Testing the regularity of a smooth signal

ALEXANDRA CARPENTIER

Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, CB3 0WB Cambridge, United Kingdom. E-mail: a.carpentier@statslab.cam.ac.uk

We develop a test to determine whether a function lying in a fixed L_2-Sobolev-type ball of smoothness t, and generating a noisy signal, is in fact of a given smoothness $s \geq t$ or not. While it is impossible to construct a uniformly consistent test for this problem on *every* function of smoothness t, it becomes possible if we remove a sufficiently large region of the set of functions of smoothness t. The functions that we remove are functions of smoothness strictly smaller than s, but that are very close to s-smooth functions. A lower bound on the size of this region has been proved to be of order $n^{-t/(2t+1/2)}$, and in this paper, we provide a test that is consistent after the removal of a region of such a size. Even though the null hypothesis is composite, the size of the region we remove does not depend on the complexity of the null hypothesis.

Keywords: functional analysis; minimax bounds; non-parametric composite testing problem

1. Introduction

We consider in this paper a composite testing problem in the non-parametric Gaussian regression setting. Assuming that the unknown regression function f lies in a given smoothness class (indexed by t), we want to decide whether f is in fact in a much more regular class (indexed by $s \geq t$), by constructing a suitable test. More precisely, we consider the setting of testing between two fixed L_2 Sobolev-type classes, which we define formally in Section 2 below.

Let $\Sigma(t, B)$ be the L_2-Sobolev-type ball of functions in $[0, 1]$ of smoothness t and radius B, and let $\Sigma(s, B)$ with $s > t$ be a sub-model (i.e., $\Sigma(s, B) \subset \Sigma(t, B)$). We assume that we have observations generated according to a Gaussian non-parametric model with underlying function f, at noise level n, where $f \in \Sigma(s, B)$ or $f \in \Sigma(t, B) \setminus \Sigma(s, B)$.

For $G \subset L_2$, set $\|f - G\|_2 = \inf_{g \in G} \|f - g\|_2$. We define for $\rho_n \geq 0$ the sets

$$\tilde{\Sigma}(t, B, \rho_n) = \left\{ f \in \Sigma(t, B) : \left\| f - \Sigma(s, B) \right\|_2 \geq \rho_n \right\}.$$

Note that these sets are separated away from $\Sigma(s, B)$ whenever $\rho_n > 0$. They correspond to $\Sigma(t, B) \setminus \Sigma(s, B)$ where we have removed some critical functions, very close to functions in $\Sigma(s, B)$.

We are interested in the composite testing problem:

$$H_0 : f \in \Sigma(s, B) \quad \text{vs.} \quad H_1 : f \in \tilde{\Sigma}(t, B, \rho_n). \tag{1.1}$$

More precisely, we want to know the minimal order of magnitude of ρ_n that enables the construction of a uniformly consistent test Ψ_n between H_0 and H_1, that is, of a test such that there exists N that depends on H_0, H_1 and α only such that for any $n \geq N$,

$$\sup_{f \in H_0} \mathbb{E}_f \Psi_n + \sup_{f \in H_1} \mathbb{E}_f (1 - \Psi_n) \leq \alpha.$$

Two topics that are closely related to this question have been thoroughly studied. The first one is non-parametric signal detection where $H_0 = \{0\}$. The second is the creation of adaptive and honest non-parametric confidence bands around functions.

Let us first recall the results obtained in signal detection where one wishes to test

$$H_0 : f = 0 \quad \text{vs.} \quad H_1 : f \in \big\{ f \in \Sigma(t, B) : \| f - 0 \|_2 \geq \rho_n \big\}. \tag{1.2}$$

As in any testing problem, in order to obtain a uniformly consistent test, the model has to be restricted such that the elements in H_0 are not too close to the ones in H_1. This explains the presence of the separation by ρ_n. Ingster [23,24], Spokoiny [37] and Ingster and Suslina [25] prove that the minimal order of ρ_n that enables the existence of a consistent test in the above problem is

$$\rho_n \geq Dn^{-t/(2t+1/2)}.$$

For ρ_n of this order, the authors also build a consistent test for the testing problem (1.2). They prove that the testing problem is equivalent to testing whether the sum of the squares of the means of independent (or close to independent) sub-Gaussian random variables is null or not, and the usual χ_2-test theory applies. The size ρ_n of the separation area is related to the minimax rate of estimation of the L_2 norm of f under the alternative hypothesis. A question that arises is how the results change when the null hypothesis is a composite hypothesis, in our case an infinite dimensional Sobolev-type ball.

The testing problem described in equation (1.1) is also closely connected to the problem of the creation of confidence bands around functions – see, for instance, Hoffmann and Lepski [20], Juditsky and Lambert-Lacroix [26], Hoffmann and Nickl [21], Bull and Nickl [9] where this relation is made clear. Despite the fact that there exists a quite complete and satisfying theory for adaptive non-parametric estimation – see, for example, Lepski [29], Donoho *et al.* [13], Barron *et al.* [3], Tsybakov [38] – the theory of adaptive confidence sets has some fundamental limitations. Indeed, one has to remove critical regions from the parameter space in order to construct honest adaptive confidence sets, see Low [30], Cai and Low [10], Hoffmann and Nickl [21], Bull and Nickl [9]. In the paper Bull and Nickl [9], the problem of L_2-adaptive and honest confidence sets is considered and in the course of the proofs, the authors establish that in the testing problem (1.1), ρ_n can be taken of the order

$$\rho_n \geq D \max\big(n^{-t/(2t+1/2)}, n^{-s/(2s+1)} \big),$$

for D large enough depending on the level of the test and on s, t. On the other hand, they prove in the case of density estimation (we provide a proof of this fact in our setting, see Theorem 3.2 below) that the lower bound for ρ_n is

$$\rho_n \geq D' n^{-t/(2t+1/2)},$$

for some D' positive; otherwise there exists no consistent test for the problem (1.1). In the case $s < 2t$, the upper and lower bound do not match (which in the context of confidence sets is unimportant, see Baraud [1], Cai and Low [11], Robins and Van Der Vaart [36], Bull and Nickl [9] related results).

From the point of view of hypothesis testing, the case $s < 2t$ is in fact of particular interest, as it implicitly addresses the question whether the complexity of the null hypothesis should influence the separation rate in non-parametric composite testing problems. When $s \geq 2t$, the rate of estimation in the null hypothesis is of order of the separation rate, and a reduction to a singleton null hypothesis is (intuitively) always possible as shown by the infimum test considered in Bull and Nickl [9]. For $s < 2t$, new ideas seem to be required.

To the best of our knowledge, the classical literature on non-parametric hypothesis testing does not answer this question. A majority of papers consider the case of a *singleton, or a parametric (finite dimensional) null hypothesis*, see Ingster [23], Ingster and Suslina [25], Spokoiny [37], Lepski and Spokoiny [28], Horowitz and Spokoiny [22], Pouet [34], Fromont and Laurent [16]. In this case, the null hypothesis is reducible to a finite union of singletons. The papers that do *not* consider the case of a simple null hypothesis, such as Dümbgen and Spokoiny [14], Juditsky and Nemirovski [27], Baraud *et al.* [2], consider settings where it is provable that the separation rate ρ_n must be of the same order as the estimation rate in the alternative hypothesis ($\rho_n \simeq n^{-t/(2t+1)}$ up to some $\log(n)$ factor). In particular the gap between estimation and testing rate from which the problem studied in the present paper arises does not exist, and plug-in tests that are based on the distance between an estimate of the function and the null hypothesis, are optimal in these cases. Blanchard *et al.* [8] consider a general multiple testing problem where they test a continuum of null hypotheses. As in Bull and Nickl [9], their separation rate depends on the complexity of the null hypothesis. The papers [17] and [32] consider a composite a non-parametric testing problem, and an approach based on an infimum test. For the same reason as in the paper [9], the complexity of the null hypothesis affects the separation rates they obtain. Finally the papers [4,15] consider directly the problem of smoothness testing (or smoothness estimation for [15]). However, their perspective is different and the assumptions they make are very restrictive (for instance, piecewise smoothness, see [15]).

In this paper, we demonstrate that the complexity of the null hypothesis does not influence the separation rate at least in the testing problem (1.1). More precisely, we prove that it is possible to build a test that is uniformly consistent with a separation rate

$$\rho_n \simeq n^{-t/(2t+1/2)}.$$

The test we propose uses the geometric structure of the Sobolev-type balls combined with a simple multiple testing idea, and is straightforward to implement. Our proofs rely on the specific structure of this problem, and in general whether or not the complexity of H_0 influences the separation rate depends heavily on the problem at hand.

Section 2 formalises the setting and notations that we consider. Section 3 provides the test and the main Theorems. Proofs are given in Sections 4 and 5.

2. Setting

Denote by $L_2([0, 1]) = L_2$ the space of functions defined on $[0, 1]$ such that $\|f\|_2^2 = \int_0^1 |f(x)|^2 \, dx < +\infty$, where $\|\cdot\|_2$ is the usual L_2 norm. For any functions $(f, g) \in L_2$, we consider the usual scalar product $\langle f, g \rangle = \int_0^1 f(x)g(x) \, dx$.

2.1. Wavelet basis

Let $S \geq 0$. We consider the Cohen–Daubechies–Vial wavelet basis on $[0, 1]$ with S first null moments (see Cohen *et al.* [12]), that we write

$$\{\phi_k, k \in Z_{J_0}, \psi_{l,k}, l > J_0, l \in \mathbb{N}, k \in Z_l\},$$

where $J_0 \equiv J_0(S) \in \mathbb{N}^*$ is a constant that grows with S (see Cohen *et al.* [12]), where $\forall l \geq J_0, Z_l \subset \mathbb{Z}$, and where $\forall k' \in Z_{J_0}, \forall l > J_0, \forall k \in Z_l, \phi_{k'}$ and $\psi_{l,k}$ are functions from $[0, 1]$ to $\mathbb{R}$.

The Cohen–Daubechies–Vial wavelet basis is an orthonormal basis of functions on $[0, 1]$. It is also such that

$$\forall l \geq J_0 + 1, \qquad |Z_l| = 2^l \quad \text{and} \quad z_0 \equiv Z_0(s) = |Z_{J_0}| < \infty,$$

where $\forall l \geq J_0$, $|Z_l|$ is the number of elements in the set Z_l. Note that the constant z_0 grows with S in the definition of the Cohen–Daubechies–Vial wavelet basis, and is such that $z_0 \geq 1$. We write $\forall k \in Z_{J_0}, \psi_{J_0,k} = \phi_k$ in order to simplify notations.

For any function $f \in L_2$, we consider the sequence $a \equiv a(f)$ of coefficients such that $\forall l \geq J_0, \forall k \in Z_l$,

$$a_{l,k} = \int_0^1 \psi_{l,k}(x) f(x) \, dx = \langle \psi_{l,k}, f \rangle.$$

The functions $f \in L_2$ have the representation

$$f = \sum_{l \geq J_0} \sum_{k \in Z_l} \psi_{l,k} \langle \psi_{l,k}, f \rangle = \sum_{l \geq J_0} \sum_{k \in Z_l} a_{l,k} \psi_{l,k}. \tag{2.1}$$

We moreover write for any $J \geq J_0$

$$\Pi_{V_J}(f) = \sum_{J_0 \leq l \leq J} \sum_{k \in Z_l} a_{l,k} \psi_{l,k}$$

the projection of f onto $V_J = \text{span}(\psi_{l,k}, J_0 \leq l \leq J, k \in Z_l)$ (where for any $A \subset L_2$, $\text{span}(A)$ is the vectorial sub-space generated by the functions in A). We also write

$$\Pi_{W_J}(f) = \sum_{k \in Z_J} a_{J,k} \psi_{J,k}$$

the projection of f onto $W_J = \text{span}(\psi_{J,k}, k \in Z_J)$.

2.2. Besov spaces

We consider, for $r > 0$, the $(r, 2, \infty)$-Besov (Nikolskii) norms

$$\|f\|_{r,2,\infty} = \sup_{l \geq J_0} \left(2^{lr} \left| \langle f, \psi_{l,\cdot} \rangle \right|_{l_2} \right),$$

where $|u|_{l_2} = (\sum_i u_i^2)^{1/2}$ is the sequential l_2 norm, and l_2 is the associated sequential space.

The associated $(r, 2, \infty)$-Besov (Nikolskii) spaces are defined as

$$B_{r,2,\infty} = \{ f \in L^2 : \|f\|_{r,2,\infty} < +\infty \}.$$

We write for a given $r > 0$ and a given $B > 0$ the $B_{r,2,\infty}$ Besov ball of smoothness r and radius B as

$$\Sigma(r, B) := \{ f \in B_{r,2,\infty} : \|f\|_{r,2,\infty} < B \}.$$

Since the wavelet basis we considered to build the $(r, 2, \infty)$-Besov spaces is the Cohen–Daubechies–Vial wavelets with S first null moments, the defined $(r, 2, \infty)$ Besov spaces correspond to the functional $(r, 2, \infty)$-Besov spaces (Sobolev-type spaces) for any $r \leq S$, see Meyer [31] and Härdle *et al.* [18].

Remark. We chose to consider the Cohen–Daubechies–Vial wavelet basis for simplicity and clarity in presentation, but any orthonormal wavelet basis that is such that (i) the number of wavelets $|Z_l|$ at each level l is bounded by a constant time 2^l and (ii) the basis can be used to characterize the functional $(r, 2, \infty)$-Besov spaces (Sobolev-type spaces), could have been used.

2.3. Observation scheme

Let $n > 0$. The data is a realisation of a Gaussian process defined for any $x \in [0, 1]$ as

$$dY^{(n)}(x) = f(x)\,dx + \frac{dB_x}{\sqrt{n}},$$

where $(B_x)_{x \in [0,1]}$ is a standard Brownian motion, and $f \in L_2$ is the function of interest.

Let us write for any $l \geq J_0$ and $k \in Z_l$ the associated wavelet coefficients as

$$\hat{a}_{l,k} = \langle \psi_{l,k}, dY^{(n)} \rangle = \int_0^1 \psi_{l,k}(x) f(x)\,dx + \frac{1}{\sqrt{n}} \int_0^1 \psi_{l,k}(x)\,dB_x, \quad \text{and} \quad a_{l,k} = \langle \psi_{l,k}, f \rangle,$$

where for any $g \in L_2$, $\int_0^1 g(x)\,dB_x$ is the usual stochastic integral, and is as such distributed as a Gaussian random variable of mean 0 and variance $\|g\|_2^2$. Since the Cohen–Daubechies–Vial wavelet basis is orthonormal, the coefficients $(\hat{a}_{l,k})_{l \geq J_0, k \in Z_l}$ are jointly Gaussian random variables such that

$$(\hat{a}_{l,k})_{l \geq J_0, k \in Z_l} \sim \mathcal{N}\left((a_{l,k})_{l \geq J_0, k \in Z_l}, \left(\frac{1}{n}\mathbf{1}\{l = l', k = k'\} \right)_{l \geq J_0, k \in Z_l, l' \geq J_0, k' \in Z_{l'}} \right),$$

where $\mathcal{N}(\mu, \sigma^2)$ is the normal distribution of mean μ and variance-covariance σ^2 (and where we write $X \sim \mathcal{N}(\mu, \sigma^2)$ for stating that X is such a Gaussian distribution) and where $\mathbf{1}\{\cdot\}$ is the usual indicator function.

We consider the wavelet estimate of f:

$$\hat{f}_n = \sum_{l \geq J_0} \sum_k \hat{a}_{l,k} \psi_{l,k}.$$

This estimate is of infinite variance in L_2, hence projected estimates

$$\hat{f}_n(j) := \Pi_{V_j} \hat{f}_n,$$

have to be considered.

In the sequel, we write Pr_f (respectively $\mathbb{E}_f$, and $\mathbb{V}_f$) the probability (respectively expectation, and variance) under the law of $dY^{(n)}$ when the function underlying the data is f. When no confusion is likely to arise, we write simply Pr (respectively, $\mathbb{E}$ and $\mathbb{V}$).

Remark. The spaces $B_{r,2,\infty}$ are slightly larger than the usual Sobolev spaces, see Bergh and Löfström [5] and Besov *et al.* [6]. They are however the natural objects to consider for a smoothness test, since they are the largest Besov spaces where adaptive estimation remains possible (see Donoho *et al.* [13] and Bull and Nickl [9]). Indeed, one can prove that there exists an estimate $\tilde{f}_n(Y^{(n)})$ of f such that for any $S \geq r > 1/2$ and $B > 0$, we have

$$\sup_{f \in \Sigma(r,B)} \mathbb{E}\|\tilde{f}_n - f\|_2 \leq \mathrm{O}\big(n^{-r/(2r+1)}\big),$$

see, for instance, Theorem 2 in the paper Bull and Nickl [9] (with some simple modifications needed for the regression situation considered in the present paper).

3. Testing problem

3.1. Formulation of the testing problem

Let $S \geq s > t > 0$ (we choose the Cohen–Daubechies–Vial wavelet basis with S first null moments with S larger than s). We want to test whether f is in $\Sigma(s, B)$, or whether f is outside this ball, i.e., in $\Sigma(t, B) \setminus \Sigma(s, B)$. This is generally impossible to do uniformly and functions that are t smooth but too close from s smooth functions (such that the L_2 distance between these functions and the Sobolev-type ball of smoothness s is small) have to be removed.

Let us first define the restriction of the sets $\Sigma(t, B)$ to sets that are separated away from $\Sigma(s, B)$ by some minimal distance $\rho_n > 0$:

$$\tilde{\Sigma}(t, B, \rho_n) = \big\{ f \in \Sigma(t, B) \colon \big\| f - \Sigma(s, B) \big\|_2 \geq \rho_n \big\},$$

where we remind that for any set $G \subset L_2$, we have $\|f - G\|_2 = \inf_{g \in G} \|f - g\|_2$.

The testing problem is the following

$$H_0 \colon f \in \Sigma(s, B) \quad \text{vs.} \quad H_1 \colon f \in \tilde{\Sigma}(t, B, \rho_n).$$

When no confusion is likely to arise, we will use the short-hand notation $f \in H_0$ for $f \in \Sigma(s, B)$, and $f \in H_1$ for $f \in \tilde{\Sigma}(t, B, \rho_n)$.

3.2. Main results

Let $j \geq J_0$ such that $j = \lfloor 1/(2t + 1/2) \log(n)/\log(2) \rfloor$, where $\lfloor \cdot \rfloor$ is the integer part of a real number. In particular, this definition implies that $n^{1/(2t+1/2)}/2 \leq 2^j \leq n^{1/(2t+1/2)}$.

Consider for any $J_0 < l \leq j$ the test statistics

$$T_n(l) = \|\Pi_{W_l} \hat{f}_n\|_2^2 - \frac{2^l}{n}, \quad \text{and} \quad T_n(J_0) = \|\Pi_{W_{J_0}} \hat{f}_n\|_2^2 - \frac{z_0}{n}. \tag{3.1}$$

These quantities $T_n(l)$ are estimates of $\|\Pi_{W_l}(f)\|_2^2$ across all levels $J_0 \leq l \leq j$. Concerning levels $l > j$, even in the worst case of smoothness t, the L_2 norm of the function at these levels is smaller than $n^{-t/(2t+1/2)}$, that is, $\|f - \Pi_{V_j}(f)\|_2 = O(n^{-t/(2t+1/2)})$. This implies that one does not need to control for what happens at these levels.

Let $\alpha > 0$ be the desired level of the test. Consider the positive constants $t_n(l)$ such that for any $J_0 \leq l \leq j$

$$t_n(l)^2 = \left(\frac{B}{2^{ls}} + \frac{\tau_l}{2}\right)^2 = \frac{B^2}{2^{2ls}} + \frac{B}{2^{ls}}\tau_l + \frac{\tau_l^2}{4}, \tag{3.2}$$

where the sequence $(\tau_l)_{J_0 \leq l \leq j}$ is such that for any $J_0 < l \leq j$

$$\tau_l \equiv \tau_{n,l} = 24\sqrt{\frac{z_0}{\alpha}}\frac{2^{(j+l)/8}}{\sqrt{n}}, \quad \text{and} \quad \tau_{J_0} \equiv \tau_{n,J_0} = 24\sqrt{\frac{z_0}{\alpha}}\frac{1}{\sqrt{n}}. \tag{3.3}$$

We consider the test:

$$\Psi_n(\alpha) = 1 - \prod_{J_0 \leq l \leq j} \mathbf{1}\{T_n(l) < t_n(l)^2\},$$

where we remind that $\mathbf{1}\{\cdot\}$ is the usual indicator function. We reject H_0 as soon as the test statistic at one of the levels $J_0 \leq l \leq j$ indicates a too large Besov norm. The intuition behind this test is that f belonging to $\Sigma(s, B)$ is equivalent to $\|\Pi_{V_l}(f)\|_{s,2,\infty}$ being smaller than or equal to B for any $l \geq J_0$. As explained before, we do not need to be too concerned by what happens for $l > j$. In the case $J_0 \leq l \leq j$, each statistic $T_n(l)$ is designed to test this. We illustrate this in Figure 1.

We provide the following definition of consistency for a test, following the line of work of Ingster and Suslina [25].

Definition 3.1 (α-consistency). *Let $\alpha > 0$ and H_0, H_1 be two hypotheses (functional sets). Let $\Psi_n(Y^{(n)}, H_0, H_1, \alpha)$ be a test, that is to say a measurable function taking values in $\{0, 1\}$. We say that Ψ_n is α-consistent if we have for any $n > 0$*

$$\sup_{f \in H_0} \mathbb{E}_f \Psi_n + \sup_{f \in H_1} \mathbb{E}_f(1 - \Psi_n) \leq \alpha.$$
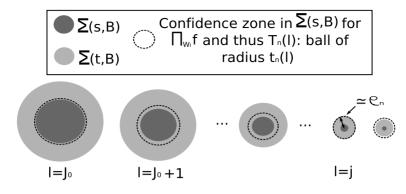
We now state the main result of this paper.

Figure 1. Illustration of the testing problem.

Theorem 3.1. *Let $\alpha > 0$. The test $\Psi_n(\alpha)$ is an α-consistent test for discriminating between H_0 and H_1 and for $\rho_n = \tilde{C}(\alpha)n^{-t/(2t+1/2)}$, where $\tilde{C}(\alpha) = 24(\frac{2^t B}{\sqrt{1-2^{-2t}}} + 19)\sqrt{\frac{1}{\alpha}}$.*

The proof of this theorem is in Section 4. The region we had to remove so that $\Psi_n(\alpha)$ is α-consistent could not have been taken significantly smaller, as stated in the next theorem.

Theorem 3.2. *Let $1 > \alpha \geq 0$. There exists no α-consistent test for discriminating between H_0 and H_1 and for $\rho_n = \tilde{D}(\alpha)n^{-t/(2t+1/2)}$, where $\tilde{D}(\alpha) = \min((\frac{1-\alpha}{2})^{1/4}, B)$.*

The proof of Theorem 3.2 is in Section 5. It is very similar to the proofs in papers Ingster [23] and Bull and Nickl [9] (the proof in paper Bull and Nickl [9] holds in the more involved case of density estimation).

We would like to emphasise that the test Ψ_n, in addition to being rather simple conceptually, is quite easy to implement since it requires only the computation of (significantly) less than n integrals/sums – the empirical coefficients – and less than $\log(n)$ sums of squares of these coefficients. It can replace the more complicated infimum test considered in the paper Bull and Nickl [9] for the creation of adaptive and honest confidence bands.

3.3. Alternative settings

We provided in the last subsection a consistent test on a model that could not have been taken significantly larger. This test was constructed in the rather simplistic setting of non-parametric Gaussian homoscedastic regression with normalised variance. But in many cases (see, e.g., Reiß [35] and Nussbaum [33]), it has been proven that it generalises rather well to more realistic and complex settings. The concern in our case, however, is that we heavily rely on the *homoscedasticity* assumption with known variance of the noise. Indeed, we subtract the constant part induced by this variance in the estimates of $T_n(l)$ in equation (3.1). This part is much larger than the de-
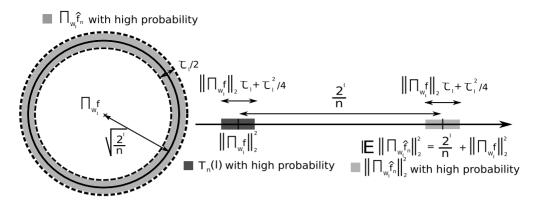
Figure 2. Statistics $T_n(l)$ and the removal of the expectation of the square of the expectation of the noise.

viations (in high probability) of $\|\Pi_{W_l} \hat{f}_n\|_2^2$ around its mean, and it is thus crucial to remove it. We illustrate this in Figure 2.

There is however a way around this problem that we discuss now, as well as generalizations to more complex settings.

Heteroscedastic non-parametric Gaussian regression. Assume now that the data are generated according to the process

$$dY^{(n)}(x) = f(x)\,dx + \frac{\sigma(x)\,dB_x}{\sqrt{n}},$$

where $(B_x)_{x \in [0,1]}$ is a standard Brownian motion, and $f, \sigma \in L_2$. Since the function σ is unknown, we cannot apply the technique we described. However, if we know a upper bound on $\|\sigma\|_2$, it is still possible to solve this problem with a very similar technique.

The modification goes as follows. We start by dividing the initial sample in two sub-samples of equal size $n/2$. Then we compute the empirical estimates of the function in these two samples and write $\hat{f}_n^{(1)}$ and $\hat{f}_n^{(2)}$ for the estimates of the function computed in each of the two halves. We then define the statistics $\hat{T}_n(l)$ (which play the same role as the $T_n(l)$) as

$$\hat{T}_n(l) = \langle \Pi_{W_l} \hat{f}_n^{(1)}, \Pi_{W_l} \hat{f}_n^{(2)} \rangle. \tag{3.4}$$

Since $\hat{f}_n^{(1)}$ and $\hat{f}_n^{(2)}$ are independent estimates of f, the additional term that comes from the expectation of the square of the noise (the variance) disappears and it is possible to prove that this newly defined $\hat{T}_n(l)$ concentrates around $\|\Pi_{W_l} f\|_2^2$ with an error of same order as in Lemma 4.2 below. This implies that we can test in a similar way and derive similar results.

Regression, density estimation and autoregressive model. The settings of non-parametric regression (with noise that can be non-Gaussian), of non-parametric density estimation, and of non-parametric auto-regressive model (AR(1)) are not too different from the heteroscedastic setting under a given set of assumptions (that, e.g., the noise on the data is sub-Gaussian and that the design is adapted for regression, and that, e.g., the regression function/density is bounded,

see Bull and Nickl [9]. This follows from the asymptotic equivalence between these models and non-parametric Gaussian regression (again, see, e.g., Reiß [35] and Nussbaum [33]).

- In the regression setting, we assume that the n data $(X_i, Y_i)_{i \leq n}$ are

$$Y_i = f(X_i) + \sigma(X_i)\varepsilon_i,$$

where ε_i are independent random variables of mean 0 and variance 1. Based on these data, we can compute also estimates for the wavelet coefficients of f as

$$\hat{a}_{l,k} = \frac{1}{n}\sum_{i=1}^{n} Y_i \psi_{l,k}(X_i),$$

and thus estimate f. Then we can follow the procedure described in the setting of heteroscedastic non-parametric Gaussian regression (equation (3.4)). However, one needs to be careful in this setting since the design (i.e., position of the X_i) is crucial. Indeed, wavelets are very localised functions and estimating the wavelet coefficients in a reasonably accurate way requires that the points X_i are spread over the whole domain, that is to say that there are enough points in each region of the domain. In particular, a standard random design will fail in this case, see Härdle *et al.* [18].
- In the density estimation setting, we assume that the n data generated by f are $(X_i)_i$, and estimate the wavelet coefficients of f as

$$\hat{a}_{l,k} = \frac{1}{n}\sum_{i=1}^{n} \psi_{l,k}(X_i),$$

and thus estimate f. Then we can follow the procedure described in the setting of heteroscedastic non-parametric Gaussian regression (equation (3.4)).
- We consider finally the non-parametric autoregressive model with memory 1 (or AR(1)). The output $(X_i)_{i \leq n}$ of an AR(1) can be described as follows:

$$X_{i+1} = f(X_i) + \sigma(X_i)\varepsilon_i.$$

After sub-sampling the data at random in order to make them close to independent, one can go back to the regression setting, and apply the same method (see, e.g., Hoffmann [19] for equivalence of this setting and regression setting after sub-sampling).

4. Proof of Theorem 3.1

This section contains a proof of Theorem 3.1.

4.1. Decomposition of the problem

The statistics $T_n(l)$ are unbiased estimates of $\|\Pi_{W_l}(f)\|_2^2$ for any $J_0 \leq l \leq j$, as explained later in this section. Assuming this, the next lemma explains why the test Ψ_n that we described is a reasonable thing to do.

Lemma 4.1. *Let* $(\tau_l)_{J_0 \leq l \leq j}$ *be a sequence of positive real numbers. Assume that*

$$\rho_n \geq \left(4 \frac{B}{\sqrt{1 - 2^{-2t}}} 2^{-jt} + \frac{4}{3} \sum_{J_0 \leq l \leq j} \tau_l \right).$$

Then we have

- $f \in H_0 \Rightarrow \max_{J_0 \leq l \leq j} (\|\Pi_{W_l}(f)\|_2 - \frac{B}{2^{ls}}) \leq 0.$
- $f \in H_1 \Rightarrow \max_{J_0 \leq l \leq j} (\|\Pi_{W_l}(f)\|_2 - \frac{B}{2^{ls}} - \tau_l) > 0.$

Proof. *Under the null Hypothesis H_0.* If f is in $\Sigma(s, B)$, then by definition of the Besov spaces

$$\|\Pi_{V_j} f\|_{s,2,\infty} \leq B,$$

which implies by definition of the $\| \cdot \|_{0,2,\infty}$ norm that

$$\sup_{J_0 \leq l \leq j} \left(\|\Pi_{W_l} f\|_{0,2,\infty} - \frac{B}{2^{ls}} \right) \leq 0.$$

This implies by Parseval's identity, and since $\|\Pi_{W_l} f\|_{0,2,\infty} = \|\Pi_{W_l} f\|_2$

$$\sup_{J_0 \leq l \leq j} \left(\|\Pi_{W_l} f\|_2 - \frac{B}{2^{ls}} \right) = \sup_{J_0 \leq l \leq j} \left(\|\Pi_{W_l} f\|_{0,2,\infty} - \frac{B}{2^{ls}} \right) \leq 0.$$

Under the alternative Hypothesis H_1. Assume that f is in $\tilde{\Sigma}(t, B, \rho_n)$. By triangular inequality, we have

$$\inf_{g \in \Sigma(s,B)} \|f - g\|_2 \leq \inf_{g \in \Sigma(s,B)} \left\| \Pi_{V_j}(f) - g \right\|_2 + \left\| f - \Pi_{V_j}(f) \right\|_2$$

$$\leq \inf_{g \in \Sigma(s,B)} \left\| \Pi_{V_j}(f) - g \right\|_2 + \frac{B}{\sqrt{1 - 2^{-2t}}} 2^{-jt},$$

since by definition of the $(t, 2, \infty)$ Besov space, we know that

$$\left\| f - \Pi_{V_j}(f) \right\|_2 \leq \sqrt{\sum_{l=j+1}^{\infty} 2^{-2lt} B^2} \leq \frac{B}{\sqrt{1 - 2^{-2t}}} 2^{-jt}.$$

We thus have, since $\rho_n \leq \inf_{g \in \Sigma(s,B)} \|f - g\|_2$ by definition of $\tilde{\Sigma}(t, B, \rho_n)$, and since $\rho_n \geq (4 \frac{B}{\sqrt{1-2^{-2t}}} 2^{-jt} + 4/3 \sum_{J_0 \leq l \leq j} \tau_l)$

$$3\rho_n/4 \leq \rho_n - \frac{B}{\sqrt{1 - 2^{-2t}}} 2^{-jt} \leq \inf_{g \in \Sigma(s,B)} \left\| \Pi_{V_j}(f) - g \right\|_2. \tag{4.1}$$

Let us write $(a_{l,k})_{l,k}$ the coefficients of f and $(b_{l,k})_{l,k}$ the coefficients of the minimiser g. We have by definition of $\Sigma(s, B)$, by the triangular inequality and by Parseval's identity

$$\inf_{g \in \Sigma(s,B)} \left\| \Pi_{V_j}(f) - g \right\|_2$$

$$\leq \inf_{g \in \Sigma(s,B)} \sum_{l=J_0}^{j} \left\| \Pi_{W_l}(f) - g \right\|_2$$

$$= \inf_{(b_{l,k})_{l,k} : \forall l \geq J_0, 2^{ls} \|b_{l,\cdot}\|_{l_2} \leq B} \left(\sum_{l=J_0}^{j} \sqrt{\sum_{k \in Z_l} (a_{l,k} - b_{l,k})^2} + \sum_{l=j+1}^{\infty} \sqrt{\sum_{k \in Z_l} b_{l,k}^2} \right)$$

$$= \sum_{l=J_0}^{j} \inf_{(b_{l,k})_{l,k} : 2^{ls} \|b_{l,\cdot}\|_{l_2} \leq B} \sqrt{\sum_{k \in Z_l} (a_{l,k} - b_{l,k})^2},$$

since the constraints defining the minimisation problems involved do not interact across the levels l. The last equation, together with equation (4.1), implies that

$$3\rho_n/4 \leq \sum_{l=J_0}^{j} \inf_{(b_{l,k})_{l,k} : 2^{ls} \|b_{l,\cdot}\|_{l_2} \leq B} \sqrt{\sum_{k \in Z_l} (a_{l,k} - b_{l,k})^2}.$$

By definition, $\rho_n \geq 4/3 \sum_{l=J_0}^{j} \tau_l$, so the last equation implies that

$$\sum_{l=J_0}^{j} \tau_l \leq \sum_{l=J_0}^{j} \inf_{(b_{l,k})_{l,k} : 2^{ls} \|b_{l,\cdot}\|_{l_2} \leq B} \sqrt{\sum_{k \in Z_l} (a_{l,k} - b_{l,k})^2}. \qquad (4.2)$$

At least one of the τ_l's has to be less than or equal to

$$\inf_{(b_{l,k})_{l,k} : 2^{ls} \|b_{l,\cdot}\|_{l_2} \leq B} \sqrt{\sum_{k \in Z_l} (a_{l,k} - b_{l,k})^2},$$

as otherwise $\sum_{l=J_0}^{j} \tau_l$ would exceed the right-hand side in equation (4.2). Let $J_0 \leq l \leq j$ be one of these indexes, we have

$$\tau_l \leq \inf_{(b_{l,k})_{l,k} : 2^{ls} \|b_{l,\cdot}\|_{l_2} \leq B} \sqrt{\sum_{k \in Z_l} (a_{l,k} - b_{l,k})^2}$$

$$\leq \max \left(0, \sqrt{\sum_{k \in Z_l} a_{l,k}^2} - \frac{B}{2^{ls}} \right)$$

$$\leq \left\| \Pi_{W_l}(f) \right\|_2 - \frac{B}{2^{ls}}$$

since by definition of the Euclidian ball, for any $u \in l_2$, we have $\inf_{v \in l_2: \|v\|_{l_2}=1} \|u - v\|_{l_2} = \max(0, \|u\|_{l_2} - 1)$.

This concludes the proof. $\qquad\square$

4.2. Convergence tools for $T_n(l)$

The next lemma is a standard and also rather weak concentration inequality (see, e.g., Birgé [7] for similar results).

Lemma 4.2. *Let* $\Delta > 0$. *Then*

$$\Pr\left\{\forall l : J_0 \le l \le j, \left|T_n(l) - \|\Pi_{W_l} f\|_2^2\right| \ge 4\sqrt{\frac{3z_0}{\Delta}\left(\frac{2^{(j+l)/2}}{n^2} + 2^{l/4}\frac{\|\Pi_{W_l} f\|_2^2}{n}\right)}\right\} \le \Delta.$$

Proof. Let $J_0 < l \le j$. Note first that by Parseval's identity, we have $\|\Pi_{W_l}\hat{f}_n\|_2^2 = \sum_k \hat{a}_{l,k}^2$. Then we have by definition $T_n(l) = \sum_k \hat{a}_{l,k}^2 - \frac{2^l}{n}$.

We have $\hat{a}_{l,k} = a_{l,k} + \hat{a}_{l,k} - a_{l,k}$ where $\hat{a}_{l,k} - a_{l,k} \sim \mathcal{N}(0, 1/n)$ (by assumption of the Gaussian model), and thus we have

$$\mathbb{E}|\hat{a}_{l,k}|^2 = \frac{1}{n} + a_{l,k}^2.$$

Also since for any constant $m \in \mathbb{R}$, and for $G \sim \mathcal{N}(0, 1)$,

$$\mathbb{V}(G + m)^2 = \mathbb{E}(G^2 + 2Gm - 1)^2 = \mathbb{E}(G^4 + 4G^2m^2 + 1 - 2G^2) = 4m^2 + 2 \le 4(1 + m^2),$$

we have

$$\mathbb{V}|\hat{a}_{l,k}|^2 \le 4\left(\frac{1}{n^2} + \frac{a_{l,k}^2}{n}\right).$$

This implies since the $\hat{a}_{l,k}$ are independent Gaussian random variables

$$\mathbb{E}\left(\sum_{k \in Z_l} \hat{a}_{l,k}^2\right) = \sum_{k \in Z_l} a_{l,k}^2 + \frac{2^l}{n},$$

and

$$\mathbb{V}\left(\sum_{k \in Z_l} \hat{a}_{l,k}^2\right) \le 4\left(\frac{2^l}{n^2} + \frac{\sum_{k \in Z_l} a_{l,k}^2}{n}\right).$$

This implies by Chebyshev's inequality that for any $\delta_l > 0$, we have

$$\Pr\left\{\left|\sum_{k \in Z_l} \hat{a}_{l,k}^2 - \frac{2^l}{n} - \sum_{k \in Z_l} a_{l,k}^2\right| \ge \sqrt{\frac{1}{\delta_l}4\left(\frac{2^l}{n^2} + \frac{\sum_{k \in Z_l} a_{l,k}^2}{n}\right)}\right\} \le \delta_l$$

and since $\|\Pi_{W_l}\hat{f}_n\|_2^2 = \sum_{k \in Z_l} \hat{a}_{l,k}^2$ and $\|\Pi_{W_l}f\|_2^2 = \sum_{k \in Z_l} a_{l,k}^2$ that

$$\Pr\left\{\left|\|\Pi_{W_l}\hat{f}_n\|_2^2 - \frac{2^l}{n} - \|\Pi_{W_l}f\|_2^2\right| \geq \sqrt{\frac{1}{\delta_l}4\left(\frac{2^l}{n^2} + \frac{\|\Pi_{W_l}f\|_2^2}{n}\right)}\right\} \leq \delta_l.$$

In the same way (since there are z_0 terms in Z_{J_0}), we have for $l = J_0$, that for any $\delta_{J_0} > 0$

$$\Pr\left\{\left|\|\Pi_{W_{J_0}}\hat{f}_n\|_2^2 - \frac{z_0}{n} - \|\Pi_{W_{J_0}}f\|_2^2\right| \geq \sqrt{\frac{1}{\delta_{J_0}}4\left(\frac{z_0}{n^2} + \frac{\|\Pi_{W_{J_0}}f\|_2^2}{n}\right)}\right\} \leq \delta_{J_0}.$$

These two last results imply by definition of $T_n(l)$, that for any $J_0 \leq l \leq j$

$$\Pr\left\{\left|T_n(l) - \|\Pi_{W_l}f\|_2^2\right| \geq \sqrt{\frac{1}{\delta_l}4\left(\frac{2^l}{n^2} + \frac{\|\Pi_{W_l}f\|_2^2}{n}\right)}\right\} \leq \delta_l,$$

and

$$\Pr\left\{\left|T_n(J_0) - \|\Pi_{W_{J_0}}f\|_2^2\right| \geq \sqrt{\frac{1}{\delta_{J_0}}4\left(\frac{z_0}{n^2} + \frac{\|\Pi_{W_{J_0}}f\|_2^2}{n}\right)}\right\} \leq \delta_{J_0}.$$

These results imply by an union bound over all $J_0 \leq l \leq j$, that we have

$$\Pr\left\{\forall l : J_0 < l \leq j, \left|T_n(l) - \|\Pi_{W_l}f\|_2^2\right| \geq \sqrt{\frac{1}{\delta_l}4\left(\frac{2^l}{n^2} + \frac{\|\Pi_{W_l}f\|_2^2}{n}\right)},\right.$$
$$\left.\left|T_n(J_0) - \|\Pi_{W_{J_0}}f\|_2^2\right| \geq \sqrt{\frac{1}{\delta_{J_0}}4\left(\frac{z_0}{n^2} + \frac{\|\Pi_{W_{J_0}}f\|_2^2}{n}\right)}\right\} \leq \sum_{J_0 \leq l \leq j} \delta_l.$$

Set for any $J_0 < l \leq j$, $\delta_l = (2^{-(j-l)/2} + 2^{-l/4})\Delta/12$, and $\delta_{J_0} = \Delta/12$. Then

$$\Pr\left\{\forall l : J_0 < l \leq j, \left|T_n(l) - \|\Pi_{W_l}f\|_2^2\right| \geq 4\sqrt{\frac{3}{\Delta}\left(\frac{2^{(j+l)/2}}{n^2} + 2^{l/4}\frac{\|\Pi_{W_l}f\|_2^2}{n}\right)},\right.$$
$$\left.\left|T_n(J_0) - \|\Pi_{W_{J_0}}f\|_2^2\right| \geq 4\sqrt{\frac{3}{\Delta}\left(\frac{z_0}{n^2} + \frac{\|\Pi_{W_{J_0}}f\|_2^2}{n}\right)}\right\} \leq \sum_{J_0 \leq l \leq j} \delta_l \leq \Delta,$$

since

$$\sum_{J_0 \leq l \leq j} \delta_l \leq \frac{\Delta}{12} + \frac{\Delta}{12}\sum_{1 \leq l \leq j}(2^{-(j-l)/2} + 2^{-l/4}) \leq \frac{\Delta}{12}\left(1 + \frac{1}{1 - 2^{-1/2}} + \frac{1}{1 - 2^{-1/4}}\right) \leq \Delta.$$

Since $z_0 \geq 1$, we have

$$\Pr\left\{\forall l : J_0 \leq l \leq j, \left|T_n(l) - \|\Pi_{W_l} f\|_2^2\right| \geq 4\sqrt{\frac{3z_0}{\Delta}\left(\frac{2^{(j+l)/2}}{n^2} + 2^{l/4}\frac{\|\Pi_{W_l} f\|_2^2}{n}\right)}\right\} \leq \Delta,$$

which concludes the proof. $\qquad\square$

4.3. Study of the test

Set $c \equiv c(\alpha) = 24\sqrt{\frac{z_0}{\alpha}}$, where we remind that $\alpha > 0$ is the desired level of the test. By definition of the quantities τ_l (equation (3.3)), we have for any $J_0 < l \leq j$

$$\tau_l \equiv \tau_{n,l} = c\frac{2^{(j+l)/8}}{\sqrt{n}}, \quad \text{and} \quad \tau_{J_0} \equiv \tau_{n,J_0} = c\frac{1}{\sqrt{n}}.$$

We thus have

$$\sum_{l=J_0}^{j} \tau_l \leq \sum_{l=0}^{j} c\frac{2^{(j+l)/8}}{\sqrt{n}} \leq \frac{c}{\sqrt{n}}2^{j/4}\left(1 + \frac{1}{1 - 2^{-1/8}}\right) \leq 14cn^{-t/(2t+1/2)}. \tag{4.3}$$

Also, by definition of $\tilde{C}(\alpha)$ in Theorem 3.1, we have

$$\rho_n = c\left(\frac{2^t B}{\sqrt{1 - 2^{-2t}}} + 19\right)n^{-t/(2t+1/2)}.$$

In particular this implies together with equation (4.3), and since $2^j \leq 2^t n^{t/(2t+1/2)}$, that

$$\rho_n \geq c\frac{B}{\sqrt{1 - 2^{-2t}}}2^{-jt} + \frac{4}{3}\sum_{J_0 \leq l \leq j} \tau_l. \tag{4.4}$$

4.3.1. *Null hypothesis*

Since $f \in \Sigma(s, B)$, by Lemma 4.1,

$$\max_{J_0 \leq l \leq j}\left(\|\Pi_{W_l} f\|_2 - \frac{B}{2^{ls}}\right) \leq 0.$$

Thus by Lemma 4.2, we have with probability at least $1 - \alpha/2$ that for any $J_0 \leq l \leq j$

$$T_n(l) \leq \|\Pi_{W_l} f\|_2^2 + 4\sqrt{\frac{6z_0}{\alpha}\left(\frac{2^{(l+j)/2}}{n^2} + 2^{l/4}\frac{\|\Pi_{W_l} f\|_2^2}{n}\right)}$$

$$\leq \frac{B}{2^{ls}}\left(\frac{B}{2^{ls}} + 4 \times 2^{l/4}\sqrt{\frac{6z_0}{\alpha n}}\right) + 4\sqrt{\frac{6z_0}{\alpha}}\frac{2^{(j+l)/4}}{n}$$

$$\leq \frac{B^2}{2^{2ls}} + 4\frac{B}{2^{ls}}\sqrt{\frac{6z_0}{\alpha}}\frac{2^{(j+l)/8}}{n^{1/2}} + 4\sqrt{\frac{6z_0}{\alpha}}\frac{2^{(j+l)/4}}{n}$$

$$\leq \left(\frac{B}{2^{ls}} + 4\sqrt{\frac{6z_0}{\alpha}}\frac{2^{(j+l)/8}}{n^{1/2}}\right)^2$$

$$\leq \left(\frac{B}{2^{ls}} + \tau_l/\sqrt{6}\right)^2 < t_n(l)^2,$$

since $c = 24\sqrt{\frac{z_0}{\alpha}}$, and by definition of $t_n(l)$ (see equation (3.2)).

So with probability at least $1 - \alpha/2$, we have $\Psi_n = 0$ under H_0.

4.3.2. *Alternative hypothesis*

The sequence $(\tau_l)_l$, and ρ_n verify the assumptions of Lemma 4.1 (see equation (4.4)).

If H_1 is verified, then

$$\max_{J_0 \leq l \leq j}\left(\|\Pi_{W_l}(f)\|_2 - \frac{B}{2^{ls}} - \tau_l\right) > 0,$$

see Lemma 4.1. So there exists $J_0 \leq l \leq j$ such that

$$\|\Pi_{W_l}(f)\|_2 \geq \frac{B}{2^{ls}} + \tau_l.$$

By Lemma 4.2, we have with probability at least $1 - \alpha/2$ that for this l

$$T_n(l) \geq \|\Pi_{W_l}f\|_2^2 - 4\sqrt{\frac{6z_0}{\alpha}\left(\frac{2^{(j+l)/2}}{n^2} + 2^{l/4}\frac{\|\Pi_{W_l}f\|_2^2}{n}\right)}$$

$$\geq \left(\frac{B}{2^{ls}} + \tau_l\right)\left(\frac{B}{2^{ls}} + \tau_l - 4\sqrt{\frac{6z_0}{\alpha}}\frac{2^{l/4}}{n^{1/2}}\right) - 4\sqrt{\frac{6z_0}{\alpha}}\frac{2^{(j+l)/2}}{n^2}$$

$$\geq \left(\frac{B}{2^{ls}} + \tau_l\right)\left(\frac{B}{2^{ls}} + \tau_l/2\right) - 4\sqrt{\frac{6z_0}{\alpha}}\frac{2^{(j+l)/4}}{n}$$

$$\geq \frac{B^2}{2^{2ls}} + \frac{B}{2^{ls}}\tau_l + \tau_l^2/2 - 4\sqrt{\frac{6z_0}{\alpha}}\frac{2^{(j+l)/4}}{n}$$

$$\geq \frac{B^2}{2^{2ls}} + \frac{B}{2^{ls}}\tau_l + \tau_l^2/4$$

$$\geq \left(\frac{B}{2^{ls}} + \tau_l/2\right)^2 = t_n(l)^2.$$

since $c = 24\sqrt{\frac{z_0}{\alpha}}$, and by definition of $t_n(l)$ (see equation (3.2)).

So with probability at least $1 - \alpha/2$, we have $\Psi_n = 1$ under H_1.

Conclusion on the test Ψ_n. All the inequalities developed earlier are true for any f in H_0 or H_1 with constants depending only on s, t, B, α and the supremum over f in H_0 and H_1 of the error of type one and two are bounded by $\alpha/2$. Finally, the test Ψ_n of errors of type 1 and 2 bounded by $\alpha/2$ distinguishes between H_0 and H_1 with condition $\rho_n = 24\sqrt{\frac{z_0}{\alpha}}(\frac{2^t B}{\sqrt{1-2^{-2t}}} + 19)n^{-t/(2t+1/2)}$. This implies that

$$\sup_{f \in \Sigma(s,B)} \mathbb{E}_f \Psi_n + \sup_{f \in \tilde{\Sigma}(t,B,\rho_n)} \mathbb{E}_f(1 - \Psi_n) \leq \alpha.$$

5. Proof of Theorem 3.2

Let $B > 0$, $s > t > 0$, $\min(1, B) > \upsilon > 0$, and $j \in \mathbb{N}^*$ such that $j = \lfloor 1/(2t + 1/2) \log(n)/\log(2) \rfloor$, where $\lfloor \cdot \rfloor$ is the integer part of a real number. In particular, this definition implies that $n^{1/(2t+1/2)}/2 \leq 2^j \leq n^{1/(2t+1/2)}$.

Step 1: *Definition of a testing problem on some large set*. Define the set

$$I \equiv I_j = \big\{(\alpha_{l,k})_{l \geq J_0, k \in Z_l} : \forall l \neq j, \alpha_{l,k} = 0, \alpha_{j,k} \in \{-1, 1\}\big\}.$$

Consider the sequence of coefficients indexed by a given $\alpha \in I$ as

$$a_{l,k}^{(\alpha)} = \upsilon a \alpha_{l,k},$$

where $a = \frac{1}{\sqrt{n}2^{j/4}}$. Consider the function associated to $a^{(\alpha)}$ that we write $f^{(\alpha)}$ and that we define as

$$f^{(\alpha)} = \sum_{l=J_0}^{\infty} \sum_{k \in Z_l} a_{l,k}^{(\alpha)} \psi_{l,k} = \sum_{k \in Z_j} a_{j,k}^{(\alpha)} \psi_{j,k}.$$

Consider the testing problem

$$H_0: f = 0 \quad \text{vs.} \quad H_1: f = f^{(\alpha)}, \qquad \alpha \in I. \tag{5.1}$$

Step 2: *Quantity of interest*. An observation in the white noise model is equivalent, by sufficiency considerations, to an observation of empirical coefficients: equivalently to having access to the process $Y^{(n)}$, we have access to the empirical coefficients $(\hat{a}_{l,k})_{l,k}$ (where $\hat{a}_{l,k} = \int \psi_{l,k} dY^{(n)}$) and each of these coefficients are independent $\mathcal{N}(a_{l,k}, 1/n)$. Let Ψ be a test, i.e., some measurable function (according to the empirical coefficients) taking values in $\{0, 1\}$.

We have for any $\eta > 0$ (using the notations Pr_0 and $\mathbb{E}_0$ for the probability and expectation when the data are generated with $f = 0$)

$$\mathbb{E}_0[\Psi] + \sup_{f^{(\alpha)}, \alpha \in I} \mathbb{E}_{f^{(\alpha)}}[1 - \Psi] \geq \mathbb{E}_0[\Psi] + \frac{1}{|I|} \sum_{\alpha \in I} \mathbb{E}_{f^{(\alpha)}}[1 - \Psi]$$

$$\geq \mathbb{E}_0\big[\mathbf{1}\{\Psi = 1\}\big] + \mathbf{1}\{\Psi = 0\}Z \tag{5.2}$$

$$\geq (1 - \eta)\text{Pr}_0(Z \geq 1 - \eta),$$

where $Z = \frac{1}{|I|} \sum_{\alpha \in I} \prod_{l,k} \frac{dP_{l,k}^{(\alpha)}}{dP_{l,k}^0}$, where $dP_{l,k}^{(\alpha)}$ is the density of $\hat{a}_{l,k}$ when the function generating the data is $f^{(\alpha)}$, and $dP_{l,k}^0$ is the density of $\hat{a}_{l,k}$ when the function generating the data is 0 (this holds since the $(\hat{a}_{l,k})_{l,k}$ are independent).

More precisely, we have since the $(\hat{a}_{l,k})_{l,k}$ are independent $\mathcal{N}(a_{l,k}, 1/n)$

$$
Z\big((x_k)_k\big) \equiv Z\big((x_{l,k})_{l,k}\big) = \frac{1}{|I|} \sum_{\alpha \in I} \prod_{l,k} \frac{\exp((-n/2)(x_{l,k} - a_{l,k}^{(\alpha)})^2)}{\exp((-n/2)x_{l,k}^2)}
$$

$$
= \frac{1}{|I|} \sum_{\alpha \in I} \prod_{k \in Z_j} \exp(nx_k a_k^{(\alpha)}) \exp\left(-\frac{n}{2}(a_k^{(\alpha)})^2\right),
$$

where $(x_k)_k \equiv (x_{j,k})_k$ and $(a_k^{(\alpha)})_k \equiv (a_k^{(\alpha)})_k$. In the rest of the proof, we write also $(\alpha_k)_k \equiv (\alpha_{j,k})_k$ in order to simplify notations.

By Markov and Cauchy Schwarz's inequality

$$
\text{Pr}_0(Z \geq 1 - \eta) \geq 1 - \frac{\mathbb{E}_0|Z - 1|}{\eta} \geq 1 - \frac{\sqrt{\mathbb{E}_0(Z - 1)^2}}{\eta}. \tag{5.3}
$$

Step 3: *Study of the term in Z.* We have by definition of Z

$$
\mathbb{E}_0\big[(Z - 1)^2\big]
$$

$$
= \int_{x_1,\ldots,x_{2j}} \left(\frac{1}{|I|} \sum_{\alpha \in I} \prod_k \exp(x_k n a_k^{(\alpha)}) \exp\left(-\frac{n}{2}(a_k^{(\alpha)})^2\right) - 1\right)^2
$$

$$
\times \prod_k \frac{1}{\sqrt{2n\pi}} \exp\left(-\frac{n}{2}(x_k)^2\right) dx_1 \cdots x_{2j}
$$

$$
= \int_{x_1,\ldots,x_{2j}} \left(\frac{1}{|I|} \sum_{\alpha \in I} \prod_k \exp(x_k n a_k^{(\alpha)}) \exp\left(-\frac{n}{2}(a_k^{(\alpha)})^2\right)\right)^2
$$

$$
\times \prod_k \frac{1}{\sqrt{2n\pi}} \exp\left(-\frac{n}{2}(x_k)^2\right) dx_1 \cdots x_{2j}
$$

$$
- 2 \int_{x_1,\ldots,x_{2j}} \frac{1}{|I|} \sum_{\alpha \in I} \prod_k \exp(x_k n a_k^{(\alpha)}) \exp\left(-\frac{n}{2}(a_k^{(\alpha)})^2\right)
$$

$$
\times \prod_k \frac{1}{\sqrt{2n\pi}} \exp\left(-\frac{n}{2}(x_k)^2\right) dx_1 \cdots x_{2j} + 1
$$

$$
= \int_{x_1,\ldots,x_{2j}} \left(\frac{1}{|I|} \sum_{\alpha \in I} \prod_k \exp(x_k n a_k^{(\alpha)}) \exp\left(-\frac{n}{2}(a_k^{(\alpha)})^2\right)\right)^2
$$

$$\times \prod_k \frac{1}{\sqrt{2n\pi}} \exp\left(-\frac{n}{2}(x_k)^2\right) dx_1 \cdots x_{2^j}$$

$$- 2\frac{1}{|I|} \sum_{\alpha \in I} \prod_k \int_{x_k} \frac{1}{\sqrt{2n\pi}} \exp\left(-\frac{n}{2}(x_k - a_k^{(\alpha)})^2\right) dx_k + 1$$

$$= \int_{x_1,\ldots,x_{2^j}} \left(\frac{1}{|I|} \sum_{\alpha \in I} \prod_k \exp\left(x_k n a_k^{(\alpha)}\right) \exp\left(-\frac{n}{2}(a_k^{(\alpha)})^2\right)\right)^2$$

$$\times \prod_k \frac{1}{\sqrt{2n\pi}} \exp\left(-\frac{n}{2}(x_k)^2\right) dx_1 \cdots x_{2^j} - 1$$

by Fubini–Tonelli. This implies by developing the first term that

$$\mathbb{E}_0\left[(Z-1)^2\right]$$

$$= \frac{1}{|I|^2} \left(\sum_{\alpha,\alpha' \in I} \int_{x_1 \ldots, x_{2^j}} \prod_k \exp\left(x_k n(a_k^{(\alpha)} + a_k^{(\alpha')})\right) \exp\left(-\frac{n}{2}\left((a_k^{(\alpha)})^2 + (a_k^{(\alpha')})^2\right)\right)\right.$$

$$\left. \times \frac{1}{\sqrt{2n\pi}} \exp\left(-\frac{n}{2}(x_k)^2\right) dx_1 \cdots dx_{2^j}\right) - 1$$

$$= \frac{1}{|I|^2} \left(\sum_{\alpha,\alpha' \in I} \prod_k \int_{x_k} \exp\left(x_k n(a_k^{(\alpha)} + a_k^{(\alpha')})\right) \exp\left(-\frac{n}{2}\left((a_k^{(\alpha)})^2 + (a_k^{(\alpha')})^2\right)\right)\right.$$

$$\left. \times \frac{1}{\sqrt{2n\pi}} \exp\left(-\frac{n}{2}(x_k)^2\right) dx_k\right) - 1$$

$$= \frac{1}{|I|^2} \left(\sum_{\alpha,\alpha' \in I} \prod_k \int_{x_k} \exp\left(x_k n \upsilon a(\alpha_k + \alpha_k')\right) \exp\left(-n\upsilon^2 a^2\right) \frac{1}{\sqrt{2n\pi}} \exp\left(-\frac{n}{2}(x_k)^2\right) dx_k\right) - 1.$$

This implies by integrating depending on the respective values of α_k and α_k' that

$$\mathbb{E}_0\left[(Z-1)^2\right]$$

$$= \frac{1}{|I|^2} \left[\sum_{\alpha,\alpha' \in I} \prod_k \left(\exp\left(n\upsilon^2 a^2\right) \mathbf{1}\{\alpha_k = \alpha_k' = 1\} \int_{x_k} \frac{1}{\sqrt{2n\pi}} \exp\left(-\frac{n}{2}(x_k - 2\upsilon a)^2\right) dx_k\right.\right.$$

$$+ \exp\left(n\upsilon^2 a^2\right) \mathbf{1}\{\alpha_k = \alpha_k' = -1\}$$

$$\times \int_{x_k} \frac{1}{\sqrt{2n\pi}} \exp\left(-\frac{n}{2}(x_k + 2\upsilon a)^2\right) dx_k$$

$$+ \exp\left(-n\upsilon^2 a^2\right) \mathbf{1}\{\alpha_k \neq \alpha_k'\} \left.\left.\int_{x_k} \frac{1}{\sqrt{2n\pi}} \exp\left(-\frac{n}{2}x_k^2\right) dx_k\right)\right] - 1 \tag{5.4}$$

$$= \frac{1}{|I|^2} \left(\sum_{\alpha,\alpha' \in I} \prod_k \left(\exp\left(-n\upsilon^2 a^2\right)\left(1 - \mathbf{1}\{\alpha_k \neq \alpha_k'\}\right) + \exp\left(n\upsilon^2 a^2\right) \mathbf{1}\{\alpha_k \neq \alpha_k'\}\right)\right) - 1.$$

Since the α and α' take respectively all possible values in $\{-1,1\}^{2^j}$, by definition of the expectation, and by replacing α and α' by R and R' in the formula, we have

$$\frac{1}{|I|^2}\sum_{\alpha,\alpha'\in I}[\cdot]=\mathbb{E}_{(R_i)_i,(R'_j)_j}[\cdot],$$

where the $(R_i)_i$, $(R'_j)_j$ are two sequences of i.i.d. Rademacher random variables that are also independent of each other, and where $\mathbb{E}_{(R_i)_i,(R'_j)_j}[\cdot]$ is the expectation according to these random variables. This implies together with equation (5.4) that

$$\mathbb{E}_0\big[(Z-1)^2\big]$$

$$=\mathbb{E}_{(R_i)_i,(R'_j)_j}\Big[\prod_k\big(\exp(-n\upsilon^2a^2)\big(1-\mathbf{1}\{R_k\neq R'_k\}\big)+\exp(n\upsilon^2a^2)\mathbf{1}\{R_k\neq R'_k\}\big)\Big]-1$$

$$=\prod_k\mathbb{E}_{R_k,R'_k}\big[\exp(-n\upsilon^2a^2)\big(1-\mathbf{1}\{R_k\neq R'_k\}\big)+\exp(n\upsilon^2a^2)\mathbf{1}\{R_k\neq R'_k\}\big]-1,$$

since all R_k, R'_k are independent of each other. Moreover, $\mathbf{1}\{R_k\neq R'_k\}$ is a Bernoulli random variable of parameter $1/2$ (since the two Rademacher are independent), which implies

$$\mathbb{E}_0\big[(Z-1)^2\big]=\prod_k\mathbb{E}_B\big[\exp(-n\upsilon^2a^2)(1-B)+\exp(n\upsilon^2a^2)B\big]-1$$

$$=\big(\mathbb{E}_B\big[\exp(-n\upsilon^2a^2)(1-B)+\exp(n\upsilon^2a^2)B\big]\big)^{2^j}-1,$$

where $\mathbb{E}_B[\cdot]$ is the expectation according to a Bernoulli random variable with parameter $1/2$. The last equation implies

$$\mathbb{E}_0\big[(Z-1)^2\big]=\left(\frac{\exp(-n\upsilon^2a^2)+\exp(n\upsilon^2a^2)}{2}\right)^{2^j}-1$$

$$\leq\left(\frac{1-n\upsilon^2a^2+(n\upsilon^2a^2)^2+1+n\upsilon^2a^2+(n\upsilon^2a^2)^2}{2}\right)^{2^j}-1$$

$$\leq\big(1+(n\upsilon^2a^2)^2\big)^{2^j}-1,$$

since for any $|u|\leq 1$, we have $\exp(u)\leq 1+u+u^2$. Since $a^2=\frac{1}{n2^{j/2}}$, we have

$$\mathbb{E}_0\big[(Z-1)^2\big]\leq\left(1+\frac{\upsilon^4}{2^j}\right)^{2^j}-1$$

$$\leq\left(\exp\Big(\frac{\upsilon^4}{2^j}\Big)\right)^{2^j}-1=\exp(\upsilon^4)-1$$

$$\leq 1+2\upsilon^4-1=2\upsilon^4,$$

since for any $0 \leq u \leq 1$, we have $1 + u \leq \exp(u) \leq 1 + 2u$.

Step 4: *Conclusion on the testing problem (5.1).* By combining this with equations (5.2), (5.3), we know that for n large enough

$$\mathbb{E}_0[\Psi] + \sup_{f^{(\alpha)}, \alpha \in I} \mathbb{E}_{f^{(\alpha)}}[1 - \Psi] \geq 1 - 2\upsilon^4,$$

and since this holds with any Ψ, we have

$$\inf_{\Psi} \left[\mathbb{E}_0[\Psi] + \sup_{f^{(\alpha)}, \alpha \in I} \mathbb{E}_{f^{(\alpha)}}[1 - \Psi] \right] \geq 1 - 2\upsilon^4,$$

where $\inf_{\Psi}$ is the infimum over measurable tests Ψ. This implies that there is no $1 - 2\upsilon^4$ consistent test for test (5.1) (and it holds for any $0 \leq \upsilon < 1$).

Step 5: *Translation of this result in terms of the test (1.1).* Set

$$\rho_n = \frac{\upsilon n^{-t/(2t+1/2)}}{2}.$$

Since $\upsilon \leq B$,

$$\left\| f^{(\alpha)} \right\|_{t,2,\infty} = \sqrt{\sum_{k \in Z_l} (a_k^{(\alpha)})^2 2^{2jt}} = \upsilon \leq B,$$

so $f^{(\alpha)} \in \Sigma(t, B)$.

Also since $\forall \alpha \in I$, only the jth first coefficients of $f^{(\alpha)}$ are non-zero (i.e., $f^{(\alpha)} = \Pi_{W_j}(f^{(\alpha)}) = \sum_{k \in Z_j} a_{j,k}^{(\alpha)} \psi_{j,k}$), then by definition of $\Sigma(s, B)$

$$
\begin{aligned}
\left\| f^{(\alpha)} - \Sigma(s, B) \right\|_2 &= \inf_{(b_{l,k})_{l,k} : 2^{ls} \|b_{l,\cdot}\|_{l_2} \leq B} \sqrt{\sum_{l,k} (a_{l,k}^{(\alpha)} - b_{l,k})^2} \\
&= \inf_{(b_{l,k})_{l,k} : 2^{ls} \|b_{l,\cdot}\|_{l_2} \leq B} \sqrt{\sum_{k \in Z_j} (a_{j,k}^{(\alpha)} - b_{j,k})^2 + \sum_{l \neq j, k \in Z_l} b_{l,k}^2} \\
&= \inf_{(b_k)_k : 2^{js} \|b\|_{l_2} \leq B} \sqrt{\sum_{k \in Z_j} (a_{j,k}^{(\alpha)} - b_k)^2} \\
&= \max\left(0, \sqrt{\sum_{k \in Z_j} (a_{j,k}^{(\alpha)})^2} - B 2^{-js} \right) \\
&= \max\left(0, \left\| \Pi_{W_j} f^{(\alpha)} \right\|_2 - B 2^{-js} \right).
\end{aligned}
$$

Since by definition of the Euclidian ball, for any $u \in l_2$, we have $\inf_{v \in l_2 : \|v\|_{l_2} = 1} \|u - v\|_{l_2} = \max(0, \|u\|_{l_2} - 1)$.

We thus have $\forall \alpha \in I$, and for all n large enough

$$\left\| f^{(\alpha)} - \Sigma(s, B) \right\|_2 \geq \left\| \Pi_{W_j} f^{(\alpha)} \right\|_2 - B2^{-js}$$
$$\geq \upsilon n^{-t/(2t+1/2)} - B n^{-s/(2t+1/2)}$$
$$\geq \frac{\upsilon n^{-t/(2t+1/2)}}{2}$$

by triangular inequality and since for any $g \in \Sigma(s, B)$, $\|\Pi_{W_j}(g)\|_2 \leq 2^{-s} B n^{-s/(2t+1/2)} \leq \upsilon/(2) n^{t/(2t+1/2)}$ for n large enough, since $s > t$. This together with the fact that $f^{(\alpha)} \in \Sigma(t, B)$ implies that $\forall \alpha \in I$, $f^{(\alpha)} \in \tilde{\Sigma}(t, B, \rho_n)$.

We know that $0 \in \Sigma(s, B)$, and that $\forall \alpha$, $f^{(\alpha)} \in \tilde{\Sigma}(t, B, \rho_n)$ (by the previous equations). This implies that the testing problem (5.1) is a strictly easier problem than the testing problem (1.1), that is, that

$$\inf_{\Psi} \left[\mathbb{E}_0[\Psi] + \sup_{f^{(\alpha)}, \alpha \in I} \mathbb{E}_{f^{(\alpha)}}[1 - \Psi] \right] \leq \inf_{\Psi} \left[\sup_{f \in \Sigma(s, B)} \mathbb{E}_f[\Psi] + \sup_{f \in \tilde{\Sigma}(t, B, \rho_n)} \mathbb{E}_f[1 - \Psi] \right].$$

We know that there is no $1 - 2\upsilon^4$ consistent test for the test (5.1) and hence, there is no $1 - 2\upsilon^4$ consistent test for test (1.1) (and it holds for any $0 \leq \upsilon < 1$).

Acknowledgements

References

[1] Baraud, Y. (2004). Confidence balls in Gaussian regression. *Ann. Statist.* **32** 528–551. MR2060168

[2] Baraud, Y., Huet, S. and Laurent, B. (2005). Testing convex hypotheses on the mean of a Gaussian vector. Application to testing qualitative hypotheses on a regression function. *Ann. Statist.* **33** 214–257. MR2157802

[3] Barron, A., Birgé, L. and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** 301–413. MR1679028

[4] Belitser, E. and Enikeeva, F. (2008). Empirical Bayesian test of the smoothness. *Math. Methods Statist.* **17** 1–18. MR2400361

[5] Bergh, J. and Löfström, J. (1976). *Interpolation Spaces. An Introduction*. Berlin: Springer. MR0482275

[6] Besov, O.V., Il'in, V.P., Nikol'skiĭ, S.M. and Mikhaĭlovich, S. (1978). *Integral Representations of Functions and Imbedding Theorems*, *Vol.* 1. New York: Halsted Press. MR0519341

[7] Birgé, L. (2001). An alternative point of view on Lepski's method. In *State of the Art in Probability and Statistics* (*Leiden*, 1999). *Institute of Mathematical Statistics Lecture Notes – Monograph Series* **36** 113–133. Beachwood, OH: IMS. MR1836557

[8] Blanchard, G., Delattre, S. and Roquain, E. (2014). Testing over a continuum of null hypotheses with False Discovery Rate control. *Bernoulli* **20** 304–333. MR3160584

[9] Bull, A.D. and Nickl, R. (2013). Adaptive confidence sets in L^2. *Probab. Theory Related Fields* **156** 889–919. MR3078289

[10] Cai, T.T. and Low, M.G. (2004). An adaptation theory for nonparametric confidence intervals. *Ann. Statist.* **32** 1805–1840. MR2102494

[11] Cai, T.T. and Low, M.G. (2006). Adaptive confidence balls. *Ann. Statist.* **34** 202–228. MR2275240

[12] Cohen, A., Daubechies, I. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.* **1** 54–81. MR1256527

[13] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508–539. MR1394974

[14] Dümbgen, L. and Spokoiny, V.G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** 124–152. MR1833961

[15] Dziedziul, K. and Ćmiel, B. (2014). Density smoothness estimation problem using a wavelet approach. *ESAIM Probab. Statist.* **18** 130–144. MR3143736

[16] Fromont, M. and Laurent, B. (2006). Adaptive goodness-of-fit tests in a density model. *Ann. Statist.* **34** 680–720. MR2281881

[17] Gayraud, G. and Pouet, C. (2005). Adaptive minimax testing in the discrete regression scheme. *Probab. Theory Related Fields* **133** 531–558. MR2197113

[18] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications. Lecture Notes in Statistics* **129**. New York: Springer. MR1618204

[19] Hoffmann, M. (1999). On nonparametric estimation in nonlinear AR(1)-models. *Statist. Probab. Lett.* **44** 29–45. MR1706307

[20] Hoffmann, M. and Lepski, O. (2002). Random rates in anisotropic regression. *Ann. Statist.* **30** 325–396. MR1902892

[21] Hoffmann, M. and Nickl, R. (2011). On adaptive inference and confidence bands. *Ann. Statist.* **39** 2383–2409. MR2906872

[22] Horowitz, J.L. and Spokoiny, V.G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69** 599–631. MR1828537

[23] Ingster, Yu.I. (1986). Minimax testing of nonparametric hypotheses on a distribution density in the l_p metrics. *Theory Probab. Appl.* **31** 333–337.

[24] Ingster, Yu.I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Math. Methods Statist.* **2** 85–114. MR1257978

[25] Ingster, Yu.I. and Suslina, I.A. (2003). *Nonparametric Goodness-of-fit Testing Under Gaussian Models. Lecture Notes in Statistics* **169**. New York: Springer. MR1991446

[26] Juditsky, A. and Lambert-Lacroix, S. (2003). Nonparametric confidence set estimation. *Math. Methods Statist.* **12** 410–428. MR2054156

[27] Juditsky, A. and Nemirovski, A. (2002). On nonparametric tests of positivity/monotonicity/convexity. *Ann. Statist.* **30** 498–527. MR1902897

[28] Lepski, O.V. and Spokoiny, V.G. (1999). Minimax nonparametric hypothesis testing: The case of an inhomogeneous alternative. *Bernoulli* **5** 333–358. MR1681702

[29] Lepskiǐ, O.V. (1992). On problems of adaptive estimation in white Gaussian noise. In *Topics in Nonparametric Estimation. Adv. Soviet Math.* **12** 87–106. Providence, RI: Amer. Math. Soc. MR1191692

[30] Low, M.G. (1997). On nonparametric confidence intervals. *Ann. Statist.* **25** 2547–2554. MR1604412

[31] Y. Meyer (1992). Wavelets and applications. In *Proceedings of the Second International Conference Held in Marseille, May* 1989. *RMA: Research Notes in Applied Mathematics* **20**. Paris: Masson. MR1276521

[32] Nickl, R. and van de Geer, S. (2013). Confidence sets in sparse regression. *Ann. Statist.* **41** 2852–2876. MR3161450

[33] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430. MR1425959

[34] Pouet, C. (2002). Test asymptotiquement minimax pour une hypothèse nulle composite dans le modèle de densité. *C. R. Math. Acad. Sci. Paris* **334** 913–916. MR1909939

[35] Reiß, M. (2008). Asymptotic equivalence for nonparametric regression with multivariate and random design. *Ann. Statist.* **36** 1957–1982. MR2435461

[36] Robins, J. and van der Vaart, A. (2006). Adaptive nonparametric confidence sets. *Ann. Statist.* **34** 229–253. MR2275241

[37] Spokoiny, V.G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24** 2477–2498. MR1425962

[38] Tsybakov, A.B. (2004). *Introduction à L'estimation Non-paramétrique. Mathématiques & Applications* (*Berlin*) [*Mathematics & Applications*] **41**. Berlin: Springer. MR2013911