Probab. Theory Relat. Fields 133, 531–558 (2005)
Digital Object Identifier (DOI) 10.1007/s00440-005-0445-4

Ghislaine Gayraud · Christophe Pouet

Adaptive minimax testing in the discrete regression scheme

Received: 19 August 2003 / Revised version: 24 February 2005 /
Published online: 6 June 2005 –

Abstract. We consider the problem of testing hypotheses on the regression function from n observations on the regular grid on [0, 1]. We wish to test the null hypothesis that the regression function belongs to a given functional class (parametric or even nonparametric) against a composite nonparametric alternative. The functions under the alternative are separated in the L_2-norm from any function in the null hypothesis. We assume that the regression function belongs to a wide range of Hölder classes but as the smoothness parameter of the regression function is unknown, an adaptive approach is considered. It leads to an optimal and unavoidable loss of order $\sqrt{\log(\log n)}$ in the minimax rate of testing compared with the non-adaptive setting. We propose a smoothness-free test that achieves the optimal rate, and finally we prove the lower bound showing that no test can be consistent if in the distance between the functions under the null hypothesis and those in the alternative, the loss is of order smaller than the optimal loss.

1. Introduction

Discrete regression models are frequently used in statistics and econometrics. If the model is misspecified, it can lead to very inaccurate results. Therefore, the problem of hypothesis testing about regression function f is essential since it allows us to check the specification of the model. Nonparametric methods seem to be natural if you do not have in mind specific departures from the model. There have been a number of papers, initiated by Ingster (1982), concerning nonparametric hypothesis testing on the signal function in the regression model or in the Gaussian white noise model, via an asymptotical minimax approach : Ermakov (1990), Ingster (1990, 1993), Härdle and Mammen (1993), Lepskii (1993), Suslina (1993), Hart (1997), Härdle and Kneip (1999), Lepskii and Spokoiny (1999), Lepskii and Tsybakov (2000), Gayraud and Pouet (2001). In all of these papers, the authors specify a functional class to which f belongs; the considered functional classes

G. Gayraud: CREST, Timbre J340, 3 av. P. Larousse, 92241 Malakoff Cedex, and Laboratoire de Mathématiques Raphael Salem, Université de Rouen, France.
e-mail: gayraud@ensae.fr

C. Pouet: Centre de Mathématiques et Informatique , Université de Provence, 39, rue Joliot Curie, 13453 Marseille Cedex 13, France. e-mail: pouet@cmi.univ-mrs.fr

Mathematics Subject Classification (2000): 62G10, 62G08; Secondary 62C20

Key words or phrases: Adaptive testing – Composite null hypothesis – Nonparametric hypotheses testing – Minimax rate of testing – Regression model

are characterized by a smoothness parameter. Then, the goal of these papers is to determine the minimal (optimal) distance between the null hypothesis and the set of alternatives for which testing with prescribed error probabilities is still possible. We refer the reader to the paper of Lepskii and Tsybakov (2000) for an accurate definition of the minimax setup testing, including also the definition of the exact separation constants. However, the test procedures depend heavily on the smoothness assumption which is typically unknown as well as the regression function f : this makes them unnatural and unattractive from a practical point of view. It is our first motivation to extend the non-adaptive case of our previous paper (Gayraud and Pouet 2001) to the adaptive case i.e. the case where the smoothness parameter is also supposed unknown. Other papers must be mentioned here in the context of adaptive hypothesis testing about the regression function : in Baraud, Huet and Laurent (2003), no assumption on f is required since the distance they used to separate the null hypothesis and the alternative is a discrete one; it avoids to quantify the approximation of the L_2-norm, when it is used as the distance as in our study, by a discrete sum of squared terms. The same occurs in Horowitz and Spokoiny (2001) and in Fromont and Lévy-Leduc (2003). Horowitz and Spokoiny (2001) give results for a composite null hypothesis and for Hölder spaces. Here, we consider a functional class of Hölderian functions with a smoothness parameter ranging over a wide scale. But we are more general than Horowitz and Spokoiny (2001) in the choice of the class under the null hypothesis since they take a parametric family of given functions whereas in our study only a control of the entropy of the class which could even be nonparametric is required. Fromont and Lévy-Leduc (2003) consider the problem of periodic signal detection in a Gaussian fixed design regression framework, assuming that the signal belongs to some periodic Sobolev balls. Fan, Zhang and Zhang (2001) give adaptive results when the alternatives lie in a range of Sobolev ball; their test statistic is based on generalized likelihood ratio and they obtain the asymptotic distribution of their test statistic under the null hypothesis, which is free of any nuisance parameter (this result is referred as Wilks phenomenon). Fan and Zhang (2004) extend Fan, Zhang and Zhang (2001) paper in the sense that they consider unspecified distribution of the errors instead of assuming a parametric structure for their distribution. The closest paper to our approach is Spokoiny (1996) and also its extension Spokoiny (1998) (the nontrivial extension of this second paper is to consider an arbitrary L_r–norm instead of the L_2–norm), although its model is the Gaussian white noise and its null hypothesis is defined by the unique null function. In our study, we do not specify the law of the errors except for the resolution of the lower bound and we get the upper bound under an assumption on the control of the tail distribution of the errors. Moreover Spokoiny (1996, 1998) consider a collection of Besov balls; it allows to decompose the regression function f on an orthonormal basis of wavelets which gives interesting properties of independence for the resolution of the lower bound.

To summarize, this paper is devoted to obtain the optimal minimax adaptive rates in testing $\mathcal{F}_0$, a given functional class of regression functions, against a functional class composed of Hölder balls. Our study gives the possibility to consider a rich class $\mathcal{F}_0$ since only a control of its entropy is required : parametric but also nonparametric functional classes are allowed. On the contrary of other papers (Horowitz and Spokoiny (2001), Guerre and Lavergne (2002)) in which a parametric family under the null hypothesis is considered, no existence of an estimate under the null hypothesis is needed. Concerning the upper bound, we do not specify the law of the errors; just a control of the tail distribution is required. It is worth mentioning an interesting link between the distribution tail of the errors and the entropy of the class $\mathcal{F}_0$. Besides, we prove that the adaptive case leads to a loss of efficiency of order $\sqrt{\log(\log n)}$.

The paper is organized as follows. In Section 2, we state the testing problem and the minimax approach to solve it. In Section 3, we explain the proposed adaptive test procedure. Section 4 is devoted to the statement of the results. In Section 5, some comments are given, in addition with some figures which illustrate the link between the distribution tail of the error and the control of the entropy of $\mathcal{F}_0$. The proofs are postponed in Section 6 and those of the lemmas are given in Appendix.

2. Model and adaptive minimax framework

We consider the usual regression model

$$Y_i = f(x_i) + \xi_i, \ i = 1, \dots, n,$$

where f defined on [0, 1] belongs to the class $\Sigma(\beta, C, M)$, $\beta > 1/4$, $C > 0$, $M > 0$, which is defined as follows :

$$\Sigma(\beta, C, M) = \begin{cases} \{f \in H(\beta, C) \ : \ \|f\|_\infty \le M\} & \text{if } \frac{1}{4} < \beta \le 1, \\ \{f \in H(\beta, C) \ : \ \|f\|_\infty \le M, \ \|f'\|_\infty \le M\} & \text{if } \beta > 1, \end{cases}$$

where $H(\beta, C)$ denotes the Hölderian class of functions with β as the smoothness parameter and C as the Hölder constant and $\|\cdot\|_\infty$ denotes the supremum norm. Besides, the real random variables ξ_i, $i \in \{1, \dots, n\}$ are i.i.d. with zero mean and unknown variance $\sigma^2 > 0$, and the points x_i are deterministic equispaced on [0, 1] and numbered in such a way that $|x_i - x_{i-1}| = \dfrac{1}{n}$, $\forall i = 2, \dots, n$.
Given the sample $Y_1, \dots, Y_n$, if we suppose β known and for a given functional class $\mathcal{F}_0$, one could consider the following test problem

$$H_0 : f \in \mathcal{F}_0 \subset \Sigma(\beta, C, M),$$

against

$$H_1 : f \in \Lambda(v(n, \beta)) = \left\{ f \in \Sigma(\beta, C, M) : \inf_{f_0 \in \mathcal{F}_0} \|f - f_0\|_2 \ge v(n, \beta) \right\},$$

where $\|\cdot\|_2$ denotes the L_2-norm and $v(n, \beta)$ is a sequence of positive numbers decreasing to zero as n goes to infinity. Note that $\Lambda(v(n, \beta))$ is defined by three parameters : the class $\Sigma(\beta, C, M)$, the L_2–norm and $v(n, \beta)$. It can be shown (Ingster 1993) that given $\Sigma(\beta, C, M)$ and the L_2-norm, $v(n, \beta)$ cannot be chosen in an arbitrary way. It turns out that, if $v(n, \beta)$ is too small, then it is not possible to test the hypothesis H_0 against H_1 with a given summarized errors of the first and the second type. On the other hand, if $v(n, \beta)$ is very large, such a testing is possible;

assuming β known, the problem is to find the smallest sequence $v(n,\beta)$ for which such a test is still possible and to indicate the corresponding test functions. More precisely, let Δ_n be a *test statistic* that is an arbitrary function with values 0, 1 which is measurable with respect to $Y_1,\dots,Y_n$ and such that we accept H_0 if $\Delta_n = 0$ and we reject H_0 if $\Delta_n = 1$. This smallest sequence $v(n,\beta)$ is said to be the minimax rate of testing if it satisfies relations (1), (2) in Definition 1.

Definition 1. *(1) For any given $\tau_1 \in (0,1)$, there exists a positive constant a such that as n large enough*

$$\inf_{\tilde{\Delta}_n}\{\sup_{f_0\in\mathcal{F}_0}\mathbf{P}_{f_0}(\tilde{\Delta}_n = 1) + \sup_{f\in\Lambda(av(n,\beta))}\mathbf{P}_f(\tilde{\Delta}_n = 0)\} \geq \tau_1 + o_n(1), \tag{1}$$

where $\inf_{\tilde{\Delta}_n}$ is the infimum over all possible test functions; relation (1) is called relation of the lower bound.

(2) For any given τ_2 in $(0,1)$, there exist a positive constant A and test functions Δ_n such that as n large enough

$$\sup_{f_0\in\mathcal{F}_0}\mathbf{P}_{f_0}(\Delta_n = 1) + \sup_{f\in\Lambda(Av(n,\beta))}\mathbf{P}_f(\Delta_n = 0) \leq \tau_2 + o_n(1), \tag{2}$$

Relation (2) is called relation of the upper bound.

In the case of β known, the minimax rate of testing is $v(n,\beta) = n^{-2\beta/(4\beta+1)}$ and the test function which achieves the minimax rate of testing is a statistic which depends obviously on β (Gayraud and Pouet, 2001). But since the regression function itself is unknown the a priori knowledge of its smoothness β could appear unrealistic.

The purpose of this paper is to solve the previous problem of testing in an adaptive framework i.e. supposing that β is unknown. Following the terminology of Spokoiny (1996), we assume that β belongs to a set $\mathcal{T} = \{\beta : 1/4 < \beta_\star \leq \beta \leq \beta^\star\}$, where $\beta_\star$ and $\beta^\star$ are fixed. We suppose that the given functional class $\mathcal{F}_0$ under H_0 is included in the greatest regular class $\Sigma(\beta^\star, C, M)$. We thus give the smallest sequence which separates the null hypothesis and the alternatives when the alternatives contain the whole scale of classes $\Sigma(\beta, C, M)$, $\beta \in \mathcal{T}$ so that we extend the result of Gayraud and Pouet (2001).

More precisely, we search for universal test functions Δ_n (free of β) such that for n large enough and for any given α_0 in $(0,1)$, it exists a positive constant $A > 0$ such that

$$\sup_{f_0\in\mathcal{F}_0}\mathbf{P}_{f_0}(\Delta_n = 1) + \sup_{\beta\in\mathcal{T}}\sup_{f\in\Lambda(Av(nt_n^{-1},\beta))}\mathbf{P}_f(\Delta_n = 0) \leq \alpha_0 + o_n(1) \tag{3}$$

where t_n is a sequence of positive numbers increasing to infinity with n as slow as possible.

One could ask if there is a possibility to get (3) with alternatives defined in a closer neighborhood of H_0 for example with an order of $o(v(nt_n^{-1},\beta))$ i.e. if one could distinguish the null hypothesis and the alternative when the alternative is much closer to the null hypothesis. The answer is negative and it is done by proving the relation of the lower bound i.e. for n large enough and for any $\alpha_1 \in (0,1)$, there exists a positive constant a such that

$$\inf_{\tilde{\Delta}_n}\{\sup_{f_0\in\mathcal{F}_0} \mathbb{P}_{f_0}(\tilde{\Delta}_n = 1) + \sup_{\beta\in\mathcal{T}} \sup_{f\in\Lambda(av(nt_n^{-1},\beta))} \mathbb{P}_f(\tilde{\Delta}_n = 0)\} \geq \alpha_1 + o_n(1), \quad (4)$$

where $\inf_{\tilde{\Delta}_n}$ denotes the infimum under all possible test statistics.

Actually, relations (3) and (4) mean that the non–adaptive rate $v(n,\beta)$ is contaminated by the term t_n in the adaptive setting and it could not be avoided.

Definition 2. *The sequence $v(nt_n^{-1},\beta)$ satisfying relations (3) and (4) is said to be the adaptive minimax rate of testing. And the test functions Δ_n for which relation (3) holds is said to be the adaptive minimax test functions.*

3. Test Procedure

In the case of unknown β in $\mathcal{T}$, and for some f_0 in $\mathcal{F}_0$, we consider

$$T_{n,\beta,f_0} = \frac{1}{\sqrt{m_\beta}} \sum_{k=1}^{m_\beta} \frac{m_\beta}{n\hat{\sigma}_n^2} \sum_{\substack{i\in I_k,\ j\in I_k\\ i\neq j}} (Y_i - f_0(x_i))(Y_j - f_0(x_j)), \quad (5)$$

where $m_\beta = (t_n^{-1}n)^{2/(4\beta+1)}$ with $t_n = \sqrt{\log(\log n)}$, $\hat{\sigma}_n^2 = \frac{1}{2(n-1)}\sum_{i=2}^n (Y_i - Y_{i-1})^2$ is an estimate of σ^2 and $I_k = \{i : x_i \in A_k\}$, with $A_k = [\frac{k-1}{m_\beta}, \frac{k}{m_\beta}[,\ \forall k \in \{1,\ldots,m_\beta - 1\}$ and $A_{m_\beta} = [\frac{m_\beta - 1}{m_\beta}, 1]$. The quantity m_β is supposed to be integer, if it is not the case, m_β will be chosen equal to the integer part of $(t_n^{-1}n)^{2/(4\beta+1)}$; therefore one could prove that if any β_1 and β_2 in $\mathcal{T}$ are much closer in distance than the order $\frac{1}{\log n}$, both of them provide the same m_{β_1}. Consequently, the range of adaptation $\mathcal{T}$ can be translated into a range $\mathcal{M}$ of the form $\mathcal{M} = \{m_\beta : m_{\beta^\star} \leq m_\beta \leq m_{\beta_\star}\}$ in which two consecutive m_{β_j} and $m_{\beta_{j+1}}$ correspond to a subset $\mathcal{T}^\star \subset \mathcal{T}$ of $\beta \in \mathcal{T}$ which are at least distant from $c\frac{1}{\log n}$ where c is an adapted positive constant which makes $m_{\beta_{j+1}}$ an integer and it depends on $\beta_\star$, $\beta^\star$ and n. Since $\beta_\star$ and $\beta^\star$ are fixed, the cardinality of the set $\mathcal{M}$ is of order $\log n$ as n large enough. Now we are able to define the adaptive test :

$$\Delta_n = \mathbb{I}_{\sup_{m_\beta\in\mathcal{M}} \inf_{f_0\in\mathcal{F}_0} T_{n,\beta,f_0} \geq \rho_n}, \quad (6)$$

where $\rho_n = \sqrt{8\log(\log n)}$ and T_{n,β,f_0} is given by relation (5).

The heuristic idea of this test procedure is the following : if β known, the test statistic $\hat{\sigma}_n^2 \frac{\sqrt{\tilde{m}_\beta}}{n} T_{n,\beta,f_0}$ estimates the squared L_2-norm $\|f - f_0\|_2^2$. Under H_1 this quantity must be greater than $\tilde{m}_\beta^{-2\beta}$ (the bias error), and than $\frac{\sqrt{\tilde{m}_\beta}}{n}$ and $\frac{\|f-f_0\|_2}{\sqrt{n}}$ (the stochastic errors), in order to detect the signals in the alternative. This leads to an optimal choice of $\tilde{m}_\beta = n^{\frac{2}{4\beta+1}}$ which appears in the non-adaptive setting. If β unknown, since the range of adaptation in β is of order $\log n$ and since the tail of the stochastic term has a gaussian behavior, the minimal loss of optimality due to adaptation is an extra $\log(\log n)$ term. This leads to an optimal choice of the bandwidth $m_\beta = \left(t_n^{-1}\, n\right)^{\frac{2}{4\beta+1}}$, where $t_n = \sqrt{\log(\log n)}$.

It is to be noted that the optimal test procedure cannot be the plug-in of an efficient estimator for $\|f - f_0\|_2^2$. Indeed Fan (1991) gives an efficient estimator of the squared L_2-norm which entails a rate of testing of order $n^{-\frac{1}{4}}$. This rate is worse than the minimax rate of testing. As it is mentioned in Lepski, Nemirovski and Spokoiny (1999), estimating the L_2-norm is not equivalent to estimating the squared L_2-norm. Thus our test procedure relies on an efficient estimator of the L_2-norm.

4. Assumptions and Results

Assume the following :

(A1.Sup) The random variables $\xi_1, \ldots, \xi_n$ are i.i.d., with zero mean, an unknown finite variance $\sigma^2 > 0$ and an unknown finite fourth order moment.

(A1.Inf) The random variables $\xi_1, \ldots, \xi_n$ are i.i.d. centered Gaussian with variance equal to $\sigma^2 > 0$.

(A2.Sup) $\lim_{y\to+\infty} \mathbb{P}\left(|\xi_i| \geq y\right) y^{\frac{1}{\kappa}(2+\frac{1}{2\beta_\star})} = 0,\ \forall i\ \in \{1, \ldots, n\}$, where $\kappa < 1$ is a positive constant.

(A3) The δ–entropy of the class $\mathcal{F}_0$ calculated in the supremum norm is bounded by δ^{-r}, where r is less than $\dfrac{2\beta_\star(1-\kappa)(1+4\beta^\star)}{(16+8\kappa)\beta_\star\beta^\star + 2(2\beta^\star + \kappa\beta_\star)}$.

Remark 1. Assumption (A2.Sup) deals with the distribution tail of the ξ_i's. One must note that there exists a relation between the weight of the distribution tail (A2.Sup) and the entropy of the class $\mathcal{F}_0$. This point will be developed in Section 5.

Theorem 1 deals with the upper bound. It states that our test can achieve asymptotically any given first-type and second-type errors under some conditions.

Theorem 1. (Behavior of Δ_n under the null hypothesis and under the alternative) *Let Δ_n be the test functions defined by relation (6). Then,*

- *according to (A1.Sup) and as n goes to infinity,*

$$\sup_{f_0\in\mathcal{F}_0} \mathbb{P}_{f_0}(\Delta_n = 1) = o_n(1), \tag{7}$$

- *according to assumptions (A1.Sup), (A2.Sup) and (A3), there exists a constant $A > 0$ large enough such that as n goes to infinity*

$$\sup_{\beta\in\mathcal{T}} \sup_{f\in\Lambda(Av(nt_n^{-1},\beta))} \mathbb{P}_f(\Delta_n = 0) = o_n(1), \tag{8}$$

where $t_n = \sqrt{\log\log n}$ is the loss due to adaptivity and $v\left(nt_n^{-1}, \beta\right) = \left(nt_n^{-1}\right)^{\frac{-2\beta}{4\beta+1}}$ is the adaptive minimax rate of testing.

This means that relation (3) holds with $\alpha_0 = 0$.

Theorem 2 deals with the lower bound which is stated in the Gaussian case. It gives the optimal rate of testing H_0 against the range of alternatives when $\beta \in \mathcal{T}$, which could not be improved without losing the distinguishability between the hypotheses.

Theorem 2. (Lower bound) *Suppose that (A1.Inf)–(A3) hold and that there exists* $\tilde{f}_0 \in \mathcal{F}_0$ *such that* $\exists (C', M') : \ C' < C, \ M' < M$ *such that* $\tilde{f}_0$ *belongs to* $\Sigma(\beta, C', M')$*, for all* $\beta \in \mathcal{T}$*. Then, for any given* $\alpha_2 \in (0, 1)$*, there exists a positive constant a such that as n large enough,*

$$\inf_{\tilde{\Delta}_n}\{ \sup_{f_0 \in \mathcal{F}_0} \mathbb{P}_{f_0}(\tilde{\Delta}_n = 1) + \sup_{\beta \in \mathcal{T}} \sup_{f \in \Lambda(av(nt_n^{-1}, \beta))} \mathbb{P}_f(\tilde{\Delta}_n = 0)\} \geq \alpha_2 + o_n(1), \quad (9)$$

where $\inf_{\tilde{\Delta}_n}$ *means the infimum under all possible test functions.*

Proposition 1. *Under Assumption (A1.Sup), the estimate* $\hat{\sigma}_n^2$ *satisfies*

$$\forall \gamma > 0 \ \sup_{\beta \in \mathcal{T}} \sup_{f \in \Sigma(\beta, C, M)} \mathbb{P}_f(|\frac{\hat{\sigma}_n^2}{\sigma^2} - 1| > \gamma) \underset{n \to \infty}{\longrightarrow} 0.$$

5. Comments

Calculation of the test statistics

The test statistics defined in (5) is rather general. One can wonder whether it can be implemented in practice as $\inf_{f_0 \in \mathcal{F}_0}$ has to be computed. When the null hypothesis is parametric, the calculation can be done. For instance, let us consider $\mathcal{F}_0$ as the class of polynoms of order less than $p \in \mathbb{N}$. One can see that

$$T_{n,\beta,f_0} = \frac{1}{\sqrt{m_\beta}} \sum_{k=1}^{m_\beta} \frac{m_\beta}{n\hat{\sigma}_n^2} \left(\sum_{\substack{i,j \in I_k \\ j \neq i}} Y_i Y_j + \sum_{\substack{i,j \in I_k \\ j \neq i}} f_0(x_i) f_0(x_j) - 2 \sum_{\substack{i,j \in I_k \\ j \neq i}} f_0(x_i) Y_j \right)$$

$$= \frac{1}{\sqrt{m_\beta}} \sum_{k=1}^{m_\beta} \frac{m_\beta}{n\hat{\sigma}_n^2} \left(\sum_{\substack{i,j \in I_k \\ j \neq i}} Y_i Y_j + \left(\sum_{i \in I_k} f_0(x_i) \right)^2 - \sum_{i \in I_k} f_0^2(x_i) \right.$$
$$\left. -2 \left(\sum_{i \in I_k} f_0(x_i) \right) \left(\sum_{j \in I_k} Y_j \right) + 2 \sum_{i \in I_k} f_0(x_i) Y_i \right).$$

The first summand plays no role in the minimization. Thus one has to minimize

$$\left(\sum_{i \in I_k} f_0(x_i) \right)^2 - \sum_{i \in I_k} f_0^2(x_i) - 2 \left(\sum_{i \in I_k} f_0(x_i) \right) \left(\sum_{j \in I_k} Y_j \right) + 2 \sum_{i \in I_k} f_0(x_i) Y_i,$$

where the quantities $\left(\sum_{i\in I_k} f_0(x_i)\right)^2, \sum_{i\in I_k} f_0^2(x_i), \sum_{i\in I_k} f_0(x_i)$ and $\sum_{i\in I_k} f_0(x_i)\, Y_i$ can be easily computed in the case of polynoms.

For example, let us choose $p = 2$ and denote $\theta_0, \theta_1, \theta_2$ the coefficients of the polynoms. It leads to minimize over $\theta_0, \theta_1, \theta_2 \in \Theta \subset \mathbb{R}^3$ (Θ must satisfy Assumption (A3)) the following expression

$$\left(\theta_0 \mathrm{card}\,(I_k) + \theta_1 \sum_{i\in I_k} x_i + \theta_2 \sum_{i\in I_k} x_i^2\right)^2 - \frac{1}{2}\sum_{i\in I_k}\left(\theta_0 + \theta_1 x_i + \theta_2 x_i^2\right)^2$$
$$-2\left((\theta_0 \mathrm{card}\,(I_k) + \theta_1 \sum_{i\in I_k} x_i + \theta_2 \sum_{i\in I_k} x_i^2 \sum_{j\in I_k}) Y_j\right.$$
$$\left. + 2\theta_0 \sum_{i\in I_k} Y_i + 2\theta_1 \sum_{i\in I_k} x_i Y_i + 2\theta_2 \sum_{i\in I_k} x_i^2 Y_i\right),$$

where $\mathrm{card}\,(I_k) = \frac{n}{m_\beta}$. One can see that the minimization can be made with a computer or even by hand.

In the case of more complicated parametric families one can use numerical methods (see Clarke (1990), Lemaréchal (1976) or Hiriart-Urruty and Lemaréchal (1993a–1993b)). For instance, non-differentiable objective functions are associated with non-differentiable regression functions lying in the class defined under the null hypothesis (non-differentiability with respect to the parameter). It has to be stressed that even in this case the test statistics can be computed with numerical methods (subgradient and ε-subgradient methods). This generalizes the results obtained by Horowitz and Spokoiny (2001) who make differentiability assumptions for the regression function.

In the case of nonparametric null hypothesis, the computation of the test statistics is much more involved and has to take into account the structure of the associated class. As one can see in the proofs, the main device is to define a grid on $\mathcal{F}_0$ which approximates well the functions in $\mathcal{F}_0$.

Lower bound

The construction of the lower bound is rather classical. Yet, there is a significative difference with Spokoiny (1996). In his paper, the parametric subset in the alternatives is built using wavelets. It provides good property of independence through the orthogonality between the different levels of resolution. Here we have no natural basis for constructing the parametric family in the alternative, which makes the proof rather technical. As we are considering the discrete regression model, it is almost useless to build the parametric subset in the alternatives through an orthogonal basis such as wavelets. Indeed, the advantage to use wavelets in the Gaussian noise model is to get independence for the sets of random variables corresponding to different values for the smoothness parameter. Here the orthogonality does not lead to independence of the random variables. To cope with the dependence we met, we split the main sum over the range of β into several sums where the summands are almost independent.

About the distribution of the test statistics under the null hypothesis

In Fan, Zhang and Zhang (2001) and in Fan and Zhang (2004), the interesting Wilks phenomenon occurs. Here, this property is not needed since one just want to bound from above the first type error, therefore the asymptotic distribution under the null hypothesis of the test statistics is not required. However the upper bound of the first type error displays a probability which is free of the underlying regression function.

Link between the smoothness parameter, the entropy and the distribution tail of the errors

In this part, the link between the smoothness parameter, the entropy and the distribution tail of the errors is highlighted by several points of view. First, some general comments : one could consider the distribution of the errors as well as some parameters characterizing the class $\mathcal{F}_0$ as nuisance parameters. But unlike Fan, Zhang and Zhang (2001), Horowitz and Spokoiny (2001) and Fan and Zhang (2004), those parameters do not need to be estimated. Actually, the effect of the nuisance parameters appears in the constraints on the entropy, the distribution tail of the errors and the smoothness parameter. For instance the more nuisance parameters there are in $\mathcal{F}_0$, the stronger are the constraints on the distribution tail of the errors and the smoothness parameter.

Second, three figures below illustrate the link between the smoothness, the entropy and the tail distribution.

Figure 1. When the behavior of the tail is known (i.e. κ is given), there is a relationship between the minimum smoothness, also the maximum one and the entropy of $\mathcal{F}_0$. In the case of $\kappa = 0.95$ (heavy tail), r must be much smaller than in the case of $\kappa = 0.05$ (light tail). This phenomenon indicates that the behavior of the tail plays an important role in the problem. If a class is rich, light tails are needed in order to provide good localization. The role of $\beta_\star$ and $\beta^\star$ are important in the small values domain. When $\beta_\star$ is large, it seems to be almost no difference for r. We have to keep in mind another condition : $\mathcal{F}_0$ is included in $H(\beta^\star, C)$. It leads to a link between r and $\beta^\star$ defined by $r < \dfrac{1}{\beta^\star}$ (see Chapter 15 in Lorentz, Golitschek and Makovoz (1996) for the calculation of the entropy of a Hölder ball).

Figure 2. We suppose that the true smoothness is $\beta_0 = 1.2$ and we illustrate the effect of the smoothness range on r and κ. First, in all cases the heavier the tail is, the poorer the class $\mathcal{F}_0$ has to be. A small range for β allows large values for r. Also, we can see that the choice of $\beta_\star$ is much more important than the choice

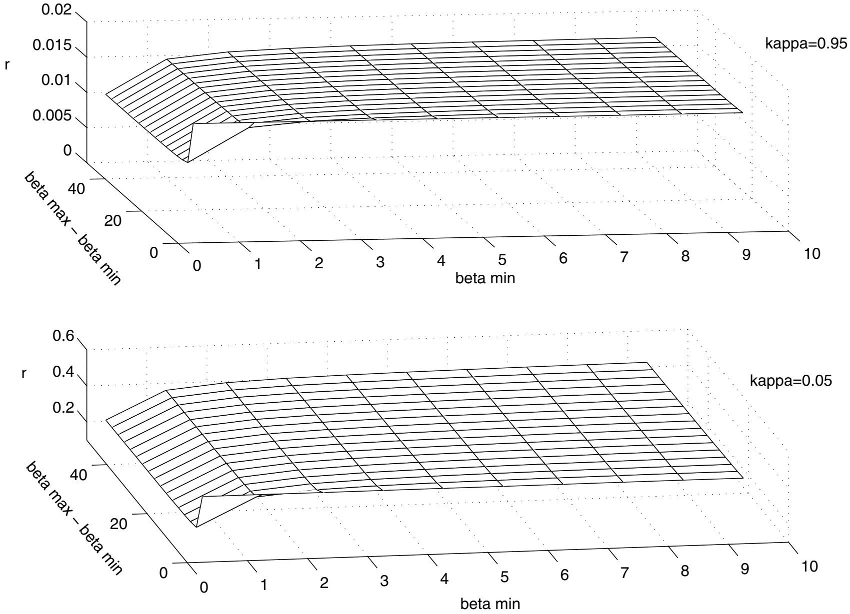

Fig. 1. κ fixed and r as the function of $\beta_\star$ and $\beta^\star - \beta_\star$

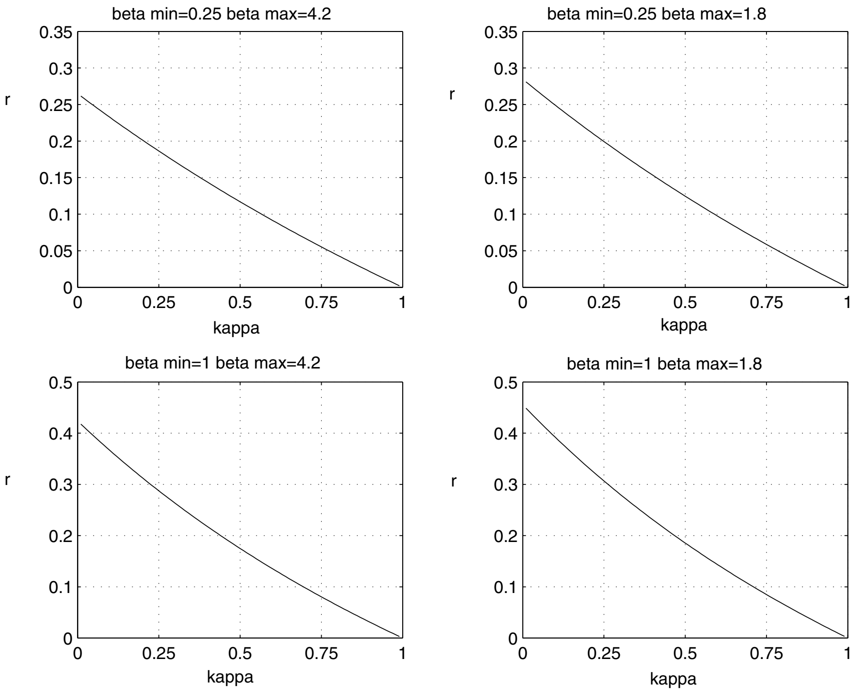

Fig. 2. $\beta_0 = 1.2$ fixed and r as the function of κ

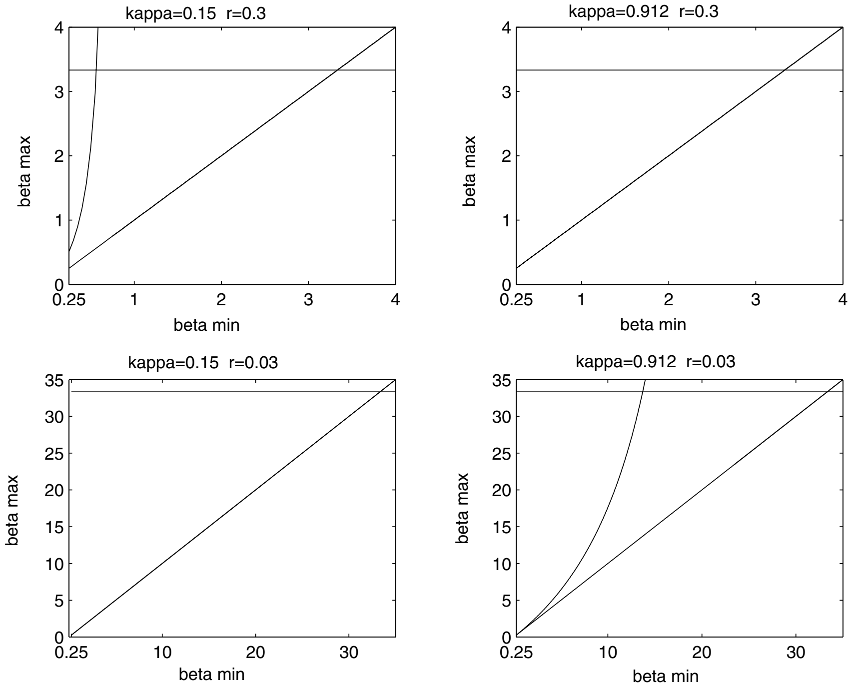

Fig. 3. κ and r fixed, the range for $\beta^\star$ as function of $\beta_\star$

of $\beta^\star$. Thus the statistician has to be more careful in the choice of $\beta_\star$ than in the choice of $\beta^\star$.

Figure 3. We consider different situations that can be encountered. We show the effect of the assumptions (A2.Sup) and (A3) on the choice of $\beta_\star$ and $\beta^\star$. For a given pair of (κ, r), the allowed area for $(\beta_\star, \beta^\star)$ is the intersection between the part of the plane under the horizontal line ($\beta^\star < \frac{1}{r}$), the one over the line $\beta_\star = \beta^\star$ since $\beta_\star \leq \beta^\star$ and the one under the third curve (Assumption (A3)). The range of adaptation is restricted by those assumptions. For example it is impossible to consider a rich class ($r = 0.3$) and a heavy tail ($\kappa = 0.912$) since Assumption (A3) is never satisfied (this is the reason why it does not appear on the graph). On the contrary for a poor class ($r = 0.03$) and a light tail ($\kappa = 0.15$), Assumption (A3) is always satisfied (this is the reason why it does not appear in the graph), we have a wide choice of $(\beta_\star, \beta^\star)$. For the two other graphs (a rich class and a light tail or a poor class and a heavy tail), Assumption (A3) plays a role for small values of $\beta_\star$.

6. Proofs

6.1. Proof of Theorem 1

The first type error : for any function f_0 in $\mathcal{F}_0$, we get

$$
\begin{aligned}
\mathbb{P}_{f_0}(\Delta_n = 1) &= \mathbb{P}_{f_0}\left(\sup_{m_\beta \in \mathcal{M}} \inf_{f \in \mathcal{F}_0} T_{n,\beta,f} \geq \rho_n\right) \\
&\leq \mathbb{P}_{f_0}\left(\sup_{m_\beta \in \mathcal{M}} T_{n,\beta,f_0} \geq \rho_n\right) \\
&\leq \mathbb{P}_{f_0}\left(\{\sup_{m_\beta \in \mathcal{M}} \frac{\hat{\sigma}_n^2}{\sigma^2} T_{n,\beta,f_0} \geq \frac{\hat{\sigma}_n^2}{\sigma^2}\rho_n\} \cap \{|\frac{\hat{\sigma}_n^2}{\sigma^2} - 1| \leq \gamma\}\right) \\
&\quad + \mathbb{P}_{f_0}\left(|\frac{\hat{\sigma}_n^2}{\sigma^2} - 1| > \gamma\right) \\
&\leq \sum_{m_\beta \in \mathcal{M}} \mathbb{P}_{f_0}\left(\frac{\hat{\sigma}_n^2}{\sigma^2} T_{n,\beta,f_0} \geq \rho_n(1-\gamma)\right) + \mathbb{P}_{f_0}\left(|\frac{\hat{\sigma}_n^2}{\sigma^2} - 1| > \gamma\right).
\end{aligned} \tag{10}
$$

Due to Proposition 1, the second term in right hand side of (10) is of order $o_n(1)$ as n large enough, and since (10) holds for any $\gamma > 0$, we take it equal to $\frac{1}{2}$. Next, we focus only on the first term in the right hand side of (10).

Note that under $\mathbb{P}_{f_0}$, the test statistics can be written in the following way

$$
\begin{aligned}
\frac{\hat{\sigma}_n^2}{\sigma^2} T_{n,\beta,f_0} &= \frac{1}{\sigma^2} \frac{1}{\sqrt{m_\beta}} \sum_{k=1}^{m_\beta} \frac{m_\beta}{n} \sum_{\substack{i \in I_k,\ j \in I_k \\ i \neq j}} \xi_i \xi_j, \\
&= \frac{1}{\sigma^2} \frac{1}{\sqrt{m_\beta}} \sum_{k=1}^{m_\beta} \eta_k,
\end{aligned} \tag{11}
$$

where $\eta_k = \frac{m_\beta}{n} \sum_{\substack{i \in I_k,\ j \in I_k \\ i \neq j}} \xi_i \xi_j$ are independent random variables. In order to bound the probability, we are going to use the Berry-Esseen inequality (Petrov (1996), Theorem 5.4 p. 149). First, let us check the conditions.
The random variables η_k are independent and centered.
The variance of η_k is

$$
\operatorname{Var}(\eta_k) = 2\left(1 - \frac{m_\beta}{n}\right)\sigma^4,
$$

and the sum of the variances is

$$
2m_\beta\left(1 - \frac{m_\beta}{n}\right)\sigma^4.
$$

Finally, the third order absolute moment of the random variable η_k is of interest. Let us compute a upper bound for the fourth order moment since $\mathbb{E}(|\eta_k|^3) \leq \mathbb{E}(|\eta_k|^4)^{3/4}$.

$$\mathbb{E}\left((\eta_k)^4\right) = \left(\frac{m_\beta}{n}\right)^4 \sum_{\substack{i_1,j_1,i_2,j_2,i_3,j_3,i_4,j_4,\in I_k \\ j_1\neq i_1, j_2\neq i_2, j_3\neq i_3, j_4\neq i_4}} \mathbb{E}\left[\left(\xi_{i_1}\xi_{j_1}\right)\left(\xi_{i_2}\xi_{j_2}\right)\left(\xi_{i_3}\xi_{j_3}\right)\left(\xi_{i_4}\xi_{j_4}\right)\right].$$

Here, note that the power of any variable ξ_i can be only 2, 3 or 4. Indeed the non-zero summands are

$$\mathbb{E}(\xi_{i_1}^4\xi_{i_2}^4), \quad \mathbb{E}(\xi_{i_1}^4\xi_{i_2}^2\xi_{i_3}^2), \quad \mathbb{E}(\xi_{i_1}^3\xi_{i_2}^3\xi_{i_3}^2), \quad \mathbb{E}(\xi_{i_1}^2\xi_{i_2}^2\xi_{i_3}^2\xi_{i_4}^2).$$

The numbers of the summands described above is less than $C_1\left(\frac{n}{m_\beta}\right)^4$ (with C_1 a fixed positive constant). Thus, the fourth order moment is bounded by a fixed constant which does not depend on the values of n, m_β, β. The third order moment exists and is also bounded by a fixed constant which does not depend on the values of n, m_β, β.
Denote

$$L_{m_\beta} = \left(\sum_{k=1}^{m_\beta} \mathrm{Var}(\eta_k)\right)^{-3/2} \sum_{k=1}^{m_\beta} \mathbb{E}(|\eta_k|^3),$$

and note that $L_{m_\beta} \leq \dfrac{C_1 m_\beta}{\left(2m_\beta\left(1-\frac{m_\beta}{n}\right)\sigma^4\right)^{\frac{3}{2}}}$. Now, the result of Berry-Esseen inequality (Theorem 5.4 in Petrov (1996)) entails

$$\mathbb{P}\left(\frac{1}{\sqrt{m_\beta}}\sum_{k=1}^{m_\beta}\eta_k > \frac{\rho_n}{2}\right) \leq 1-\Phi\left(\frac{\rho_n}{2}\right) + \frac{C_2 m_\beta}{\left(m_\beta\left(1-\frac{m_\beta}{n}\right)\sigma^4\right)^{\frac{3}{2}}},$$

where C_2 is an absolute positive constant. The well-known inequality for the standard Gaussian distribution function leads to

$$\begin{aligned}\mathbb{P}\left(\frac{1}{\sqrt{m_\beta}}\sum_{k=1}^{m_\beta}\eta_k > \frac{\rho_n}{2}\right) &\leq \frac{2}{\rho_n}e^{-\frac{\rho_n^2}{8}} + \frac{C_2 m_\beta}{\left(m_\beta\left(1-\frac{m_\beta}{n}\right)\sigma^4\right)^{\frac{3}{2}}}, \\ &\leq \frac{2}{\rho_n \log n} + \frac{C_2}{\sqrt{m_\beta}\left(\left(1-\frac{m_\beta}{n}\right)\sigma^4\right)^{\frac{3}{2}}}.\end{aligned}$$

Thus a rough upper bound for the sum is

$$\sum_{m_\beta\in\mathcal{M}} \mathbb{P}_{f_0}\left(\frac{\hat{\sigma}_n^2}{\sigma^2}T_{n,\beta,f_0} \geq \frac{\rho_n}{2}\right) \leq \frac{C_1 c\log(n)}{\left(t_n^{-1}n\right)^{\frac{1}{4\beta^*+1}}} + \frac{2c\log(n)}{\rho_n \log n}. \tag{12}$$

The right-hand side of (12) goes to 0 as n goes to $+\infty$.

The second type error : For all $\beta \in \mathcal{T}$, we have

$$\begin{aligned}
&\sup_{f\in\Lambda(Av(nt_n^{-1},\beta))} \mathbb{P}_f\big(\sup_{m_\beta\in\mathcal{M}} \inf_{f_0\in\mathcal{F}_0} T_{n,\beta,f_0} \le \rho_n\big) \\
&\le \sup_{f\in\Lambda(Av(nt_n^{-1},\beta))} \mathbb{P}_f\big(\big|\frac{\hat{\sigma}_n^2}{\sigma^2} - 1\big| > \gamma\big) \\
&\quad + \sup_{f\in\Lambda(Av(nt_n^{-1},\beta))} \mathbb{P}_f\left(\sup_{m_\beta\in\mathcal{M}} \inf_{f_0\in\mathcal{F}_0} \frac{\hat{\sigma}_n^2}{\sigma^2} T_{n,\beta,f_0} \le \rho_n(1+\gamma)\right). \qquad (13)
\end{aligned}$$

Due to Proposition 1, the first term in the right hand side of (13) is $o_n(1)$ as n large enough and we could take γ equal to 1; therefore we focus only on the second term in the right hand side of (13) :

$$\begin{aligned}
&\sup_{f\in\Lambda(Av(nt_n^{-1},\beta))} \mathbb{P}_f\left(\sup_{m_\beta\in\mathcal{M}} \inf_{f_0\in\mathcal{F}_0} \frac{\hat{\sigma}_n^2}{\sigma^2} T_{n,\beta,f_0} \le 2\rho_n\right) \\
&\le \sup_{f\in\Lambda(Av(nt_n^{-1},\beta))} \mathbb{P}_f\left(\inf_{f_0\in\mathcal{F}_0} \frac{\hat{\sigma}_n^2}{\sigma^2} T_{n,\beta_L,f_0} \le 2\rho_n\right) \\
&= \sup_{f\in\Lambda(Av(nt_n^{-1},\beta))} \mathbb{P}_f\left(\inf_{f_0\in\mathcal{F}_0} \frac{\sqrt{m_L}}{n}\frac{\hat{\sigma}_n^2}{\sigma^2} T_{n,\beta_L,f_0} \le \frac{2\rho_n\sqrt{m_L}}{n}\right), \qquad (14)
\end{aligned}$$

where β_L in T_{n,β_L,f_0} is $\beta_j \in \mathcal{T}^\star$ since the true regularity β of the function f, belongs to $[\beta_j, \beta_{j+1}]$, m_L is associated with β_L. Under $f \in \Lambda(Av(nt_n^{-1},\beta))$, rewrite $\frac{\sqrt{m_L}}{n}\frac{\hat{\sigma}_n^2}{\sigma^2} T_{n,\beta_L,f_0}$ in the following way

$$\begin{aligned}
\frac{\sqrt{m_L}}{n}\frac{\hat{\sigma}_n^2}{\sigma^2} T_{n,\beta_L,f_0} &= \frac{1}{m_L}\sum_{k=1}^{m_L} \frac{m_L^2}{\sigma^2 n^2} \\
&\quad \left(\sum_{\substack{i\in I_k,\ j\in I_k\\ i\neq j}} (\xi_i + f(x_i) - f_0(x_i))\left(\xi_j + f\left(x_j\right) - f_0\left(x_j\right)\right)\right) \\
&= \frac{1}{m_L}\sum_{k=1}^{m_L} \frac{m_L^2}{n^2\sigma^2}\left(\sum_{\substack{i\in I_k,\ j\in I_k\\ i\neq j}} \xi_i\xi_j\right) \\
&\quad +2\frac{1}{m_L}\sum_{k=1}^{m_L} \frac{m_L^2}{n^2\sigma^2}\left(\sum_{\substack{i\in I_k,\ j\in I_k\\ i\neq j}} (f(x_i) - f_0(x_i))\xi_j\right)
\end{aligned}$$

$$+\frac{1}{m_L}\sum_{k=1}^{m_L}\frac{m_L^2}{\sigma^2 n^2}\left(\sum_{\substack{i\in I_k,\ j\in I_k\\ i\neq j}}(f(x_i)-f_0(x_i))(f(x_j)-f_0(x_j))\right).$$

Set

$$T_{n,1,\beta_L}=\frac{1}{m_L}\sum_{k=1}^{m_L}\frac{m_L^2}{n^2\sigma^2}\left(\sum_{\substack{i\in I_k,\ j\in I_k\\ i\neq j}}\xi_i\xi_j\right),$$

$$T_{n,2,\beta_L,f_0}=2\frac{1}{m_L}\sum_{k=1}^{m_L}\frac{m_L^2}{\sigma^2 n^2}\left(\sum_{\substack{i\in I_k,\ j\in I_k\\ i\neq j}}(f(x_i)-f_0(x_i))\xi_j\right),$$

$$T_{n,3,\beta_L,f_0}=\frac{1}{m_L}\sum_{k=1}^{m_L}\frac{m_L^2}{n^2\sigma^2}\left(\sum_{\substack{i\in I_k,\ j\in I_k\\ i\neq j}}(f(x_i)-f_0(x_i))(f(x_j)-f_0(x_j))\right).$$

Now, define the event

$$D_{1,\beta}=\left\{\inf_{f_0\in\mathcal{F}_0}\left(T_{n,2,\beta_L,f_0}+T_{n,3,\beta_L,f_0}\right)<\frac{5}{2}\rho_n\frac{\sqrt{m_L}}{n}\right\},$$

and consider again relation (14), $\forall f\in\Lambda(Av(nt_n^{-1},\beta))$,

$$\begin{aligned}
&\mathbf{P}_f\left(\inf_{f_0\in\mathcal{F}_0}\frac{\sqrt{m_L}}{n}\frac{\hat\sigma_n^2}{\sigma^2}T_{n,\beta_L,f_0}\leq 2\rho_n\frac{\sqrt{m_L}}{n}\right)\\
&=\mathbf{P}_f\left(T_{n,1,\beta_L}+\inf_{f_0\in\mathcal{F}_0}\left(T_{n,2,\beta_L,f_0}+T_{n,3,\beta_L,f_0}\right)\leq 2\rho_n\frac{\sqrt{m_L}}{n}\right)\\
&\leq\mathbf{P}_f\left(\left\{T_{n,1,\beta_L}+\inf_{f_0\in\mathcal{F}_0}\left(T_{n,2,\beta_L,f_0}+T_{n,3,\beta_L,f_0}\right)\leq 2\rho_n\frac{\sqrt{m_L}}{n}\right\}\cap D_{1,\beta}^c\right)\\
&\quad+\mathbf{P}_f\left(D_{1,\beta}\right)\\
&\leq\mathbf{P}\left(\frac{1}{\sqrt{m_L}}\sum_{k=1}^{m_L}\eta_k\leq-\frac{\rho_n}{2}\right)+\mathbf{P}_f\left(D_{1,\beta}\right),
\end{aligned}\tag{15}$$

where the η_k's are defined in relation (11) in the proof of the first type error. Note that the first term in the right-hand side (RHS) of (15) is independent of f and is similar to the one considered in the proof of the first type error; therefore acting exactly as in the proof of the first type error, the limit as n goes to infinity of $\mathbb{P}\left(\frac{1}{\sqrt{m_L}}\sum_{k=1}^{m_L}\eta_k \le -\frac{\rho_n}{2}\right)$ is $o_n(1)$. The proof follows provided that for all $\beta \in \mathcal{T}$,

$$\sup_{f\in\Lambda(Av(nt_n^{-1},\beta))} \mathbb{P}_f(D_{1,\beta}) = o_n(1), \qquad \text{as } n \to +\infty.$$

Lemma 1. *For a sufficiently large positive constant A and under (A1.Sup)–(A2.Sup)–(A3), we get*

$$\sup_{\beta\in\mathcal{T}} \sup_{f\in\Lambda(Av(nt_n^{-1},\beta))} \mathbb{P}_f\left(D_{1,\beta}\right) = o_n(1), \qquad \text{as } n \to +\infty.$$

The proof of Lemma 1 is given in Appendix. From (13), (14), (15) and applying Lemma 1, there exists a positive constant A large enough such that relation (8) holds.

6.2. Proof of Theorem 2

First, let ϕ be a function belonging to all $H(\beta, 1)$, for $\beta \in \mathcal{T}$, with support in $\mathbb{R}$ which satisfies $\phi(x) = 0$ for $x \notin]0, 1[$, $\int_0^1 \phi(x)\,dx = 0$ and $\int_0^1 \phi(x)^2\,dx = \lambda^2$.

Let $\kappa_{n,\beta}^2 = \frac{[1 + (v(nt_n^{-1}, \beta))^{1/(4\beta)}]}{\lambda^2}\left(av(nt_n^{-1}, \beta)\right)^2$. Denote $\Xi_\beta = \{-1;+1\}^{m_\beta}$, where $m_\beta = [h_\beta^{-1}]$ ($[\cdot]$ denotes the integer part), h_β will be chosen later and is a sequence of positive numbers decreasing to zero as n goes to infinity. Define $\eta_\beta = \left(\eta_{1,\beta}, \dots, \eta_{m_\beta,\beta}\right) \in \Xi_\beta$, $z_{k,\beta} = (k-1)h_\beta$ and $A_{k,\beta} = \left[z_{k,\beta}, z_{k+1,\beta}\right]$ for $k = 1, \dots, m_\beta - 1$ and $A_{m_\beta,\beta} = \left[z_{m_\beta-1,\beta}, 1\right]$. Let $I_{k,\beta} = \left\{i : x_i \in A_{k,\beta}\right\}$ for $k = \{1, \dots, m_\beta\}$. Set

$$f_{\eta_\beta}(x) = \tilde{f}_0(x) + \sum_{k=1}^{m_\beta} \eta_{k,\beta}\kappa_{n,\beta}\phi\left(\frac{x - z_{k,\beta}}{h_\beta}\right),$$

where $\tilde{f}_0$ is the function belonging to $\mathcal{F}_0$ whose properties are specified in Theorem 2. The considered parametric set of functions, $\mathcal{F}_\beta$, is $\left\{f_{\eta_\beta}, \eta_\beta \in \Xi_\beta\right\}$. Clearly, there is a one-to-one application between Ξ_β and $\mathcal{F}_\beta$.

Second, let us define the set of β's say $\mathcal{T}'$ on which we will consider the whole parametric family $\mathcal{F}$. Recall that $\beta^\star$ is the greatest regularity and consider $h_0 := h_{\beta^\star} = (nt_n^{-1})^{-2/(4\beta^\star+1)}$ the length of each $A_{k,\beta^\star}$, $k \in \{1, \dots, m_{\beta^\star}\}$. Next, let us define the scale on h_β, $\beta \in \mathcal{T}'$ such as $h_0 = 2h_{\beta_1}$ and $h_{\beta_j} = 2h_{\beta_{j+1}}$, for all $j \in \{1, \dots, c\log n\}$, where c is a positive constant depending only on $\beta_\star$, $\beta^\star$ and n. Then the whole parametric family is defined as follows

$$\mathcal{F} = \bigcup_{j\in\{0,\dots,c\log n\}} \mathcal{F}_{\beta_j}.$$

We have to check two conditions on the set $\mathcal{F}$: for all $j \in \{0, \ldots, c \log n\}$,

$$\sup_{0 \le x \le 1} \left| f_{\eta_{\beta_j}}(x) \right| \le M, \text{ and also if } \beta_j > 1 \sup_{0 \le x \le 1} \left| f'_{\eta_{\beta_j}}(x) \right| \le M, \tag{16}$$

$$f_{\eta_{\beta_j}} \in \Sigma\left(\beta_j, L, M\right). \tag{17}$$

Relations (16) and (17) are derived from hypotheses on ϕ and $\tilde{f}_0$ exactly as relations (16) and (17) in Gayraud and Pouet (2001).

Third, we define a probability measure π on $\mathcal{F}$ as

$$\pi = \frac{1}{1 + c \log n} \sum_{j \in \{0, \ldots, c \log n\}} \pi_{\beta_j},$$

where π_{β_j} is the probability measure on $\mathcal{F}_{\beta_j}$. For all $j \in \{0, \ldots, c \log n\}$, let $\eta_1, \ldots, \eta_{m_{\beta_j}}$ be i.i.d. taking values 1 and -1 with probability $1/2$, and let π_{β_j} be the corresponding probability measure. Note that π is such that

$$\frac{d\mathbb{P}_\pi}{d\mathbb{P}_{\tilde{f}_0}} = \frac{1}{1 + c \log n} \sum_{j \in \{0, \ldots, c \log n\}} \frac{d\mathbb{P}_{\pi_{\beta_j}}}{d\mathbb{P}_{\tilde{f}_0}},$$

$$\text{where} \quad \frac{d\mathbb{P}_{\pi_{\beta_j}}}{d\mathbb{P}_{\tilde{f}_0}} = E_{\pi_{\beta_j}} \left[\frac{d\mathbb{P}_{f_{\eta_{\beta_j}}}}{d\mathbb{P}_{\tilde{f}_0}} \right] \quad \text{and} \quad \mathbb{P}_\pi = \int \mathbb{P}_f \, d\pi(f)$$

Fourth, for any $j \in \{0, \ldots, c \log n\}$, denote $\Lambda_{n,\beta_j} = \mathcal{F}_{\beta_j} \cap \Lambda(av(nt_n^{-1}, \beta_j))$. For any test function $\tilde{\Delta}_n$, the term $\sup_{f_0 \in \mathcal{F}_0} \mathbb{P}_{f_0}\left(\tilde{\Delta}_n = 1\right) + \sup_{j \in \{0, \ldots, c \log n\}} \sup_{f \in \Lambda_{n,\beta_j}} \mathbb{P}_f\left(\tilde{\Delta}_n = 0\right)$ is bounded from below by

$$\begin{aligned}
&\mathbb{P}_{\tilde{f}_0}\left(\tilde{\Delta}_n = 1\right) + \frac{1}{c \log n + 1} \sum_{j \in \{0, \ldots, c \log n\}} \sup_{f \in \Lambda_{n,\beta_j}} \mathbb{P}_f\left(\tilde{\Delta}_n = 0\right) \\
&\quad \ge \mathbb{P}_{\tilde{f}_0}\left(\tilde{\Delta}_n = 1\right) + \frac{1}{c \log n + 1} \sum_{j \in \{0, \ldots, c \log n\}} \int_{\Lambda_{n,\beta_j}} \mathbb{P}_f\left(\tilde{\Delta}_n = 0\right) d\pi_{\beta_j}(f) \\
&\quad = \mathbb{P}_{\tilde{f}_0}\left(\tilde{\Delta}_n = 1\right) + \frac{1}{c \log n + 1} \sum_{j \in \{0, \ldots, c \log n\}} \left\{ \int \mathbb{P}_f\left(\tilde{\Delta}_n = 0\right) d\pi_{\beta_j}(f) \right. \\
&\qquad \left. - \int_{(\Lambda_{n,\beta_j})^c} \mathbb{P}_f\left(\tilde{\Delta}_n = 0\right) d\pi_{\beta_j}(f) \right\} \\
&\quad \ge 1 - \mathrm{var}\left(\mathbb{P}_{\tilde{f}_0}, \mathbb{P}_\pi\right) - \frac{1}{c \log n + 1} \sum_{j \in \{0, \ldots, c \log n\}} \int_{(\Lambda_{n,\beta_j})^c} d\pi_{\beta_j}(f) \\
&\quad = (1 - \mathrm{var}\left(\mathbb{P}_{\tilde{f}_0}, \mathbb{P}_\pi\right))(1 + o_n(1)), \quad \text{as } n \text{ large enough},
\end{aligned} \tag{18}$$

where var (P, Q) is the total variation distance between the distributions P and Q (see Le Cam (1986), Chapter 4 for definition and properties) and where relation (18) holds provided that

$$\lim_{n\to\infty} \inf_{j\in\{0,\dots,c\log n\}} \pi_{\beta_j}(f \in \Lambda_{n,\beta_j}) = 1, \tag{19}$$

and

$$\lim_{n\to\infty} 1 - \mathrm{var}\left(\mathbb{P}_{\tilde{f}_0}, \mathbb{P}_\pi\right) \neq 0. \tag{20}$$

Relation (20) is proven later. The proof of (19) follows from relation (19) in Gayraud and Pouet (2001) : we replace ψ_n by $v(nt_n^{-1}, \beta_j)$ and we consider the subset $\mathcal{F}^\star_{0,j} \subset \mathcal{F}_0$ defined by $\mathcal{F}^\star_{0,j} = \{f \in \mathcal{F}_0 : \|\tilde{f}_0 - f\|_2 \leq 2\kappa_{n,\beta_j}\lambda\}$ instead of Θ'. We construct an n^{-b}-net $\mathcal{N}^\star_{0,\beta_j}$ in the supremum norm on the subset $\mathcal{F}^\star_{0,j}$ with $b > 1/2$ and following the calculations until the inequality (22) in Gayraud and Pouet (2001), we get

$$\inf_{j\in\{0,\dots,c\log n\}} \pi_{\beta_j}(f \in \Lambda_{n,\beta_j}) \geq 1 - \exp(n^{br})\exp(-n^{\frac{1}{4\beta^\star+1}}), \tag{21}$$

where the last inequality is coarse since the diameter of the set $\mathcal{F}^\star_{0,j}$ is not taken into account to calculate the entropy. The constraint $b < \dfrac{1}{r(4\beta^\star+1)}$ is compatible with $b > \frac{1}{2}$ since a lower bound for the term $\dfrac{1}{r(4\beta^\star+1)}$ is $\dfrac{1}{2r}$ which is strictly larger than $\dfrac{1}{2}$ (equivalent to $r < 1$). Therefore, the RHS of (21) tends to 1 as n goes to infinity : relation (19) is then satisfied.

Fifth, we define a partition on $\mathcal{T}'$ into M_n sets $\mathcal{T}_l$, each containing N_n β's with the following property :

$$\text{for all } \beta, \gamma \in \mathcal{T}_l : |\beta - \gamma| \geq c_1 \frac{\log(\log n)}{\log n},$$

where c_1 is a positive integer. Then, N_n is of order $\dfrac{\log n}{\log(\log n)}$ and M_n is of order $\log(\log n)$.

To conclude, it remains to prove that relation (20) holds or equivalently that $\mathrm{var}\left(\mathbb{P}_{\tilde{f}_0}, \mathbb{P}_\pi\right)$ goes to zero as n goes to infinity. We have,

$$\mathrm{var}\left(\mathbb{P}_{\tilde{f}_0}, \mathbb{P}_\pi\right) = \frac{1}{2}\int \left|\frac{d\mathbb{P}_\pi}{d\mathbb{P}_{\tilde{f}_0}} - 1\right| d\mathbb{P}_{\tilde{f}_0}.$$

$$\begin{aligned}
\int \left|\frac{d\mathbb{P}_\pi}{d\mathbb{P}_{\tilde{f}_0}} - 1\right| d\mathbb{P}_{\tilde{f}_0} &\leq \frac{1}{M_n}\sum_{l=1}^{M_n}\int \left|\frac{1}{N_n}\sum_{\beta\in\mathcal{T}_l}\frac{d\mathbb{P}_{\pi_\beta}}{d\mathbb{P}_{\tilde{f}_0}} - 1\right| d\mathbb{P}_{\tilde{f}_0}, \\
&\leq \frac{1}{M_n}\sum_{l=1}^{M_n}\sqrt{\int \left(\frac{1}{N_n}\sum_{\beta\in\mathcal{T}_l}\frac{d\mathbb{P}_{\pi_\beta}}{d\mathbb{P}_{\tilde{f}_0}} - 1\right)^2 d\mathbb{P}_{\tilde{f}_0}}.
\end{aligned}$$

We focus only on the quantity

$$\int \left(\frac{1}{N_n} \sum_{\beta \in \mathcal{T}_l} \frac{\mathrm{d}\mathbf{P}_{\pi_\beta}}{\mathrm{d}\mathbf{P}_{\tilde{f}_0}} - 1 \right)^2 \mathrm{d}\mathbf{P}_{\tilde{f}_0} = T_{1,n} + T_{2,n},$$

$$\text{where} \quad T_{1,n} = \frac{1}{N_n^2} \sum_{\beta \in \mathcal{T}_l} \int \left(\frac{\mathrm{d}\mathbf{P}_{\pi_\beta}}{\mathrm{d}\mathbf{P}_{\tilde{f}_0}} - 1 \right)^2 \mathrm{d}\mathbf{P}_{\tilde{f}_0}$$

$$\text{and} \quad T_{2,n} = \frac{1}{N_n^2} \sum_{\substack{\beta,\gamma \in \mathcal{T}_l \\ \beta \neq \gamma}} \int \left(\frac{\mathrm{d}\mathbf{P}_{\pi_\beta}}{\mathrm{d}\mathbf{P}_{\tilde{f}_0}} - 1 \right) \left(\frac{\mathrm{d}\mathbf{P}_{\pi_\gamma}}{\mathrm{d}\mathbf{P}_{\tilde{f}_0}} - 1 \right) \mathrm{d}\mathbf{P}_{\tilde{f}_0}.$$

Let us study separately $T_{1,n}$ and $T_{2,n}$. One must first note that for any $l \in \{1, \dots, M_n\}$ and for any β in $\mathcal{T}_l$,

$$\begin{aligned}
\frac{\mathrm{d}\mathbf{P}_{\pi_\beta}}{\mathrm{d}\mathbf{P}_{\tilde{f}_0}} &= E_{\pi_\beta} \left(\frac{\mathrm{d}\mathbf{P}_{f_{\eta_\beta}}}{\mathrm{d}\mathbf{P}_{\tilde{f}_0}} \right) \\
&= E_{\pi_\beta} \left(\prod_{i=1}^{n} \exp \left(\frac{1}{\sigma^2} \kappa_{n,\beta} (Y_i - \tilde{f}_0(x_i)) \sum_{s=1}^{m_\beta} \eta_{s,\beta} \phi \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) \right.\right. \\
&\qquad \left.\left. - \frac{1}{2\sigma^2} \kappa_{n,\beta}^2 \left(\sum_{s=1}^{m_\beta} \eta_{s,\beta} \phi \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) \right)^2 \right) \right) \\
&= \prod_{s=1}^{m_\beta} \exp \left(- \frac{1}{2\sigma^2} \kappa_{n,\beta}^2 \sum_{i \in I_{s,\beta}} \phi^2 \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) \right) \\
&\qquad E_{\pi_\beta} \left(\exp \left(\frac{1}{\sigma^2} \kappa_{n,\beta} \sum_{i \in I_{s,\beta}} (Y_i - \tilde{f}_0(x_i)) \eta_{s,\beta} \phi \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) \right) \right) \qquad (22) \\
&= \prod_{s=1}^{m_\beta} \exp \left(- \frac{1}{2\sigma^2} \kappa_{n,\beta}^2 \sum_{i \in I_{s,\beta}} \phi^2 \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) \right) \\
&\qquad \cosh \left(\frac{1}{\sigma^2} \kappa_{n,\beta} \sum_{i \in I_{s,\beta}} (Y_i - \tilde{f}_0(x_i)) \phi \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) \right). \qquad (23)
\end{aligned}$$

Set $\frac{1}{\sigma^2} \kappa_{n,\beta} \sum_{i \in I_{s,\beta}} (Y_i - \tilde{f}_0(x_i)) \phi \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) = Z_s \kappa_{n,\beta} \frac{1}{\sigma} \sqrt{\sum_{i \in I_{s,\beta}} \phi^2 \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right)}$, where Z_s is under $\mathbf{P}_{\tilde{f}_0}$ a standard normal real variable. Next applying the following equality,

$$\int \cosh^2(ux) \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{x^2}{2} \right) dx = \frac{1}{2} \left(1 + \exp \left(2u^2 \right) \right),$$

we obtain

$$\int \left(\frac{\mathrm{d}\mathbb{P}_{\pi_\beta}}{\mathrm{d}\mathbb{P}_{\tilde{f}_0}} \right)^2 \mathrm{d}\mathbb{P}_{\tilde{f}_0} = \prod_{s=1}^{m_\beta} \cosh \left(\kappa_{n,\beta}^2 \frac{1}{\sigma^2} \sum_{i \in I_{s,\beta}} \phi^2 \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) \right).$$

Then, from (23) and applying a classical majoration concerning cosh (see Ingster 1993), we get

$$\int \left(\frac{\mathrm{d}\mathbb{P}_{\pi_\beta}}{\mathrm{d}\mathbb{P}_{\tilde{f}_0}} \right)^2 \mathrm{d}\mathbb{P}_{\tilde{f}_0} \leq \prod_{s=1}^{m_\beta} \exp \left(c_2 \kappa_{n,\beta}^4 \frac{1}{\sigma^4} \left(\sum_{i \in I_{s,\beta}} \phi^2 \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) \right)^2 \right). \quad (24)$$

Then, since ϕ is bounded, the choice of m_β and $\kappa_{n,\beta}$ yields the order of the right–hand side term in (24) as n large enough

$$\exp \left(c_3 m_\beta \kappa_{n,\beta}^4 \frac{1}{\sigma^4} \frac{n^2}{m_\beta^2} \right) = O\left((\log n)^{c_4} \right),$$

where c_4 is a positive constant depending on σ, $\|\phi\|_\infty$, λ and a. Therefore if we choose a small enough so that c_4 is strictly less than one, we get

$$T_{1,n} \underset{n \to +\infty}{\longrightarrow} 0. \quad (25)$$

Rewrite $T_{2,n}$ as follows

$$\begin{aligned} T_{2,n} &= \frac{1}{N_n^2} \sum_{\substack{\beta,\gamma \in \mathcal{T}_l \\ \beta \neq \gamma}} \int \left(\frac{\mathrm{d}\mathbb{P}_{\pi_\beta}}{\mathrm{d}\mathbb{P}_{\tilde{f}_0}} \frac{\mathrm{d}\mathbb{P}_{\pi_\gamma}}{\mathrm{d}\mathbb{P}_{\tilde{f}_0}} - \frac{\mathrm{d}\mathbb{P}_{\pi_\beta}}{\mathrm{d}\mathbb{P}_{\tilde{f}_0}} - \frac{\mathrm{d}\mathbb{P}_{\pi_\gamma}}{\mathrm{d}\mathbb{P}_{\tilde{f}_0}} + 1 \right) \mathrm{d}\mathbb{P}_{\tilde{f}_0} \\ &= \frac{1}{N_n^2} \sum_{\substack{\beta,\gamma \in \mathcal{T}_l \\ \beta \neq \gamma}} \int \left(\frac{\mathrm{d}\mathbb{P}_{\pi_\beta}}{\mathrm{d}\mathbb{P}_{\tilde{f}_0}} \frac{\mathrm{d}\mathbb{P}\pi_\gamma}{\mathrm{d}\mathbb{P}_{\tilde{f}_0}} - 1 \right) \mathrm{d}\mathbb{P}_{\tilde{f}_0}. \end{aligned}$$

Since the number of pairs $(\beta, \gamma) \in \mathcal{T}_l \times \mathcal{T}_l$ is of order N_n^2 as n is large enough, we want to treat for any pair $\mathcal{T}_l \times \mathcal{T}_l$, the term $\int \left(\frac{\mathrm{d}\mathbb{P}_{\pi_\beta}}{\mathrm{d}\mathbb{P}_{\tilde{f}_0}} \frac{\mathrm{d}\mathbb{P}_{\pi_\gamma}}{\mathrm{d}\mathbb{P}_{\tilde{f}_0}} \right) \mathrm{d}\mathbb{P}_{\tilde{f}_0}$ which will be denoted $T_{2,\gamma,\beta,n}$ and we want to show that for any pair $(\beta, \gamma) \in \mathcal{T}_l \times \mathcal{T}_l$, the term $T_{2,\gamma,\beta,n}$ goes to 1 as n is increasing which implies that

$$\begin{aligned} T_{2,n} = \frac{1}{N_n^2} \sum_{\substack{\beta,\gamma \in \mathcal{T}_l \\ \beta \neq \gamma}} \int \left(\frac{\mathrm{d}\mathbb{P}_{\pi_\beta}}{\mathrm{d}\mathbb{P}_{\tilde{f}_0}} \frac{\mathrm{d}\mathbb{P}_{\pi_\gamma}}{\mathrm{d}\mathbb{P}_{\tilde{f}_0}} - 1 \right) \mathrm{d}\mathbb{P}_{\tilde{f}_0} &\leq \sup_{\substack{\beta,\gamma \in \mathcal{T}_l \\ \beta \neq \gamma}} T_{2,\gamma,\beta,n} - 1 \\ &= o_n(1), \text{ as } n \to +\infty. \quad (26) \end{aligned}$$

Without loss of generality, suppose that $\beta < \gamma$ so that by construction, there exists a positive integer $u_{\gamma,\beta}$ satisfying $h_\gamma = u_{\gamma,\beta} h_\beta$ and $m_\gamma < m_\beta$. For all $k \in$

$\{1, \ldots, m_\gamma\}$, the interval $A_{k,\gamma}$ contains $u_{\gamma,\beta}$ intervals $(A_{s,\beta})_s$, then denote $I^{k,\gamma}_{s,\beta} = \{i : x_i \in A_{s,\beta} \subset A_{k,\gamma}\}$ and obviously $\mathrm{card}(I^{k,\gamma}_{s,\beta}) = \dfrac{n}{m_\beta}$. Denote $(A_{s,\beta,\gamma})_s$ the intervals included in $A_{k,\gamma}$ and denote also by $z_{s,\beta,\gamma}$ the endpoints of these $(A_{s,\beta,\gamma})_s$. From previous calculations (22) and applying Fubini, we get

$$
\begin{aligned}
T_{2,\gamma,\beta,n} &= \int \prod_{s=1}^{m_\beta} \exp\left(-\frac{1}{2\sigma^2}\kappa^2_{n,\beta}\sum_{i\in I_{s,\beta}}\phi^2\left(\frac{x_i - z_{s,\beta}}{h_\beta}\right)\right) \\
&\quad E_{\pi_\beta}\left(\exp\left(\frac{1}{\sigma^2}\kappa_{n,\beta}\sum_{i\in I_{s,\beta}}(Y_i - \tilde f_0(x_i))\eta_{s,\beta}\phi\left(\frac{x_i - z_{s,\beta}}{h_\beta}\right)\right)\right) \\
&\quad \times \prod_{k=1}^{m_\gamma} \exp\left(-\frac{1}{2\sigma^2}\kappa^2_{n,\gamma}\sum_{i\in I_{k,\gamma}}\phi^2\left(\frac{x_i - z_{k,\gamma}}{h_\gamma}\right)\right) \\
&\quad E_{\pi_\gamma}\left(\exp\left(\frac{1}{\sigma^2}\kappa_{n,\gamma}\sum_{i\in I_{k,\gamma}}(Y_i - \tilde f_0(x_i))\eta_{k,\gamma}\phi\left(\frac{x_i - z_{k,\gamma}}{h_\gamma}\right)\right)\right) \mathrm{d}\mathbb{P}_{\tilde f_0} \\
&= \prod_{k=1}^{m_\gamma}\prod_{s=1}^{u_{\gamma,\beta}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i\in I^{k,\gamma}_{s,\beta}}\left\{\kappa^2_{n,\beta}\phi^2\left(\frac{x_i - z_{s,\beta,\gamma}}{h_\beta}\right) + \kappa^2_{n,\gamma}\phi^2\left(\frac{x_i - z_{k,\gamma}}{h_\gamma}\right)\right\}\right) \\
&\quad \times E_{\pi_\beta}E_{\pi_\gamma}\int \exp\left(\frac{1}{\sigma^2}\sum_{i\in I^{k,\gamma}_{s,\beta}}(Y_i - \tilde f_0(x_i))\{\kappa_{n,\beta}\eta_{s,\beta}\phi\left(\frac{x_i - z_{s,\beta,\gamma}}{h_\beta}\right)\right. \\
&\qquad \left. + \kappa_{n,\gamma}\eta_{k,\gamma}\phi\left(\frac{x_i - z_{k,\gamma}}{h_\gamma}\right)\}\right) \mathrm{d}\mathbb{P}_{\tilde f_0} \\
&= \prod_{k=1}^{m_\gamma}\prod_{s=1}^{u_{\gamma,\beta}} E_{\pi_\beta}E_{\pi_\gamma}\exp\left(\frac{1}{\sigma^2}\sum_{i\in I^{k,\gamma}_{s,\beta}}\kappa_{n,\beta}\eta_{s,\beta}\phi\left(\frac{x_i - z_{s,\beta,\gamma}}{h_\beta}\right)\kappa_{n,\gamma}\eta_{k,\gamma}\right. \\
&\quad \left.\phi\left(\frac{x_i - z_{k,\gamma}}{h_\gamma}\right)\right) \\
&= \prod_{k=1}^{m_\gamma}\prod_{s=1}^{u_{\gamma,\beta}} \cosh\left(\frac{1}{\sigma^2}\sum_{i\in I^{k,\gamma}_{s,\beta}}\kappa_{n,\beta}\phi\left(\frac{x_i - z_{s,\beta,\gamma}}{h_\beta}\right)\kappa_{n,\gamma}\phi\left(\frac{x_i - z_{k,\gamma}}{h_\gamma}\right)\right) \\
&\le \exp\left(c_5 m_\gamma u_{\gamma,\beta}\kappa^2_{n,\beta}\kappa^2_{n,\gamma}\frac{n^2}{m^2_\beta}\right), \qquad (27)
\end{aligned}
$$

where the last inequality holds due to (24), and the positive constant c_5 depends on $\|\phi\|_\infty$ and σ^2. Since for any β we have $\kappa^2_{n,\gamma}\dfrac{n}{\sqrt{m_\beta}} = c_6 t_n$ where c_6 is a positive constant depending on a and λ, the right hand side in (27) becomes

$$\begin{aligned}
\exp\left(c_6 m_\gamma u_{\gamma,\beta}\kappa_{n,\beta}^2\kappa_{n,\gamma}^2\frac{n^2}{m_\beta^2}\right) &= \exp\left(c_6 m_\gamma \frac{m_\beta}{m_\gamma}\kappa_{n,\beta}^2\kappa_{n,\gamma}^2\frac{n}{\sqrt{m_\beta}}\frac{n}{\sqrt{m_\gamma}}\frac{1}{m_\beta}\sqrt{\frac{m\gamma}{m_\beta}}\right)\\
&= \exp\left(c_7 t_n^2\sqrt{\frac{m\gamma}{m_\beta}}\right)\\
&= \exp\left(c_7 t_n^2\sqrt{\frac{m\gamma}{m_\beta}}\right)\\
&= o_n\left((\log(\log n))(\log n)^{-\delta}\right) \text{ as } n\to\infty, \qquad (28)
\end{aligned}$$

where δ is a positive constant depending on the positive integer appearing in the difference separating β and γ, and c_7 depends on c_5 and c_6. Relation (28) implies (26), which together with (25) yield the result and if we take a sufficiently small (9) holds with $\alpha_2 = 1$.

6.3. Proof of Proposition 1

For any $\beta \in \mathcal{T}$ and for any f in $\Sigma(\beta, C, M)$, we get

$$\begin{aligned}
\hat{\sigma}_n^2 &= \frac{1}{2(n-1)}\sum_{i=2}^{n}(Y_i - Y_{i-1})^2\\
&= \frac{1}{2(n-1)}\sum_{i=2}^{n}(\xi_i - \xi_{i-1})^2 + \frac{1}{2(n-1)}\sum_{i=2}^{n}(f(x_i) - f(x_{i-1}))^2\\
&\quad + \frac{1}{(n-1)}\sum_{i=2}^{n}(f(x_i) - f(x_{i-1}))(\xi_i - \xi_{i-1}). \qquad (29)
\end{aligned}$$

It can be shown that under (A1.Sup), the first part in the RHS of (29) tends to σ^2 in probability as n goes to infinity. Indeed, if we split the first term in the RHS of (29) into two terms which depend on the quantities $(\xi_i - \xi_{i-1})^2$ with i even and with i odd respectively, then both terms are empirical mean of i.i.d. variables with expectation equal to $\sigma^2/2$; the weak large numbers law achieves the proof. For the third part, Chebyshev's inequality yields for any $\beta \in \mathcal{T}$ and for any f in $\Sigma(\beta, C, M)$ and any $\tau > 0$:

$$\mathbb{P}_f\left(\frac{1}{(n-1)}\sum_{i=2}^{n}(\xi_i - \xi_{i-1})(f(x_i) - f(x_{i-1})) > \tau\right) = o_n(1) \text{ as } n \text{ goes to } +\infty.$$

Finally, for the second part,

$$|f(x_i) - f(x_{i-1})|^2 \leq \begin{cases} C^2|x_i - x_{i-1}|^{2\beta} \leq C^2 n^{-2\beta_\star} & \text{if } \beta \leq 1,\\ M^2|x_i - x_{i-1}|^2 \leq M^2 n^{-2} & \text{if } \beta > 1.\end{cases}$$

Therefore,

$$\frac{1}{2(n-1)}\sum_{i=2}^{n}(f(x_i)-f(x_{i-1}))^2 \le \frac{n-1}{2(n-1)}(\sup(M,C))^2 n^{-2\beta_\star},$$

which goes to zero as $n \to +\infty$.

APPENDIX

Proof of Lemma 1. For all $\beta \in \mathcal{T}$ and for all $f \in \Lambda\left(Av(nt_n^{-1},\beta)\right)$, denote $\beta_L \in \mathcal{T}^\star$ satisfying $\beta_L \le \beta < \beta_{L+1}$, where β_{L+1} is also in $\mathcal{T}^\star$; the proof of the lemma will be made in three steps :

- (a) First, we replace the infimum over $\mathcal{F}_0$ by an infimum over a set $\mathcal{N}_0$ whose cardinality is controlled by Assumption (A3). The set $\mathcal{N}_0$ will be defined later.
- (b) Second, we obtain a lower bound for $T_{n,3,\beta_L,f_0^N}$, $f_0^N \in \mathcal{N}_0$.
- (c) Third, we use Bernstein's inequality.

(a) According to Assumption (A3), for all $f_0 \in \mathcal{F}_0$ there exists $f_0^N \in \mathcal{N}_0$ such that $\sup_{x\in[0,1]} |f_0(x) - f_0^N(x)| \le \delta$, with $\delta = n^{-4\beta^\star/(4\beta^\star+1)-\kappa 2\beta_\star/(4\beta_\star+1)}$. The functional set $\mathcal{N}_0$ is the smallest set of functions f_0^N which are needed to cover $\mathcal{F}_0$. Then,

$$\begin{aligned}|T_{n,2,\beta_L,f_0} - T_{n,2,\beta_L,f_0^N}| &= \left|2\frac{1}{m_L}\sum_{k=1}^{m_L}\frac{m_L^2}{\sigma^2 n^2}\left(\sum_{i\in I_k, j\in I_k, i\neq j}[f_0^N(x_i) - f_0(x_i)]\xi_j\right)\right| \\ &\le \frac{2}{\sigma^2}\delta \sup_{i\in\{1,\dots,n\}}|\xi_i|. \end{aligned} \tag{30}$$

The same occurs for $T_{n,3,\beta_L,f_0}$ i.e.

$$\begin{aligned}|T_{n,3,\beta_L,f_0} - T_{n,3,\beta_L,f_0^N}| &= \Bigg|\frac{1}{m_L}\sum_{k=1}^{m_L}\frac{m_L^2}{\sigma^2 n^2}\sum_{i\in I_k,\ j\in I_k,\ i\neq j}[f(x_i) - f_0^N(x_i)] \\ &\quad [f(x_j) - f_0^N(x_j)] - \\ &\quad [f(x_i) - f_0^N(x_i) + f_0^N(x_i) - f_0(x_i)][f(x_j) \\ &\quad - f_0^N(x_j) + f_0^N(x_j) - f_0(x_j)]\Bigg| \\ &\le \delta^2\frac{2}{\sigma^2} + \delta\frac{4}{\sigma^2}(2M+\delta) \\ &= \delta^2\frac{6}{\sigma^2} + \delta\frac{8M}{\sigma^2}, \end{aligned} \tag{31}$$

where M is one of the constants which characterizes the functional class under the alternative. Thus relations (30) and (31) lead to

$$\begin{aligned}&\inf_{f_0\in\mathcal{F}_0}\left(T_{n,2,\beta_L,f_0} + T_{n,3,\beta_L,f_0}\right) \\ &\quad \ge \inf_{f_0^N\in\mathcal{N}_0}\left(T_{n,2,\beta_L,f_0^N} + T_{n,3,\beta_L,f_0^N} - \frac{2}{\sigma^2}\delta\sup_{i=1,\dots,n}|\xi_i| - \delta^2\frac{6}{\sigma^2} - \delta\frac{8M}{\sigma^2}\right).\end{aligned}$$

Denote $B(n) = n^{\kappa \frac{2\beta_\star}{4\beta_\star+1}}$, where $\kappa < 1$ is the positive constant defined in Assumption (A2.Sup). Next, consider $D_2 = \{ \sup_{i=1,\dots,n} |\xi_i| > B(n)\}$, then by Assumption (A2.Sup) and the choice of $B(n)$, $\mathbb{P}(D_2) \to 0$ as $n \to +\infty$. Moreover the choice of $B(n)$ entails that $\frac{2}{\sigma^2}\delta B(n) + \delta^2 \frac{6}{\sigma^2} + \delta \frac{8M}{\sigma^2} = o_n(\rho_n \frac{\sqrt{m_L}}{n})$. To conclude step (a), just note that for all $\beta \in \mathcal{T}$ and uniformly over $\Lambda(Av(nt_n^{-1}, \beta))$, we have as n large enough

$$\mathbb{P}_f\left(D_{1,\beta}\right) \leq \mathbb{P}_f\left(\inf_{f_0^N \in \mathcal{N}_0} \left(T_{n,2,\beta_L,f_0^N} + T_{n,3,\beta_L,f_0^N}\right) < 3\rho_n \frac{\sqrt{m_L}}{n}\right). \quad (32)$$

(b) For all $\beta \in \mathcal{T}$ and for all $f \in \Lambda(Av(nt_n^{-1}, \beta))$, we get following from Gayraud and Pouet (2001), in both cases $\beta > 1$ and $\beta \leq 1$,

$$\begin{aligned} T_{n,3,\beta_L,f_0^N} &= T_{n,3,\beta_L,f_0^N} - T_{n,3,\beta_L,f_0} + T_{n,3,\beta_L,f_0} \\ &\geq \tilde{C}^2 A^2 m_L^{-2\beta} - \delta \\ &\geq \frac{\tilde{C}^2}{4} A^2 m_\beta^{-2\beta}, \end{aligned} \quad (33)$$

where the last inequality holds due to the value of δ and provided that $\beta_L \leq \beta$; the positive constant $\tilde{C}$ depends on β, β_L, C, M. For simplicity sake, denote $c^2 = \frac{\tilde{C}^2}{4}$. Then from (33),

$$\begin{aligned} \inf_{f_0^N \in \mathcal{N}_0} \left(T_{n,2,\beta_L,f_0^N} + T_{n,3,\beta_L,f_0^N}\right) &\geq cAm_\beta^{-\beta} \inf_{f_0^N \in \mathcal{N}_0} \left[\frac{T_{n,2,\beta_L,f_0^N}}{\sqrt{T_{n,3,\beta_L,f_0^N}}} + \sqrt{T_{n,3,\beta_L,f_0^N}}\right] \\ &\geq cAm_\beta^{-\beta} \left[\inf_{f_0^N \in \mathcal{N}_0} \frac{T_{n,2,\beta_L,f_0^N}}{\sqrt{T_{n,3,\beta_L f_0^N}}} + cAm_\beta^{-\beta}\right]. \end{aligned}$$

Then, as $-\sup(-x) = \inf(x)$, we obtain

$$\begin{aligned} \mathbb{P}_f\left(D_{1,\beta}\right) &\leq \mathbb{P}_f\left(\frac{3\rho_n \frac{\sqrt{m_L}}{n}}{cAm_\beta^{-\beta}} \geq -\sup_{f_0^N \in \mathcal{N}_0} \left|\frac{T_{n,2,\beta_L,f_0^N}}{\sqrt{T_{n,3,\beta_L,f_0^N}}}\right| + cAm_\beta^{-\beta}\right) \\ &= \mathbb{P}_f\left(\sup_{f_0^N \in \mathcal{N}_0} \left|\frac{T_{n,2,\beta_L,f_0^N}}{\sqrt{T_{n,3,\beta_L,f_0^N}}}\right| \geq cAm_\beta^{-\beta} - \frac{3\sqrt{8}(\log(\log n))^{1/2}}{cAm_\beta^{-\beta}} \frac{\sqrt{m_L}}{n}\right) \\ &\leq \sum_{f_0^N \in \mathcal{N}_0} \mathbb{P}_f\left(\left|\frac{T_{n,2,\beta_L,f_0^N}}{\sqrt{T_{n,3,\beta_L,f_0^N}}}\right| \geq cAm_\beta^{-\beta} - \frac{3\sqrt{8}(\log(\log n))^{1/2}\sqrt{m_L}}{ncAm_\beta^{-\beta}}\right). \end{aligned} \quad (34)$$

(c) Let us consider the event $D_{3,\beta} = \left\{\left|\frac{T_{n,2,\beta_L,f_0^N}}{\sqrt{T_{n,3,\beta_L,f_0^N}}}\right| \geq cAm_\beta^{-\beta} - \frac{3\sqrt{8}(\log(\log n))^{1/2}\sqrt{m_L}}{ncAm_\beta^{-\beta}}\right\}$. We use the following lemma to prove that the event is negligible in probability.

Lemma 2. *For all $\beta \in \mathcal{T}$ and for any $f \in \Lambda(Av(nt_n^{-1}, \beta))$, we get as n large enough,*

$$\sum_{f_0^N \in \mathcal{N}_0} \mathbb{P}_f(D_{3,\beta} \cap D_2^c) = o_n(1),$$

where D_2^c is the complement of the event $D_2 = \{ \sup_{i=1,\dots,n} |\xi_i| > B(n)\}$ and $B(n) = n^{\kappa \frac{2\beta_\star}{4\beta_\star+1}}$.

Then, from relation (34) in step (b), and applying Lemma 2, Lemma 1 is proved.

Proof of Lemma 2. The main device is Bernstein's inequality. Let

$$Z_j = \frac{m_L}{n^2\sigma^2} \sum_{\substack{i \in I_k \\ i \neq j}} \frac{f(x_i) - f_0^N(x_i)}{\sqrt{T_{n,3,\beta_L,f_0^N}}} \xi_j, \quad \text{if } j \in I_k.$$

The random variables $Z_1, \dots, Z_n$ are independent with zero mean since $\mathbb{E}(\xi_i) = 0,\ \forall i \in \{1, \dots, n\}$. Similarly to Gayraud and Pouet (2001), we obtain an upper bound for Z_j, for all $j = 1, \dots, n$, that is

$$\left|Z_j\right| \leq K_1 B(n) \frac{\sqrt{m_L}}{\sigma^2 n}, \tag{35}$$

where K_1 is a positive constant. Also, we get

$$\sum_{j=1}^{n} \mathbb{E}_f \left(\left(Z_j - \mathbb{E}_f\left(Z_j\right)\right)^2 \right) \leq \frac{K_2}{n}, \tag{36}$$

where K_2 is a positive constant.

Next from (35) and (36) and applying Bernstein's inequality, we get an exponential upper bound,

$$\begin{aligned} &\mathbb{P}_f \left(\left| \frac{T_{n,2,\beta_L,f_0^N}}{\sqrt{T_{n,3,\beta_L,f_0^N}}} \right| \geq cAm_\beta^{-\beta} - \frac{3\sqrt{8}(\log(\log n))^{1/2}\sqrt{m_L}}{ncAm_\beta^{-\beta}} \right) \\ &\leq \exp\left(- \frac{\left(cAm_\beta^{-\beta} - \frac{3\sqrt{8}(\log(\log n))^{1/2}\sqrt{m_L}}{ncAm_\beta^{-\beta}} \right)^2}{\frac{K_2}{n} + \frac{(cAm_\beta^{-\beta} - \frac{3\sqrt{8}(\log(\log n))^{1/2}\sqrt{m_L}}{ncAm_\beta^{-\beta}}) K_1 \frac{B(n)\sqrt{m_L}}{\sigma^2 n}}{3}} \right), \end{aligned} \tag{37}$$

where $B(n) = n^{\kappa \frac{2\beta_\star}{4\beta_\star+1}}$ and with κ satisfying Assumption (A2.Sup). Such a choice for $B(n)$ yields $\mathbb{P}(\sup_{i=1,\dots,n} |\xi_i| > B(n)) = o_n(1)$ as $n \to +\infty$.

Let us consider the squared term in the numerator of the exponential term in the previous inequality and let us denote it N^2 and show that N is of order $m_\beta^{-\beta}$:

$$
\begin{aligned}
N &= m_\beta^{-\beta}[cA - 3\sqrt{8}t_n^{(4\beta+1)/(4\beta+1)}(cA)^{-1}n^{1/(4\beta_L+1)}n^{-(4\beta+1)/(4\beta+1)}n^{4\beta/(4\beta+1)} \\
&\quad t_n^{-1/(4\beta_L+1)}t_n^{-4\beta/(4\beta+1)}] \\
&= m_\beta^{-\beta}[cA - 3\sqrt{8}(cA)^{-1}n^{1/(4\beta_L+1)}n^{-1/(4\beta+1)}t_n^{-1/(4\beta_L+1)}t_n^{1/(4\beta+1)}] \\
&= m_\beta^{-\beta}[cA - 3\sqrt{8}(cA)^{-1}\exp\{\log(n)4(\beta-\beta_L)/\iota\}t_n^{-4(\beta-\beta_L)/\iota}] \\
&= m_\beta^{-\beta}[cA - 3\sqrt{8}(cA)^{-1}\exp\{4a_0\}\exp\{-2(\log(\log(\log n)))(\beta-\beta_L)/\iota\}] \\
&\geq m_\beta^{-\beta}c\frac{A}{2},
\end{aligned}
$$

where $\iota = (4\beta_L+1)(4\beta+1)$ and a_0 are positive constants and where the last relation holds since we can choose A sufficiently large enough and since the last exponential term in the above equation is a constant smaller than one as n is large enough. Actually choose A sufficiently large such that $cA - 3\sqrt{8}(cA)^{-1}\exp\{4a_0\}\exp\{-4/\iota\} \geq A/2$. Then we obtain

$$
\frac{\left(cAm_\beta^{-\beta} - \frac{3\sqrt{8}\sqrt{\log(\log n)m_L}}{ncAm_\beta^{-\beta}}\right)^2}{\frac{K_2}{n}} \sim n^{1/(4\beta+1)}t_n^{4\beta/(4\beta+1)},
$$

$$
\frac{\left(cAm_\beta^{-\beta} - \frac{3\sqrt{8}\sqrt{\log(\log n)m_L}}{ncAm_\beta^{-\beta}}\right)^2}{\frac{\left(cAm_\beta^{-\beta} - \frac{3\sqrt{8}\sqrt{\log(\log n)m_L}}{ncAm_\beta^{-\beta}}\right)K_1\frac{B(n)\sqrt{m_L}}{\sigma^2 n}}{3}} \sim n^{(2\beta)/(4\beta+1)}n^{4(\beta_L-\beta)/\iota}(B(n))^{-1}
$$

$$
t_n^{2\beta/(4\beta+1)}t_n^{1/(4\beta_L+1)}.
$$

The choice of m, t_n and $B(n)$ entails that the right-hand side in (37), which is independent of the regression function f, is decreasing to 0 exponentially. Finally, due to the previous argument and due to Assumption (A3), Lemma 2 follows. More precisely, the RHS in (37) is at least of order

$$
\exp(-n^{\frac{2\beta_\star}{4\beta_\star+1}-\kappa\frac{2\beta_\star}{4\beta_\star+1}}),
$$

which would be compared to the cardinality of $\mathcal{N}_0$ which is $\exp(\delta^{-r})$, with $\delta = n^{-\frac{4\beta^\star}{4\beta^\star+1}-\kappa\frac{2\beta_\star}{4\beta_\star+1}}$. Denote $R(\beta_\star,\beta^\star) = \frac{4\beta^\star}{4\beta^\star+1} + \kappa\frac{2\beta_\star}{4\beta_\star+1} = \frac{(16+8\kappa)\beta_\star\beta^\star+2(2\beta^\star+\kappa\beta_\star)}{(4\beta_\star+1)(4\beta^\star+1)}$. Then the relation between r, and $\beta^\star$, $\beta_\star$ and κ is

$$
\begin{aligned}
R(\beta_\star,\beta^\star)r &< \frac{2\beta_\star(1-\kappa)}{4\beta_\star+1} \\
&\Leftrightarrow \\
r &< \frac{2\beta_\star(1-\kappa)(1+4\beta^\star)}{(16+8\kappa)\beta_\star\beta^\star+2(2\beta^\star+\kappa\beta_\star)}. \qquad (38)
\end{aligned}
$$

Relation (38) is exactly the required condition in Assumption (A3).

Acknowledgements. The thoughtful comments of two anonymous referees, leading to significant improvement of this paper, are gratefully acknowledged. We are also grateful to Oleg Lepskii for helpful discussions.

References

1. Baraud, Y., Huet, S., Laurent, B.: Adaptive tests of linear hypothesis by model selection. Ann. Stat., **31** (1), 225–251 (2003)
2. Clarke, F. H.: Optimization and Nonsmooth Analysis. Second edition. Classics in Applied Mathematics, 5. Society for Industrial and Applied Mathematics (SIAM), Philadelphia 1990
3. Ermakov, M.S.: Minimax detection of a signal in a white Gaussian noise. Theory Probab. Appl., **35**, 667–679 (1990)
4. Fan, J.: On the estimation of quadratic functionals. Ann. Stat., **19** (3), 1273–1294 (1991)
5. Fan, J, Zhang, C., Zhang, J.: Generalized likelihood ratio statistics and Wilks phenomenon. Ann. Stat., **29** (1), 153–193 (2001)
6. Fan, J., Zhang, J.: Sieve empirical likelihood ratio tests for nonparametric functions. Ann. Stat., **32** (5), 1858–1907 (2004)
7. Fromont, M., Lévy-Leduc, C.: Adaptive tests for periodic signal detection with applications to laser vibrometry. Prépublication d'Orsay, **2003–76**, 2003
8. Gayraud, G., Pouet, Ch.: Minimax Testing composite null hypotheses in the discrete regression scheme. Math. Methods Stat., **4**, 375–394 (2001)
9. Guerre, E., Lavergne, P.: Optimal minimax rates for nonparametric specification testing in regression models. Econom. Theory, **18**, 1139–1171 (2002)
10. Härdle, W., Kneip, A.: Testing a regression model when we have smooth alternatives in mind. Scand. J. Stat., **26** (2), 221–238 (1999)
11. Härdle, W., Mammen, E.: Comparing nonparametric versus parametric fits. Ann. Stat., **21** (4), 1926–1947 (1993)
12. Hart, J.D.: Nonparametric Smoothing and Lack-of-fit Tests. Springer, N.Y. e.a. 1997
13. Hiriart-Urruty, J.-B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms. I. Fundamentals. Grundlehren der Mathematischen Wissenschaften, 305. Springer-Verlag, Berlin 1993a
14. Hiriart-Urruty, J.-B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms. II. Advanced Theory and Bundle Methods. Grundlehren der Mathematischen Wissenschaften, 306. Springer-Verlag, Berlin 1993b
15. Horowitz, J.L., Spokoiny, V.G.: An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. Econometrica, **69**, 599–631 (2001)
16. Ingster, Yu.I.: On minimax nonparametric detection of a signal in Gaussian white noise. Probl. Inf. Trans., **18** (2), 61–73 (1982)
17. Ingster, Yu.I.: Minimax detection of signals in L_p–metrics. Zap. Nauchn. Semin. LOMI, **184**, 152–168 (1990)
18. Ingster, Yu.I.: Asymptotically minimax hypothesis testing for nonparametric alternatives. I, II, III., Math. Methods Stat., **2**, 85–114; 1710-189; 249–268 (1993)
19. Le Cam, L.M.: Asymptotic Methods in Statistical Decision theory, Springer Series in Statistics. Springer, New–York 1986
20. Lemaréchal, C.: Nondifferentiable optimisation subgradient and ε-subgradient methods. Optimization and operations research. Proc. Conf., Oberwolfach, (1975), pp. 191–199. Lecture Notes in Econom. Math. Systems, Vol. 117, Springer, Berlin, 1976

21. Lepskii, O.V.: On asymptotically exact testing of nonparametric hypotheses. Discussion paper, Université de Louvain 1993
22. Lepskii, O.V., Nemirovski, A., Spokoiny, V.G.: On estimation of the L_r norm of a regression function. Probab. Theory Relat. Fields, **113**, 221–253 (1999)
23. Lepskii, O.V., Spokoiny, V.G.: Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. Bernoulli, **5**, 333–358 (1999)
24. Lepskii, O.V., Tsybakov, A.B.: Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. Probab. Theory Relat. Fields, **117**, 17–48 (2000)
25. Lorentz, G.G., Golitschek, M.V., Makovoz, Y.: Constructive Approximation : advanced problems. Springer Verlag, Berlin 1996
26. Petrov, V.V.: Limit Theorems of Probability Theory. Oxford University Press, New-York 1996
27. Spokoiny, V.G.: Adaptive hypothesis testing using wavelets. Ann. Stat., **24** (6), 2477–2498 (1996)
28. Spokoiny, V.G.: Adaptive and spatially adaptive testing of a nonparametric hypothesis. Math. Methods Stat., **7** (3), 245–273 (1998)
29. Suslina, I.A.: Minimax detection of a signal for l_p–ball. Zap. Nauchn. Semin. POMI, **207**, 127–137 (1993)