n)}(d(f, f_0) > M \epsilon_n | X^n)
\leq P_f^{(n)}(\mathcal{H}_n^c) + P_f^{(n)} \phi_n + P_f^{(n)} \Pi(\mathcal{F}_n) \exp \left( -(C + 4) n \epsilon_n^2 \right)(1 - \phi_n)^1_{\mathcal{H}_n},
\]
where the first two terms go to 0. The last term has bound
\[
P_f^{(n)} \Pi(d(f, f_0) > M\epsilon_n|X^n)(1 - \phi_n)1_{\mathcal{H}_n}
\]
\[
\leq \exp \left((C + 2)n\epsilon_n^2\right) P_f^{(n)} \int_{\mathcal{F}_n} \frac{P_f^{(n)}}{P_{f_0}^{(n)}} (X^n)(1 - \phi_n)(X^n) d\Pi(f)
\]
\[
+ \exp \left((C + 2)n\epsilon_n^2\right) P_f^{(n)} \int_{\mathcal{F}_n} \frac{P_f^{(n)}}{P_{f_0}^{(n)}} (X^n) d\Pi(f)
\]
\[
\leq \exp \left((C + 2)n\epsilon_n^2\right) \sup_{\{f \in \mathcal{F}_n : d(f, f_0) > M\epsilon_n\}} P_f^{(n)} (1 - \phi_n)
\]
\[
+ \exp \left((C + 2)n\epsilon_n^2\right) \sup_{\{f \in \mathcal{F}_n \cap \text{supp}(\Pi) : d(f, f_0) > M\epsilon_n\}} P_f^{(n)} (1 - \phi_n)
\]
\[
\leq \exp \left(-(LM^2 - C - 2)n\epsilon_n^2\right) + 2 \exp \left(-2n\epsilon_n^2\right).
\]

We pick \(M\) satisfying \(M > \sqrt{L^{-1}(C + 2)}\), and then every term goes to 0. The proof is complete.

### 6. Proofs of Auxiliary Results.

#### 6.1. Proofs of Some Technical Lemmas

We present the proofs of Lemma 5.2, Lemma 5.3 and Lemma 5.4 below.

**Proof of Lemma 5.2.** For the Gaussian measure \(\Pi_{A\alpha}^{K}\), let \(H_{A\alpha}\) be its reproducing kernel Hilbert space (RKHS) with norm \(\| \cdot \|_{H_{A\alpha}}\) defined by
\[
\|\theta\|_{H_{A\alpha}}^2 = \sum_{k=1}^{K} A^{-1}_{\alpha,k} \|\theta_k\|^2.
\]

Define the quantity
\[
\phi_{\theta_0}^{A\alpha}(\epsilon) = \inf_{\theta \in H_{A\alpha} : \|\theta - \theta_0\| < \epsilon} \|\theta\|_{H_{A\alpha}}^2 - \log \Pi_{A\alpha}^{K} \left\{ \sum_{k=1}^{K} \|\theta_k\|^2 \leq \epsilon^2 \right\}.
\]

According to Lemma 5.3 in van der Vaart and van Zanten (2008),
\[
(6.1) \quad \phi_{\theta_0}^{A\alpha}(\epsilon_n) \leq - \log \Pi_{A\alpha}^{K} \left\{ \sum_{k=1}^{K} \|\theta_k - \theta_{0k}\|^2 \leq \epsilon_n^2 \right\} \leq \phi_{\theta_0}^{A\alpha}(\epsilon_n/2).
\]
Define \( b_j = j^{-(2\alpha+1)} \) and \( a_j = \sum_k j \in B_k \} A_{\alpha,k} \). Let \( Z_j \) be i.i.d. Gaussian sequence. We first use Lemma 5.1 to compare \( \mathbb{P}\left( \sum_j a_j Z_j^2 \leq \epsilon_n^2 \right) \) with \( \mathbb{P}\left( \sum_j b_j Z_j^2 \leq \epsilon_n^2 \right) \). According to the definition of \( A_{\alpha,k} \), we have

\[
e^{-(2\alpha+1)(k+1)} \leq A_{\alpha,k} \leq e^{-2\alpha-1} \]

for each \( k \).

Therefore, for each \( j \in B_k \),

\[
a_j b_j \leq e^{(2\alpha+1)(k+1)} e^{-2\alpha-1} \leq e^{2\alpha+1},
\]

and

\[
b_j a_j \leq e^{-(2\alpha+1)(k+1)} \leq e^{2\alpha+1}.
\]

The bound does not depend on \( k \). Thus, by Lemma 5.1 we have

\[
\bar{\Pi}_K^{A_\alpha} \left\{ \sum_{k=1}^K \theta_k^2 \leq \epsilon_n^2 \right\} \geq \mathbb{P}\left( \sum_j a_j Z_j^2 \leq \epsilon_n^2 \right) \geq \mathbb{P}\left( \sum_j b_j Z_j^2 \leq e^{-2\alpha-1} \epsilon_n^2 \right).
\]

By Zolotarev (1986), there exists \( C > 0 \) such that

\[
\mathbb{P}\left( \sum_j b_j Z_j^2 \leq e^{-2\alpha-1} \epsilon_n^2 \right) \geq \exp \left( -Cn\epsilon_n^2 \right).
\]

Thus, we have

\[
(6.2) \quad \bar{\Pi}_K^{A_\alpha} \left\{ \sum_{k=1}^K \theta_k^2 \leq \epsilon_n^2 \right\} \geq \exp \left( -Cn\epsilon_n^2 \right).
\]

Then, we calculate the RKHS approximation of \( \theta_0 \).

\[
\inf_{\theta \in \mathcal{H}_{A_\alpha}} \|\theta\|^2_{\mathcal{H}_{A_\alpha}} \leq \|\theta_0\|^2_{\mathcal{H}_{A_\alpha}} = \sum_{k=1}^K A_{\alpha,k}^{-1} \|\theta_{0 k}\|^2 \leq \sum_{j=1}^J a_j^{-1} \theta_{0j}^2 \\
\leq e^{2\alpha+1} \sum_{j=1}^J b_j^{-1} \theta_{0j}^2 = e^{2\alpha+1} \sum_{j=1}^J \theta_{0j}^2 \leq e^{2\alpha+1} J \sum_{j=1}^J \theta_{0j}^2 \\
\leq Q^2 e^{2\alpha+1} J \leq Cn\epsilon_n^2.
\]

Combining (6.2), we have \( \phi_{\theta_0}^{A_\alpha}(\epsilon_n) \leq Cn\epsilon_n^2 \). By (6.1), we reach the desired conclusion.
Proof of Lemma 5.3. We need (2.1) to lower bound $\mathbb{P}(V_{\alpha,R})$,

$$
\mathbb{P}(V_{\alpha,R}) = \prod_{k=1}^{K} \mathbb{P}(R^{-1}A_{\alpha,k} \leq A_k \leq RA_{\alpha,k}) = \prod_{k=1}^{K} \int_{R^{-1}A_{\alpha,k}}^{RA_{\alpha,k}} g_k(t) dt
$$

$$
\geq \prod_{k=1}^{K} e^{-c_1 k} \left( e^{-k} - R^{-1}A_{\alpha,k} \right) \geq \prod_{k=1}^{K} \exp \left( -c' e^k \right)
$$

$$
= \exp \left( -c' \sum_{k=1}^{K} e^k \right) \geq \exp \left( -C' e^{K} \right) \geq \exp \left( -C'' n \epsilon^2 \right).
$$

Proof of Lemma 5.4. We have

$$
\bar{\Pi} \left\{ \sum_{j>J} (\theta_j - \theta_{0j})^2 \leq \frac{\epsilon_n^2}{2} \right\} = \bar{\Pi} \left\{ \sum_{k>K} ||\theta_k - \theta_{0k}||^2 \leq \frac{\epsilon_n^2}{2} \right\}
$$

$$
\geq \bar{\Pi} \left\{ \sum_{k>K} ||\theta_k||^2 \leq \frac{\epsilon_n^2}{8} \sum_{k>K} ||\theta_{0k}||^2 \leq \frac{\epsilon_n^2}{8} \right\} = \bar{\Pi} \left\{ \sum_{k>K} ||\theta_k||^2 \leq \frac{\epsilon_n^2}{8} \right\}
$$

$$
\geq 1 - \frac{8}{\epsilon_n^2} \sum_{k>K} n_k \mathbb{E} ||\theta_k||^2 = 1 - \frac{8}{\epsilon_n^2} \sum_{k>K} n_k \mathbb{E} A_k.
$$

The second equality above is because

$$
\sum_{k>K} ||\theta_{0k}||^2 = \sum_{j>J} \theta_{0j}^2 \leq J^{-2\alpha} \sum_{j>J} j^{2\alpha} \theta_{0j}^2 \leq \frac{1}{8} \frac{\epsilon_n^2}{n}. \tag{2.2}
$$

The last inequality is Markov inequality. By (2.2),

$$
\sum_{k>K} n_k \mathbb{E} A_k \leq \sum_{k>K} e^{k+1} e^{-c_2 k^2} \leq e^{-CK^2} = O(n^{-1}),
$$

for sufficiently large $n$, and therefore,

$$
\bar{\Pi} \left\{ \sum_{j>J} (\theta_j - \theta_{0j})^2 \leq \frac{\epsilon_n^2}{2} \right\} \to 1.
$$

6.2. Proofs of Some Testing Results.
6.2.1. Proof of Lemma 5.5. We divide the alternative set into rings

\[
\{ f \in \mathcal{F}_n \cap \text{supp}(\Pi) : d(f, f_0) > M\epsilon_n \} \subset \bigcup_{l>M} \{ f \in \mathcal{F}_n \cap \text{supp}(\Pi) : l\epsilon_n < d(f, f_0) \leq (l+1)\epsilon_n \}.
\]

For each ring indexed by \( l \), we cover it with balls of radius \( \xi l\epsilon_n \). Denote \( N(\delta, \mathcal{H}, \rho) \) to be the covering number of \( \mathcal{H} \) with \( \delta \)-balls under distance \( \rho \).
The following proposition bounds the covering number of each ring. It will be proved in the supplementary material.

**Proposition 6.1.** For each integer \( l > M \) with sufficiently large \( M \), we have

\[
\log N(\xi l\epsilon_n, \{ f \in \mathcal{F}_n \cap \text{supp}(\Pi) : l\epsilon_n < d(f, f_0) \leq (l+1)\epsilon_n \}, d) \leq C_{(\beta, b, \xi)} n^2\epsilon_n^2,
\]

with some constant \( C_{(\beta, b, \xi)} > 0 \).

By the conclusion of Proposition 6.1, for each \( l > M \), there exists \( \{ \phi_{li} \}_{i=1}^{N_l} \subset \{ f \in \mathcal{F}_n \cap \text{supp}(\Pi) : l\epsilon_n < d(f, f_0) \leq (l+1)\epsilon_n \} \), such that

\[
\{ f \in \mathcal{F}_n \cap \text{supp}(\Pi) : l\epsilon_n < d(f, f_0) \leq (l+1)\epsilon_n \} \subset \bigcup_{i=1}^{N_l} \{ f \in \mathcal{F}_n \cap \text{supp}(\Pi) : d(f, f_{li}) \leq \xi \epsilon_n \},
\]

with \( N_l \) bounded by \( C_{(\beta, b, \xi)} n^2\epsilon_n^2 \). Since for each \( f_{li} \), \( d(f_{li}, f_0) > l\epsilon_n \), we have

\[
\{ f \in \mathcal{F}_n \cap \text{supp}(\Pi) : d(f, f_{li}) \leq \xi \epsilon_n \} \subset \{ f \in \mathcal{F}_n \cap \text{supp}(\Pi) : d(f, f_{li}) \leq \xi d(f_0, f_{li}) \}.
\]

The final decomposition of the alternative set is

\[
\{ f \in \mathcal{F}_n \cap \text{supp}(\Pi) : d(f, f_0) > M\epsilon_n \} \subset \bigcup_{l>M} \bigcup_{i=1}^{N_l} \{ f \in \mathcal{F}_n \cap \text{supp}(\Pi) : d(f, f_{li}) \leq \xi d(f_0, f_{li}) \}.
\]

According to the testing property (2.9) and (2.10), there exists \( \phi_{li} \) such that

\[
P_{f_0}^{(n)}(\phi_{li}) \leq \exp \left( -Ll^2 n\epsilon_n^2 \right),
\]

\[
\sup_{\{ f \in \text{supp}(\Pi) : d(f, f_{li}) \leq \xi d(f_0, f_{li}) \}} P_{f}^{(n)}(1 - \phi_{li}) \leq \exp \left( -Ll^2 n\epsilon_n^2 \right).
\]
Define $\phi = \max_{l>M} \max_{1 \leq i \leq N_l} \phi_{li}$, and its error bound is

$$P_{f_0}^{(n)}(\phi) \leq \sum_{l>M} \sum_{i=1}^{N_l} P_{f_0}^{(n)} \phi_{li}$$

$$\leq \sum_{l>M} \sum_{i=1}^{N_l} \exp \left( -Ll^2n\epsilon_n^2 \right)$$

$$\leq \exp \left( C_{(\beta,b,\xi)} n\epsilon_n^2 \right) \sum_{l>M} \exp \left( -Ll^2n\epsilon_n^2 \right)$$

$$\leq 2 \exp \left( -LM^2n\epsilon_n^2 + C_{(\beta,b,\xi)} n\epsilon_n^2 \right)$$

$$\leq 2 \exp \left( -\frac{1}{2} LM^2n\epsilon_n^2 \right),$$

for sufficiently large $M$. We also have

$$\sup_{\{f \in \text{supp}(\Pi) \cap F_n: d(f,f_0)>M\epsilon_n\}} P_{f}^{(n)}(1-\phi) \leq \exp \left( -LM^2n\epsilon_n^2 \right).$$

Thus, the proof is complete.

6.2.2. Proof of Lemma 3.4. We first state the Bernstein’s inequality in van der Vaart (1998, Page 285), which will be used in the proof.

**Lemma 6.1.** Let $X_1, ..., X_n$ be i.i.d. observations. For any bounded, measurable function $f$, the Bernstein’s inequality holds:

$$P \left( n^{-1/2} \left| \sum_{i=1}^{n} \left( f(X_i) - \mathbb{E}f(X_i) \right) \right| > x \right) \leq 2 \exp \left( -\frac{1}{4} \mathbb{E}f^2(X_1) + \frac{x^2}{\|f\|_\infty^2 / \sqrt{n}} \right),$$

for each $x > 0$.

**Proof of Lemma 3.4.** For design points $X_1, ..., X_n$ i.i.d. from $U[0,1]$, we define $P_n$ to be the associated empirical distribution. Our analysis first condition on the design points. Define the testing function to be

$$\phi_n = \left\{ \sum_{i=1}^{n} Y_i(f_1(X_i) - f_0(X_i)) \geq \frac{1}{2} \sum_{i=1}^{n} f_1^2(X_i) - \frac{1}{2} \sum_{i=1}^{n} f_0^2(X_i) \right\},$$

with the testing statistic to be

$$T_n = \sum_{i=1}^{n} Y_i(f_1(X_i) - f_0(X_i)) - \frac{1}{2} \sum_{i=1}^{n} f_1^2(X_i) + \frac{1}{2} \sum_{i=1}^{n} f_0^2(X_i).$$
Lemma 6.1, for any $f \in \{X_1, \ldots, X_n\}$, we need $f$ which implies the unconditional bound

$$P_{f_0}^{(n)}(\phi_n|X_1, \ldots, X_n) \leq 1 - \Phi\left(\sqrt{n}(\mathbb{P}_n(f_1 - f_0)^2)^{1/2}/2\right),$$

which we need $f$ which implies the unconditional bound

$$P_{f_0}^{(n)}(\phi_n|X_1, \ldots, X_n) \leq 1 - \Phi\left(\sqrt{n}(\mathbb{P}_n(f_1 - f_0)^2)^{1/2}/4\right).$$

We can also find the distribution of $T_n$ under $P_{f}^{(n)}(\cdot|X_1, \ldots, X_n)$. As long as $f$ satisfies $\mathbb{P}_n(f - f_0)^2 \leq \frac{1}{16}\mathbb{P}_n(f_1 - f_0)^2$, the mean is bounded below by $\frac{1}{4}\mathbb{P}_n(f_0 - f_1)^2$. Therefore,

$$\sup_{\{f : \mathbb{P}_n(f - f_0)^2 \leq \frac{1}{4}\mathbb{P}_n(f_1 - f_0)^2\}} P_{f}^{(n)}(1 - \phi_n|X_1, \ldots, X_n) \leq 1 - \Phi\left(\sqrt{n}(\mathbb{P}_n(f_1 - f_0)^2)^{1/2}/4\right).$$

This error probability is conditioning on the design points. To derive a bound without such conditioning, we need $f$ satisfy $||f - f_0||^2 \leq \frac{1}{32}||f_1 - f_0||^2$. By Lemma 6.1, for any $f \in \{f : ||f - f_0||^2 \leq \frac{1}{32}||f_1 - f_0||^2\}$, we have

$$\mathbb{P}\left(\mathbb{P}_n(f - f_0)^2 > \frac{1}{16}\mathbb{P}_n(f_1 - f_0)^2\right) \leq \exp\left\{-\frac{n\left(\frac{1}{16}||f_1 - f_0||^2 - ||f - f_0||^2\right)^2}{4\mathbb{P}\left((f - f_0)^2 - \frac{1}{16}(f_1 - f_0)^2\right)^2 + 5B^2\left(\frac{1}{16}||f_1 - f_0||^2 - ||f - f_0||^2\right)}\right\}$$

$$\leq \exp\left\{-\frac{n\left(\frac{1}{16}||f_1 - f_0||^2 - ||f - f_0||^2\right)}{4\mathbb{P}\left((f - f_0)^2 - \frac{1}{16}(f_1 - f_0)^2\right)^2 + 5B^2\left(\frac{1}{16}||f_1 - f_0||^2 - ||f - f_0||^2\right)}\right\}$$

$$\leq \exp\left\{-\frac{n\left(\frac{1}{16}||f_1 - f_0||^2 - ||f - f_0||^2\right)}{20B^2\left(||f - f_0||^2 + \frac{1}{16}||f_1 - f_0||^2\right) + 5B^2\left(\frac{1}{16}||f_1 - f_0||^2 - ||f - f_0||^2\right)}\right\}$$

$$\leq \exp\left\{-\frac{n||f_1 - f_0||^2}{2080B^2}\right\}.$$
Therefore,

\[
\sup_{\{f: \|f-f_0\|^2 \leq \frac{1}{16} \|f_1-f_0\|^2\}} \mathbb{P}\left( \mathbb{P}_n(f-f_0)^2 > \frac{1}{16} \mathbb{P}_n(f_1-f_0)^2 \right) \leq \exp\left\{ - \frac{n \|f_1-f_0\|^2}{2080B^2} \right\} .
\]

Using this result, we can bound the unconditional error probability by

\[
\sup_{\{f: \|f-f_0\|^2 \leq \frac{1}{16} \|f_1-f_0\|^2\}} P_f^{(n)}(1 - \phi_n) 
\leq \sup_{\{f: \|f-f_0\|^2 \leq \frac{1}{16} \|f_1-f_0\|^2\}} \mathbb{P}\left( \left\{ \mathbb{P}_n(f-f_0)^2 \leq \frac{1}{16} \mathbb{P}_n(f_1-f_0)^2 \right\} P_f^{(n)}(1 - \phi_n|X_1,\ldots,X_n) \right) 
+ \sup_{\{f: \|f-f_0\|^2 \leq \frac{1}{16} \|f_1-f_0\|^2\}} \mathbb{P}\left( \mathbb{P}_n(f-f_0)^2 > \frac{1}{16} \mathbb{P}_n(f_1-f_0)^2 \right) 
\leq \mathbb{P}\left( 1 - \Phi\left( \sqrt{n}\left( \mathbb{P}_n(f_1-f_0)^2 \right)^{1/2} / 4 \right) \right) + \exp\left\{ - \frac{n \|f_1-f_0\|^2}{2080B^2} \right\} .
\]

Now, to bound both error probability, it is sufficient to bound

\[
\mathbb{P}\left( 1 - \Phi\left( \sqrt{n}\left( \mathbb{P}_n(f_1-f_0)^2 \right)^{1/2} / 4 \right) \right) .
\]

Using Bernstein’s inequality again, we have

\[
\mathbb{P}\left( 1 - \Phi\left( \sqrt{n}\left( \mathbb{P}_n(f_1-f_0)^2 \right)^{1/2} / 4 \right) \right) 
\leq \mathbb{P}\left\{ \left\{ \mathbb{P}_n(f_1-f_0)^2 \geq \frac{1}{4} \|f_1-f_0\|^2 \right\} \left( 1 - \Phi\left( \sqrt{n}\left( \mathbb{P}_n(f_1-f_0)^2 \right)^{1/2} / 4 \right) \right) \right\} 
+ \mathbb{P}\left( \mathbb{P}_n(f_1-f_0)^2 < \frac{1}{4} \|f_1-f_0\|^2 \right) 
\leq 1 - \Phi\left( \sqrt{n}\|f_1-f_0\| / 8 \right) + \exp\left\{ - \frac{1}{4} \mathbb{P}(f_1-f_0)^4 + 2B^2\|f_1-f_0\|^2 \right\} 
\leq 1 - \Phi\left( \sqrt{n}\|f_1-f_0\| / 8 \right) + \exp\left\{ - \frac{1}{4} \frac{n \|f_1-f_0\|^4}{B^2\|f_1-f_0\|^2 + 2B^2\|f_1-f_0\|^2} \right\} 
\leq \exp\left( - \frac{n \|f_1-f_0\|^2}{128} \right) + \exp\left( - \frac{n \|f_1-f_0\|^2}{24B^2} \right),
\]

where the treatment of the Gaussian tail is the same as what we did for the white noise model. Thus, the proof is complete. \[\square\]
6.2.3. **Proof of Lemma 3.7.** Remember for the spectral density estimation, we have observation \((X_1, \ldots, X_n) \sim P^{(n)}_f = N\left(0, \Gamma_n(g)\right)\), where \(g = e^f\).

Define \(\|\Gamma_n(g)\|_F^2 = \text{tr}(\Gamma_n(g)\Gamma_n(g)^T)\) to be the matrix Frobenius norm. We first present a testing result under Frobenius norm. The following lemma is a special version of Lemma 5.9 in Gao and Zhou (2013).

**Lemma 6.2.** Let \(\mathcal{M}\) be the covariance matrix class

\[
\mathcal{M} = \{\Gamma = (\gamma_{ij})_{n \times n} : \Gamma = \Gamma^T, L^{-1} \leq \lambda_{\min}(\Gamma) \leq \lambda_{\max}(\Gamma) \leq L\}.
\]

For any two covariance matrices \(\Gamma_0, \Gamma_1 \in \mathcal{M}\), there exists a testing function \(\phi\), such that for \(n\) large enough,

\[
P_{\Gamma_0} \phi \leq \exp\left(-C\|\Gamma_0 - \Gamma_1\|_F^2\right),
\]

\[
\sup_{\{\Gamma \in \mathcal{M} : \|\Gamma - \Gamma_1\|_F \leq \delta\}} P_{\Gamma} (1 - \phi) \leq \exp\left(-C\|\Gamma_0 - \Gamma_1\|_F^2\right),
\]

where \(P_{\Gamma} = N(0, \Gamma)\) is a \(n\)-variate Gaussian distribution, and \(C, \delta\) are constants only depending on \(L\).

The testing result for \(l^2\)-norm can be established by exploring the equivalence between \(l^2\)-norm and Frobenius norm. The following two lemmas are proved in the supplementary material.

**Lemma 6.3.** Given any \(f_1, f_2\) and \(g_1 = e^{f_1}\) and \(g_2 = e^{f_2}\), we have

\[
2\pi n\|g_1 - g_2\|^2 - \frac{n}{2} \sum_{|h| \geq n/2} (\eta_{1h} - \eta_{2h})^2 \leq \|\Gamma_n(g_1) - \Gamma_n(g_2)\|_F^2 \leq 2\pi n\|g_1 - g_2\|^2.
\]

**Lemma 6.4.** As long as \(\sum_{j=0}^{\infty} j|\theta_j| < B\), there exists \(B' > 0\), such that

\[
|\eta_h| \leq B'|h|^{-1}, \quad \text{for all } h.
\]

**Proof of Lemma 3.7.** For any \(f \in D\), it is uniformly bounded. Therefore, the spectral density \(g = e^f\) is uniformly bounded from up and below. The spectrum of the covariance matrix \(\Gamma(g)\) is also uniformly bounded. There exists sufficiently large \(L\), such that the support of \(\Pi\) is a subset of the matrix class \(\mathcal{M}\) defined in Lemma 6.2. Consider the following testing problem

\[
H_0 : f = f_0, \quad H_1 : f \in \text{supp}(\Pi) \text{ and } ||f - f_1|| \leq \xi||f_0 - f_1||.
\]
We use the notations $g = e^f$, $\Gamma = \Gamma(g)$, and $g_i = e^{f_i}$, $\Gamma_i = \Gamma(g_i)$ for $i = 0, 1$. There exists $b > 0$ such that $b^{-1}||f_0 - f_1|| \leq ||g_0 - g_1|| \leq b||f_0 - f_1||$. The alternative set is

$$\{ f \in \text{supp}(\Pi) : ||f - f_1|| \leq \xi||f_0 - f_1|| \} \subset \{ f \in \text{supp}(\Pi) : ||g - g_1|| \leq b^2\xi||g_0 - g_1|| \}.$$

By Lemma 6.3,

$$2\pi n||g_0 - g_1||^2 - \frac{n}{2} \sum_{|h| > n/2} (\eta_{1h} - \eta_{2h})^2 \leq ||\Gamma_0 - \Gamma_1||_F^2 \leq 2\pi n||g_0 - g_1||^2$$

By Lemma 6.4, there is $C > 0$ such that $\frac{n}{2} \sum_{|h| > n/2} (\eta_{1h} - \eta_{2h})^2 \leq C$. Thus,

$$2\pi n||g_0 - g_1||^2 - C \leq ||\Gamma_0 - \Gamma_1||_F^2 \leq 2\pi n||g_0 - g_1||^2.$$

Therefore, the alternative set is a subset of

$$\left\{ f \in \text{supp}(\Gamma) : ||\Gamma - \Gamma_1||_F^2 \leq b^2\xi\left(||\Gamma_0 - \Gamma_1||_F^2 + C\right) \right\}.$$

Since $||\Gamma_0 - \Gamma_1||_F^2 \geq C'n\epsilon_n^2 \to \infty$, $C < ||\Gamma_0 - \Gamma_1||_F^2$ for a sufficiently large $n$. Thus, the alternative set is contained in

$$\left\{ f \in \text{supp}(\Gamma) : ||\Gamma - \Gamma_1||_F^2 \leq 2b^2\xi||\Gamma_0 - \Gamma_1||_F^2 \right\}.$$

Choose $\xi < \delta(2b^2)^{-1}$, and according to Lemma 6.2, there exists a testing function $\phi$, such that

$$P_{f_0}^{(n)} \phi \leq \exp\left(-C||\Gamma_1 - \Gamma_0||_F^2\right),$$

$$\sup_{\{f \in \text{supp}(\Pi) : ||f - f_1|| \leq \xi||f_0 - f_1||\}} P_f^{(n)}(1 - \phi) \leq \exp\left(-C||\Gamma_1 - \Gamma_0||_F^2\right).$$

The final conclusion follows the relation

$$||\Gamma_0 - \Gamma_1||_F^2 \geq C'n||f_0 - f_1||^2 - C \geq 1/2 C'n ||f_0 - f_1||^2,$$

as $n||f_0 - f_1||^2 \geq n\epsilon_n^2 \to \infty$. 

References.

[1] Barron, A. R. (1988). The exponential convergence of posterior probabilities with implications for Bayes estimators of density functions. Univ. of Illinois.

[2] Barron, A. R. (1989). Uniformly powerful goodness of fit tests. The Annals of Statistics 17, 107–124.
[3] Barron, A., Schervish, M. J. and Wasserman, L. (1999). The consistency of posterior distributions in nonparametric problems. *The Annals of Statistics* **27** 536–561.

[4] Barron, A. R. and Sheu, C.-H. (1991). Approximation of density functions by sequences of exponential families. *The Annals of Statistics* **13** 1347–1369.

[5] Brown, L. D. and Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *The Annals of Statistics* **24** 2384–2398.

[6] Brown, L. D., Cai, T. T., Low, M. G. and Zhang, C.-H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. *The Annals of Statistics* **30** 688–707.

[7] Brown, L. D., Carter, A. V., Low, M. G. and Zhang, C.-H. (2004). Equivalence theory for density estimation, Poisson processes and Gaussian white noise with drift. *The Annals of Statistics* **32** 2074–2097.

[8] Cai, T. T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *The Annals of Statistics* **27** 898–924.

[9] Cai, T. T. (2008). On information pooling, adaptability and superefficiency in nonparametric function estimation. *Journal of Multivariate Analysis* **99** 421–436.

[10] Cai, T. T. and Zhou, H. H. (2009). A data-driven block thresholding approach to wavelet estimation. *The Annals of Statistics* **37** 569–595.

[11] de Jonge, R. and van Zanten, J. (2010). Adaptive nonparametric Bayesian inference using location-scale mixture priors. *The Annals of Statistics* **38** 3300–3320.

[12] Efroimovich, S. Y. (1986). Nonparametric estimation of a density of unknown smoothness. *Theory of Probability & Its Applications* **30** 557–568.

[13] Gao, C. and Zhou, H. H. (2013). Rate-optimal posterior contraction for sparse PCA. *arXiv preprint arXiv:1312.0142*.

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

[15] Ghosal, S., Lember, J. and van der Vaart, A. (2008). Nonparametric Bayesian model selection and averaging. *Electronic Journal of Statistics* **2** 63–89.

[16] Ghosal, S. and van der Vaart, A. (2007). Convergence rates of posterior distributions for noniid observations. *The Annals of Statistics* **35** 192–223.

[17] Golubev, G. K., Nussbaum, M. and Zhou, H. H. (2010). Asymptotic equivalence of spectral density estimation and Gaussian white noise. *The Annals of Statistics* **38** 181–214.

[18] Hall, P., Kerkyacharian, G. and Picard, D. (1998). Block threshold rules for curve estimation using kernel and wavelet methods. *The Annals of Statistics* **26** 922–942.

[19] Kruijer, W., Rousseau, J. and van der Vaart, A. (2010). Adaptive Bayesian density estimation with location-scale mixtures. *Electronic Journal of Statistics* **4** 1225–1257.

[20] Kruijer, W. and van der Vaart, A. (2008). Posterior convergence rates for Dirichlet mixtures of beta densities. *Journal of Statistical Planning and Inference* **138** 1981–1992.

[21] Le Cam, L. (1973). Convergence of estimates under dimensionality restrictions. *The Annals of Statistics* **38**–53.

[22] Mason, D. M. and Zhou, H. (2012). Quantile coupling inequalities and their applications. *Probability Surveys* **9** 439–479.

[23] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *The Annals of Statistics* **2399–2430.

[24] Pollard, D. (1990). Empirical processes: theory and applications. In *NSF-CBMS...*
APPENDIX A: PROOF OF PROPOSITION 2.1

According to the definition, (2.1) is obviously true for $c_1 = 1$. We also have

$$
\int_{-k^2}^{\infty} g_k(t) dt = \left( e^{-k} - e^{-k^2} \right) \exp \left( -e^k \right) \\
\leq \exp \left( - (k + e^k) \right) \\
\leq \exp \left( - e^k \right),
$$
and thus (2.3) is true for $c_3 = 1$. We finally check (2.2).

$$\int_0^\infty t g_k(t) dt = \int_0^{e^{-k^2}} t g_k(t) dt + \int_{e^{-k^2}}^e t g_k(t) dt \leq e^{-2k^2} T_k + e^{-2k} \exp(-e^k)$$

$$\leq e^{-2k^2} (2e^{k^2} + \exp(-e^k)) + e^{-2k} \exp(-e^k)$$

$$\leq 2 \exp(-k^2) + \exp(-2k^2 - e^k) + \exp(-2k - e^k)$$

$$\leq 4e^{-k^2}.$$ 

Thus, (2.2) is true for $c_2 = 1$.

**APPENDIX B: PROOF OF PROPOSITION 6.1**

We use $|| \cdot ||$ to denote the $l^2$ norm in $\mathbb{R}^{N_\beta}$, where $N_\beta$ is defined by $N_\beta = [(n\beta^{-1})^{2\alpha+1}]$. According to the definition of $\mathcal{F}_n$, for any $\theta_1, \theta_2 \in \mathcal{F}_n$, we have

$$\sqrt{\sum_{j>N_\beta} (\theta_{1j} - \theta_{2j})^2} \leq \sqrt{\sum_{j>N_\beta} (\theta_{1j} - \theta_{0j})^2} + \sqrt{\sum_{j>N_\beta} (\theta_{0j} - \theta_{2j})^2} \leq 2\epsilon_n.$$ 

Combining the equivalence between $d$ and $|| \cdot ||$, a $|| \cdot ||_*$-ball is contained in a $d$-ball. That is, for any $\theta^* \in \mathcal{F}_n \cap \text{supp}(\Pi)$, we have

$$\{ \theta \in \mathcal{F}_n \cap \text{supp}(\Pi) : d(\theta, \theta^*) \leq l\xi \epsilon_n \} \supset \{ \theta \in \mathcal{F}_n \cap \text{supp}(\Pi) : ||\theta - \theta^*||_* \leq (b^{-1}l\xi - 2)\epsilon_n \}.$$ 

We bound each $d$-ring by

$$\{ f \in \mathcal{F}_n \cap \text{supp}(\Pi) : l\epsilon_n < d(f, f_0) \leq (l + 1)\epsilon_n \} \supset \{ f \in \mathcal{F}_n : ||f - f_0|| \leq b(l + 1)\epsilon_n \} \supset \{ \theta \in \mathcal{F}_n : ||\theta - \theta_0||_* \leq b(l + 1)\epsilon_n \}.$$ 

Therefore, we have

$$\log N\left(\xi l\epsilon_n, \{ f \in \mathcal{F}_n \cap \text{supp}(\Pi) : l\epsilon_n < d(f, f_0) \leq (l + 1)\epsilon_n \}, d \right) \leq \log N\left((b^{-1}l\xi - 2)\epsilon_n, \{ \theta \in \mathcal{F}_n : ||\theta - \theta_0||_* \leq b(l + 1)\epsilon_n \}, || \cdot ||_* \right) \leq N_\beta \log \left(\frac{6b(l + 1)}{b^{-1}l\xi - 2}\right),$$
where the last inequality is a covering number calculation in $\mathbb{R}^{N_\beta}$, due to Lemma 4.1 of Pollard (1990). Since $l > M$, for $M$ sufficiently large, the above quantity can be upper bounded by

$$N_\beta \log \left( \frac{12b(l + 1)}{b^{-1}l\xi} \right) \leq N_\beta \log \left( \frac{24b^2\xi^{-1}}{n\epsilon^2} \right) \leq C(\beta, b, \xi) n\epsilon^2.$$  

Thus, the proof is complete.

APPENDIX C: PROOFS OF SOME AUXILIARY LEMMAS

The proof of the Lemma 3.1 and Lemma 3.5 are similar. We present the first proof and sketch the second.

**Proof of Lemma 3.1.** We keep using the notations in the proof of Theorem 2.1. We are going to prove (2.5) and (2.6) for the prior $\Pi$. By the property of conditioning, we have

$$\Pi \left\{ \sum_{j=1}^{\infty} (\theta_{0j} - \theta_j)^2 \leq \epsilon_n^2 \right\} \geq \Pi \left\{ \sum_{j=1}^{\infty} (\theta_{0j} - \theta_j)^2 \leq \epsilon_n^2, \sum_{j=1}^{\infty} |\theta_j| < B \right\}$$

$$\geq \Pi \left\{ \sum_{j=1}^{J_\alpha} (\theta_{0j} - \theta_j)^2 \leq \epsilon_n^2/2, \sum_{j=1}^{J_\alpha} |\theta_j| < B/2 \right\} + \Pi \left\{ \sum_{j=J_\alpha}^{\infty} (\theta_{0j} - \theta_j)^2 \leq \epsilon_n^2/2, \sum_{j=J_\alpha}^{\infty} |\theta_j| < B/2 \right\}$$

$$= \Pi \left\{ \sum_{j=1}^{J_\alpha} (\theta_{0j} - \theta_j)^2 \leq \epsilon_n^2/2, \sum_{j=1}^{J_\alpha} |\theta_j| < B/2 \right\} \Pi \left\{ \sum_{j=J_\alpha}^{\infty} (\theta_{0j} - \theta_j)^2 \leq \epsilon_n^2/2, \sum_{j=J_\alpha}^{\infty} |\theta_j| < B/2 \right\},$$

where we redefine $J_\alpha$ by $J_\alpha = Gn^{2\alpha+1}$, and we choose $G$ large enough such that $G^{-2\alpha}Q^2 \leq 1/8$. We round the number $J_\alpha$ to the nearest boundary of block so that independence of blocks can be used in the last equality. For the second term in the previous display,

$$\Pi \left\{ \sum_{j>J_\alpha} (\theta_{0j} - \theta_j)^2 \leq \epsilon_n^2/2, \sum_{j>J_\alpha} |\theta_j| < B/2 \right\} \geq \Pi \left\{ \sum_{j>J_\alpha} \theta_j^2 \leq \epsilon_n^2/8, \sum_{j>J_\alpha} |\theta_j| < B/2 \right\}$$

$$\geq 1 - \Pi \left\{ \sum_{j>J_\alpha} \theta_j^2 > \epsilon_n^2/8 \right\} - \Pi \left\{ \sum_{j>J_\alpha} |\theta_j| > B/2 \right\}.$$
where the second inequality is because

\[
2 \sum_{j > J_\alpha} \theta_{0j}^2 \leq J_\alpha^{-2\alpha} Q^2 \leq \epsilon_n^2 G^{-2\alpha} Q^2 \leq \frac{\epsilon_n^2}{8}.
\]

Since \( \mathbb{E} A_k \leq e^{-c_2 k^2} \) from (2.2), it is easy to show by Markov inequality,

\[
\bar{\Pi} \left\{ \sum_{j > J_\alpha} \theta_{0j}^2 > \frac{\epsilon_n^2}{8} \right\} + \bar{\Pi} \left\{ \sum_{j > J_\alpha} |\theta_j| > \frac{B}{2} \right\} \rightarrow 0.
\]

Therefore,

\[
\Pi \left\{ \sum_{j=1}^\infty (\theta_{0j} - \theta_j)^2 \leq \epsilon_n^2 \right\} \geq \frac{1}{2} \bar{\Pi} \left\{ \sum_{j=1}^{J_\alpha} (\theta_{0j} - \theta_j)^2 \leq \frac{\epsilon_n^2}{2} \sum_{j=1}^{J_\alpha} |\theta_j| < \frac{B}{2} \right\}
\]

\[
= \frac{1}{2} \bar{\Pi} \left\{ \sum_{j=1}^{J_\alpha} (\theta_{0j} - \theta_j)^2 \leq \frac{\epsilon_n^2}{2} \right\}.
\]

The last equality is because

\[
(C.1) \quad \left\{ \sum_{j=1}^{J_\alpha} (\theta_{0j} - \theta_j)^2 \leq \frac{\epsilon_n^2}{2} \right\} \subset \left\{ \sum_{j=1}^{J_\alpha} |\theta_j| < \frac{B}{2} \right\},
\]

which is from calculation

\[
\sum_{j=1}^{J_\alpha} |\theta_j| \leq \sum_{j=1}^{J_\alpha} \left| \theta_j - \theta_{0j} \right| + \sum_{j=1}^{J_\alpha} |\theta_{0j}| \leq \sqrt{J_\alpha} \left( \sum_{j=1}^{J_\alpha} (\theta_{0j} - \theta_j)^2 \right)^{1/2} + \left( \sum_{j=1}^{J_\alpha} j^{2\alpha^*} \theta_{0j}^2 \right)^{1/2} \left( \sum_{j=1}^{J_\alpha} j^{-2\alpha^*} \right)^{1/2}
\]

\[
\leq \sqrt{J_\alpha} \epsilon_n^2 + Q \gamma = \left( G n^{-\frac{2\alpha}{2\alpha + 1}} \right)^{1/2} + \frac{B}{3} \leq \frac{B}{2},
\]

where we have used the assumption \( \alpha \geq \alpha^* > 1/2 \) and \( 3Q \gamma \leq B \). Use the same method in the proof of Theorem 2.1, it can be shown that

\[
\bar{\Pi} \left\{ \sum_{j=1}^{J_\alpha} (\theta_{0j} - \theta_j)^2 \leq \frac{\epsilon_n^2}{2} \right\} \geq \exp \left( -C n \epsilon_n^2 \right).
\]

Therefore, (2.5) is proved for \( \Pi \). We proceed to bound \( \Pi(F_n^c) \), which is relatively easy.

\[
\Pi(F_n^c) = \frac{\bar{\Pi} \left( F_n^c \cap \left\{ \sum_{j=1}^{\infty} |\theta_j| < B \right\} \right)}{\bar{\Pi} \left\{ \sum_{j=1}^{\infty} |\theta_j| < B \right\}} \leq \frac{\bar{\Pi} \left( F_n^c \right)}{\bar{\Pi} \left\{ \sum_{j=1}^{\infty} |\theta_j| < B \right\}}.
\]
Notice the denominator $\bar{\Pi} \left\{ \sum_{j=1}^{\infty} |\theta_j| < B \right\}$ is a positive constant independent of $n$. Therefore, we can bound $\Pi(F^c_n)$ by the same argument in the proof of Theorem 2.1 and obtain the same bound as that of $\bar{\Pi}(F^c_n)$ by choosing sufficiently small $\beta$ in (2.4). The proof is complete.

**Proof of Lemma 3.5.** The only difference of the proof from that of Lemma 3.1 is (C.1). Here we need

$$\left\{ \sum_{j=1}^{J_\alpha} (\theta_{0j} - \theta_j)^2 \leq \frac{\epsilon^2}{2} \right\} \subset \left\{ \sum_{j=1}^{J_\alpha} |\theta_j| < \frac{B}{2} \right\},$$

which is from

$$\sum_{j=1}^{J_\alpha} j|\theta_j| \leq \sum_{j=1}^{J_\alpha} j|\theta_j - \theta_{0j}| + \sum_{j=1}^{J_\alpha} j|\theta_{0j}|$$

$$\leq J_\alpha^{3/2} \left( \sum_{j=1}^{J_\alpha} (\theta_j - \theta_{0j})^2 \right)^{1/2} + \left( \sum_{j=1}^{J_\alpha} j^{2\alpha^*} \theta_{0j}^2 \right)^{1/2} \left( \sum_{j=1}^{J_\alpha} j^{-2\alpha^*+2} \right)^{1/2}$$

$$\leq \sqrt{J_\alpha^{3/2} \epsilon_n^2} + Q\gamma = \left( G n^{-\frac{2\alpha^*+3}{2\alpha^*+1}} \right)^{1/2} + \frac{B}{3} \leq \frac{B}{2}.$$

Repeat other parts in the proof of Lemma 3.1, we reach the desired conclusion.

**Proof of Lemma 3.2.** For $p_f = \frac{e^f}{\int e^f}$, we denote $\psi(f) = \log \int e^f$, and then we have $p_f = \exp(f - \psi(f))$. Notice that

$$||f||_\infty = \left\| \sum_j \theta_j \phi_j \right\|_\infty \leq \sqrt{2} \sum_j |\theta_j| \leq \sqrt{2}B,$$

which implies $\exp(-2\sqrt{2}B) \leq p_f(x) \leq \exp(2\sqrt{2}B)$ for any $x$. We use an inequality from Lemma 1 in Barron and Sheu (1991). For any constant $c$,

$$D(P_{f_1}, P_{f_2}) \leq \frac{1}{2} e^{||\log(p_{f_1}/p_{f_2}) - c||_\infty} \int p_{f_1}(x) \left( \log \frac{p_{f_1}(x)}{p_{f_2}(x)} - c \right)^2 dx.$$

Choose $c = \psi(f_1) - \psi(f_2)$, we have

$$H^2(P_{f_1}, P_{f_2}) \leq D(P_{f_1}, P_{f_2}) \leq e^{2\sqrt{2}B} \int p_{f_1}(f_1 - f_2)^2 \leq e^{4\sqrt{2}B} ||f_1 - f_2||^2.$$
Also, by using the reverse version of the inequality in Lemma 1 in Barron and Sheu (1991), we have
\[
V(P_{f_1}, P_{f_2}) \leq e^{\frac{1}{2}||\log(p_{f_1}/p_{f_2})||_\infty} D(P_{f_1}, P_{f_2}) \leq e^{6\sqrt{B}||f_1 - f_2||^2}.
\]
On the other hand,
\[
V(P_{f_1}, P_{f_2}) = \int p_{f_1} \left( f_1 - f_2 - \psi(f_1) + \psi(f_2) \right)^2 \geq e^{-2\sqrt{B}} \int \left( f_1 - f_2 - \psi(f_1) + \psi(f_2) \right)^2 = e^{-2\sqrt{B}} \int (f_1 - f_2)^2 + e^{-2\sqrt{B}} \left( \psi(f_1) - \psi(f_2) \right)^2 - 2e^{-2\sqrt{B}} \left( \psi(f_1) - \psi(f_2) \right) \int (f_1 - f_2) \geq e^{-2\sqrt{B}} ||f_1 - f_2||^2,
\]
because \( \int (f_1 - f_2) = \int f_1 - \int f_2 = 0 \). Using Lemma 8.2 in Ghosal, Ghosh and van der Vaart (2000), we have
\[
D(P_{f_1}, P_{f_2}) \leq 2H^2(P_{f_1}, P_{f_2}) \left\| \frac{P_{f_1}}{P_{f_2}} \right\|_\infty \leq 2e^{4\sqrt{B}} H^2(P_{f_1}, P_{f_2}).
\]
Therefore,
\[
H^2(P_{f_1}, P_{f_2}) \geq \frac{1}{2} e^{-4\sqrt{B}} D(P_{f_1}, P_{f_2}) \geq \frac{1}{2} e^{-6\sqrt{B}} V(P_{f_1}, P_{f_2}) \geq \frac{1}{2} e^{-8\sqrt{B}} ||f_1 - f_2||^2.
\]
The proof is complete. □

**Proof of Lemma 3.6.** Consider the multivariate Gaussian distribution \( P_{\Gamma} = N(0, \Gamma) \). Then, we have
\[
D(P_{\Gamma_1}, P_{\Gamma_2}) = P_{\Gamma_1} \left( -\frac{1}{2} \log \det(\Gamma_1^{-1}) - \frac{1}{2} \text{tr}\left( \Gamma_1^{-1} - \Gamma_2^{-1} \right) XX^T \right) = -\frac{1}{2} \log \det(\Gamma_1^{-1}) - \frac{1}{2} \text{tr}\left( I - \Gamma_1^{-1} \right) = -\frac{1}{2} \log \det \left( \Gamma_2^{-1/2} \Gamma_1 \Gamma_2^{-1/2} \right) + \frac{1}{2} \text{tr}\left( \Gamma_2^{-1/2} \Gamma_1 \Gamma_2^{-1/2} - I \right) \leq \frac{1}{4} ||\Gamma_2^{-1/2} \Gamma_1 \Gamma_2^{-1/2} - I||^2_F \leq \frac{1}{4} ||\Gamma_2^{-1/2}||^2_F ||\Gamma_1 - \Gamma_2||^2_F.
where we use $\| \cdot \|$ to denote the matrix spectral norm here. We also have
\[
V(P_{\Gamma_1}, P_{\Gamma_2}) = \text{Var}_{P_{\Gamma}} \left( \frac{1}{2} X^T (\Gamma_1^{-1} - \Gamma_2^{-1}) X \right)
\]
\[
= \frac{1}{4} \text{Var}_{P_{\Gamma}} \left( X^T (\Gamma_1^{-1} - \Gamma_2^{-1}) X \right)
\]
\[
= \frac{1}{2} \| I - \Gamma_1^{1/2} \Gamma_2^{-1/2} \Gamma_1^{1/2} \|_F^2
\]
\[
= \frac{1}{2} \| I - \Gamma_2^{-1/2} \Gamma_1 \Gamma_2^{-1/2} \|_F^2
\]
\[
\leq \frac{1}{2} \| \Gamma_1^{-1} \|_F^2 \| \Gamma_1 - \Gamma_2 \|_F^2.
\]

In the spectral density estimation, we have $P_{f_1}^{(n)} = P_{\Gamma_1}$ and $P_{f_2}^{(n)} = P_{\Gamma_2}$ with $\Gamma_1 = \Gamma(g_1)$ and $\Gamma_2 = \Gamma(g_2)$. Then, we have $\| \Gamma_1^{-1} \|_F^2 \leq (2\pi)^{-1} \| g_1^{-1} \|_\infty \leq e^B$. We also have $\| \Gamma_1 - \Gamma_2 \|_F^2 = \sum_{|h| < n} (n - |h|)(\eta_{1h} - \eta_{2h})^2 \leq 2\pi n \| g_1 - g_2 \|_F^2 \leq C_B \| f_1 - f_2 \|_F^2$. Therefore, we have
\[
D(P_{\Gamma_1}, P_{\Gamma_2}) \leq bn \| f_1 - f_2 \|_F^2 \quad \text{and} \quad V(P_{\Gamma_1}, P_{\Gamma_2}) \leq bn \| f_1 - f_2 \|_F^2,
\]
with $b = \frac{1}{2} e^{2B} C_B$. □

**Proof of Lemma 5.1.** Notice
\[
\mathbb{P} \left( \sum_j a_j (Z_j - c_j)^2 \leq \epsilon^2 \right) = \mathbb{P} \left( \sum_j \frac{a_j}{b_j} b_j (Z_j - c_j)^2 \leq \epsilon^2 \right)
\]
\[
\leq \mathbb{P} \left( R^{-1} \sum_j b_j (Z_j - c_j)^2 \leq \epsilon^2 \right),
\]
where the inequality is because
\[
\sum_j \frac{a_j}{b_j} b_j (Z_j - c_j)^2 \geq R^{-1} \sum_j b_j (Z_j - c_j)^2.
\]
The other side is similar. Thus, the proof is complete. □

**Proof of Lemma 6.3.** By definition,
\[
\| \Gamma_n(g_1) - \Gamma_n(g_2) \|_F^2 = \sum_{|h| < n} (n - |h|)(\eta_{1h} - \eta_{2h})^2.
\]
The upper bound is by
\[
\sum_{|h| < n} (n - |h|)(\eta_{1h} - \eta_{2h})^2 \leq n \sum_{|h| < n} (\eta_{1h} - \eta_{2h})^2 \leq 2\pi n \| g_1 - g_2 \|_F^2.
\]
The lower bound is given by
\[
\sum_{|h|<n} (n - |h|)(\eta_1h - \eta_2h)^2 \geq \sum_{|h|<n/2} (n - |h|)(\eta_1h - \eta_2h)^2 
\geq \frac{n}{2} \sum_{|h|<n/2} (\eta_1h - \eta_2h)^2 
= \pi n |g_1 - g_2|^2 - \frac{n}{2} \left( \sum_{|h|\geq n/2} (\eta_1h - \eta_2h)^2 \right).
\]
Thus, the proof is complete.

**Proof of Lemma 6.4.** We use the notation \(\phi_j(t) = \cos(jt)\) and \(\psi_j(t) = \sin(jt)\) for each \(j\). According to the setting, \(f = \sum_j \theta_j \phi_j\) and \(g = \sum_j \eta_j \phi_j\). Since we assume \(\alpha > 3/2\), the derivatives of both \(f\) and \(g\) exist. Using the relation \(gf' = g'\), we have
\[
(C.2) \quad \left(\sum_j j \theta_j \psi_j\right) \left(\sum_j \eta_j \phi_j\right) = \sum_j j \eta_j \psi_j.
\]
Using the relation \(\psi_m \phi_n = \frac{1}{\sqrt{2}} \left(\psi_{m+n} + \psi_{m-n}\right)\), the left side of (C.2) is
\[
\sum_{m,n} m \theta_m \eta_n \psi_m \phi_n = \frac{1}{\sqrt{2}} \sum_{m,n} m \theta_m \eta_n \psi_{m+n} + \frac{1}{\sqrt{2}} \sum_{m,n} m \theta_m \eta_n \psi_{m-n}
= \frac{1}{\sqrt{2}} \sum_{k=2}^{\infty} \left( \sum_{m+n=k} m \theta_m \eta_n \right) \psi_k 
+ \frac{1}{\sqrt{2}} \sum_{k=1}^{\infty} \left( \sum_{m-n=k} m \theta_m \eta_n - \sum_{n-m=k} m \theta_m \eta_n \right) \psi_k
= \frac{1}{\sqrt{2}} \sum_{k=1}^{\infty} \left( \sum_{l=1}^{k-1} l \theta_l \eta_{k-l} + \sum_{l=k+1}^{\infty} l \theta_l \eta_{l-k} - \sum_{l=1}^{\infty} l \theta_l \eta_{l+k} \right) \psi_k.
\]
Since \(\{\psi_k\}\) is orthogonal, we must have
\[
k \eta_k = \frac{1}{\sqrt{2}} \sum_{l=1}^{k-1} l \theta_l \eta_{k-l} + \frac{1}{\sqrt{2}} \sum_{l=k+1}^{\infty} l \theta_l \eta_{l-k} - \frac{1}{\sqrt{2}} \sum_{l=1}^{\infty} l \theta_l \eta_{l+k}
= \frac{1}{\sqrt{2}} \sum_{l=1}^{\infty} l \theta_l \eta_{k-l} - \frac{1}{\sqrt{2}} \sum_{l=1}^{\infty} l \theta_l \eta_{l+k}
= \frac{1}{\sqrt{2}} \sum_{l=1}^{\infty} l \theta_l (\eta_{k-l} - \eta_{k+l}),
\]

which establishes the relation between \( \{\eta_j\} \) and \( \{\theta_j\} \). For each \( j \),

\[
|\eta_j| = \left| \int g\phi_j \right| \leq e^B.
\]

Therefore,

\[
|k\eta_k| = \left| \frac{1}{\sqrt{2}} \sum_{i=1}^{\infty} l\theta_l(\eta_{k-l} - \eta_{k+l}) \right| \leq \sqrt{2}e^B \sum_{i=1}^{\infty} |l\theta_l| \leq \sqrt{2}Be^B.
\]

The proof is complete.

APPENDIX D: PROOFS OF LEMMA 4.1, THEOREM 4.1 AND COROLLARY 4.1

Proof of Lemma 4.1. Define \( K_n = \{ D(P_{f_0}^{(n)}, P_{f}^{(n)}) \leq bn\epsilon_n^2 \} \) and the renormalized prior \( \tilde{\Pi}(A) = \frac{\Pi(A \cap K_n)}{\Pi(K_n)} \). We have

\[
P_{f_0}^{(n)}(\mathcal{H}_n^c) \leq P_f^{(n)} \left( \int \frac{P_{f_0}^{(n)}}{P_f^{(n)}}(X^n)d\tilde{\Pi}(f) \leq \exp \left( -(b+1)n\epsilon_n^2 \right) \right)
\]

\[
\leq P_{f_0}^{(n)} \left( \int \frac{P_{f_0}^{(n)}}{P_f^{(n)}}(X^n)d\tilde{\Pi}(f) \geq (b+1)n\epsilon_n^2 \right),
\]

where the last inequality is Jensen’s inequality. From now on, we prove each statistical model in Section 3 respectively. First, for the density estimation model, define \( Y_i = \int \log \frac{P_{f_0}^{(n)}}{P_f^{(n)}}(X_i)d\tilde{\Pi}(f) \). Then, it is easy to see that

\[
\int \log \frac{P_{f_0}^{(n)}}{P_f^{(n)}}(X^n)d\tilde{\Pi}(f) = \sum_{i=1}^{n} Y_i \quad \text{and} \quad P_{f_0}Y_i \leq bn\epsilon_n^2.
\]

Since \( f, f_0 \in D \), \( Y_i \) is bounded as \( |Y_i| \leq 4\sqrt{2}B \) for \( i = 1, 2, ..., n \). Using Hoeffding’s inequality, we have

\[
P_{f_0}^{(n)}(\mathcal{H}_n^c) \leq P_{f_0}^{(n)} \left( \frac{1}{n} \sum_{i=1}^{n} (Y_i - \mathbb{E}Y_i) \geq c_n^2 \right)
\]

\[
\leq \exp \left( -Cn\epsilon_n^2 \right).
\]

Next, we consider Gaussian sequence model. By definition,

\[
\log \frac{P_{f_0}^{(n)}}{P_f^{(n)}}(X^n) = \sqrt{n} \sum_{i=1}^{n} Z_i(\theta_{i0} - \theta_i) + \frac{n}{2}||\theta_0 - \theta||^2,
\]
where $Z_i = \sqrt{n}(X_i - \theta_{i0}) \sim N(0,1)$ under $P_{f_0}^{(n)}$. Letting $E_{\tilde{\Pi}}$ be the expectation under $\tilde{\Pi}$, we have

$$P_{f_0}^{(n)}(H_n^c) \leq \mathbb{P} \left( \sum_{i=1}^{\infty} Z_i(\theta_{i0} - E_{\tilde{\Pi}}\theta_i) \geq \sqrt{ne_n^2} \right) \leq \mathbb{P} \left( \sum_{i=1}^{\infty} Z_i(\theta_{i0} - E_{\tilde{\Pi}}\theta_i) \geq \frac{\sqrt{ne_n^2}}{\sqrt{2b}} \right) \leq \exp\left( -\bar{C}ne_n^2 \right).$$

The calculation for the white noise model is the same. We consider Gaussian regression now. By definition, we have

$$\log \mathbb{P}_{f_0}^{(n)}(X^n) = \sum_{i=1}^{n} Z_i \left( f_0(X_i) - f(X_i) \right) + \frac{1}{2} \sum_{i=1}^{n} \left( f_0(X_i) - f(X_i) \right)^2,$$

where $Z_i = Y_i - f_0(X_i) \sim N(0,1)$ under $P_{f_0}^{(n)}$. Define

$$Y_i = Z_i \left( f_0(X_i) - \int f(X_i) d\tilde{\Pi}(f) \right) + \frac{1}{2} \int \left( f_0(X_i) - f(X_i) \right)^2 d\tilde{\Pi}(f).$$

Then, we have $\int \log \mathbb{P}_{f_0}^{(n)}(X^n) d\tilde{\Pi}(f) = \sum_{i=1}^{n} Y_i$ and $P_{f_0}^{(n)} Y_i \leq bne_n^2$. Since $f, f_0 \in D$, we have $||f||_\infty \vee ||f_0||_\infty \leq \sqrt{2B}$, $Y_i$ is sub-Gaussian random variable. Hence,

$$P_{f_0}^{(n)}(H_n^c) \leq P_{f_0}^{(n)} \left( \frac{1}{n} \sum_{i=1}^{n} (Y_i - EY_i) \geq \epsilon_n^2 \right) \leq \exp\left( -\bar{C}ne_n^2 \right).$$

The case for spectral density estimation falls into the general Gaussian covariance matrix estimation theory. The proof is similar to the proof of Lemma 5.1 in Gao and Zhou (2013), and is omitted here. \[\Box\]

**Proof of Theorem 4.1.** This is just repeating the argument in the proof of Theorem 2.2 by using an improved bound for $P_{f_0}^{(n)}(H_n^c)$ provided by Lemma 4.1. \[\Box\]

**Proof of Corollary 4.1.** In the cases of density estimation, Gaussian regression and spectral density estimation, the norm $|| \cdot ||$ is bounded in the support of the prior. Therefore, the conclusion follows the same argument in
Ghosal, Ghosh and van der Vaart (2000). In the cases of Gaussian sequence model and white noise model, we have

\[ P_f^0 \| \mathbb{E}_{\tilde{\Pi}}(\theta | X^n) - \theta_0 \|^2 \]
\[ \leq P_f^0 \mathbb{E}_{\tilde{\Pi}}(\| \theta - \theta_0 \|^2 | X^n) \]
\[ \leq \sum_{j \geq M} P_f^0 \mathbb{E}_{\tilde{\Pi}}(\| \theta - \theta_0 \|^2 I_{A_j} | X^n), \]

where \( A_j = \{ j \epsilon_n^2 < \| \theta - \theta_0 \|^2 \leq (j + 1) \epsilon_n^2 \} \) for each \( j \). Bounding each summand by \((j + 1) \epsilon_n^2 P_f^0 \tilde{\Pi}(A_j | X^n)\), the proof is complete. The details are omitted here.

APPENDIX E: PROOFS OF THEOREM 4.2 AND THEOREM 4.3

The proof of Theorem 4.2 mimics the proof of Theorem 2.1. For each \( k \), we redefine

\[ A_{\alpha,k} = 2^{-2\alpha k} - 2^{-2\alpha(k+1)} \]

Define \( K \) to be the smallest integer such that

\[ K \geq 1 \frac{\log_2 \left( \frac{4^\alpha}{4^\alpha - 1} 8Q^2 \right)}{2\alpha + 1} \log_2(n). \]

**Lemma E.1.** For any \( \theta \in B_{p,q}^\alpha(Q) \), with \( \alpha > 0, p \geq 2, q \geq 1, Q > 0 \), we have

\[ \sum_{k \geq K} \| \theta_k \|^2 \leq \frac{4^\alpha}{4^\alpha - 1} Q^2 2^{-2\alpha K}, \]

for any \( K \).

**Lemma E.2.** For any \( \alpha > 0 \), and \( \theta_0 \in B_{p,q}^\alpha(Q) \), there exists \( C_3 > 0 \), such that

\[ \tilde{\Pi}_{K}^{A_{\alpha}} \left\{ \sum_{j=1}^{\infty} (\theta_j - \theta_{0j})^2 \leq \epsilon_n^2 \right\} \geq 2^{-C_3 n \epsilon_n^4}. \]

**Lemma E.3.** For each \( k \), let \( A_k \sim g_k \), with \( \{g_k\} \) satisfying (2.1)-(2.3). we have

\[ \mathbb{P}(V_{\alpha}) \geq 2^{-C_2 n \epsilon_n^2}. \]
Lemma E.4. For $K$ defined above, and $\theta_0 \in B_{p,q}^\alpha(Q)$, we have
\[ \Pi \left\{ \sum_{k>K} \|\theta_k - \theta_{0k}\|^2 \leq \frac{\epsilon_n^2}{2} \right\} \geq \frac{1}{2}, \]
for sufficiently large $n$.

Proof of Theorem 4.2. The proof is the same as the proof of Theorem 2.1 by combining the above lemmas.

Proof of Theorem 4.3. This is a direct implication of Theorem 2.2.

Among the above four lemmas, we only prove Lemma E.1 and Lemma E.2. The proof of the other two are the same as the proof of Lemma 5.3 and Lemma 5.4.

Proof of Lemma E.1. For any $\theta \in B_{p,q}^\alpha(Q)$, since $B_{p,q}^\alpha(Q) \subset B_{2,\infty}^\alpha(Q)$, we have $\theta \in B_{2,\infty}^\alpha(Q)$, where
\[ B_{2,\infty}^\alpha(Q) = \left\{ \theta : \max_k \left( 2^{\alpha k} \|\theta_k\| \right) \leq Q \right\}. \]
Thus,
\[ \sum_{k \geq K} \|\theta_k\|^2 \leq Q^2 \sum_{k \geq K} 2^{-2\alpha k} = \frac{4^\alpha}{4^\alpha - 1} Q^2 2^{-2\alpha K}. \]

Proof of Lemma E.2. The proof is essentially the same as in the proof of Lemma 5.2. The only slight difference is the approximation of $\theta_0$ by the RKHS of the Gaussian process. For each $k$, we have
\[ 2^{-(2\alpha+1)(k+1)} \leq A_{\alpha,k} \leq 2^{-(2\alpha+1)k}. \]
Thus, using the fact that $\theta_0 \in B_{p,q}^\alpha(Q) \subset B_{2,\infty}^\alpha(Q)$, we have
\[ \|\theta_0\|_{H^{\alpha}}^2 = \sum_{k=1}^K A_{\alpha,k}^{-1} \|\theta_{0k}\|^2 \leq Q^2 \sum_{k=1}^K 2^{2\alpha + k + 1} \leq Q^2 2^{2\alpha + 2K} \leq C_n \epsilon_n^2, \]
for some $C > 0$. 

Department of Statistics
Yale University
E-mail: chao.gao@yale.edu
huibin.zhou@yale.edu