

The Bayes Lepski's Method and Credible Bands through Volume of Tubular Neighborhoods

William Weimin Yoo and Aad W. van der Vaart
Leiden University

Abstract

For a general class of priors based on random series basis expansion, we develop the Bayes Lepski's method to estimate unknown regression function. In this approach, the series truncation point is determined based on a stopping rule that balances the posterior mean bias and the posterior standard deviation. Equipped with this mechanism, we present a method to construct adaptive Bayesian credible bands, where this statistical task is reformulated into a problem in geometry, and the band's radius is computed based on finding the volume of certain tubular neighborhood embedded on a unit sphere. We consider two special cases involving B-splines and wavelets, and discuss some interesting consequences such as the uncertainty principle and self-similarity. Lastly, we show how to program the Bayes Lepski stopping rule on a computer, and numerical simulations in conjunction with our theoretical investigations concur that this is a promising Bayesian uncertainty quantification procedure.

Keywords: Bayes Lepski, volume of tube, adaptive credible bands, supremum norm posterior contraction, B-splines, CDV wavelets, uncertainty principle, self-similar.

MSC2010 classifications: Primary 62G15 62G05; secondary 62G08 62C10

1 Introduction

Uncertainty quantification is now an important research direction in Bayesian nonparametrics, and substantial efforts have been made to answer the question whether Bayesian credible sets are indeed frequentist confidence regions, especially in infinite-dimensional settings. As Bayesian uncertainty quantification can be conducted automatically once the posterior distribution is derived from our modeling assumptions, a positive answer to this question will give statisticians a powerful alternative to other commonly used procedures such as the bootstrap.

In this paper, we will study the issue of constructing adaptive Bayesian simultaneous credible bands that will have both high levels of credibility and also coverage probability in the frequentist sense. To begin our investigation, let us consider the nonparametric regression model

$$Y_i = f(X_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

where Y_i is the response variable, X_i is the covariate, and the errors $\varepsilon_1, \dots, \varepsilon_n$ are independent and identically distributed (i.i.d.) as $N(0, \sigma^2)$ with unknown $0 < \sigma < \infty$. Each X_i takes values in some rectangular region in \mathbb{R} , which is assumed to be $[0, 1]$ without loss of generality.

Credible bands are constructed based on the supremum norm distance. They are easily visualized and they allow one to immediately see the extent of uncertainty as delimited by the bands. It is for this reason that credible bands are more statistically meaningful than for example L_2 -type credible sets. Several methods have been proposed in the Bayesian literature. We have the fixed radius L_∞ -ball around the posterior mean through B-splines random series prior ([23]) and scaled Brownian motion prior (Chapter 3 of [19]), the technique of “gluing” pointwise intervals for linear functionals to produce bands ([1, 20]) and the intersection of two balls through the multiscale approach ([16, 2]).

Estimation and uncertainty quantification for nonparametric models, such as the regression problem considered in this paper, depend crucially on the choice of the tuning parameter in any proposed statistical procedure. Let us work with random series priors by projecting f onto a general class of J -dimensional basis functions and eliciting normal priors on the coefficients. We know from minimax theory that the optimal J when loss is measured using the L_∞ -distance is different from the case of L_2 -loss. Specifically, if the true regression function f_0 is α -Hölder smooth, then J must be of the order $(n/\log n)^{1/(2\alpha+1)}$ for L_∞ -norm and $n^{1/(2\alpha+1)}$ for L_2 -norm. If we assign J a prior or estimate it using empirical Bayes, the resulting posterior will concentrate around the L_2 -optimal J , due to L_2 -norm being the intrinsic metric for the likelihood function. Therefore for inference tasks based on the L_∞ -distance such as constructing Bayesian credible bands, the usual Bayesian procedures will not “choose” the appropriate L_∞ -norm J and this discrepancy causes the credible bands to have inadequate coverage probabilities.

To circumvent this problem, we propose a procedure to choose the correct L_∞ -norm J by using only summaries of the posterior distribution such as its mean and variance. The dynamics of bias-variance tradeoff dictates that the tuning parameter J is chosen to balance the supremum norm bias of the posterior mean and the posterior standard deviation. Based on this observation, we design a Lepski’s type Bayesian stopping rule by successively reducing the value of J until the point when the posterior mean bias starts to dominate the posterior standard deviation. If we adopt the framework that all Bayesian estimation and inference must be done through the posterior distribution, then this procedure is admissible within this framework since we are using posterior quantities to decide when to stop. Despite the intuitive nature of Lepski’s type procedure, it has since its inception ([11, 12]) acquired the reputation of being impractical to use in actual practice, and consequently has only remained as a theoretical tool for statisticians to construct adaptive procedures.

In this paper, we hope to shatter this perception by showing that the aforementioned Bayes Lepski’s method can be programmed on a computer to do actual tuning parameter selection. As the optimal J is increasing in n , we tend to stop sooner when sample size is larger, and this is in stark contrast to established methods such as leave-one-out cross validation. It will be seen that pairwise differences in posterior means are used as a proxy for posterior mean bias, and due to their nature of being shrinkage estimators results in better performance (in terms of coverage probabilities) when compared to their frequentist counterpart. More importantly as a Bayesian procedure, the Bayes Lepski’s method alleviates the need to design complicated reversible jump Markov Chain Monte Carlo algorithms

if J was assigned a prior, and it moreover ensures that the correct J is chosen based on the loss/distance used in the problem.

Equipped with this mechanism, we present a method to construct Bayesian credible bands with guaranteed coverage probabilities, where this statistical inference task is reformulated into a problem in geometry, and the band's radius is computed based on finding the volume of certain tubular neighborhoods embedded on a unit sphere. In many cases, this volume can be explicitly calculated and it yields closed-form formulas for radius computation. Thus, the volume of tube approach bypasses the need to sample repeatedly from the posterior to estimate the radius empirically, and coupled with the Bayes Lepski's method introduced previously, results in an efficient procedure for adaptive Bayesian credible band construction with coverage guarantees. The idea of using volume of geometrical objects to compute statistical quantities goes back to [7] in the context of hypothesis testing. Subsequent authors used and generalized this idea to other applications in frequentist statistics, for example volume tests of significance, simultaneous inference, projection pursuit regression and others (see [15, 10, 9]). The Bayes Lepski volume of tube method presented above was partly inspired by these ideas and we address some important issues that were not considered by these authors, in particular tuning parameter selection (through Bayes Lepski) and self-similarity.

We apply our general result to B-splines and wavelets, and this inadvertently led us to some counterintuitive notions in uncertainty quantification. It should be intuitively clear that the Bayes Lepski's procedure cannot choose J that is too small as to make the model bias unmanageable. If we were using B-splines of order q as our prior, this event is prevented by requiring that the true function be a certain distance away from the space of q -order polynomial splines. However if the truth turns out to be a polynomial splines of the same order, then it is at distance zero from this space and we are not able to quantify uncertainty honestly for this truth, despite the fact that we have correctly modelled it using B-splines. This seems to give rise to the phenomenon that the more accurately we are able to model and estimate the truth, the less we are able to access its quality and statistical uncertainty and vice versa, where such statement is reminiscence of the uncertainty principle in quantum mechanics.

The paper is structured as follows. In the next section we state some notations used in this paper. In Section 3, we introduce the Bayes Lepski's method and discuss prior and model assumptions. This is then followed by a general result on L_∞ -posterior contraction. In Section 4, we first give an exposition on the volume of tube method to bound tail probabilities in the spirit of [9], we then introduce our Bayesian credible band construction and present the main results on its coverage probability and size. Consequently, we specialize our results to B-splines (Section 5) and wavelets (Section 6). In Section 7, we discuss practical implementation and present simulation results. Proofs are in Section 8 while the Appendix (Section 9) contains some useful auxiliary results that are of independent interests.

2 Notations and Definitions

For two sequences a_n, b_n and when $n \rightarrow \infty$, we say that $a_n = O(b_n)$ or $a_n \lesssim b_n$ if $a_n \leq Cb_n$ for some constant $C > 0$ not depending on n ; and we say $a_n = o(b_n)$ if $a_n/b_n \rightarrow 0$. The real line is \mathbb{R} while positive integers are \mathbb{N} . Let us denote the $L_p(D)$ -norms of f as $\|f\|_p = (\int_D |f(x)|^p dx)^{1/p}$, with $\|f\|_\infty = \sup_{x \in D} |f(x)|$ for $p = \infty$. For the vector/series version ℓ_p -norms, we have $\|\mathbf{x}\|_p = (\sum_i |x_i|^p)^{1/p}$ and $\|\mathbf{x}\|_\infty = \max_i |x_i|$. As in standard practice, we write $\|\mathbf{x}\|_2$ simply as $\|\mathbf{x}\|$. Let $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ be the maximum and minimum eigenvalues of a matrix \mathbf{A} . The induced matrix norm is defined as $\|\mathbf{A}\|_{(p,q)} = \sup_{\mathbf{x} \neq 0} \|\mathbf{A}\mathbf{x}\|_q / \|\mathbf{x}\|_p$. Then it is known that $\|\mathbf{A}\|_{(2,2)} = \lambda_{\max}^{1/2}(\mathbf{A}^T \mathbf{A})$ is the spectral norm, $\|\mathbf{A}\|_{(\infty,\infty)} = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|$ is the max of absolute value of row sums, and $\|\mathbf{A}\|_{(1,1)} = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|$ is the max of absolute value of column sums. By Hölder's inequality, we have the relation $\|\mathbf{A}\|_{(2,2)} \leq \sqrt{\|\mathbf{A}\|_{(1,1)} \|\mathbf{A}\|_{(\infty,\infty)}}$. If \mathbf{A} is symmetric, then $\|\mathbf{A}\|_{(2,2)} = |\lambda_{\max}(\mathbf{A})|$ and $\|\mathbf{A}\|_{(1,1)} = \|\mathbf{A}\|_{(\infty,\infty)}$, which in view of the relation above gives $\|\mathbf{A}\|_{(2,2)} \leq \|\mathbf{A}\|_{(\infty,\infty)}$.

3 The Bayes Lepski's method

For $f \in L_2([0, 1])$, let us project f onto a lower J -dimensional linear subspace spanned by the basis functions $\mathbf{a}_J(x) = (a_1(x), \dots, a_J(x))^T$. These bases are compactly supported on some interval in \mathbb{R} and they are linearly independent. Note that we do not require them to be orthonormal. We can write this projection as $f(x) = \mathbf{a}_J(x)^T \boldsymbol{\theta}$ and the associated linear operator arising from such projection as K_J . We then say that the bases are v -regular if $K_J(p) = p$ when p is a polynomial of degree v or less, i.e., K_J reproduces polynomials of degree v or less.

Let us choose end points $1 \leq j_{\min} \leq j_{\max} \leq n$ such that $j_{\min} = (n/\log n)^{1/(2v+1)}$ and $j_{\max} = o(n/\log n)$. Write the candidate set as $\mathcal{J} := [j_{\min}, j_{\max}] \cap \mathbb{N}$. In what follows, we will only consider J belonging to \mathcal{J} , as the latter serves as a reservoir of admissible J for our model.

Next, we endow a Gaussian process prior on f by eliciting a normal prior on the basis coefficients $\boldsymbol{\theta} | \sigma \sim N(\boldsymbol{\eta}, \sigma^2 \boldsymbol{\Omega})$. We take $\|\boldsymbol{\eta}\|_\infty < \infty$ and select the prior covariance matrix such that for some constants $0 < c_1 \leq c_2 < \infty$,

$$c_1 \leq \lambda_{\min}(\boldsymbol{\Omega}) \leq \lambda_{\max}(\boldsymbol{\Omega}) \leq c_2. \quad (3.1)$$

Let $\mathbf{A} = (\mathbf{a}_J(X_1)^T, \dots, \mathbf{a}_J(X_n)^T)^T$ be the basis matrix constructed by evaluating the basis vector $\mathbf{a}_J(x)$ at the covariates $(X_1, \dots, X_n)^T$. Since the bases are linearly independent, the Gram matrix $\mathbf{A}^T \mathbf{A}$ is nonsingular and we choose the covariates such that for some constants $0 < C_1 \leq C_2 < \infty$,

$$C_1 \lambda_{\mathbf{A}, J} \leq \lambda_{\min}(\mathbf{A}^T \mathbf{A}) \leq \lambda_{\max}(\mathbf{A}^T \mathbf{A}) \leq C_2 \lambda_{\mathbf{A}, J} \quad (3.2)$$

as $n \rightarrow \infty$, where the common eigenvalue bound $\lambda_{\mathbf{A}, J} \rightarrow \infty$ as $n \rightarrow \infty$ for any $J \in \mathcal{J}$, and we assume it is decreasing in J for any fixed n . We further assume that $\mathbf{A}^T \mathbf{A}$ is m -banded

in the sense that $(\mathbf{A}^T \mathbf{A})_{ij} = 0$ if $|i - j| > m$ for some fixed m not depending on n . This is a reasonable assumption in view of the fact that most bases used in applications such as B-splines and Daubechies wavelets produce banded Gram matrices. In fact, bandedness arises naturally for compactly supported basis systems that are constructed by shifting and scaling a base/template function. It will be seen that the shrinkage factor $(\mathbf{A}^T \mathbf{A} + \mathbf{\Omega}^{-1})^{-1}$ will be present in both the posterior mean and variance, and to exploit the bandedness of $\mathbf{A}^T \mathbf{A}$ in our proofs and also during actual computations, we let the prior precision matrix $\mathbf{\Omega}^{-1}$ be r -banded with r possibly differing from m .

Let us denote the uniform ℓ_1 - and ℓ_2 -norms of the basis vector \mathbf{a}_J as

$$l_{1,J} := \sup_{x \in [0,1]} \|\mathbf{a}_J(x)\|_1 \quad \text{and} \quad l_{2,J} := \sup_{x \in [0,1]} \|\mathbf{a}_J(x)\|, \quad (3.3)$$

and we know that $l_{1,J} \geq l_{2,J}$. For any $J \in \mathcal{J}$, we require the common eigenvalue bound be

$$\lambda_{\mathbf{A},J} \geq l_{1,J}^2. \quad (3.4)$$

To deal with unknown σ , we will use empirical Bayes by maximizing the marginal likelihood $\mathbf{Y} | \sigma \sim \mathbf{N}[\mathbf{A}\boldsymbol{\eta}, \sigma^2(\mathbf{A}\mathbf{\Omega}\mathbf{A}^T + \mathbf{I})]$ to get

$$\hat{\sigma}_J^2 = n^{-1}(\mathbf{Y} - \mathbf{A}\boldsymbol{\eta})^T (\mathbf{A}\mathbf{\Omega}\mathbf{A}^T + \mathbf{I})^{-1} (\mathbf{Y} - \mathbf{A}\boldsymbol{\eta}). \quad (3.5)$$

For some large enough constant $\tau > 0$, we estimate the optimal number of basis J in \mathcal{J} by

$$\hat{j}_n = \min \left\{ j \in \mathcal{J} : \|\mathbf{E}_j(f|\mathbf{Y}) - \mathbf{E}_i(f|\mathbf{Y})\|_\infty \leq \tau \hat{\sigma}_i l_{1,i} \sqrt{\frac{\log i}{\lambda_{\mathbf{A},i}}}, \forall i > j, i \in \mathcal{J} \right\}. \quad (3.6)$$

Here $\mathbf{E}_J[f(x)|\mathbf{Y}]$ is the posterior mean by making its dependence on number of basis J explicit, which by conjugacy is

$$\mathbf{E}_J[f(x)|\mathbf{Y}] = \mathbf{a}_J(x)^T (\mathbf{A}^T \mathbf{A} + \mathbf{\Omega}^{-1})^{-1} (\mathbf{A}^T \mathbf{Y} + \mathbf{\Omega}^{-1} \boldsymbol{\eta}). \quad (3.7)$$

In words, our stopping rule is when the less-than inequality of (3.6) is flipped as we take successively smaller $j \in \mathcal{J}$, and set \hat{j}_n as the previous element when this occurs. Note that since the optimal J is increasing in n , we tend to stop sooner when sample size is larger. Compare this to established methods such as leave-one-out cross validation, where we have to iterate across all $J \in \mathcal{J}$ and execution time for each iteration increases exponentially with n . Although we begin with a large J , computation is still manageable due to the bandedness of $\mathbf{A}^T \mathbf{A}$ and compact support of the basis function, which has only a fixed number of nonzero values at any given x .

If the minimum is over an empty set, i.e., the sup-norm difference in posterior means is always larger than the right hand side on \mathcal{J} , we can either let $\hat{j}_n = j_{\max}$ by default or increase j_{\max} until an appropriate \hat{j}_n is chosen. On the other hand, if this difference is always less than the right hand side over all \mathcal{J} , we conclude that j_{\min} is too large and we enlarge \mathcal{J} by increasing the regularity ν of the bases. We then substitute J for \hat{j}_n in the conditional posterior distribution $\Pi_J(f|\mathbf{Y}, \sigma = \sigma_J)$, and we write this Lepski's (Gaussian) posterior simply as $\Pi(f|\mathbf{Y})$.

Remark 3.1. *The form given in (3.6) (especially the right hand side) is asymptotic in nature, and it is used to derive large sample results for credible band coverage. In finite samples, some extra considerations have to be taken into account, for example inducing undersmoothing by encouraging earlier stopping. We do not do sample splitting as we feel that it is more natural to achieve the same aim by finding a good finite sample proxy for the right hand side. More implementation issues will be discussed in Section 7 on simulations.*

To study Bayesian procedures from a frequentist viewpoint, let us assume the existence of a true regression function $f_0 : [0, 1] \rightarrow \mathbb{R}$ such that the corresponding true model has the following assumption:

Global assumption on true model:

Under the true distribution P_0 , $Y_i = f_0(X_i) + \varepsilon_i, i = 1, \dots, n$, where ε_i are i.i.d. Gaussian with mean 0 and finite variance $\sigma_0^2 > 0$ for $i = 1, \dots, n$.

3.1 Preliminary: adaptive contraction rates in L_∞

Let \mathcal{G} be a Banach ball defined through the norm $\|\cdot\|_{\mathcal{G}}$, such that for any $g \in \mathcal{G}$, $\|g\|_{\mathcal{G}} \leq R$ for some $R > 0$, and there exists a $\boldsymbol{\theta}_0 \in \mathbb{R}^J$ where for some constant $C_0 > 0$ depending on the basis used, we have

$$\|\mathbf{a}_J(\cdot)^T \boldsymbol{\theta}_0 - g\|_\infty \leq C_0 \|g\|_{\mathcal{G}} h(J), \quad (3.8)$$

with h being a nonincreasing function such that $h(j_{\min}) \rightarrow 0$ as $n \rightarrow \infty$, and $\|\boldsymbol{\theta}_0\|_\infty < \infty$. Moreover for $f_0 \in \mathcal{G}$, let us define

$$j_n^* := \min \left\{ j \in \mathcal{J} : \|f_0\|_{\mathcal{G}} h(j) \leq l_{1,j} \sqrt{\frac{\log j}{\lambda_{\mathbf{A},j}}} \right\}. \quad (3.9)$$

Here if $j \geq j_n^*$, then the posterior mean bias $\|f_0\|_{\mathcal{G}} h(j) \leq l_{1,j} \sqrt{\log(j)/\lambda_{\mathbf{A},j}}$; while this inequality will reverse when $j < j_n^*$. Therefore we deduce that j_n^* solves $\|f_0\|_{\mathcal{G}} h(j) = l_{1,j} \sqrt{\log(j)/\lambda_{\mathbf{A},j}}$ up to some universal constants not depending on n , i.e., j_n^* balances the order of the posterior mean bias and the posterior standard deviation as represented by the expression on the right hand side. In this context, the Bayes Lepski's stopping rule in (3.6) can be seen as a data driven procedure to estimate this true j_n^* .

For all results in this section, we formulate our rates in the bias-variance-tradeoff fashion without explicit reference to the smoothness scale that f_0 might belong to. A key requirement before establishing contraction rate is consistency of the empirical Bayes estimate $\hat{\sigma}_J, J \in \mathcal{J}$:

Proposition 3.2 (Consistency). *For any $J \in \mathcal{J}$, there are constants $\xi, C, Q > 0$ such that as $n \rightarrow \infty$,*

$$\inf_{f_0 \in \mathcal{G}} P_0(\sigma_0 - \xi \delta_{n,J} \leq \hat{\sigma}_J \leq \sigma_0 + \xi \delta_{n,J}) \geq 1 - C e^{-QJ},$$

where $\delta_{n,J} = \sqrt{J/n} + \|f_0\|_{\mathcal{G}} h(J) \rightarrow 0$.

Note that this is an exponential type inequality, as the usual route through mean square error and Markov's inequality prove to be inadequate for L_∞ -contraction, this is due to the need of controlling simultaneously $\widehat{\sigma}_i$ for all $i > j$ in the stopping rule of (3.6). Another key ingredient is to ensure that the Bayes Lepski procedure will not stop sooner before reaching the true j_n^* with probability tending to 1.

Proposition 3.3. *For large enough τ , there exists a constant $\mu > 2$ such that for $n \rightarrow \infty$,*

$$\sup_{f_0 \in \mathcal{G}} P_0(\widehat{j}_n > j_n^*) \lesssim \frac{1}{(j_n^*)^{\mu-2}}.$$

These two key preliminaries will then enable us to establish the following result on adaptive sup-norm posterior contraction.

Theorem 3.4. *As $n \rightarrow \infty$, there are constants $\xi, M > 0$ such that*

$$\sup_{f_0 \in \mathcal{G}} E_0 \Pi \left[\|f - f_0\|_\infty > \xi l_{1, j_n^*} \sqrt{\frac{\log j_n^*}{\lambda_{\mathcal{A}, j_n^*}}} \middle| \mathbf{Y} \right] \lesssim (j_n^*)^{-M}. \quad (3.10)$$

The rate above is in abstract form that is applicable to general basis systems that fulfil our modeling assumptions. Concrete rates will be given in Sections 5 and 6 to B-splines and wavelets respectively. Using the same techniques utilized in Corollary 4.5 of [24], we further deduce that the posterior mean as a point estimator converges to f_0 at the rate indicated above uniformly over $[0, 1]$ in P_0 -probability.

4 Adaptive credible bands through geometry

In Bayesian nonparametrics, the standard construction of credible bands is a L_∞ -ball of fixed radius around the posterior mean. However in many problems, it is more natural to allow the width to depend on the regression function's domain such that its radius is a function of $x \in [0, 1]$. This enables the bands to adapt to local characteristics of the curve to be estimated and hence is expected to give better uncertainty quantification. As an example, for some fixed $J \in \mathcal{J}$ and a given credibility level $1 - \gamma$, we can construct bands by using the posterior mean as center and some multiple of the posterior standard deviation as radius as follows:

$$\mathcal{C}_J := \left\{ f : |f(x) - E_J[f(x)|\mathbf{Y}]| \leq w_{\gamma, J} \sqrt{\text{Var}_J[f(x)|\mathbf{Y}]}, \forall x \in [0, 1] \right\},$$

and we choose the quantile $w_{\gamma, J} > 0$ such that the posterior probability of this set is at least $1 - \gamma$. For our prior formulation $f(x) = \mathbf{a}_J(x)^T \boldsymbol{\theta}$, the computation of $w_{\gamma, J}$ involves estimating

$$\Pi \left[\sup_{x \in [0, 1]} \left| \frac{\mathbf{a}_J(x)^T [\boldsymbol{\theta} - E_J(\boldsymbol{\theta}|\mathbf{Y})]}{\sqrt{\mathbf{a}_J(x)^T \text{Var}_J(\boldsymbol{\theta}|\mathbf{Y}) \mathbf{a}_J(x)}} \right| > w_{\gamma, J} \middle| \mathbf{Y} \right]. \quad (4.1)$$

Now if we have a sharp upper bound $U(w_{\gamma,J})$ to this tail probability, then we find $w_{\gamma,J}$ such that it solves $U(w_{\gamma,J}) = \gamma$.

With this choice, let us now access the coverage (in the frequentist sense) of the Bayesian credible band \mathcal{C}_J by estimating $P_0(f_0 \notin \mathcal{C}_J)$. By selecting $J \in \mathcal{J}$ in an optimal fashion (e.g., through the Bayes Lepski's method), it will be seen that this quantity is bounded above up to some negligible terms by another tail probability of the form

$$P_0 \left[\sup_{x \in [0,1]} \left| \frac{\mathbf{a}_J(x)^T \{E_J(\boldsymbol{\theta}|\mathbf{Y}) - E_0 E_J(\boldsymbol{\theta}|\mathbf{Y})\}}{\sqrt{\mathbf{a}_J(x)^T \text{Var}_0 E_J(\boldsymbol{\theta}|\mathbf{Y}) \mathbf{a}_J(x)}} \right| > w_{\gamma,J} \right]. \quad (4.2)$$

Again if we have a sharp upper bound $G(w_{\gamma,J})$ to the above and establish that $G(w_{\gamma,J}) \leq U(w_{\gamma,J}) \leq \gamma$, then Bayesian credible band \mathcal{C}_J of credibility at least $1 - \gamma$ corresponds to a frequentist confidence band of confidence level at least $1 - \gamma$. Therefore to achieve this almost exact correspondence between Bayesian and frequentist quantification of uncertainty, it is then imperative that we are able to get sharp upper bounds for tail probabilities of “standardized” Gaussian processes in supremum norm.

4.1 Tail probabilities and the volume-of-tube formula

In this subsection, let us temporarily leave the regression framework and work with general Gaussian processes generated through a random series. We will divide our ordeal into 3 steps and will revisit the issue of Bayesian credible band construction in the next subsection.

(i) *Recast statistical problem for geometric formulation*

Let $\mathbf{X} \sim N_J(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and define

$$W := \sup_{x \in [0,1]} \frac{\mathbf{a}_J(x)^T (\mathbf{X} - \boldsymbol{\mu})}{\sqrt{\mathbf{a}_J(x)^T \boldsymbol{\Sigma} \mathbf{a}_J(x)}}.$$

Our task is then to estimate as accurately as possible $P(W > w)$, which with appropriate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ corresponds to the two tail probabilities mentioned above in (4.1) and (4.2) (since $P(|W| > w) \leq 2P(W > w)$). We further decompose $W = QR$ such that $R^2 = (\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})$ and

$$\begin{aligned} Q &= \sup_{x \in [0,1]} \frac{\mathbf{a}_J(x)^T (\mathbf{X} - \boldsymbol{\mu})}{\sqrt{\mathbf{a}_J(x)^T \boldsymbol{\Sigma} \mathbf{a}_J(x)}} \frac{1}{\sqrt{(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})}} \\ &= \sup_{x \in [0,1]} \underbrace{\frac{[\boldsymbol{\Sigma}^{1/2} \mathbf{a}_J(x)]^T}{\|\boldsymbol{\Sigma}^{1/2} \mathbf{a}_J(x)\|}}_{\beta_J(x)} \underbrace{\frac{\boldsymbol{\Sigma}^{-1/2} (\mathbf{X} - \boldsymbol{\mu})}{\|\boldsymbol{\Sigma}^{-1/2} (\mathbf{X} - \boldsymbol{\mu})\|}}_U =: \sup_{x \in [0,1]} \langle \beta_J(x), U \rangle. \end{aligned}$$

Since $\boldsymbol{\Sigma}^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) \sim N_J(\mathbf{0}, \mathbf{I})$, it follows that U is uniformly distributed on the unit sphere S^{J-1} ; while β_J is a curve that maps $[0, 1]$ to S^{J-1} . Observe that Q is independent of R , with

R^2 distributed as chi-square with J degrees of freedom χ_J^2 . Therefore,

$$P(W > w) = \int_w^\infty P\left(\sup_{x \in [0,1]} \langle \beta_J(x), U \rangle > wr^{-1}\right) p_{\chi_J^2}(r) dr, \quad (4.3)$$

where $p_{\chi_J^2}$ is the density function of χ_J^2 , and we compute the tail probability of W by using chi-square mixtures. Thus, our task is to evaluate the probability in the integrand and this is where concepts from geometry comes into play.

(ii) *Tubular neighborhoods and the geometric connection*

Let u be some point on the unit sphere S^{J-1} , and we define a metric d to measure distance between u and the curve β_J by finding the nearest point on β_J to u in ℓ_2 -norm, such that $d(u, \beta_J)^2 = \inf_{x \in [0,1]} \|u - \beta_J(x)\|^2$. Since $\|u\|^2 = 1 = \|\beta_J(x)\|^2$ for any x , this squared distance can also be written as $2(1 - \sup_{x \in [0,1]} \langle \beta_J(x), u \rangle)$. Keeping this relation in mind, let us define tubular neighborhoods parameterized by θ around β_J as

$$\mathcal{T}_\theta = \left\{ u \in S^{J-1} : d(u, \beta_J) \leq [2(1 - \cos \theta)]^{1/2} \right\} = \left\{ u \in S^{J-1} : \sup_{x \in [0,1]} \langle \beta_J(x), u \rangle \geq \cos \theta \right\},$$

and we call θ the geodesic or angular radius that controls the (geodesic) width of these tubes. To visualize these objects, see Figure 1.

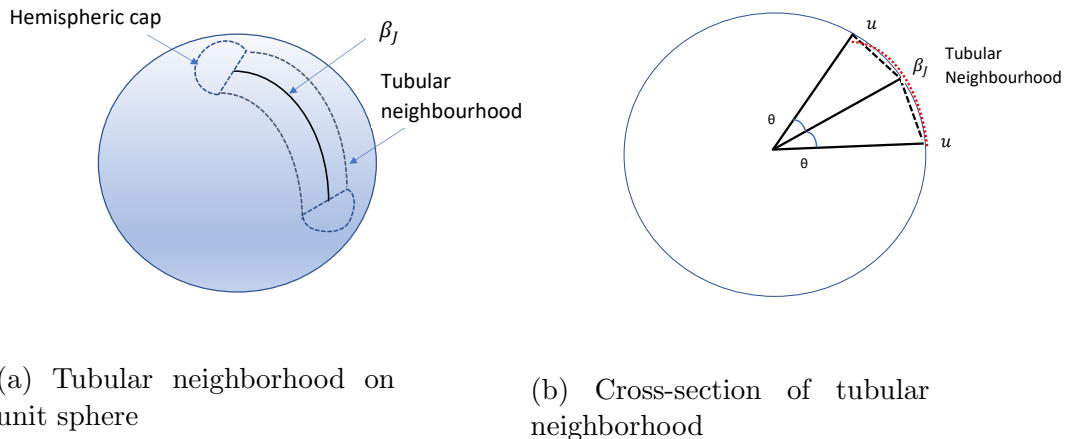


Figure 1: Tubular neighborhood around curve β_J

For a set A in \mathbb{R}^J , write $V(A)$ to be the volume of A with respect to the Lebesgue measure. Now recall that $V(S^{J-1}) = 2\pi^{J/2}/\Gamma(J/2)$. Let U be a random variable uniformly distributed on S^{J-1} (such as the U in (4.3) above), then

$$P\left(\sup_{x \in [0,1]} \langle \beta_J(x), U \rangle \geq \cos \theta\right) = P(U \in \mathcal{T}_\theta) = \frac{V(\mathcal{T}_\theta)}{V(S^{J-1})} = \frac{\Gamma(J/2)}{2\pi^{J/2}} V(\mathcal{T}_\theta), \quad (4.4)$$

and what remains is to find the volume of the tube \mathcal{T}_θ .

(iii) *Volume-of-tube formula*

The volume of \mathcal{T}_θ depends on the shape the tube can possibly take, and this in turn is affected by 3 situations: β_J is closed (a loop), or open with endpoints at 0 and 1, or that the tube \mathcal{T}_θ self overlaps. Among these circumstances, the maximal volume is achieved when β_J is open with boundaries. In that case, the enveloping tube has 3 pieces (see Figure 1a), the main “cylinder” around the curve whose volume is the curve’s length $|\beta_J|$ times its cross-sectional area, and two hemispheric caps at the end points. The exact formula for this upper bound was calculated by [15]:

Theorem 4.1 (Naiman). *Let $\beta_J : [0, 1] \rightarrow S^{J-1}$ be a continuously differentiable curve, with nowhere vanishing first derivative, and of finite length. Then we have for any $0 \leq \theta \leq 2\pi$,*

$$V(\mathcal{T}_\theta) \leq |\beta_J|V(\mathcal{B}_{J-2}) \sin^{J-2} \theta + V(S^{J-2}) \int_{\cos \theta}^1 (1 - z^2)^{(J-3)/2} dz,$$

where $|\beta_J|$ is the arc length of β_J and \mathcal{B}_{J-2} is the unit ball in \mathbb{R}^{J-2} .

This volume inequality is sharp, in fact for open β_J , the above will be an equality when θ is smaller than some critical radius, which is the threshold for tube self-overlap. Since $V(\mathcal{B}_{J-2}) = \pi^{(J-2)/2}/\Gamma(J/2)$ and $V(S^{J-2}) = 2\pi^{(J-1)/2}/\Gamma[(J-1)/2]$, we can perform a change of variable $u = z^2$ in the integral above and write (4.4) as

$$P \left(\sup_{x \in [0,1]} \langle \beta_J(x), U \rangle \geq \cos \theta \right) \leq \frac{|\beta_J|}{2\pi} \sin^{J-2} \theta + \frac{\Gamma(J/2)}{2\pi^{1/2} \Gamma[(J-1)/2]} P \left[B \left(\frac{1}{2}, \frac{J-1}{2} \right) \geq \cos^2 \theta \right]$$

where $B(\alpha_1, \alpha_2)$ is a beta distributed random variable with shape and scale parameters α_1 and α_2 respectively. Choose θ such that $\cos \theta = wr^{-1}$ and note that $\sin \theta = (1 - \cos^2 \theta)^{1/2} = (1 - w^2 r^{-2})^{1/2}$, if we plug this bound back to (4.3) and compute the chi-square mixtures, we will arrive at

$$P \left(\sup_{x \in [0,1]} \frac{\mathbf{a}_J(x)^T (\mathbf{X} - \boldsymbol{\mu})}{\sqrt{\mathbf{a}_J(x)^T \boldsymbol{\Sigma} \mathbf{a}_J(x)}} > w \right) \leq \frac{|\beta_J|}{2\pi} e^{-w^2/2} + 1 - \Phi(w), \quad (4.5)$$

with $|\beta_J| = \int_0^1 \|\beta'_J(x)\| dx$ for $\beta_J(x) = \boldsymbol{\Sigma}^{1/2} \mathbf{a}_J(x) / \|\boldsymbol{\Sigma}^{1/2} \mathbf{a}_J(x)\|$ and $\Phi(\cdot)$ is the cumulative distribution function of a standard normal.

4.2 Variable width Bayesian credible bands

Let us return to our variable width credible band in the beginning. For $J \in \mathcal{J}$, we construct

$$\mathcal{C}_J := \left\{ f : |f(x) - \mathbb{E}_J[f(x)|\mathbf{Y}]| \leq w_{\gamma,J} \sqrt{\text{Var}_J[f(x)|\mathbf{Y}]}, \forall x \in [0, 1] \right\}.$$

Motivated by the previous exposition on volume of tubes, for any $J \in \mathcal{J}$ and a given credibility level $1 - \gamma$, we choose $w_{\gamma,J}$ such that it solves

$$\gamma = \frac{|\beta_J|}{\pi} e^{-w_{\gamma,J}^2/2} + 2[1 - \Phi(w_{\gamma,J})], \quad (4.6)$$

where the arc length $|\beta_J| = \int_0^1 \|\beta'_J(x)\| dx$ under our hierarchical Bayes prior is

$$\begin{aligned} |\beta_J| &= \int_0^1 \left\{ \frac{d}{dx} \left(\frac{\mathbf{M}^{1/2} \mathbf{a}_J(x)^T}{[\mathbf{a}_J(x)^T \mathbf{M} \mathbf{a}_J(x)^T]^{1/2}} \right) \frac{d}{dx} \left(\frac{\mathbf{M}^{1/2} \mathbf{a}_J(x)}{[\mathbf{a}_J(x)^T \mathbf{M} \mathbf{a}_J(x)^T]^{1/2}} \right) \right\}^{1/2} dx \\ &= \int_0^1 \frac{[\mathbf{a}_J(x)^T \mathbf{M} \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{M} \mathbf{a}_J(x)]^{1/2}}{[\mathbf{a}_J(x)^T \mathbf{M} \mathbf{a}_J(x)]^{3/2}} dx, \end{aligned} \quad (4.7)$$

for $\mathbf{M} := (\mathbf{A}^T \mathbf{A} + \mathbf{\Omega}^{-1})^{-1}$ and $\mathbf{R} := \dot{\mathbf{a}}_J(x) \mathbf{a}_J(x)^T - \mathbf{a}_J(x) \dot{\mathbf{a}}_J(x)^T$ where we denote $\dot{\mathbf{a}}_J(x) = (a'_1(x), \dots, a'_J(x))^T$. To complete our construction, we choose $J = \hat{j}_n$ by the Bayes Lepski's stopping rule given in (3.6) to give a $1 - \gamma$ adaptive simultaneous credible band $\mathcal{C}_{\hat{j}_n}$.

Now by setting $\boldsymbol{\mu} = \mathbb{E}_{\hat{j}_n}(\boldsymbol{\theta} | \mathbf{Y})$ and $\boldsymbol{\Sigma} = \text{Var}_{\hat{j}_n}(\boldsymbol{\theta} | \mathbf{Y})$, we can use the tail estimate of (4.5) to bound (4.1) and ensure that our credible band $\mathcal{C}_{\hat{j}_n}$ has at least $1 - \gamma$ credibility:

$$\Pi \left[\sup_{0 \leq x \leq 1} \frac{|f(x) - \mathbb{E}_{\hat{j}_n}[f(x) | \mathbf{Y}]|}{\sqrt{\text{Var}_{\hat{j}_n}[f(x) | \mathbf{Y}]}} \geq w_{\gamma, \hat{j}_n} \mid \mathbf{Y} \right] \leq \frac{|\beta_{\hat{j}_n}|}{\pi} e^{-w_{\gamma, \hat{j}_n}^2/2} + 2[1 - \Phi(w_{\gamma, \hat{j}_n})] = \gamma.$$

Is $\mathcal{C}_{\hat{j}_n}$ also a confidence band of level at least $1 - \gamma$? The theorem below answers this in the affirmative. Let us define another curve $\beta_{0,J} : [0, 1] \rightarrow S^{J-1}$ such that its arc length has the same functional form as in (4.7) but with \mathbf{M} replaced by $\mathbf{M}_0 := (\mathbf{A}^T \mathbf{A} + \mathbf{\Omega}^{-1})^{-1} \mathbf{A}^T \mathbf{A} (\mathbf{A}^T \mathbf{A} + \mathbf{\Omega}^{-1})^{-1}$.

Theorem 4.2 (Credible band coverage). *Suppose the following 3 assumptions hold ($n \rightarrow \infty$):*

1. *There exists a set \mathcal{F} such that $\inf_{f_0 \in \mathcal{G} \cap \mathcal{F}} P_0(\hat{j}_n \geq j_n^*/\kappa) \rightarrow 1$ for some constant $\kappa \geq 1$.*
2. *The arc length $|\beta_J| \geq \max\{|\beta_{0,J}|, C J\}$ for any $J \in \mathcal{J}$ and some constant $C > 0$.*
3. *$(\log \hat{j}_n)^{-1/2} \left(\inf_{x \in [0,1]} \|\mathbf{a}_{\hat{j}_n}(x)\| \right)^{-1} \left[l_{1, \hat{j}_n} \lambda_{\mathbf{A}, \hat{j}_n}^{-1/2} + \lambda_{\mathbf{A}, \hat{j}_n}^{1/2} h(\hat{j}_n) \right] = o_{P_0}(1)$ under the event $\{j_n^*/\kappa \leq \hat{j}_n \leq j_n^*\}$.*

Then as $n \rightarrow \infty$, we have

$$\inf_{f_0 \in \mathcal{G} \cap \mathcal{F}} P_0 \left(f_0 \in \mathcal{C}_{\hat{j}_n} \right) \geq 1 - \gamma + o(1).$$

The inequality in the coverage statement dictates that this is a conservative procedure. The 3 conditions are satisfied for the most commonly used basis systems and we verify them using B-spline and wavelets in the next two sections. The first condition says that

the Bayes Lepski's rule cannot stop too late and choose J that is too small as to make the approximation bias unmanageable. As we will discuss in detail later in Section 5.1, this give rise to the counterintuitive notion that we need a certain amount of error in order to quantify uncertainty honestly. The second condition says that the arc length under our model/posterior must be longer than the one under the true distribution, so that credible level under the posterior can be translated to confidence level when coverage is measured. At the same time, the model arc length must be larger (up to some constant) than $J \in \mathcal{J}$ so that in the last condition, the size of the quantile $w_{\gamma, J}$ is of the order larger than the standardized approximation bias, where the seemingly technical appearance is the result of reducing the aforementioned condition to one formulated by the basis components only.

Therefore, if the Bayes Lepski rule do not stop too early or too late, we have the following result on the size of the credible band generated through the volume of tube method.

Corollary 4.3 (Length of radius function). *Let $r(x) := w_{\gamma, \hat{j}_n} \sqrt{\text{Var}_{\hat{j}_n}[f(x)|\mathbf{Y}]}$ be the radius function for the variable width credible band. Let assumption 1 of Theorem 4.2 hold for some set \mathcal{F} and constant $\kappa \geq 1$ and assume in addition that $|\beta_J| \asymp J$ for any $J \in \mathcal{J}$. Then as $n \rightarrow \infty$,*

$$\inf_{f_0 \in \mathcal{G} \cap \mathcal{F}} P_0 \left[\left\| \mathbf{a}_{j_n^*/\kappa}(x) \right\| \sqrt{\frac{\log(j_n^*)}{\lambda_{\mathbf{A}, j_n^*/\kappa}}} \lesssim r(x) \lesssim \left\| \mathbf{a}_{j_n^*}(x) \right\| \sqrt{\frac{\log(j_n^*)}{\lambda_{\mathbf{A}, j_n^*}}}, \forall x \in [0, 1] \right] \rightarrow 1.$$

Since κ is a constant not depending on n , the upper and lower bounds will have the same asymptotic order, this further implies that the radius cannot concentrate to quickly to the 0 function, and it must do so at least at a certain rate indicated above.

To showcase the applicability of the methods discussed, let us consider two special but widely used classes of basis functions, with the B-splines as the non-orthonormal example and the wavelets serving as the canonical orthonormal example.

5 Application I: B-splines bases

Suppose we project f onto the J -dimensional space of polynomial splines spanned by B-splines basis functions $B_{j,q}(x), j = 1, \dots, J$. Each of this basis is a piecewise polynomial defined on the knot sequence $\mathcal{T} = \{0 = t_0 < t_1, \dots, t_N < t_{N+1} = 1\}$, such that $B_{j,q}$ restricted to a knot interval is a polynomial of order q (or degree $q - 1$), and $B_{j,q}$ is $q - 2$ times differentiable at the knot points. We select the knots such that their distribution is quasi-uniform, in the sense that $\max_{1 \leq k \leq N+1} (t_k - t_{k-1}) \lesssim \min_{1 \leq k \leq N+1} (t_k - t_{k-1})$, i.e., the max knot increment is of the same order as the min knot increment. Here N is the number of interior knots and it is related to the number of basis by $J = N + q$.

Let $\mathbf{b}_{J,q}(x) = (B_{1,q}(x), \dots, B_{J,q}(x))^T$. Our hierarchical prior can then be formulated as $f(x) = \mathbf{b}_{J,q}(x)^T \boldsymbol{\theta}$ with $\boldsymbol{\theta} | \sigma \sim N_J(\boldsymbol{\eta}, \sigma^2 \boldsymbol{\Omega})$ and σ^2 estimated by empirical Bayes (3.5). Construct the B-spline basis matrix as $\mathbf{B} = (\mathbf{b}_{J,q}(X_1)^T, \dots, \mathbf{b}_{J,q}(X_n)^T)^T$. By Theorem 4.18 of [18], we know that B-splines are linearly independent and hence $\mathbf{B}^T \mathbf{B}$ is invertible. For a

fixed cumulative distribution function $F(x)$ with positive and continuous density on $[0, 1]$, let us choose the covariates such that

$$\sup_{x \in [0,1]} |F_n(x) - F(x)| = o(N^{-1}),$$

where $F_n(x) = n^{-1} \sum_{i=1}^n \mathbb{1}_{[0, X_i]}(x)$ is the empirical distribution of (X_1, \dots, X_n) . Then by Lemma A.9 of [23], there exist constants $0 < C_1 \leq C_2 < \infty$ such that

$$C_1(n/J) \leq \lambda_{\min}(\mathbf{B}^T \mathbf{B}) \leq \lambda_{\max}(\mathbf{B}^T \mathbf{B}) \leq C_2(n/J), \quad (5.1)$$

and hence $\lambda_{\mathbf{B}, J} = n/J$ in this case. Now since $\sum_{j=1}^J B_{j,q}(x) = 1$ (partition of unity) and $0 \leq B_{j,q}(x) \leq 1$ for any j and x , it follows that $l_{1,J} = 1$. Therefore if $J \leq n$, then $\lambda_{\mathbf{B}, J} \geq l_{1,J}^2$. For any $X_i \in [0, 1]$, we can always find a k such that $X_i \in [t_{k-1}, t_k]$, and since B-splines have compact support, this implies that only q -consecutive $B_{k,q}(X_i), \dots, B_{k+q-1,q}(X_i)$ are nonzero. Therefore if $|u - v| > q$, then $B_{u,q}(X_i)B_{v,q}(X_i) = 0$ and hence $(\mathbf{B}^T \mathbf{B})_{u,v} = 0$. As a result, $\mathbf{B}^T \mathbf{B}$ is q -banded.

By Theorem 6.18 of [18], there is a linear operator K_J that maps f onto the space of J -dimensional polynomial splines, and K_J reproduces polynomials of order q . Hence we choose the lower point in the candidate set as $j_{\min} = (n/\log n)^{1/(2q+1)}$, and this further implies that we adapt α up to q . Thus, we estimate the optimal number of B-splines bases as (for some large enough τ)

$$\hat{j}_n = \min \left\{ j \in \mathcal{J} : \|E_j(f|\mathbf{Y}) - E_i(f|\mathbf{Y})\|_{\infty} \leq \tau \hat{\sigma}_i \sqrt{\frac{i \log i}{n}}, \forall i > j, i \in \mathcal{J} \right\}.$$

Let \mathcal{H}^{α} denote the Hölder space of order $\alpha > 0$ consisting of functions $g : [0, 1] \rightarrow \mathbb{R}$ such that $\|g\|_{\mathcal{H}^{\alpha}} < \infty$, where the Hölder norm is defined as

$$\|g\|_{\mathcal{H}^{\alpha}} = \max_{0 \leq k < [\alpha]} \sup_{x \in [0,1]} |g^{(k)}(x)| + \sup_{x, y \in [0,1]: x \neq y} \frac{|g^{([\alpha])}(x) - g^{([\alpha])}(y)|}{|x - y|^{\alpha - [\alpha]}},$$

and $[\alpha]$ is the integer part of α . We then take $\mathcal{G} = \{g : \|g\|_{\mathcal{H}^{\alpha}} \leq R\} =: \mathcal{H}^{\alpha}(R)$ to be the Hölder ball of radius $R > 0$. By Theorem 22 in Chapter XII of [3] or Theorem 6.31 of [18], we know that for any $g \in \mathcal{H}^{\alpha}(R)$, there exists a $\boldsymbol{\theta}_0 \in \mathbb{R}^J$ and a constant $C_q > 0$ depending on q such that

$$\|g - \mathbf{b}_{J,q}(\cdot)^T \boldsymbol{\theta}_0\|_{\infty} \leq C_q \|g\|_{\mathcal{H}^{\alpha}} J^{-\alpha}. \quad (5.2)$$

Thus $h(x) = x^{-\alpha}$ for the B-splines case and it follows that $h(j_{\min}) = (\log n/n)^{\alpha/(2q+1)} \rightarrow 0$ as $n \rightarrow \infty$. By (7.2) of [23], we have $\|\boldsymbol{\theta}_0\|_{\infty} < \infty$. Suppose the true regression function $f_0 \in \mathcal{H}^{\alpha}(R)$ for some unknown $\alpha > 0$. Then j_n^* defined in (3.9) balances the posterior mean bias $\|f_0\|_{\mathcal{H}^{\alpha}} J^{-\alpha}$ on one side with the posterior standard deviation $\sqrt{J \log(J)/n}$ on the

other, and we deduce that $j_n^* \asymp \|f_0\|_{\mathcal{H}^\alpha}^{2/(2\alpha+1)} (n/\log n)^{1/(2\alpha+1)}$. Therefore Theorem 3.4 when specializing to B-splines bases yield

$$\sup_{0 < \alpha \leq q} \sup_{f_0 \in \mathcal{H}^\alpha(R)} \mathbb{E}_0 \Pi \left[\|f - f_0\|_\infty > \xi (\log n/n)^{\alpha/(2\alpha+1)} | \mathbf{Y} \right] \lesssim (\log n/n)^M \quad (5.3)$$

for some constants $R, \xi, M > 0$. When compared to Theorem 4.4 of [23], our result is fully rate adaptive and the ξ appearing in the radius is a constant rather than some arbitrary sequence going to infinity.

We now turn to the problem of constructing credible bands. Using the volume-of-tube method discussed in the previous section, we construct

$$\mathcal{B}_n := \left\{ f : |f(x) - \mathbb{E}_{\hat{j}_n}[f(x)|\mathbf{Y}]| \leq w_{\gamma, \hat{j}_n} \sqrt{\text{Var}_{\hat{j}_n}[f(x)|\mathbf{Y}]}, \forall x \in [0, 1] \right\}, \quad (5.4)$$

where the posterior mean and variance with the Bayes Lepski's stopping rule $J = \hat{j}_n$ is

$$\mathbb{E}_{\hat{j}_n}[f(x)|\mathbf{Y}] = \mathbf{b}_{\hat{j}_n, q}(x)^T (\mathbf{B}^T \mathbf{B} + \mathbf{\Omega}^{-1})^{-1} (\mathbf{B}^T \mathbf{Y} + \mathbf{\Omega}^{-1} \boldsymbol{\eta}), \quad (5.5)$$

$$\text{Var}_{\hat{j}_n}[f(x)|\mathbf{Y}] = \hat{\sigma}_{\hat{j}_n}^2 \mathbf{b}_{\hat{j}_n, q}(x)^T (\mathbf{B}^T \mathbf{B} + \mathbf{\Omega}^{-1})^{-1} \mathbf{b}_{\hat{j}_n, q}(x), \quad (5.6)$$

and we choose w_{γ, \hat{j}_n} such that it solves $\gamma = |\beta_{\hat{j}_n}| \pi^{-1} e^{-w_{\gamma, \hat{j}_n}^2/2} + 2[1 - \Phi(w_{\gamma, \hat{j}_n})]$. The expression for $|\beta_{\hat{j}_n}|$ can be found by substituting $\mathbf{b}_{\hat{j}_n, q}(x)$ for $\mathbf{a}_{\hat{j}_n}(x)$ and \mathbf{B} for \mathbf{A} in (4.7). For B-splines however, we can further use (8) of Chapter X in [3] to write explicitly $d\mathbf{b}_{\hat{j}_n, q}(x)/dx$ and this in turn gives

$$|\beta_{\hat{j}_n}| = \int_0^1 \frac{[\mathbf{b}_{\hat{j}_n, q}(x)^T \mathbf{M} \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{M} \mathbf{b}_{\hat{j}_n, q}(x)]^{1/2}}{[\mathbf{b}_{\hat{j}_n, q}(x)^T \mathbf{M} \mathbf{b}_{\hat{j}_n, q}(x)]^{3/2}} dx \quad (5.7)$$

for $\mathbf{M} := (\mathbf{B}^T \mathbf{B} + \mathbf{\Omega}^{-1})^{-1}$ and $\mathbf{R} := \mathbf{W} \mathbf{b}_{\hat{j}_n, q-1}(x) \mathbf{b}_{\hat{j}_n, q}(x)^T - \mathbf{b}_{\hat{j}_n, q}(x) \mathbf{b}_{\hat{j}_n, q-1}(x)^T \mathbf{W}^T$, where \mathbf{W} , a matrix of dimension $\hat{j}_n \times (\hat{j}_n - 1)$, is $q - 1$ times

$$\begin{pmatrix} -(t_1 - t_{2-q})^{-1} & 0 & 0 & \cdots & 0 & 0 \\ (t_1 - t_{2-q})^{-1} & -(t_2 - t_{3-q})^{-1} & 0 & \cdots & 0 & 0 \\ 0 & (t_2 - t_{3-q})^{-1} & -(t_3 - t_{4-q})^{-1} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & (t_{\hat{j}_n-1} - t_{\hat{j}_n-q})^{-1} \end{pmatrix}. \quad (5.8)$$

To apply Theorem 4.2 and conclude that \mathcal{B}_n has high coverage probability in the frequentist sense, we proceed to verify the 3 conditions needed.

Let us first verify the last condition, which dictates that the quantile w_{γ, \hat{j}_n} is of the order larger than the standardized posterior mean bias. Now by Lemma 9.6, we will have $\inf_{x \in [0, 1]} \|\mathbf{b}_{\hat{j}_n, q}(x)\| \geq q^{-1}$. Note that under the event $\{j_n^*/\kappa \leq \hat{j}_n \leq j_n^*\}$ with $\kappa \geq 1$ coming

from the first assumption, we deduce that $(\widehat{j}_n/n)^{1/2} \asymp \widehat{j}_n^{-\alpha}$ and $\widehat{j}/n = o_{P_0}(1)$. Thus if we intersect with the aforementioned event, the third condition reads

$$(\log \widehat{j}_n)^{-1/2} q \left[\sqrt{\frac{\widehat{j}_n}{n}} + \sqrt{\frac{n}{\widehat{j}_n} \widehat{j}_n^{-\alpha}} \right] \lesssim (\log n)^{-1/2} (1 + o_{P_0}(1)) = o_{P_0}(1),$$

as $n \rightarrow \infty$. Now for the second arc length condition, Lemmas 5.1 and 5.2 below jointly say that B-splines do indeed satisfy the arc length inequality for any $J \in \mathcal{J}$ as $n \rightarrow \infty$.

Lemma 5.1. *Let $|\beta_J| = \int_0^1 \|\beta'_J(x)\| dx$ be the arc length of the curve $\beta_J : [0, 1] \rightarrow S^{J-1}$ given in (5.7). Then as $n \rightarrow \infty$,*

$$|\beta_J| \asymp J, \quad \text{for any } J \in \mathcal{J}.$$

Lemma 5.2. *There exist constants $\xi_1, \xi_2, \xi_3 > 0$ such that*

$$|\beta_{0,J}| \leq \frac{[1 - \xi_1(J/n) + \xi_2(J/n)^2]^{1/2}}{[1 - \xi_3(J/n)]^{3/2}} |\beta_J|.$$

Consequently for any $J \in \mathcal{J}$, $|\beta_{0,J}| \leq |\beta_J|$ as $n \rightarrow \infty$.

It now remains to verify the first condition dictating a lower bound for the Bayes Lepski's estimate \widehat{j}_n . To this end, let us describe the set \mathcal{F} suited for this purpose. First, some notations. Let δ_q be some positive constant depending only on q and \mathcal{T} a quasi-uniform partition of $[0, 1]$. Denote $\Delta_{\mathcal{T}}$ to be the maximum knot increment associated with this partition \mathcal{T} . Furthermore, let \mathbb{T} be the collection of all quasi-uniform partitions of $[0, 1]$. We then take our set \mathcal{F} to be

$$\mathcal{F}^\alpha = \bigcap_{\mathcal{T} \in \mathbb{T}} \left\{ f : \inf_{p \in \mathcal{P}_q(\mathcal{T})} \|f - p\|_\infty \geq \delta_q \|f\|_{\mathcal{H}^\alpha} \Delta_{\mathcal{T}}^\alpha \right\}, \quad (5.9)$$

where $0 < \alpha \leq q$ and $\mathcal{P}_q(\mathcal{T})$ is the space of polynomial splines of order q with knots in \mathcal{T} . This set together with the Hölder ball $\mathcal{H}^\alpha(R)$ then make it possible to establish the following lemma and at the same time fulfils the first lower bound condition.

Lemma 5.3. *Let $\kappa \in \mathbb{N}$ be some constant that depends on δ_q, α and τ where τ is a large enough positive constant contained in the stopping rule \widehat{j}_n of (3.6). Then for any $R > 0$, there exist constants $\mu > 1$ and $Q > 0$ such that*

$$\sup_{0 < \alpha \leq q} \sup_{f_0 \in \mathcal{F}^\alpha} P_0(\widehat{j}_n < j_n^*/\kappa) \lesssim \frac{1}{j_{\min}^{\mu-1}} + j_n^* e^{-Qj_n^*},$$

and consequently in view of Proposition 3.3, we have as $n \rightarrow \infty$,

$$\inf_{0 < \alpha \leq q} \inf_{f_0 \in \mathcal{H}^\alpha(R) \cap \mathcal{F}^\alpha} P_0(j_n^*/\kappa \leq \widehat{j}_n \leq j_n^*) \rightarrow 1. \quad (5.10)$$

As a result, we can appeal to Theorem 4.2 and conclude that for any $R > 0$,

$$\inf_{0 < \alpha \leq q} \inf_{f_0 \in \mathcal{H}^\alpha(R) \cap \mathcal{F}^\alpha} P_0(f_0 \in \mathcal{B}_n) \geq 1 - \gamma + o(1). \quad (5.11)$$

By Lemmas 5.1, 5.3 and 9.6, we have the following B-splines version of Corollary 4.3 on the size of the radius function

$$\inf_{0 < \alpha \leq q} \inf_{f_0 \in \mathcal{H}^\alpha(R) \cap \mathcal{F}^\alpha} P_0[r(x) \asymp (\log n/n)^{\alpha/(2\alpha+1)}, \forall x \in [0, 1]] \rightarrow 1. \quad (5.12)$$

We note that the uniform length of the radius is exactly equal up to some constant to the adaptive posterior contraction rate in (5.3), and we say that this Bayesian uncertainty quantification procedure is honest.

5.1 The uncertainty principle and self-similar functions

In this subsection, we will investigate the function space \mathcal{F}^α in order to gain a deeper understanding as to its nature and its implication towards statistical estimation and uncertainty quantification. In a quick glance, the restriction imposed on \mathcal{F}^α involves a lower bound on the approximation power of polynomial splines, and it further suggests that we can only quantify uncertainty for functions that are at least a certain distance away from the space of splines.

Now if we were using posterior based on random series of order q B-splines, and the true regression function f_0 turns out to be a polynomial spline of the same order. Then we “hit the right spot” and our posterior will concentrate around f_0 at the optimal rate $(\log n/n)^{\alpha/(2\alpha+1)}$, given that we have correctly modelled the truth. However for this truth f_0 , we have $\inf_{p \in \mathcal{P}_q(\mathcal{T})} \|f_0 - p\|_\infty = 0$ and there is nothing to prevent \hat{j}_n to stop too late and choose a $J \in \mathcal{J}$ that is too small. Consequently, we will not be able to quantify uncertainty honestly owing to the fact that the posterior mean bias has a much larger order than the posterior standard deviation. This example, although extreme, serves to shed light on a fundamental limitation in statistics. *The more accurately we are able to model and estimate the truth, the less we are able to access its quality and statistical uncertainty and vice versa.* Limitation of this type is reminiscence to the uncertainty principle in quantum mechanics, where it states that the more precisely we know the position of a particle, the less precisely its momentum can be determined and vice versa.

If $f_0 \in \mathcal{F}^\alpha$, we know that by Theorem 22 in Chapter XII of [3] or Theorem 6.31 of [18], there is a $p_0 \in \mathcal{P}_q(\mathcal{T})$ and some constant $C_q > 0$ depending only on f_0 and q such that

$$\delta_q \|f_0\|_{\mathcal{H}^\alpha} \Delta_{\mathcal{T}}^\alpha \leq \inf_{p \in \mathcal{P}_q(\mathcal{T})} \|f_0 - p\|_\infty \leq \|f_0 - p_0\|_\infty \leq C_q \|f_0\|_{\mathcal{H}^\alpha} \Delta_{\mathcal{T}}^\alpha. \quad (5.13)$$

In this formulation, condition (5.9) can be interpreted as being self-similar in regularity across different levels of quasi-uniform knot partition $\mathcal{T} \in \mathbb{T}$, or in other words, functions that have the same approximation power by splines across different quasi-uniform knot partitions of $[0, 1]$. In fact, it is even possible to redefine (5.13) in a recursive version so as to remove

its dependence on α . In view of Lemma 6.17 of [18], we can construct a sequence of quasi-uniform knot partitions $\mathcal{T}_1 \subset \mathcal{T}_2 \subset \dots \subset \mathcal{T}_M$ such that $\Delta_{\mathcal{T}_{k+1}}/2 \leq \Delta_{\mathcal{T}_k} \leq 3\Delta_{\mathcal{T}_{k+1}}/2$. Thus in view (5.13), it holds that

$$\begin{aligned} \inf_{p \in \mathcal{P}_q(\mathcal{T}_{k+1})} \|f_0 - p\|_\infty &\geq \delta_q \|f_0\|_{\mathcal{H}^\alpha} \Delta_{\mathcal{T}_{k+1}}^\alpha \geq \frac{\delta_q}{C_q} \inf_{p \in \mathcal{P}_q(\mathcal{T}_k)} \|f_0 - p\|_\infty \left(\frac{\Delta_{\mathcal{T}_{k+1}}}{\Delta_{\mathcal{T}_k}} \right)^\alpha \\ &\geq Q \inf_{p \in \mathcal{P}_q(\mathcal{T}_k)} \|f_0 - p\|_\infty, \end{aligned}$$

with $Q := (2/3)^\alpha \delta_q / C_q$. Naturally, we expect that $\inf_{p \in \mathcal{P}_q(\mathcal{T}_{k+1})} \|f_0 - p\|_\infty \leq \inf_{p \in \mathcal{P}_q(\mathcal{T}_k)} \|f_0 - p\|_\infty$ since finer knot partitions will result in better approximations. Functions in \mathcal{F}^α however are able to invert this natural order by incurring an extra constant. Note that self-referencing in splines approximation error can be seen as the L_∞ -splines analogue of the polished tail condition introduced in Definition 3.2 of [21].

Now are sets like (5.9) a reasonable construction in practice? Here, we give two classes of functions that satisfy the lower bound of (5.9).

Example 1: Let $m \in \mathbb{N}$ be such that $t_{m+1} - t_m = \Delta_{\mathcal{T}}$. Let $\|g^{(\alpha)}\|_\infty < \infty$ for some $\alpha \in \mathbb{N}$, g is not a polynomial spline of order q or less, with g and its derivatives up to $\alpha - 1$ are 0 at its boundary 0 (or at 1 with slight modification of the argument below). We construct

$$f_0(x) = \begin{cases} \Delta_{\mathcal{T}}^\alpha g\left(\frac{x - t_m}{t_{m+1} - t_m}\right), & \text{for } t_m \leq x \leq t_{m+1}, \\ 0, & \text{otherwise.} \end{cases}$$

Let us denote the L_∞ -Sobolev space $L_\infty^\alpha := \{f : \sum_{k=0}^\alpha \|f^{(k)}\|_\infty < \infty\}$. By Taylor's expansion on t_m and the assumption that g and its derivatives up to $\alpha - 1$ vanish at 0, it follows that $f_0(x) = f_0^{(\alpha)}(\xi)(x - t_m)^\alpha / \alpha!$ for some ξ in between x and t_m . Therefore since our domain is $[0, 1]$ and $\|f_0^{(\alpha)}\|_\infty < \infty$ because of $\|g^{(\alpha)}\|_\infty < \infty$, it follows that $f_0 \in L_\infty^\alpha$. By Remark 1 of [4], we have the embedding $L_\infty^\alpha \subset \mathcal{B}_{\infty, \infty}^\alpha$, where the latter Besov space is equivalent to \mathcal{H}^α . Therefore, we conclude $\|f_0\|_{\mathcal{H}^\alpha} \leq C$ for some constant $C > 0$.

Define $\beta := \inf_{p \in \mathcal{P}_q(\mathcal{T})} \|g - p\|_\infty$ and denote $\bar{p}(x) := \Delta_{\mathcal{T}}^\alpha p[(x - t_m)/(t_{m+1} - t_m)]$. Because g is not a spline of order q , we have by a change of variable that

$$0 < \beta \leq \sup_{x \in [0, 1]} |g(x) - p(x)| = \Delta_{\mathcal{T}}^{-\alpha} \sup_{x \in [t_m, t_{m+1}]} |f(x) - \bar{p}(x)|.$$

Now since the inequality above holds for any arbitrary $p \in \mathcal{P}_q(\mathcal{T})$ and \bar{p} is also a polynomial spline, we can conclude that

$$\inf_{p \in \mathcal{P}_q(\mathcal{T})} \|f_0 - p\|_\infty \geq \Delta_{\mathcal{T}}^\alpha (\beta/C) C \geq \delta_q \|f_0\|_{\mathcal{H}^\alpha([0, 1])} \Delta_{\mathcal{T}}^\alpha,$$

where $\delta_q := \beta/C$ depends only on f_0 and q .

Another important class of examples are those functions who exhibit self-similarity in scale, such that the function looks approximately the same at different scaling of its domain.

Example 2: Let f_0 be self-similar (or invariant) in scale of order α , i.e., $f_0(x/\lambda) = \lambda^{-\alpha} f_0(x)$ for any $\lambda > 0$, and f_0 is not a polynomial spline of order q . For $p \in \mathcal{P}_q(\mathcal{T})$, let $\widehat{p}(x) = \lambda^\alpha p(x/\lambda)$. By a change of variables, we have

$$\sup_{0 \leq x \leq 1} |f_0(x) - p(x)| = \lambda^{-\alpha} \sup_{0 \leq x \leq \lambda} |\lambda^\alpha f_0(x/\lambda) - \lambda^\alpha p(x/\lambda)| = \lambda^{-\alpha} \sup_{0 \leq x \leq \lambda} |f_0(x) - \widehat{p}(x)|.$$

Note that \widehat{p} is also a polynomial splines of order q . Since the above equality is valid for any $p \in \mathcal{P}_q(\mathcal{T})$ and f_0 is not a polynomial splines of order q ,

$$0 < \delta_q := \inf_{p \in \mathcal{P}_q(\mathcal{T})} \|f_0 - p\|_{L_\infty[0,1]} \leq \lambda^{-\alpha} \inf_{p \in \mathcal{P}_q(\mathcal{T})} \|f_0 - p\|_{L_\infty[0,\lambda]},$$

where δ_q is some constant depending only on q and f_0 . The lower bound can then be recovered through setting $\lambda = \|f_0\|_{\mathcal{H}^\alpha}^{1/\alpha} \Delta_{\mathcal{T}}$. This notion of self-similarity can be substantially generalized to include being invariant with respect to certain transformations of the domain, for example see (2.3) of [8].

6 Application II: Wavelet bases

Since our domain is $[0, 1]$, we will use the Cohen-Daubechies-Vial (CDV) father and mother wavelets φ and ψ . These wavelets are compactly supported and boundary corrected. Assume that the wavelets are r -regular, in the sense that $|(d^p/dx^p)\varphi(x)| \leq C_m(1 + |x|)^{-m}$ for some constants $C_m, m \in \mathbb{N}$ and $0 \leq p \leq r$. We construct an orthonormal basis for $L_2([0, 1])$ by dilating and translating these wavelets as follows: $\varphi_{j,k}(x) = 2^{j/2}\varphi(2^jx - k)$ and $\psi_{j,k}(x) = 2^{j/2}\psi(2^jx - k)$. We then project f onto the 2^J -dimensional CDV wavelet bases as

$$f(x) = \sum_{k=0}^{2^N-1} \vartheta_k \varphi_{N,k}(x) + \sum_{j=N}^{J-1} \sum_{k=0}^{2^j-1} \theta_{j,k} \psi_{j,k}(x) =: \boldsymbol{\psi}_J(x)^T \boldsymbol{\theta},$$

where we have concatenated the father and mother wavelets into a single vector of functions $\boldsymbol{\psi}_J(x)$, and their coefficients collected into $\boldsymbol{\theta}$. The total number of wavelets is 2^J which we will identify with the “ J ” used in previous sections. Here we choose N such that $2^N \geq 2r$. As before, we put prior $\boldsymbol{\theta}|\sigma \sim \mathcal{N}(\boldsymbol{\eta}, \sigma^2 \boldsymbol{\Omega})$ and estimate σ^2 by empirical Bayes (3.5). Write the wavelet basis matrix as $\boldsymbol{\Psi} = (\boldsymbol{\psi}_J(X_1)^T, \dots, \boldsymbol{\psi}_J(X_n)^T)^T$. We choose the covariates such that $\|F_n - F\|_\infty = O(1/n)$. Since $2^{j_{\max}} = o(n)$ by assumption, Lemma 6.4 of [24] tells us that there exist constants $C_1, C_2 > 0$ such that

$$C_1 n \leq \lambda_{\min}(\boldsymbol{\Psi}^T \boldsymbol{\Psi}) \leq \lambda_{\max}(\boldsymbol{\Psi}^T \boldsymbol{\Psi}) \leq C_2 n, \quad (6.1)$$

and we have $\lambda_{\boldsymbol{\Psi}, J} = n$ in this case. Since the wavelets are compact supported and uniformly bounded, we have $l_{1,J} \lesssim 2^{J/2}$ (Lemma 9.7). If $2^J \leq n$, then $\lambda_{\boldsymbol{\Psi}, J} \gtrsim l_{1,J}^2$. Suppose the support of ψ is $[a, b]$, then for any x and j , $\psi_{j,k}(x) \neq 0$ for $2^jx - b \leq k \leq 2^jx - a$ and only a fixed number (not depending on j) of consecutive k 's for $\psi_{j,k}$ are nonzero. Thus as in the B-splines case before, $\boldsymbol{\Psi}^T \boldsymbol{\Psi}$ is banded.

Define $K_J(x, y) = \sum_k \varphi_{J,k}(x)\varphi_{J,k}(y)$ and let $K_J(f)(x) = \int K_J(x, y)f(y)dy$ for any $f \in L_2([0, 1])$ be the wavelet projection operator. By Theorem 4 of Section 2.6 [14], we know that $K_J(p) = p$ for any polynomial p of degree r or less. Thus, $2^{j_{\min}} = (n/\log n)^{1/(2r+1)}$ and we adapt α up to r . Then in the present case

$$\widehat{j}_n = \min \left\{ j \in \mathcal{J} : \|E_j(f|\mathbf{Y}) - E_i(f|\mathbf{Y})\|_\infty \leq \tau \widehat{\sigma}_i \sqrt{\frac{i2^i}{n}}, \forall i > j, i \in \mathcal{J} \right\}. \quad (6.2)$$

Let $\mathcal{B}_{\infty, \infty}^\alpha$ be the Besov space consisting of functions $g : [0, 1] \rightarrow \mathbb{R}$ such that $\|g\|_{\mathcal{B}_{\infty, \infty}^\alpha} < \infty$, where the Besov norm is defined as

$$\|g\|_{\mathcal{B}_{\infty, \infty}^\alpha} = \max_{0 \leq k \leq 2^N - 1} |\langle g, \varphi_{N,k} \rangle| + \max_{j \geq N} 2^{j(\alpha+1/2)} \max_{0 \leq k \leq 2^j - 1} |\langle g, \psi_{j,k} \rangle|. \quad (6.3)$$

Here we take $\mathcal{G} = \{g : \|g\|_{\mathcal{B}_{\infty, \infty}^\alpha} \leq R\} =: \mathcal{B}_{\infty, \infty}^\alpha(R)$ to be the Besov ball of radius R . By Proposition 4.3.8 of [5], we have for any $g \in \mathcal{B}_{\infty, \infty}^\alpha(R)$ that there is a constant $C_r > 0$ depending on r such that

$$\|K_J(g) - g\|_\infty \leq C_r \|g\|_{\mathcal{B}_{\infty, \infty}^\alpha} 2^{-J\alpha}, \quad (6.4)$$

and $h(x) = 2^{-x\alpha}$ in this case. Hence we see that $h(j_{\min}) = (\log n/n)^{\alpha/(2r+1)} \rightarrow 0$ as $n \rightarrow \infty$. This wavelet projection can be represented as $K_J(g) = \langle \boldsymbol{\psi}_J, \boldsymbol{\theta}_0 \rangle$ for some $\boldsymbol{\theta}_0 \in \mathbb{R}^{2^J}$ and by virtue of Proposition 4.3.5 of [5], we have $\|\boldsymbol{\theta}_0\|_\infty < \infty$. Suppose $f_0 \in \mathcal{B}_{\infty, \infty}^\alpha(R)$ with $\alpha > 0$ unknown. Then j_n^* of (3.9) balances $\|f_0\|_{\mathcal{B}_{\infty, \infty}^\alpha} 2^{-j\alpha}$ and $\sqrt{2^j j/n}$ to yield $2^{j_n^*} \asymp \|f_0\|_{\mathcal{B}_{\infty, \infty}^\alpha}^{2/(2\alpha+1)} (n/\log n)^{1/(2\alpha+1)}$. As a result, the wavelet version of Theorem 3.4 is

$$\sup_{0 < \alpha \leq r} \sup_{f_0 \in \mathcal{B}_{\infty, \infty}^\alpha(R)} E_0 \Pi[\|f - f_0\|_\infty > \xi (\log n/n)^{\alpha/(2\alpha+1)} | \mathbf{Y}] \lesssim (\log n/n)^M \quad (6.5)$$

as $n \rightarrow \infty$, for some constants $R, \xi, M > 0$.

The wavelet credible band constructed using the volume-of-tube method is

$$\mathcal{W}_n := \left\{ f : |f(x) - E_{\widehat{j}_n}[f(x)|\mathbf{Y}]| \leq w_{\gamma, \widehat{j}_n} \sqrt{\text{Var}_{\widehat{j}_n}[f(x)|\mathbf{Y}]}, \forall x \in [0, 1] \right\},$$

where the posterior mean and variance is

$$E_{\widehat{j}_n}[f(x)|\mathbf{Y}] = \boldsymbol{\psi}_{\widehat{j}_n}(x)^T (\boldsymbol{\Psi}^T \boldsymbol{\Psi} + \boldsymbol{\Omega}^{-1})^{-1} (\boldsymbol{\Psi}^T \mathbf{Y} + \boldsymbol{\Omega}^{-1} \boldsymbol{\eta}), \quad (6.6)$$

$$\text{Var}_{\widehat{j}_n}[f(x)|\mathbf{Y}] = \widehat{\sigma}_{\widehat{j}_n}^2 \boldsymbol{\psi}_{\widehat{j}_n}(x)^T (\boldsymbol{\Psi}^T \boldsymbol{\Psi} + \boldsymbol{\Omega}^{-1})^{-1} \boldsymbol{\psi}_{\widehat{j}_n}(x), \quad (6.7)$$

and we choose $w_{\gamma, \widehat{j}_n}$ such that it solves $\gamma = |\beta_{\widehat{j}_n}| \pi^{-1} e^{-w_{\gamma, \widehat{j}_n}^2} + 2[1 - \Phi(w_{\gamma, \widehat{j}_n})]$, where the arc length is

$$|\beta_{\widehat{j}_n}| = \int_0^1 \frac{[\boldsymbol{\psi}_{\widehat{j}_n}(x)^T \mathbf{M} \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{M} \boldsymbol{\psi}_{\widehat{j}_n}(x)]^{1/2}}{[\boldsymbol{\psi}_{\widehat{j}_n}(x)^T \mathbf{M} \boldsymbol{\psi}_{\widehat{j}_n}(x)]^{3/2}} dx, \quad (6.8)$$

for $\mathbf{M} := (\Psi^T \Psi + \Omega^{-1})^{-1}$ and $\mathbf{R} := \dot{\psi}_{\hat{j}_n}(x) \psi_{\hat{j}_n}(x)^T - \psi_{\hat{j}_n}(x) \dot{\psi}_{\hat{j}_n}(x)^T$. As before, we proceed to verify the 3 conditions. Let us remind again the reader that J used in the previous sections now translates to 2^J and the event $\{j_n^*/\kappa \leq \hat{j}_n \leq j_n^*\}$ appearing in conditions 1 and 3 now is $\{j_n^* - \kappa \leq \hat{j}_n \leq j_n^*\}$.

We start with the third condition. By applying Lemma 9.7, we have $\inf_{x \in [0,1]} \|\psi_{\hat{j}_n}(x)\| \gtrsim 2^{\hat{j}_n/2}$. Then under event $\{j_n^* - \kappa \leq \hat{j}_n \leq j_n^*\}$, we have $(2^{\hat{j}_n}/n)^{1/2} \asymp 2^{-\alpha \hat{j}_n}$ and the third condition is

$$(\hat{j}_n)^{-1/2} 2^{-\hat{j}_n/2} \left[\sqrt{\frac{2^{\hat{j}_n}}{n}} + \sqrt{n} 2^{-\alpha \hat{j}_n} \right] \lesssim (\log n)^{-1/2} \left(\frac{1}{\sqrt{n}} + 1 \right) = o_{P_0}(1),$$

as $n \rightarrow \infty$. The second condition is fulfilled due to the following two lemmas.

Lemma 6.1. *Let $|\beta_J| = \int_0^1 \|\beta'_J(x)\| dx$ be the arc length of the curve $\beta_J : [0, 1] \rightarrow S^{J-1}$ given in (6.8). Then as $n \rightarrow \infty$,*

$$|\beta_J| \asymp 2^J \quad \text{for any } J \in \mathcal{J}.$$

Lemma 6.2. *For any $J \in \mathcal{J}$, there exist constants $\xi_1, \xi_2, \xi_3 > 0$ such that*

$$|\beta_{0,J}| \leq \frac{[1 - \xi_1 n^{-1} + \xi_2 n^{-2}]^{1/2}}{[1 - \xi_3 n^{-1}]^{3/2}} |\beta_J|.$$

Consequently, $|\beta_{0,J}| \leq |\beta_J|$ as $n \rightarrow \infty$.

Let us now describe the set \mathcal{F} used in the first condition. Define δ_r to be a positive constant that depends on r the regularity of the CDV wavelets. Let $\{V_j : j \in \mathbb{Z}\}$ generate a multiresolution analysis of $L_2([0, 1])$ associated with these wavelets, such that $V_{j-1} \subset V_j$, $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$ and $\bigcup_{j \in \mathbb{Z}} V_j$ is dense in $L_2([0, 1])$. For a fixed $J_0 \in \mathbb{N}$, we take \mathcal{F} to be

$$\mathcal{F}^\alpha = \bigcap_{j \geq J_0} \left\{ f : \inf_{p \in V_j} \|p - f\|_\infty \geq \delta_r \|f\|_{\mathcal{B}_{\infty, \infty}^\alpha} 2^{-j\alpha} \right\}. \quad (6.9)$$

Equipped with this set, we can use the same reasoning as in the proof of Lemma 5.3 to deduce the following lemma, which serves to fulfil the first condition of Theorem 4.2 for the CDV wavelet bases.

Lemma 6.3. *Let $\kappa \in \mathbb{N}$ be some large enough constant depending on δ_r, α and τ . Then for any $R > 0$ and $j_{\min} \geq J_0$, there are constants $\mu > 0$ such that*

$$\sup_{0 < \alpha \leq r} \sup_{f_0 \in \mathcal{F}^\alpha} P_0(\hat{j}_n < j_n^* - \kappa) \lesssim e^{-\mu j_{\min}},$$

and consequently in view of Lemma 4.4, we have as $n \rightarrow \infty$,

$$\inf_{0 < \alpha \leq r} \inf_{f_0 \in \mathcal{B}_{\infty, \infty}^\alpha(R) \cap \mathcal{F}^\alpha} P_0(j_n^* - \kappa \leq \hat{j}_n \leq j_n^*) \rightarrow 1. \quad (6.10)$$

Now in view of (6.4), the self-referencing version is

$$\inf_{p \in \mathcal{V}_{j+1}} \|p - f_0\|_\infty \geq \delta_r \|f_0\|_{\mathcal{B}_{\infty, \infty}^\alpha} 2^{-(j+1)\alpha} \geq \frac{\delta_r 2^{-\alpha}}{C_r} \|K_j(f_0) - f_0\|_\infty \geq G \inf_{p \in \mathcal{V}_j} \|p - f_0\|_\infty,$$

for $G := \delta_r 2^{-\alpha} / C_r$. As before, the uncertainty principle applies here in the wavelet case. That is, if we used a posterior based on wavelet random series at resolution j and f_0 so happened is in \mathcal{V}_j in the multiresolution analysis, then we are “correct” in estimating the truth, and our posterior should contract to this truth at the optimal rate. However, $\inf_{p \in \mathcal{V}_j} \|p - f_0\| = 0$ and we will not be able to quantify uncertainty honestly for this truth.

As a result, we can invoke Theorem 4.2 and conclude that for any $R > 0$,

$$\inf_{0 < \alpha \leq r} \inf_{f_0 \in \mathcal{B}_{\infty, \infty}^\alpha(R) \cap \mathcal{F}^\alpha} P_0(f_0 \in \mathcal{W}_n) \geq 1 - \gamma + o(1). \quad (6.11)$$

By Lemmas 6.1, 6.3 and 9.7, we have the following CDV wavelet version of Corollary 4.3 on the size of the radius function

$$\inf_{0 < \alpha \leq r} \inf_{f_0 \in \mathcal{B}_{\infty, \infty}^\alpha(R) \cap \mathcal{F}^\alpha} P_0 [r(x) \asymp (\log n/n)^{\alpha/(2\alpha+1)}, \forall x \in [0, 1]] \rightarrow 1. \quad (6.12)$$

7 Simulations

The right hand side of the Bayes Lepski stopping rule (3.6) is a measure of posterior variation in sup-norm, and it serves to check the increase in posterior mean bias on the left hand side. The expression presented however is asymptotic and the use of (maximal) eigenvalues at each iteration makes it not suitable for fast computations. Moreover, to use this rule and obtain sufficient coverage for our bands, we need to induce some form of undersmoothing such that we will on average choose J that is larger than necessary. The difficulty part is to figure out a way to do this without knowing the true smoothness of f_0 .

To achieve this, we need to find a good finite sample proxy on the right hand side, such that its asymptotic formula coincides with the one in (3.6). There are several different choices for such proxies with the same large sample property, and we choose the one which tends to give smaller values on average. For our simulations, we found that $V_i := \hat{\sigma}_i \sqrt{\inf_{x \in [0, 1]} \mathbf{a}_i(x)^T (\mathbf{A}^T \mathbf{A} + \mathbf{\Omega}^{-1})^{-1} \mathbf{a}_i(x) \sqrt{\log j}}$ works. Here we set $\tau = 1$ for the sake of algorithmic simplicity, since a large τ is there for purely technical reason (to ensure exponent of certain powers of n is negative and so they approach 0 as $n \rightarrow \infty$). Based on this line of reasoning, we then substitute V_i as the right hand side of (3.6), and this give the following practical version in pseudocode for credible bands computation using the Bayes Lepski procedure.

```

Initialize  $j_{\max} = n/\log^2 n$  and  $j_{\min} = (n/\log n)^{1/(2v+1)}$ ;
Set  $j = j_{\max} - 1$ ;
while  $j \geq j_{\min}$  do
  rule  $\leftarrow 0$ ;
  for  $i: j < i \leq j_{\max}$  do
    | rule  $\leftarrow$  rule  $+\mathbb{1}\{\|E_j(f|\mathbf{Y}) - E_i(f|\mathbf{Y})\|_\infty > V_i\}$ ;
  end
  if rule  $> 0$  then
    |  $\hat{j}_n \leftarrow j + 1$ ;
    | break;
  else
    |  $j \leftarrow j - 1$ ;
  end
end

```

Algorithm 1: The Bayes Lepski algorithm (undersmoothing version)

We store values of $E_i(f|\mathbf{Y})$ and V_i at each iteration and they are used again for comparison at a lower j if needed. The computation of these two quantities is made easier by the bandedness of $\mathbf{A}^T \mathbf{A}$ and the compact support property of the bases utilized.

The true function is taken to be $f_0(x) = 2x - x^3 + \exp\{-50(x - 0.5)^2\}$ for $x \in [0, 1]$, and we observed $Y_i = f_0(X_i) + \varepsilon_i$ such that $X_i = i/n, i = 1, \dots, n$ and $\sigma_0^2 = 0.1$. We will use cubic B-splines ($q = 4$) with uniformly distributed knots. For the Gaussian prior on the coefficients, we set $\boldsymbol{\eta} = \mathbf{0}$ and $\boldsymbol{\Omega} = 10\mathbf{I}$. We estimate σ_0^2 using empirical Bayes with formula given in (3.5), and we compute $J = \hat{j}_n$ using the Bayes Lepski algorithm given above. Once this is done, we construct the B-splines volume of tube credible band as in (5.4), and we calculate the quantile w_{γ, \hat{j}_n} by computing the arc length (5.7) and solving the equation relating them. We set the credibility level $1 - \gamma$ as 0.95 and see whether the resulting credible band has at least that amount of coverage probability. To highlight the benefits of the Bayes Lepski and volume of tube construction, we compare it against two other methods: the frequentist version where least squares is used instead of the posterior mean, and the L_∞ -ball around the posterior mean with (non-inflated) fixed radius that was considered in [23]. All computations were carried in the statistical software R and the code can be found in the first author's web page (type William Weimin Yoo in search engines).

Table 1 shows coverage probabilities computed for all 3 methods, where these probabilities are calculated empirically by repeating the experiment 1000 times and recording whether f_0 lies entirely within the credible bands for each Monte Carlo iterate.

All methods have at least 0.95 coverage for large samples and the results are comparable in this case. For very small samples, the Bayes Lepski (and also the Bayes fixed radius) has slightly higher coverage than the frequentist Lepski procedure. At first glance, it seems that the Bayes Lepski and the fixed radius method have comparable performance. But before making any hasty conclusions, let us take a look at the 1000 Monte Carlo mean radius for each method visualized as box-plots in Figure 2.

Table 1: Coverage probabilities for all 3 methods. Nominal level is 0.95.

n	50	100	300	500	1000	2000
Bayes Lepski	0.908	0.937	0.943	0.958	0.955	0.957
Frequentist	0.87	0.924	0.945	0.947	0.956	0.958
Fixed radius	0.919	0.938	0.953	0.955	0.941	0.946

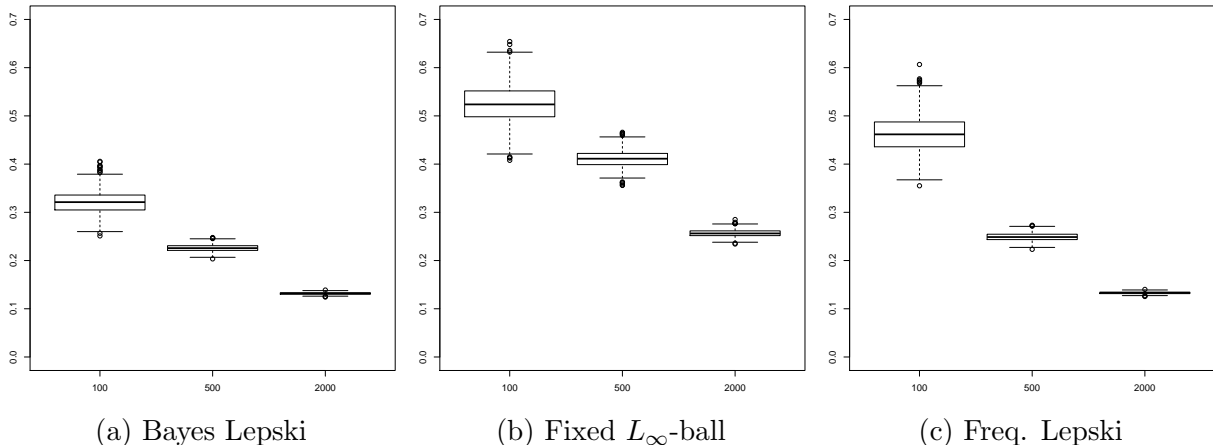


Figure 2: 1000 Monte Carlo mean radius for $n = 100, 500, 2000$

We see that the Bayes Lepski with volume of tube method has narrower bands on average when compared to the fixed L_∞ -ball (roughly half of its size). The frequentist volume of tube method has slightly larger bands for small sample sizes i.e., $n = 100$ but is otherwise comparable to the Bayesian counterpart for larger sample sizes considered. For all methods, the radius decreases with n as expected. The table and figure together show that the slight deterioration in coverage and the sudden increase in radius of the frequentist Lepski at $n \leq 100$ is due to random fluctuations associated with small sample sizes. It appears that the presence of shrinkage through $\mathbf{\Omega} = 10\mathbf{I}$ stabilizes this erratic fluctuations and resulting in a more predictable behavior in terms of the rate the radius decreases to 0. Among the three methods, the Bayes fixed radius takes the longest time to compute the bands, this is due to the fact that it must further estimate the quantile empirically by drawing samples from the posterior, while the volume of tube method provides an explicit expression to compute the quantile through arc length calculations.

Simulation results in conjunction with our theoretical investigations concur that the Bayes Lepski volume of tube methods is a very promising Bayesian uncertainty quantification procedure. In particular, the Bayes Lepski rule by itself is flexible enough to be deployed in other statistical applications where tuning parameters need to be determined, for example the number of components in mixture models or spike-and-slab prior weights in high-dimensional linear regression. We envision that this rule together with the volume of tube method can be applied to other modeling situations, and this includes binary or Poisson regression and

other generalized linear models. We hope to pursue these research directions in the future.

8 Proofs

Let us record here the common eigenvalue bound assumption and its corollary for ease of reference since they are used frequently in the proofs. Recall that:

$$\lambda_{\mathbf{A},J} \lesssim \lambda_{\min}(\mathbf{A}^T \mathbf{A}) \leq \lambda_{\max}(\mathbf{A}^T \mathbf{A}) \lesssim \lambda_{\mathbf{A},J}. \quad (8.1)$$

Observe that for symmetric and positive definite \mathbf{T} , $\|\mathbf{T}\|_{(2,2)} = \lambda_{\max}(\mathbf{T})$ and $\lambda_{\max}(\mathbf{T}^{-1}) = \lambda_{\min}^{-1}(\mathbf{T})$ and vice versa. Now for $J \in \mathcal{J}$ in the candidate set, $\lambda_{\mathbf{A},J}$ will dominate any constants for large enough n by assumption, and these facts together with the prior assumption in (3.1) imply that as $n \rightarrow \infty$,

$$\lambda_{\mathbf{A},J}^{-1} \lesssim \lambda_{\min} \{(\mathbf{A}^T \mathbf{A} + \boldsymbol{\Omega}^{-1})^{-1}\} \leq \lambda_{\max} \{(\mathbf{A}^T \mathbf{A} + \boldsymbol{\Omega}^{-1})^{-1}\} \lesssim \lambda_{\mathbf{A},J}^{-1}. \quad (8.2)$$

For notational simplicity, let us write $\tilde{f}_J(x) = \mathbb{E}_J[f(x)|\mathbf{Y}]$ and $\tilde{V}_J(x) = \text{Var}_J[f(x)|\mathbf{Y}]$. Also, denote $\hat{\mathbf{a}}_J(x) = (a'_1(x), \dots, a'_J(x))^T$. For $J \in \mathcal{J}$, define $\mathcal{U}_n = \{\sigma : |\sigma - \sigma_0| \leq \xi \delta_{n,J}\}$ to be a shrinking neighborhood around σ_0 with $\delta_{n,J} = o(1)$ and the constant $\xi > 0$ given in Proposition 3.2.

Proof of Proposition 3.2. Let us define $\mathbf{U} = n^{-1}(\mathbf{A}\boldsymbol{\Omega}\mathbf{A}^T + \mathbf{I}_n)^{-1}$ and $\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{F}_0 \sim \text{N}(0, \sigma_0^2 \mathbf{I}_n)$ where $\mathbf{F}_0 = (f_0(X_1), \dots, f_0(X_n))^T$. For two square matrices of equal size, we say $\mathbf{A} < \mathbf{B}$ if $\mathbf{B} - \mathbf{A}$ is positive definite. Now by adding and subtracting \mathbf{F}_0 and $\mathbb{E}_0(\boldsymbol{\varepsilon}^T \mathbf{U} \boldsymbol{\varepsilon})$, we have $|\hat{\sigma}_J^2 - \sigma_0^2|$ is

$$|\boldsymbol{\varepsilon}^T \mathbf{U} \boldsymbol{\varepsilon} - \mathbb{E}_0(\boldsymbol{\varepsilon}^T \mathbf{U} \boldsymbol{\varepsilon}) + 2(\mathbf{F}_0 - \mathbf{A}\boldsymbol{\eta})^T \mathbf{U} \boldsymbol{\varepsilon} + (\mathbf{F}_0 - \mathbf{A}\boldsymbol{\eta})^T \mathbf{U} (\mathbf{F}_0 - \mathbf{A}\boldsymbol{\eta}) + \mathbb{E}_0(\boldsymbol{\varepsilon}^T \mathbf{U} \boldsymbol{\varepsilon}) - \sigma_0^2|.$$

Using the triangle inequality and the fact that $\mathbb{E}_0(\boldsymbol{\varepsilon}^T \mathbf{U} \boldsymbol{\varepsilon}) = \mathbb{E}_0 \text{tr}(\boldsymbol{\varepsilon}^T \mathbf{U} \boldsymbol{\varepsilon}) = \text{tr} \mathbf{U} \mathbb{E}_0(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T) = \sigma_0^2 \text{tr}(\mathbf{U})$, the above is bounded by

$$|\boldsymbol{\varepsilon}^T \mathbf{U} \boldsymbol{\varepsilon} - \mathbb{E}_0(\boldsymbol{\varepsilon}^T \mathbf{U} \boldsymbol{\varepsilon})| + 2|(\mathbf{F}_0 - \mathbf{A}\boldsymbol{\eta})^T \mathbf{U} \boldsymbol{\varepsilon}| + B_J,$$

where $B_J := |\sigma_0^2[\text{tr}(\mathbf{U}) - 1] + (\mathbf{F}_0 - \mathbf{A}\boldsymbol{\eta})^T \mathbf{U} (\mathbf{F}_0 - \mathbf{A}\boldsymbol{\eta})|$. Now using the fact that for two random variables X, Y , $P(X + Y > a) \leq P(X > a/2) + P(Y > a/2)$ for any a , we can write

$$\begin{aligned} P_0(|\hat{\sigma}_J^2 - \sigma_0^2| > M\delta_{n,J}) &\leq P_0\left(|\boldsymbol{\varepsilon}^T \mathbf{U} \boldsymbol{\varepsilon} - \mathbb{E}_0(\boldsymbol{\varepsilon}^T \mathbf{U} \boldsymbol{\varepsilon})| > \frac{M\delta_{n,J} - B_J}{2}\right) \\ &\quad + P_0\left(|(\mathbf{F}_0 - \mathbf{A}\boldsymbol{\eta})^T \mathbf{U} \boldsymbol{\varepsilon}| > \frac{M\delta_{n,J} - B_J}{4}\right). \end{aligned} \quad (8.3)$$

Let us first work on the second term and we start by bounding B_J . Note that since $\mathbf{U} < n^{-1} \mathbf{I}_n$, we have $\text{tr}(\mathbf{U}) < 1$. Then in view of the inequality $(\mathbf{x} + \mathbf{y})^T \mathbf{T} (\mathbf{x} + \mathbf{y}) \leq 2\mathbf{x}^T \mathbf{T} \mathbf{x} + 2\mathbf{y}^T \mathbf{T} \mathbf{y}$ for any positive definite \mathbf{T} , we have

$$B_J \leq \sigma_0^2[1 - \text{tr}(\mathbf{U})] + 2(\mathbf{F}_0 - \mathbf{A}\boldsymbol{\theta}_0)^T \mathbf{U} (\mathbf{F}_0 - \mathbf{A}\boldsymbol{\theta}_0) + 2(\boldsymbol{\theta}_0 - \boldsymbol{\eta})^T \mathbf{A}^T \mathbf{U} \mathbf{A} (\boldsymbol{\theta}_0 - \boldsymbol{\eta}), \quad (8.4)$$

with $\boldsymbol{\theta}_0$ taken from (3.8). To proceed, we need the following result called the binomial inverse theorem (Theorem 18.2.8 of [6]), and it says that for matrices $\mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}$ of conformable dimensions, we will have

$$(\mathbf{B} + \mathbf{CDE})^{-1} = \mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{C}(\mathbf{D}^{-1} + \mathbf{EB}^{-1}\mathbf{C})^{-1}\mathbf{EB}^{-1}.$$

Hence by applying the above twice to $n\mathbf{U}$, we get

$$n\mathbf{U} = \mathbf{I}_n - \mathbf{A}(\mathbf{A}^T\mathbf{A} + \boldsymbol{\Omega}^{-1})^{-1}\mathbf{A}^T = \mathbf{I}_n - \mathbf{P}_\mathbf{A} + \mathbf{V}, \quad (8.5)$$

where $\mathbf{P}_\mathbf{A} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is the orthogonal projection matrix and $\mathbf{V} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}[\boldsymbol{\Omega} + (\mathbf{A}^T\mathbf{A})^{-1}]^{-1}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$.

For the first term in (8.4), we use (8.5) to write

$$1 - \text{tr}(\mathbf{U}) = n^{-1}\text{tr}(\mathbf{I}_n - n\mathbf{U}) = n^{-1}\text{tr}(\mathbf{P}_\mathbf{A} - \mathbf{V}) \leq n^{-1}\text{rank}(\mathbf{A}) \leq J/n$$

because \mathbf{V} is positive definite. Since $\mathbf{U} < n^{-1}\mathbf{I}_n$, we have $\lambda_{\max}(\mathbf{U}) < n^{-1}$, then using the inequality $\mathbf{x}^T\mathbf{T}\mathbf{x} \leq \lambda_{\max}(\mathbf{T})\|\mathbf{x}\|^2$ for any square matrix \mathbf{T} , we can bound the second term of (8.4) by

$$\lambda_{\max}(\mathbf{U})\|\mathbf{F}_0 - \mathbf{A}\boldsymbol{\theta}_0\|^2 \leq \|\mathbf{F}_0 - \mathbf{A}\boldsymbol{\theta}_0\|_\infty^2 \leq C_0^2\|f_0\|_{\mathcal{G}}^2h(J)^2,$$

in view of (3.8) in the definition of \mathcal{G} . Next, observe that $(\mathbf{I}_n - \mathbf{P}_\mathbf{A})\mathbf{A} = \mathbf{0}$ and $\mathbf{A}^T\mathbf{V}\mathbf{A} = [\boldsymbol{\Omega} + (\mathbf{A}^T\mathbf{A})^{-1}]^{-1} < \boldsymbol{\Omega}^{-1}$. These facts then enable us to write the last term in (8.4) as

$$n^{-1}(\boldsymbol{\theta}_0 - \boldsymbol{\eta})^T\mathbf{A}^T\mathbf{V}\mathbf{A}(\boldsymbol{\theta}_0 - \boldsymbol{\eta}) \leq n^{-1}\lambda_{\max}(\boldsymbol{\Omega}^{-1})\|\boldsymbol{\theta}_0 - \boldsymbol{\eta}\|^2 \leq (J/n)\lambda_{\min}^{-1}(\boldsymbol{\Omega})\|\boldsymbol{\theta}_0 - \boldsymbol{\eta}\|_\infty^2 \lesssim J/n,$$

since $\|\boldsymbol{\eta}\|_\infty < \infty$ and $\lambda_{\min}^{-1}(\boldsymbol{\Omega}) \leq c_1^{-1}$ by our prior assumption in (3.1), and $\|\boldsymbol{\theta}_0\|_\infty < \infty$ is from the definition of \mathcal{G} . By considering these bounds together, we conclude that $B_J \leq Q_1[J/n + \|f_0\|_{\mathcal{G}}^2h(J)^2]$ for some constant $Q_1 > 0$.

Returning to the original task of controlling the second term in (8.3), which we will do through the tail bound of a standard normal, we first note $(\mathbf{F}_0 - \mathbf{A}\boldsymbol{\eta})^T\mathbf{U}\boldsymbol{\varepsilon} \sim \text{N}[0, \sigma_0^2(\mathbf{F}_0 - \mathbf{A}\boldsymbol{\eta})^T\mathbf{U}^2(\mathbf{F}_0 - \mathbf{A}\boldsymbol{\eta})]$. This variance can in turn be decomposed into two quadratic forms as in the last two terms of B_J in (8.4) above but with \mathbf{U}^2 in the middle, and the first is

$$\|\mathbf{U}(\mathbf{F}_0 - \mathbf{A}\boldsymbol{\theta}_0)\|^2 \leq n\lambda_{\max}^2(\mathbf{U})\|\mathbf{F}_0 - \mathbf{A}\boldsymbol{\theta}_0\|_\infty^2 \leq n^{-1}C_0^2\|f_0\|_{\mathcal{G}}^2h(J)^2,$$

while the second is $\|\mathbf{A}\mathbf{U}(\boldsymbol{\theta}_0 - \boldsymbol{\eta})\|^2$. To deal with this term, recognize that $\mathbf{I}_n - \mathbf{P}_\mathbf{A}$ is idempotent and projects into the null space of \mathbf{A} , it then holds that $n^2\mathbf{A}^T\mathbf{U}^2\mathbf{A} = \mathbf{A}^T(\mathbf{I}_n - \mathbf{P}_\mathbf{A} + \mathbf{V})^2\mathbf{A} = \mathbf{A}^T\mathbf{V}^2\mathbf{A} < [\boldsymbol{\Omega} + (\mathbf{A}^T\mathbf{A})^{-1}]^{-1} < \boldsymbol{\Omega}^{-1}$. Thus,

$$\|\mathbf{A}\mathbf{U}(\boldsymbol{\theta}_0 - \boldsymbol{\eta})\|^2 \leq \lambda_{\max}(\mathbf{A}^T\mathbf{U}\mathbf{A})\|\boldsymbol{\theta}_0 - \boldsymbol{\eta}\|^2 \lesssim J/n^2.$$

As a conclusion, our analysis yield $\text{Var}_0[(\mathbf{F}_0 - \mathbf{A}\boldsymbol{\eta})^T\mathbf{U}\boldsymbol{\varepsilon}] \leq Q_2n^{-1}[J/n + \|f_0\|_{\mathcal{G}}^2h(J)^2]$ for some constant $Q_2 > 0$.

Let us take $\delta_{n,J} = \sqrt{J/n} + \|f_0\|_{\mathcal{G}}h(J)$ and note that $\delta_{n,J} \gg B_J$ for all $J \in \mathcal{J}$. Then using the tail bound $P(|Z| > z) \leq 2e^{-z^2/(2\sigma^2)}$ for $Z \sim N(0, \sigma^2)$ and $z \geq 0$, we deduce that

$$P_0 \left(|(\mathbf{F}_0 - \mathbf{A}\boldsymbol{\eta})^T \mathbf{U}\boldsymbol{\varepsilon}| > \frac{M\delta_{n,J} - B_J}{4} \right) \leq 2 \exp \left\{ -\frac{(M\delta_{n,J}/2)^2}{8Q_2 n^{-1} [J/n + \|f_0\|_{\mathcal{G}}^2 h(J)^2]} \right\},$$

and this is further bounded above by $2e^{-Q_3 n}$ for some constant $Q_3 > 0$ when n is large enough. It now remains to derive an equivalent exponential bound for the first term in (8.3) and this is accomplished via the Hanson-Wright inequality for quadratic forms (Theorem 1.1 of [17]), i.e., for $t \geq 0$,

$$P_0 (|\boldsymbol{\varepsilon}^T \mathbf{U}\boldsymbol{\varepsilon} - \mathbb{E}_0(\boldsymbol{\varepsilon}^T \mathbf{U}\boldsymbol{\varepsilon})| > t) \leq 2 \exp \left[-Q_4 \min \left(\frac{t^2}{K^4 \|\mathbf{U}\|_{\text{HS}}^2}, \frac{t}{K^2 \|\mathbf{U}\|_{(2,2)}} \right) \right], \quad (8.6)$$

where $Q_4 > 0$ is some constant, K is an upper bound for the Orlicz norm with Orlicz function $\psi_2(x) = e^{x^2} - 1$ of ε_i , which we can take $K = \sqrt{8/3}\sigma_0$ by direct calculations, and $\|\mathbf{U}\|_{\text{HS}}$ is the Hilbert-Schmidt or Frobenius norm of \mathbf{U} . We know that $\|\mathbf{U}\|_{(2,2)} < n^{-1}$, and $\|\mathbf{U}\|_{\text{HS}}^2 = \text{tr}(\mathbf{U}^2) = n^{-2} \text{tr}(\mathbf{I}_n - \mathbf{P}_{\mathbf{A}} + \mathbf{V}^2) = n^{-2}[n - J + \text{tr}(\mathbf{V}^2)]$ since $\mathbf{I}_n - \mathbf{P}_{\mathbf{A}}$ is idempotent, of rank $n - J$, and projects into the null space of \mathbf{A} . Then by using the cyclic permutation of the trace operator and Lemma 9.8,

$$\begin{aligned} \text{tr}(\mathbf{V}^2) &\leq \lambda_{\max} \{(\mathbf{A}^T \mathbf{A})^{-1}\} \text{tr} \{[\boldsymbol{\Omega} + (\mathbf{A}^T \mathbf{A})^{-1}]^{-1} (\mathbf{A}^T \mathbf{A})^{-1} [\boldsymbol{\Omega} + (\mathbf{A}^T \mathbf{A})^{-1}]^{-1}\} \\ &\leq \lambda_{\min}^{-1}(\mathbf{A}^T \mathbf{A}) \text{tr}(\boldsymbol{\Omega}^{-1}) \lesssim J/\lambda_{\mathbf{A},J}, \end{aligned}$$

where we deduce the second inequality by adding and subtracting $\boldsymbol{\Omega}$ to $(\mathbf{A}^T \mathbf{A})^{-1}$, and had used (8.1) with (3.1) for the last inequality. As a result, $\|\mathbf{U}\|_{\text{HS}}^2 \lesssim n^{-1}$. By collecting the intermediate results so far and substitute $t = (M\delta_{n,J} - B_J)/2$ into (8.6), we can conclude that the first term in (8.3) is bounded by $2e^{-Q_5 J}$ for some constant $Q_5 > 0$. Together with the normal tail bound above,

$$P_0(|\hat{\sigma}_J^2 - \sigma_0^2| > M\delta_{n,J}) \leq 2e^{-Q_5 J} + 2^{-Q_3 n} \lesssim e^{-QJ},$$

for some constant $Q > 0$ as $n \rightarrow \infty$. The statement for $\hat{\sigma}_J$ is then implied by

$$\left| \sqrt{\hat{\sigma}_J^2} - \sqrt{\sigma_0^2} \right| = \frac{|\hat{\sigma}_J^2 - \sigma_0^2|}{\sqrt{\hat{\sigma}_J^2} + \sqrt{\sigma_0^2}} \leq \frac{|\hat{\sigma}_J^2 - \sigma_0^2|}{\sigma_0}.$$

The claim that $\delta_{n,J} \rightarrow 0$ can be seen by $\delta_{n,J} \leq \sqrt{j_{\max}/n} + Rh(j_{\min}) = o[(\log n)^{-1/2}] + o(1)$ by definition of j_{\max} and the assumption $h(j_{\min}) = o(1)$. \square

Proof of Proposition 3.3. Let $j \in \mathcal{J}$ such that $j > j_n^*$. Then if $\hat{j}_n = j$, this implies that for the previous element $j - 1$, the corresponding sup-norm difference in posterior means must be greater than the $\tau\hat{\sigma}_{i,l_{1,i}}\sqrt{\log(i)/\lambda_{\mathbf{A},i}}$ threshold. It follows that using a union bound, we will have

$$P_0(\hat{j}_n = j) \leq \sum_{i \in \mathcal{J}: i \geq j} \left[P_0 \left(\|\tilde{f}_{j-1} - \tilde{f}_i\|_{\infty} > \tau\hat{\sigma}_{i,l_{1,i}}\sqrt{\frac{\log i}{\lambda_{\mathbf{A},i}}}, \hat{\sigma}_i \in \mathcal{U}_n \right) + P_0(\hat{\sigma}_i \notin \mathcal{U}_n) \right]. \quad (8.7)$$

The second term is $O(e^{-Qj})$ for some constant $Q > 0$ by virtue of Proposition 3.2. Now by the triangle inequality,

$$\|\tilde{f}_{j-1} - \tilde{f}_i\|_\infty \leq \|\tilde{f}_{j-1} - \mathbb{E}_0 \tilde{f}_{j-1}\|_\infty + \|\tilde{f}_i - \mathbb{E}_0 \tilde{f}_i\|_\infty + \|\mathbb{E}_0 \tilde{f}_{j-1} - \mathbb{E}_0 \tilde{f}_i\|_\infty.$$

Observe that $l_{1,i}$ is increasing in i by definition, $\lambda_{\mathbf{A},i}$ and $h(i)$ are decreasing in i by assumption. Then by another application of the triangle inequality and using Lemma 9.1 with $i > j - 1 \geq j_n^*$, the last term on the right hand side is bounded above by

$$\begin{aligned} \|\mathbb{E}_0 \tilde{f}_{j-1} - f_0\|_\infty + \|\mathbb{E}_0 \tilde{f}_i - f_0\|_\infty &\lesssim \frac{l_{1,j-1}}{\lambda_{\mathbf{A},j-1}} + \frac{l_{1,i}}{\lambda_{\mathbf{A},i}} + \|f_0\|_{\mathcal{G}}[h(j-1) + h(i)] \\ &\lesssim \|f_0\|_{\mathcal{G}} h(j_n^*) + \frac{l_{1,i}}{\lambda_{\mathbf{A},i}} \leq \xi l_{1,i} \sqrt{\frac{\log i}{\lambda_{\mathbf{A},i}}} \end{aligned}$$

when n is large enough. The last inequality follows from the definition of j_n^* given in (3.9) and $\xi > 0$ is some universal constant. Suppose n is large enough so that $\hat{\sigma}_i > \sigma_0/2$ over $\hat{\sigma}_i \in \mathcal{U}_n$. If we choose τ large enough so that $\tau > (2/\sigma_0)\xi$, then the first term on the right hand side of (8.7) is bounded above by

$$\begin{aligned} &\sum_{i \in \mathcal{J}: i \geq j} P_0 \left(\|\tilde{f}_{j-1} - \mathbb{E}_0 \tilde{f}_{j-1}\|_\infty > \frac{\tau(\sigma_0/2) - \xi}{2} l_{1,i} \sqrt{\frac{\log i}{\lambda_{\mathbf{A},i}}} \right) \\ &+ \sum_{i \in \mathcal{J}: i \geq j} P_0 \left(\|\tilde{f}_i - \mathbb{E}_0 \tilde{f}_i\|_\infty > \frac{\tau(\sigma_0/2) - \xi}{2} l_{1,i} \sqrt{\frac{\log i}{\lambda_{\mathbf{A},i}}} \right). \end{aligned} \quad (8.8)$$

Apply Lemma 9.3 by taking $x = \mu \log(i)$ for both the cases $J = j - 1$ and $J = i$ above, such that $i > j - 1$ and for any $2 < \mu \leq [\tau(\sigma_0/2) - \xi - 2G_1]^2 / (8G_2)$ when τ is large enough and $G_1, G_2 > 0$ some constants, it follows that (8.8) is bounded above by $2 \sum_{i \in \mathcal{J}: i \geq j} e^{-\mu \log i}$. Suppose n is large enough so that $j_n^* \geq 2$, then using the fact that $\sum_{k=i}^I k^{-\beta} \leq \int_{i-1}^I x^{-\beta} dx$ for $\beta > 0$, we deduce that

$$P_0(\hat{j}_n > j_n^*) = \sum_{j > j_n^*}^{j_{\max}} P_0(\hat{j}_n = j) \lesssim \sum_{j > j_n^*}^{j_{\max}} \sum_{i=j}^{j_{\max}} e^{-\mu \log i} + \sum_{j > j_n^*}^{j_{\max}} e^{-Qj} \lesssim \frac{1}{(j_n^*)^{\mu-2}}. \quad \square$$

In what follows, let $\mathbf{M} = (\mathbf{A}^T \mathbf{A} + \mathbf{\Omega}^{-1})^{-1}$ and $\mathbf{M}_0 = (\mathbf{A}^T \mathbf{A} + \mathbf{\Omega}^{-1})^{-1} \mathbf{A}^T \mathbf{A} (\mathbf{A}^T \mathbf{A} + \mathbf{\Omega}^{-1})^{-1}$.

Proof of Theorem 3.4. Let $\epsilon_n = l_{1,j_n^*} \sqrt{\log(j_n^*)} / \lambda_{\mathbf{A},j_n^*}$ and ξ some constant to be determined below, then by the law of total probability,

$$\begin{aligned} \mathbb{E}_0 \Pi(\|f - f_0\|_\infty > \xi \epsilon_n | \mathbf{Y}) &\leq \mathbb{E}_0 \Pi(\|f - f_0\|_\infty > \xi \epsilon_n | \mathbf{Y}) \mathbb{1}_{\{\hat{\sigma}_{\hat{j}_n} \in \mathcal{U}_n, \hat{j}_n \leq j_n^*\}} \\ &+ P_0(\hat{\sigma}_{\hat{j}_n} \notin \mathcal{U}_n, \hat{j}_n \leq j_n^*) + P_0(\hat{j}_n > j_n^*). \end{aligned} \quad (8.9)$$

To bound the first term, we use the triangle inequality to write

$$\|f - f_0\|_\infty \leq \|f - \tilde{f}_{\hat{j}_n}\|_\infty + \|\tilde{f}_{\hat{j}_n} - \tilde{f}_{j_n^*}\|_\infty + \|\tilde{f}_{j_n^*} - \mathbb{E}_0 \tilde{f}_{j_n^*}\|_\infty + \|\mathbb{E}_0 \tilde{f}_{j_n^*} - f_0\|_\infty. \quad (8.10)$$

By the definition of \hat{j}_n in (3.6) and intersecting with the event $\{\hat{\sigma}_{\hat{j}_n} \in \mathcal{U}_n, \hat{j}_n \leq j_n^*\}$, the second term is bounded above by $\tau(\sigma_0 + o(1))\epsilon_n \leq C_1\tau\epsilon_n$ for some constant $C_1 > 0$ in P_0 -probability. Applying Lemma 9.3 for $k = j_n^*$ and $x = \log j_n^*$, the third term is with at least $1 - (j_n^*)^{-1}$ P_0 -probability, bounded from above by $(G_1 + \sqrt{2G_2})\epsilon_n$ for some constants $G_1, G_2 > 0$. By Lemma 9.1 and the definition of j_n^* in (3.9), it follows that the last term is of the order $l_{1,j_n^*}/\lambda_{\mathbf{A},j_n^*} + \|f_0\|_{\mathcal{G}h(j_n^*)} \leq C_2\epsilon_n$ for some constant $C_2 > 0$ and when n is large enough.

Note that $U := f - \tilde{f}_{\hat{j}_n}$ is a centered Gaussian process under the conditional posterior with covariance kernel $\hat{\sigma}_{\hat{j}_n}^2 \mathbf{a}_{\hat{j}_n}(x)^T \mathbf{M} \mathbf{a}_{\hat{j}_n}(y)$ for $x, y \in [0, 1]$. In view of (8.10) and the bounds established in the previous paragraph, the first term in (8.9) is bounded above by $\mathbb{E}_0 \Pi [\|U\|_\infty > (\xi - C_1\tau - G_1 - \sqrt{2G_2} - C_2)\epsilon_n | \mathbf{Y}] \mathbb{1}_{\{\hat{\sigma}_{\hat{j}_n} \in \mathcal{U}_n, \hat{j}_n \leq j_n^*\}}$. Now by Lemma 9.2 and intersecting the event $\{\hat{\sigma}_{\hat{j}_n} \in \mathcal{U}_n, \hat{j}_n \leq j_n^*\}$, we have that $\mathbb{E}(\|U\|_\infty | \mathbf{Y}) \lesssim (\sigma_0 + o(1))l_{1,\hat{j}_n} [\log(\hat{j}_n)/\lambda_{\mathbf{A},\hat{j}_n}]^{1/2} \leq C_3\epsilon_n$ in P_0 -probability for some constant $C_3 > 0$, since $l_{1,J}$ is increasing in J by definition and $\lambda_{\mathbf{A},J}$ is decreasing in J by assumption.

Define $\nu^2 = \sup_{x \in [0,1]} \text{Var}[U(x) | \mathbf{Y}]$. Take $\xi > C_1\tau + G_1 + \sqrt{2G_2} + C_2 + C_3 =: \xi_0$. Then by the Borell's inequality (see Proposition A.2.1 of [22]), the first term in (8.9) is further bounded above by

$$\begin{aligned} & \mathbb{E}_0 \Pi [\|U\|_\infty - \mathbb{E}(\|U\|_\infty | \mathbf{Y}) > (\xi - \xi_0)\epsilon_n | \mathbf{Y}] \mathbb{1}_{\{\hat{\sigma}_{\hat{j}_n} \in \mathcal{U}_n, \hat{j}_n \leq j_n^*\}} \\ & \leq 2 \int_{\{\hat{\sigma}_{\hat{j}_n} \in \mathcal{U}_n, \hat{j}_n \leq j_n^*\}} \exp\{-(\xi - \xi_0)^2 \epsilon_n^2 / (2\nu^2)\} dP_0. \end{aligned}$$

Using the inequality $\mathbf{y}^T \mathbf{T} \mathbf{y} \leq \lambda_{\max}(\mathbf{T}) \|\mathbf{y}\|^2$ for any square matrix \mathbf{T} and (8.2), we know that under the event $\{\hat{\sigma}_{\hat{j}_n} \in \mathcal{U}_n, \hat{j}_n \leq j_n^*\}$,

$$\nu^2 \leq (\sigma_0^2 + o(1)) \lambda_{\max}(\mathbf{M}) l_{2,\hat{j}_n}^2 \lesssim l_{2,j_n^*}^2 / \lambda_{\mathbf{A},j_n^*}, \quad (8.11)$$

since $l_{2,J}$ is increasing in J by definition while $\lambda_{\mathbf{A},J}$ is decreasing in J by assumption. Therefore, $\epsilon_n^2/\nu^2 \gtrsim \log j_n^*$. As a result, the first term on the right hand side of (8.9) approaches 0 at the rate some power of $1/j_n^*$ when $\xi > \xi_0$ as $n \rightarrow \infty$.

For the second term in (8.9), we use a union bound and Proposition 3.2 to write

$$P_0(\hat{\sigma}_{\hat{j}_n} \notin \mathcal{U}_n, \hat{j}_n \leq j_n^*) \leq \sum_{j \leq j_n^*} P_0(\hat{\sigma}_j \notin \mathcal{U}_n, \hat{j}_n = j) \lesssim \sum_{j=j_{\min}}^{j_n^*} e^{-Qj} \lesssim e^{-Qj_{\min}}$$

for some constant $Q > 0$, and this tends to 0 as $n \rightarrow \infty$ by the definition of j_{\min} . It now remains to show that the last term in (8.9) goes to 0 at some power of $1/j_n^*$, and this is ensured by Proposition 3.3. \square

Proof of Theorem 4.2. We work instead with the complement

$$P_0(f_0 \notin \mathcal{C}_{\hat{j}_n}) = P_0 \left(\sup_{0 \leq x \leq 1} \frac{|f_0(x) - \tilde{f}_{\hat{j}_n}(x)|}{\sqrt{\tilde{V}_{\hat{j}_n}(x)}} > w_{\gamma, \hat{j}_n} \right), \quad (8.12)$$

and show that it is bounded above by $\gamma + o(1)$ for any $f_0 \in \mathcal{G} \cap \mathcal{F}$. Now by the triangle inequality, we have $|f_0(x) - \tilde{f}_{\hat{j}_n}(x)| \leq |f_0(x) - \mathbb{E}_0 \tilde{f}_{\hat{j}_n}(x)| + |\mathbb{E}_0 \tilde{f}_{\hat{j}_n}(x) - \tilde{f}_{\hat{j}_n}(x)|$. Then using the facts that $\sup(f + g) \leq \sup f + \sup g$ and $\sup fg = \sup f \sup g$, it follows that the right hand side of (8.12) is bounded above by

$$P_0 \left[\sup_x \sqrt{\frac{\text{Var}_0 \tilde{f}_{\hat{j}_n}(x)}{\tilde{V}_{\hat{j}_n}(x)}} \sup_x \frac{|\tilde{f}_{\hat{j}_n}(x) - \mathbb{E}_0 \tilde{f}_{\hat{j}_n}(x)|}{\sqrt{\text{Var}_0 \tilde{f}_{\hat{j}_n}(x)}} > w_{\gamma, \hat{j}_n} - \sup_x \frac{|f_0(x) - \mathbb{E}_0 \tilde{f}_{\hat{j}_n}(x)|}{\sqrt{\tilde{V}_{\hat{j}_n}(x)}} \right]$$

Now notice that \mathbf{M}_0 can be reexpressed as $\mathbf{M} - \mathbf{M}\mathbf{\Omega}^{-1}\mathbf{M}$ and consequently, $\mathbf{y}^T \mathbf{M}_0 \mathbf{y} \leq \mathbf{y}^T \mathbf{M} \mathbf{y}$ for any vector \mathbf{y} , i.e., $\mathbf{M} - \mathbf{M}_0$ is nonnegative definite. Take $\mathbf{y} = \mathbf{a}_{\hat{j}_n}(x)$ and the definition of the variances yield

$$\sup_{0 \leq x \leq 1} \sqrt{\frac{\text{Var}_0 \tilde{f}_{\hat{j}_n}(x)}{\tilde{V}_{\hat{j}_n}(x)}} \leq \frac{\sigma_0}{\hat{\sigma}_{\hat{j}_n}}.$$

As a result, the right hand side of (8.12) is bounded from above by

$$P_0 \left(\sup_x \frac{|\tilde{f}_{\hat{j}_n}(x) - \mathbb{E}_0 \tilde{f}_{\hat{j}_n}(x)|}{\sqrt{\text{Var}_0 \tilde{f}_{\hat{j}_n}(x)}} > w_{\gamma, \hat{j}_n} \frac{\hat{\sigma}_{\hat{j}_n}}{\sigma_0} \left\{ 1 - w_{\gamma, \hat{j}_n}^{-1} \sup_x \frac{|f_0(x) - \mathbb{E}_0 \tilde{f}_{\hat{j}_n}(x)|}{\sqrt{\tilde{V}_{\hat{j}_n}(x)}} \right\} \right). \quad (8.13)$$

By applying the inequality $\mathbf{y}^T \mathbf{T} \mathbf{y} \geq \lambda_{\min}(\mathbf{T}) \|\mathbf{y}\|^2$ for any square matrix \mathbf{T} , we obtain

$$\tilde{V}_{\hat{j}_n}(x) \geq \hat{\sigma}_{\hat{j}_n}^2 \lambda_{\min}(\mathbf{M}) \|\mathbf{a}_{\hat{j}_n}(x)\|^2 \gtrsim \hat{\sigma}_{\hat{j}_n}^2 \inf_{x \in [0,1]} \|\mathbf{a}_{\hat{j}_n}(x)\|^2 / \lambda_{\mathbf{A}, \hat{j}_n} \quad (8.14)$$

where we have used the lower bound in (8.2). In view of Lemma 9.1, the bias $|f_0(x) - \mathbb{E}_0 \tilde{f}_{\hat{j}_n}(x)| \lesssim l_{1, \hat{j}_n} / \lambda_{\mathbf{A}, \hat{j}_n} + Rh(\hat{j}_n)$ for any $f_0 \in \mathcal{G}$ and uniformly over $x \in [0, 1]$. From the definition of w_{γ, \hat{j}_n} given in (4.6) and the fact that $0 \leq \Phi(x) \leq 1$ for any x ,

$$\gamma = |\beta_{\hat{j}_n}| \pi^{-1} e^{-w_{\gamma, \hat{j}_n}^2 / 2} + 2[1 - \Phi(w_{\gamma, \hat{j}_n})] \geq |\beta_{\hat{j}_n}| \pi^{-1} e^{-w_{\gamma, \hat{j}_n}^2 / 2}, \quad (8.15)$$

and rearranging gives $w_{\gamma, \hat{j}_n} \geq \sqrt{2 \log(\gamma^{-1} \pi^{-1} |\beta_{\hat{j}_n}|)}$. Therefore by the second assumption $|\beta_{\hat{j}_n}| \gtrsim \hat{j}_n$, we will have $w_{\gamma, \hat{j}_n} \gtrsim [\log(\hat{j}_n)]^{1/2}$. Collecting all these results,

$$1 - w_{\gamma, \hat{j}_n}^{-1} \sup_x \frac{|f_0(x) - \mathbb{E}_0 \tilde{f}_{\hat{j}_n}(x)|}{\sqrt{\tilde{V}_{\hat{j}_n}(x)}} > 1 - C \frac{[\log(\hat{j}_n)]^{-1/2}}{\hat{\sigma}_{\hat{j}_n} \inf_{x \in [0,1]} \|\mathbf{a}_{\hat{j}_n}(x)\|} \left(\frac{l_{1, \hat{j}_n}}{\sqrt{\lambda_{\mathbf{A}, \hat{j}_n}}} + \sqrt{\lambda_{\mathbf{A}, \hat{j}_n}} h(\hat{j}_n) \right),$$

where $C > 0$ is some constant. Let us define \mathcal{E} to be the event in (8.13) but with the right hand side replaced by the lower bound above. We know that under the event $\mathcal{I} := \{j_n^*/\kappa \leq \hat{j}_n \leq j_n^*\}$ with the constant $\kappa \geq 1$ arising from the first assumption, the right hand side above is $1 + o_{P_0}(1)$ by the third assumption. Then by the law of total probability, we can further bound the right hand side of (8.13) from above by

$$P_0 \left(\mathcal{E} \cap \mathcal{I} \cap \{\hat{\sigma}_{\hat{j}_n} \in \mathcal{U}_n\} \right) + P_0(\hat{\sigma}_{\hat{j}_n} \notin \mathcal{U}_n, \mathcal{I}) + P_0(\mathcal{I}^c). \quad (8.16)$$

Keeping in mind the aforementioned established bounds and using (4.5) with $\boldsymbol{\mu} = \mathbb{E}_0 \mathbb{E}_j(\boldsymbol{\theta} | \mathbf{Y})$ and $\boldsymbol{\Sigma} = \text{Var}_0 \mathbb{E}_j(\boldsymbol{\theta} | \mathbf{Y})$, we can bound the first term by

$$\begin{aligned} & \sum_{j=j_n^*/\kappa}^{j_n^*} P_0 \left[\sup_{0 \leq x \leq 1} \frac{|f_0(x) - \mathbb{E}_0 \tilde{f}_{\hat{j}_n}(x)|}{\sqrt{\text{Var}_0 \tilde{f}_{\hat{j}_n}(x)}} > w_{\gamma, \hat{j}_n} \mid \hat{j}_n = j \right] P_0(\hat{j}_n = j) \\ & \leq \max_{j_n^*/\kappa \leq j \leq j_n^*} \left(\frac{|\beta_{0,j}|}{\pi} e^{-w_{\gamma,j}^2} + 2[1 - \Phi(w_{\gamma,j})] \right) \sum_{j=j_n^*/\kappa}^{j_n^*} P_0(\hat{j}_n = j) \leq \gamma \end{aligned}$$

as $n \rightarrow \infty$, since $|\beta_{0,j}| \leq |\beta_j|$ for $j \in [j_n^*/\kappa, j_n^*] \subset \mathcal{J}$ by the second assumption, the definition of $w_{\gamma,j}$ in (4.6), and the first assumption for the last sum.

To complete the proof, we bound the second term in (8.16) by

$$\sum_{j=j_n^*/\kappa}^{j_n^*} P_0(\hat{\sigma}_j \notin \mathcal{U}_n) \lesssim \sum_{j=j_n^*/\kappa}^{j_n^*} e^{-Qj} \lesssim e^{-Qj_n^*/\kappa},$$

in view of Proposition 3.2 for some constant $Q > 0$. The last term of (8.16) is $o(1)$ follows from the first assumption. \square

Proof of Corollary 4.3. Let us define the event $\mathcal{I} := \{j_n^*/\kappa \leq \hat{j}_n \leq j_n^*\}$. For the upper bound, we have by the tail probability estimate of a standard normal $1 - \Phi(x) \leq e^{-x^2/2}$ that

$$\gamma = |\beta_{\hat{j}_n}| \pi^{-1} e^{-w_{\gamma, \hat{j}_n}^2/2} + 2[1 - \Phi(w_{\gamma, \hat{j}_n})] \leq e^{-w_{\gamma, \hat{j}_n}^2/2} (|\beta_{\hat{j}_n}| \pi^{-1} + 2).$$

By rearranging we see that $w_{\gamma, \hat{j}_n} \leq \sqrt{2 \log(|\beta_{\hat{j}_n}| \gamma^{-1} \pi^{-1} + 2\gamma^{-1})}$. Since $|\beta_{\hat{j}_n}| \lesssim \hat{j}_n$ by assumption, it follows that $w_{\gamma, \hat{j}_n} \lesssim (\log \hat{j}_n)^{1/2}$. Using the same bound as in (8.11), it holds that $\tilde{V}_{\hat{j}_n}(x) \lesssim \hat{\sigma}_{\hat{j}_n}^2 \|\mathbf{a}_{\hat{j}_n}(x)\|^2 / \lambda_{\mathbf{A}, \hat{j}_n}$. Therefore under the event $\mathcal{I} \cap \{\hat{\sigma}_{\hat{j}_n} \in \mathcal{U}_n\}$, we will have $r(x) \lesssim \|\mathbf{a}_{j_n^*/\kappa}(x)\| \sqrt{\log(j_n^*) / \lambda_{\mathbf{A}, j_n^*/\kappa}}$ for any $x \in [0, 1]$, since $\|\mathbf{a}_J(x)\|$ is increasing in J by definition and $\lambda_{\mathbf{A}, J}$ is decreasing in J by assumption.

Now $w_{\gamma, \hat{j}_n} \gtrsim (\log \hat{j}_n)^{1/2}$ was established in the proof of Theorem 4.2 because of (8.15). Using the same argument in (8.14), we will have $\tilde{V}_{\hat{j}_n}(x) \gtrsim \hat{\sigma}_{\hat{j}_n}^2 \|\mathbf{a}_{\hat{j}_n}(x)\|^2 / \lambda_{\mathbf{A}, \hat{j}_n}$. Therefore under the event $\mathcal{I} \cap \{\hat{\sigma}_{\hat{j}_n} \in \mathcal{U}_n\}$, we deduce $r(x) \gtrsim \|\mathbf{a}_{j_n^*/\kappa}(x)\| \sqrt{\log(j_n^*) / \lambda_{\mathbf{A}, j_n^*/\kappa}}$ for n large enough. The rest of the proof can be completed by the law of total probability as in (8.16) in the proof of Theorem 4.2, with \mathcal{E} the complement of the event to be proven, i.e., $r(x)$ does not lie in between the upper and lower bounds stated in the corollary for some $x \in [0, 1]$. \square

It is now convenient to note down the specialization of (8.2) to the B-splines and wavelet cases for any $J \in \mathcal{J}$:

$$J/n \lesssim \lambda_{\min} \{(\mathbf{B}^T \mathbf{B} + \mathbf{\Omega}^{-1})^{-1}\} \leq \lambda_{\max} \{(\mathbf{B}^T \mathbf{B} + \mathbf{\Omega}^{-1})^{-1}\} \lesssim J/n, \quad (8.17)$$

$$n^{-1} \lesssim \lambda_{\min} \{(\mathbf{\Psi}^T \mathbf{\Psi} + \mathbf{\Omega}^{-1})^{-1}\} \leq \lambda_{\max} \{(\mathbf{\Psi}^T \mathbf{\Psi} + \mathbf{\Omega}^{-1})^{-1}\} \lesssim n^{-1}. \quad (8.18)$$

Proof of Lemmas 5.1 and 6.1. For the reader's convenience, let us reproduce the arc length formula for general basis functions (4.7):

$$|\beta_J| = \int_0^1 \frac{[\mathbf{a}_J(x)^T \mathbf{M} \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{M} \mathbf{a}_J(x)]^{1/2}}{[\mathbf{a}_J(x)^T \mathbf{M} \mathbf{a}_J(x)]^{3/2}} dx,$$

where $\mathbf{R} := \dot{\mathbf{a}}_J(x) \mathbf{a}_J(x)^T - \mathbf{a}_J(x) \dot{\mathbf{a}}_J(x)^T$. We start with the lower bound of $|\beta_J|$, which requires a lower bound for the numerator and an upper bound for the denominator. By (8.2), we see that $\mathbf{a}_J(x)^T \mathbf{M} \mathbf{a}_J(x) \leq \lambda_{\max}(\mathbf{M}) l_{2,J}^2$ for any x . Therefore,

$$|\beta_J| \geq \lambda_{\max}^{-3/2}(\mathbf{M}) l_{2,J}^{-3} \int_0^1 [\mathbf{a}_J(x)^T \mathbf{M} \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{M} \mathbf{a}_J(x)]^{1/2} dx.$$

For B-splines, we know that $\lambda_{\max}(\mathbf{M}) \lesssim J/n$ by (8.17) and $l_{2,J} \leq 1$ by Lemma 9.6. Then in view of (9.4) of Lemma 9.5 with $\mathbf{\Sigma} = \mathbf{M}$ in the integrand, we will obtain

$$|\beta_J| \gtrsim (n/J)^{3/2} \lambda_{\min}(\mathbf{M})^{1/2} \lambda_{\min}(\mathbf{M}) J \gtrsim J,$$

where we have appealed again to (8.17) to lower bound the minimum eigenvalue. The wavelet case is similar, but now $\lambda_{\max}(\mathbf{M}) \lesssim n^{-1}$ by (8.18) and $l_{2,J} \lesssim 2^{J/2}$ by Lemma 9.7. By invoking (9.5) of Lemma 9.5 with $\mathbf{\Sigma} = \mathbf{M}$, we see that

$$|\beta_J| \gtrsim n^{3/2} 2^{-3J/2} \lambda_{\min}(\mathbf{M})^{1/2} \lambda_{\min}(\mathbf{M}) 2^{5J/2} \gtrsim 2^J.$$

For the upper bound of $|\beta_J|$, we need a lower bound for the denominator and (8.2) gives $\mathbf{a}_J(x)^T \mathbf{M} \mathbf{a}_J(x) \geq \lambda_{\min}(\mathbf{M}) \inf_{x \in [0,1]} \|\mathbf{a}_J(x)\|^2$. Thus,

$$|\beta_J| \leq \lambda_{\min}^{-3/2}(\mathbf{M}) \left(\inf_{x \in [0,1]} \|\mathbf{a}_J(x)\| \right)^{-3} \int_0^1 [\mathbf{a}_J(x)^T \mathbf{M} \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{M} \mathbf{a}_J(x)]^{1/2} dx.$$

For B-splines, we have $\lambda_{\min}(\mathbf{M}) \gtrsim J/n$ by (8.17) and $\inf_{x \in [0,1]} \|\mathbf{a}_J(x)\| \geq q^{-1}$ by Lemma 9.6. Subsequently by (9.4) of Lemma 9.5,

$$|\beta_J| \lesssim (n/J)^{3/2} \lambda_{\max}(\mathbf{M})^{1/2} \lambda_{\max}(\mathbf{M}) J \lesssim J,$$

again by invoking (8.17). For wavelets however, we have $\lambda_{\min}(\mathbf{M}) \gtrsim n^{-1}$ by (8.18) and $\inf_{x \in [0,1]} \|\mathbf{a}_J(x)\| \gtrsim 2^{J/2}$ by Lemma 9.7. Consequently by (9.5) of Lemma 9.5,

$$|\beta_J| \lesssim n^{3/2} 2^{-3J/2} \lambda_{\max}(\mathbf{M})^{1/2} \lambda_{\max}(\mathbf{M}) 2^{5J/2} \lesssim 2^J,$$

by appealing once more to (8.18). □

Proof of Lemmas 5.2 and 6.2. The expression $|\beta_{0,J}|$ has the same functional form as $|\beta_J|$ in (4.7) except that \mathbf{M} is replaced by \mathbf{M}_0 , giving (for ease of reference)

$$|\beta_{0,J}| = \int_0^1 \frac{[\mathbf{a}_J(x)^T \mathbf{M}_0 \mathbf{R}^T \mathbf{M}_0 \mathbf{R} \mathbf{M}_0 \mathbf{a}_J(x)]^{1/2}}{[\mathbf{a}_J(x)^T \mathbf{M}_0 \mathbf{a}_J(x)]^{3/2}} dx,$$

where $\mathbf{R} := \dot{\mathbf{a}}_J(x) \mathbf{a}_J(x)^T - \mathbf{a}_J(x) \dot{\mathbf{a}}_J(x)^T$. We then need to upper bound the numerator and lower bound the denominator. Before we begin, note that $\mathbf{M}_0 = \mathbf{M} - \mathbf{M} \boldsymbol{\Omega}^{-1} \mathbf{M}$ and since $\boldsymbol{\Omega}$ is positive definite, we have $\mathbf{M}_0 \leq \mathbf{M}$.

By our assumption on the prior matrix, we have $\lambda_{\max}(\boldsymbol{\Omega}^{-1}) \leq c_1^{-1}$. Then

$$\mathbf{a}_J(x)^T \mathbf{M} \boldsymbol{\Omega}^{-1} \mathbf{M} \mathbf{a}_J(x) \leq \lambda_{\max}(\boldsymbol{\Omega}^{-1}) \lambda_{\max}(\mathbf{M})^2 \|\mathbf{a}_J(x)\|^2 \lesssim \lambda_{\max}(\mathbf{M})^2 l_{2,J}^2;$$

while $\mathbf{a}_J(x)^T \mathbf{M} \mathbf{a}_J(x) \geq \lambda_{\min}(\mathbf{M}) \inf_{x \in [0,1]} \|\mathbf{a}_J(x)\|^2$. Putting these two together and using the relation between \mathbf{M}_0 and \mathbf{M} , we conclude that

$$\begin{aligned} \mathbf{a}_J(x)^T \mathbf{M}_0 \mathbf{a}_J(x) &= \mathbf{a}_J(x)^T \mathbf{M} \mathbf{a}_J(x) \left(1 - \frac{\mathbf{a}_J(x)^T \mathbf{M} \boldsymbol{\Omega}^{-1} \mathbf{M} \mathbf{a}_J(x)}{\mathbf{a}_J(x)^T \mathbf{M} \mathbf{a}_J(x)} \right) \\ &\geq \mathbf{a}_J(x)^T \mathbf{M} \mathbf{a}_J(x) \left[1 - \xi_3 \frac{\lambda_{\max}(\mathbf{M})^2 l_{2,J}^2}{\lambda_{\min}(\mathbf{M}) \inf_{x \in [0,1]} \|\mathbf{a}_J(x)\|^2} \right] \end{aligned}$$

for some constant $\xi_3 > 0$. For B-splines, use (8.17) and Lemma 9.6 to deduce that the expression inside the square brackets reduces to $1 - \xi_3(J/n)$. For the wavelets, use instead (8.18) and Lemma 9.7 to see that it is now $1 - \xi_3 n^{-1}$.

We now deal with the numerator. Now use $\mathbf{M}_0 \leq \mathbf{M}$ on the middle \mathbf{M}_0 and substituting $\mathbf{M}_0 = \mathbf{M} - \mathbf{M} \boldsymbol{\Omega}^{-1} \mathbf{M}$ on the other two \mathbf{M}_0 's flanking the quadratic form, we know that $\mathbf{a}_J(x)^T \mathbf{M}_0 \mathbf{R}^T \mathbf{M}_0 \mathbf{R} \mathbf{M}_0 \mathbf{a}_J(x)$ will be bounded above by $\mathbf{a}_J(x)^T \mathbf{M} \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{M} \mathbf{a}_J(x)$ times

$$1 - \frac{2\mathbf{a}_J(x)^T \mathbf{M} \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{M} \boldsymbol{\Omega}^{-1} \mathbf{M} \mathbf{a}_J(x)}{\mathbf{a}_J(x)^T \mathbf{M} \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{M} \mathbf{a}_J(x)} + \frac{\mathbf{a}_J(x)^T \mathbf{M} \boldsymbol{\Omega}^{-1} \mathbf{M} \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{M} \boldsymbol{\Omega}^{-1} \mathbf{M} \mathbf{a}_J(x)}{\mathbf{a}_J(x)^T \mathbf{M} \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{M} \mathbf{a}_J(x)}. \quad (8.19)$$

By Corollary 9.5 with $\boldsymbol{\Sigma} = \mathbf{M}$, we have for $\mu(J) = J^2$ (B-splines) or $\mu(J) = 2^{5J}$ (CDV wavelets) that

$$\mu(J) \lambda_{\min}(\mathbf{M})^3 \lesssim \mathbf{a}_J(x)^T \mathbf{M} \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{M} \mathbf{a}_J(x) \lesssim \mu(J) \lambda_{\max}(\mathbf{M})^3. \quad (8.20)$$

By another application of Lemma 9.4 with $\boldsymbol{\Sigma} = \mathbf{M} \boldsymbol{\Omega}^{-1} \mathbf{M}$, we can further deduce that $\mathbf{a}_J(x)^T \mathbf{M} \boldsymbol{\Omega}^{-1} \mathbf{M} \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{M} \boldsymbol{\Omega}^{-1} \mathbf{M} \mathbf{a}_J(x) \lesssim \mu(J) \lambda_{\max}(\mathbf{M}) \lambda_{\max}(\mathbf{M} \boldsymbol{\Omega}^{-1} \mathbf{M})^2$. Note that since $\mathbf{M} \boldsymbol{\Omega}^{-1} \mathbf{M}$ is positive definite, we have by the sub-multiplicative property of the $\|\cdot\|_{(2,2)}$ -norm that $\lambda_{\max}(\mathbf{M} \boldsymbol{\Omega}^{-1} \mathbf{M}) = \|\mathbf{M} \boldsymbol{\Omega}^{-1} \mathbf{M}\|_{(2,2)} \leq \|\mathbf{M}\|_{(2,2)}^2 \|\boldsymbol{\Omega}^{-1}\|_{(2,2)} = \lambda_{\max}(\mathbf{M})^2 \lambda_{\max}(\boldsymbol{\Omega}^{-1})$. In conjunction with (8.20), the second ratio in (8.19) is then bounded up to some constant multiple by $\lambda_{\max}(\mathbf{M})^5 \lambda_{\max}(\boldsymbol{\Omega}^{-1})^2 \lambda_{\min}^{-3}(\mathbf{M})$. For B-splines, this is further bounded by $\xi_2(J/n)^2$ for some constant $\xi_2 > 0$ in view of (8.17); while this will be $\xi_2 n^{-2}$ for the wavelet case in view of (8.18).

It now remains to bound the first ratio in (8.19). The numerator of this ratio consists of a quadratic form that is asymmetric in the sense that the $\mathbf{\Omega}^{-1}$ on the right flank is matched by \mathbf{M} and not $\mathbf{\Omega}^{-1}$, and hence inequality such as Corollary 9.5 is not directly applicable. To work around this, we use Lemma 9.8 by exploiting the cyclic permutation of the trace operator to rearrange the order of these matrices, and expelling the asymmetric matrices out so that the reduced quadratic form is symmetric. Here goes our argument. The numerator (a scalar quantity) is equal to $\text{tr}[\mathbf{a}_J(x)^T \mathbf{M} \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{M} \mathbf{\Omega}^{-1} \mathbf{M} \mathbf{a}_J(x)]$ and by rearrangement and twice application of Lemma 9.8 is

$$\begin{aligned} \text{tr} [\mathbf{\Omega}^{-1} \mathbf{M} \mathbf{a}_J(x) \mathbf{a}_J(x)^T \mathbf{M} \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{M}] &\geq \lambda_{\min}(\mathbf{\Omega}^{-1}) \text{tr} [\mathbf{M} \mathbf{a}_J(x) \mathbf{a}_J(x)^T \mathbf{M} \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{M}] \\ &\geq \lambda_{\min}(\mathbf{\Omega}^{-1}) \lambda_{\min}(\mathbf{M}) \mathbf{a}_J(x)^T \mathbf{M} \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{M} \mathbf{a}_J(x) \\ &\gtrsim \lambda_{\min}(\mathbf{\Omega}^{-1}) \lambda_{\min}(\mathbf{M})^4 \mu(J), \end{aligned}$$

where the last inequality follows from the lower bound of (8.20) above. Then in view of the upper bound in (8.20), we see that the first ratio of (8.19) is greater up to some constant than $\lambda_{\min}(\mathbf{\Omega}^{-1}) \lambda_{\min}(\mathbf{M})^4 \lambda_{\max}^{-3}(\mathbf{M})$. For B-splines, this is greater than $\xi_1(J/n)$ as a result of (8.17); and for the wavelets, it will be $\xi_1 n^{-1}$ by (8.18). We will obtain our result by combining the upper and lower bounds established for the two ratios, and also noting that the integral integrates to 1 since our bounds are uniform in $x \in [0, 1]$. \square

Proof of Lemma 5.3. Let $j \in \mathcal{J} = [j_{\min}, j_{\max}] \cap \mathbb{N}$ such that $j < j_n^*/\kappa$ for some constant $\kappa \in \mathbb{N}$ to be chosen below. By the definition of \hat{j}_n in (3.6), it follows that

$$P_0(\hat{j}_n = j) \leq P_0 \left(\|\tilde{f}_j - \tilde{f}_{j_n^*}\|_{\infty} \leq \tau \hat{\sigma}_{j_n^*} \sqrt{\frac{j_n^* \log j_n^*}{n}}, \hat{\sigma}_{j_n^*} \in \mathcal{U}_n \right) + P_0(\hat{\sigma}_{j_n^*} \notin \mathcal{U}_n). \quad (8.21)$$

The second term is $O(e^{-Qj_n^*})$ for some constant $Q > 0$ in view of Proposition 3.2. For the first term, we apply the reverse triangle inequality twice to obtain

$$\|\tilde{f}_j - \tilde{f}_{j_n^*}\|_{\infty} \geq \|\mathbf{E}_0 \tilde{f}_j - f_0\|_{\infty} - \|\mathbf{E}_0 \tilde{f}_{j_n^*} - f_0\|_{\infty} - \|\tilde{f}_j - \mathbf{E}_0 \tilde{f}_j - \tilde{f}_{j_n^*} + \mathbf{E}_0 \tilde{f}_{j_n^*}\|_{\infty}.$$

Observe that $\mathbf{E}_0 \tilde{f}_j(x) = \sum_{k=1}^j \mathbf{E}_0 \mathbf{E}(\theta_k | \mathbf{Y}) B_{k,q}(x)$ is a polynomial spline of order q with quasi-uniform knots in $\mathcal{T}_n := \{0 = t_0 < t_1 < \dots < t_N < t_{N+1} = 1\}$. Let us denote $\Delta_{\mathcal{T}_n} := \max_{1 \leq i \leq N+1} (t_i - t_{i-1})$ to be the corresponding maximum knot increment. Since $f_0 \in \mathcal{F}^{\alpha}$ (in view of (5.9)), the first term above is bounded below by

$$\begin{aligned} \inf_{p \in \mathcal{P}_q(\mathcal{T}_n)} \|p - f_0\|_{\infty} &\geq \delta_q \|f_0\|_{\mathcal{H}^{\alpha}([0,1])} \Delta_{\mathcal{T}_n}^{\alpha} \gtrsim \delta_q \|f_0\|_{\mathcal{H}^{\alpha}([0,1])} j^{-\alpha} \\ &\gtrsim \delta_q (\kappa/2)^{\alpha} \|f_0\|_{\mathcal{H}^{\alpha}([0,1])} (j_n^*/2)^{-\alpha} \geq C_1 \delta_q \kappa^{\alpha} \sqrt{j_n^* \log(j_n^*)/n}, \end{aligned}$$

for some constant $C_1 > 0$. Here, the second inequality follows from the fact $\Delta_{\mathcal{T}_n} \asymp N^{-1} \asymp j^{-1}$ since the spacings between knots are of the same order and they sum to one. The third and last inequalities follow from the assumption $j < j_n^*/\kappa$ and (3.9), where $\|f_0\|_{\mathcal{H}^{\alpha}([0,1])} i^{-\alpha} > \sqrt{i \log(i)/n}$ if $i < j_n^*$. By Lemma 9.1 restricted to B-splines bases and in view of the

definition of j_n^* given in (3.9), it holds that $\|\mathbb{E}_0 \tilde{f}_{j_n^*} - f_0\|_\infty \lesssim j_n^*/n + \|f_0\|_{\mathcal{H}^\alpha([0,1])} (j_n^*)^{-\alpha} \leq C_2 \sqrt{j_n^* \log(j_n^*)/n}$ for some constant $C_2 > 0$. Suppose n is large enough so that $\hat{\sigma}_{j_n^*} \leq 2\sigma_0$ for $\hat{\sigma}_{j_n^*} \in \mathcal{U}_n$. Then by combining the bounds established, the first term on the right hand side of (8.21) is bounded above by

$$\begin{aligned} & P_0 \left[\|\tilde{f}_j - \mathbb{E}_0 \tilde{f}_j - \tilde{f}_{j_n^*} + \mathbb{E}_0 \tilde{f}_{j_n^*}\|_\infty > (C_1 \delta_q \kappa^\alpha - C_2 - 2\sigma_0 \tau) \sqrt{\frac{j_n^* \log j_n^*}{n}} \right] \\ & \leq P_0 \left[\|\tilde{f}_j - \mathbb{E}_0 \tilde{f}_j\|_\infty > \frac{\sqrt{\kappa}(C_1 \delta_q \kappa^\alpha - C_2 - 2\sigma_0 \tau)}{2} \sqrt{\frac{j \log j}{n}} \right] \\ & \quad + P_0 \left[\|\tilde{f}_{j_n^*} - \mathbb{E}_0 \tilde{f}_{j_n^*}\|_\infty > \frac{\sqrt{\kappa}(C_1 \delta_q \kappa^\alpha - C_2 - 2\sigma_0 \tau)}{2} \sqrt{\frac{j \log j}{n}} \right], \end{aligned}$$

in view of the triangle inequality and since $j < j_n^*/\kappa$. We then take the constant $\kappa \in \mathbb{N}$, which depends only on δ_q, α and τ , to be large enough so that $C_1 \delta_q \kappa^\alpha - C_2 - 2\sigma_0 \tau > 0$. In addition, we take n large enough so that $j_{\min} \geq 2$. Applying Lemma 9.3 by letting $x = \mu \log j$ for both terms $J = j$ and $J = j_n^*$ such that $1 < \mu \leq [\sqrt{\kappa}(C_1 \delta_q \kappa^\alpha - C_2 - 2\sigma_0 \tau) - 2G_1]^2 / (8G_2)$ with $G_1, G_2 > 0$ some constants, we can then conclude that

$$P_0(\hat{j}_n < j_n^*/\kappa) = \sum_{j=j_{\min}}^{\lfloor j_n^*/\kappa \rfloor} P_0(\hat{j}_n = j) \lesssim \sum_{j=j_{\min}}^{j_{\max}} e^{-\mu \log j} + \sum_{j=j_{\min}}^{\lfloor j_n^*/\kappa \rfloor} e^{-Qj_n^*} \lesssim \frac{1}{j_{\min}^{\mu-1}} + j_n^* e^{-Qj_n^*},$$

where we have used the relation $\sum_{k=i}^{\infty} k^{-\beta} \leq \int_{i-1}^{\infty} x^{-\beta} dx$ for $\beta > 0$ in the last inequality. Therefore, the right hand side tends to 0 since $j_{\min} = (n/\log n)^{1/(2q+1)} \rightarrow \infty$ and $j_n^* e^{-Qj_n^*} \rightarrow 0$ as $n \rightarrow \infty$. \square

9 Appendix

Lemma 9.1. *For any $x \in [0, 1]$, $J \in \mathcal{J}$, and uniformly over $f_0 \in \mathcal{G}$,*

$$|\mathbb{E}_0 \mathbb{E}_J[f(x)|\mathbf{Y}] - f_0(x)| \lesssim l_{1,J}/\lambda_{\mathbf{A},J} + \|f_0\|_{\mathcal{G}} h(J).$$

Proof. By definition (3.8), we know that for any $f_0 \in \mathcal{G}$, there exists a $\boldsymbol{\theta}_0$ such that

$$|\mathbb{E}_0 \tilde{f}_J(x) - f_0(x)| \leq |\mathbb{E}_0 \tilde{f}_J(x) - \mathbf{a}_J(x)^T \boldsymbol{\theta}_0| + C_0 \|f_0\|_{\mathcal{G}} h(J).$$

In view of (3.7), the first term is $|\mathbf{a}_J(x)^T \mathbf{M}[\mathbf{A}^T \mathbf{F}_0 + \boldsymbol{\Omega}^{-1} \boldsymbol{\eta} - \mathbf{M}^{-1} \boldsymbol{\theta}_0]|$. By Hölder's inequality and the sub-multiplicative property of the $\|\cdot\|_{(1,1)}$ and $\|\cdot\|_{(\infty,\infty)}$ -norms, this first term can be bounded above by

$$\begin{aligned} & \|\mathbf{a}_J(x)^T \mathbf{M} \mathbf{A}^T\|_1 \|\mathbf{F}_0 - \mathbf{A} \boldsymbol{\theta}_0\|_\infty + \|\mathbf{a}_J(x)\|_1 \|\mathbf{M} \boldsymbol{\Omega}^{-1} (\boldsymbol{\eta} - \boldsymbol{\theta}_0)\|_\infty \\ & \leq l_{1,J} \|\mathbf{M}\|_{(1,1)} \|\mathbf{A}^T\|_{(1,1)} \|f_0 - \mathbf{a}_J(\cdot)^T \boldsymbol{\theta}_0\|_\infty + l_{1,J} \|\mathbf{M}\|_{(\infty,\infty)} \|\boldsymbol{\Omega}^{-1}\|_{(\infty,\infty)} (\|\boldsymbol{\eta}\|_\infty + \|\boldsymbol{\theta}_0\|_\infty). \end{aligned}$$

By our prior assumption, $\|\boldsymbol{\eta}\|_\infty = O(1)$. By definition of \mathcal{G} , $\|\boldsymbol{\theta}_0\|_\infty = O(1)$ and $\|f_0 - \mathbf{a}_J(\cdot)^T \boldsymbol{\theta}_0\|_\infty \leq C_0 \|f_0\|_{\mathcal{G}} h(J)$. Using Lemma 9.9 with $g(x) = x^{-1}$, we have $\|\boldsymbol{\Omega}^{-1}\|_{(\infty, \infty)} = O(1)$ and $\|\mathbf{M}\|_{(\infty, \infty)} \lesssim \lambda_{\mathbf{A}, J}^{-1}$ by exploiting the bandedness of both $\mathbf{A}^T \mathbf{A}$ and $\boldsymbol{\Omega}^{-1}$. Now since \mathbf{M} is symmetric, its (absolute value of) row and column sums are equal. Thus, $\|\mathbf{M}\|_{(1, 1)}$ (max of absolute value of column sums) is equal to $\|\mathbf{M}\|_{(\infty, \infty)}$ (max of absolute value of row sums) and this is $O(\lambda_{\mathbf{A}, J}^{-1})$ as established previously. Furthermore, $\|\mathbf{A}^T\|_{(1, 1)} = \max_{1 \leq i \leq n} \sum_{j=1}^J |a_j(X_i)| \leq l_{1, J}$. Collecting these results together yield

$$|\mathbb{E}_0 \tilde{f}_J(x) - f_0(x)| \lesssim l_{1, J} / \lambda_{\mathbf{A}, J} + l_{1, J}^2 \lambda_{\mathbf{A}, J}^{-1} \|f_0\|_{\mathcal{G}} h(J) + \|f_0\|_{\mathcal{G}} h(J) \lesssim l_{1, J} / \lambda_{\mathbf{A}, J} + \|f_0\|_{\mathcal{G}} h(J),$$

since $l_{1, J} \leq \sqrt{\lambda_{\mathbf{A}, J}}$ by the assumption in (3.4). \square

Lemma 9.2. *For any $2 \leq J \leq n$,*

$$\mathbb{E}[\|f - \mathbb{E}_J(f|\mathbf{Y})\|_\infty | \mathbf{Y}] \lesssim \hat{\sigma}_J l_{1, J} \sqrt{\log(J) / \lambda_{\mathbf{A}, J}}.$$

Proof. For any $x \in [0, 1]$, $[f(x) - \tilde{f}_J(x) | \mathbf{Y}]$ is equal in distribution as $\hat{\sigma}_J \mathbf{a}_J(x)^T (\mathbf{A}^T \mathbf{A} + \boldsymbol{\Omega}^{-1})^{-1/2} \mathbf{Z}$ for $\mathbf{Z} \sim N_J(\mathbf{0}, \mathbf{I}_J)$. Then by Hölder's inequality,

$$\mathbb{E}[\|f - \tilde{f}_J\|_\infty | \mathbf{Y}] \leq \hat{\sigma}_J l_{1, J} \|(\mathbf{A}^T \mathbf{A} + \boldsymbol{\Omega}^{-1})^{-1/2}\|_{(\infty, \infty)} \mathbb{E}\|\mathbf{Z}\|_\infty.$$

Utilizing Lemma 9.9 with $g(x) = x^{-1/2}$ and the bandedness of both $\mathbf{A}^T \mathbf{A}$ and $\boldsymbol{\Omega}^{-1}$, we deduce that $\|(\mathbf{A}^T \mathbf{A} + \boldsymbol{\Omega}^{-1})^{-1/2}\|_{(\infty, \infty)} \lesssim \lambda_{\mathbf{A}, J}^{-1/2}$. Then by Lemma 2.3.4 of [5], it follows that $\mathbb{E}\|\mathbf{Z}\|_\infty \lesssim \sqrt{\log J}$ for $J \geq 2$, and the upper bound is proved by multiplying all the bounds established. \square

Lemma 9.3. *For any $2 \leq J \leq n$ and $x \geq 0$, there exist constants $G_1, G_2 > 0$ such that*

$$P_0 \left(\|\mathbb{E}_J(f|\mathbf{Y}) - \mathbb{E}_0 \mathbb{E}_J(f|\mathbf{Y})\|_\infty \geq G_1 l_{1, J} \sqrt{\frac{\log J}{\lambda_{\mathbf{A}, J}}} + l_{2, J} \sqrt{2G_2 \frac{x}{\lambda_{\mathbf{A}, J}}} \right) \leq e^{-x}. \quad (9.1)$$

Proof. By our global assumption on the true model, $\tilde{f}_J - \mathbb{E}_0 \tilde{f}_J$ is under P_0 a mean 0 Gaussian process, such that its variance function is

$$\text{Var}_0 \tilde{f}_J(x) = \sigma_0^2 \mathbf{a}_J(x)^T \mathbf{M}_0 \mathbf{a}_J(x) \leq \sigma_0^2 \lambda_{\max}(\mathbf{M}) \|\mathbf{a}_J(x)\|^2 \leq G_2 l_{2, J}^2 / \lambda_{\mathbf{A}, J}$$

for any $x \in [0, 1]$ since $\mathbf{M}_0 \leq \mathbf{M}$, the fact $\mathbf{y}^T \mathbf{T} \mathbf{y} \leq \lambda_{\max}(\mathbf{T}) \|\mathbf{y}\|^2$ for any square matrix \mathbf{T} , and the last inequality from (8.2). By the Borell's inequality (see Proposition A.2.1 from [22] or Theorem 2.5.8 in [5]), we have

$$P_0 \left(\|\tilde{f}_J - \mathbb{E}_0 \tilde{f}_J\|_\infty \geq \mathbb{E}_0 \|\tilde{f}_J - \mathbb{E}_0 \tilde{f}_J\|_\infty + \sqrt{2G_2 l_{2, J}^2 \lambda_{\mathbf{A}, J}^{-1} x} \right) \leq e^{-x}. \quad (9.2)$$

Define $\boldsymbol{\eta} = \mathbf{A}^T \boldsymbol{\varepsilon}$. Then $\tilde{f}_J(x) - \mathbb{E}_0 \tilde{f}_J(x)$ can be written as $\mathbf{a}_J(x)^T \mathbf{M} \boldsymbol{\eta}$ for any $x \in [0, 1]$, and it follows from Hölder's inequality that

$$\mathbb{E}_0 \|\tilde{f}_J - \mathbb{E}_0 \tilde{f}_J\|_\infty \leq \sup_{x \in [0, 1]} \|\mathbf{a}_J(x)\|_1 \|\mathbf{M}\|_{(\infty, \infty)} \mathbb{E}_0 \|\boldsymbol{\eta}\|_\infty \lesssim l_{1, J} \lambda_{\mathbf{A}, J}^{-1} \mathbb{E}_0 \|\boldsymbol{\eta}\|_\infty,$$

where we have applied Lemma 9.9 with $g(x) = x^{-1}$ to bound $\|\mathbf{M}\|_{(\infty, \infty)} \lesssim \lambda_{\mathbf{A}, J}^{-1}$. Note that by our global assumption on the true model, $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma_0^2 \mathbf{A}^T \mathbf{A})$ under P_0 . Now by Lemma 2.3.4 of [5], we have for $Z_m \sim \mathcal{N}(0, 1)$, $m = 1, \dots, J$, i.i.d. that

$$\mathbb{E}_0 \|\boldsymbol{\eta}\|_\infty \leq \max_{1 \leq m \leq J} \sqrt{\sigma_0^2 (\mathbf{A}^T \mathbf{A})_{m, m}} \mathbb{E} \left(\max_{1 \leq m \leq J} |Z_m| \right) \lesssim \sqrt{\lambda_{\mathbf{A}, J} \log(J)},$$

and used (8.2) to bound $(\mathbf{A}^T \mathbf{A})_{m, m}$. Thus, $\mathbb{E}_0 \|\tilde{f}_J - \mathbb{E}_0 \tilde{f}_J\|_\infty \leq G_1 l_{1, J} \sqrt{\log(J) / \lambda_{\mathbf{A}, J}}$ for some constant $G_1 > 0$. Substituting this bound back into (9.2) yields the desired result. \square

Lemma 9.4. *For any $x \in [0, 1]$ and any symmetric matrix $\boldsymbol{\Sigma}$ of conformable dimensions,*

$$\mathbf{a}_J(x)^T \boldsymbol{\Sigma}^T \mathbf{R}^T \mathbf{M} \mathbf{R} \boldsymbol{\Sigma} \mathbf{a}_J(x) \leq 4 \lambda_{\max}(\mathbf{M}) \lambda_{\max}(\boldsymbol{\Sigma})^2 \|\mathbf{a}_J(x)\|^4 \|\dot{\mathbf{a}}_J(x)\|^2.$$

For the lower bound, assume that for any $x \in [0, 1]$ there is a $\mathbf{U}(x)$ such that $\|\mathbf{U}(x)\|^2 \leq \xi$ for some $\xi > 0$ that may depend on J , and $\langle \mathbf{U}(x), \mathbf{a}_J(x) \rangle = 0$, then

$$\mathbf{a}_J(x)^T \boldsymbol{\Sigma}^T \mathbf{R}^T \mathbf{M} \mathbf{R} \boldsymbol{\Sigma} \mathbf{a}_J(x) \geq \lambda_{\min}(\mathbf{M}) \lambda_{\min}(\boldsymbol{\Sigma})^2 \|\mathbf{a}_J(x)\|^4 \langle \mathbf{U}(x), \dot{\mathbf{a}}_J(x) \rangle^2 \xi^{-1}.$$

Proof. Since the quadratic form in question is greater than $\lambda_{\min}(\mathbf{M}) \mathbf{a}_J(x)^T \boldsymbol{\Sigma}^T \mathbf{R}^T \mathbf{R} \boldsymbol{\Sigma} \mathbf{a}_J(x)$ and less than $\lambda_{\max}(\mathbf{M}) \mathbf{a}_J(x)^T \boldsymbol{\Sigma}^T \mathbf{R}^T \mathbf{R} \boldsymbol{\Sigma} \mathbf{a}_J(x)$, we can reduce the problem to finding upper and lower bounds for $\mathbf{a}_J(x)^T \boldsymbol{\Sigma}^T \mathbf{R}^T \mathbf{R} \boldsymbol{\Sigma} \mathbf{a}_J(x)$.

For the upper bound, we begin by defining functions $g_1(x) := \dot{\mathbf{a}}_J(x)^T \boldsymbol{\Sigma} \mathbf{a}_J(x)$ and $g_2(x) := \mathbf{a}_J(x)^T \boldsymbol{\Sigma} \mathbf{a}_J(x)$. Using the fact that $\|\mathbf{x} - \mathbf{y}\|^2 \leq 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2$ (Cauchy-Schwartz inequality and geometric-arithmetric mean inequality $\sqrt{ab} \leq (a + b)/2$), we have

$$\begin{aligned} \mathbf{a}_J(x)^T \boldsymbol{\Sigma}^T \mathbf{R}^T \mathbf{R} \boldsymbol{\Sigma} \mathbf{a}_J(x) &= \|g_1(x) \mathbf{a}_J(x) - g_2(x) \dot{\mathbf{a}}_J(x)\|^2 \\ &\leq 2g_1(x)^2 \|\mathbf{a}_J(x)\|^2 + 2g_2(x)^2 \|\dot{\mathbf{a}}_J(x)\|^2. \end{aligned} \quad (9.3)$$

Now $g_2(x) \leq \lambda_{\max}(\boldsymbol{\Sigma}) \|\mathbf{a}_J(x)\|^2$. Then by the Cauchy-Schwartz inequality and norm submultiplicative property, we have $|g_1(x)| \leq \|\dot{\mathbf{a}}_J(x)\| \lambda_{\max}(\boldsymbol{\Sigma}) \|\mathbf{a}_J(x)\|$. These two bounds therefore imply that

$$\mathbf{a}_J(x)^T \boldsymbol{\Sigma}^T \mathbf{R}^T \mathbf{R} \boldsymbol{\Sigma} \mathbf{a}_J(x) \leq 4 \|\mathbf{a}_J(x)\|^4 \|\dot{\mathbf{a}}_J(x)\|^2 \lambda_{\max}(\boldsymbol{\Sigma})^2.$$

For the lower bound, we know that by assumption $\|\mathbf{U}(x)\|^2 \leq \xi$ and $\langle \mathbf{U}(x), \mathbf{a}_J(x) \rangle = 0$. Since $g_2(x) \geq \lambda_{\min}(\boldsymbol{\Sigma}) \|\mathbf{a}_J(x)\|^2$, then in view of (9.3) and the Cauchy-Schwartz inequality,

$$\begin{aligned} \mathbf{a}_J(x)^T \boldsymbol{\Sigma}^T \mathbf{R}^T \mathbf{R} \boldsymbol{\Sigma} \mathbf{a}_J(x) \xi &\geq \|g_1(x) \mathbf{a}_J(x) - g_2(x) \dot{\mathbf{a}}_J(x)\|^2 \|\mathbf{U}(x)\|^2 \\ &\geq \langle \mathbf{U}(x), g_1(x) \mathbf{a}_J(x) - g_2(x) \dot{\mathbf{a}}_J(x) \rangle^2 \\ &= [g_1(x) \langle \mathbf{U}(x), \mathbf{a}_J(x) \rangle - g_2(x) \langle \mathbf{U}(x), \dot{\mathbf{a}}_J(x) \rangle]^2 \\ &\geq \lambda_{\min}(\boldsymbol{\Sigma})^2 \|\mathbf{a}_J(x)\|^4 \langle \mathbf{U}(x), \dot{\mathbf{a}}_J(x) \rangle^2. \end{aligned} \quad \square$$

Corollary 9.5. *By specializing Lemma 9.4 to B-splines, we will obtain:*

$$J^2 \lambda_{\min}(\mathbf{M}) \lambda_{\min}(\mathbf{\Sigma})^2 \lesssim \mathbf{b}_{J,q}(x)^T \mathbf{\Sigma}^T \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{\Sigma} \mathbf{b}_{J,q}(x) \lesssim J^2 \lambda_{\max}(\mathbf{M}) \lambda_{\max}(\mathbf{\Sigma})^2; \quad (9.4)$$

and for the wavelet bases:

$$2^{5J} \lambda_{\min}(\mathbf{M}) \lambda_{\min}(\mathbf{\Sigma})^2 \lesssim \boldsymbol{\psi}_J(x)^T \mathbf{\Sigma}^T \mathbf{R}^T \mathbf{M} \mathbf{R} \mathbf{\Sigma} \boldsymbol{\psi}_J(x) \lesssim 2^{5J} \lambda_{\max}(\mathbf{M}) \lambda_{\max}(\mathbf{\Sigma})^2. \quad (9.5)$$

Proof. By Lemma 9.6, we see that $\|\mathbf{b}_{J,q}(x)\|^4 \leq 1$ for any $x \in [0, 1]$. By (8) of Chapter X in [3] and Lemma 9.6, we know that

$$\|\dot{\mathbf{b}}_{J,q}(x)\|^2 = \|\mathbf{W} \mathbf{b}_{J,q-1}(x)\|^2 \leq \lambda_{\max}(\mathbf{W}^T \mathbf{W}) \|\mathbf{b}_{J,q-1}(x)\|^2 \leq \|\mathbf{W}^T \mathbf{W}\|_{(2,2)} \leq \|\mathbf{W}^T \mathbf{W}\|_{(\infty, \infty)},$$

with \mathbf{W} given in (5.8). Note that $\mathbf{W}^T \mathbf{W}$ is 3-banded where its i th diagonal entry is $2(q-1)^2 / (t_i - t_{i+1-q})^2$ and its off-diagonals are of the form $-(q-1)^2 (t_i - t_{i+1-q})^{-1} (t_{i+1} - t_{i+2-q})^{-1}$. Therefore by the quasi-uniformity of knots,

$$\|\mathbf{W}^T \mathbf{W}\|_{(\infty, \infty)} \lesssim \frac{1}{\min_{1 \leq k \leq N} (t_k - t_{k-1})^2} \lesssim \frac{1}{\Delta_T^2} \lesssim J^2.$$

Substituting this bound back yields the stated upper bound.

For the lower bound, we need to construct a vector that is orthogonal to $\mathbf{b}_{J,q}(x)$, and we use a technique developed by [25] for this purpose. Now for any $x \in [0, 1]$, we can always find an index i_x such that $x \in [t_{i_x-1}, t_{i_x}]$. Define $\mathbf{U}(x) = (u_1(x), \dots, u_J(x))^T$ such that

$$u_j(x) = \begin{cases} 1 - B_{i_x,q}(x), & j = i_x, \\ -B_{i_x,q}(x), & j = i_x + 1, \dots, i_x + q - 1, \\ 0, & \text{otherwise.} \end{cases}$$

Using the fact that $0 \leq B_{j,q}(x) \leq 1$ for any x and j , we have

$$\|\mathbf{U}(x)\|^2 = (1 - B_{i_x,q}(x))^2 + (q-1) B_{i_x,q}(x)^2 \leq q, \quad (9.6)$$

and we can take $\xi = q$. By the compact support of B-splines i.e., $B_{j,q}(x) = 0$ for $j < i_x$ or $j > i_x + q - 1$ and the partition of unity $\sum_j B_{j,q} = 1$, we obtain

$$\begin{aligned} \langle \mathbf{U}(x), \mathbf{b}_{J,q}(x) \rangle &= \sum_{j=i_x}^{i_x+q-1} u_j(x) B_{j,q}(x) = (1 - B_{i_x,q}(x)) B_{i_x,q}(x) - B_{i_x,q}(x) \sum_{j=i_x+1}^{i_x+q-1} B_{j,q}(x) \\ &= (1 - B_{i_x,q}(x)) B_{i_x,q}(x) - B_{i_x,q}(x) (1 - B_{i_x,q}(x)) = 0. \end{aligned} \quad (9.7)$$

Furthermore, using the derivative formula for B-splines as encoded in (5.8) and the compact support of $B_{j-1,q}(x)$,

$$\begin{aligned} \langle \mathbf{U}(x), \mathbf{W} \mathbf{b}_{J,q-1}(x) \rangle &= (q-1) \sum_{j=i_x}^{i_x+q-2} \frac{u_{j+1}(x) - u_j(x)}{t_j - t_{j-q+1}} B_{j,q-1}(x) \\ &= (q-1) \frac{-B_{i_x,q}(x) - (1 - B_{i_x,q}(x))}{t_{i_x} - t_{i_x-q+1}} B_{i_x,q-1}(x) \\ &= -(q-1) \frac{B_{i_x,q-1}(x)}{t_{i_x} - t_{i_x-q+1}}. \end{aligned} \quad (9.8)$$

Now observe that $t_{i_x} - t_{i_x+q-1} \leq (q-1)\Delta\tau \lesssim (q-1)J^{-1}$ from the quasi-uniformity of the knots, and note that since $x \in [t_{i_x-1}, t_{i_x}]$ and $B_{i_x, q-1}(x)$ is supported on $x \in [t_{i_x-q+1}, t_{i_x}]$, it follows that $B_{i_x, q-1}(x) > 0$ and hence $\min_{0 \leq x \leq 1} B_{i_x, q-1}(x)^2$ is some positive constant not depending on J . Squaring both sides of (9.8) while keeping in mind of the auxiliary facts discussed, we see that $\langle \mathbf{U}(x), \dot{\mathbf{b}}_{J,q}(x) \rangle^2 \gtrsim q^{-2} \min_{0 \leq x \leq 1} B_{i_x, q-1}(x)^2 J^2 \lambda_{\min}(\boldsymbol{\Sigma})^2$ for any $x \in [0, 1]$. The lower bound in the statement then follows from this bound and $\inf_{x \in [0, 1]} \|\mathbf{b}_{J,q}(x)\|^4 \geq q^{-2}$ by Lemma 9.6.

For the wavelets, we can without loss of generality assume that $\mathbf{R} \neq \mathbf{0}$ because if it were the zero matrix, then the statements of the lemma become vacuously true. Let us start with the upper bound. By Lemma 9.7, it holds that $\|\boldsymbol{\psi}_J(x)\|^4 \lesssim 2^{2J}$ for any $x \in [0, 1]$. By the property of CDV wavelets, we know that ψ' is uniformly bounded, and since there are compactly supported,

$$\|\dot{\boldsymbol{\psi}}_J(x)\|^2 = \sum_{j=N}^{J-1} \sum_{k=0}^{2^j-1} \left[2^{j/2} 2^j \psi'(2^j x - k) \right]^2 \lesssim \sum_{j=N}^{J-1} 2^{3j} \lesssim 2^{3J}.$$

Hence the 2^{5J} -factor comes from multiplying these two bounds.

For the lower bound, first note that for a fixed j , $\psi_{j,k}(x)$ is nonzero for only a finite number of k 's due again to its compact support property. For the sake of subsequent argument, let us concretely enumerate these nonzero spatial shifts as $k = i_x, i_x + 1, \dots, i_x + m - 1$ for some $m > 1$ following the notation for the B-splines case discussed previously. We then construct $\mathbf{U}(x) = (\mathbf{U}_N^T(x), \dots, \mathbf{U}_{J-1}^T(x))^T$ where the k th entry for the j th vector $\mathbf{U}_j(x)$ is

$$u_{j,k}(x) = \begin{cases} \psi_{j, i_x+1}(x), & k = i_x, \\ -\psi_{j, i_x}(x), & k = i_x + 1, \\ 0, & \text{otherwise.} \end{cases}$$

Then its ℓ_2 -norm is

$$\|\mathbf{U}(x)\|^2 = \sum_{j=N}^{J-1} [2^{j/2} \psi(2^j x - i_x - 1)]^2 + \sum_{j=N}^{J-1} [2^{j/2} \psi(2^j x - i_x)]^2 \leq C_1 2^J,$$

for some constant $C_1 > 0$. Hence we take $\xi = C_1 2^J$. Orthogonality can be checked by

$$\langle \mathbf{U}(x), \boldsymbol{\psi}_J(x) \rangle = \sum_{j=N}^{J-1} \sum_{k=0}^{2^j-1} u_{j,k}(x) \psi_{j,k}(x) = \sum_{j=N}^{J-1} [\psi_{j, i_x+1}(x) \psi_{j, i_x}(x) - \psi_{j, i_x}(x) \psi_{j, i_x+1}(x)] = 0.$$

Likewise, we have

$$\begin{aligned} \langle \mathbf{U}(x), \dot{\boldsymbol{\psi}}_J(x) \rangle &= \sum_{j=N}^{J-1} \sum_{k=0}^{2^j-1} u_{j,k}(x) \psi'_{j,k}(x) = \sum_{j=N}^{J-1} \left[\psi_{j, i_x+1}(x) \psi'_{j, i_x}(x) - \psi_{j, i_x}(x) \psi'_{j, i_x+1}(x) \right] \\ &= \sum_{j=N}^{J-1} 2^{2j} \underbrace{\left[\psi(2^j x - i_x - 1) \psi'(2^j x - i_x) - \psi(2^j x - i_x) \psi'(2^j x - i_x - 1) \right]}_{\text{call this } C_2}. \end{aligned}$$

Now since $\mathbf{R} \neq \mathbf{0}$ by assumption, it follows that $C_2 \neq 0$. Therefore,

$$\langle \mathbf{U}(x), \dot{\boldsymbol{\psi}}_J(x) \rangle^2 = C_2^2 \left(\sum_{j=N}^{J-1} 2^{2j} \right)^2 \geq C_2^2 \sum_{j=N}^{J-1} 2^{4j} \geq (C_2^2/16) 2^{4J}.$$

By Lemma 9.7, we have $\|\boldsymbol{\psi}_J(x)\|^4 \gtrsim 2^{2J}$. With everything now in place, the lower bound of Lemma 9.4 takes the form of (up to some constant multiple) $\lambda_{\min}(\mathbf{M})\lambda_{\min}(\boldsymbol{\Sigma})^2 2^{2J} 2^{4J} 2^{-J}$ and this gives the lower bound for the present lemma. \square

Lemma 9.6 (L_2 -norm of B-splines). *For any $x \in [0, 1]$ and any $1 \leq J \leq n$,*

$$q^{-1} \leq \|\mathbf{b}_{J,q}(x)\|^2 \leq 1.$$

Proof. For any $x \in [0, 1]$, we can always find a positive integer i_x such that $x \in [t_{i_x-1}, t_{i_x}]$. Using the compact support of B-splines $B_{j,q}(x) = 0$ for $j < i_x$ or $j > i_x + q - 1$, we can write

$$\|\mathbf{b}_{J,q}(x)\|^2 = \sum_{j=1}^J B_{j,q}(x)^2 = \sum_{j=i_x}^{i_x+q-1} B_{j,q}(x)^2 \geq \frac{1}{q} \left(\sum_{j=i_x}^{i_x+q-1} B_{j,q}(x) \right)^2 = \frac{1}{q},$$

where we have used Jensen's inequality in the lower bound and the partition of unity property $\sum_j B_{j,q}(x) = 1$ for any $x \in [0, 1]$. For the upper bound, note that since $0 \leq B_{j,q}(x) \leq 1$ for all j , it holds

$$\|\mathbf{b}_{J,q}(x)\|^2 = \sum_{j=1}^J B_{j,q}(x)^2 \leq \sum_{j=1}^J B_{j,q}(x) = 1,$$

again using the B-spline partition of unity. \square

Lemma 9.7 (L_2 -norm of wavelets). *For any $x \in [0, 1]$ and any $1 \leq J \leq n$,*

$$\|\boldsymbol{\psi}_J(x)\| \asymp 2^{J/2}.$$

Proof. For any $x \in [0, 1]$, choose k_x such that $2^{J-1}x - k_x$ is in the support of ψ , so that $\psi(2^{J-1}x - k_x)$ is a constant strictly larger than 0. Then

$$\|\boldsymbol{\psi}_J(x)\|^2 = \sum_{j=N}^{J-1} \sum_{k=0}^{2^j-1} \psi_{j,k}(x)^2 \geq \sum_{k=0}^{2^{J-1}-1} \psi_{J-1,k}(x)^2 \geq (0.5) 2^J \psi(2^{J-1}x - k_x)^2 \gtrsim 2^J.$$

For the upper bound, note that since ψ is compact supported and bounded, we have $\|\sum_{k=0}^{2^j-1} \psi_{j,k}(\cdot)\|_\infty \lesssim 2^{j/2}$ and $|\psi_{j,k}(x)| \lesssim 2^{j/2}$ for any $x \in [0, 1]$, therefore

$$\|\boldsymbol{\psi}_J(x)\|^2 \leq \max_{N \leq j \leq J-1} \max_{0 \leq k \leq 2^j-1} |\psi_{j,k}(x)| \sum_{j=N}^{J-1} \sum_{k=0}^{2^j-1} \psi_{j,k}(x) \lesssim 2^J. \quad \square$$

The following result relates the trace of matrix products to min and max eigenvalues of one of its product factor.

Lemma 9.8. *For any $J \times J$ non-negative definite matrices \mathbf{A}, \mathbf{B} ,*

$$\lambda_{\min}(\mathbf{A})\text{tr}(\mathbf{B}) \leq \text{tr}(\mathbf{AB}) \leq \lambda_{\max}(\mathbf{A})\text{tr}(\mathbf{B}).$$

Proof. By eigen-decomposition, $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T$ for some orthonormal matrix \mathbf{P} and $\mathbf{\Lambda} = \text{Diag}\{\lambda_1(\mathbf{A}), \dots, \lambda_J(\mathbf{A})\}$ are the eigenvalues. Using the cyclic permutation of the trace operator, we have

$$\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{P}\mathbf{\Lambda}\mathbf{P}^T\mathbf{B}) = \text{tr}(\mathbf{\Lambda}\mathbf{P}^T\mathbf{B}\mathbf{P}) = \sum_{i=1}^J \lambda_i(\mathbf{A})(\mathbf{P}^T\mathbf{B}\mathbf{P})_{ii}.$$

Note that $(\mathbf{P}^T\mathbf{B}\mathbf{P})_{ii} \geq 0$ since \mathbf{B} is non-negative definite. Therefore, the right hand side above is bounded above by $\lambda_{\max}(\mathbf{A})\text{tr}(\mathbf{P}^T\mathbf{B}\mathbf{P})$ and below by $\lambda_{\min}(\mathbf{A})\text{tr}(\mathbf{P}^T\mathbf{B}\mathbf{P})$. The result follows by invoking the cyclic permutation of the trace operator again and the fact $\mathbf{P}\mathbf{P}^T = \mathbf{I}$ to conclude that $\text{tr}(\mathbf{P}^T\mathbf{B}\mathbf{P}) = \text{tr}(\mathbf{B}\mathbf{P}\mathbf{P}^T) = \text{tr}(\mathbf{B})$. \square

Lemma 9.9. *Let \mathbf{H} be a $J \times J$ symmetric, positive definite and w -banded such that $h_{ij} = 0$ if $|i - j| > w$. Assume that the eigenvalues of \mathbf{H} are contained in $[a\tau_m, b\tau_m]$ for some fixed $0 < a < b < \infty$ and some sequence τ_m . Furthermore, let the function g be analytic on $[a\tau_m, b\tau_m]$. Then,*

$$\|g(\tau_m^{-1}\mathbf{H})\|_{(\infty, \infty)} = O(1).$$

Proof. First observe that if \mathbf{U} is a -banded and \mathbf{V} is b -banded, then \mathbf{UV} is $a + b$ -banded. Indeed, $(\mathbf{UV})_{i,j} = \sum_{k=1}^J u_{ik}v_{kj} \neq 0$ when at least one term in the sum is not zero. Therefore, both u_{ik} and v_{kj} are nonzero for some k . By bandedness of the matrices involved, it must be that $|i - k| \leq a$ and $|j - k| \leq b$, and thus by the triangle inequality $|i - j| \leq a + b$, implying that \mathbf{UV} is $a + b$ -banded. By an induction argument, we deduce that \mathbf{U}^n is na -banded.

Because we can always scale \mathbf{H} by τ_m^{-1} such that its eigenvalues are in $[a, b]$, we let $\tau_m = 1$ without loss of generality. Let p_n be a polynomial function of degree n , and by the previous argument, we know that $p_n(\mathbf{H})$ is nw -banded. We will use a result by Bernstein (see Theorem 73 of [13]) on the best polynomial approximation to analytic functions. For any f analytic on $[a, b]$, there exist constants $C_0 > 0, \delta < 1$ such that

$$\inf_{p \in \mathcal{P}} \|f - p\|_{\infty} \leq C_0 \delta^{n+1}, \quad (9.9)$$

where \mathcal{P} is the space of polynomials of degree n (or order $n + 1$). Here C_0 and δ depend on half-axes of certain ellipse containing $[a, b]$ and also the supremum of f on this ellipse. Now by spectral theory and the result above,

$$\|g(\mathbf{H}) - p_n(\mathbf{H})\|_{(2,2)} = \sup_{x \in \Lambda(\mathbf{H})} |g(x) - p_n(x)| \leq C_0 \delta^{n+1},$$

where $\Lambda(\mathbf{H})$ is the set of eigenvalues of \mathbf{H} . As mentioned before, $p_n(\mathbf{H})_{i,j} = 0$ for $|i - j| > nw$. Now suppose $i \neq j$ and choose n such that $nw < |i - j| \leq (n + 1)w$, then

$$|g(\mathbf{H})_{i,j}| = |g(\mathbf{H})_{i,j} - p_n(\mathbf{H})_{i,j}| \leq \|g(\mathbf{H}) - p_n(\mathbf{H})\|_{(2,2)} \leq C_0 \delta^{|i-j|/w},$$

since $\delta < 1$. For the case $i = j$, we have $g(\mathbf{H})_{ii} \leq \|g(\mathbf{H})\|_{(2,2)}$. Combining both cases, we conclude $|g(\mathbf{H})_{i,j}| \leq \max\{C_0, \|g(\mathbf{H})\|_{(2,2)}\} \delta^{|i-j|/w}$. Again recalling that $\delta < 1$,

$$\|g(\mathbf{H})\|_{(\infty,\infty)} \leq \max\{C_0, \|g(\mathbf{H})\|_{(2,2)}\} \max_{1 \leq i \leq J} \sum_{j=1}^J \delta^{|i-j|/w} \lesssim 1 + 2 \sum_{j=1}^{\infty} \delta^{j/w} < \infty. \quad \square$$

References

- [1] Castillo, I. and Nickl, R. (2013). Nonparametric Bernstein-von Mises theorems in Gaussian white noise. *Ann. Statist.*, 41(4):1999–2028.
- [2] Castillo, I. and Nickl, R. (2014). On the Bernstein-von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.*, 42(5):1941–1969.
- [3] de Boor, C. (2001). *A Practical Guide to Splines*. Springer-Verlag New York, Inc., revised edition.
- [4] Giné, E. and Nickl, R. (2011). Rates of contraction for posterior distributions in L^r -metrics, $1 \leq r \leq \infty$. *Ann. Statist.*, 39(6):2883–2911.
- [5] Giné, E. and Nickl, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press.
- [6] Harville, D. A. (1997). *Matrix Algebra From a Statistician's Perspective*. Springer, New York.
- [7] Hotelling, H. (1939). Tubes and spheres in n -spaces, and a class of statistical problems. *American Journal of Mathematics*, 61(2):440–460.
- [8] Jaffard, S. (1997). Multifractal formalism for functions part II: Self-similar functions. *Siam J. Math. Anal.*, 28(4):971–998.
- [9] Johansen, S. and Johnstone, I. M. (1990). Hotelling's theorem on the volume of tubes: some illustrations in simultaneous inference and data analysis. *Ann. Statist.*, 18(2):652–684.
- [10] Knowles, M. D. (1988). *Simultaneous Confidence Bands for Random Functions*. PhD thesis, Stanford University.
- [11] Lepski, O. V. (1990). On a problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.*, 35(3):454–466.

- [12] Lepski, O. V., Mammen, E., and Spokoiny, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.*, 25(3):929–947.
- [13] Meinardus, G. (1967). *Approximation of Functions: Theory and Numerical Methods*. Springer-Verlag, New York. Translated by Larry L. Schumaker.
- [14] Meyer, Y. (1992). *Wavelets and Operators*. Cambridge University Press.
- [15] Naiman, D. Q. (1986). Conservative confidence bands in curvilinear regression. *Ann. Statist.*, 14(3):896–906.
- [16] Ray, K. (2017). Adaptive Bernstein-von Mises theorems in Gaussian white noise. *Ann. Statist.* To appear.
- [17] Rudelson, M. and Vershynin, R. (2013). Hanson-Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.*, 18(82):1–9.
- [18] Schumaker, L. (2007). *Spline Functions: Basic Theory*. Cambridge University Press, New York, third edition.
- [19] Sniekers, S. (2015). *Credible sets in nonparametric regression*. PhD thesis, Leiden University.
- [20] Szabó, B. (2015). On Bayesian based adaptive confidence sets for linear functionals. *Bayesian Statistics from Methods to Models and Applications*, pages 91–105.
- [21] Szabó, B., van der Vaart, A. W., and van Zanten, J. H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.*, 43(4):1391–1428.
- [22] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Process With Applications to Statistics*. Springer-Verlag New York, Inc.
- [23] Yoo, W. W. and Ghosal, S. (2016). Supremum norm posterior contraction and credible sets for nonparametric multivariate regression. *Ann. Statist.*, 44(3):1069–1102.
- [24] Yoo, W. W., Rousseau, J., and Rivoirard, V. (2017). Adaptive supremum norm posterior contraction: spike-and-slab priors and anisotropic Besov spaces. arXiv: 1708.01909 [math.ST].
- [25] Zhou, S., Shen, X., and Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist.*, 26(5):1760–1782.