The Annals of Statistics
2003, Vol. 31, No. 1, 225–251

ADAPTIVE TESTS OF LINEAR HYPOTHESES BY MODEL SELECTION

BY Y. BARAUD, S. HUET AND B. LAURENT

Ecole Normale Superieure, DMA, INRA, Laboratoire de Biométrie and Université Paris-Sud

We propose a new test, based on model selection methods, for testing that the expectation of a Gaussian vector with n independent components belongs to a linear subspace of $\mathbb{R}^n$ against a nonparametric alternative. The testing procedure is available when the variance of the observations is unknown and does not depend on any prior information on the alternative. The properties of the test are nonasymptotic and we prove that the test is rate optimal [up to a possible $\log(n)$ factor] over various classes of alternatives simultaneously. We also provide a simulation study in order to evaluate the procedure when the purpose is to test goodness-of-fit in a regression model.

1. Introduction.

We consider the regression model

$$Y = f + \sigma\varepsilon, \tag{1}$$

where f is an unknown vector of $\mathbb{R}^n$, ε a Gaussian vector with i.i.d. components distributed as standard Gaussian random variables, $\varepsilon_i \sim \mathcal{N}(0,1)$, and σ some unknown positive quantity. Let V be some linear subspace of $\mathbb{R}^n$. The aim of this paper is to propose a test of the null hypothesis "f belongs to V" against the alternative that it does not under no prior information on f.

The testing procedure can be described in the following way. We consider a finite collection of linear subspaces of $V^\perp$, $\{S_m, m \in \mathcal{M}\}$, such that for each m, $S_m \neq V^\perp$ and $S_m \neq \{0\}$. The index set $\mathcal{M}$ is allowed to depend on the number of observations n. Our testing procedure consists of doing several tests. Let $\{\alpha_m, m \in \mathcal{M}\}$ be a suitable collection of numbers in $]0,1[$, we consider for each $m \in \mathcal{M}$ the Fisher test of level α_m of the null hypothesis

$$\mathrm{H}_0\colon\ f \in V \quad \text{against the alternative} \quad \mathrm{H}_{1,m}\colon\ f \in (V + S_m) \setminus V$$

and we decide to reject the null hypothesis if one of the Fisher tests does.

In the particular case where σ is known and equal to $1/\sqrt{n}$, our statistical framework is related to the Gaussian sequence model. In this latter framework, Spokoiny (1996) proposed a procedure for testing $f = 0$ against $f \neq 0$ which has the property of achieving the minimax rate of testing [up to an unavoidable

Received June 1999; revised December 2001.

AMS 2000 subject classifications. Primary 62G10; secondary 62G20.

Key words and phrases. Adaptive test, model selection, linear hypothesis, minimax hypothesis testing, nonparametric alternative, goodness-of-fit, nonparametric regression, Fisher test, Fisher's quantiles.

$\log\log(n)$ factor] for a wide range of Besov classes. The procedure is based on thresholding techniques, the choice of the threshold being judiciously calibrated for the problem at hand.

When f takes the particular form $(F(x_1), \dots, F(x_n))^T$, for some function F (we shall call this framework the functional regression model), many tests of linear hypotheses have been proposed. Some are based on nonparametric function estimation methods, a very natural idea being to compare a nonparametric estimation of F to a parametric one computed under the null hypothesis. We refer, for example, to Staniswalis and Severini (1991), Müller (1992), Härdle and Mammen (1993), Hart (1997), Chen (1994) and Eubank and LaRiccia (1993) for an overview. Such methods present several drawbacks. First, when the nonparametric estimation is based on some a priori choice of the regularity of F, the same holds for the testing procedure. In particular, the issue of the test depends upon this choice. Besides, a data-driven choice may affect the level of the test. Dette and Munk (1998) proposed a test based on the estimation of some $\mathbb{L}^2$-distance between F and the null. The procedure does not depend on the choice of a smoothing parameter. In addition the asymptotic distribution of the test statistic under the null approximates well its distribution for a small sample size. However, in the papers previously mentioned, the authors give no hint as to how their procedures perform from the minimax point of view.

In a recent paper, Horowitz and Spokoiny (2001) proposed a test for parametric hypotheses against nonparametric alternatives in the functional regression model with heteroscedastic errors, that does not require any prior knowledge of the smoothness of F. Their procedure rejects the null hypothesis if for some bandwidth among a grid, the distance between the nonparametric kernel estimator and the kernel-smoothed parametric estimator of F under the null hypothesis is large. This so-defined test achieves asymptotically the desired level. It is rate optimal among adaptive procedures over Hölder classes of alternatives and achieves nearly the parametric rate of testing for directional alternatives.

Several methods based on model selection by penalized criteria, originally used for estimation inference, have been proposed in order to avoid using a prior assumption on f. The test proposed by Eubank and Hart (1992) is based on a penalized criterion which is related to the well-known Mallows C_p and has the property to achieve the parametric rate of testing over directional alternatives. This criterion (and related ones) is used for nonparametric estimation in the regression framework. We refer, for example, to the work of Birgé and Massart (2001) in the context of Gaussian errors and to Baraud (2000a) under weaker assumptions.

Another way to handle the problem of testing $f \in V$ is to consider the coordinates of the vector Y onto an orthonormal basis of $V^{\perp}$. Then, the problem of testing $f \in V$ amounts to testing that these coordinates are distributed as i.i.d. centered Gaussian random variables of common variance σ^2. By so doing, one shifts the original problem of goodness-of-fit in the regression framework with known variance to the problem of goodness-of-fit in the density framework. In this latter framework, testing procedures based on penalized criteria which are akin to Schwarz's have been proposed by Inglot and Ledwina (1996). Schwarz's criterion (and related ones) can be used for estimation issues in the density model; see, for example, Nishii (1988) and Castellan (2000) for more recent work.

Our procedure is based on a penalized criterion which is connected to that proposed by Laurent and Massart (2000) for the estimation of a quadratic functional in the regression framework. Such a connection will be highlighted in Section 2.3. A key point of our procedure is that it permits mixing several kinds of linear spaces S_m, increasing thus our chance to capture some specific feature of the discrepancy between f and the null for a fixed value of n. Consequently, when the collection of S_m's is suitably chosen, we show that the testing procedure is powerful over a large class of alternatives, both from the theoretical and practical point of view. The theory developed in this paper is restricted to the case where the errors are i.i.d. Gaussian. However the procedure offers the following advantages. The properties of the testing procedure are nonasymptotic. For each n the test has the desired level and we characterize a set of vectors over which our test is powerful. We establish that the procedure is both minimax over some classes of alternatives (up to a logarithmic factor) and $1/\sqrt{n}$-consistent over directional ones. Its performance from the practical point of view is illustrated in simulation studies. It is shown there that the procedure is very easy to implement and robust with respect to some discrepancy to the Gaussian assumption.

The paper is organized as follows: in Section 2 we present the testing procedure. Our main result is stated in Section 3 and rates of testing are given in Section 4. The simulation studies are presented in Section 5. Sections 6–9 contain the proofs.

Let us now introduce some notation that is repeatedly used throughout the paper. In the sequel, $\mathbb{P}_f$ denotes the law of the observation Y and d is the dimension of V. For s, t in $\mathbb{R}^n$ we set $\|s\|_n^2 = \sum_{i=1}^n s_i^2/n$ and $d_n(t,s) = \|s - t\|_n$. For each $m \in \mathcal{M}$, we set $V_m = V \oplus S_m$ and, respectively, denote by D_m and N_m the dimension of S_m and $V_m^\perp$. For any linear subspace W of $\mathbb{R}^n$ we denote by Π_W the orthogonal projector onto W (with respect to the Euclidean norm). In order to keep the notation as short as possible we set for each $m \in \mathcal{M}$, $\Pi_m = \Pi_{S_m}$. For any $u \in \mathbb{R}$, $\bar{\Phi}(u)$, $\bar{\chi}_D(u)$, $\bar{F}_{D,N}(u)$ denote, respectively, the probability for a standard Gaussian variable, a chi-square with D degrees of freedom and a Fisher with D and N degrees of freedom to be larger than u.

2. The testing procedure.

2.1. *Description of the procedure.* Let us fix some $\alpha \in]0,1[$. Let $n \geq d + 2$ and consider a finite collection of linear spaces $\{S_m, m \in \mathcal{M}\}$ of $V^\perp$ such that for all $m \in \mathcal{M}$, $1 \leq D_m \leq n - d - 1$. We set

$$T_\alpha = \sup_{m \in \mathcal{M}} \left\{ \frac{N_m \|\Pi_m Y\|_n^2}{D_m \|Y - \Pi_{V_m} Y\|_n^2} - \bar{F}^{-1}_{D_m,N_m}(\alpha_m) \right\}, \tag{2}$$

where $\{\alpha_m,\ m \in \mathcal{M}\}$ is a collection of numbers in $]0,1[$ such that

$$\forall f \in V, \qquad \mathbb{P}_f(T_\alpha > 0) \le \alpha. \tag{3}$$

We reject the null hypothesis when T_α is positive. In the sequel, we choose the collection $\{\alpha_m,\ m \in \mathcal{M}\}$ in accordance with one of the two following procedures:

P1. For all $m \in \mathcal{M}$, $\alpha_m = a_n$ where a_n is the α-quantile of the random variable

$$\inf_{m \in \mathcal{M}} \bar{F}_{D_m,N_m}\left(\frac{N_m \|\Pi_m \varepsilon\|_n^2}{D_m \|\varepsilon - \Pi_{V_m}\varepsilon\|_n^2}\right).$$

P2. The α_m's satisfy the equality

$$\sum_{m \in \mathcal{M}} \alpha_m = \alpha.$$

2.2. *Behavior of the test under the null hypothesis.* The test associated with procedure P1 has the property to be of size α. More precisely, we have

$$\forall f \in V, \qquad \mathbb{P}_f(T_\alpha > 0) = \alpha. \tag{4}$$

This result follows from the fact that a_n satisfies

$$\mathbb{P}\left(\sup_{m \in \mathcal{M}} \left\{\frac{N_m \|\Pi_m \varepsilon\|_n^2}{D_m \|\varepsilon - \Pi_{V_m}\varepsilon\|_n^2} - \bar{F}^{-1}_{D_m,N_m}(a_n)\right\} > 0\right) = \alpha, \tag{5}$$

and that for $f \in V$, $\Pi_m Y = \Pi_m \varepsilon$ and $Y - \Pi_{V_m} Y = \varepsilon - \Pi_{V_m}\varepsilon$.

The choice of the α_m's as proposed in procedure P2 allows us to deal with the computation of the α-quantile a_n. In fact no preliminary computation is required at all. A test associated to procedure P2 leads to a Bonferonni test for which we only have (3) by arguing as follows:

$$\begin{aligned}
\mathbb{P}_f(T_\alpha > 0) &\le \sum_{m \in \mathcal{M}} \mathbb{P}_f\left(\frac{N_m \|\Pi_m Y\|_n^2}{D_m \|Y - \Pi_{V_m} Y\|_n^2} > \bar{F}^{-1}_{D_m,N_m}(\alpha_m)\right) \\
&\le \sum_{m \in \mathcal{M}} \alpha_m = \alpha.
\end{aligned}$$

The procedure is therefore conservative. Another drawback lies in the fact that the choice of α_m's is arbitrary, which gives a Bayesian flavor to the procedure. Nevertheless, we see in Section 5 that for suitable S_m's associated with the choice $\alpha_m = \alpha/|\mathcal{M}|$ for all $m \in \mathcal{M}$, the procedure P2 is powerful and, because of its conservative character, more robust with respect to a discrepancy to Gaussianity.

2.3. *Connection with model selection.* The aim of this section is to explain the underlying ideas of the procedure we propose and to relate it to model selection. In order to simplify our task, we assume that the quantity σ is known and that $V = \{0\}$. Let us start with a collection of linear subspaces of $\mathbb{R}^n$, $\{S_m,\ m \in \mathcal{M} = \{0, 1, \dots, n\}\}$, which is totally ordered for inclusion and such that for each $m \in \mathcal{M}$, $\dim(S_m) = m$ (note that we add 0 to the index set $\mathcal{M}$). As σ is known, our test statistic T_α can be replaced by the simpler one,

$$\hat{T}_\alpha = \sup_{m \in \mathcal{M}\setminus\{0\}} \left[\|\Pi_m Y\|_n^2 - \frac{\sigma^2}{n} \bar{\chi}_m^{-1}(\alpha_m) \right], \tag{6}$$

where $\bar{\chi}_m^{-1}(u)$ denotes the $1-u$ quantile of a χ^2-random variable with m degrees of freedom. Similarly, we reject the null $f = 0$ if $\hat{T}_\alpha > 0$. The connection with model selection is made as follows: Note that the critical region is equivalently defined as $\widehat{m} \geq 1$, where $\widehat{m}$ is a maximizer among $\mathcal{M}$ of the penalized criterion $\mathrm{Crit}(m)$ given by

$$\mathrm{Crit}(m) = \|\Pi_m Y\|_n^2 - \mathrm{pen}(m), \tag{7}$$

with pen defined here by

$$\mathrm{pen}(0) = 0 \quad \text{and} \quad \forall m \in \mathcal{M} \setminus \{0\},\ \mathrm{pen}(m) = \bar{\chi}_m^{-1}(\alpha_m)\frac{\sigma^2}{n}. \tag{8}$$

The test proposed by Eubank and Hart (1992) is based on the same critical region with a penalty function pen defined as $\mathrm{pen}(m) = \sigma^2 m C_\alpha / n$. The constant C_α is greater than 1 and is calibrated for the test to be asymptotically of level α. Therefore, the difference between Eubank and Hart's procedure and ours lies in the choice of the penalty function. The penalty function chosen by Eubank and Hart is related to Mallows' C_p. Criteria based on penalty functions of the form $C\sigma^2 m/n$ with $C > 1$ aim at selecting a "good" model for the purpose of estimating f with respect to quadratic loss [see Baraud (2000a) or Birgé and Massart (2001)]. In contrast, the penalty function involved in our procedure is related to that proposed in Laurent and Massart (2000) and is calibrated for the purpose of estimating the quadratic functional $\|f\|_n^2 = d_n^2(f, V)$.

3. The power of the test. Let α and β be two numbers in $]0, 1[$, and let $\{S_m,\ m \in \mathcal{M}\}$ and $\{\alpha_m,\ m \in \mathcal{M}\}$ be the collections defined in Section 2. The aim of this section is to describe a set of $\mathbb{R}^n$-vectors over which the test defined in Section 2.1 is powerful.

Let us first introduce some quantities that depend on α_m, β, D_m and N_m. For each $u > 0$ and $m \in \mathcal{M}$, let us set $L_m = \log(1/\alpha_m)$, $L = \log(2/\beta)$, $k_m = 2\exp(4L_m/N_m)$,

$$K_m(u) = 1 + 2\sqrt{\frac{u}{N_m}} + 2k_m\frac{u}{N_m},$$

$$\Lambda_1(m) = 2.5\big(1 + K_m(L_m) \vee k_m\big)\frac{D_m + L_m}{N_m},$$

$$\Lambda_2(m) = 2.5\sqrt{1 + K_m^2(L)}\left(1 + \sqrt{\frac{D_m}{N_m}}\right),$$

$$\Lambda_3(m) = 2.5\left[\left(\frac{k_m K_m(L)}{2}\right) \vee 5\right]\left(1 + 2\frac{D_m}{N_m}\right).$$

Under the following condition the quantities $\Lambda_1(m)$, $\Lambda_2(m)$ and $\Lambda_3(m)$ behave like constants.

($\mathrm{H}_{\mathcal{M}}$) For all $m \in \mathcal{M}$, $\alpha_m \geq \exp(-N_m/10)$ and $\beta \geq 2\exp(-N_m/21)$.

These conditions are usually met for reasonable choices of the collections $\{S_m,\ m \in \mathcal{M}\}$ and $\{\alpha_m,\ m \in \mathcal{M}\}$. Let us now give our main result.

THEOREM 1. *Assume that $n \geq d + 2$ and let T_α be the test statistic defined by* (2). *Then $\mathbb{P}_f(T_\alpha > 0) \geq 1 - \beta$ for all f belonging to the set*

$$\mathcal{F}_n(\beta) = \left\{f \in \mathbb{R}^n,\ d_n^2(f, V) \geq \inf_{m \in \mathcal{M}} \Delta(f, m)\right\},$$

where

$$\Delta(f, m) = [1 + \Lambda_1(m)]d_n^2(\Pi_{V^\perp} f, S_m) + \left[\Lambda_2(m)\sqrt{D_m \log\left(\frac{2}{\alpha_m \beta}\right)} + \Lambda_3(m)\log\left(\frac{2}{\alpha_m \beta}\right)\right]\frac{\sigma^2}{n}.$$

Moreover, under Condition ($\mathrm{H}_{\mathcal{M}}$) *the following inequalities hold*:

$$\Lambda_1(m) \leq 10\frac{D_m + L_m}{N_m}, \qquad \Lambda_2(m) \leq 5\left(1 + \sqrt{\frac{D_m}{N_m}}\right)$$

and

$$\Lambda_3(m) \leq 12.5\left(1 + 2\frac{D_m}{N_m}\right).$$

COMMENT. For the sake of simplicity, let us assume that $V = \{0\}$ and that the α_m's are chosen to be all equal to $\alpha/|\mathcal{M}|$ (procedure P2). As already mentioned, the test described in Section 2 consists of doing several Fisher tests of $f = 0$ ($V = \{0\}$ here) against $f \in S_m$ at level α_m and of rejecting the null hypothesis if one of the Fisher tests does. Then a natural question arises: what advantage should be expected of such a procedure compared to a classical Fisher test? In order to answer this question, let us fix some $m \in \mathcal{M}$ and consider the test statistic

$$T_m = \left\{\frac{N_m\|\Pi_m Y\|_n^2}{D_m\|Y - \Pi_{V_m} Y\|_n^2} - \bar{F}_{D_m, N_m}^{-1}(\alpha)\right\}.$$

By applying Theorem 1 with $\mathcal{M} = \{m\}$ and $\alpha_m = \alpha$, we obtain that the Fisher test $\phi_m = \mathbb{1}\{T_m > 0\}$ has power greater than $1 - \beta$ over the set $\mathcal{F}_m(\beta)$ of vectors f satisfying

$$\|f\|_n^2 \geq C_1 d_n^2(f, S_m) + C_2\left[\sqrt{D_m \log\left(\frac{2}{\alpha\beta}\right)} + \log\left(\frac{2}{\alpha\beta}\right)\right]\frac{\sigma^2}{n}, \tag{9}$$

where C_1 and C_2 behave like constants with respect to n if the ratio $(D_m + \log(1/\alpha))/N_m$ remains bounded, and where C_1 is close to 1 if this ratio is small. Inequality (9) gives a good description of the class of functions f over which this test is powerful since the following result holds.

PROPOSITION 1. *Let $\beta \in]0, 1 - \alpha[$ and*

$$\theta(\alpha, \beta) = \sqrt{2 \log\left(1 + 4(1 - \alpha - \beta)^2\right)}.$$

If

$$\|f\|_n^2 \leq d_n^2(f, S_m) + \theta(\alpha, \beta)\sqrt{D_m}\frac{\sigma^2}{n},$$

then $\mathbb{P}_f(T_m > 0) \leq 1 - \beta$.

If we now consider the test statistic T_α defined by (2), we obtain from Theorem 1 that the corresponding test has power greater than $1 - \beta$ over the set of vectors for which there exists $m \in \mathcal{M}$ such that

$$\|f\|_n^2 \geq C_1 d_n^2(f, S_m) + C_2\left[\sqrt{D_m \log\left(\frac{2}{\alpha_m\beta}\right)} + \log\left(\frac{2}{\alpha_m\beta}\right)\right]\frac{\sigma^2}{n}. \tag{10}$$

The difference between (9) and (10) lies in the fact that $\log(1/\alpha)$ has been replaced by $\log(1/\alpha_m)$. For typical collections of S_m's and α_m's, the difference $\log(1/\alpha_m) - \log(1/\alpha)$ is of order $\log(n)$ or $\log\log(n)$. Therefore, we deduce from (10) that for each $\beta \in]0, 1 - \alpha[$, the test based on T_α has power greater than $1 - \beta$ over a class of vectors which is close to $\bigcup_{m \in \mathcal{M}} \mathcal{F}_m(\beta)$. It follows that for each $f \neq 0$ the power of this test under $\mathbb{P}_f$ is comparable to the power of the best of the Fisher tests among the family $\{\phi_m, m \in \mathcal{M}\}$.

4. Rates of testing.

In this section, we consider the case where $V = \{0\}$ and establish rates of testing in the two following statistical models.

The functional regression model. Let $x_1, \ldots, x_n$ be deterministic points in $[0, 1]$ and let F be some unknown function of the Hilbert space $\mathbb{L}_2([0, 1], dx)$. We consider the regression model given by (1) where

$$f = \left(F(x_1), \ldots, F(x_n)\right)^T$$

and where σ^2 is unknown.

The truncated Gaussian sequence model. Let $(\phi_j)_{j\geq 1}$ be some Hilbert basis of $\mathbb{L}_2([0,1], dx)$ endowed with the usual inner product denoted by $\langle\cdot,\cdot\rangle$, and let F be some function in this space. We consider the regression model given by (1) where

$$f = (\langle F, \phi_1\rangle, \ldots, \langle F, \phi_n\rangle)^T$$

and $\sigma^2 = \tau^2/n$, where τ is unknown.

If τ is known and equals 1, the truncated Gaussian sequence model is related to the Gaussian sequence model for which one observes the whole sequence $(Y_i,\ i \geq 1)$ given by

$$Y_i = f_i + \frac{1}{\sqrt{n}}\varepsilon_i, \qquad i \in \mathbb{N} \setminus \{0\}.$$

This model is the ideal framework for doing statistical inference in the Gaussian white noise model by filtering the latter onto the ϕ_j's. For a suitable choice of the basis $(\phi_j)_{j\geq 1}$, for example, the Fourier or a wavelet basis, it is possible to establish connections between the regularity of the function F and the fact that the sequence of its coefficients belongs to some sets such as ellipsoids, l_p-balls and so forth. Such results are recalled in Laurent and Massart (2000).

The main difference between the functional regression model and the truncated Gaussian sequence model lies in the meaning of the vector f. In the functional regression model, the main feature of $f = (F(x_1), \ldots, F(x_n))^T$ is that the differences between its successive components are all the smaller because F is regular. In contrast, the main feature of f in the truncated Gaussian sequence model lies in the fact that its components tend to decrease. These particular features can be taken into account by the choice of the collection of S_m's.

4.1. *The functional regression model.* In this section, we take full advantage of the possibility of mixing several linear spaces in the collection $\{S_m,\ m \in \mathcal{M}\}$, each of them being appropriate for detecting specific alternatives. We show that for an adequate choice of the collection, we obtain a test that is both rate optimal (in some sense) over the s-Hölderian balls of radius R, simultaneously for all $s \in]1/4, 1]$ and $R > 0$, and that is able to detect local alternatives at the parametric rate of testing. For this aim, we introduce the following collections of α_m's and S_m's.

1. For each $k \in \mathcal{M}_1 = \{2^j,\ j \geq 0\} \cap \{1, \ldots, [n/2]\}$ ($[x]$ denotes the integer part of x), we define $\alpha_{(k,1)} = \alpha/(2|\mathcal{M}_1|)$ and $S_{(k,1)}$ as the linear space spanned by the vectors
$$\left\{\left(\mathbb{1}_{](j-1)/k, j/k]}(x_1), \ldots, \mathbb{1}_{](j-1)/(k), j/k]}(x_n)\right)^T,\ j = 1, \ldots, k\right\}.$$
2. Let $(\phi_j)_{j\geq 1}$ be a Hilbert basis of $\mathbb{L}_2([0,1], dx)$. For each $k \in \mathcal{M}_2 = \{1, \ldots, n\}$ we define $\alpha_{(k,2)} = 3\alpha/(\pi^2 k^2)$ and $S_{(k,2)}$ as the linear space spanned by the vector $(\phi_k(x_1), \ldots, \phi_k(x_n))^T$.

In the sequel, we set $\mathcal{M} = \{(k,1),\ k \in \mathcal{M}_1\} \cup \{(k,2),\ k \in \mathcal{M}_2\}$ and consider the collections $\{S_m,\ m \in \mathcal{M}\}$ and $\{\alpha_m,\ m \in \mathcal{M}\}$. For $R > 0$ and $s \in]0,1]$, we define $\mathcal{H}_s(R) \subset \mathbb{R}^n$ as

$$\mathcal{H}_s(R) = \{(F(x_1), \dots, F(x_n))^T / F \in H_s(R)\},$$

where

$$H_s(R) = \{F : [0,1] \to \mathbb{R},\ \forall (x,y) \in [0,1]^2,\ |F(x) - F(y)| \le R|x-y|^s\}.$$

COROLLARY 1. *Let α and β be two numbers in $]0,1[$. Let T_α be the test statistic defined by* (2). *Then the following holds*:

(i) *Assume that*

$$n \ge 42 \max\left\{\log\left(\frac{45}{\alpha}\right), \log\left(\frac{2}{\beta}\right)\right\} \tag{11}$$

and that $R^2 \ge \sqrt{\log\log(n)}\sigma^2/n$. Then there exists a constant $C(\alpha,\beta)$ depending only on α and β such that for all $s \in]0,1]$ and for all $f \in \mathcal{H}_s(R)$, satisfying $\|f\|_n^2 \ge C(\alpha,\beta)\rho_{n,1}^2$ with

$$\rho_{n,1}^2 = R^{2/(1+4s)}\left(\frac{\sqrt{\log\log(n)}}{n}\sigma^2\right)^{4s/(1+4s)} + R^2 n^{-2s} + \frac{\log\log(n)}{n}\sigma^2,$$

we have $\mathbb{P}_f(T_\alpha > 0) \ge 1 - \beta$.

(ii) *Let $F \in \mathbb{L}_2([0,1], dx)$ and $f = (F(x_1), \dots, F(x_n))^T$. Assume that the x_i's satisfy for all $k \ge 1$,*

$$\lim_{n \to +\infty} \frac{1}{n}\sum_{i=1}^n F(x_i)\phi_k(x_i) = \int_0^1 F(x)\phi_k(x)\,dx,$$

$$\lim_{n \to +\infty} \frac{1}{n}\sum_{i=1}^n \phi_k^2(x_i) = 1, \qquad \lim_{n \to +\infty} \frac{1}{n}\sum_{i=1}^n F^2(x_i) = \int_0^1 F^2(x)\,dx.$$

If for some $k_0 \ge 1$, $(\int_0^1 F(x)\phi_{k_0}(x)\,dx)^2 > -17.5\log(\alpha_{(k_0,2)}\beta/2)\sigma^2$, then for n large enough,

$$\mathbb{P}_{f/\sqrt{n}}(T_\alpha > 0) \ge 1 - \beta.$$

COMMENTS. The results of this corollary are obtained by taking the union of the two collections $\{S_{(k,1)},\ k \in \mathcal{M}_1\}$ and $\{S_{(k,2)},\ k \in \mathcal{M}_2\}$. It can be seen in the proof that the first collection allows us to obtain (i) and the second one (ii). By mixing the two collections, we increase the performance of the test.

In case (i), note that the rate of testing is free from any assumption on the x_i's. This rate is of order $(\sqrt{\log\log(n)}/n)^{2s/(1+4s)}$ for $s > 1/4$. In the Gaussian white noise model, the rate $n^{-2s/(1+4s)}$ is known to be minimax over s-Hölderian balls [Ingster (1993)]. The loss of efficiency of a $\log\log(n)$ factor is because the test is adaptive, which means that it does not use any prior knowledge on s and R [for further details see Spokoiny (1996)].

When $s < 1/4$, the rate of testing is of order n^{-s} and we do not know whether this rate is optimal or not. If the variance were known, the procedure based on $\widehat{T}_\alpha$ defined at (6) would lead to a rate of testing of order $n^{-1/4}$, whatever the value of $s < 1/4$ [see Baraud, Huet and Laurent (2001)]. This rate is known to be minimax on related sets of vectors as proved in Baraud (2000b).

The result presented in (ii) is analogous to that previously given by Eubank and Hart (1992). Namely, we get the parametric rate $1/\sqrt{n}$ in directional alternatives.

For each $k \in \mathcal{M}_1$, $S_{(k,1)}$ is related to the functional space generated by the piecewise constant functions over the intervals $](j-1)/k, j/k]$, $j = 1, \dots, k$. The proof of (i) is based on the fact that these functional spaces have good approximation properties in sup-norm with respect to the functions belonging to $H_s(R)$. In the same way, by using a collection of linear spaces, $S_{(k,1)}$, which is now related to piecewise polynomials of degree not larger than $r \geq 1$, one would derive some nearly minimax results over the class of functions F on $[0, 1]$ such that their derivatives of order r belong to $H_s(R)$.

4.2. *The truncated Gaussian sequence model.* The aim of this section is to establish uniform separation rates over l_p-balls defined as

$$\mathcal{E}_{p,s}(R) = \left\{ f \in \mathbb{R}^n, \left(\sum_{i=1}^{n} i^{ps} |f_i|^p \right)^{1/p} \leq R \right\} \tag{12}$$

with $s, p, R > 0$.

Let us now introduce the collections $\{S_m, m \in \mathcal{M}\}$ and $\{\alpha_m, m \in \mathcal{M}\}$. Let $(e_1, \dots, e_n)$ be the canonical basis of $\mathbb{R}^n$. For each $k \in \mathcal{M}_1 = \{2^j, j \geq 0\} \cap \{1, \dots, [n/2]\}$, let $S_{(k,1)}$ be the linear space generated by $\{e_1, \dots, e_k\}$ and for all $k \in \mathcal{M}_2 = \{1, \dots, n\}$, let $S_{(k,2)}$ be the space generated by the vector e_k. We set $\mathcal{M} = \{(k, 1), k \in \mathcal{M}_1\} \cup \{(k, 2), k \in \mathcal{M}_2\}$. Let $k_0 = \sup(\mathcal{M}_1)$; for all $k \in \mathcal{M}_1$ and $k \neq k_0$, we set $\alpha_{(k,1)} = \alpha/(4|\mathcal{M}_1|)$ and $\alpha_{(k_0,1)} = \alpha/4$. For all $k \in \mathcal{M}_2$, we set $\alpha_{(k,2)} = \alpha/(2n)$. One observes

$$Y_i = f_i + \frac{\tau}{\sqrt{n}} \varepsilon_i, \qquad i = 1, \dots, n,$$

where the ε_i's are i.i.d. standard Gaussian variables. In this framework we establish rates of testing relative to the Euclidean norm of f, say, $n\|f\|_n^2$.

COROLLARY 2. *Let α and β be two numbers in* $]0, 1[$ *and assume that*

$$n \geq 42 \max\left\{ \log\left(\frac{90}{\alpha}\right), \log\left(\frac{2}{\beta}\right) \right\}. \tag{13}$$

Let T_α *be the test statistic defined by* (2) *and* $\mathcal{E}_{p,s}(R)$ *the set defined by* (12). *Let* $s > (1/2 - 1/p)_+$ *and*

$$s' = \begin{cases} s - 1/2 + 1/p, & \text{if } p \geq 2, \\ s - 1/4 + 1/(2p), & \text{if } p < 2. \end{cases}$$

Then there exists a constant C *depending only on* α, β, s *and* p, *such that the following holds*:

(i) *For all* $p \geq 2$, *for all* $R > 0$ *such that* $R^2 \geq \sqrt{\log\log(n)}\tau^2/n$, *and for all* $f \in \mathcal{E}_{p,s}(R)$, *satisfying* $n\|f\|_n^2 \geq C\rho_{n,2}^2$ *with*

$$\rho_{n,2}^2 = \left[R^{2/(1+4s')}\left(\frac{\sqrt{\log\log(n)}}{n}\right)^{4s'/(1+4s')}\tau^2\right] \wedge \left[\frac{\tau^2}{\sqrt{n}}\right] + R^2 n^{-2s'} + \frac{\log\log n}{n}\tau^2,$$

we obtain

$$\mathbb{P}_f(T_\alpha > 0) \geq 1 - \beta.$$

(ii) *For all* $p \in]0, 2[$, $R \geq \tau$ *and for all* $f \in \mathcal{E}_{p,s}(R)$ *such that*

$$n\|f\|_n^2 \geq C\left\{\left[\left(R^{4/p}\tau^{2-p}\right)^{1/(1+4s')} l_n^{(1/p-1/2)/(1+4s')}\left(\frac{\tau^2\sqrt{\log\log(n)}}{n}\right)^{4s'/(1+4s')}\right] \wedge \left[\frac{\tau^2}{\sqrt{n}}\right] + R^2(\log(n))^{1-p/2} n^{-ps-1+p/2} + \frac{\log\log(n)}{n}\tau^2\right\},$$

where $l_n = \log(n)/\log\log(n)$, *we obtain*

$$\mathbb{P}_f(T_\alpha > 0) \geq 1 - \beta.$$

COMMENTS. Before coming to the comments on the rates given in Corollary 2, we mention that in case (ii) the condition $R \geq \tau$ could be weakened at the price of more technicalities in the proof.

Let us now turn to these comments. In the statistical framework considered here, the problem of giving minimax rates of testing under no prior knowledge on τ remains, to our knowledge, open. In the sequel, we shall compare the uniform rates of testing established in Corollary 2 to lower bounds which were established when τ is known. As we shall see those rates coincide [up to a possible $\log\log(n)$ or $\log(n)$ factor] when s is large enough, showing thus that the testing procedure is rate optimal in those cases. However, for small values of s, those rates differ and we do not know whether the procedure is rate optimal or not.

When $p \geq 2$ and $s' > 1/4$, the rate of testing is of order $(\frac{\sqrt{\log\log(n)}}{n})^{2s'/(1+4s')}$ which is the minimax rate of testing up to a logarithmic factor; see Lepski and Spokoiny (1999) and Ingster and Suslina (1998).

When $p = 2$, it was shown by Spokoiny (1996) that the extra logarithmic factor is due to the adaptive property of the test. Moreover, the rate of testing

$R^{2/1+4s}(\sqrt{\log\log n/n})^{4s/(1+4s)}$ corresponds to the rate obtained by Laurent and Massart (2000) for estimating $n\|f\|_n^2$ in the Gaussian sequence model.

When $p < 2$ and $s > 1/2 - 1/(2p)$, one obtains the rate $n^{-2s'/(1+4s')}$ up to a logarithmic factor. The rate $n^{-2s'/(1+4s')}$ is known to be minimax for Besov bodies from Lepski and Spokoiny (1999) and Ingster and Suslina (1998). The rate we get differs from that obtained by Spokoiny (1996) by a $\log(n)$ factor.

When $p < 2$ and $s < 1/2 - 1/(2p)$, or when $p \geq 2$ and $s < 1/4$, the rate of testing we get is of order $n^{-(p\wedge 2)(s+1/2-1/p)}$ [up to a $\log(n)$ factor]. We do not know if this rate is optimal or not.

5. Simulation studies in the functional regression model. We carry out a simulation study in order to evaluate the performance of our procedure both when the errors are normally distributed and when they are not. We compare its performance with those for the testing procedures proposed by Horowitz and Spokoiny (2001), Eubank and Hart (1992) and Dette and Munk (1998). We consider three simulation experiments.

1. The first simulation experiment was performed by Dette and Munk (1998). They considered standard normally distributed errors and five regression functions F,

$$F_c(x) = 1 + c\cos(10\pi x),$$

   for $c = 0$, 0.25, 0.5, 0.75, 1. We test that the function F is constant at level $\alpha = 5\%$; the null hypothesis is "f belongs to V" where V equals V_{cste}, the linear space of $\mathbb{R}^n$ with dimension $d = 1$ spanned by the vector $(1, \ldots, 1)^T$. The number of observations n equals 100, and for all $i = 1, \ldots, n$, $x_i = (i - 0.5)/n$.
2. The second simulation experiment was performed by Horowitz and Spokoiny (2001). They considered three distributions of the errors ε_i, $i = 1, \ldots, n$.
   (a) The Gaussian distribution: $\varepsilon_i \sim \mathcal{N}(0, 4)$.
   (b) The mixture of Gaussian distributions: ε_i is distributed as $\pi X_1 + (1-\pi)X_2$ where π is distributed as a Bernoulli variable with expectation 0.9, X_1 and X_2 are centered Gaussian variables with variance, respectively, equal to 2.43 and 25, π, and X_1 and X_2 are independent. This distribution has heavy tails.
   (c) The Type I distribution: ε_i has density $(s/2)f_X(\mu + (s/2)x)$ where $f_X(x) = \exp\{-x - \exp(-x)\}$, and where μ and s^2 are the expectation and the variance of a variable X with density f_X. These ε_i's are centered variables with variance 4. This distribution is asymmetrical.

   By considering the distributions (b) and (c), we evaluate the robustness of our procedure with respect to the Gaussian assumption.

   Three regression functions F are considered:

$$F_0(x) = 1 + x,$$

$$F_\tau(x) = 1 + x + \frac{5}{\tau}\phi\left(\frac{x}{\tau}\right) \qquad \text{with } \tau = 0.25 \text{ and } \tau = 1,$$

where ϕ is the density of a standard Gaussian variable. When $\tau = 0.25$ the regression function F_τ presents a peak, and when $\tau = 1$, F_τ presents a small bump. We test the linearity of the function F at level $\alpha = 5\%$: the null hypothesis is "f belongs to V" where V equals V_{lin}, the linear space of $\mathbb{R}^n$ with dimension $d = 2$ spanned by the vectors $(1, \ldots, 1)^T$ and $(x_1, \ldots, x_n)^T$.

The number n of observations equals 250. The x_i's are simulated once and for all as centered Gaussian variables with variance equal to 25 and are constrained to lie in the interval $[\Phi^{-1}(0.05), \Phi^{-1}(0.95)]$, where Φ is the distribution function of a standard Gaussian variable.

3. The third simulation experiment is similar to the first one except that we consider four regression functions F:

$$F_\delta(x) = -\delta(x - 0.1)\mathbb{1}_{x \le 0.1}$$

for $\delta = 0$, 20, 30, 40.

5.1. *Piecewise constant functions and trigonometric polynomials.* The testing procedure depends on the choice of the collections $\{S_m, m \in \mathcal{M}\}$ and $\{\alpha_m, m \in \mathcal{M}\}$.

The collection $\{S_m, m \in \mathcal{M}\}$. We consider the spaces S_m based on piecewise functions and trigonometric polynomials and two collections of indices $\mathcal{M}$. More precisely, for each $k = 1, \ldots, n$, we consider the spaces $S'_{(k,\text{pc})}$ of piecewise constant functions based on the intervals $\{](l-1)/k, l/k], l = 1, \ldots, k\}$ and $S'_{(k,\text{tp})}$ the space of trigonometric polynomials of degree not larger than k. For each $\delta \in \{\text{pc}, \text{tp}\}$, and for each $k = 1, \ldots, n$, $S_{(k,\delta)}$ is the linear subspace of $\mathbb{R}^n$ defined as the orthogonal projection of $\{(F(x_1), \ldots, F(x_n))^T, F \in S'_{(k,\delta)}\}$ onto $V^\perp$. Now let us set $J_n = [\log(n/2)/\log(2)]$, and define for $\delta \in \{\text{pc}, \text{tp}\}$,

$$\mathcal{M}_{\text{dya},\delta} = \{k \in \{1, \ldots, n\}, \dim(S_{(k,\delta)}) \in \{2^j, 0 \le j \le J_n\}\},$$

$$\mathcal{M}_{\text{all},\delta} = \{k \in \{1, \ldots, n\}, \dim(S_{(k,\delta)}) \in \{1, 2, 3, \ldots, 2^{J_n}\}\},$$

$$\mathcal{M}_{\text{dya}} = \mathcal{M}_{\text{dya,pc}} \cup \mathcal{M}_{\text{dya,tp}}, \quad \text{and} \quad \mathcal{M}_{\text{all}} = \mathcal{M}_{\text{all,pc}} \cup \mathcal{M}_{\text{all,tp}}.$$

The collection $\{\alpha_m, m \in \mathcal{M}\}$. We consider the procedures P1 and P2 defined in Section 2.1. The quantity a_n is calculated by simulation. When we are using the procedure P2, the α_m's equal $\alpha/|\mathcal{M}|$ where $|\mathcal{M}|$ denotes the cardinality of $\mathcal{M}$. The test statistic is equal to T_α defined by (2) and is denoted $T_{\text{P},\mathcal{M}}$ with $\text{P} \in \{\text{P1}, \text{P2}\}$ and

$$\mathcal{M} \in \{\mathcal{M}_{\text{dya,pc}}, \mathcal{M}_{\text{dya,tp}}, \mathcal{M}_{\text{all,pc}}, \mathcal{M}_{\text{all,tp}}, \mathcal{M}_{\text{dya}}, \mathcal{M}_{\text{all}}\}.$$

Let us now comment on these choices.

1. For a given $\mathcal{M}$, the test based on $T_{\mathrm{P1},\mathcal{M}}$ is more powerful than the test based on $T_{\mathrm{P2},\mathcal{M}}$. This comes from the fact that a_n is not smaller than the $\alpha/|\mathcal{M}|$. Since the calculation of a_n requires additional computation, it is worth comparing the procedures P1 and P2 in the simulation study.
2. We introduce two collections $\mathcal{M}_{\mathrm{dya}}$ and $\mathcal{M}_{\mathrm{all}}$ in order to evaluate the dependency of our testing procedure with respect to the size of the collection. We restrict ourselves to the procedure P1 for which both tests are of size α.
3. As already mentioned our procedure is driven by the idea that if F is close to one of the functional spaces $S'_{(k,\delta)}$ the test rejects the null with probability close to 1. Because we have no prior information on F, we propose mixing several kinds of functional spaces as our procedure permits. We could also mix linear spaces based on polynomials, piecewise polynomials, wavelets and so on.

5.2. *Description of other procedures.* Let us now describe the procedures with which we compare ours.

The test proposed by Horowitz and Spokoiny (2001). They proposed an adaptive procedure for testing that the regression function belongs to some parametric family of functions. Their procedure rejects the null hypothesis if for some bandwidth among a grid, the distance between the nonparametric kernel estimator and the kernel smoothed parametric estimator of F under the null hypothesis is large. The quantiles of their test statistic are estimated by a bootstrap method.

The test proposed by Eubank and Hart (1992). They proposed a test based on the penalized criterion defined by (7). Let us recall that the null hypothesis is rejected if $\widehat{m}$, the maximizer of $\mathrm{Crit}(m)$, is greater than 1. They proposed to choose the quantities Π_m and $\mathrm{pen}(m)$ as follows.

1. In the first simulation experiment, where the null hypothesis is $f \in V_{\mathrm{cste}}$, Π_m is the projector on the space generated by the cosine functions $\{\sqrt{2}\cos(\pi j x),\ j = 1,\dots,m\}$ calculated in $x_i = (i-0.5)/n$ for $i = 1,\dots,n$.

   In the second simulation experiment, where the null hypothesis is $f \in V_{\mathrm{lin}}$, Π_m is the projector on the space generated by the m last columns of P_m, where P_m is the matrix of ortho-normalized polynomials $1, x, x^2, \dots, x^{m+1}$ with respect to the design points.
2. The penalty function $\mathrm{pen}(m)$ is defined by $\mathrm{pen}(m) = \widehat{\sigma}^2 C_\alpha m/n$ where $\widehat{\sigma}^2$ is the variance estimator proposed by Gasser, Sroka and Jennen-Steinmetz (1986) and $C_\alpha = 4.18$. This value of C_α, given by Eubank and Hart (1992), ensures that the test is asymptotically of level $\alpha = 5\%$.

The test proposed by Dette and Munk (1998). They proposed a test based on estimation of the minimal distance between f and V_{cste}, say $\widehat{M}_n^2$. In the particular case considered in the second simulation study, the null hypothesis is rejected if

$$\sqrt{\frac{9n}{17}}\frac{\widehat{M}_n^2}{\widehat{\sigma}^2} > \bar{\Phi}^{-1}(\alpha),$$

where, for some (possibly negative) weights w_i, $i = 1, \dots, n$, satisfying $\sum_{i=1}^n w_i = n$,

$$\widehat{M}_n^2 = \frac{1}{n}\sum_{i=1}^n w_i Y_i^2 - \frac{n-1}{n}\widehat{\sigma}^2 - \left|\frac{1}{n}\sum_{i=1}^n w_i Y_i\right|^2.$$

5.3. *Results of the simulations.* The results of the first simulation study are given in Table 1. The percentages of rejection of the null hypothesis given under the columns DM-test and EH-test are reported from the paper by Dette and Munk (1998). They used 5000 simulations for estimating the percentages of rejection. The results corresponding to our procedure are based on 4000 simulations. We compare procedures P1 and P2 when $\mathcal{M}$ equals $\mathcal{M}_{\text{dya}}$ and we evaluate the effect of the size of the collection $\mathcal{M}$ when the chosen procedure is P1.

The results of the second simulation study are given in Table 2. The percentages of rejection of the null hypothesis given under the column HS-test are reported from the paper by Horowitz and Spokoiny (2001). They used 1000 simulations for estimating the level of the test and 250 simulations for estimating the power. The results corresponding to the procedure proposed by Eubank and Hart (1992) and to our procedure are based on 4000 simulations. We evaluate the performance of the procedures P1 and P2 for non-Gaussian errors when $\mathcal{M} = \mathcal{M}_{\text{dya}}$.

TABLE 1
First simulation study: percentages of rejection and value of the signal/noise ratio. The nominal level of the test is $\alpha = 5\%$

Null hypothesis is true					
c	**DM-test**	**EH-test**	$\mathbf{T}_{\mathbf{P1},\mathcal{M}_{\text{all}}}$	$\mathbf{T}_{\mathbf{P1},\mathcal{M}_{\text{dya}}}$	$\mathbf{T}_{\mathbf{P2},\mathcal{M}_{\text{dya}}}$
0	0.044	0.056	0.050	0.053	0.024

Null hypothesis is false						
c	**signal/noise**	**DM-test**	**EH-test**	$\mathbf{T}_{\mathbf{P1},\mathcal{M}_{\text{all}}}$	$\mathbf{T}_{\mathbf{P1},\mathcal{M}_{\text{dya}}}$	$\mathbf{T}_{\mathbf{P2},\mathcal{M}_{\text{dya}}}$
0.25	0.18	0.07	0.06	0.10	0.08	0.05
0.50	0.35	0.19	0.10	0.36	0.30	0.23
0.75	0.53	0.49	0.40	0.83	0.74	0.66
1	0.71	0.82	0.86	0.99	0.98	0.96

TABLE 2

Second simulation study: percentages of rejection and value of the signal/noise ratio equal to $d_n(f, V)/\sigma$. The nominal level of the test is $\alpha = 5\%$

Null hypothesis is true				
Error distribution	**HS-test**	**EH-test**	$\mathbf{T}_{\mathbf{P1},\mathcal{M}_{\mathbf{dya}}}$	$\mathbf{T}_{\mathbf{P2},\mathcal{M}_{\mathbf{dya}}}$
Normal	0.066	0.054	0.054	0.036
Mixture	0.054	0.058	0.071	0.055
Type I	0.055	0.058	0.060	0.040
Null hypothesis is false: $\tau = 1$, signal/noise $= 0.31$				
Error distribution	**HS-test**	**EH-test**	$\mathbf{T}_{\mathbf{P1},\mathcal{M}_{\mathbf{dya}}}$	$\mathbf{T}_{\mathbf{P2},\mathcal{M}_{\mathbf{dya}}}$
Normal	0.79	0.87	0.96	0.94
Mixture	0.80	0.88	0.95	0.93
Type I	0.82	0.89	0.96	0.94
Null hypothesis is false: $\tau = 0.25$, signal/noise $= 0.69$				
Error distribution	**HS-test**	**EH-test**	$\mathbf{T}_{\mathbf{P1},\mathcal{M}_{\mathbf{dya}}}$	$\mathbf{T}_{\mathbf{P2},\mathcal{M}_{\mathbf{dya}}}$
Normal	0.92	0.95	1	1
Mixture	0.93	0.94	1	1
Type I	0.97	0.95	1	1

In Table 3 we study the effect of mixing two collections of linear subspaces. This comparison is based on the three simulation studies with Gaussian errors. We compare the percentages of rejection obtained for the procedure P1 when we consider in the collection either the piecewise constant functions, or the trigonometric polynomials, or a mixture of these collections.

As expected, under the null hypothesis, the percentage of rejection equals 5% when we use the procedure P1 and is less than 5% when we use the procedure P2. For the error distributions "mixture" and "Type I," the levels of both procedures increase slightly; see Table 2. As expected, the conservative procedure P2 is more robust than P1.

In Table 1 the test based on $T_{\mathrm{P1},\mathcal{M}_{\mathrm{all}}}$ has the greatest power. The superiority of $T_{\mathrm{P1},\mathcal{M}_{\mathrm{all}}}$ over $T_{\mathrm{P1},\mathcal{M}_{\mathrm{dya}}}$ is significant for $c = 0.5$ and $c = 0.75$, where the power of the tests is far from 1.

The test based on the statistic $T_{\mathrm{P2},\mathcal{M}_{\mathrm{dya}}}$ has the smallest power among the procedures we proposed, but is more powerful than the procedures proposed by Eubank and Hart (1992), Horowitz and Spokoiny (2001) and Dette and Munk (1998); see Tables 1 and 2. Though the testing procedure based on the statistic $T_{\mathrm{P2},\mathcal{M}_{\mathrm{dya}}}$ is conservative, it seems to be a good compromise for practical issues; its implementation is very easy and its performance is good.

TABLE 3

Effect of mixing two collections: percentages of rejection and value of the signal/noise ratio. The nominal level of the test is $\alpha = 5\%$

First simulation study				
c	**signal/noise**	$\mathbf{T}_{\mathbf{P1},\mathcal{M}_{\mathbf{dya,pc}}}$	$\mathbf{T}_{\mathbf{P1},\mathcal{M}_{\mathbf{dya,tp}}}$	$\mathbf{T}_{\mathbf{P1},\mathcal{M}_{\mathbf{dya}}}$
0	0	0.054	0.054	0.053
0.25	0.18	0.08	0.08	0.08
0.50	0.35	0.22	0.32	0.30
0.75	0.53	0.58	0.78	0.74
1	0.71	0.90	0.99	0.98
Second simulation study				
F	**signal/noise**	$\mathbf{T}_{\mathbf{P1},\mathcal{M}_{\mathbf{dya,pc}}}$	$\mathbf{T}_{\mathbf{P1},\mathcal{M}_{\mathbf{dya,tp}}}$	$\mathbf{T}_{\mathbf{P1},\mathcal{M}_{\mathbf{dya}}}$
$1+x$	0	0.048	0.045	0.054
$F_1(x)$	0.31	0.96	0.94	0.96
Third simulation study				
$\boldsymbol{\delta}$	**signal/noise**	$\mathbf{T}_{\mathbf{P1},\mathcal{M}_{\mathbf{dya,pc}}}$	$\mathbf{T}_{\mathbf{P1},\mathcal{M}_{\mathbf{dya,tp}}}$	$\mathbf{T}_{\mathbf{P1},\mathcal{M}_{\mathbf{dya}}}$
0	0	0.050	0.049	0.053
20	0.35	0.36	0.31	0.36
30	0.53	0.78	0.69	0.77
40	0.70	0.97	0.94	0.97

In Table 3 we see that the power of the test based on $T_{P1,\mathcal{M}_{\rm dya}}$ is equal (or nearly equal) to the largest power of the tests based on $T_{P1,\mathcal{M}_{\rm dya,pc}}$ and $T_{P1,\mathcal{M}_{\rm dya,tp}}$. Thus, by mixing several collections of linear spaces, one can improve the performance of the test. This should in fact be recommended in practice.

6. Proof of Theorem 1. For the sake of simplicity and to keep our formulae as short as possible we assume that $\sigma^2 = 1$ and work with the Euclidean norm of $\mathbb{R}^n$, denoted by $\|\cdot\|$, instead of $\|\cdot\|_n = \|\cdot\|/\sqrt{n}$. We shall denote by $d(\cdot,\cdot)$ the distance $\sqrt{n}d_n(\cdot,\cdot)$.

By definition of T_α, for any $f \in \mathbb{R}^n$, $\mathbb{P}_f(T_\alpha \le 0) \le \inf_{m\in\mathcal{M}} P_f(m)$ where

$$P_f(m) = \mathbb{P}_f\left(\frac{N_m\|\Pi_m Y\|^2}{D_m\|Y - \Pi_{V_m}Y\|^2} \le \bar{F}^{-1}_{D_m,N_m}(\alpha_m)\right).$$

Let us denote by $Q(a, D, u)$ the $1-u$ quantile of a noncentral χ^2 variable with D degrees of freedom and noncentrality parameter a. For each $m \in \mathcal{M}$, the random variables $\|\Pi_m Y\|^2$ and $\|Y - \Pi_{V_m}Y\|^2$ are distributed as noncentral χ^2 variables with, respectively, D_m and N_m degrees of freedom and noncentrality parameters

$\|\Pi_m f\|^2$ and $\|f - \Pi_{V_m} f\|^2$. Hence,

$$P_f(m) \le \mathbb{P}_f\left(\|\Pi_m Y\|^2 \le \frac{D_m}{N_m}\bar{F}^{-1}_{D_m,N_m}(\alpha_m) Q\left(\|f - \Pi_{V_m} f\|^2, N_m, \frac{\beta}{2}\right)\right) + \frac{\beta}{2}$$

and therefore, $\mathbb{P}_f(T_\alpha \le 0) \le \beta$ if for some m in $\mathcal{M}$,

$$\frac{D_m}{N_m}\bar{F}^{-1}_{D_m,N_m}(\alpha_m) Q\left(\|f - \Pi_{V_m} f\|^2, N_m, \frac{\beta}{2}\right) \le Q\left(\|\Pi_m f\|^2, D_m, 1 - \frac{\beta}{2}\right). \tag{14}$$

Thanks to Lemma 1 in Birgé (2001) we know that the following inequalities hold for all $u \in]0, 1[$:

$$Q(a, D, u) \le D + a + 2\sqrt{(D + 2a)\log(1/u)} + 2\log(1/u), \tag{15}$$

$$Q(a, D, 1 - u) \ge D + a - 2\sqrt{(D + 2a)\log(1/u)}. \tag{16}$$

By using that $k_m \ge 2$ and the inequalities $\sqrt{u + v} \le \sqrt{u} + \sqrt{v}$, $2\sqrt{uv} \le \theta u + \theta^{-1} v$, which hold for all positive numbers u, v, θ, we derive that

$$Q\left(\|f - \Pi_{V_m} f\|^2, N_m, \frac{\beta}{2}\right) \le N_m + 2\|f - \Pi_{V_m} f\|^2 + 2\sqrt{N_m L} + 2k_m L \tag{17}$$

and

$$Q\left(\|\Pi_m f\|^2, D_m, 1 - \frac{\beta}{2}\right) \ge D_m + \frac{4}{5}\|\Pi_m f\|^2 - 2\sqrt{D_m L} - 10L. \tag{18}$$

Since $\|\Pi_m f\|^2 = d^2(f, V) - \|f - \Pi_{V_m} f\|^2$, (14) is satisfied as soon as

$$\begin{aligned} d^2(f, V) \ge \delta_m(f) = {} & \left(1 + 2.5\frac{D_m}{N_m}\bar{F}^{-1}_{D_m,N_m}(\alpha_m)\right)\|f - \Pi_{V_m} f\|^2 \\ & + \frac{5}{4} D_m \bar{F}^{-1}_{D_m,N_m}(\alpha_m)\left(1 + 2\sqrt{\frac{L}{N_m}} + 2k_m\frac{L}{N_m}\right) \\ & - \frac{5}{4}(D_m - 2\sqrt{D_m L} - 10L). \end{aligned}$$

Hence, it remains to show that $\delta_m(f) \le n\Delta(f, m)$. Let us now set $r_m = D_m/N_m$. To bound $D_m \bar{F}^{-1}_{D_m,N_m}(\alpha_m)$ we use the following lemma proved in Section 9.

LEMMA 1. *Let $u \in]0, 1[$ and $\bar{F}^{-1}_{D,N}(u)$ be the $1 - u$ quantile of a Fisher random variable with D and N degrees of freedom. Then we have*

$$\begin{aligned} D\bar{F}^{-1}_{D,N}(u) \le {} & D + 2\sqrt{D\left(1 + \frac{D}{N}\right)\log\left(\frac{1}{u}\right)} \\ & + \left(1 + 2\frac{D}{N}\right)\frac{N}{2}\left[\exp\left(\frac{4}{N}\log\left(\frac{1}{u}\right)\right) - 1\right]. \end{aligned} \tag{19}$$

By using the inequality, $\exp(u) - 1 \le u\exp(u)$ which holds for all $u > 0$, we derive from (19) that

$$D_m \bar{F}^{-1}_{D_m,N_m}(\alpha_m) \le D_m + 2\sqrt{D_m L_m}(1+\sqrt{r_m}) + k_m(1+2r_m)L_m$$

$$\le D_m\left(1 + 2\sqrt{\frac{L_m}{N_m}} + 2k_m\frac{L_m}{N_m}\right) + 2\sqrt{D_m L_m} + k_m L_m \tag{20}$$

$$\le [1 + K_m(L_m) \vee k_m](D_m + L_m). \tag{21}$$

Using now (17), (18), (20), (21) and the inequality $\lambda\sqrt{u} + \gamma\sqrt{v} \le \sqrt{\lambda^2+\gamma^2} \times \sqrt{u+v}$, which holds for any positive numbers λ, γ, u, v, we get

$$\begin{aligned}
\frac{4}{5}\delta_m(f) &\le \frac{4}{5}[1+\Lambda_1(m)]\|f - \Pi_{V_m} f\|^2 - (D_m - 2\sqrt{D_m L} - 10L)\\
&\quad + \big(D_m + 2\sqrt{D_m L_m}(1+\sqrt{r_m}) + k_m(1+2r_m)L_m\big)\\
&\quad \times \left(1 + 2\sqrt{\frac{L}{N_m}} + 2k_m\frac{L}{N_m}\right)\\
&\le \frac{4}{5}[1+\Lambda_1(m)]\|f - \Pi_{V_m} f\|^2 + D_m + 2\sqrt{D_m}\sqrt{r_m L} + 2k_m r_m L\\
&\quad + 2K_m(L)(1+\sqrt{r_m})\sqrt{D_m L_m} + K_m(L)k_m(1+2r_m)L_m\\
&\quad - D_m + 2\sqrt{D_m L} + 10L\\
&\le \frac{4}{5}[1+\Lambda_1(m)]\|f - \Pi_{V_m} f\|^2\\
&\quad + 2\sqrt{1+K_m^2(L)}(1+\sqrt{r_m})\sqrt{D_m(L_m+L)}\\
&\quad + 2[(k_m K_m(L)/2) \vee 5](1+2r_m)(L_m+L).
\end{aligned}$$

Hence, $\delta_m(f) \le n\Delta(f,m)$ and the first part of Theorem 1 follows. The inequalities on $\Lambda_i(m)$'s for $i = 1, 2, 3$ derive easily from the definition of the Λ_i's noting that under $(\mathrm{H}_{\mathcal{M}})$ we have that $L_m \le N_m/10$ and $L \le N_m/21$ for all $m \in \mathcal{M}$.

7. Proof of Proposition 1. In order to avoid technicalities, we shall only prove this result in the case where σ^2 is known and equals 1 by considering the test statistic

$$\hat{T}_m = \hat{T}_m(Y) = \|\Pi_m Y\|_n^2 - \frac{1}{n}\bar{\chi}_m^{-1}(\alpha).$$

The proof of the result for the test statistic T_m is based on analogous arguments.

Let us prove that for all $f \in \mathbb{R}^n$ such that

$$\|f\|_n^2 \le \|f - \Pi_m f\|_n^2 + \theta(\alpha,\beta)\sqrt{D_m}\frac{\sigma^2}{n}, \tag{22}$$

we have $\mathbb{P}_f(T_m \le 0) \ge \beta$.

By using Proposition 1 in Baraud (2000b), we know that for all $\rho \le \theta(\alpha,\beta)\sqrt{D_m}$ there exists some $g_m \in S_m$ such that $n\|g_m\|_n^2 = \rho$ and $\mathbb{P}_{g_m}(\widehat{T}_m < 0) \ge \beta$. For any orthogonal transformation U keeping S_m invariant, Π_m and U commute leading to the identity $\widehat{T}_m(Y) = \widehat{T}_m(UY)$. Thus, for all $f_m \in S_m$ such that $n\|f_m\|_n^2 = \rho$, we have

$$\mathbb{P}_{g_m}(\widehat{T}_m \le 0) = \mathbb{P}_{f_m}(\widehat{T}_m \le 0) \ge \beta.$$

Now for all $f \in \mathbb{R}^n$ satisfying (22), we have that

$$n\|\Pi_m f\|_n^2 = n\|f\|_n^2 - n\|f - \Pi_m f\|_n^2 \le \theta(\alpha,\beta)\sqrt{D_m}$$

and therefore

$$\mathbb{P}_f(\widehat{T}_m \le 0) = \mathbb{P}_{\Pi_m f}(\widehat{T}_m \le 0) \ge \beta.$$

8. Proof of Corollaries 1 and 2. Throughout this section, C denotes some constant that may vary from line to line. The dependency of C with respect to various quantities is specified by the notation $C(\cdot)$.

8.1. *Proof of Corollary* 1(i). The proof is divided into consecutive steps.

Step 1. There exists a universal constant Λ such that for all $m \in \mathcal{M}_1 \times \{1\}$ and for all $i = 1, 2, 3$, $\Lambda_i(m) \le \Lambda$ (one can take $\Lambda = 37.5$).

Let us first check that condition ($\mathrm{H}_{\mathcal{M}}$) in Theorem 1 holds. Since

$$|\mathcal{M}_1| \le 1 + \log(n/2)/\log(2) \le \log(n)/\log(2),$$

we have for all $m \in \mathcal{M}_1 \times \{1\}$,

$$\alpha_m = \frac{\alpha}{2|\mathcal{M}_1|} \ge \frac{\alpha\log(2)}{2\log(n)}. \tag{23}$$

Since $D_m \le n/2$, condition ($\mathrm{H}_{\mathcal{M}}$) therefore holds as soon as

$$\alpha \ge 2\log(n)\exp(-n/21)/\log(2) \quad \text{and} \quad \beta \ge 2\exp(-n/42).$$

Clearly, the latter inequality is fulfilled under (11). As for the former, noticing that for all $x \ge 1$, $\log(x) \le x \le 42\exp(x/42 - 1)$, we obtain that

$$2\frac{\log(n)}{\log(2)}\exp(-n/21) \le \frac{84}{e\log(2)}\exp(-n/42) \le 45\exp(-n/42),$$

which leads to the result under (11). Since condition ($\mathrm{H}_{\mathcal{M}}$) holds we obtain the result claimed in Step 1 by using the inequalities on $\Lambda_i(m)$ for $i = 1, 2, 3$ and the fact that for all $m \in \mathcal{M}$, $D_m \le N_m$ and $L_m \le N_m/10$.

Step 2. For all sequences of points $(x_1, x_2, \ldots, x_n) \in [0,1]^n$, $f \in \mathcal{H}_s(R)$ and $k \ge 1$,

$$d_n^2(f, S_{(k,1)}) \le R^2 k^{-2s}. \tag{24}$$

Let us set $I_{j,k} = \{i,\ x_i \in](j-1)/k, j/k]\}$, $i(j,k) = \inf I_{j,k}$ when $I_{j,k} \neq \varnothing$ and f^k the $\mathbb{R}^n$ vector defined as $f_i^k = f_{i(j,k)} = F(x_{i(j,k)})$ if $i \in I_{j,k}$. Clearly f^k belongs to $S_{(k,1)}$ and we have for all $f \in \mathcal{H}_s(R)$,

$$
\begin{aligned}
d_n^2(f, S_{(k,1)}) &\le d_n^2(f, f^k)\\
&\le \frac{1}{n}\sum_{j=1}^{k}\sum_{i\in I_{j,k}} |F(x_i) - F(x_{i(j,k)})|^2 \le R^2 k^{-2s}.
\end{aligned}
$$

Step 3. For each $f \in \mathcal{H}_s(R)$ let

$$
\rho_n^2(f) = \inf_{k\in\mathcal{M}_1} \Delta\big(f, (k,1)\big), \tag{25}
$$

where for each $m \in \mathcal{M}$, $\Delta(f, m)$ is defined in Theorem 1. There exists a constant $C(\alpha,\beta)$ such that $\rho_n^2(f) \le C(\alpha,\beta)\rho_{n,1}^2$.

We deduce from (23) that for all $m \in \mathcal{M}_1 \times \{1\}$,

$$
\log\left(\frac{2}{\alpha_m \beta}\right) \le C(\alpha,\beta)\log\log(n).
$$

Setting $L_n = \log\log(n)$, we deduce from Step 1 and Step 2 that for all $f \in \mathcal{H}_s(R)$,

$$
C^{-1}(\alpha,\beta)\rho_n^2(f) \le \inf_{k\in\mathcal{M}_1}\left[R^2k^{-2s} + \sqrt{kL_n}\frac{\sigma^2}{n}\right] + L_n\frac{\sigma^2}{n}. \tag{26}
$$

Note that $R^2k^{-2s} \le \sqrt{kL_n}\sigma^2/n$ if and only if

$$
k \ge k^* = \left(\frac{R^2 n}{\sigma^2\sqrt{L_n}}\right)^{2/(1+4s)}.
$$

Under the assumption on R we know that $k^* \ge 1$. Let us now distinguish between two cases. If there exists $k' \in \mathcal{M}_1$ such that $k^* \le k'$ one can take $k' \le 2k^*$ and then

$$
\begin{aligned}
\inf_{k\in\mathcal{M}_1}\left(R^2k^{-2s} + \sqrt{kL_n}\frac{\sigma^2}{n}\right) &\le 2\sqrt{k'L_n}\frac{\sigma^2}{n}\\
&\le 2\sqrt{2}R^{2/(1+4s)}\left(\frac{\sqrt{\log\log(n)}}{n}\sigma^2\right)^{4s/(1+4s)}.
\end{aligned} \tag{27}
$$

Else, we take $k' \in \mathcal{M}_1$ such that $n/4 \le k' \le n/2$. Since $k' \le k^* \wedge n/2$ and since $s \le 1$, we obtain that

$$
\inf_{k\in\mathcal{M}_1}\left(R^2k^{-2s} + \sqrt{kL_n}\frac{\sigma^2}{n}\right) \le 2R^2k'^{-2s} \le 32R^2n^{-2s}. \tag{28}
$$

By gathering (26)–(28) we conclude the proof of Step 3. We conclude the proof of case (i) by using Theorem 1 and Step 3.

8.2. *Proof of Corollary* 1(ii). Let us set $f^n = f/\sqrt{n}$ and $r_n^2(k_0) = \sum_{i=1}^n \phi_{k_0}^2(x_i)/n$. Since

$$
\begin{aligned}
&n\|f^n\|_n^2 - nd_n^2\Big(f^n, S_{(k_0,2)}\Big)\\
&\quad = n\left(\frac{1}{nr_n(k_0)}\sum_{i=1}^n f_i^n\phi_{k_0}(x_i)\right)^2 = \frac{1}{r_n^2(k_0)}\left(\frac{1}{n}\sum_{i=1}^n F(x_i)\phi_{k_0}(x_i)\right)^2,
\end{aligned}
\tag{29}
$$

by using Theorem 1 and the fact that $d_n^2(f^n, S_{(k_0,2)}) \le \|f^n\|_n^2 = \|f\|_n^2/n$, it suffices to show that for n large enough the right-hand side of (29) is larger than

$$\Lambda_1(k_0,2)\|f\|_n^2 + \big(\Lambda_2(k_0,2) + \Lambda_3(k_0,2)\big)\sigma^2\log\left(\frac{2}{\alpha_{(k_0,2)}\beta}\right).$$

This inequality is clearly satisfied for n large enough since under the assumption of Corollary 1,

$$\lim_{n\to+\infty}\Lambda_1(k_0,2)\|f\|_n^2 = 0$$

and

$$\limsup_{n\to+\infty}\big(\Lambda_2(k_0,2)+\Lambda_3(k_0,2)\big) \le 17.5.$$

8.3. *Proof of Corollary* 2 *for* $p \ge 2$. In the sequel, we shall use the fact that there exists a universal constant Λ such that for all $m \in \mathcal{M}$ and $i = 1, 2, 3$, $\Lambda_i(m) \le \Lambda$. Such a result is obtained by using (13) and by arguing as in the proof of Corollary 1. We divide the proof into consecutive steps.

Step 1. For each $f \in \mathcal{E}_{p,s}(R)$ and each $k \in \mathcal{M}_1$,

$$nd_n^2(f, S_{(k,1)}) \le C(s,p)R^2k^{-2s'}.$$

By using Hölder's inequality we have that

$$
\begin{aligned}
nd_n^2(f, S_{k,1}) &= \sum_{i=k+1}^n f_i^2 i^{2s} i^{-2s} \le \left(\sum_{i=k+1}^n |f_i|^p i^{ps}\right)^{2/p}\left(\sum_{i=k+1}^n i^{-2qs}\right)^{1/q}\\
&\le R^2\left(\sum_{i=k+1}^n i^{-2qs}\right)^{1/q},
\end{aligned}
$$

where $1/q = 1 - 2/p$. The condition $s > 1/2 - 1/p$ ensures that the series $\sum_{i>k} i^{-2qs}$ converges and since

$$\sum_{i>k} i^{-2qs} \le \int_k^\infty \frac{dx}{x^{2qs}} \le \frac{k^{1-2qs}}{1-2qs}, \tag{30}$$

the result follows.

Step 2. For each $f \in \mathcal{E}_{p,s}(R)$ let

$$\rho_n^2(f) = \inf_{k \in \mathcal{M}_1} \Delta\big(f, (k, 1)\big).$$

There exists a constant $C = C(\alpha, \beta, s, p)$ such that $n\rho_n^2(f) \le C\rho_{n,2}^2$.

By using Theorem 1, we have that

$$C^{-1} n\rho_n^2(f) \le \inf_{k \in \mathcal{M}_1} \left(R^2 k^{-2s'} + \sqrt{k \log(1/\alpha_{(k,1)})}\frac{\tau^2}{n}\right) + \frac{\log\log(n)}{n}\tau^2.$$

On the one hand, arguing as in the proof of Corollary 1, case (i) with s' in place of s, $\sigma^2 = \tau^2/n$ and introducing

$$k^\star = \left(\frac{nR^2}{\tau^2\sqrt{L_n}}\right)^{2/(1+4s')} \ge 1$$

in place of k^* where σ has been replaced by τ, one obtains that

$$\begin{aligned} C^{-1} &\inf_{k \in \mathcal{M}_1} \left(R^2 k^{-2s'} + \sqrt{k \log(1/\alpha_{(k,1)})}\frac{\tau^2}{n}\right) \\ &\le R^2 n^{-2s'} + R^{2/(1+4s')}\left(\frac{\tau^2\sqrt{\log\log(n)}}{n}\right)^{4s'/(1+4s')}. \end{aligned}$$

On the other hand, choosing $k = k_0 = \sup \mathcal{M}_1$ ($n/4 \le k_0 \le n/2$, $\alpha_{(k_0,1)} = \alpha/4$), we obtain that

$$C^{-1} \inf_{k \in \mathcal{M}_1} \left(R^2 k^{-2s'} + \sqrt{k \log(1/\alpha_{(k_0,1)})\frac{\tau^2}{n}}\right) \le R^2 n^{-2s'} + \frac{\tau^2}{\sqrt{n}}.$$

By taking the infimum between those two bounds, the result follows.

8.4. *Proof of Corollary* 2 *for* $p < 2$. For all $k \in \{1, \dots, n\}$, $D_{(k,2)} = 1$, $\log(1/\alpha_{(k,2)}) = \log(2n/\alpha)$ and so

$$\Lambda_1(k, 2) \le 10\left(\frac{1}{n-1} + \frac{\log(2n/\alpha)}{n-1}\right) \le C(\alpha)\frac{\log(n)}{n}.$$

Hence for each f,

$$\begin{aligned} n\Delta\big(f, (k, 2)\big) &\le \big(1 + \Lambda_1(k, 2)\big)\sum_{i \ne k} f_i^2 + C(\alpha, \beta)\frac{\log(n)}{n}\tau^2 \\ &\le \sum_{i \ne k} f_i^2 + C_1(\alpha, \beta)\frac{\log(n)}{n}(n\|f\|_n^2 + \tau^2). \end{aligned}$$

Then we distinguish between two cases. Whether there exists some $k \in \{1, \dots, n\}$ such that

$$f_k^2 \ge C_1(\alpha, \beta)\frac{\log(n)}{n}(n\|f\|_n^2 + \tau^2),$$

then $\|f\|_n^2 \geq \Delta(f, (k, 2))$ and by Theorem 1 we get that

$$\mathbb{P}_f(T_\alpha > 0) \geq 1 - \beta.$$

Or for all $k \in \{1, \ldots, n\}$,

$$f_k^2 \leq C_1(\alpha, \beta) \frac{\log(n)}{n} (n\|f\|_n^2 + \tau^2).$$

By using the subadditivity of the function $x \mapsto x^{p/2}$ for $p < 2$ we obtain that for each $f \in \mathcal{E}_{p,s}(R)$,

$$n\|f\|_n^2 \leq \sum_{i=1}^n i^{2s} f_i^2 \leq \left(\sum_{i=1}^n i^{ps} |f_i|^p \right)^{2/p} \leq R^2$$

and therefore, τ^2 being smaller than R^2, we deduce that for all $i \in \{1, \ldots, n\}$,

$$f_i^2 \leq 2C_1(\alpha, \beta) R^2 \frac{\log(n)}{n}. \tag{31}$$

Thus, for any $f \in \mathcal{E}_{p,s}(R)$ satisfying (31) and for any $k \in \mathcal{M}_1$,

$$\begin{aligned}
&nd_n^2(f, S_{(k,1)}) \\
&\quad = \sum_{i>k} f_i^2 = \sum_{i>k} |f_i|^{2-p} |f_i|^p \\
&\quad \leq C(\alpha, \beta, p) \left(\frac{\log(n)}{n} \right)^{1-p/2} R^{2-p} \sum_{i>k} |f_i|^p i^{ps} i^{-ps} \\
&\quad \leq C(\alpha, \beta, p) R^2 \left(\frac{\log(n)}{n} \right)^{1-p/2} k^{-ps}.
\end{aligned}$$

We compute an upper bound for the quantity $n\rho_n^2(f)$ for those f satisfying (31) by arguing as in the proof of the case $p \geq 2$ with

$$k^\star = \left(\frac{R^2 n^{p/2}}{\tau^2 \sqrt{L_n}} (\log(n))^{1-p/2} \right)^{2/(1+2ps)}$$

and then obtain the desired result.

9. Proof of Lemma 1. Let $U_{D,N}$ be a Fisher random variable with D and N degrees of freedom. Let X_D and Y_N be independent random variables distributed as a χ^2 with respective degrees of freedom D and N. The random variable $DU_{D,N}$ has the same distribution as that of NX_D/Y_N. Let us fix $u > D$ and $\lambda \in [0, 1/(2N)[$. The computation of the Laplace transform of a χ^2 random variable allows one to derive the following inequalities:

$$
\begin{aligned}
\mathbb{P}[DU_{D,N} \geq u] &= \mathbb{P}[NX_D - uY_N > 0] \leq \mathbb{E}[\exp(\lambda N X_D - \lambda u Y_N)] \\
&= \exp\biggl[-\frac{D}{2}\log(1-2N\lambda) - \frac{N}{2}\log(1+2\lambda u)\biggr].
\end{aligned}
\tag{32}
$$

By minimizing the right-hand side of (32) with respect to λ (the minimum being achieved for $\lambda^* = (u-D)/(2u(N+D)) \in [0, 1/(2N)[$) we obtain that

$$\mathbb{P}[DU_{D,N} \geq u] \leq \exp[\phi_{D,N}(u)],$$

where

$$\phi_{D,N}(u) \leq \biggl[-\frac{N+D}{2}\log\Bigl(1+\frac{u-D}{N+D}\Bigr) + \frac{D}{2}\log\Bigl(1+\frac{u-D}{D}\Bigr)\biggr].$$

In the sequel we set

$$u(t) = D + 2\sqrt{D\Bigl(1+\frac{D}{N}\Bigr)t} + \frac{1}{2}\Bigl(1+2\frac{D}{N}\Bigr)N\bigl(\exp(4t/N)-1\bigr).$$

We obtain (19) by proving that for all $t \geq 0$, $\phi_{D,N}(u(t)) \leq -t$. Since $u(0) = D$, it suffices to show that for all $t > 0$, $u'(t)\phi'_{D,N}(u(t)) \leq -1$ and as $\phi'_{D,N}(u) = -N(u-D)/(2u(N+u))$ it remains to check that for all $t > 0$,

$$u'(t)\bigl(u(t)-D\bigr) \geq \frac{2}{N}u^2(t) + 2u(t). \tag{33}$$

Let us set $e(t) = \exp(4t/N) - 1$ and $r = D/N$. With this notation we have

$$
\begin{aligned}
u(t) - D &= 2\sqrt{Nr(1+r)t} + 0.5(1+2r)Ne(t), \\
u'(t) &= \sqrt{Nr(1+r)/t} + 2(1+2r)\bigl(1+e(t)\bigr), \\
\frac{2}{N}u^2(t) + 2u(t) &= \frac{2}{N}\bigl(Nr + 2\sqrt{Nr(1+r)t} + 0.5(1+2r)Ne(t)\bigr)^2 \\
&\quad + 2\bigl(Nr + 2\sqrt{Nr(1+r)t} + 0.5(1+2r)Ne(t)\bigr)
\end{aligned}
$$

and (33) becomes

$$\frac{N}{2}(1+2r)^2e^2(t) - 8r(1+r)t + \frac{N}{2}\sqrt{Nr(1+r)}(1+2r)\frac{e(t)}{\sqrt{t}} \geq 0, \tag{34}$$

the terms in $\sqrt{t}e(t), e(t), \sqrt{t}$ and the constant vanishing. Since for all t, $e(t) \geq 4t/N$ by setting $T = \sqrt{t}$, we obtain that (34) is satisfied if for all $T > 0$,

$$P(T) = \frac{4(1+2r)^2}{N}T^3 - 4r(1+r)T + (1+2r)\sqrt{Nr(1+r)} \geq 0.$$

The function P admits a minimum for $T^* = \sqrt{3^{-1}Nr(1+r)}/(1+2r)$ for which

$$P(T^*) = \frac{\sqrt{Nr(1+r)}}{1+2r}\left((1+2r)^2 - \frac{8r(1+r)}{3\sqrt{3}}\right) \geq 0,$$

which concludes the proof of Lemma 1.

REFERENCES

BARAUD, Y. (2000a). Model selection for regression on a fixed design. *Probab. Theory Related Fields* **117** 467–493.

BARAUD, Y. (2000b). Nonasymptotic minimax rates of testing in signal detection. Technical Report 00.25, Ecole Normale Supérieure, Paris.

BARAUD, Y., HUET, S. and LAURENT, B. (2001). Nonparametric smoothing and lack-of-fit tests. In *Goodness-of-fit Tests and Model Validity* (C. Huber-Carol, N. Balakrishnan, M. S. Nikulin and M. Mesbah, eds.) 193–204. Birkhäuser, Boston.

BIRGÉ, L. (2001). An alternative point of view on Lepski's method. In *State of the Art in Probability and Statistics* (*Leiden, 1999*) 113–133. IMS, Beachwood, OH.

BIRGÉ, L. and MASSART, P. (2001). Gaussian model selection. *J. Eur. Math. Soc.* **3** 203–268.

CASTELLAN, G. (2000). Density estimation via exponential model selection. Technical Report 00.25, Univ. Paris XI, Orsay.

CHEN, J.-C. (1994). Testing for no effect in nonparametric regression via spline smoothing techniques. *Ann. Inst. Statist. Math.* **46** 251–265.

DETTE, H. and MUNK, A. (1998). Validation of linear regression models. *Ann. Statist.* **26** 778–800.

EUBANK, R. L. and HART, J. D. (1992). Testing goodness-of-fit in regression via order selection criteria. *Ann. Statist.* **20** 1412–1425.

EUBANK, R. L. and LARICCIA, V. N. (1993). Testing for no effect in nonparametric regression. *J. Statist. Plann. Inference* **36** 1–14.

GASSER, T., SROKA, L. and JENNEN-STEINMETZ, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73** 625–633.

HÄRDLE, W. and MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947.

HART, J. D. (1997). *Nonparametric Smoothing and Lack-of-fit Tests*. Springer, New-York.

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.

INGLOT, T. and LEDWINA, T. (1996). Asymptotic optimality of data-driven Neyman's tests for uniformity. *Ann. Statist.* **24** 1982–2019.

INGSTER, YU. I. (1993). Asymptotically minimax testing for nonparametric alternatives I. *Math. Methods Statist.* **2** 85–114.

INGSTER, YU. I. and SUSLINA, I. A. (1998). Minimax detection of a signal for Besov bodies and balls. *Problems Inform. Transmission* **34** 48–59.

LAURENT, B. and MASSART, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** 1302–1338.

LEPSKI, O. V. and SPOKOINY, V. G. (1999). Minimax nonparametric hypothesis testing: The case of inhomogeneous alternative. *Bernoulli* **5** 333–358.

MÜLLER, H.-G. (1992). Goodness-of-fit diagnostics for regression models. *Scand. J. Statist.* **19** 157–172.

NISHII, R. (1988). Maximum likelihood principle and model selection when the true model is unspecified. *J. Multivariate Anal.* **27** 392–403.

SPOKOINY, V. G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24** 2477–2498.

STANISWALIS, J. G. and SEVERINI, T. A. (1991). Diagnostics for assessing regression models. *J. Amer. Statist. Assoc.* **86** 684–692.

Y. BARAUD
ECOLE NORMALE SUPERIEURE, DMA
45, RUE D'ULM
75230 PARIS CEDEX 05
FRANCE
E-MAIL: yannick.baraud@ens.fr

S. HUET
INRA, LABORATOIRE DE BIOMÉTRIE
78352 JOUY-EN-JOSAS CEDEX
FRANCE

B. LAURENT
UNIVERSITÉ PARIS-SUD
BAT. 45
91450 ORSAY CEDEX
FRANCE