

## Honest Bayesian confidence sets for the L2-norm

***Citation for published version (APA):***

Szabó, B. T., Vaart, van der, A. W., & Zanten, van, J. H. (2013). *Honest Bayesian confidence sets for the L2-norm*. (arXiv.org; Vol. 1311.7474 [math.ST]). s.n.

***Document status and date:***

Published: 01/01/2013

***Document Version:***

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

***Please check the document version of this publication:***

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

***General rights***

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

***Take down policy***

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

# Honest Bayesian confidence sets for the $L^2$ -norm

Botond Szabó<sup>a,\*</sup>, Aad van der Vaart<sup>b</sup>, Harry van Zanten<sup>c</sup>

<sup>a</sup>*Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, Netherlands*

<sup>b</sup>*Leiden University, P.O. Box 9512, 2300 RA Leiden, Netherlands*

<sup>c</sup>*University of Amsterdam, P.O. Box 94248,  
1090 GE Amsterdam, Netherlands*

---

## Abstract

We investigate the problem of constructing Bayesian credible sets that are honest and adaptive for the  $L^2$ -loss over a scale of Sobolev classes with regularity ranging between  $[D, 2D]$ , for some given  $D$  in the context of the signal-in-white-noise model. We consider a scale of prior distributions indexed by a regularity hyper-parameter and choose the hyper-parameter both by marginal likelihood empirical Bayes and by hierarchical Bayes method, respectively. Next we consider a ball centered around the corresponding posterior mean with prescribed posterior probability. We show by theory and examples that both the empirical Bayes and the hierarchical Bayes credible sets give misleading, overconfident uncertainty quantification for certain oddly behaving truth. Then we construct a new empirical Bayes method based on risk estimation, which provides the correct uncertainty quantification and optimal size.

*Keywords:* Credible sets, coverage, uncertainty quantification

---

## 1. Introduction

In Bayesian nonparametrics it is common to visualize the uncertainty of the posterior distribution by plotting the credible sets, i.e. sets accumulating a large fraction (typically 95%) of the posterior mass. These sets are often used in practice to quantify the uncertainty of a given estimate. They can be especially useful when the construction of confidence sets is not possible due to computational complexity or lack of theoretical results. However, the frequentist interpretation of the credible sets at the moment is rather unclear. In the present paper we investigate the asymptotic frequentist behaviour of Bayesian credible sets in the

---

\*Corresponding author

*Email addresses:* [b.szabo@tue.nl](mailto:b.szabo@tue.nl) (Botond Szabó), [avdvaart@math.leidenuniv.nl](mailto:avdvaart@math.leidenuniv.nl) (Aad van der Vaart), [hvzanten@uva.nl](mailto:hvzanten@uva.nl) (Harry van Zanten)

<sup>1</sup>Research supported by the Netherlands Organization for Scientific Research (NWO). The research leading to these results has received funding from the European Research Council under ERC Grant Agreement 320637.

<sup>2</sup>AMS 2000 subject classifications: Primary 62G15,62G05; secondary 62G20

context of the signal-in-white-noise model. We consider the sequence formulation

$$X_i = \theta_{0,i} + \frac{1}{\sqrt{n}}Z_i, \quad \text{for all } i = 1, 2, \dots \quad (1.1)$$

where  $X = (X_1, X_2, \dots)$  is the observed infinite sequence,  $Z_i$  are independent standard normal distributed random variables and  $\theta_0 = (\theta_{0,1}, \theta_{0,2}, \dots)$  is the unknown infinite dimensional parameter of interest. The signal-in-white-noise model is a relatively simple and tractable model but can be used at the same time as a platform to investigate more difficult statistical problems. For instance it is asymptotically equivalent with the regression model [1], and the density function estimation problem [16]. We expect our finding to generalize to the preceding models, however this does not follow straightforward from the asymptotic equivalence of the models, since we consider a particular method.

We assume that the true signal  $\theta_0$  belongs to a collection of nested submodels  $\cup_{\beta \in [D_1, D_2]} \Theta^\beta$ , for fixed  $D_1$  and  $D_2$ , and  $\Theta^{\beta_2} \subset \Theta^{\beta_1}$  for  $\beta_1 < \beta_2$ . Considering any norm  $\|\cdot\|$  a confidence set  $\hat{C}_n$  corresponding this norm is called honest over  $\cup_{\beta \in [D_1, D_2]} \Theta^\beta$  if, for a level  $\gamma > 0$ ,

$$\liminf_{n \rightarrow \infty} \inf_{\theta_0 \in \cup_{\beta \in [D_1, D_2]} \Theta^\beta} \mathbb{P}_{\theta_0}(\theta_0 \in \hat{C}_n) \geq 1 - \gamma$$

and rate adaptive if for all  $\beta \in [D_1, D_2]$

$$\liminf_{n \rightarrow \infty} \inf_{\theta_0 \in \Theta^\beta} \mathbb{P}_{\theta_0}(\|\hat{C}_n\| \leq C_\beta r_{n,\beta}) \geq 1 - \gamma, \quad (1.2)$$

where  $r_{n,\beta}$  denotes the minimax rate corresponding the norm  $\|\cdot\|$  and class  $\Theta^\beta$ , and the constant  $C_\beta$  depends only on the parameter  $\beta$ . Honesty (uniformity in  $\theta_0$ ) is a relatively strong, but essential requirement. A pointwise (not uniform in  $\theta_0$ ) confidence set has very limited applicability in practice, because in this case we know that for large enough  $n$  the confidence set contains the true parameter with high probability, but what “large enough” means highly depends on the unknown parameter itself. Therefore pointwise asymptotic confidence sets provide only theoretical quantification of the uncertainty, in practice these sets are essentially uninformative.

It was shown in [12] and [19] that the size of honest confidence sets over  $\cup_{\beta \in [D_1, D_2]} \Theta^\beta$  is bounded below by the maximum of the minimax rate of estimation  $r_{n,\beta}$  of  $\theta \in \Theta^\beta$  and the minimax testing rate  $\varepsilon_{n,D_1}$  of  $\theta \in \Theta^\beta$  against the alternative hypothesis  $\{\theta \in \Theta^{D_1} : \|\theta - \Theta^\beta\| \geq \varepsilon_{n,D_1}\}$ . Usually the testing rate  $\varepsilon_{n,D_1}$  depends only on the larger submodel  $\Theta_{D_1}$  and it is typically not bigger than the rate of estimation  $r_{n,D_1}$  (over the submodel  $\Theta^{D_1}$ ). Therefore for the existence of honest and adaptive confidence sets over a collection of submodels  $\cup_{\beta \in [D_1, D_2]} \Theta^\beta$  we require that the minimax estimation rate over the smallest submodel  $\Theta^{D_2}$  is not smaller than the minimax testing rate of  $\theta_0 \in \Theta^{D_2}$  against the largest submodel  $\Theta^{D_1}$ , i.e.  $\varepsilon_{n,D_1} \leq r_{n,D_2}$ .

As a first example, consider the supremum-norm and the corresponding Sobolev ball

$$\Theta_\infty^\beta(M) = \{\theta \in \ell^2 : \sup |\theta_i i^\beta| \leq M\}.$$

The minimax testing rate for  $\theta_0 \in \Theta_\infty^\beta(M)$ ,  $\beta > D_1$ , against the largest submodel  $\Theta_\infty^{D_1}(M)$  is  $\varepsilon_{n,D_1} = (n/\log n)^{-D_1/(1+2D_1)}$ , see (3.128) of [10]. Furthermore the minimax rate for estimating  $\theta_0 \in \Theta^\beta(M)$  is  $r_{n,\beta} = (n/\log n)^{-\beta/(1+2\beta)}$ , see (2.87) of [10]. Following from the lower bound introduced in [12] and [19] the size of a honest confidence set over  $\Theta_\infty^\beta(M)$ , with  $\beta > D_1$ , is bounded from below by  $(n/\log n)^{-D_1/(1+2D_1)} \gg (n/\log n)^{-\beta/(1+2\beta)}$ . Therefore honesty and adaptivity can not hold at the same time for any choice of the parameters  $D_1 < D_2$ . Similar results were concluded for the  $L^\infty$ -loss in various other settings, see for instance [4], [7] and [15]. To achieve honesty and adaptivity at the same time for  $L^\infty$ -loss some additional constraints have to be introduced, see for instance [9], [17], [8], [2].

However, the situation is rather different if we consider the  $\ell^2$ -loss and Sobolev balls

$$S^\beta(M) = \{\theta \in \ell^2 : \sum_i \theta_i^2 i^{2\beta} \leq M\}.$$

In this case the minimax testing rate for  $\theta_0 \in S^\beta(M)$ ,  $\beta > D_1$ , against the alternative hypothesis  $S^{D_1}(M)$  is of order  $n^{-D_1/(1/2+2D_1)}$ ; see Theorem 2.1 or 3.1 of [11] or (3.128) of [10]. Furthermore following from [18] the minimax rate for estimating  $\theta_0 \in S^\beta(M)$  is a constant multiplier of  $n^{-\beta/(1+2\beta)}$ . These bounds suggest that the size of the honest confidence sets for  $\beta \in [D_1, D_2]$  can be of the order  $n^{-\beta/(1+2\beta)} \vee n^{-D_1/(1/2+2D_1)}$ . For  $\beta \leq 2D_1$  this means the minimax bound  $n^{-\beta/(1+2\beta)}$  while for  $\beta > 2D_1$  a sub-optimal rate  $n^{-D_1/(1/2+2D_1)}$ . In our work we focus on the special case  $D_2 = 2D_1$ , where the size of the honest confidence sets are bounded from below by the minimax rate  $n^{-\beta/(1+2\beta)}$ , hence adaptation is possible. The existence of honest and adaptive confidence sets over  $\cup_{\beta \in [D, 2D]} S^\beta(M)$  were shown in [19], [3].

In this article we investigate whether Bayesian methods can reproduce the frequentist results and provide adaptive and honest confidence sets for the  $\ell^2$ -loss over the collection of Sobolev balls  $\cup_{\beta \in [D, 2D]} S^\beta(M)$ . First we consider the empirical Bayes method based on marginal likelihood estimation of the regularity parameter and show that although the size of the credible sets achieve the optimal rate, the honesty requirement will not be fulfilled. We construct certain oddly behaving signals  $\theta_0$  for which the marginal likelihood empirical Bayes method provides credible sets with coverage tending to zero, i.e. honesty (and also pointwise coverage) fails. A technical explanation of the preceding phenomenon relies on the bias-variance trade-off. In the marginal likelihood empirical Bayes method, for certain irregular signals  $\theta_0$  the bias can dominate the posterior spread and the variance of the posterior mean, which leads to a coverage probability close to zero.

Next, we consider the hierarchical Bayes method with arbitrary hyper-prior distribution over  $[D, 2D]$ . We show that the full Bayes method performs similarly to the marginal likelihood empirical Bayes method, in the sense that the hierarchical Bayes credible sets are not honest over  $\cup_{\beta \in [D, 2D]} S^\beta(M)$  (or in the case of a degenerate hyper-prior distribution at  $D$  they are not adaptive). This result is perhaps not surprising in the light of [20] and [14] where we have investigated the close relationship of these two techniques.

The negative results stated above show that the standard Bayesian techniques fail to achieve the frequentist limits. However, by modifying the empirical Bayes procedure one can construct optimally behaving credible sets. We introduce a new empirical Bayes method

based on risk estimation, which provides honest credible sets and achieves adaptivity as well. As the first step of the technique we give an estimator for the squared bias and compute the posterior variance. Then we balance out these two quantities to get high coverage and at the same time optimal size for the credible sets. The method is based on the findings of [19].

The main message of this paper is that the choice of the statistical method has to be in accordance with the goal one wants to achieve. For instance if one evaluates the performance of the posterior mean with the mean integrated squared error then it could happen that the likelihood based procedures attain sub-optimal behaviour. A possible explanation of this phenomenon relies on that the mean integrated squared error is connected to the  $\ell^2$ -loss function, while the likelihood based methods, like the marginal likelihood empirical Bayes method and the hierarchical Bayes method, are related to the Kullback-Leibler divergence.

As mentioned above it is not possible to be honest and rate adaptive at the same time considering the  $L^2$ -loss function over the whole range of submodels  $\cup_{\beta>D} S^\beta(M)$ . However, by removing an asymptotically negligible set of signals from the collection of nested submodels  $\cup_{\beta>D} S^\beta(M)$  the construction of honest and adaptive confidence sets is possible, see [3]. We discuss briefly the extension of our risk based empirical Bayes method to cover this more general case as well. A slightly different direction to achieve honesty is to introduce additional constraints on the signals, similarly to the  $L^\infty$ -loss case. In [21] we introduced a new constraint, the ‘‘polished tail’’ condition, and showed that under this restriction the marginal likelihood empirical Bayes method produces honest confidence sets.

The remainder of the paper is organized as follows. In Section 2 we describe in details the marginal likelihood empirical Bayes and full Bayes procedures and state that the credible sets based on the preceding techniques have frequentist coverage tending to zero for certain true signals. Then we introduce a new empirical Bayes technique, which provides adaptive and honest confidence sets. In Section 3 we demonstrate both the negative and the positive findings by simulating the credible sets for an irregular, oddly behaving function. The proofs of the theorems of Section 2 are given in Sections 4 and 5. The proofs of additional auxiliary lemmas and theorems are deferred to Sections 6 and 7.

### 1.1. Notation

The  $\ell^2$  norm of an element  $\theta \in \ell^2$  is denoted by  $\|\theta\| = (\sum_{i=1}^{\infty} \theta_i^2)^{1/2}$ . For two real sequences  $a_n$  and  $b_n$  the notation  $a_n \lesssim b_n$  means that  $a_n/b_n$  is bounded, and  $a_n \ll b_n$  that  $a_n/b_n \rightarrow 0$ . For two real numbers  $a$  and  $b$ , the notations  $a \vee b$  and  $a \wedge b$  denote the maximum and the minimum of the numbers, respectively. If  $a$  denotes the empty set and  $b$  is a real number, then both  $a \vee b$  and  $a \wedge b$  are taken to be equal to  $b$ . We denote the distribution of the infinite sequence  $X$  corresponding to  $\theta_0$  in (1.1) by  $P_{\theta_0}$  and the corresponding expectation and variance by  $E_{\theta_0}$  and  $\text{var}_{\theta_0}$ , respectively.

## 2. Main result

### 2.1. Model

To make inference about the unknown sequence  $\theta_0$  in the signal-in-white-noise model (1.1) we endow it with a prior distribution

$$\Pi_\alpha(\cdot) = \bigotimes_{i=1}^{\infty} N(0, i^{-1-2\alpha}), \quad (2.1)$$

where the parameter  $\alpha > 0$  denotes the regularity level. The corresponding posterior distribution  $\theta|X \sim \Pi_\alpha(\cdot|X)$  can be easily computed

$$\Pi_\alpha(\cdot|X) = \bigotimes_{i=1}^{\infty} N\left(\frac{n}{i^{1+2\alpha} + n} X_i, \frac{1}{i^{1+2\alpha} + n}\right). \quad (2.2)$$

The optimal choice of the hyper-parameter  $\alpha = \beta$  leads to posterior contraction rate  $n^{-\beta/(1+2\beta)}$ , while for other choices we get sub-optimal contraction rates; see [13] and [5]. Since the regularity parameter  $\beta \in [D, 2D]$  of the truth  $\theta_0$  is usually not available one has to use a data driven method to choose  $\alpha$ .

### 2.2. Marginal likelihood empirical Bayes method

The first adaptive Bayes method we deal with is the marginal likelihood empirical Bayes method. In the Bayesian setting, described by the conditional distributions  $\theta|\alpha \sim \Pi_\alpha$  and  $X|(\theta, \alpha) \sim \otimes_i N(\theta_i, 1/n)$ , it holds that

$$X|\alpha \sim \bigotimes_{i=1}^{\infty} N(0, i^{-1-2\alpha} + 1/n).$$

The corresponding log-likelihood for  $\alpha$  (relative to an infinite product of  $N(0, 1/n)$ -distributions) is given by

$$\ell_n(\alpha) = -\frac{1}{2} \sum_{i=1}^{\infty} \left( \log \left( 1 + \frac{n}{i^{1+2\alpha}} \right) - \frac{n^2}{i^{1+2\alpha} + n} X_i^2 \right). \quad (2.3)$$

We consider the maximum likelihood estimator, i.e. the estimator  $\hat{\alpha}_n$  which maximizes the preceding marginal log-likelihood function in the interval  $[D, 2D]$ , formally

$$\hat{\alpha}_n = \arg \max_{\alpha \in [D, 2D]} \ell_n(\alpha).$$

Then the *empirical Bayes posterior* is defined as the random measure  $\Pi_{\hat{\alpha}_n}(\cdot|X)$  obtained by substituting  $\hat{\alpha}_n$  for  $\alpha$  in the posterior distribution given in (2.2), i.e.

$$\Pi_{\hat{\alpha}_n}(A|X) = \Pi_\alpha(A|X) \Big|_{\alpha=\hat{\alpha}_n} \quad (2.4)$$

for measurable subsets  $A \subset \ell^2$ .

For fixed  $\alpha > 0$  the posterior distribution (2.2) is Gaussian with variance independent of the data, hence for given  $\gamma \in (0, 1)$  there exists a deterministic radius  $r_{n,\gamma}(\alpha)$  such that the ball around the posterior mean  $\hat{\theta}_\alpha = \hat{\theta}_{n,\alpha}$  contains  $1 - \gamma$  fraction of the posterior mass:

$$\Pi_\alpha(\theta : \|\theta - \hat{\theta}_{n,\alpha}\| \leq r_{n,\gamma}(\alpha) | X) = 1 - \gamma. \quad (2.5)$$

In the empirical Bayes method we replace the fixed parameter  $\alpha$  with the estimator  $\hat{\alpha}_n$ . We investigate the resulting credible ball (possibly) blown up by a constant multiplier  $L > 0$

$$\hat{C}_n^E(L) = \{\theta : \|\theta - \hat{\theta}_{n,\alpha}\| \leq L r_{n,\gamma}(\hat{\alpha}_n)\}. \quad (2.6)$$

From [21] follows that the radius of the empirical Bayes credible sets (2.6) is rate adaptive over a collection of Sobolev balls  $\cup_{\beta \in [D, 2D]} S^\beta(M)$ . Here we are interested whether the credible sets are also honest at the same time over  $\cup_{\beta \in [D, 2D]} S^\beta(M)$ . Unfortunately the answer is negative to this question. By adapting Theorem 3.1 of [21] to the present setting we can show that for any regularity parameter  $\beta \in [D, 2D)$  there exists a sequence  $\theta_0 \in S^\beta(M)$  such that the coverage of the marginal likelihood empirical Bayes credible sets tends to zero along a subsequence.

**Theorem 2.1.** *Take an arbitrary  $\beta \in [D, 2D)$ ,  $\beta' \in [D, \beta)$  and  $M > 0$ . Furthermore, take a sequence of positive integers  $n_j$  such that  $n_1 \geq 2$  and  $n_j \gg n_{j-1}^{1+4D}$ , let  $K > 0$  and define the sequence  $\theta_0 = (\theta_{0,1}, \theta_{0,2}, \dots)$  by*

$$\theta_{0,i}^2 = \begin{cases} K i^{-1-2\beta}, & \text{if } n_j^{1/(1+2\beta)} \leq i < 2n_j^{1/(1+2\beta)} \text{ for every } j = 1, 2, \dots \\ 0, & \text{else.} \end{cases} \quad (2.7)$$

*Then the constant  $K$  can be chosen such that  $\theta_0 \in S^{\beta'}(M)$  and for every  $L > 0$  the coverage of the credible set  $\hat{C}_n^E(L)$  defined in (2.6) tends to zero, i.e.  $P_{\theta_0}(\theta_0 \in \hat{C}_{n_j}^E(L)) \rightarrow 0$  as  $j$  tends to infinity.*

The proof of the theorem follows the line of the proof of Theorem 3.1 of [21] tailored to the present set up. The main difference between Theorem 2.1 and Theorem 3.1 of [21] is that in the present setup we have the prior information that the true smoothness lies in the interval  $\beta \in [D, 2D]$ . However, the maximizer of the marginal likelihood function can easily fall outside this interval, for instance in the case of (2.7). Therefore, in Theorem 2.1 we are not concerned with the asymptotic performance of the global maximizer of the likelihood function like in Theorem 3.1 of [21], but rather the local maximizer in the interval  $[D, 2D]$ . We defer the proof to the appendix, Section 7.1.

### 2.3. Hierarchical Bayes

In the hierarchical, full Bayes method the hyper-parameter  $\alpha$  in (2.1) is endowed with a hyper-prior distribution  $\lambda$ . Then the hierarchical prior distribution takes the form

$$\Pi(d\theta) = \int_D^{2D} \Pi_\alpha(d\theta) \lambda(d\alpha).$$

We consider a ball around the hierarchical posterior mean  $\hat{\theta}_n$  with radius  $\hat{r}_{n,\gamma}$  such that it accumulates a fraction  $1 - \gamma$  of the posterior mass. In the construction of the hierarchical Bayes credible sets we introduce some additional flexibility by (possibly) blowing up the ball with a constant factor  $L$

$$\hat{C}_n^H(L) = \{\theta : \|\theta - \hat{\theta}_n\| \leq L\hat{r}_{n,\gamma}\}. \quad (2.8)$$

Similarly to the marginal likelihood empirical Bayes method the hierarchical Bayes method also chooses for certain oddly behaving sequences  $\theta_0$  a sub-optimal hyper-parameter by concentrating the hyper-posterior distribution around it. Therefore, the hierarchical Bayes credible sets (2.8) are not honest and/or have sub-optimal size.

**Theorem 2.2.** *For any choice of the hyper-prior  $\lambda$  there exist  $\beta \in [D, 2D]$  and  $\theta_0 \in S^\beta(M)$  (for arbitrary  $M > 0$ ) such that for every  $L > 0$  the hierarchical Bayes credible set defined in (2.8), has sub-optimal size  $\|\hat{C}_n^H(L)\| \gg n^{-\beta/(1+2\beta)}$  with  $P_{\theta_0}$ -probability tending to one and/or has frequentist coverage tending to zero along a subsequence.*

*Proof.* See Section 4. □

In our setup the main difference between the marginal likelihood empirical Bayes technique and the hierarchical Bayes method is that the former is conditionally Gaussian given the observations while the latter has much more complicated distribution. In the proof of Theorem 2.2 we use the results on the asymptotic behaviour of the maximum likelihood estimator  $\hat{\alpha}_n$ , but the different nature of the posterior distributions requires separate analysis.

#### 2.4. Risk based empirical Bayes method

The main problem with the marginal likelihood empirical Bayes method is that the estimator maximizes the likelihood function instead of minimizing the estimated mean squared error of the posterior mean. This could cause a wrong bias-variance trade-off and therefore bad coverage result. In the present section we aim to correct this problem and give another estimator for the hyper-parameter which provides adaptive and honest empirical Bayes credible sets over  $\cup_{\beta \in [D, 2D]} S^\beta(M)$ . The idea of our estimator relies on the technique introduced in [19].

First we give an estimator for the squared norm of the bias  $B_n^2(\alpha; \theta_0) = \|\theta_0 - E_{\theta_0} \hat{\theta}_{n,\alpha}\|^2 = \sum_{i=1}^{\infty} i^{2+4\alpha} \theta_{0,i}^2 / (i^{1+2\alpha} + n)^2$  with fixed hyper-parameter  $\alpha$ :

$$\hat{B}_{n,k_n}^2(\alpha) = \sum_{i=1}^{k_n} \frac{i^{2+4\alpha}}{(i^{1+2\alpha} + n)^2} (X_i^2 - \frac{1}{n}),$$

where the sequence  $k_n$  will be specified later. One can observe that the expected value of the preceding estimator is

$$B_{n,k_n}^2(\alpha; \theta_0) := E_{\theta_0} \hat{B}_{n,k_n}^2(\alpha) = \sum_{i=1}^{k_n} i^{2+4\alpha} \theta_{0,i}^2 / (i^{1+2\alpha} + n)^2. \quad (2.9)$$



Hence for  $\theta_0 \in S^\beta(M)$  the bias of the estimator  $\hat{B}_{n,k_n}^2(\alpha)$  is bounded above by

$$\sum_{i=k_n+1}^{\infty} i^{2+4\alpha} \theta_{0,i}^2 / (i^{1+2\alpha} + n)^2 \leq (k_n + 1)^{-2\beta} \sum_{i=k_n+1}^{\infty} i^{2\beta} \theta_{0,i}^2 \leq M k_n^{-2\beta}. \quad (2.10)$$

For the choice  $k_n = n^{1/(1/2+2D)}$  the right hand side of the previous display is further bounded from above by  $M n^{-4D/(1+4D)}$  for any  $\beta \in [D, 2D]$ .

With the help of the preceding estimator  $\hat{B}_{n,k_n}(\alpha)$  we define

$$\tilde{\alpha}_n = \inf\{\alpha \geq D : \hat{B}_{n,k_n}(\alpha) \geq C_1 n^{-\alpha/(1+2\alpha)}\} \wedge ((2D - C_0/\log n) \vee D), \quad (2.11)$$

with

$$C_1 > 0 \quad \text{and} \quad C_0 = (1 + 4D)^2 \log(25C_1^{-2}\gamma^{-1})/2 \vee 0. \quad (2.12)$$

The parameter  $C_1$  controls the degree of under smoothing; a smaller choice for the parameter  $C_1$  results in a smaller estimator for the regularity parameter  $\tilde{\alpha}_n$ . The parameter  $C_0$  controls the behaviour of the estimator close to the upper bound  $2D$ . It is a monotonically decreasing function of  $C_1$ .

We use this estimator in the empirical Bayes procedure and define the risk based empirical Bayes posterior by substituting  $\tilde{\alpha}_n$  defined in (2.11) for  $\alpha$  in the posterior (2.2). The risk based empirical Bayes credible sets are constructed as

$$\hat{C}_n^R(L) := \{\theta : \|\theta - \hat{\theta}_{n,\tilde{\alpha}_n}\| < L r_{n,\gamma}(\tilde{\alpha}_n)\}, \quad (2.13)$$

where  $L$  is a scaling parameter,  $r_{n,\gamma}$  is the radius of a  $1 - \gamma$  credible ball for fixed  $\alpha$  defined in (2.5) and  $\tilde{\alpha}_n$  is the new estimator of the hyper-parameter. We show that the credible sets defined in (2.13) are honest over the collection of Sobolev balls  $\cup_{\beta \in [D, 2D]} S^\beta(M)$  and rate adaptive over  $S^\beta(M)$  for all  $\beta \in [D, 2D]$ .

**Theorem 2.3.** *For arbitrary positive parameters  $D, M, C_1$  and  $\gamma$  the credible sets defined in (2.13) with the constant factor  $L \geq \sqrt{8(1 + 3^{1+4D})(\sqrt{6} + \sqrt{2(C_1^2 + M)})}$  are honest*

$$\liminf_{n \rightarrow \infty} \inf_{\theta_0 \in \cup_{\beta \in [D, 2D]} S^\beta(M)} P_{\theta_0}(\theta_0 \in \hat{C}_n^R(L)) \geq 1 - \gamma. \quad (2.14)$$

Furthermore the radius of the credible set is rate adaptive, i.e. for all  $\beta \in [D, 2D]$

$$\liminf_{n \rightarrow \infty} \inf_{\theta_0 \in S^\beta(M)} P_{\theta_0}(r_{n,\gamma}(\hat{\alpha}_n) \leq K n^{-\beta/(1+2\beta)}) \geq 1 - \gamma \quad (2.15)$$

with

$$K = \sqrt{3 + 2/D} \exp\left(\frac{2[(1/2 + \beta) \log(2M/C_1^2) \vee C_0]}{(1/2 + 2D)^2}\right).$$

*Proof.* See Section 5. □

The scaling parameter  $L$ , similarly to  $C_1$  (given in (2.11)), also controls the degree of under smoothing. It can be seen that a smaller choice of the parameter  $C_1$  results in a smaller value for the scaling factor  $L$ . At the same time the radius of the credible ball (2.15) is monotonically increasing as  $C_1$  goes to zero.

**Remark 2.1.** *From the definition of  $\tilde{\alpha}_n$  one can see that for regularity parameter  $\beta > 2D$  the estimator of the hyperparameter  $\tilde{\alpha}_n$  “undersmooths” the truth, i.e. chooses smaller regularity parameter ( $\tilde{\alpha}_n \leq 2D$ ) than the true regularity. Therefore the size of the credible set will be sub-optimal, but the coverage statement holds (actually we get asymptotically conservative coverage one in this case).*

### 2.5. Extension to the case $D_2 > 2D_1$

A natural question is the performance of Bayesian procedures for  $D_2 > 2D_1$ , i.e. without the assumption  $\beta \in [D_1, 2D_1]$ . As we have discussed it already in the introduction, the construction of adaptive and honest confidence sets is impossible in this case. Therefore there is no hope for a Bayesian based procedure to provide confidence sets with good coverage properties and rate adaptive size. However, by introducing some additional constraints, like in [3], the construction of adaptive and honest confidence sets is possible. We show that by slightly adapting our Bayesian based method (introduced in Section 2.4) we can achieve the frequentist limit.

For given  $0 < D_1 < D_2 < \infty$  as a first step we define the grid

$$B = \{\beta_m\}_{m=1}^N = \{D_1, 2D_1, 4D_1, \dots, 2^{N-1}D_1\},$$

for  $2^{(N-1)}D_1 \leq D_2 < 2^N D_1$  and with the help of the grid we introduce the notation, for  $\beta \in B \setminus \{\beta_N\}$ ,

$$\tilde{S}(\beta, \rho, M) = \tilde{S}(\beta, B, \rho_n(\beta), M) = \{\theta \in S^\beta(M) : \|\theta - S^\alpha(M)\| \geq \rho_n(\beta), \forall \alpha > \beta, \alpha \in B\},$$

denoting the collection of  $\beta$ -regular signals which are at least

$$\rho_n(\beta) := n^{-\beta/(1/2+2\beta)}$$

far away from any Sobolev ball  $S^\alpha(M)$  (for  $\alpha > \beta$  and  $\alpha \in B$ ). Then we can take the union of such sets

$$\mathcal{P}_n(M, B) = S^{\beta_N}(M) \bigcup \left( \bigcup_{\beta \in B \setminus \{\beta_N\}} \tilde{S}(\beta, \rho, M) \right).$$

One can observe that the above set tends to  $S^{D_1}(M)$  as  $n$  goes to infinity, hence the “left out signals” asymptotically vanish. Furthermore from Theorem 1 and Theorem 5 of [3] follows that  $\mathcal{P}_n(M, B)$  is the largest set on which the construction of adaptive and honest confidence sets is possible in  $\ell_2$ -norm (for known radius parameter  $M > 0$ ).

Next we introduce a slight modification (following the technique in [3]) of our risk based empirical Bayes method which allows us to reproduce the frequentist results. Using the

adapted version of the test given in (17) of [3] to our setting we can give an estimator  $\hat{\beta}_n \in B$  such that for any  $\theta_0 \in \mathcal{P}_n(M, B)$  we have

$$P_{\theta_0}(\theta_0 \in \cup_{\beta \in [\hat{\beta}_n, 2\hat{\beta}_n]} S^\beta(M)) \geq 1 - \gamma/2.$$

Then we modify the estimator  $\tilde{\alpha}_n$  given in (2.11) by plugging in  $\hat{k}_n = n^{1/(1/2+2\hat{\beta}_n)}$ . Along the lines of the proof of Theorem 5 of [3] and Theorem 2.3 it can be shown that the so defined extended risk based empirical Bayes method provides adaptive and honest confidence sets over  $\mathcal{P}_n(M, B)$ .

## 2.6. Discussion

The marginal likelihood empirical Bayes method and hierarchical Bayes method are closely related. They differ only in that the empirical Bayes method takes the maximizer of the marginal Bayesian likelihood whereas the hierarchical Bayes approach equips this marginal likelihood with a hyper-prior. Therefore the sub-optimal behaviour of the marginal likelihood empirical Bayes approach (at least in our setting) leads to a sub-optimal behaviour of the hierarchical Bayes technique. The situation is not as bad as it looks at first sight, by appropriate choice of the hyper-parameter the construction of adaptive and honest credible sets is possible over  $\cup_{\beta \in [D, 2D]} S^\beta(M)$ . However, these credible sets have no close relation with the full Bayes method. The applied estimator is based on balancing out the bias and variance, and therefore it is substantially different from the marginal likelihood empirical Bayes method and hence from the hierarchical (full) Bayes method.

Another important question is the practical applicability of the derived risk based empirical Bayes method. First of all we note that similarly to the marginal likelihood empirical Bayes method it is computationally substantially faster than the hierarchical Bayes method (except perhaps in case of conjugate hyper-priors). Furthermore following from Remark 2.1 the sets  $\hat{C}_n^R$  can be used in practice for uncertainty quantification even in the case  $\beta > 2D$  in the sense that they contain the truth with high probability for large enough  $n$  (they just do not achieve the optimal size which is perhaps the smaller problem). Of course if one wants to achieve full adaptation then one can apply the extended method of Section 2.5 at the price of throwing out certain (asymptotically vanishing) subsets. For large values of  $n$  the constant multipliers do not play an important role, therefore the choice of the parameter  $C_1$  in (2.12) and as a consequence  $C_0$  and  $L$  has no real effect on the behaviour of the credible sets. For small sample size the parameter  $C_1$  plays a crucial role. A smaller choice of  $C_1$  leads to better coverage results but a larger set size, while a larger  $C_1$  results in a worse coverage level but smaller set size. Therefore (depending on the noise level) we recommend to use smaller values of  $C_1$ .

Finally, we note that our results (after slightly adapting the proofs) also hold for other regularity classes, for instance hyperrectangles. However, in the present paper we consider only Sobolev balls for better tractability and for better connection with the frequentist literature.

### 3. Simulation study

To illustrate our findings we consider the functional formulation of the signal-in-white-noise model

$$X_t = \int_0^t \theta_0(s) ds + (1/\sqrt{n})W_t, \quad t \in [0, 1],$$

where  $X_t$  is the noise observation,  $\theta_0$  the unknown function of interest and  $W_t$  denotes the Wiener process. Then we simulate data from this model for  $\theta_0$  given by its Fourier coefficients

$$\theta_{0,i} = \begin{cases} \sin(i)10^{-1.7}, & \text{if } i = 10..20, \\ 3 \sin(i)100^{-1.7}, & \text{if } i = 100..150, \\ i^{-1.2}, & \text{if } i = 4^{4^j} \dots 2 \times 4^{4^j} \text{ for any } j = 2, \dots, \\ 0, & \text{else.} \end{cases}$$

with respect to the eigen basis  $\varphi_i(t) = \sqrt{2} \cos((i - 1/2)\pi t)$ . We note that the function  $\theta_0$  is contained in the Sobolev ball  $S^{\beta'}(M)$  with any  $\beta' < \beta := 1.2$  and sufficiently large  $M$ . Furthermore we assume the prior knowledge that  $\beta \in [1, 2]$ .

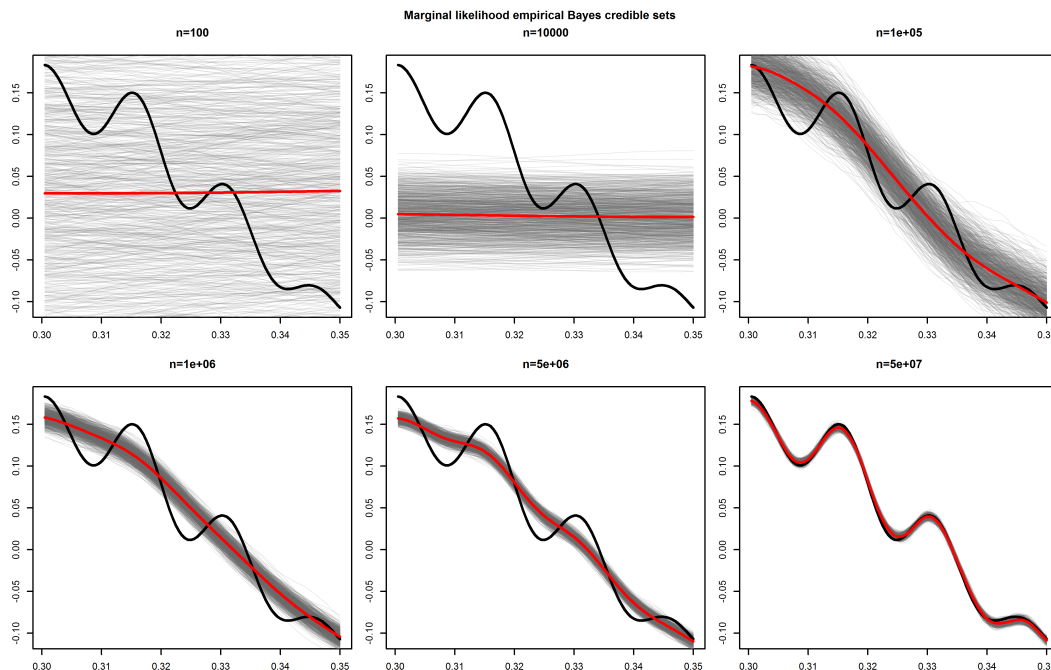


Figure 1: Marginal likelihood empirical Bayes credible sets for unregular function. The true function is drawn in black, the posterior mean in red and the credible set in grey. We have  $n = 10^2, 10^4, 10^5, 10^6, 5 \times 10^6$  and  $5 \times 10^7$ .

In Figure 1 we visualize the 95% marginal likelihood empirical Bayes credible sets defined in (2.6) for  $n = 10^2, 10^4, 10^5, 10^6, 5 \times 10^6$  and  $5 \times 10^7$ , respectively. We note that in Figure

1 and also in the following figures we plot the zoomed in version of the functions to the interval  $[0.3, 0.35]$ . The true function is drawn in black, the posterior mean in blue and the grey area is the collection of the 95% closest out of 800 draws to the posterior mean from the posterior distribution, which gives a good indication of the 95%-credible set. In the simulation study we do not blow up the credible sets by a constant factor, i.e. we worked with  $L = 1$  in all three cases. The figure indicates that the credible set fails to cover the true function along a subsequence.

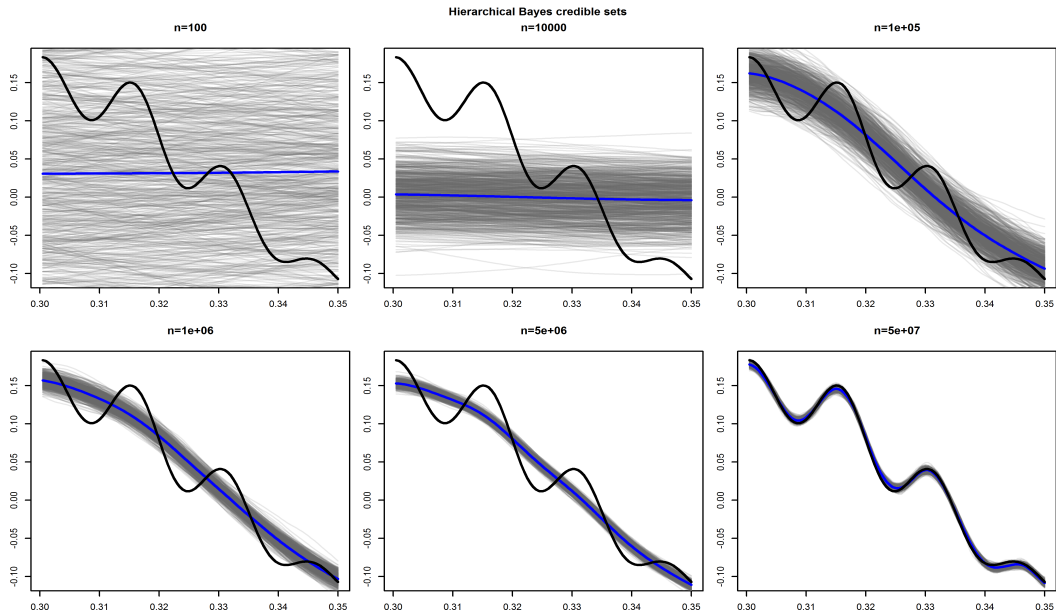


Figure 2: Hierarchical Bayes credible sets. The true function is drawn in black, the posterior mean in blue and the credible sets in grey. We have  $n = 10^2, 10^4, 10^5, 10^6, 5 \times 10^6$  and  $5 \times 10^7$ .

To demonstrate that the hierarchical Bayes method has bad coverage performance along a subsequence we plot the credible set given in (2.8) with uniform hyper-prior  $\lambda$  on  $[D, 2D]$  in Figure 2. Since the posterior distribution can not be computed explicitly we used an MCMC method generating draws from the hierarchical posterior distribution. As a first step we truncate the infinite dimensional vector  $\theta_0$  to its first  $N = n^{1/(1/2+2D)}$  coefficients  $\theta_0^N$ . This way the approximation  $\|\theta - \theta^N\|$  is of smaller order than the contraction rate. Then we apply a Metropolis within Gibbs sampling algorithm for sampling draws  $(\alpha, \theta^N)$  from the posterior distribution. We alternate between draws  $\theta^N | (\alpha, X)$  which can be done explicitly and  $\alpha | (\theta^N, X)$ . For the latter we use standard Metropolis-Hastings algorithm with uniform proposal distribution over  $[D, 2D]$  for  $\alpha$ . We choose a 3200 iteration long burn in period and then sample 800 draws. The implementation is straightforward hence we omit further description of the algorithm. One can see that similarly to the marginal likelihood empirical Bayes method the true function (in black) is not included in the credible sets for  $n = 10^4, 10^6$  and  $5 \times 10^6$ .

Finally in an attempt to illustrate the better coverage property of the risk based empirical Bayes credible sets we use the same simulated data as in the preceding two cases and plot

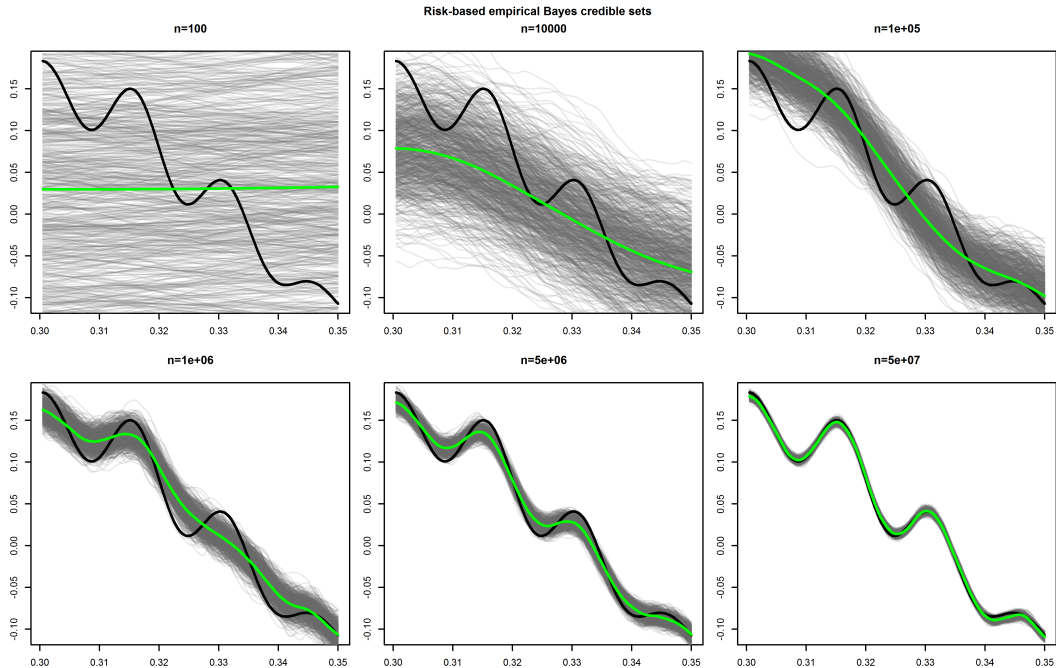


Figure 3: Risk based empirical Bayes credible sets. The true function is drawn in black, the posterior mean in green and the credible sets in grey. We have  $n = 10^2, 10^4, 10^5, 10^6, 5 \times 10^6$  and  $5 \times 10^7$ .

in Figure 3 the corresponding 95%-credible set. In the particular example we choose the parameter  $C_1$  (given in (2.11)) to be  $C_1 = 1/3$ , hence the constant  $C_0$  defined in (2.12) is 105.1. In the example we take  $C_0$  to be zero, because this way we avoid very conservative credible sets, i.e. credible sets from overly under smoothed posterior distributions ( $\tilde{\alpha}_n = D$ ). It can be clearly seen that the grey areas cover the true function  $\theta_0$  plotted in black even in the critical cases  $n = 10^4, 10^6$  and  $5 \times 10^6$  where the other two Bayesian procedures failed.

#### 4. Proof of Theorem 2.2

In the case that the hyper prior  $\lambda$  is the Dirac measure concentrated at  $\alpha_0 \in [D, 2D]$ , the hierarchical posterior distribution will be  $\Pi_{\alpha_0}(\cdot | X)$  see (2.2). By choosing  $\beta \neq \alpha_0$  and  $\theta_0 \in S^\beta(M)$  the posterior distribution  $\Pi_{\alpha_0}(\cdot | X)$  achieves a sub-optimal contraction rate around  $\theta_0$  following from [6], hence we get sub-optimal size for the credible set and/or coverage tending to zero.

It remains to deal with non-degenerate hyper-prior distributions  $\lambda$ , for which there exist some constants  $c > 0$  and  $D^* \in (D, 2D)$  such that  $\lambda(\alpha \in [D^*, 2D]) > c$ . One can see that for any given  $n$  there exists a parameter  $D' = D'(n) \in [D_1, 2D - 1/(\log n)]$  such that  $\lambda([D' + 1/(2 \log n), D' + 1/(\log n)]) > c/(2 \log n)$ . Furthermore take any  $\beta \in (D, D^*)$  and define  $\theta_0$  as in (2.7) (with small enough constant  $K$  as in Theorem 2.1). Finally let us introduce the notation

$$W_n(\alpha) = \hat{\theta}_\alpha - E_{\theta_0} \hat{\theta}_{n,\alpha}, \quad \text{and} \quad B_n(\alpha; \theta_0) = E_{\theta_0} \hat{\theta}_{n,\alpha} - \theta_0, \quad (4.1)$$

for the centered posterior mean and bias, respectively.

From triangle inequality we have

$$\begin{aligned} & \mathbb{P}_{\theta_0}(\|\hat{\theta}_{n_j} - \theta_0\| \leq L\hat{r}_{n_j,\gamma}) \leq \\ & \mathbb{P}_{\theta_0}\left(\|B_{n_j}(D', \theta_0)\| \leq \|\hat{\theta}_{n_j} - \hat{\theta}_{n_j,D'}\| + \|W_{n_j}(D')\| + L\hat{r}_{n_j,\gamma}\right), \end{aligned}$$

To prove that the right hand side of the preceding display tends to zero it is sufficient to show that there exist positive constants  $A_1, A_2, A_3, A_4$  such that

$$\|B_{n_j}(D'; \theta_0)\|^2 \geq A_1 n_j^{-2\beta/(1+2\beta)}, \quad (4.2)$$

$$\mathbb{P}_{\theta_0}(\|W_{n_j}(D')\|^2 \leq A_2 n_j^{-2D'/(1+2D')}) \rightarrow 1, \quad (4.3)$$

$$\mathbb{P}_{\theta_0}\left(\sup_{\alpha \in [D', 2D]} \|\hat{\theta}_{n_j} - \hat{\theta}_{n_j,\alpha}\|^2 \leq A_3 n_j^{-2D'/(1+2D')}\right) \rightarrow 1, \quad (4.4)$$

$$\mathbb{P}_{\theta_0}(\hat{r}_{n_j,\gamma}^2 \leq A_4 n_j^{-2D'/(1+2D')}) \rightarrow 1, \quad (4.5)$$

since  $n_j^{-2D'/(1+2D')} \ll n_j^{-2\beta/(1+2\beta)}$ . We note that (4.4) can be replaced by a weaker assertion but for the proof of (4.5) we need it in the present form. The proof of assertion (4.3) follows from (5.9) of [21] with  $\alpha = D'$  and  $p = 0$ , i.e.

$$\inf_{\theta_0 \in \ell^2} \mathbb{P}_{\theta_0}\left(\sup_{\alpha \geq \alpha_1} \|\hat{\theta}_{n,\alpha} - \mathbb{E}_{\theta_0} \hat{\theta}_{n,\alpha}\|^2 \leq 6n^{-2\alpha_1/(1+2\alpha_1)}\right) \rightarrow 1. \quad (4.6)$$

For (4.2) we have following from the explicit expression for  $\hat{\theta}_{D'}$  in (2.2), the inequality  $D' > \beta$  and the definition of  $\theta_0$  given in (2.7) with the notation  $N_j = n_j^{1/(1+2\beta)}$  that

$$\|B_{n_j}(D'; \theta_0)\|^2 \geq \sum_{N_j \leq i < 2N_j} \frac{i^{2+4D'} \theta_{0,i}^2}{(i^{1+2D'} + n_j)^2} \geq \frac{1}{4} \sum_{N_j \leq i < 2N_j} \theta_{0,i}^2 \gtrsim n_j^{-2\beta/(1+2\beta)}. \quad (4.7)$$

For the proofs of assertions (4.4) and (4.5) we refer to Sections 4.1 and 4.2, respectively.

#### 4.1. Proof of assertion (4.4)

By Jensen's inequality, Fubini's theorem and triangle inequality one can obtain that

$$\begin{aligned} & \sup_{\alpha_1 \in [D', 2D]} \|\hat{\theta}_{n_j} - \hat{\theta}_{n_j,\alpha_1}\|^2 = \sup_{\alpha_1 \in [D', 2D]} \left\| \int_D^{2D} (\hat{\theta}_{n_j,\alpha} - \hat{\theta}_{n_j,\alpha_1}) \lambda(d\alpha|X) \right\|^2 \\ & \leq \sup_{\alpha_1 \in [D', 2D]} \sum_{i=1}^{\infty} \int_D^{2D} (\hat{\theta}_{n_j,\alpha,i} - \hat{\theta}_{n_j,\alpha_1,i})^2 \lambda(d\alpha|X) \\ & \leq \sup_{\alpha_1, \alpha_2 \in [D', 2D]} \|\hat{\theta}_{n_j,\alpha_1} - \hat{\theta}_{n_j,\alpha_2}\|^2 \lambda(\alpha \in [D', 2D]|X) \\ & \quad + \sup_{\alpha_1, \alpha_2 \in [D, D']} \|\hat{\theta}_{n_j,\alpha_1} - \hat{\theta}_{n_j,\alpha_2}\|^2 \lambda(\alpha \in [D, D']|X). \end{aligned} \quad (4.8)$$

Next we show that both terms on the right hand side of (4.8) are bounded above by constant times  $n_j^{-2D'/(1+2D')}$  with  $P_{\theta_0}$ -probability tending to one.

Starting with the first term, we use the trivial bound 1 for the hyper-posterior probability. Furthermore we have

$$\begin{aligned} \sup_{\alpha_1, \alpha_2 \in [D', 2D]} \|\hat{\theta}_{n_j, \alpha_1} - \hat{\theta}_{n_j, \alpha_2}\|^2 &\leq \sup_{\alpha_1, \alpha_2 \in [D', 2D]} \|\mathbb{E}_{\theta_0} \hat{\theta}_{n_j, \alpha_1} - \mathbb{E}_{\theta_0} \hat{\theta}_{n_j, \alpha_2}\|^2 \\ &\quad + 2 \sup_{\alpha \in [D', 2D]} \|\hat{\theta}_{n_j, \alpha} - \mathbb{E}_{\theta_0} \hat{\theta}_{n_j, \alpha}\|^2. \end{aligned} \quad (4.9)$$

From (4.6) with  $\alpha_1 = D'$  the second term on the right hand side of (4.9) is bounded above by a multiple of  $n_j^{-2D'/(1+2D')}$  with  $P_{\theta_0}$ -probability tending to one. The first term on the right hand side of (4.9) can be written as  $\sup_{\alpha_1, \alpha_2 \in [D', 2D]} \sum_{i=1}^{\infty} (f_i(\alpha_1) - f_i(\alpha_2))^2$  for  $f_i(\alpha) = n_j \theta_{0,i} / (n_j + i^{1+2\alpha})$ . The derivative of  $f_i(\alpha)$  is  $-2 \log(i) i^{1+2\alpha} n_j \theta_{0,i} / (n_j + i^{1+2\alpha})^2$ . Writing the difference as the integral of  $f'_i(\alpha)$ , applying Cauchy-Schwarz inequality to its squares and then interchanging the sum and the integral we get that

$$\begin{aligned} \sum_{i=1}^{\infty} (f_i(\alpha_1) - f_i(\alpha_2))^2 &= \sum_{i=1}^{\infty} \left( \int_{\alpha_1}^{\alpha_2} f'_i(\alpha) d\alpha \right)^2 \leq \sum_{i=1}^{\infty} (\alpha_2 - \alpha_1) \int_{\alpha_1}^{\alpha_2} (f'_i(\alpha))^2 d\alpha \\ &= (\alpha_2 - \alpha_1) \int_{\alpha_1}^{\alpha_2} \sum_{i=1}^{\infty} (f'_i(\alpha))^2 d\alpha \leq (\alpha_2 - \alpha_1)^2 \sup_{\alpha \in [\alpha_1, \alpha_2]} \sum_{i=1}^{\infty} (f'_i(\alpha))^2 \\ &\leq 4D^2 \sup_{\alpha \in [D', 2D]} \sum_{i=1}^{\infty} \frac{n_j^2 \theta_{0,i}^2 (\log i)^2 i^{2+4\alpha}}{(n_j + i^{1+2\alpha})^4} \end{aligned} \quad (4.10)$$

Using the definition of  $\theta_0$  given in (2.7) and the lower bounds  $i^{1+2\alpha}$  and  $n_j$  for the term  $n_j + i^{1+2\alpha}$  in the denominator, the expression in the preceding display is bounded from above by constant times

$$\sum_{1 \leq i \leq 2n_{j-1}^{1/(1+2\beta)}} n_j^{-2} (\log i)^2 i^{1+8D-2\beta} + \sum_{i \geq n_j^{1/(1+2\beta)}} n_j^2 (\log i)^2 i^{-(3+4D'+2\beta)}. \quad (4.11)$$

The first term is bounded above by a multiple of  $(\log n_{j-1})^2 n_{j-1}^{(2+8D-2\beta)/(1+2\beta)} / n_j^2$ , which tends to zero faster than  $n_{j-1}^{-2D'/(1+2D')}$  following from the assumptions  $D < \beta < D'$  and  $n_j \geq n_{j-1}^{1+4D}$ . By Lemma 7.1 the second term of (4.11) is bounded above by

$$n_j^2 (\log n_j)^2 n_j^{-(2+4D'+2\beta)/(1+2\beta)} \ll (\log n_j)^2 n_j^{-2D'/(1+2\beta)} \ll n_j^{-2D'/(1+2D')},$$

following from  $\beta < D^* < D'$ .

It remained to deal with the second term on the right hand side of (4.8). Following from (4.9), (4.6) with  $\alpha_1 = D$  and triangle inequality we have with  $P_{\theta_0}$ -probability tending to one that

$$\sup_{\alpha_1, \alpha_2 \in [D, 2D]} \|\hat{\theta}_{n_j, \alpha_1} - \hat{\theta}_{n_j, \alpha_2}\|^2 \leq 4 \sup_{\alpha \in [D, 2D]} \|\mathbb{E}_{\theta_0} \hat{\theta}_{n_j, \alpha}\|^2 + 12n_j^{-D/(1+2D)} \leq 4M + o(1).$$



In Section 4.3 we prove that

$$\mathbb{E}_{\theta_0} \lambda(\alpha \in [D, D'] | X) \lesssim \exp(-c_1 n_j^{1/(1+4D)}) \log n_j, \quad (4.12)$$

for some constant  $c_1$ . Therefore by applying Markov's inequality one can obtain that the second term on the right hand side of (4.8) has smaller rate than any polynomial with  $\mathbb{P}_{\theta_0}$ -probability tending to one, which concludes the proof.

#### 4.2. Proof of assertion (4.5)

First we give a lower bound for the hierarchical posterior probability of the credible ball centered around the hierarchical posterior mean with radius  $\hat{r}_{n_j, \gamma}$ .

$$\Pi(\theta : \|\theta - \hat{\theta}_{n_j}\| < \hat{r}_{n_j, \gamma} | X) \geq \int_{D'}^{2D} \Pi_\alpha(\theta : \|\theta - \hat{\theta}_{n_j}\| < \hat{r}_{n_j, \gamma} | X) \lambda(d\alpha | X).$$

Then by applying triangle inequality one can observe that the right hand side of the preceding display is bounded from below by

$$\begin{aligned} & \int_{D'}^{2D} \Pi_\alpha(\theta : \|\theta - \hat{\theta}_{n_j, \alpha}\| + \|\hat{\theta}_{n_j} - \hat{\theta}_{n_j, \alpha}\| < \hat{r}_{n_j, \gamma} | X) \lambda(d\alpha | X) \geq \\ & \inf_{\alpha \in [D', 2D]} \Pi_\alpha(\theta : \|\theta - \hat{\theta}_{n_j, \alpha}\| + \|\hat{\theta}_{n_j} - \hat{\theta}_{n_j, \alpha}\| < \hat{r}_{n_j, \gamma} | X) \lambda(\alpha \in [D', 2D] | X). \end{aligned} \quad (4.13)$$

From assertion (4.12) and by applying Markov's inequality follows that the probability  $\lambda(\alpha \in [D', 2D] | X)$  is bigger than  $(1 - \gamma)/(1 - \gamma/2)$  with  $\mathbb{P}_{\theta_0}$ -probability tending to one. Therefore to prove (4.5) it is sufficient to show that there exists some large enough constant  $C$  such that with  $\mathbb{P}_{\theta_0}$ -probability tending to one

$$\inf_{\alpha \in [D', 2D]} \Pi_\alpha(\theta : \|\theta - \hat{\theta}_{n_j, \alpha}\| + \|\hat{\theta}_{n_j} - \hat{\theta}_{n_j, \alpha}\| < C n_j^{-D'/(1+2D')} | X) \geq 1 - \gamma/2. \quad (4.14)$$

Following from (5.7) and (6.1) of [21] we have for any compact interval  $[\alpha_1, \alpha_2]$  and for  $n \geq (10(1 + 3^{1+2\alpha_2})/(1 - \gamma))^{1+2\alpha_2}$  that

$$c n^{-\alpha_2/(1+2\alpha_2)} \leq \inf_{\alpha \in [\alpha_1, \alpha_2]} r_{n, \gamma}(\alpha) \leq \sup_{\alpha \in [\alpha_1, \alpha_2]} r_{n, \gamma}(\alpha) \leq C n^{-\alpha_1/(1+2\alpha_1)}, \quad (4.15)$$

with

$$c = (1 + 3^{1+2\alpha_2})^{-1/2} / \sqrt{2}, \quad \text{and} \quad C = \sqrt{3 + 2/\alpha_1}.$$

Therefore one can observe that for a fixed hyper-parameter  $\alpha \geq D'$  the radius of the  $1 - \gamma/2$ -credible ball is bounded above by a multiple of  $n_j^{-\alpha/(1+2\alpha)} \leq n_j^{-D'/(1+2D')}$ . Together with (4.4) this concludes the proof.

### 4.3. Proof of assertion (4.12)

In the proof of Theorem 2.1 we have shown that the likelihood function is monotone increasing on the interval  $[D, 2D] \supset [D, D' + 1/(\log n_j)]$ . Then by replacing  $\underline{\alpha}_n$  with  $D'$  and taking  $p = 0$  in the second paragraph of the proof of Theorem 2.5 of [14] we get that

$$\mathbb{E}_{\theta_0} \lambda(\alpha \in [D, D'] | X) \leq \frac{\exp(-Kn_j^{1/(1+2D')}/(1+2D'))}{\lambda(\alpha \in [D' + 1/(2 \log n_j), D' + 1/(\log n_j)])}. \quad (4.16)$$

From the definition of  $D' = D'(n_j)$  follows that the denominator on the right hand side of (4.16) is at least  $c/(2 \log n_j)$ . Therefore the right hand side of (4.16) is bounded above by a multiple of  $\exp(-K_1 n_j^{1/(1+4D)}) \log n_j$ .

## 5. Proof of Theorem 2.3

As a first step we investigate the behaviour of the new estimator of the hyper-parameter  $\tilde{\alpha}_n$ . We introduce the notation

$$\underline{\alpha}_n = \inf\{\alpha \geq D : B_{n,k_n}^2(\alpha; \theta_0) \geq (C_1^2/2)n^{-\frac{2\alpha}{1+2\alpha}}\} \wedge ((2D - C_0/\log n) \vee D), \quad (5.1)$$

$$\bar{\alpha}_n = \inf\{\alpha \geq D : B_{n,k_n}^2(\alpha; \theta_0) \geq 2C_1^2 n^{-\frac{2\alpha}{1+2\alpha}}\} \wedge ((2D - C_0/\log n) \vee D), \quad (5.2)$$

where  $B_{n,k_n}^2(\alpha; \theta_0)$  is defined in (2.9). The next lemma says that with high probability the estimator  $\tilde{\alpha}_n$  is going to be in the interval  $[\underline{\alpha}_n, \bar{\alpha}_n]$ .

**Lemma 5.1.** *For every positive  $\gamma$  and  $C_1$  and the constant  $C_0$  defined in (2.12) we have*

$$\liminf_{n \rightarrow \infty} \inf_{\theta_0 \in S^D(M)} P_{\theta_0}(\underline{\alpha}_n < \tilde{\alpha}_n < \bar{\alpha}_n) \geq 1 - \gamma/2. \quad (5.3)$$

Furthermore for all  $\beta \in [D, 2D]$

$$\inf_{\theta_0 \in S^{\beta}(M)} \underline{\alpha}_n > \beta - [(1/2 + \beta) \log(2M/C_1^2) \vee C_0] / \log n \quad (5.4)$$

*Proof.* See Section 6. □

Now we are ready to deal with the honest coverage assertion (2.14). From Lemma 5.1 we have

$$\inf_{\theta_0 \in S^D(M)} P_{\theta_0}(\theta_0 \in \hat{C}_n^R(L)) \geq \inf_{\theta_0 \in S^D(M)} P_{\theta_0} \left( \sup_{\underline{\alpha}_n \leq \alpha \leq \bar{\alpha}_n} \|\hat{\theta}_{n,\alpha} - \theta_0\| \leq L \inf_{\underline{\alpha}_n \leq \alpha \leq \bar{\alpha}_n} r_{n,\gamma}(\alpha) \right) - \gamma/2 + o(1).$$

Let us denote by  $B_n(\alpha; \theta_0)$  the bias  $\mathbb{E}_{\theta_0} \hat{\theta}_{n,\alpha} - \theta_0$  and by  $W_n(\alpha)$  the centered posterior mean  $\hat{\theta}_{n,\alpha} - \mathbb{E}_{\theta_0} \hat{\theta}_{n,\alpha}$ . From triangle inequality follows that for (2.14) it suffices to prove

$$\inf_{\theta_0 \in S^D(M)} P_{\theta_0} \left( \sup_{\underline{\alpha}_n \leq \alpha \leq \bar{\alpha}_n} \|W_n(\alpha)\| \leq \inf_{\underline{\alpha}_n \leq \alpha \leq \bar{\alpha}_n} L r_{n,\gamma}(\alpha) - \sup_{\underline{\alpha}_n \leq \alpha \leq \bar{\alpha}_n} \|B_n(\alpha; \theta_0)\| \right) \rightarrow 1.$$

To establish the preceding convergence we show that

$$\inf_{\alpha \in [\underline{\alpha}_n, \bar{\alpha}_n]} r_{n,\gamma}(\alpha) > (\sqrt{8(1+3^{1+4D})})^{-1} n^{-\underline{\alpha}_n/(1+2\underline{\alpha}_n)}, \quad (5.5)$$

$$\sup_{\alpha \in [\underline{\alpha}_n, \bar{\alpha}_n]} \|B_n(\alpha; \theta_0)\| \leq \sqrt{2(C_1^2 + M)} n^{-\underline{\alpha}_n/(1+2\underline{\alpha}_n)}, \quad (5.6)$$

$$\inf_{\theta \in S^D(M)} \mathbb{P}_{\theta_0} \left( \sup_{\underline{\alpha}_n \leq \alpha \leq \bar{\alpha}_n} \|W_n(\alpha)\| \leq \sqrt{6} n^{-\underline{\alpha}_n/(1+2\underline{\alpha}_n)} \right) \rightarrow 1. \quad (5.7)$$

Assertion (5.7) follows from (4.6) with  $\alpha_1 = \underline{\alpha}_n$ . For assertion (5.5) following from (4.15) it suffices to prove that

$$n^{-\underline{\alpha}_n/(1+2\underline{\alpha}_n)} \leq 2n^{-\bar{\alpha}_n/(1+2\bar{\alpha}_n)}. \quad (5.8)$$

If  $\underline{\alpha}_n = 2D - C_0/\log n$  or  $\bar{\alpha}_n = D$ , then necessarily  $\bar{\alpha}_n = 2D - C_0/\log n$  or  $\underline{\alpha}_n = D$  holds, respectively and therefore  $n^{-\underline{\alpha}_n/(1+2\underline{\alpha}_n)} = n^{-\bar{\alpha}_n/(1+2\bar{\alpha}_n)}$ . For  $\underline{\alpha}_n < 2D - C_0/\log n$  and  $D < \bar{\alpha}_n$  we have

$$(C_1^2/2)n^{-2\underline{\alpha}_n/(1+2\underline{\alpha}_n)} \leq B_{n,k_n}^2(\underline{\alpha}_n; \theta_0) \leq B_{n,k_n}^2(\bar{\alpha}_n; \theta_0) \leq 2C_1^2 n^{-2\bar{\alpha}_n/(1+2\bar{\alpha}_n)}.$$

To prove assertion (5.6) we divide the sum in  $\|B_n(\alpha; \theta_0)\|^2$  into two parts, from one to  $k_n = n^{1/(1/2+2D)}$  and from  $k_n + 1$  to infinity. From (2.10) we get that the second sum is bounded above by  $Mn^{-4D/(1+4D)}$ . To bound the sum from one to  $k_n$  we distinguish two cases. For  $\bar{\alpha}_n > D$  we have  $\|B_{n,k_n}(\bar{\alpha}_n, \theta_0)\|^2 \leq 2C_1^2 n^{-2\bar{\alpha}_n/(1+2\bar{\alpha}_n)}$  and for  $\bar{\alpha}_n = D$

$$\begin{aligned} \|B_{n,k_n}(D; \theta_0)\|^2 &= \sum_{i=1}^{k_n} \frac{i^{2+4D} \theta_{0,i}^2}{(i^{1+2D} + n)^2} \leq \sum_{i=1}^{n^{1/(1+2D)}} \frac{i^{2+2D} i^{2D} \theta_{0,i}^2}{n^2} + \sum_{i=n^{1/(1+2D)}}^{k_n} \frac{i^{2D} \theta_{0,i}^2}{i^{2D}} \\ &\leq n^{-\frac{2D}{1+2D}} \sum_{i=1}^{\infty} i^{2D} \theta_{0,i}^2 \leq Mn^{-\frac{2D}{1+2D}}. \end{aligned} \quad (5.9)$$

Finally we prove assertion (2.15). Since  $\tilde{\alpha}_n \in [\underline{\alpha}_n, \bar{\alpha}_n]$  with probability bigger than  $1-\gamma/2$ , from (4.15) follows that with probability bigger than  $1-\gamma/2$  the radius of the empirical Bayes credible ball is bounded above by  $(3+2/D)^{1/2} n^{-\underline{\alpha}_n/(1+2\underline{\alpha}_n)}$ . Then by applying assertion (5.4) and Lemma 7.2 we can conclude the proof.

## 6. Proof of Lemma 5.1

First we deal with assertion (5.3) and show separately that both  $\bar{\alpha}_n \geq \tilde{\alpha}_n$  and  $\underline{\alpha}_n \leq \tilde{\alpha}_n$  hold with probability bigger than  $1-\gamma/4$  uniformly over  $S^D(M)$ .

We start with the upper bound  $\bar{\alpha}_n \geq \tilde{\alpha}_n$ . Following from the definition of  $\tilde{\alpha}_n$  we have to deal only with the case  $\bar{\alpha}_n < 2D - C_0/\log n$ , where

$$B_{n,k_n}^2(\bar{\alpha}_n; \theta_0) \geq 2C_1^2 n^{-2\bar{\alpha}_n/(1+2\bar{\alpha}_n)} \quad (6.1)$$

holds. Then it is sufficient to prove that for  $\alpha \leq \bar{\alpha}_n$  and  $n \geq (M \vee 4C_1^2)^{(1+2D)(1+4D)}$

$$\text{var}_0 \left( \hat{B}_{n,k_n}^2(\alpha, \theta_0) \right) \leq 5n^{-8D/(1+4D)}, \quad (6.2)$$

because by (6.1) and (6.2)

$$\begin{aligned} \hat{B}_{n,k_n}^2(\bar{\alpha}_n) - C_1^2 n^{-\frac{2\bar{\alpha}_n}{1+2\bar{\alpha}_n}} &\geq C_1^2 n^{-\frac{2\bar{\alpha}_n}{1+2\bar{\alpha}_n}} + \hat{B}_{n,k_n}^2(\bar{\alpha}_n) - B_{n,k_n}^2(\bar{\alpha}_n; \theta_0) \\ &> C_1^2 n^{-\frac{2\bar{\alpha}_n}{1+2\bar{\alpha}_n}} - \sqrt{20/\gamma} n^{-\frac{4D}{1+4D}} \end{aligned}$$

holds with probability bigger than  $1 - \gamma/4$  following from (2.9) and Chebyshev's inequality. From Lemma 7.2 and  $\bar{\alpha}_n \leq 2D - C_0/\log n$  one can see that the right hand side of the preceding display is bounded below by  $(C_1^2 e^{2C_0/(1+4D)^2} - \sqrt{20/\gamma}) n^{-4D/(1+4D)}$ , which is positive following from the definition of  $C_0$  and therefore  $\bar{\alpha}_n \geq \tilde{\alpha}_n$ .

To show assertion (6.2) we note that  $\text{var}_{\theta_0} X_i^2 = 2\theta_{0,i}^2/n + 4/n^2$  and  $B_{n,k_n}^2(\alpha; \theta_0)$  is monotonically increasing, hence

$$\text{var}_{\theta_0} \left( \hat{B}_{n,k_n}^2(\alpha) \right) = \sum_{i=1}^{k_n} \frac{i^{4+8\alpha}}{(i^{1+2\alpha} + n)^4} \left( \frac{2\theta_{0,i}^2}{n} + \frac{4}{n^2} \right) \leq \frac{2B_{n,k_n}^2(\bar{\alpha}_n; \theta_0)}{n} + \frac{4k_n}{n^2}, \quad (6.3)$$

for all  $\alpha \leq \bar{\alpha}_n$ . By the choice  $k_n = n^{2/(1+4D)}$  the second term on the right hand side of (6.3) is bounded above by  $4n^{-8D/(1+4D)}$ . The first term of (6.3) for  $\bar{\alpha}_n > D$  is bounded by a multiple of  $n^{1/(1+2\bar{\alpha}_n)-2}$  and for  $\bar{\alpha}_n = D$  it is bounded by  $Mn^{-(1+4D)/(1+2D)}$  following from (5.9). Since both of the preceding rates are faster than  $n^{-8D/(1+4D)}$  this concludes the proof of assertion (6.2).

Next we deal with the lower bound  $\tilde{\alpha}_n \geq \underline{\alpha}_n$ , which holds trivially for  $\underline{\alpha}_n = D$ . Assume that  $\underline{\alpha}_n > D$  and denote the set of parameters satisfying this inequality by  $\Theta_n \subset S^D(M)$ . From triangle inequality we have

$$\begin{aligned} \sup_{\alpha \in [D, \underline{\alpha}_n]} \hat{B}_{n,k_n}^2(\alpha) - C_1^2 n^{-\frac{2\alpha}{1+2\alpha}} &\leq \sup_{\alpha \in [D, \underline{\alpha}_n]} \left( |\hat{B}_{n,k_n}^2(\alpha) - B_{n,k_n}^2(\alpha; \theta_0)| - (C_1^2/2)n^{-\frac{2\alpha}{1+2\alpha}} \right) \\ &\quad + \sup_{\alpha \in [D, \underline{\alpha}_n]} \left( B_{n,k_n}^2(\alpha; \theta_0) - (C_1^2/2)n^{-\frac{2\alpha}{1+2\alpha}} \right). \end{aligned} \quad (6.4)$$

For  $\theta_0 \in \Theta_n$  following from the definition of  $\underline{\alpha}_n$  the second term on the right hand side is non-positive. Next we show that

$$\sup_{\alpha \in [D, \underline{\alpha}_n]} \left( |\hat{B}_{n,k_n}^2(\alpha) - B_{n,k_n}^2(\alpha; \theta_0)| - (C_1^2/2)n^{-\frac{2\alpha}{1+2\alpha}} \right) \leq (24/\gamma)n^{-4D/(1+4D)} - (C_1^2/2)n^{-\frac{2\underline{\alpha}_n}{1+2\underline{\alpha}_n}}, \quad (6.5)$$

with probability bigger  $1 - \gamma/4$ . Then by Lemma 7.2 one can obtain that the right hand side of (6.5) and hence the right hand side of (6.4) is bounded above by  $[24/\gamma - (C_1^2/2)e^{2C_0/(1+4D)^2}]n^{-4D/(1+4D)} < 0$ . Therefore with probability bigger than  $1 - \gamma/4$  the lower bound  $\underline{\alpha}_n \leq \tilde{\alpha}_n$  holds.

To prove (6.5) following from Markov's inequality it suffices to show that for large enough  $n$  (depending only on  $D, C_1$  and  $\gamma$ )

$$\sup_{\theta_0 \in \Theta_n} \mathbb{E}_{\theta_0} \sup_{\alpha \in [D, \underline{\alpha}_n]} |\hat{B}_{n, k_n}^2(\alpha) - B_{n, k_n}^2(\alpha; \theta_0)| \leq 6n^{-4D/(1+4D)}.$$

By Corollary 2.2.5 [22] applied with  $\psi(x) = x^2$  the preceding inequality holds if

$$n^{8D/(1+4D)} \sup_{\theta_0 \in \Theta_n} \sup_{\alpha \in [D, \underline{\alpha}_n]} \text{var}_{\theta_0} (\hat{B}_{n, k_n}^2(\alpha)) \leq 5 \quad (6.6)$$

$$\sup_{\theta_0 \in \Theta_n} \int_0^{\text{diam}_n} \sqrt{N(\varepsilon, [D, \underline{\alpha}_n], d_n)} d\varepsilon \rightarrow 0, \quad (6.7)$$

where  $d_n$  is the semimetric defined by

$$d_n^2(\alpha_1, \alpha_2) = n^{8D/(1+4D)} \text{var}_{\theta_0} (\hat{B}_{n, k_n}^2(\alpha_1) - \hat{B}_{n, k_n}^2(\alpha_2)),$$

$N(\varepsilon, A, d)$  is the covering number of the set  $A$  with  $\varepsilon$ -balls relative to the semimetric  $d$  and  $\text{diam}_n$  is the diameter of the interval  $[D, \underline{\alpha}_n]$  relative to  $d_n$ .

The first assertion (6.6) follows immediately from (6.2). From triangle inequality one can observe that the diameter  $\text{diam}_n$  is bounded above by  $2\sqrt{5}$ . Furthermore for any  $\theta_0 \in \Theta_n$  and  $\alpha \in [D, \underline{\alpha}_n]$

$$\sup_{\theta_0 \in \Theta_n} B_{n, k_n}^2(\alpha; \theta_0) \leq 2C_1^2 n^{-2\alpha/(1+2\alpha)} \leq 2C_1^2 n^{-2D/(1+2D)}.$$

Therefore from Lemma 6.1 follows that for  $\underline{\alpha}_n \geq \alpha_2 > \alpha_1$  the semi-metric satisfies  $d_n(\alpha_1, \alpha_2) \lesssim \log(n)n^{-1/[2(1+4D)(1+2D)]} |\alpha_1 - \alpha_2|$ . Then the covering number is bounded above by

$$N(\varepsilon, [D, \underline{\alpha}_n], d_n) \lesssim \log(n)n^{-1/[2(1+4D)(1+2D)]} / \varepsilon,$$

hence the integral

$$\sup_{\theta_0 \in \Theta_n} \int_0^{2\sqrt{5}} \sqrt{N(\varepsilon, [D, \underline{\alpha}_n], d_n)} d\varepsilon \rightarrow 0.$$

It remained to prove assertion (5.4). For  $C > 0$  we have

$$\begin{aligned} \sup_{\theta \in S^\beta(M)} B_{n, k_n}^2\left(\beta - \frac{C}{\log n}, \theta\right) &\leq \sum_{1 \leq i \leq n} \frac{i^{2+2\beta-4C/\log n} i^{2\beta} \theta_{0,i}^2}{n^2} + \sum_{i > n} \frac{i^{2\beta} \theta_{0,i}^2}{i^{2\beta}} \\ &\leq \|\theta_0\|_\beta^2 n^{-\frac{2\beta}{1+2\beta-2C/\log n}} < M e^{-2C/(1+2\beta)} n^{-\frac{2\beta-2C/\log n}{1+2\beta-2C/\log n}}. \end{aligned}$$

For  $e^C \geq (2M/C_1^2)^{1/2+\beta}$  the constant multiplier on the right hand side is bounded above by  $C_1^2/2$ . Since the function  $B_{n, k_n}^2(\alpha; \theta_0)$  is monotonically increasing and  $f_n(\alpha) = n^{-2\alpha/(1+2\alpha)}$  is strictly monotonically decreasing one can conclude that for  $\alpha < \beta - C_1/\log n$  the inequality  $B_{n, k_n}^2(\alpha; \theta_0) < (C_1^2/2)n^{-2\alpha/(1+2\alpha)}$  holds and hence  $\underline{\alpha}_n \geq \beta - [(1/2+\beta) \log(2M/C_1^2) \vee C_0] / \log n$ .

**Lemma 6.1.** For any  $D \leq \alpha_1 < \alpha_2 \leq 2D$  we

$$\sup_{\theta_0 \in S^D(M)} \text{var}_{\theta_0} (\hat{B}_{n,k_n}^2(\alpha_1) - \hat{B}_{n,k_n}^2(\alpha_2)) \lesssim (\alpha_2 - \alpha_1)^2 (\log n)^2 (B_{n,k_n}^2(\alpha_2; \theta_0)/n + n^{-\frac{1+4D}{1+2D}})$$

*Proof.* The left hand side of the inequality in the lemma is equal to

$$\sum_{i=1}^{k_n} \left( \frac{i^{2+4\alpha_1}}{(i^{1+2\alpha_1} + n)^2} - \frac{i^{2+4\alpha_2}}{(i^{1+2\alpha_2} + n)^2} \right)^2 \text{var}_{\theta_0} X_i^2. \quad (6.8)$$

Furthermore the derivative of the function  $f_i(\alpha) = i^{2+4\alpha}(i^{1+2\alpha} + n)^{-2}$  satisfies

$$|f_i(\alpha)'| \leq 4n \log(i) i^{2+4\alpha} (i^{1+2\alpha} + n)^{-3}.$$

Applying  $\text{var}_{\theta_0} X_i^2 = 2\theta_{0,i}^2/n + 4/n^2$ , the preceding upper bound for  $|f_i(\alpha)'|$  and Cauchy-Schwarz inequality, similarly to (4.10) we get that (6.8) is bounded above by a multiple of

$$\begin{aligned} & (\alpha_2 - \alpha_1)^2 \log^2(n) \sup_{\alpha \in [\alpha_1, \alpha_2]} \frac{1}{n} \sum_{i=1}^{k_n} \frac{n^2 i^{4+8\alpha} \theta_{0,i}^2}{(i^{1+2\alpha} + n)^6} \\ & + (\alpha_2 - \alpha_1)^2 \log^2(n) \sup_{\alpha \in [\alpha_1, \alpha_2]} \frac{1}{n^2} \sum_{i=1}^{k_n} \frac{n^2 i^{4+8\alpha}}{(i^{1+2\alpha} + n)^6}. \end{aligned}$$

One can obtain that the first term on the right hand side is bounded above by constant times  $(\alpha_2 - \alpha_1)^2 \log^2(n) \sup_{\alpha \in [\alpha_1, \alpha_2]} B_{n,k_n}^2(\alpha; \theta_0)/n$ , while the second term by a multiple of  $(\alpha_2 - \alpha_1)^2 (\log n)^2 n^{-2+1/(1+2\alpha)}$ . The assertion of the lemma follows from the monotonically increasing property of  $B_{n,k_n}^2(\alpha; \theta_0)$ .  $\square$

## 7. Appendix

In this section we collected the proof of Theorem 2.1 (which is based on the proof of Theorem 3.1 of [21]), and additional auxiliary lemmas.

### 7.1. Proof of Theorem 2.1

First of all it is easy to see that for given  $M > 0$  and for arbitrary parameter  $\beta' \in [D, \beta]$  we can choose the constant  $K$  small enough (depending only on  $M$  and  $\beta - \beta'$ ) that  $\theta_0 \in S^{\beta'}(M)$ .

Then following from [14] and [20] we define the function

$$h_n(\alpha; \theta_0) = \frac{1 + 2\alpha}{n^{1/(1+2\alpha)} \log n} \sum_{i=1}^{\infty} \frac{n^2 i^{1+2\alpha} \log(i) \theta_{0,i}^2}{(i^{1+2\alpha} + n)^2}, \quad (7.1)$$

and introduce the variable

$$\underline{\alpha}_{0,n} = \inf\{\alpha > 0 : h_n(\alpha; \theta_0) > 1/16\} \wedge \sqrt{\log n}.$$

From Theorem 2.2 of [14] with  $p = 0$  follows that with probability tending to one the likelihood function is monotonically increasing for  $\alpha \leq \underline{\alpha}_{0,n} + 1/\log n$ , hence the maximum in  $[D, 2D]$  is taken for some  $\alpha \geq (\underline{\alpha}_{0,n} \vee D) \wedge 2D$ . We show below that  $\underline{\alpha}_{0,n_j} > 2D$ , hence with  $P_{\theta_0}$ -probability tending to one  $\hat{\alpha}_{n_j} = 2D$ .

Using the notation (4.1) we have that  $\theta_0 \in \hat{C}_n^E(L)$  if and only if  $\|\hat{\theta}_{\hat{\alpha}_n} - \theta_0\| \leq Lr_{n,\gamma}(\hat{\alpha}_n)$ , which implies that  $\|B_n(\hat{\alpha}_n; \theta_0)\| \leq Lr_{n,\gamma}(\hat{\alpha}_n) + \|W_n(\hat{\alpha}_n)\|$ . Combined with  $P_{\theta_0}(\hat{\alpha}_{n_j} = 2D) \rightarrow 1$  it follows that  $P_{\theta_0}(\theta_0 \in \hat{C}_n^E(L))$  is bounded above by

$$P_{\theta_0} \left( \|B_n(2D, \theta_0)\| \leq Lr_{n,\gamma}(2D) + \|W_n(2D)\| \right) + o(1). \quad (7.2)$$

Assertions (4.6) with  $\alpha_1 = 2D$ , (4.7) with  $D' = 2D$  and (4.15) with  $\alpha_1 = 2D$  show that

$$\begin{aligned} \sup_{\theta_0 \in \ell_2} r_{n,\gamma}(2D) &\lesssim n^{-2D/(1+4D)}, \\ \|B_n(2D; \theta_0)\|^2 &\gtrsim n^{-2\beta/(1+2\beta)} \\ \inf_{\theta_0 \in \ell_2} P_{\theta_0} \left( \|W(2D)\| \leq Cn^{-2D/(1+4D)} \right) &\rightarrow 1. \end{aligned}$$

Thus we deduce that the expression to the left of the inequality sign in (7.2) is of larger order than the expression to the right, whence the probability tends to zero along the subsequence  $n_j$ .

Finally we prove the claim that  $\underline{\alpha}_{0,n_j} \geq 2D$ , by showing that  $h_{n_j}(\alpha; \theta_0) < 1/16$  for all  $\alpha < 2D$ . Let  $N_j = n_j^{1/(1+2\beta)}$  then we have  $h_{n_j}(\alpha; \theta_0) \leq A_1 + A_2 + A_3$  for

$$\begin{aligned} A_1 &= \frac{1+2\alpha}{n_j^{1/(1+2\alpha)} \log n_j} \sum_{i \leq 2N_{j-1}} K i^{2\alpha-2\beta} \log i, \\ A_2 &= \frac{1+2\alpha}{n_j^{1/(1+2\alpha)} \log n_j} \sum_{N_j \leq i < 2N_j} \frac{n_j^2 i^{2\alpha-2\beta} (\log i) K 2^{-1-2\beta}}{(i^{1+2\alpha} + n_j)^2}, \\ A_3 &= \frac{1+2\alpha}{n_j^{1/(1+2\alpha)} \log n_j} \sum_{i \geq N_{j+1}} K n_j^2 i^{-2-2\alpha-2\beta} (\log i). \end{aligned}$$

For  $\alpha < 2D$ , so that  $i^{2\alpha-2\beta} \leq i^{2D}$ ,

$$A_1 \lesssim n_j^{-\frac{1}{1+2\alpha}} N_{j-1}^{1+2D} \lesssim n_{j-1} n_j^{-\frac{1}{1+4D}} \rightarrow 0,$$

since  $n_{j-1}^{1+4D} \ll n_j$ . By Lemma 7.1 the third term satisfies

$$A_3 \lesssim \frac{\log N_{j+1}}{\log n_j} n_j^{\frac{1+4\alpha}{1+2\alpha}} N_{j+1}^{-(1+2\alpha+2\beta)} \lesssim \frac{(\log n_{j+1})/n_{j+1}}{(\log n_j)/n_j} n_j^{\frac{2\alpha}{1+2\alpha}} n_{j+1}^{-\frac{2\alpha}{1+2\beta}}.$$

Because  $n_j^{1+4D} \ll n_{j+1}$  (and  $f(x) = (\log x)/x = 0$  is monotone decreasing for  $x \geq e$ ), this is also easily seen to vanish as  $j \rightarrow \infty$ . The term  $i^{1+2\alpha} + n_j$  in the denominator of the sum in

$A_2$  can be bounded below both by  $i^{1+2\alpha}$  and by  $n_j$ , and there are at most  $N_j$  terms in the sum. This shows that

$$\begin{aligned} A_2 &\lesssim \frac{n_j^{-1/(1+2\alpha)}}{\log n_j} N_j \left( \frac{n_j^2}{N_j^{1+2\alpha}} \wedge (2N_j)^{1+2\alpha} \right) \log(2N_j) N_j^{-1-2\beta} K \\ &\lesssim K \left( n_j^{\frac{1+2\beta-2\alpha}{1+2\beta} - \frac{1}{1+2\alpha}} \wedge n_j^{\frac{1+2\alpha-2\beta}{1+2\beta} - \frac{1}{1+2\alpha}} \right). \end{aligned}$$

The exponents of  $n_j$  in both terms in the minimum are equal to 0 at  $\alpha = \beta$ . For  $\alpha \geq \beta$  the first exponent is negative, whereas the second exponent is increasing in  $\alpha$  and hence negative for  $\alpha < \beta$ . It follows that  $A_2 \lesssim K$ .

Putting things together we see that  $\limsup_{j \rightarrow \infty} \sup_{\alpha \leq 2D} h_{n_j}(\alpha; \theta_0)$  can be made arbitrarily small by choosing  $K$  sufficiently small.

## 7.2. Auxiliary lemmas

We collected the auxiliary lemmas in this section.

**Lemma 7.1** (Lemma 10.4 of [21]). *For  $k > 0$ ,  $m \geq 0$ , and  $N \geq e^{2m/k}$ ,*

$$\sum_{i > N} i^{-1-k} (\log i)^m \leq (1/N + 2/k) (\log N)^m N^{-k}.$$

**Lemma 7.2.** *For the function  $f_n(\alpha) = n^{-2\alpha/(1+2\alpha)}$ ,  $\alpha \in [D, 2D]$ ,  $K > 0$  and  $n \geq e^{K/4}$  we have*

$$\begin{aligned} e^{2K/(1+4D)^2} f_n(\alpha) &\leq f_n(\alpha - K/\log n) \leq e^{2K/(1/2+2D)^2} f_n(\alpha), \\ e^{-2K/(1+2D)^2} f_n(\alpha) &\leq f_n(\alpha + K/\log n) \leq e^{-2K/(1+4D)^2} f_n(\alpha). \end{aligned}$$

*Proof.* Since  $(\log f_n(\alpha))' = -2(\log n)/(1+2\alpha)^2$  we have

$$\begin{aligned} \log \frac{f_n(\alpha - K/\log n)}{f_n(\alpha)} &\geq (-K/\log n)(-2 \log n)/(1+2\alpha)^2 \geq K/(1+4D)^2, \\ \log \frac{f_n(\alpha + K/\log n)}{f_n(\alpha)} &\leq K/(1+2\alpha - 2K/\log n)^2 \leq K/(1/2+2D)^2, \end{aligned}$$

where the second inequality in the second line holds for  $n \geq e^{K/4}$ . The second part of the lemma follows similarly.  $\square$

## References

- [1] BROWN, L. D., AND LOW, M. G. Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24**, 6 (1996), 2384–2398.
- [2] BULL, A. Honest adaptive confidence bands and self-similar functions. *Electron. J. Statist.* **6** (2012), 1490–1516.
- [3] BULL, A., AND NICKL, R. Adaptive confidence sets in  $l^2$ . *Probability Theory and Related Fields* **156**, 3-4 (2013), 889–919.



- [4] CAI, T. T., AND LOW, M. G. An adaptation theory for nonparametric confidence intervals. *Ann. Statist.* 32, 5 (2004), 1805–1840.
- [5] CASTILLO, I. Lower bounds for posterior rates with Gaussian process priors. *Electron. J. Stat.* 2 (2008), 1281–1299.
- [6] CASTILLO, I. Lower bounds for posterior rates with Gaussian process priors. *Electron. J. Stat.* 2 (2008), 1281–1299.
- [7] GENOVESE, C., AND WASSERMAN, L. Adaptive confidence bands. *Ann. Statist.* 36, 2 (2008), 875–905.
- [8] GINÉ, E., AND NICKL, R. Confidence bands in density estimation. *Ann. Statist.* 38, 2 (2010), 1122–1170.
- [9] HENGARTNER, N. W., AND B., S. P. Finite-sample confidence envelopes for shape-restricted densities. *Ann. Statist.* 23, 2 (1995), 525–550.
- [10] INGSTER, Y., AND SUSLINA, I. A. *Nonparametric goodness-of-fit testing under Gaussian models*, vol. 169. Springer, 2003.
- [11] INGSTER, Y. I. Asymptotic minimax nonparametric testing for independent sample density hypothesis. *Zapiski Nauchnykh Seminarov POMI* 136 (1984), 74–96.
- [12] JUDITSKY, A., AND LAMBERT-LACROIX, S. On nonparametric confidence set estimation. *Math. Meth. of Stat* 19, 4 (2003), 410–428.
- [13] KNAPIK, B., VAN DER VAART, A. W., AND VAN ZANTEN, J. H. Bayesian inverse problems with gaussian priors. *Ann. Statist.* 39, 5 (2011), 2626–2657.
- [14] KNAPIK, B. T., SZABÓ, B. T., VAN DER VAART, A. W., AND VAN ZANTEN, J. H. Bayes procedures for adaptive inference in inverse problems for the white noise model. *ArXiv e-prints* (Sept. 2012).
- [15] LOW, M. G. On nonparametric confidence intervals. *Ann. Statist.* 25, 6 (1997), 2547–2554.
- [16] NUSSBAUM, M. Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* 24, 6 (1996), 2399–2430.
- [17] PICARD, D., AND TRIBOULEY, K. Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.* 28, 1 (2000), 298–335.
- [18] PINSKER, M. S. Optimal filtration of square-integrable signals in Gaussian noise.
- [19] ROBINS, J., AND VAN DER VAART, A. W. Adaptive nonparametric confidence sets. *Ann. Statist.* 34, 1 (2006), 229–253.
- [20] SZABO, B., VAN DER VAART, A., AND ZANTEN, J. Empirical Bayes scaling of Gaussian priors in the white noise model. *Electron. J. Statist.* 7 (2013), 991–1018.
- [21] SZABO, B. T., VAN DER VAART, A. W., AND VAN ZANTEN, J. H. Frequentist coverage of adaptive nonparametric Bayesian credible sets. *ArXiv e-prints* (Oct. 2013).
- [22] VAN DER VAART, A. W., AND WELLNER, J. A. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.