ISSN 1066-5307, Mathematical Methods of Statistics, 2008, Vol. 17, No. 1, pp. 1–18. © Allerton Press, Inc., 2008.

Empirical Bayesian Test of the Smoothness

E. Belitser[1*] and F. Enikeeva[2]

[1] *Mathematical Institute, Utrecht University, The Netherlands*

[2] *EURANDOM, Eindhoven, The Netherlands,*
and
Institute for Information Transmission Problems of RAS, Moscow, Russia

Received June 18, 2007; in final form, November 27, 2007

Abstract—In the context of adaptive nonparametric curve estimation a common assumption is that a function (signal) to estimate belongs to a nested family of functional classes. These classes are often parametrized by a quantity representing the smoothness of the signal. It has already been realized by many that the problem of estimating the smoothness is not sensible. What can then be inferred about the smoothness? The paper attempts to answer this question. We consider implications of our results to hypothesis testing about the smoothness and smoothness classification problem. The test statistic is based on the empirical Bayes approach, i.e., it is the marginalized maximum likelihood estimator of the smoothness parameter for an appropriate prior distribution on the unknown signal.

Key words: empirical Bayes approach, hypothesis testing, smoothness parameter, white noise model.

2000 Mathematics Subject Classification: primary 62G10; secondary 62C10, 62G20, 62C12.

DOI: 10.3103/S1066530708010018

1. INTRODUCTION

Suppose we observe Gaussian data $X = X^{(n)} = (X_i)_{i\in\mathbb{N}}$, where $X_i \sim \mathcal{N}(\theta_i, n^{-1})$, the X_i's are independent, $\theta = (\theta_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ is an unknown parameter. This model is the sequential version of the Gaussian white noise model $dY(t) = f(t)\,dt + n^{-1/2}\,dW(t)$, $t \in [0, 1]$, where $f \in \mathcal{L}_2[0, 1] = \mathcal{L}_2$ is an unknown signal and W is the standard Brownian motion. If $\theta \in \ell_2 = \{\theta\colon \sum_{k=1}^{\infty} \theta_k^2 < \infty\}$, the infinite-dimensional parameter θ can be regarded as a sequence of the Fourier coefficients of $f \in \mathcal{L}_2$ with respect to some orthonormal basis in $\mathcal{L}_2$. Sometimes we will call θ a signal. We assume that $\theta \in \Theta \subseteq \ell_2$, where $\Theta = \cup_{\beta\in\mathcal{B}}\Theta_\beta$ and $\beta \in \mathcal{B}$ has the meaning of smoothness parameter. Here we consider only one-dimensional $\beta \in \mathcal{B} \subseteq \mathbb{R}_+ = [0, +\infty)$ and a family of Sobolev type sets $\{\Theta_\beta\}_{\beta\in\mathcal{B}}$: $\Theta_{\beta_2} \subseteq \Theta_{\beta_1}$ if $\beta_1 \le \beta_2$. Our goal is to make an inference on the smoothness of the parameter θ. More precisely, we are going to test the hypothesis about the smoothness of θ.

The white noise model attracted attention in the last few decades. Its comprehensive treatments can be found in [19] and [22]. Besides being of interest in its own (the problem of recovering a signal transmitted over a communication channel with Gaussian white noise of intensity $n^{-1/2}$), the white noise model turns out to be a mathematical idealization of some other nonparametric models. For instance, the white noise model arises as a limiting experiment as $n \to \infty$, for the model of n i.i.d. observations with unknown density and for the regression model (see [29] and [6]). On the other hand, this model captures the statistical essence of the original model and preserves its main features in a pure form; cf. [22]. Most of the statistical problems are studied in asymptotic setup from the viewpoint of increasing information $n \to \infty$. In fact, one deals with a sequence of models parametrized by n. Though non-asymptotic estimation problems are also very important, they are often not tractable mathematically. Besides, very often, non-asymptotic results become interesting and useful only for a

*E-mail: belitser@math.uu.nl

sufficiently large value of the information parameter n, i.e., they are essentially of asymptotic nature. Our approach is also primarily asymptotic. However, the intention is to derive non-asymptotic results as well (where at all feasible), to be able to evaluate precisely the influence of different quantities and constants on the quality of the inference. To simplify the notation in this paper, we omit sometimes the dependence of relevant quantities on n.

Many statistical problems for the white noise model have been already studied in the literature: signal estimation under different norms, estimation of a functional of the signal, hypothesis testing about the signal, construction of confidence sets. We name just a few references: [19], [31], [20], [14], [3], [18], [22] (see references therein), [23], [30], [5], [8], [24], [21] (see references therein), [33] (see references therein). To compare different statistical inference procedures, one can use the minimax approach, oracle inequalities, maxisets. A typical approach to the problems mentioned above is to assume that the unknown signal θ belongs to some set $\Theta_\beta \subset \ell_2$ indexed by $\beta \in \mathcal{B}$, which represents the smoothness. If the parameter β is known, then we are in a single model situation and we can use this knowledge in making inference about θ; for instance, signal estimation, functional estimation, testing hypothesis, confidence set. If the parameter β is unknown (multiple model situation: $\theta \in \cup_{\beta\in\mathcal{B}}\Theta_\beta$), an adaptation problem arises.

In the last two decades, several adaptation methods (primarily for the estimation problem) have been developed, to name a few: blockwise method (see, e.g., [15], [8], [7]), Lepski's method [25], [26], [27], [28], wavelet thresholding ([13], later developed in many other papers), penalization method ([1] and further references therein, [23], [5]), Bayesian methods ([2], [4], [17]). Some methods are designed for rather specific settings: e.g., blockwise method for the white noise sequence model with the mean squared risk. Some of them are more general, e.g., Lepski's method, which could be extended to different settings (various risk functions, multidimensional case) and even to different statistical problems: estimation of a functional of a signal, the problem of adaptive hypothesis testing.

One of the ingredients of some adaptation methods mentioned above (as Bayesian methods, Lepski's method, and the method of penalized estimators) is the problem of data-based choice $\hat\beta = \hat\beta(X^{(n)})$ for the structural parameter $\beta \in \mathcal{B}$ which marks the smoothness. One can thus regard this attendant problem as the smoothness selection problem (or the model selection problem). Typically, in a single model situation a standard (optimal in some sense) inference procedure on θ is available, i.e., in fact one has a family of nonadaptive inference procedures parametrized by $\beta \in \mathcal{B}$ at one's disposal. Then a good smoothness selection method combined with this family of procedures should lead to a good adaptive inference procedure simply by choosing the inference procedure with the selected smoothness. Ideally, we would like our adaptive inference procedure to be of the same quality as if we knew the true $\beta \in \mathcal{B}$ for which $\theta \in \Theta_\beta$, or, if this is impossible, with the smallest possible loss of quality. Actually, even if we know that $\theta \in \Theta_{\beta_1}$ for some $\beta_1 \in \mathcal{B}$ and there is an optimal (in some sense) inference procedure available for this situation, it may still be more advantageous to use an adaptive procedure instead. Indeed, a good smoothness selection method may pick some other $\beta_2 \neq \beta_1$ which may lead to a better quality simply because the underlying θ may also satisfy $\theta \in \Theta_{\beta_2}$. Even if $\theta \notin \Theta_{\beta_2}$, it still may be "very close" to Θ_{β_2}, so that the quality of the procedure corresponding to β_2 is better.

It is a folklore belief that it is impossible "to estimate the smoothness". We, however, deliberately avoid words "estimation of the smoothness" and use the term "smoothness selection" instead. The point is that the problem of selecting the smoothness, on its own, does not really make sense, since it is not quite clear how to characterize the amount of smoothness that a particular signal has (in other words, which $\beta \in \mathcal{B}$ is the most appropriate to a certain θ) and how to compare different smoothness selection methods if we do not specify for what purpose we need to select the smoothness parameter $\beta \in \mathcal{B}$. Thus, the problem of smoothness selection is only sensible in connection with the underlying statistical problem.

In this paper we are trying to test the hypothesis that the parameter θ belongs to some set Θ_{β_0}, where the value $\beta_0 \in \mathcal{B}$ is known. Loosely speaking, this corresponds to testing the hypothesis $\bar\beta(\theta) \geq \beta_0$ against $\bar\beta(\theta) < \beta_0$, where $\bar\beta(\theta)$ has the meaning of smoothness of signal θ. We use a version of the empirical Bayes approach which is due to Robbins [32]. We fix the family of Bayes estimators $\hat\theta(\beta) = \hat\theta(\beta, X)$ with respect to the priors π_β, $\beta \in \mathcal{B}$, chosen in such a way that the Bayes estimator $\hat\theta(\beta)$ is rate minimax over the Sobolev ball of smoothness β in the problem of estimating the signal θ in ℓ_2-norm. In the next section we propose some heuristic guiding idea how to check whether a certain prior

π_β adequately reflects the requirement $\theta \in \Theta_\beta$. Next, we propose a smoothness selection procedure $\hat\beta = \hat\beta(X)$ based on maximizing the restricted marginal likelihood (a version of the empirical Bayes approach). Our main goals in this paper are to study the asymptotic properties of this smoothness selection method. Namely, we look at these properties from the point of view of hypothesis testing about the smoothness of the signal and discuss some applications to the smoothness classification problem.

The paper is organized as follows. Section 2 describes the empirical Bayes approach. The main results are given in Section 3. We prove auxiliary lemmas in Section 4.

2. EMPIRICAL BAYES APPROACH

Let $\{\Theta_\beta\}_{\beta\in\mathcal{B}}$, $\mathcal{B} = [\kappa, +\infty)$ for some $\kappa > 0$, be a family of Sobolev type subspaces of ℓ_2:

$$\Theta_\beta = \Big\{\theta\colon \sum_{i=1}^{\infty} i^{2\beta}\theta_i^2 < \infty\Big\}.$$

Many quantities will depend on this constant κ, but we will skip this dependence throughout the paper to make the notation easier. We suppose that $\theta \in \Theta_\beta$ for some unknown $\beta \in \mathcal{B}$.

For a particular $\theta \in \cup_{\beta\in\mathcal{B}}\Theta_\beta$ define the function

$$A_\theta(\beta) = \sum_{i=1}^{\infty} i^{2\beta}\theta_i^2, \qquad \beta \in \mathcal{B}. \tag{1}$$

It is a monotone function of β. Note that $\theta \in \Theta_\beta$ if and only if $A_\theta(\beta) < \infty$. Throughout the paper we assume that there exists $\bar\beta \in \mathcal{B}$ such that $\bar\beta = \bar\beta(\theta) = \sup\{\beta \in \mathcal{B}\colon A_\theta(\beta) < \infty\}$. We can interpret $\bar\beta = \bar\beta(\theta)$ as the *smoothness of* θ. Two possibilities may occur: either $A_\theta(\beta) \to \infty$ as $\beta \uparrow \bar\beta$ or $A_\theta(\bar\beta) < \infty$ and $A_\theta(\beta) = \infty$ for all $\beta > \bar\beta$. It is the behavior of the function $A_\theta(\beta)$ that effectively measures the smoothness of the underlying signal θ. Unless otherwise specified, we assume from now on that $A_\theta(\bar\beta) = \infty$.

The goal of this paper is to make an inference about the smoothness of the signal on the basis of the observed data X. The inference will be based on a statistic $\hat\beta(X)$ (it has an intuitive meaning of the smoothness selector), which we construct using the empirical Bayes approach. In the next section we will make this problem mathematically formal by evaluating so-called probabilities of undersmoothing and oversmoothing for this statistic. In the rest of this section we describe the construction of $\hat\beta(X)$. The idea of the approach is to put a "right" prior $\pi(\beta)$ on the parameter θ, find the marginal distribution of X, which will depend on β, and then use the marginal maximum likelihood estimator of β as the smoothness selection procedure.

We need to clarify the choice of the right family of priors $\pi(\beta)$, $\beta \in \mathcal{B}$. As is well illustrated in a series of papers by Diaconis and Freedman (see [9], [10], [11], [12] and [16]), an arbitrary choice of the prior may lead to Bayesian procedures that easily fail in infinite-dimensional problems. An appropriate prior should reflect adequately the smoothness assumption on the unknown signal. There are many ways to describe this. Here we propose the following guiding principle, which adapts to the inference problem on θ. For example, the inference problems can be estimation of θ, estimation of a functional of θ, testing hypotheses, constructing confidence set. Usually these problems come with their own performance criteria, like the rate of convergence for the estimation problem. A particular prior leads to the corresponding Bayes procedure. We can look at its performance, according to the given criteria, from the two different perspectives: frequentists ($X^{(n)} \sim P_\theta^{(n)}$) and Bayesian ($X^{(n)} \sim P_\beta^{(n)}$, marginal of $X^{(n)}$). Thus, a prior is considered to be not unreasonable (and potentially right) if it provides the same high performance, with respect to the given criteria, of the resulting Bayes procedure simultaneously under both Bayesian and frequentists formulations. For instance, in the case of an estimation problem, Bayesian estimator should be a minimax estimator, at least with respect to the convergence rate.

This principle should not be taken as a precise prescription, but rather as a starting point in the choice of "correct" priors in infinite-dimensional statistical problems. After all, one will have to investigate the performance of the resulting Bayesian procedure in each particular statistical problem in order to claim that a certain prior is right for that problem. The choice of the prior surely depends on the underlying inference problem on θ, which is in our case the problem of signal estimation in ℓ_2-norm. Thus, in this paper we consider the following version of the above principle: we take the underlying inference problem on θ to be the problem of estimating θ in ℓ_2-norm. Next, we should choose a prior leading to a Bayes estimator that is at least rate optimal in the minimax sense over the corresponding class with smoothness β. The minimax ℓ_2-rate over the Sobolev ellipsoid of smoothness β is $n^{-2\beta/(2\beta+1)}$ (see [31]) and the Bayes risk of our estimator should attain the same convergence rate. We put the following prior $\pi = \pi(\beta)$ on θ: the θ_i's are independent and for $\delta > 1 - 2\beta$

$$\theta_i \sim \mathcal{N}(0, \tau_i^2(\beta)), \qquad \tau_i^2(\beta) = \tau_i^2(\beta,\delta,n) = n^{\frac{\delta-1}{2\beta+1}} i^{-(2\beta+\delta)}, \quad i \in \mathbb{N}. \tag{2}$$

Recall the following simple fact: if $Z \mid Y \sim \mathcal{N}(Y, \sigma^2)$ and $Y \sim \mathcal{N}(\mu, \tau^2)$, then

$$Y \mid Z \sim \mathcal{N}\left(\frac{Z\tau^2 + \mu\sigma^2}{\tau^2+\sigma^2}, \frac{\tau^2\sigma^2}{\tau^2+\sigma^2}\right).$$

Let E_π denote the expectation with respect to the prior π. The Bayesian estimator of θ based on the above prior is the vector $\hat\theta = \hat\theta(\beta) = (\hat\theta_i)_{i\in\mathbb{N}}$ with components

$$\hat\theta_i = \hat\theta_i(\beta) = \mathrm{E}(\theta_i \mid X_i) = \frac{\tau_i^2(\beta) X_i}{\tau_i^2(\beta) + n^{-1}}, \qquad i \in \mathbb{N}. \tag{3}$$

The choice of the prior and the variance (2) is made according to our principle as the following lemma shows.

For $0 < p < \infty$, $0 < q < \infty$, $0 \le r < \infty$ such that $pq > r + 1$ denote

$$B(p,q,r) = \int_0^\infty \frac{u^r}{(1+u^p)^q}\, du = p^{-1}\,\mathrm{Beta}\left(q - \frac{r+1}{p}, \frac{r+1}{p}\right), \tag{4}$$

where, for $\alpha, \beta > 0$, $\mathrm{Beta}(\alpha,\beta) = \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\, du$ is the Beta function.

Lemma 1. *Let $\hat\theta$ be defined by* (3). *Then, as $n \to \infty$,*

$$\mathrm{E}_\pi\|\theta - \hat\theta\|^2 = n^{-2\beta/(2\beta+1)} B(2\beta+\delta, 1, 0)(1+o(1)),$$
$$\mathrm{E}_\theta\|\theta - \hat\theta\|^2 \le n^{-2\beta/(2\beta+1)} \big(A_\theta(\beta) C(\beta,\delta) + B(2\beta+\delta, 2, 0)\big)(1+o(1)),$$
$$\mathrm{E}_\theta\|\theta - \hat\theta\|^2 \ge B(2\beta+\delta, 2, 0) n^{-2\beta/(2\beta+1)}(1+o(1)),$$

where $C(\beta,\delta) = \frac{(1+\delta\beta^{-1})^{2(\beta+\delta)/(2\beta+\delta)}}{(2+\delta\beta^{-1})^2}$ and the function B is defined by (4).

Proof. By (2) and Lemma 9, we evaluate the Bayes risk:

$$\mathrm{E}_\pi\|\theta - \hat\theta\|^2 = \sum_{i=1}^\infty \frac{\tau_i^2(\beta) n^{-1}}{\tau_i^2(\beta) + n^{-1}} = \sum_{i=1}^\infty \frac{n^{(\delta-1)/(2\beta+1)}}{n^{(2\beta+\delta)/(2\beta+1)} + i^{2\beta+\delta}}$$
$$= n^{-2\beta/(2\beta+1)} B(2\beta+\delta, 1, 0)(1+o(1))$$

as $n \to \infty$. The frequentist risk consists of two terms

$$\mathrm{E}_\theta\|\theta - \hat\theta\|^2 = \mathrm{E}_\theta \sum_{i=1}^\infty \left(\frac{\tau_i^2(\beta) X_i}{\tau_i^2(\beta) + n^{-1}} - \theta_i\right)^2$$
$$= \sum_{i=1}^\infty \frac{n^{-2}\theta_i^2}{(\tau_i^2(\beta) + n^{-1})^2} + \sum_{i=1}^\infty \frac{n^{-1}\tau_i^4(\beta)}{(\tau_i^2(\beta) + n^{-1})^2}.$$

Using again (2) and Lemma 9, we bound these terms as follows: as $n \to \infty$,

$$\sum_{i=1}^{\infty} \frac{n^{-2}\theta_i^2}{(\tau_i^2(\beta)+n^{-1})^2} = \sum_{i=1}^{\infty} \frac{i^{2(2\beta+\delta)}\theta_i^2}{(n^{(2\beta+\delta)/(2\beta+1)}+i^{2\beta+\delta})^2}$$
$$\le A_\theta(\beta) \max_{i\in\mathbb{N}} \frac{i^{2\beta+2\delta}}{\big(n^{(2\beta+\delta)/(2\beta+1)}+i^{2\beta+\delta}\big)^2}$$
$$= A_\theta(\beta) C(\beta,\delta) n^{-2\beta/(2\beta+1)}(1+o(1)),$$
$$\sum_{i=1}^{\infty} \frac{n^{-1}\tau_i^4(\beta)}{(\tau_i^2(\beta)+n^{-1})^2} = \sum_{i=1}^{\infty} \frac{n^{(2\beta+2\delta-1)/(2\beta+1)}}{\big(n^{(2\beta+\delta)/(2\beta+1)}+i^{2\beta+\delta}\big)^2}$$
$$= n^{-2\beta/(2\beta+1)} B(2\beta+\delta,2,0)(1+o(1)).$$

The lemma is proved. □

Below we present another lemma, which justifies in a way the choice of the variance of the prior distribution. This lemma says that if θ belongs to Θ_β, then the estimator $\hat\theta$ belongs to the same set with probability one.

Lemma 2. *Let $\theta \in \Theta_\beta$ for some $\beta \in \mathcal{B}$. Then*

$$\lim_{T\to\infty} \sup_{n\ge 1} \mathrm{P}_\theta\Big\{\sum_{i=1}^{\infty} \hat\theta_i^2 i^{2\beta} > T\Big\} = 0.$$

Proof. By the Markov inequality,

$$\mathrm{P}_\theta\Big\{\sum_{i=1}^{\infty} \hat\theta_i^2 i^{2\beta} > T\Big\} \le T^{-1}\sum_{i=1}^{\infty} i^{2\beta}\mathrm{E}_\theta[\hat\theta_i^2].$$

Note that

$$\mathrm{E}_\theta[\hat\theta_i^2] = \frac{\tau_i^4(\beta)\mathrm{E}_\theta X_i^2}{(\tau_i^2(\beta)+n^{-1})^2} = \frac{n^{2(2\beta+\delta)/(2\beta+1)}(\theta_i^2+n^{-1})}{(n^{(2\beta+\delta)/(2\beta+1)}+i^{2\beta+\delta})^2}.$$

Applying Lemma 9 (see also the remark following that lemma), we evaluate

$$\mathrm{P}_\theta\Big\{\sum_{i=1}^{\infty} i^{2\beta}\hat\theta_i^2 > T\Big\} \le \frac{n^{2(2\beta+\delta)/(2\beta+1)}}{T}\sum_{i=1}^{\infty} \frac{i^{2\beta}(\theta_i^2+n^{-1})}{(n^{(2\beta+\delta)/(2\beta+1)}+i^{2\beta+\delta})^2}$$
$$\le \frac{A_\theta(\beta)}{T} + \frac{n^{2(2\beta+\delta)/(2\beta+1)-1}}{T}\sum_{i=1}^{\infty}\frac{i^{2\beta}}{(n^{(2\beta+\delta)/(2\beta+1)}+i^{2\beta+\delta})^2}$$
$$\le \frac{A_\theta(\beta)}{T} + \frac{B(2\beta+\delta,2,2\beta)+1}{T},$$

which completes the proof of the lemma. □

Recall that we have the following marginal distribution of X: X_i's are independent and $X_i \sim \mathcal{N}\big(0,\tau_i^2(\beta)+n^{-1}\big)$, $i\in\mathbb{N}$. Let $L_n(\beta)=L_n(\beta,X)$ be the marginal likelihood of the data $X=(X_i)_{i\in\mathbb{N}}$:

$$L_n(\beta) = \prod_{i=1}^{\infty} \frac{1}{\sqrt{2\pi\big(\tau_i^2(\beta)+n^{-1}\big)}}\exp\Big\{-\frac{X_i^2}{2\big(\tau_i^2(\beta)+n^{-1}\big)}\Big\}.$$

Maximizing the function $L_n(\beta)$ is equivalent to minimizing $-\log L_n(\beta)$. To avoid complications in defining the minimum of $-\log L_n(\beta)$ under the event $\{-\log L_n(\beta)=\infty\}$, it is convenient to introduce

$Z_n(\beta) = Z_n(\beta, \beta_0) = -2\log\frac{L_n(\beta)}{L_n(\beta_0)}$ (which is almost surely finite) for some reference value $\beta_0 \in \mathcal{B}$. For any set $S \subseteq \mathcal{B}$, define a marginal likelihood estimator of β restricted to the set S:

$$\hat{\beta} = \hat{\beta}(S) = \hat{\beta}(S, X, n) = \arg\min_{\beta \in S} Z_n(\beta), \tag{5}$$

which is a version of empirical Bayes approach (see [32]). This means that $Z_n(\hat{\beta}(S)) \le Z_n(\beta)$ for all $\beta \in S$ or, equivalently,

$$\sum_{i=1}^{\infty} \frac{\big(\tau_i^2(\beta_0) - \tau_i^2(\hat{\beta}(S))\big)X_i^2}{(\tau_i^2(\beta_0) + n^{-1})(\tau_i^2(\hat{\beta}(S)) + n^{-1})} + \sum_{i=1}^{\infty} \log\frac{\tau_i^2(\hat{\beta}(S)) + n^{-1}}{\tau_i^2(\beta_0) + n^{-1}}$$
$$\le \sum_{i=1}^{\infty} \frac{\big(\tau_i^2(\beta_0) - \tau_i^2(\beta)\big)X_i^2}{(\tau_i^2(\beta_0) + n^{-1})(\tau_i^2(\beta) + n^{-1})} + \sum_{i=1}^{\infty} \log\frac{\tau_i^2(\beta) + n^{-1}}{\tau_i^2(\beta_0) + n^{-1}}$$

for all $\beta \in S$. It follows also that $Z_n(\hat{\beta}(S), \beta) \le 0$ for all $\beta \in S$.

Denote for brevity

$$a_i = a_i(\beta, \beta_0) = \frac{1}{\tau_i^2(\beta) + n^{-1}} - \frac{1}{\tau_i^2(\beta_0) + n^{-1}}, \tag{6}$$

$$b_i = b_i(\beta, \beta_0) = \frac{\tau_i^2(\beta) + n^{-1}}{\tau_i^2(\beta_0) + n^{-1}} = 1 + \frac{\tau_i^2(\beta) - \tau_i^2(\beta_0)}{\tau_i^2(\beta_0) + n^{-1}}. \tag{7}$$

Then $Z_n(\beta, \beta_0) = \sum_{i=1}^{\infty} a_i(\beta, \beta_0)X_i^2 + \sum_{i=1}^{\infty} \log b_i(\beta, \beta_0)$, and for all $\beta \in S$

$$\sum_{i=1}^{\infty} a_i(\hat{\beta}(S), \beta)X_i^2 \le \sum_{i=1}^{\infty} \log\big[b_i(\hat{\beta}(S), \beta)\big]^{-1}. \tag{8}$$

Remark 1. It is not so difficult to check that the above $\hat{\beta}$ can be related to a penalized least square estimator with the penalty $\mathrm{pen}(\theta, \beta) = \sum_{i=1}^{\infty}\big[\theta_i^2\tau_i^{-2}(\beta) + \log\big(\tau_i^2(\beta) + n^{-1}\big)\big]$. Indeed,

$$\hat{\beta}(S) = \arg\min_{\beta \in S} Z_n(\beta) = \arg\min_{\beta \in S}\min_{\theta}\Big\{n\sum_{i=1}^{\infty}(X_i - \theta_i)^2 + \mathrm{pen}(\theta, \beta)\Big\}.$$

Remark 2. In fact, we could assume $\kappa = \kappa_n \downarrow 0$ as $n \to \infty$ sufficiently slowly, so that all the results still hold true.

Remark 3. From now on we will assume the set $S = S_n$ to be finite, the exact assumptions are given in the next section. We have chosen to minimize the process $Z_n(\beta)$ over some finite set $S = S_n$ to avoid unnecessary technical complications. Indeed, we could also take S to be the whole set $\mathcal{B}$ and then study the behavior of a (near) minimum point of $Z_n(\beta)$. The usual technique in such cases inspired by the empirical processes theory is to consider the minimum over some finite grid in $\mathcal{B}$ and to make sure at the same time that the increments of the process $Z_n(\beta)$ are uniformly small over small intervals (provided the process is smooth enough). We do not pursue this approach simply because it boils down to the same considerations as in the case when we restrict the minimization to the finite set S_n from the very beginning.

3. MAIN RESULTS

First introduce some notation. For a constant $Q > 0$ denote by $\Theta_\beta(\bar{\theta}, Q) \subseteq \ell_2$ the Sobolev ellipsoid of "size" Q and smoothness β around the point $\bar{\theta} \in \ell_2$: $\Theta_\beta(\bar{\theta}, Q) = \{\theta\colon A_{\theta-\bar{\theta}}(\beta) = \sum_{i=1}^{\infty} i^{2\beta}(\theta_i - \bar{\theta}_i)^2 \le Q\}$. Denote further $\Theta_\beta(Q) = \Theta_\beta(0, Q)$. For a set B, denote by $|B|$ the cardinality of B, which may be infinite. For a nonempty set $B \subseteq \mathbb{R}$, define $\lfloor x\rfloor_B = \sup\{y \in B\colon y \le x\}$ if $\inf\{B\} < x$, and $\lfloor x\rfloor_B = \inf\{B\}$ otherwise; $\lceil x\rceil_B = \inf\{y \in B\colon y \ge x\}$ if $\sup\{B\} > x$, and $\lceil x\rceil_B = \sup\{B\}$ otherwise. Note that if B is finite, then $\lfloor x\rfloor_B, \lceil x\rceil_B \in B$. Denote $\lfloor x\rfloor = \lfloor x\rfloor_{\mathbb{Z}}$. Define the sets $S_n^-(\beta) = S_n^-(\beta, S_n) = \{\beta' \in S_n\colon \beta' \le \beta\}$ and $S_n^+(\beta) = S_n^+(\beta, S_n) = \{\beta' \in S_n\colon \beta' \ge \beta\}$.

From now on we make the following assumptions.

(i) We set $\delta = 1$ in definition (2) of the prior variances τ_i^2, unless otherwise is specified.

(ii) The set S is assumed to be finite, dependent on n in such a way that $S = S_n$ forms an ε_n-net in $[\kappa, \sup\{S_n\}]$, with $\varepsilon_n = O\big(1/(\log n)\big)$ and $\sup\{S_n\} \to \infty$ as $n \to \infty$.

The requirement $\varepsilon_n = O\big(1/(\log n)\big)$ stems from the fact that if $|\beta_1 - \beta_2| = O\big(1/(\log n)\big)$ as $n \to \infty$, then $n^{2\beta_1/(2\beta_1+1)} = O\big(n^{2\beta_2/(2\beta_2+1)}\big)$ and $n^{2\beta_2/(2\beta_2+1)} = O\big(n^{2\beta_1/(2\beta_1+1)}\big)$. Later we will impose a certain upper bound on $|S_n|$. There are many possible choices of the set S_n: for example, the choice $\varepsilon_n = (\log n)n^{-1}$ and $S_n = \{\kappa + k\epsilon_n,\ k = 0, 1, \ldots, n\}$ will do.

Recall that we observe independent Gaussian data $X_i \sim \mathcal{N}(\theta_i, n^{-1})$, $i = 1, 2 \ldots$, with unknown $\theta = (\theta_i)_{i=1}^{\infty}$. Informally, we would like now to test the hypothesis H_0: the smoothness of the signal θ is at least β_0. The alternative H_1: the smoothness of the signal θ is less than β_0. Although intuitively appealing, this is not a proper hypothesis testing problem yet. It should be of the form $H_0\colon P \in \mathcal{P}_0$ against $H_1\colon P \in \mathcal{P}_1$; a family of probability measures $\mathcal{P}_0$ against another family of probability measures $\mathcal{P}_1$. In our case $X \sim P_\theta = P_\theta^{(n)}$ and we can formalize H_0 as follows: $H_0\colon P_\theta,\ \theta \in \Theta_{\beta_0}$ or $H_0\colon P_\theta,\ \theta \in \{\theta\colon \bar{\beta}(\theta) > \beta_0\}$. It would be ideal to test this hypothesis against the alternative $H_1\colon \theta \in \ell_2 \setminus \Theta_{\beta_0}$ or $H_1\colon P_\theta,\ \theta \in \{\theta\colon \bar{\beta}(\theta) \le \beta_0\}$. However, for a test to be consistent against all the above alternatives, this set of alternatives is too large and some of the alternatives are "too close" to the null hypothesis set. A typical approach in such a situation is to restrict the set of alternatives (see [21]). This actually means that we remove a sort of indifference zone from the complement of the set Θ_{β_0}.

Let us introduce a restricted set of alternatives. Define, for some nonnegative sequence Δ_n,

$$\begin{aligned} V_\theta &= V_\theta(\Delta_n) = V_\theta(\Delta_n, n) \\ &= \left\{\beta \in S_n\colon \exists \beta' = \beta'(\beta) \in S_n,\ \beta' \le \beta,\ \frac{1}{2}\sum_{i=1}^{\infty} \frac{a_i(\beta,\beta')(\theta_i^2 - \tau_i^2(\beta'))}{1 + n^{-1}a_i(\beta,\beta')} \ge \Delta_n\right\} \end{aligned}$$

and

$$\Lambda_\beta = \Lambda_\beta(\Delta_n) = \Lambda_\beta(\Delta_n, n) = \{\theta \in \ell_2\colon S_n \cap [\beta, \infty) \subseteq V_\theta(\Delta_n)\},$$

where a_i are defined in (6).

Next, introduce the decision rule

$$\psi = \psi_n(X, \beta) = \mathbf{1}\{\hat{\beta}(X) \le \beta\},$$

where $\hat{\beta}(X)$ is the marginal maximum likelihood smoothness selector (5).

We use the decision rule $\psi_n(X, \beta_0 - \delta_n)$, with an appropriately chosen sequence δ_n, $\delta_n \to 0$, to test the hypothesis

$$H_0\colon \theta \in \Theta_{\beta_0} \quad \text{against} \quad H_1\colon \theta \in \Lambda_{\beta_0 - \delta_n}.$$

Thus, the set $\Lambda_{\beta_0-\delta_n}(\Delta_n, n)$ is the set of alternatives in our testing problem and the probabilities of type I and II errors for the test $\psi_n(X, \beta_0 - \delta_n)$ are

$$\begin{aligned} \alpha_1(\theta, \beta_0 - \delta_n, n) &= \mathrm{E}_\theta \psi_n(X, \beta_0 - \delta_n) = \mathrm{P}_\theta\{\hat{\beta} \le \beta_0 - \delta_n\}, \quad \theta \in \Theta_{\beta_0}, \\ \alpha_2(\theta, \beta_0 - \delta_n, n) &= \mathrm{E}_\theta(1 - \psi_n(X, \beta_0 - \delta_n)) = \mathrm{P}_\theta\{\hat{\beta} > \beta_0 - \delta_n\}, \quad \theta \in \Lambda_{\beta_0-\delta_n}, \end{aligned}$$

respectively.

Theorem 1. *Let $\beta_0 \in \mathcal{B}$, $Q > 0$, $\delta_n = \delta_n(c) = \frac{c}{\log n}$.*

1. *For any $C > 0$ there exists a positive $C_0 = C_0(\beta_0, Q, C)$ such that for all $c \ge C_0$*

$$\sup_{\theta \in \Theta_{\beta_0}(Q)} \alpha_1(\theta, \beta_0 - \delta_n, n) = \sup_{\theta \in \Theta_{\beta_0}(Q)} \mathrm{P}_\theta\{\hat{\beta} \le \beta_0 - \delta_n\} \le |S_n^-(\beta_0 - \delta_n)| \exp\big\{-Cn^{1/(2\beta_0+1)}\big\}.$$

2. *For any $\beta \in \mathcal{B}$ (in particular, for $\beta = \beta_0 - \delta_n$)*

$$\sup_{\theta \in \Lambda_\beta(\Delta_n)} \alpha_2(\theta, \beta, n) = \sup_{\theta \in \Lambda_\beta(\Delta_n)} \mathrm{P}_\theta\{\hat{\beta} > \beta\} \le |S_n^+(\beta)| \exp\big\{-\Delta_n\big\}.$$

Proof. Recall that $\varepsilon_n = O\big(1/(\log n)\big)$ as $n \to \infty$. Therefore, if $\beta \le \beta_0 - \frac{c}{\log n}$ with $c \ge C_0$ and $C_0 = C_0(\beta_0, Q, C)$ large enough, then there exists $\beta_0' \in S_n$ such that $\beta \le \beta_0' - \frac{C_2}{\log n} \le \beta_0 - \frac{C_1 + C_2}{\log n}$ for constant $C_1 = C_1(\beta_0, Q)$ from Lemma 7 and positive $C_2 = C_2(\beta_0, C)$ to be specified later. This holds, for example, for $C_0 = 3\max\{C_1, C_2, \varepsilon_n \log n\}$.

Since $\beta \le \beta_0' \le \beta_0 - \frac{C_1}{\log n}$, by Lemma 7 we obtain that for all $n \in \mathbb{N}$

$$\sup_{\theta\in\Theta_{\beta_0}(Q)} \mathrm{P}_\theta\{\hat\beta(S_n) \le \beta_0 - \delta_n\} \le \sum_{\beta \in S_n^-(\beta_0-\delta_n)} \sup_{\theta\in\Theta_{\beta_0}(Q)} \mathrm{P}_\theta\{\hat\beta = \beta\}$$
$$\le |S_n^-(\beta_0 - \delta_n)| \exp\left\{\frac{B(1,2,0)n^{1/(2\beta_0'+1)}}{2} + \frac{5}{8} - \frac{n^{1/(\beta+\beta_0'+1)}}{16}\right\}.$$

Denote for brevity $\delta_n' = \frac{C_2}{\log n}$ and $B = B(1,2,0)$, where the function B is defined by (4). As $\beta \le \beta_0' - \delta_n'$, the expression in the exponent of the last relation can be bounded as follows:

$$\frac{B(1,2,0)n^{1/(2\beta_0'+1)}}{2} + \frac{5}{8} - \frac{n^{1/(\beta+\beta_0'+1)}}{16} = \frac{B}{2} n^{1/(2\beta_0'+1)} + \frac{5}{8} - \frac{n^{1/(2\beta_0'+1-\delta_n')}}{16}$$
$$= \left(\frac{B}{2} + \frac{5}{8} n^{-1/(2\beta_0'+1)} - \frac{1}{16} n^{\frac{\delta_n'}{(2\beta_0'+1-\delta_n)(2\beta_0'+1)}}\right) n^{1/(2\beta_0'+1)} \le \left(\frac{B}{2} + \frac{5}{8} - \frac{n^{\delta_n'/(2\beta_0'+1)^2}}{16}\right) n^{1/(2\beta_0'+1)}$$
$$= \left(\frac{B}{2} + \frac{5}{8} - \frac{1}{16}\exp\left\{\frac{C_2}{(2\beta_0'+1)^2}\right\}\right) n^{1/(2\beta_0'+1)} = -Cn^{1/(2\beta_0'+1)} \le -Cn^{1/(2\beta_0+1)}$$

with $C_2 = (2\beta_0+1)^2 \log(16(\frac{B}{2} + \frac{5}{8} + C))$. The first assertion of the theorem follows for

$$C_0 = 3\max\{C_1, C_2, \varepsilon_n \log n\}.$$

The second assertion of the theorem follows immediately from the definition of the set $\Lambda_n(\beta)$ and Lemma 6. Indeed,

$$\mathrm{P}_\theta\{\hat\beta(S_n) > \beta\} \le \sum_{\beta\in S_n^+(\beta)} \mathrm{P}_\theta\{\hat\beta = \beta\}$$
$$\le |S_n^+(\beta)| \exp\left\{\frac{1}{2}\sum_{i=1}^\infty \frac{a_i(\beta,\beta')\big(\tau_i^2(\beta') - \theta_i^2\big)}{1 + n^{-1}a_i(\beta,\beta')}\right\} \le |S_n^+(\beta)| \exp\{-\Delta_n\}.$$

□

Remark 4. Of course, according to the second assertion of the above theorem, we can make the set of alternatives $\Lambda_\beta(\Delta_n)$ larger by taking a larger $\beta > \beta_0 - \delta_n$, for instance, for $\Lambda_{\beta_0}(\Delta_n)$ instead of $\Lambda_{\beta_0-\delta_n}(\Delta_n)$. The problem is then that an indifference zone $[\beta_0 - \delta_n, \beta]$ appears for the test statistic $\hat\beta$. Namely, the above theorem provides the claimed upper bound for the probability of type II error only if $\hat\beta > \beta$ and not for $\hat\beta \in [\beta_0 - \delta_n, \beta]$.

The smaller δ_n and Δ_n, the bigger the set of alternatives $\Lambda_{\beta_0-\delta_n}(\Delta_n)$ is. On the other hand, the upper bound for the probability of type II error has the term $e^{-\Delta_n}$, so that taking Δ_n smaller makes the probability of type II error higher. Also, as the above theorem shows, the sequence δ_n has to be at least $c(\log n)^{-1}$ for a sufficiently large constant c in order to make the probability of type I error small. Thus, there is some kind of trade-off between different aspects of the problem: an improvement upon one aspect leads to the deterioration on the other.

For any $\beta < \bar\beta(\theta)$ it is reasonable to call $\mathrm{P}_\theta\big\{\hat\beta(S_n) \le \beta\big\}$ the probability of undersmoothing. Given $\theta \in \Theta_{\beta_0}$, i.e., $\beta_0 < \bar\beta(\theta)$, we see that the probability of type I error $\alpha_1(\theta, \beta_0, \delta_n, n)$ is actually the probability of undersmoothing $\mathrm{P}_\theta\big\{\hat\beta(S_n) \le \beta_0 - \delta_n\big\}$, which we would like to be converging to zero, with δ_n tuned as precisely as possible. The first assertion of the theorem claims that the probability of undersmoothing converges to zero as $n \to \infty$ for properly chosen δ_n. It says essentially that if

$\beta_0 < \bar{\beta}(\theta)$, then our selection procedure picks values β which are smaller than β_0 with exponentially small probability. Asymptotically, there is no probability mass on $(\kappa, \bar{\beta}(\theta) - \varepsilon]$.

On the other hand, if $\theta \notin \Theta_{\beta_0}$, i.e., $\beta_0 \geq \bar{\beta}(\theta)$, $\mathrm{P}_\theta\{\hat{\beta}(S_n) \geq \beta\}$ can be regarded as the probability of "oversmoothing". We would like our selection procedure to pick the "oversmoothed" values $\beta \geq \beta_0$ also with small probability: $\mathrm{P}_\theta\{\hat{\beta}(S_n) \geq \beta\} \to 0$. However, we could not establish that the probability of oversmoothing converges to zero for all θ such that $\beta_0 \geq \bar{\beta}(\theta)$ (or $\theta \notin \Theta_{\beta_0}$). We established this fact only for $\theta \in \Lambda_{\beta_0 - \delta_n}(\Delta_n)$, which is essentially a subset of the complement of Θ_{β_0} (see lemma below). Thus, there is a sort of buffer zone between Θ_{β_0} and $\Lambda_{\beta_0 - \delta_n}(\Delta_n)$ on which our selection procedure cannot distinguish. It is impossible to get rid of this uncertainty: making the buffer zone for θ smaller leads to the appearance of an indifference zone for β (see the remark above).

We give some heuristic arguments why this buffer zone should appear. Recall that our empirical Bayes selection procedure is based on the prior designed to match the Bayes and frequentists versions of the ℓ_2-risk signal estimation problem. The bias and variance of the estimator $\hat{\theta}(\beta)$ are respectively increasing and decreasing functions of β. The best choice of β is the one for which the bias and the variance terms are balanced, they should be at least of the same order. Consider now the estimator $\hat{\theta}(\hat{\beta})$. For small values of $\hat{\beta}$, the variance term of the risk will dominate the bias term, the undersmoothing situation. Big values of $\hat{\beta}$ will eventually lead to oversmoothing: bias will dominate the variance. Presumably, the buffer zone consists of those θ for which the bias and variance terms of the risk of $\hat{\theta}(\hat{\beta})$ are balanced up to the order.

At a glance it is unclear how the sets Θ_{β_0} and $\Lambda_{\beta_0 - \delta_n}$ are related to each other. If $\delta_n \to 0$ and $\Delta_n \to \infty$ as $n \to \infty$, then we should have $\Theta_{\beta_0} \cap \Lambda_{\beta_0 - \delta_n} = \varnothing$ for sufficiently large n. The following lemmas describe in some sense the relation between the sets Θ_β and Λ_β and which θ's are contained in Λ_β.

Lemma 3. *Let $\beta_0 \in \mathcal{B}$. If $\theta \in \Theta_{\beta_0}$ and $\Delta_n = cn^{1/(2\beta_0+1)}$, then there exists $N = N(\theta, \beta_0, c)$ such that $\theta \notin \Lambda_{\beta_0}(\Delta_n)$ for all $n \geq N$.*

Proof. Due to the assumptions made on the set S_n, we can assume without loss of generality that $\beta_0 \in S_n$. Indeed, since $\beta_0 < \bar{\beta}(\theta)$, also $\lceil \beta_0 \rceil_{S_n} < \bar{\beta}(\theta)$ for all sufficiently large n, and we can use $\lceil \beta_0 \rceil_{S_n}$ instead of β_0 everywhere in the proof.

For any $\beta' \leq \beta_0$ we have that $0 \leq a_i(\beta_0, \beta') \leq n$. Therefore, for any $M \in \mathbb{N}$,

$$\sum_{i=1}^{\infty} \frac{a_i(\beta_0, \beta')(\theta_i^2 - \tau_i^2(\beta'))}{1 + n^{-1} a_i(\beta_0, \beta')} \leq \sum_{i=1}^{\infty} a_i(\beta_0, \beta')\theta_i^2 \leq \sum_{i=1}^{\infty} \frac{\theta_i^2}{\tau_i^2(\beta_0) + n^{-1}}$$
$$\leq M \sum_{i=1}^{M} i^{2\beta_0}\theta_i^2 + \frac{n}{(M+1)^{2\beta_0}} \sum_{i=M+1}^{\infty} i^{2\beta_0}\theta_i^2.$$

Take $M = M_n = M_n(\beta_0, A_\theta(\beta_0) = \lfloor \varepsilon(2A_\theta(\beta_0))^{-1} n^{1/(2\beta_0+1)} \rfloor$, so that the first term in the right-hand side of the last inequality is not greater than $\varepsilon n^{1/(2\beta_0+1)}/2$. Next, since $A_\theta(\beta_0) < \infty$, there exists $N = N(\theta, \beta_0, \varepsilon)$ such that for all $n \geq N$,

$$\sum_{i=M_n+1}^{\infty} i^{2\beta_0}\theta_i^2 \leq \frac{\varepsilon}{2}(2A_\theta(\beta_0))^{-1}/\varepsilon)^{-2\beta_0}.$$

Thus above relations imply that

$$\sum_{i=1}^{\infty} \frac{a_i(\beta_0, \beta')(\theta_i^2 - \tau_i^2(\beta'))}{1 + n^{-1} a_i(\beta_0, \beta')} \leq \varepsilon n^{1/(2\beta_0+1)}$$

for all $n \geq N(\theta, \beta_0, \varepsilon)$. Take $\varepsilon = c/2$, then $\beta_0 \notin V_\theta(\Delta_n)$ and thus $\theta \notin \Lambda_{\beta_0}(\Delta_n)$, which concludes the proof of the lemma. □

The next lemma refines slightly the previous one if we assume that the points of the set S_n are distant from each other by at least $O\big((\log n)^{-1}\big)$.

Lemma 4. *Let $\beta_0 \in \mathcal{B}$. If $\theta \in \Theta_{\beta_0}$ and $\min_{\beta_1,\beta_2\in S_n} |\beta_1-\beta_2| \ge \frac{c}{\log n}$, then there exists $N = N(\theta,\beta_0,c)$ such that $\theta \notin \Lambda_{\beta_0}(0)$ for all $n \ge N$.*

Proof. Without loss of generality assume $\beta_0 \in S_n$. Note first that $0 \le a_i(\beta_0,\beta') \le n$ for any $\beta' \le \beta_0$. Therefore,

$$\sum_{i=1}^{\infty}\frac{a_i(\beta_0,\beta')(\theta_i^2-\tau_i^2(\beta'))}{1+n^{-1}a_i(\beta_0,\beta')} \le \sum_{i=1}^{\infty} a_i(\beta_0,\beta')\theta_i^2 - \frac{1}{2}\sum_{i=1}^{\infty} a_i(\beta_0,\beta')\tau_i^2(\beta').$$

Since $\min_{\beta_1,\beta_2\in S_n}|\beta_1-\beta_2| \ge \frac{c}{\log n}$, we have for any $\beta' \in S_n$ such that $\beta' < \beta_0$ that $\beta_0 \ge \beta' + \frac{c}{\log n}$. Using this, Lemmas 9 and 10, we obtain that for some $C = C(\beta_0,c)$

$$\sum_{i=1}^{\infty} a_i(\beta_0,\beta')\tau_i^2(\beta') = B(2\beta_0+1,1,2(\beta_0-\beta'))n^{1-2\beta'/(2\beta_0+1)}(1+o(1)) \\ - B(2\beta'+1,1,0)n^{1/(2\beta'+1)}(1+o(1)) \ge Cn^{1/(2\beta_0+1)}$$

for all $n \ge N_1(\beta_0,c)$. In the previous lemma, we proved that

$$\sum_{i=1}^{\infty} a_i(\beta_0,\beta')\theta_i^2 \le \varepsilon n^{1/(2\beta_0+1)}$$

for all $n \ge N_2(\theta,\beta_0,\varepsilon)$. Take $\varepsilon = C(\beta_0,c)/2$ and $N = \max\{N_1,N_2\}$, to get that $\beta_0 \notin V_\theta(0)$ and thus $\theta \notin \Lambda_{\beta_0}(0)$ for all $n \ge N$. □

Lemma 5. *Let $\beta_0 \in \mathcal{B}$. If $\beta_0 > \bar\beta(\theta)$ (i.e., $\theta \notin \Theta_{\beta_0}$), then for any $C > 0$ and any $N \in \mathbb{N}$ there exists $n = n(\theta,C) \ge N$ such that $\theta \in \Lambda_{\beta_0+1/2}(\Delta_n)$ with $\Delta_n = Cn^{1/(2\beta_0+2)}$.*

Proof. Denote $\varepsilon = \beta_0 - \bar\beta(\theta)$. Lemma 8 implies that for any $\beta',\beta \in S_n$, $\beta' \le \beta$,

$$\sum_{i=1}^{\infty} a_i(\beta,\beta')\tau_i^2(\beta') \le \sum_{i=1}^{\infty}\frac{\tau_i^2(\beta')}{\tau_i^2(\beta)+n^{-1}} \le C(\kappa)n^{1-2\beta'/(2\beta_0+1)} \tag{9}$$

for some $C(\kappa)$.

Next, for any $\beta' < \beta$, $0 \le \delta \le 1-\exp\{-2(\log 2)(\beta-\beta')\}$ and $T_n = \lfloor n^{1/(2\beta+1)}\rfloor$ we have

$$\sum_{i=1}^{\infty} a_i(\beta,\beta')\theta_i^2 \ge \delta\sum_{i=2}^{\infty}\frac{n^2 i^{2\beta+1}\theta_i^2}{(i^{2\beta+1}+n)(i^{2\beta'+1}+n)} \ge \frac{\delta}{2}\sum_{i=2}^{T_n}\frac{ni^{2\beta+1}\theta_i^2}{i^{2\beta+1}+n} \ge \frac{\delta}{4}\sum_{i=2}^{T_n} i^{2\beta+1}\theta_i^2.$$

Consider any $\beta \ge \beta_0 + \frac{1}{2} = \bar\beta(\theta) + \frac{1}{2} + \varepsilon$. Since $\sum_{i=1}^{\infty} i^{2\bar\beta+\varepsilon/2}\theta_i^2 = \infty$, for any $K > 0$ there exist infinitely many $i \in \mathbb{N}$ (subsequence $i_k \to \infty$ as $k \to \infty$) such that $i^{2\bar\beta+1+\varepsilon}\theta_i^2 \ge K$. This infinite subsequence depends of course on the constant K. Thus

$$\sum_{i=2}^{T_n} i^{2\beta+1}\theta_i^2 = \sum_{i=2}^{T_n} i^{2\beta-2\bar\beta-\varepsilon}\theta_i^2 i^{2\bar\beta+1+\varepsilon} \ge KT_n^{2\beta-2\bar\beta-\varepsilon} \ge Kn^{(2\beta-2\bar\beta-\varepsilon)/(2\beta+1)}$$

for infinitely many n. Certainly, $n^{(2\beta-2\bar\beta-\varepsilon)/(2\beta+1)} \ge n^{1-2\beta'/(2\beta+1)}$ for any $\beta' \ge \bar\beta + \frac{1+\varepsilon}{2}$. Using this and the last two relations, we have that for any $\beta \ge \bar\beta + 1/2 + \varepsilon$ there exists $\beta' \in [\bar\beta + 1/2 + \varepsilon/2, \beta)$ such that

$$\sum_{i=1}^{\infty} a_i(\beta,\beta')\theta_i^2 \ge \frac{\delta}{4}\sum_{i=2}^{T_n} i^{2\beta+1}\theta_i^2 \ge \frac{\delta K}{4}n^{1-2\beta'/(2\beta+1)} \tag{10}$$

for any $K > 0$ and infinitely many n.

Combining estimates (9) and (10), we get that for any $K > 0$ and $\beta \geq \bar{\beta} + 1/2 + \varepsilon$ there exists $\beta' \in [\bar{\beta} + 1/2 + \varepsilon/2, \beta)$ such that

$$\sum_{i=1}^{\infty} \frac{a_i(\beta_0, \beta')(\theta_i^2 - \tau_i^2(\beta'))}{1 + n^{-1} a_i(\beta_0, \beta')} \geq \sum_{i=1}^{\infty} a_i(\beta, \beta')\Big(\frac{\theta_i^2}{2} - \tau_i^2(\beta')\Big) \geq \Big(\frac{\delta K}{8} - C(\kappa)\Big) n^{1-2\beta'/(2\beta+1)} \qquad (11)$$

for infinitely many n. Let $\varepsilon_1 = \min\{\varepsilon, 1\}$. Now, for any $\beta \geq \bar{\beta}(\theta) + 1/2 + \varepsilon$, choose

$$\beta' = \beta'(\beta) = \lfloor (\bar{\beta} + 1/2 + \varepsilon_1/2)(2\beta + 1)/(2\bar{\beta} + 2 + \varepsilon_1) \rfloor_{S_n}.$$

In this case it is easy to see that $\beta' \in [\bar{\beta} + 1/2 + \varepsilon/2, \beta)$ and $\beta - \beta' \geq \varepsilon/(2(2\bar{\beta} + 2 + \varepsilon_1))$. The last inequality implies that if we take $\delta = \delta_\varepsilon = 1 - \exp\{-(\log 2)\varepsilon/(2\bar{\beta} + 2 + \varepsilon_1)\}$, then $0 \leq \delta_\varepsilon \leq 1 - \exp\{-2(\log 2)(\beta - \beta')\}$. Note further that

$$1 - 2\beta'/(2\beta + 1) \geq 1/(2\bar{\beta} + 2 + \varepsilon_1) \geq 1/(2\beta_0 - 2\varepsilon + 2 + \varepsilon_1) \geq 1/(2\beta_0 + 2).$$

Using this relation and (11), we obtain that for any $K > 0$ and any $\beta \geq \bar{\beta} + 1/2 + \varepsilon$ there exists $\beta' \in [\bar{\beta} + 1/2 + \varepsilon/2, \beta)$ such that

$$\sum_{i=1}^{\infty} \frac{a_i(\beta_0, \beta')(\theta_i^2 - \tau_i^2(\beta'))}{1 + n^{-1} a_i(\beta_0, \beta')} \geq \Big(\frac{\delta_\varepsilon K}{8} - C(\kappa)\Big) n^{1-2\beta'/(2\beta+1)} \geq \Big(\frac{\delta_\varepsilon K}{8} - C(\kappa)\Big) n^{1/(2\beta_0+2)}$$

for infinitely many n, which implies that if we take K such that $\frac{\delta_\varepsilon K}{8} - C(\kappa) \geq C$, then $\theta \in \Lambda_{\beta_0+1/2}(\Delta_n)$, with $\Delta_n = C n^{1/(2\beta_0+2)}$, for infinitely many n. □

Remark 5. Suppose we want to test

$$H_0\colon \theta \in \Theta_{\beta_0} \quad \text{against} \quad H_1\colon \theta \notin \Theta_{\beta_0 - 1/2},$$

for $\beta_0 - 1/2 \in \mathcal{B}$. Lemma 5 and Theorem 1 imply that for any $N \in \mathbb{N}$ there exists $n \geq N$ such that the probabilities of type I and II errors are both exponentially small in n, provided $|S_n| \leq C_1 \exp\{C_2 n^{1/(2\beta_0+1)}\}$ for some $C_1, C_2 > 0$.

Apart from the smoothness hypothesis testing framework, we can apply our results to the smoothness classification problem. Suppose we have to decide which smoothness value from the set S we should assign to our unknown signal θ on the basis of the observation X. Suppose we are allowed to choose only between two known values, $S = \{\beta_1, \beta_2\}$. Assume $\beta_1 < \beta_2$. If we knew θ, a reasonable oracle classifier of the signal smoothness would be $\lfloor \bar{\beta}(\theta) \rfloor_S$, that is, if $\bar{\beta}(\theta) < \beta_2$, then the oracle smoothness classifier is β_1, otherwise β_2. Consider an empirical smoothness classifier $\lfloor \hat{\beta}(X) \rfloor_S$ and the probability of its misclasiffication error: $\gamma(\beta, \beta') = \mathrm{P}_\theta(\lfloor \hat{\beta}(X) \rfloor_S = \beta)$ while $\lfloor \bar{\beta}(\theta) \rfloor_S = \beta'$, $\beta, \beta' \in S$, $\beta \neq \beta'$.

There are three cases: (a) $\beta_2 \leq \bar{\beta}(\theta)$, then $\lfloor \bar{\beta}(\theta) \rfloor_S = \beta_2$; (b) $\beta_1 \leq \bar{\beta}(\theta) < \beta_2$, then $\lfloor \bar{\beta}(\theta) \rfloor_S = \beta_1$; (c) $\bar{\beta}(\theta) < \beta_1$, then again $\lfloor \bar{\beta}(\theta) \rfloor_S = \beta_1$. Case (a) is the easiest one, the misclassification probability $\gamma(\beta_1, \beta_2)$ is exponentially small according to Theorem 1. In case (c), by Lemma 5 and Theorem 1, we derive that for any $N \in \mathbb{N}$ there exists $n \geq N$ such that the misclassification probability $\gamma(\beta_2, \beta_1)$ is exponentially small in n if $\beta_2 > \beta_1 + 1/2$.

Consider now case (b). If $\beta_2 > \bar{\beta}(\theta) + 1/2$, we are essentially in the same situation as in case (c). If $\bar{\beta}(\theta) < \beta_2 \leq \bar{\beta}(\theta) + 1/2$, our results do not provide any bound on the misclassification probability $\gamma(\beta_2, \beta_1)$. Thus if we assume that β_1 and β_2 are apart from each other by at least $1/2$, i.e., $\beta_2 > \beta_1 + 1/2$, we can apply our results only if β_1 is sufficiently (depending on the difference $\beta_2 - \beta_1$) close to $\bar{\beta}(\theta)$ so that $\beta_2 > \bar{\beta}(\theta) + 1/2$.

This uncertainty in case (b) appears because we look at the misclassification probability for the two different values of the oracle classifier $\lfloor \bar{\beta}(\theta) \rfloor_S \in S = \{\beta_1, \beta_2\}$, and not of the "true" smoothness $\bar{\beta}(\theta)$ of θ, which can actually take any value in $\mathcal{B}$. Suppose now that we want to bound the misclassification probability only when $\bar{\beta}(\theta) \in S = \{\beta_1, \beta_2\}$. Then we will have essentially only situations (a) and (c): (a) $\beta_1 < \beta_2 = \bar{\beta}(\theta)$ and (c) $\bar{\beta}(\theta) = \beta_1 < \beta_2$. In this case we can bound the misclassification probability by applying Theorem 1 in case (a) and Lemma 5 and Theorem 1 in case (c), provided $\beta_2 > \beta_1 + 1/2$ in case (c).

4. AUXILIARY RESULTS

This section provides some lemmas which we need to prove the main results.

Lemma 6. *For any* $\beta_0, \beta \in S_n$

$$\mathrm{P}_\theta\big\{\hat\beta(S_n) = \beta\big\} \le \exp\left\{\frac{1}{2}\sum_{i=1}^\infty \frac{a_i\big(\tau_i^2(\beta_0) - \theta_i^2\big)}{1 + n^{-1}a_i}\right\}$$
$$= \exp\left\{\frac{1}{2}\sum_{i=1}^\infty \frac{(\tau_i^2(\beta) - \tau_i^2(\beta_0))(\theta_i^2 - \tau_i^2(\beta_0))}{\tau_i^2(\beta)\tau_i^2(\beta_0) + 2n^{-1}\tau_i^2(\beta_0) + n^{-2}}\right\}.$$

Proof. We use here the following shorthand notation: $a_i = a_i(\beta, \beta_0)$, $b_i = b_i(\beta, \beta_0)$.

Since $\beta_0 \in S_n$, by (8) and the Markov inequality, we have

$$\mathrm{P}_\theta\{\hat\beta = \beta\} = \mathrm{P}_\theta\big\{Z_n(\beta, \beta') \le 0 \,\forall\, \beta' \in S_n\big\} \le \mathrm{P}_\theta\big\{Z_n(\beta, \beta_0) \le 0\big\}$$
$$= \mathrm{P}_\theta\bigg\{-\sum_{i=1}^\infty a_i X_i^2 \ge \sum_{i=1}^\infty \log b_i\bigg\}$$
$$\le \mathrm{E}_\theta \exp\left\{-\frac{1}{2}\sum_{i=1}^\infty a_i X_i^2\right\}\exp\left\{-\frac{1}{2}\sum_{i=1}^\infty \log b_i\right\}.$$

To compute $\mathrm{E}_\theta \exp\{-\frac{1}{2}a_i X_i^2\}$, we use the following elementary identity for a Gaussian random variable $\eta \sim \mathcal{N}(\mu, \sigma^2)$:

$$\mathrm{E}\exp\{\lambda\eta^2\} = (1 - 2\kappa\sigma^2)^{-1/2}\exp\left\{\frac{\kappa\mu^2}{1 - 2\kappa\sigma^2}\right\} \qquad \text{for} \quad \lambda < \frac{1}{2\sigma^2}.$$

Apply this equality for $\lambda = -\frac{a_i}{2}$ and $\eta = X_i$ (condition $\lambda < \frac{1}{2\sigma^2}$ corresponds to $-a_i < n$, which is always true since $|a_i| < n$ for all $i \in \mathbb{N}$):

$$\mathrm{E}_\theta \exp\left\{-\frac{1}{2}a_i X_i^2\right\} = (1 + n^{-1}a_i)^{-1/2}\exp\left\{\frac{-a_i\theta_i^2}{2(1 + n^{-1}a_i)}\right\}.$$

Combining the previous relations, we obtain

$$\mathrm{P}_\theta\{\hat\beta = \beta\} \le \prod_{i=1}^\infty \frac{b_i^{-1/2}}{(1 + n^{-1}a_i)^{1/2}}\exp\left\{\frac{-a_i\theta_i^2}{2(1 + n^{-1}a_i)}\right\}. \tag{12}$$

From definitions (6) and (7) it follows

$$\frac{b_i^{-1}}{1 + n^{-1}a_i} = 1 + \frac{a_i\tau_i^2(\beta_0)}{1 + n^{-1}a_i}.$$

Using this, the elementary inequality $1 + x \le e^x$, $x \in \mathbb{R}$, and (12), we finally arrive at

$$\mathrm{P}_\theta\{\hat\beta = \beta\} \le \exp\left\{\frac{1}{2}\sum_{i=1}^\infty \frac{a_i\big(\tau_i^2(\beta_0) - \theta_i^2\big)}{1 + n^{-1}a_i}\right\}$$
$$= \exp\left\{\frac{1}{2}\sum_{i=1}^\infty \frac{(\tau_i^2(\beta) - \tau_i^2(\beta_0))(\theta_i^2 - \tau_i^2(\beta_0))}{\tau_i^2(\beta)\tau_i^2(\beta_0) + 2n^{-1}\tau_i^2(\beta_0) + n^{-2}}\right\}.$$

□

Lemma 7. *Let $A_\theta(\beta_0) < \infty$ for some $\beta_0 \in S_n$. Then there exists an $N = N(\beta_0, \theta)$ such that for any $n \ge N$ and any $\beta \in S_n$, $\beta < \beta_0$, the inequality*

$$\mathrm{P}_\theta\big\{\hat\beta(S_n) = \beta\big\} \le \exp\left\{\frac{I(\beta_0) n^{1/(2\beta_0+1)}}{2} + \frac{5}{8} - \frac{n^{1/(\beta+\beta_0+1)}}{16}\right\}$$

holds for all $n \ge N$.

Moreover, let $\beta_0, Q > 0$. Then there exists $C_1 = C_1(\beta_0, Q)$ such that for any $\beta, \beta_0' \in S_n$, $\beta \le \beta_0' \le \beta_0 - \frac{C_1}{\log n}$, the inequality

$$\sup_{\theta \in \Theta_{\beta_0}(Q)} \mathrm{P}_\theta\big\{\hat\beta(S_n) = \beta\big\} \le \exp\left\{\frac{B(1,2,0) n^{1/(2\beta_0'+1)}}{2} + \frac{5}{8} - \frac{n^{1/(\beta+\beta_0'+1)}}{16}\right\}$$

holds for all $n \in \mathbb{N}$. Here $I(\beta_0) = B(2\beta_0+1, 2, 0)$ and $B(1,2,0)$ are defined by (4).

Proof. We make use of Lemma 6:

$$\mathrm{P}_\theta\{\hat\beta = \beta\} \le \exp\left\{\frac{1}{2}\sum_{i=1}^\infty \frac{(\tau_i^2(\beta) - \tau_i^2(\beta_0))(\theta_i^2 - \tau_i^2(\beta_0))}{\tau_i^2(\beta)\tau_i^2(\beta_0) + 2n^{-1}\tau_i^2(\beta_0) + n^{-2}}\right\} = \exp\left\{\frac{S_1 + S_2(\theta)}{2}\right\}, \tag{13}$$

where

$$S_1 = \sum_{i=1}^\infty \frac{\tau_i^4(\beta_0) - \tau_i^2(\beta)\tau_i^2(\beta_0)}{\tau_i^2(\beta)\tau_i^2(\beta_0) + 2n^{-1}\tau_i^2(\beta_0) + n^{-2}} = S_{11} - S_{12},$$

$$S_2(\theta) = \sum_{i=1}^\infty \frac{-a_i(\beta,\beta_0)\theta_i^2}{1 + n^{-1}a_i(\beta,\beta_0)} = \sum_{i=1}^\infty \frac{\big(\tau_i^2(\beta) - \tau_i^2(\beta_0)\big)\theta_i^2}{\tau_i^2(\beta)\tau_i^2(\beta_0) + 2n^{-1}\tau_i^2(\beta_0) + n^{-2}}.$$

The rest of the proof of the first assertion is similar to the corresponding part from the proof of Lemma 3.1 in Belitser and Ghosal [2], where a purely Bayesian smoothness selector was considered. First we bound the term S_1. As $\beta < \beta_0$, we have $i^{-(2\beta+1)} > i^{-(2\beta_0+1)}$ and therefore, by Lemma 9 (see also the remark after that lemma), we obtain

$$S_{11} = \sum_{i=1}^\infty \frac{i^{-2(2\beta_0+1)}}{i^{-2(\beta+\beta_0+1)} + 2n^{-1}i^{-(2\beta_0+1)} + n^{-2}} \le \sum_{i=1}^\infty \frac{i^{-2(2\beta_0+1)}}{(i^{-(2\beta_0+1)} + n^{-1})^2}$$
$$= n^2 \sum_{i=1}^\infty \frac{1}{(n + i^{2\beta_0+1})^2} \le B(2\beta_0+1,2,0) n^{1/(2\beta_0+1)} + 1.$$

To bound S_{12} from below, note first that the term $i^{-2(\beta+\beta_0+1)}$ is not less than n^{-2} for $i \le n^{1/(\beta+\beta_0+1)}$ and not less than $n^{-1}i^{-(2\beta_0+1)}$ for $i \le n^{1/(2\beta+1)}$, which includes also all $i \le n^{1/(\beta+\beta_0+1)}$, since $\beta < \beta_0$. This implies

$$S_{12} = \sum_{i=1}^\infty \frac{i^{-2(\beta_0+\beta+1)}}{i^{-2(\beta+\beta_0+1)} + 2n^{-1}i^{-(2\beta_0+1)} + n^{-2}}$$
$$\ge \sum_{i=1}^{\lfloor n^{1/(\beta+\beta_0+1)}\rfloor} \frac{i^{-2(\beta+\beta_0+1)}}{4i^{-2(\beta+\beta_0+1)}} \ge \frac{n^{1/(\beta+\beta_0+1)} - 1}{4}.$$

Combining the last two inequalities, we arrive at

$$S_1 \le B(2\beta_0+1,2,0) n^{1/(2\beta_0+1)} - \frac{n^{1/(\beta+\beta_0+1)}}{4} + \frac{5}{4}. \tag{14}$$

Now note that $\tau_i^2(\beta) > \tau_i^2(\beta_0)$ as $\beta < \beta_0$. Then, for any $m \in \mathbb{N}$, we have

$$S_2(\theta) \le \sum_{i=1}^\infty \frac{\tau_i^2(\beta)\theta_i^2}{\tau_i^2(\beta)\tau_i^2(\beta_0) + 2n^{-1}\tau_i^2(\beta_0) + n^{-2}} = \sum_{i=1}^\infty \frac{n^2 i^{2\beta_0+1}\theta_i^2}{n^2 + 2ni^{2\beta+1} + i^{2(\beta+\beta_0+1)}}$$

$$\leq \sum_{i=1}^{m} i^{2\beta_0+1}\theta_i^2 + \sum_{i=m+1}^{\infty} \frac{n^2 i^{2\beta_0}\theta_i^2}{i^{2\beta+2\beta_0+1}} \leq m\sum_{i=1}^{m} i^{2\beta_0}\theta_i^2 + \frac{n^2}{(m+1)^{2\beta+2\beta_0+1}} \sum_{i=m+1}^{\infty} i^{2\beta_0}\theta_i^2.$$

Let $C_\varepsilon = C_\varepsilon(A_\theta(\beta_0)) = \max\{1, 2A_\theta(\beta_0)/\varepsilon\}$ for some fixed $\varepsilon > 0$. Take now

$$m = m_n = m_n(\beta_0, \beta, A_\theta(\beta_0), \varepsilon) = \lfloor C_\varepsilon^{-1} n^{1/(\beta+\beta_0+1)} \rfloor,$$

so that the first term in the right-hand side of the last inequality $m_n \sum_{i=1}^{m_n} i^{2\beta_0}\theta_i^2 \leq \varepsilon n^{1/(\beta+\beta_0+1)}/2$ for all $n \in \mathbb{N}$. Next, there exists $N = N(\beta_0, \theta, \varepsilon)$ such that for any $n \geq N$

$$\sum_{i \geq M_n} i^{2\beta_0}\theta_i^2 \leq \frac{\varepsilon}{2} C_\varepsilon^{-4\beta_0-1} \leq \frac{\varepsilon}{2} C_\varepsilon^{-2\beta-2\beta_0-1},$$

with $M_n = M_n(\beta_0, A_\theta(\beta_0), \varepsilon) = C_\varepsilon^{-1} n^{1/(2\beta_0+1)}$, which implies that the second term

$$\frac{n^2}{(m_n+1)^{2\beta+2\beta_0+1}} \sum_{i=m_n+1}^{\infty} i^{2\beta_0}\theta_i^2 \leq C_\varepsilon^{2\beta+2\beta_0+1} n^{1/(\beta+\beta_0+1)} \sum_{i \geq M_n} i^{2\beta_0}\theta_i^2 \leq \frac{\varepsilon n^{1/(\beta+\beta_0+1)}}{2}$$

for all $n \geq N$, since $m_n + 1 \geq M_n$. Therefore the relation $S_2(\theta) \leq \varepsilon n^{1/(\beta+\beta_0+1)}$ holds for all $n \geq N$. We choose $\varepsilon = 1/8$ and combine this relation with (13) and (14) to finish the proof of the first assertion of the lemma.

To establish the second assertion, we repeat the arguments as above with β_0' instead of β_0. Since $S_{11} \leq B(2\beta_0'+1, 2, 0) n^{1/(2\beta_0'+1)} + 1 \leq B(1,2,0) n^{1/(2\beta_0'+1)} + 1$, similarly to (14), we get now

$$S_1 \leq B(1,2,0) n^{1/(2\beta_0'+1)} - \frac{n^{1/(\beta+\beta_0'+1)}}{4} + \frac{5}{4}. \tag{15}$$

It remains to handle the term $S_2(\theta) = S_2(\theta, \beta, \beta_0')$ uniformly over $\theta \in \Theta_{\beta_0}(Q)$. We assume that $C_\varepsilon = \max\{1, 2Q/\varepsilon\}$ for some $\varepsilon > 0$ to be chosen later and $m_n = \lfloor C_\varepsilon^{-1} n^{1/(\beta+\beta_0'+1)} \rfloor$. As before, we derive that for all $n \in \mathbb{N}$

$$S_2(\theta) \leq \frac{\varepsilon n^{1/(\beta+\beta_0'+1)}}{2} + n^{1/(\beta+\beta_0'+1)} C_\varepsilon^{2\beta+2\beta_0'+1} \sum_{i=m_n+1}^{\infty} i^{2\beta_0'}\theta_i^2.$$

If $\beta \leq \beta_0 - \frac{K}{\log n}$ with $K = K(\beta_0, Q, \varepsilon) = (2\beta_0+1)^2 \log C_\varepsilon$, then

$$\frac{C_\varepsilon}{n^{1/(\beta+\beta_0'+1)}} \leq \frac{C_\varepsilon}{n^{1/(2\beta_0+1)} n^{1/(\beta+\beta_0'+1)-1/(2\beta_0+1)}} \leq \frac{C_\varepsilon}{n^{1/(2\beta_0+1)} n^{(\beta_0-\beta)/(2\beta_0+1)^2}} \leq \frac{1}{n^{1/(2\beta_0+1)}}.$$

Therefore, as $\theta \in \Theta_{\beta_0}(Q)$, $C_\varepsilon \geq 1$, and $\beta < \beta_0' \leq \beta_0 - \frac{C_1}{\log n}$, we evaluate for all $n \in \mathbb{N}$

$$C_\varepsilon^{2\beta+2\beta_0'+1} \sum_{i=m_n+1}^{\infty} i^{2\beta_0'}\theta_i^2 \leq \frac{C_\varepsilon^{4\beta_0+1} \sum_{i=m_n+1}^{\infty} i^{2\beta_0}\theta_i^2}{(m_n+1)^{2(\beta_0-\beta_0')}} \leq \frac{C_\varepsilon^{4\beta_0+1} Q}{(C_\varepsilon^{-1} n^{1/(\beta+\beta_0'+1)})^{2C_1/\log n}}$$

$$\leq \frac{C_\varepsilon^{4\beta_0+1} Q}{(n^{1/(2\beta_0+1)})^{2C_1/\log n}} = \frac{C_\varepsilon^{4\beta_0+1} Q}{e^{2C_1/(2\beta_0+1)}} \leq \frac{\varepsilon}{2}$$

if $C_1 = C_1(\beta_0, Q, \varepsilon) = \max\{\log\big(2QC_\varepsilon^{4\beta_0+1}\varepsilon^{-1}\big)^{(2\beta_0+1)/2}, K\}$. Therefore, if $\beta \leq \beta_0' \leq \beta_0 - \frac{C_1}{\log n}$, the relation $S_2(\theta) \leq \varepsilon n^{1/(\beta+\beta_0'+1)}$ holds uniformly over $\theta \in \Theta_{\beta_0}(Q)$ and all $n \in \mathbb{N}$. Take $\varepsilon = 1/8$ and combine the last uniform bound for $S_2(\theta)$ with (13) and (15) to finish the proof of the second assertion of the lemma. □

Remark 6. An interesting question is whether there exists a sequence $\beta_n = \beta_n(\theta)$ such that $\beta_n \uparrow \bar{\beta}(\theta)$ (eventually slowly enough) and $\mathrm{P}_\theta\{\hat{\beta}(S_n) = \beta_n\} \to 0$ as $n \to \infty$.

Analyzing the exponential upper bound for $\mathrm{P}_\theta\{\hat{\beta}(S_n) = \beta_n\}$ in the above lemma, we deduce that β_n can not approach $\bar{\beta}(\theta)$ faster than at the logarithmic rate if we want this bound to converge to zero.

Indeed, a $\beta_{0,n}$ has to be chosen in this upper bound so that $\beta_n < \beta_{0,n} < \bar\beta(\theta)$, since it also has to satisfy $A_\theta(\beta_{0,n}) < \infty$. It is not so difficult to see (by the same reasoning as in the proof of the first assertion of Theorem 1) that this upper bound becomes small (of order $\exp\{-Cn^{1/(2\beta_{0,n}+1)}\} \le \exp\{-Cn^{1/(2\bar\beta+1)}\}$) only if $\beta_{0,n}$ and β_n are sufficiently distant from each other, namely, $\beta_{0,n} \ge \beta_n + c/\log n$ for some sufficiently large constant c. However, as follows from the proof, the larger $A_\theta(\beta_0)$, i.e., the closer β_0 to $\bar\beta(\theta)$, the larger the corresponding $N = N(\beta_0, \theta)$ (the one for which the bound in the lemma holds for all $n \ge N$). Therefore, even if $\beta_n \uparrow \bar\beta(\theta)$ very slowly, we still cannot conclude in general that $\mathrm{P}_\theta\{\hat\beta(S_n) = \beta_n\} \to 0$ as $n \to \infty$.

Remark 7. The first assertion of the above lemma is not claimed to be uniform with respect to θ since the inequality holds only for $n \ge N(\beta_0, \theta)$. However, if $A_{\bar\theta}(\beta_0) < \infty$, then for a sufficiently small ellipsoid size Q, the uniformity does hold. Indeed, we only need to evaluate the term $S_2(\theta)$ uniformly over $\theta \in \Theta_{\beta_0}(\bar\theta, Q)$. Now, for any $\theta \in \Theta_{\beta_0}(\bar\theta, Q)$ we have $S_2(\theta) \le 2S_2(\bar\theta) + 2S_2(\theta - \bar\theta)$. As in the proof of Lemma 7, we can find $N_1 = N_1(\beta_0, \bar\theta, \varepsilon)$ such that $S_2(\bar\theta) \le n^{1/(\beta+\beta_0+1)}\varepsilon/4$ for all $n \ge N_1$. Next, by taking $m = m_n = \lfloor (n^{1/(\beta+\beta_0+1)} \rfloor$, we derive that for any $Q < \varepsilon/4$ there exists $N_2 = N_2(\beta_0, Q)$ such that

$$S_2(\theta - \bar\theta) \le A_{\theta-\bar\theta}(\beta_0) n^{1/(\beta+\beta_0+1)} \le Qn^{1/(\beta+\beta_0+1)} \le n^{1/(\beta+\beta_0+1)}\varepsilon/4$$

for all $n \ge N_2$ for any $\theta \in \Theta_{\beta_0}(\bar\theta, Q)$. We conclude that for any $Q < \varepsilon/4$ there exists

$$N_3 = N_3(\beta_0, \bar\theta, Q, \varepsilon) = \max\{N_1(\beta_0, \bar\theta, \varepsilon), N_2(\beta_0, Q)\}$$

such that $S_2(\theta) \le \varepsilon n^{1/(\beta+\beta_0+1)}$ for all $n \ge N_3$, uniformly over $\theta \in \Theta_{\beta_0}(\bar\theta, Q)$. Take $\varepsilon = 1/8$ to derive the assertion of the lemma uniformly over $\theta \in \Theta_{\beta_0}(\bar\theta, Q)$ for any $Q < 1/32$.

Lemma 8. *For any $\beta', \beta \in \mathbb{R}$ such that $\kappa \le \beta' \le \beta$ the following inequality holds: for some $C = C(\kappa)$*

$$\sum_{i=1}^\infty \frac{\tau_i^2(\beta')}{\tau_i^2(\beta) + n^{-1}} \le C(\kappa) n^{1-2\beta'/(2\beta+1)}.$$

Proof. Since $\kappa \le \beta' \le \beta$,

$$\begin{aligned}
\sum_{i=1}^\infty \frac{\tau_i^2(\beta')}{\tau_i^2(\beta) + n^{-1}} &= \sum_{i=1}^\infty \frac{ni^{2\beta-2\beta'}}{n + i^{2\beta+1}} \\
&\le \sum_{i=1}^{\lfloor n^{1/(2\beta+1)} \rfloor - 1} i^{2\beta-2\beta'} + 2n^{(2\beta-2\beta')/(2\beta+1)} + \sum_{i=\lfloor n^{1/(2\beta+2)} \rfloor + 2}^\infty n i^{-(2\beta'+1)} \\
&\le \int_0^{\lfloor n^{1/(2\beta+1)} \rfloor} x^{2\beta-2\beta'}\, dx + 2n^{1-2\beta'/(2\beta+1)} + \int_{\lfloor n^{1/(2\beta+1)} \rfloor + 1}^\infty n x^{-(2\beta'+1)}\, dx \\
&\le \left(\frac{1}{2\beta - 2\beta' + 1} + \frac{1}{2\beta'} + 2 \right) n^{1-2\beta'/(2\beta+1)} \le C(\kappa) n^{1-2\beta'/(2\beta+1)},
\end{aligned}$$

with $C(\kappa) = 3 + (2\kappa)^{-1}$. □

Remark 8. By using Lemmas 9 and 10, one can improve the constant in the above upper bound for sufficiently large n.

Finally we prove two technical lemmas used in the proofs of other results. Let b_+ denote the nonnegative part of b. A version of the following auxiliary result is contained in [16]. As compared to Lemma 2 in [16], our lemma below provides also bounds for the second order terms suitable for our purposes.

Lemma 9. *Suppose* $0 < p < \infty$, $0 < q < \infty$, $0 \le r < \infty$. *Let* $\gamma_n \to \infty$ *as* $n \to \infty$. *If* $pq > r+1$, *then*

$$\sum_{i=1}^{\infty} \frac{i^r}{(\gamma_n + i^p)^q} = B(p,q,r)\gamma_n^{(1+r)/p-q} + \phi_n$$

and if $pq > r$, *then*

$$\max_{i \in \mathbb{N}} \frac{i^r}{(\gamma_n + i^p)^q} = D(p,q,r)\gamma_n^{(r/p)-q} + \psi_n,$$

where $\phi_n = \phi_n(p,q,r)$ *and* $\psi_n = \psi_n(p,q,r)$ *are such that*

$$|\phi_n| \le D(p,q,r)\gamma_n^{-q+\frac{r}{p}}, \qquad |\psi_n| \le C(p,q,r)\gamma_n^{-q+\frac{(r-1)_+}{p}}$$

for some constant $C(p,q,r) > 0$,

$$B(p,q,r) = \int_0^\infty \frac{u^r du}{(1+u^p)^q}$$

is defined by (4), *and*

$$D(p,q,r) = r^{r/p}(pq-r)^{q-(r/p)}(pq)^{-q} = \left(1 - \frac{r}{pq}\right)^q \left(\frac{pq}{r} - 1\right)^{-r/p},$$

with the convention $0^0 = 1$.

Remark 9. Notice that if $r \le 1$ or $pq \ge 2r$ then $0 \le D(p,q,r) \le 1$.

Proof. Denote $g(u) = \frac{u^r}{(\gamma_n + u^p)^q}$, $u > 0$. The function $g(u)$ is increasing on $u \in [0, u_{\max}]$ and decreasing on $[u_{\max}, \infty)$ with $u_{\max} = \big(r\gamma_n/(pq-r)\big)^{1/p}$. Therefore,

$$\int_0^\infty \frac{u^r}{(1+u^p)^q}\, du - g(u_{\max}) \le \sum_{i=1}^{\infty} \frac{i^r}{(\gamma_n + i^p)^q} \le \int_0^\infty \frac{u^r}{(1+u^p)^q}\, du + g(u_{\max})$$

with $g(u_{\max}) = D(p,q,r)\gamma_n^{(r/p)-q}$, which establishes the first relation. To prove the second relation, we first compute

$$g'(u) = \frac{r\gamma_n u^{r-1} - (pq-r)u^{p+r-1}}{(\gamma_n + u^p)^{q+1}}$$

and then evaluate

$$\max_{u \ge 1} |g'(u)| \le \max\left\{ \max_{u \ge 1}\left\{ \frac{r\gamma_n u^{r-1}}{(\gamma_n + u^p)^{q+1}} \right\}, \max_{u \ge 1}\left\{ \frac{(pq-r)u^{p+r-1}}{(\gamma_n + u^p)^{q+1}} \right\} \right\} \le C(p,q,r)\gamma_n^{-q+\frac{(r-1)_+}{p}}$$

for some constant $C(p,q,r) > 0$. Finally, using this bound and unimodality of the function $g(u)$ on $[0, \infty)$, we obtain

$$\left| g(u_{\max}) - \max_{i \in \mathbb{N}} \frac{i^r}{(\gamma_n + i^p)^q} \right| \le \max_{u \ge 1} |g'(u)| \le C(p,q,r)\gamma_n^{-q+\frac{(r-1)_+}{p}},$$

which completes the proof of the lemma. □

The following short lemma follows directly from the properties of Beta and Gamma functions.

Lemma 10. *Let* $r \ge 0$, $p > r+1$. *Then*

$$B(p,1,r) = \frac{\pi}{p \sin\big(\pi(r+1)/p\big)}, \qquad B(p,2,r) = \frac{\pi(p-r-1)}{p^2 \sin\big(\pi(r+1)/p\big)}.$$

Proof. From the definition of the function B we get

$$B(p,q,r) = p^{-1}\,\mathrm{Beta}\left(q - \frac{r+1}{p}, \frac{r+1}{p}\right) = \frac{\Gamma\big(q - \frac{r+1}{p}\big)\Gamma\big(\frac{r+1}{p}\big)}{p\Gamma(q)},$$

where $\Gamma(\cdot)$ is the Gamma function. The lemma follows by the following properties of the Gamma function: $\Gamma(1) = \Gamma(2) = 1$, $\Gamma(z)\Gamma(1-z) = \pi/\sin(\pi z)$ and $\Gamma(1+z) = z\Gamma(z)$. □

REFERENCES

1. A. Barron, L. Birgé, and P. Massart, "Risk Bounds for Model Selection", Probab. Theory and Related Fields **113**, 301–413 (1999).
2. E. Belitser and S. Ghosal, "Adaptive Bayesian Inference on the Mean of an Infinite-Dimensional Normal Distribution", Ann. Statist. **31**, 536–559 (2003).
3. E. Belitser and B. Levit, "On Minimax Filtering over Ellipsoids", Math. Methods Statist. **3**, 259–273 (1995).
4. E. Belitser and B. Levit, "On the Empirical Bayes Approach to Adaptive Filtering", Math. Methods Statist. **12**, 131–154 (2003).
5. L. Birgé and P. Massart, "Gaussian Model Selection", J. Eur. Math. Soc. **3**, 203–268 (2001).
6. L. D. Brown and M. G. Low, "Asymptotic Equivalence of Nonparametric Regression and White Noise", Ann. Statist. **24**, 2384–2398 (1995).
7. L. Cavalier, G. K. Golubev, D. Picard, and A. B. Tsybakov, "Oracle Inequalities for Inverse Problems", Ann. Statist. **30**, 843–874 (2002).
8. L. Cavalier and A. B. Tsybakov, "Penalized Blockwise Stein'S Method, Monotone Oracles and Sharp Adaptive Estimation", Math. Methods Statist. **10**, 247–282 (2001).
9. P. Diaconis and D. Freedman, "On Inconsistent Bayes Estimates in the Discrete Case", Ann. Statist. **11**, 1109–1118 (1983).
10. P. Diaconis and D. Freedman, "On the Consistency of Bayes Estimates (with Discussion)", Ann. Statist. **14**, 1–67 (1986).
11. P. Diaconis and D. Freedman, "On Inconsistent Bayes Estimates of Location", Ann. Statist. **14**, 68–87 (1986).
12. P. Diaconis and D. Freedman, "Consistency of Bayes Estimates for Nonparametric Regression: Normal Theory", Bernoulli **4**, 411–444 (1998).
13. D. Donoho and I. Johnstone, "Ideal spatial adaptation by wavelet shrinkage", Biometrika **81**, 425–455 (1994).
14. D. Donoho, R. Liu, and B. MacGibbon, "Minimax Risk over Hyperrectangles, and Implications", Ann. Statist. **18**, 1416–1437 (1990).
15. S. Efromovich and M. Pinsker, "Learning algorithm for nonparametric filtering", Automat. Remote Control **11**, 1434–1440 (1984).
16. D. Freedman, "On the Bernstein–von Mises Theorem with Infinite-Dimensional Parameters", Ann. Statist. **27**, 1119–1140 (1999).
17. S. Ghosal, J. Lember, and A. van der Vaart, "On Bayesian Adaptation", Acta Appl. Math. **79**, 165–175 (2003).
18. G. Golubev and B. Levit, "Asymptotically Efficient Estimation for Analytic Distributions", Math. Methods Statist. **5**, 357–368 (1996).
19. I. A. Ibragimov and R. Z. Khasminski, *Statistical Estimation: Asymptotic Theory* (Springer, New York, 1981).
20. I. A. Ibragimov and R. Z. Khasminski, "On Nonparametric Estimation of the Value of a Linear Functional in Gaussian White Noise", Theory Probab. Appl. **29**, 18–32 (1984).
21. Y. Ingster and I. Suslina, *Nonparametric Goodness-Of-Fit Testing under Gaussian Models* (Springer, New York, 2003).
22. I. Johnstone, *Function Estimation in Gaussian Noise: Sequence Models* (1999), http://www-stat.stanford.edu/∼imj/ (monograph draft).
23. B. Laurent and P. Massart, "Adaptive Estimation of a Quadratic Functional by Model Selection", Ann. Statist. **30**, 325–396 (2000).
24. O. Lepski and M. Hoffmann, "Random Rates in Anisotropic Regression", Ann. Statist. **28**, 1302–1338 (2002).
25. O. V. Lepski, "One Problem of Adaptive Estimation in Gaussian White Noise", Theory Probab. Appl. **35**, 459–470 (1990).
26. O. V. Lepski, "Asymptotic Minimax Adaptive Estimation. 1. Upper Bounds", Theory Probab. Appl. **36**, 645–659 (1991).

27. O. V. Lepski, "Asymptotic Minimax Adaptive Estimation. Statistical Model without Optimal Adaptation. Adaptive Estimators", Theory Probab. Appl. **37**, 468–481 (1992).
28. O. Lepski and V. Spokoiny, "Optimal Pointwise Adaptive Methods in Nonparametric Estimation", Ann. Statist. **25**, 2512–2546 (1997).
29. M. Nussbaum, "Asymptotic Equivalence of Density Estimation and Gaussian White Noise", Ann. Statist. **24**, 2399–2430 (1996).
30. D. Picard and K. Tribouley, "Adaptive Confidence Interval for Pointwise Curve Estimation", Ann. Statist. **28**, 298–335 (2000).
31. M. S. Pinsker, "Optimal Filtering of Square-Integrable Signals in Gaussian Noise", Problems Inform. Transmission **16**, 120–133 (1980).
32. H. Robbins, "An Empirical Bayes Approach to Statistics", in *Proc. 3rd Berkeley Symp. on Math. Statist. and Prob. 1, Berkeley* (Univ. of California Press, Berkeley, 1955), pp. 157–164.
33. A. Tsybakov, *Introduction à l'estimation non-paramétrique* (Springer, Berlin, 2004).