TESTING A LINEAR HYPOTHESIS USING HAAR TRANSFORM

SPOKOINY, V.

Weierstrass Institute for Applied Analysis and Stochastics,
Mohrenstr. 39, 10117 Berlin

and

Institute for Information Transmission Problems,
Bolshoi Karetnyi 19, Moscow 101447, Russia

1991 *Mathematics Subject Classification.* 62H25; Secondary 62G10.

Key words and phrases. regression model, Haar basis, hypothesis of linearity, nonparametric alternative .

ABSTRACT. The paper is concerned with the problem of testing a linear hypothesis about regression function. New testing procedure based on the Haar transform is proposed which is adaptive to unknown smoothness properties of the underlying function. The results show that under mild conditions on the design and smoothness of the regression function, this procedure provides with the near optimal rate of testing.

1. Introduction

Suppose we are given data $(X_i, Y_i), i = 1, \dots, n$, with $X_i \in \mathbb{R}^1$, $Y_i \in \mathbb{R}^1$, obeying the regression equation

$$Y_i = f(X_i) + \xi_i \tag{1.1}$$

where f is an unknown regression function and ξ_i are random errors.

Statistical analysis for such model may focus on the qualitative features of the underlying function f. Particularly, the signal detection problem corresponds to testing the simple zero hypothesis that f is identically zero. Another typical example is connected with the hypothesis of linearity. More generally one may consider a parametric type hypothesis about f.

To be more definitive with our exposition, we restrict ourselves to the case of the hypothesis of linearity. Using the hypothesis testing framework, we test the null hypothesis $H_0 : f$ 'is linear' versus the alternative $H_1 : f$ 'is not linear'.

The problem of testing a simple or parametrically described hypothesis is one of the classical in statistical inference, see Neyman (1937), Mann and Wald (1942), Lehmann (1957). In the present paper, we follow the approach developed in Ingster (1982 through 1993), see also Lepski and Spokoiny (1995) and Spokoiny (1996a, 1996b). The key idea of this approach is to test the null hypothesis against as large as possible class of alternatives. This leads to considering a nonparametric alternative.

The problem of testing parametric versus nonparametric regression fit was considered in Härdle and Mammen (1993), see also Härdle, Spokoiny and Sperlich (1995). But the proposed in these papers testing procedure meet a crucial for applications problem: their parameters depend on unknown smoothness properties of the alternative. The problem of adaptive testing was considered in Spokoiny (1996a, 1996b). Some wavelet-based testing procedure was proposed which is shown to be optimal (in an adaptive sense) for a wide range of function smoothness classes. But in its turn, this procedure is described for an idealized 'signal + white noise' model and only the case of a simple null is considered.

The aim of this paper is to develop an adaptive testing method which allows nonregular design and non-simple null, and which is computationally simple and stable. The latter property is attained by using of the simplest wavelet basis, namely the Haar transform.

2. Model and testing problem

In what follows we consider the observation model (1.1) $Y_i = f(X_i) + \xi_i$. The random errors ξ_i are assumed to be independent standard normal, $\xi_i \sim \mathcal{N}(0,1)$. This assumption is made to simplify the exposition. It can be relaxed in a usual way.

Note that we do not assume a random, or equidistant, or regular design because such kind of assumption seems to be inadequate for a majority of particular problems arising in statistical practice. Later we will impose some rather mild conditions which can be instructively verified in applications.

Our aim is to analyze the function f by given data. More specifically, we wish to test the hypothesis of linearity $H_0 : f$ 'is linear' that is $f(x) = a + bx$ for some constants a, b.

Let ϕ be a test i.e. a measurable function of observations with two values $0, 1$. As usual, the event $\{\phi = 0\}$ is treated as we accept the hypothesis and $\phi = 1$ means that the hypothesis is rejected. We measure the quality of a test ϕ by the corresponding error probabilities of the first and second kinds. Let $\boldsymbol{P}_f$ denote the distribution of the data $Y_1, \ldots, Y_n$ for a fixed model function f, see (1.1). Let now f_0 denote a linear function. Then the error probability of the first kind at a point f_0 is the probability under f_0 to reject the hypothesis,

$$\alpha_{f_0}(\phi) = \boldsymbol{P}_{f_0}(\phi = 1).$$

Similarly one defines the error probability of the second kind. If the function f is not linear, then

$$\beta_f(\phi) = \boldsymbol{P}_f(\phi = 0).$$

We wish to construct such a test whose error probability of the first kind does not exceed a prescribed level α_0 and which is sensitive against as large class of alternatives f as possible. Of course, no test could be sensitive against all alternatives. We consider therefore the class of alternatives f separated from the set of linear functions with distance at least ϱ,

$$\inf_{a,b} \|f(\cdot) - a - b \cdot\| \geq \varrho. \tag{2.1}$$

Here $\|\cdot\|$ means the usual L_2-norm. This condition, for any $\varrho > 0$, is also not sufficient for consistent testing, see Burnashev (1979) or Ibragimov and Khasminski (1977). One has to impose additionally some regularity (or smoothness) conditions on the underlying function f. Some typical examples are considered in Ingster (1982, 1993), Lepski and Spokoiny (1995), Spokoiny (1996b) among others where f is supposed to belong to some Hölder, Sobolev or Besov ball $\mathcal{F}$. Under such an assumption, given $\varrho > 0$, one searches for such a test ϕ that $P_{f_0}(\phi = 1) \leq \alpha_0$ for each linear f_0 and $P_f(\phi = 0) \leq \beta_0$ for every f from $\mathcal{F}$ satisfying (2.1), where α_0 and β_0 are given positive constants. The minimal value of ϱ for which such a test exists, qualifies sensitivity of testing.

One may expect that increasing in the number n of observations results in improving the sensitivity. We assume therefore that the value ϱ depends on n, $\varrho = \varrho(n)$ and we study the problem in the asymptotic set-up when n tends to infinity. We will call this sequence $\varrho(n)$ *the optimal rate of testing.*

Below we aim both to describe the optimal rate of testing $\varrho(n)$ for the considered problem of testing a linear hypothesis and to construct tests ϕ_n^* satisfying the above constraints on the error probabilities of the first and second kind. It was shown in Ingster (1982), see also Ingster (1993), Lepski and Spokoiny (1995), that even in the case of a simple null, the optimal rate of testing and the structure of rate-optimal tests depend heavily on smoothness properties of the underlying function or, in the other words, on the parameter of the function class $\mathcal{F}$. Typically no such kind of information is available. Following to Spokoiny (1996a, 1996b) we consider this problem adaptively i.e. we aim to construct such tests which do not require to know smoothness parameters but still provide at least near optimal rate of testing. It turned out, see Spokoiny (1996a), that such an adaptive property results in loss of power (or sensitivity) of testing but inessentially, by some log log-factor.

Such a procedure is described in the next section. The asymptotic properties of this procedure are discussed in Section 4. The proofs are postponed to Section 5.

3. Testing procedure

We consider the usual univariate regression model

$$Y_i = f(X_i) + \xi_i, \qquad i = 1, \dots, n, \tag{3.1}$$

We are going to test the hypothesis that the function f is linear. The similar testing problem was considered in Spokoiny (1996a), see also Spokoiny (1996b). However, there are two essential points which make us to modify slightly the method of testing. First of all, the above mentioned papers deal with so called 'signal + white noise' model which is in its turn some idealization of the regression model with the uniform random design. In view of practical applications, it would be very desirable to relax the such an assumption. Secondly, we consider now the case of composite null hypothesis in place of a simple null. This also create some technical difficulties.

In Spokoiny (1996a, 1996b) some wavelet-based testing procedure was proposed. Now we apply Haar decomposition which can be viewed as a particular (and the simplest) case of the wavelet transform. Note that any other functional basis can be applied in place of the Haar basis. Our choice was motivated by simplicity of calculating the corresponding coefficients and by its expressive power.

To begin by, we recall the construction and the main properties of the Haar transform. By I we denote the multi-index $I = (j,k)$ with $j = 0,1,2,\dots$ and $k = 0,1,\dots,2^j - 1$. By $\mathcal{I}$ we denote the set of all such multi-indices.

Let now the function $\psi(t)$ be defined by

$$\psi(t) = \begin{cases} 0 & t < 0, \\ 1 & 0 \le t < 1/2, \\ -1 & 1/2 \le t < 1, \\ 0 & t > 1. \end{cases} \tag{3.2}$$

For every $I = (j,k)$, set

$$\psi_I(t) = 2^{j/2}\psi(2^j t - k). \tag{3.3}$$

Clearly the function ψ_I is supported to the interval $[2^{-j}k, 2^{-j}(k+1)]$. It is well known that each measurable function f can be decomposed in the following way

$$f(t) = c_0 + \sum_{I \in \mathcal{I}} c_I \psi_I(t). \tag{3.4}$$

This means that the problem of recovering the function f can be transformed to the problem of estimating the coefficients c_I by given data. Since we have only n observations, it makes no sense to estimate more (in order) than n coefficients. We restrict therefore the total number of considered levels j. Let some j_1 be fixed such that

$$2^{j_1+1} \le n.$$

Set

$$\mathcal{I}_j = \{(j,k), k = 0, 1, \ldots, 2^j - 1\}$$

for the index set corresponding to j th level. We change now the infinite decomposition (3.4) by the finite approximation $\sum_{I \in \mathcal{I}(j_1)} c_I \psi_I(t)$ where the index set $\mathcal{I}(j_1)$ contains all level sets $\mathcal{I}_j$ with $j \le j_1$. Taking into account the structure of the null hypothesis, we complement the set of functions $(\psi_I,\, I \in \mathcal{I}_j)$, $j \le j_1$, with two functions $\psi_0 \equiv 1$ and $\psi_1(t) = t$, that is

$$\mathcal{I}(j_1) = \{0, 1\} + \bigcup_{j=0}^{j_1} \mathcal{I}_j. \tag{3.5}$$

The idea of the proposed procedure is quite clear. One estimates first all the coefficients $(c_I,\, I \in \mathcal{I}(j_1))$ by data. If our function f is really linear, this means that all the coefficients c_I for $I \ne 0, 1$ should be zero. This is just what we wish to verify.

Before we begin with our procedure, let us note that the functions ψ_0 and ψ_I, $I \in \mathcal{I}$, form the ortonormal basis in $L_2[0,1]$ with respect to Lebesgue measure on $[0,1]$. When dealing with real data, we change the integral by the finite sum over design points. It may occur that these functions ψ_I are no more ortonormal and are not orthogonal to each other in $L_2(\mu_n)$, where μ_n is the empirical design measure, $\mu_n(A) = \sum_{i=1}^n \mathbf{1}(X_i \in A)$. To cope with this, we change the functions ψ_I by its standardized versions ψ'_I: for $I = (j,k)$,

$$\psi'_I(t) = \lambda_I^{-1} \psi(2^j t - k), \tag{3.6}$$

where ψ is due to (3.2) and the normalizer λ_I is defined by

$$\lambda_I^2 = \sum_{i=1}^n |\psi(2^j X_i - k)|^2. \tag{3.7}$$

(Recall that $d = 1$ and hence X_i takes values in the interval $[0,1]$.) Particularly, $\lambda_0^2 = n$, $\lambda_1^2 = (X_1^2 + \ldots + X_n^2)$, and

$$\lambda_I^2 = M_I = \#\{i : X_i \in [2^{-j}k, 2^{-j}(k+1))\}, \qquad I \in \mathcal{I}.$$

In the sequel, we approximate the function f by linear combinations of the functions ψ'_I, $I \in \mathcal{I}(j_1)$. Let g be a function observed at point $X_1, \ldots, X_n$. Define

$\|g\|_n$ by

$$\|g\|_n^2 = \sum_{i=1}^{n} g^2(X_i).$$

Determine a column-vector $\boldsymbol{\theta}^*(j_1) = (\theta_I^*,\, I \in \mathcal{I}(j_1))$ as a minimizer of the error of approximation,

$$\boldsymbol{\theta}^*(j_1) = \operatorname*{arginf}_{\boldsymbol{\theta}(j_1)} \|f - \sum_{I \in \mathcal{I}(j_1)} \theta_I \psi_I'\|_n. \tag{3.8}$$

Such a vector exists by sure (probably not unique).

We begin by estimating the coefficients $(\theta_I^*,\, I \in \mathcal{I}(j_1)$ by the least square method. To get an explicit representation of the least square estimator (LSE) $\hat{\boldsymbol{\theta}}(j_1)$ of $\boldsymbol{\theta}^*(j_1)$, we introduce matrix notation.

First of all, we make an agreement to identify every function g on $\mathbb{R}^d$ with the vector $(g(X_i),\, i = 1, \dots, n)$ in $\mathbb{R}^n$. Particularly, the model function f is identified with the vector $(f(X_i),\, i = 1, \dots, n)$.

Denote by N_j the number of elements in each level j,

$$N_j = \#(\mathcal{I}_j) = 2^j$$

and let $N(j_1)$ be the total number of elements in the set $\mathcal{I}(j_1)$,

$$N(j_1) = 2 + \sum_{j=0}^{j_1} N_j = 1 + 2^{j_1+1}. \tag{3.9}$$

Introduce $n \times N(j_1)$-matrix $\Psi(j_1) = (\psi_{i,I},\, i = 1, \dots, n,\, I \in \mathcal{I}(j_1))$ with elements

$$\psi_{i,I} = \psi_I'(X_i) = \lambda_I^{-1} \psi_I(X_i), \qquad I \in \mathcal{I}(j_1),\, i = 1, \dots, n. \tag{3.10}$$

Now the approximation problem (3.8) can be rewritten in the form

$$\boldsymbol{\theta}^*(j_1) = \operatorname*{arginf}_{\boldsymbol{\theta}(j_1)} \|f - \Psi(j_1)\boldsymbol{\theta}(j_1)\|_n^2.$$

The solution to this quadratic problem can be represented as

$$\boldsymbol{\theta}^*(j_1) = \left(\Psi^T(j_1)\Psi(j_1)\right)^{-1} \Psi^T(j_1) f, \tag{3.11}$$

where the sign ^{T} means transposition. Strictly speaking, this representation is valid only if the matrix $\Psi^T(j_1)\Psi(j_1)$ is not degenerate. In the general case, one may use the similar expression for $\boldsymbol{\theta}^*(j_1)$ when understanding $\left(\Psi^T(j_1)\Psi(j_1)\right)^{-1}$ as a pseudo-inverse matrix.

Since the function f is observed with a noise, we consider the least squares estimator $\hat{\boldsymbol{\theta}}(j_1)$ of the vector $\boldsymbol{\theta}^*(j_1)$ which is defined by minimization of the sum of residuals square,

$$\hat{\boldsymbol{\theta}}(j_1) = \operatorname*{arginf}_{\boldsymbol{\theta}(j_1)} \|Y - \Psi(j_1)\boldsymbol{\theta}(j_1)\|_n^2 = \operatorname*{arginf}_{\{\theta_I \in \mathcal{I}(j_1)\}} \sum_{i=1}^{n} \left(Y_i - \sum_{I \in \mathcal{I}(j_1)} \theta_I \psi_I'(X_i) \right)^2. \tag{3.12}$$

Here Y means the column-vector with elements Y_i, $i = 1, \dots, n$.

Let $V(j_1)$ be the pseudo-inverse of $\Psi^T(j_1)\Psi(j_1)$,

$$V(j_1) = \left(\Psi^T(j_1)\Psi(j_1)\right)^{-1}.$$

Then

$$\hat{\boldsymbol{\theta}}(j_1) = V(j_1)\Psi^T(j_1)Y. \tag{3.13}$$

One gets by (3.1) in a usual way that $\hat{\boldsymbol{\theta}}(j_1)$ is a Gaussian vector with the mean $\boldsymbol{\theta}^*(j_1)$ and the covariance matrix $V(j_1)$,

$$\hat{\boldsymbol{\theta}}(j_1) \sim \mathcal{N}\left(\boldsymbol{\theta}^*(j_1), V(j_1)\right). \tag{3.14}$$

The proposed testing procedure is based on the fact that for a linear function f, all the coefficients θ^*_I, $I \neq 0, 1$, are zero and therefore, the corresponding estimates $\hat{\theta}_I$ are Gaussian zero mean random variables.

We proceed as follows. We start with $j_1 = 0$ and repeat the calculation for each j_1 till the finest resolution level $j(n)$ defined as

$$j(n) = \lfloor \log_2(n) - 1 \rfloor. \tag{3.15}$$

For each $j_1 \leq j(n)$, let $\hat{\boldsymbol{\theta}}(j_1)$ be defined by (3.12). Denote by $\hat{\boldsymbol{\theta}}_{j_1}$ the part of the vector $\hat{\boldsymbol{\theta}}(j_1)$ corresponding to the level j_1,

$$\hat{\boldsymbol{\theta}}_{j_1} = (\hat{\theta}_I,\ I \in \mathcal{I}_{j_1}).$$

At the step j_1, we analyze the subvector $\hat{\boldsymbol{\theta}}_{j_1}$ only. Following to Spokoiny (1996a), we introduce two kinds of tests: the first one, so called 'local test', analyses each term $\hat{\theta}_I$, $I \in \mathcal{I}_{j_1}$, separately; the second one is levelwise, i.e. all the estimates $\hat{\theta}_I$, $I \in \mathcal{I}_{j_1}$, are used for calculating the corresponding test statistic.

Let $v_{I,I'}$, $I, I' \in \mathcal{I}(j_1)$, be the elements of the matrix $V(j_1) = \left(\Psi^T(j_1)\Psi(j_1)\right)^{-1}$. Due to (3.14), one has under the null hypothesis $\hat{\theta}_I \sim \mathcal{N}(0, v_{I,I})$ and hence each variable $v_{I,I}^{-1/2}\hat{\theta}_I$ is standard normal (if $v_{I,I} > 0$). The local test rejects the null hypothesis if at least one such value exceeds certain logarithmic level,

$$\phi_{\mathrm{loc}}(j_1) = \mathbf{1}\left(\max_{I \in \mathcal{I}(j_1)} v_{I,I}^{-1/2}|\hat{\theta}_I| > \lambda_n\right) \tag{3.16}$$

where

$$\lambda_n = 2\sqrt{\log n}. \tag{3.17}$$

In the definition (3.16) we use the fact that $v_{I,I} = 0$ implies $\hat{\theta}_I = 0$, see (3.13), and we assume $0/0 = 0$. Note that both $v_{I,I}$ and θ_I depend on j_1. We do not show explicitly this dependence only with the aim to minimize our notation.

The local test ϕ_{loc} is very sensitive to functions f containing localized fluctuations like jumps or jumps of derivatives. The next test, which was called a χ^2-test in Ingster (1993) and 'detail' test in Spokoiny (1996a), allows to detect very small but systematic component. It is based on the standardized sum of squares of $\hat{\theta}_I$, $I \in \mathcal{I}_{j_1}$. Let V_{j_1} be the submatrix of the matrix $V(j_1)$ corresponding to the level j_1, i.e. $V_{j_1} = (v_{I,I'},\ I, I' \in \mathcal{I}_{j_1})$. In view of (3.14), the vector $\hat{\boldsymbol{\theta}}_{j_1}$ is under the null Gaussian zero mean with the covariance matrix V_{j_1}. First we consider the case when V_{j_1} is non-degenerate. Necessary corrections of the procedure for the general

case will be discussed later on. If $\det V_{j_1} \neq 0$, then the vector $\boldsymbol{\eta}_{j_1} = (\eta_I,\, I \in \mathcal{I}_{j_1})$ defined as standardization of $\hat{\boldsymbol{\theta}}_{j_1}$,

$$\boldsymbol{\eta}_{j_1} = V_{j_1}^{-1/2}\hat{\boldsymbol{\theta}}_{j_1}, \tag{3.18}$$

is under the null standard normal. We consider χ^2-type statistics

$$S_{j_1} = \|\boldsymbol{\eta}_{j_1}\|^2 = \sum_{I\in\mathcal{I}_{j_1}} \eta_I^2. \tag{3.19}$$

Obviously, for each $f \in \mathcal{F}_0$ (i.e. for a linear f), the distribution of S_{j_1} does not depend on f and we denote by $\boldsymbol{E}_0$ and $\boldsymbol{D}_0$ the corresponding expectation and variance. One has clearly

$$\begin{aligned}\boldsymbol{E}_0 S_{j_1} &= N_{j_1},\\ \boldsymbol{D}_0 S_{j_1} &= \boldsymbol{E}_0(S_{j_1} - \boldsymbol{E}_0 S_{j_1})^2 = 2N_{j_1}.\end{aligned}$$

Following to Spokoiny (1996a), we consider test statistics T_{j_1} of the form

$$T_{j_1} = \frac{S_{j_1} - \boldsymbol{E}_0 S_{j_1}}{\sqrt{\boldsymbol{D}_0 S_{j_1}}} = (2N_{j_1})^{-1/2}(S_{j_1} - N_{j_1}). \tag{3.20}$$

One may expect that at least when N_{j_1} is large enough, the value T_{j_1} is asymptotically normal. We define therefore

$$\phi_{\mathrm{detail}}(j_1) = \mathbf{1}\,(|T_{j_1}| > \lambda_n) \tag{3.21}$$

with the same λ_n as above.

In the case with $\det V_{j_1} = 0$, denote by $V_{j_1}^-$ the pseudo-inverse of V_{j_1} and set

$$S_{j_1} = \hat{\boldsymbol{\theta}}_{j_1}^T V_{j_1}^- \hat{\boldsymbol{\theta}}_{j_1}, \tag{3.22}$$

$$N'_{j_1} = \mathrm{tr}(V_{j_1}^- V_{j_1}). \tag{3.23}$$

Then clearly S_{j_1} is again a χ^2-statistic, now with N'_{j_1} degree of freedom. Particularly

$$\begin{aligned}\boldsymbol{E}_0 S_{j_1} &= N'_{j_1},\\ \boldsymbol{D}_0 S_{j_1} &= \boldsymbol{E}_0(S_{j_1} - \boldsymbol{E}_0 S_{j_1})^2 = 2N'_{j_1}.\end{aligned}$$

Further we proceed as above with T_{j_1} from (3.20) and N'_{j_1} in place of N_{j_1}.

Finally we reject the linear hypothesis H_0 if one of $\phi_{\mathrm{loc}}(j_1)$ or $\phi_{\mathrm{detail}}(j_1)$ does,

$$\phi^* = \max_{0\le j_1\le j(n)} \max\{\phi_{\mathrm{loc}}(j_1), \phi_{\mathrm{detail}}(j_1)\}. \tag{3.24}$$

4. Main results

In this section we present the results describing asymptotic properties of the proposed testing procedure. We split our results by evaluating separately the error probabilities of the first and second kinds. Such an approach is motivated by difference in necessary conditions on the model and especially on the design $X_1, \ldots, X_n$. The result describing the error probabilities $\alpha_{f_0}(\phi^*)$ of the first kind is valid under very mild assumptions on the design. This is a very important and desirable property of each 'good' testing procedure: whatever the design is, when the underlying

function is really linear, the test rejects null with a small probability. To provide with a high sensitivity of the test, we need in more strong regularity conditions on the design.

Recall that every function g on $\mathbb{R}^d$ is identified with the vector $(g(X_i),\, i = 1,\dots,n)$ in $\mathbb{R}^n$. Particularly, f is identified with $(f(X_{i,m}),\, i = 1,\dots,n)$ and ψ'_I is understood as the vector with elements $\lambda_I^{-1}\psi_I(X_{i,m})$. Recall also the notation $\|g\|_n^2 = \sum_{i=1}^n g^2(X_i)$.

Denote by $\mathcal{L}(j)$ the linear subspace in $\mathbb{R}^n$ generated by the functions $\{\psi'_I\}$, $I \in \mathcal{I}_{j'}$, $0 \le j' < j$,

$$\mathcal{L}(j) = \left\{ \sum_{j'=0}^{j-1} \sum_{I \in \mathcal{I}_{j'}} \theta_I \psi'_I \right\}.$$

By $\Pi(j+1)f$ we denote the closest to f point from $\mathcal{L}(j)$ w.r.t. the distance $\|\cdot\|_n$,

$$\Pi(j+1)f = \operatorname*{arginf}_{g \in \mathcal{L}(j)} \|f - g\|_n = \operatorname*{arginf}_{g \in \mathcal{L}(j)} \sum_{i=1}^n |f(X_i) - g(X_i)|^2.$$

We write also Π_n for $\Pi(j(n)+1)$.

Let ϕ^* be the test introduced above.

Theorem 4.1. *Let observations Y_i, X_i, $i = 1,\dots,n$, obey the regression model (1.1). If the function f is linear, then*

$$\alpha_f(\phi^*) \equiv \boldsymbol{P}_f(\phi^* = 1) \le \delta_1(n) \to 0,$$

where $\delta_1(n)$ depends on n only and $\delta_1(n) \to 0$ as $n \to \infty$.

Now we state the results concerning the sensitivity of the proposed test ϕ^*. The first assertion is purely statistical and it shows under which conditions we detect an alternative with high probability. Next we show how these conditions can be transferred into a more usual form about the rate of testing.

Proposition 4.1. *Let $\boldsymbol{\theta}_j^* = (\theta_I^*,\, I \in \mathcal{I}_j)$ be the subvector of the vector $\boldsymbol{\theta}^*(j)$ from (3.11) corresponding to j th resolution level of the first component and let V_j be the corresponding covariance submatrix. If, for some $j \le j(n)$, it holds*

$$T_j^* \equiv 2^{-(j+1)/2} \boldsymbol{\theta}_j^{*T} V_j^{-1} \boldsymbol{\theta}_j^* > 2\lambda_n,$$

then

$$\boldsymbol{P}_f(\phi_{\mathrm{detail}}(j) = 0) \le \delta(n) \to 0, n \to \infty,$$

where $\delta(n)$ depends on n only. If, for some $j \le j(n)$, it holds

$$T_{j,\infty}^* \equiv \max_{I \in \mathcal{I}_j} v_{I,I}^{-1/2} |\theta_I^*| > 2\lambda_n,$$

then

$$\boldsymbol{P}_f(\phi_{\mathrm{loc}}(j) = 0) \le \delta(n) \to 0, \qquad n \to \infty,$$

with the same $\delta(n)$.

The result of this proposition means that the test ϕ^* detects with a probability close to one any alternative for which at least one from the corresponding values T_j^* and $T_{j,\infty}^*$ exceeds the level $2\lambda_n$. Therefore, one may suppose that the error of the second kind may occur only if one has

$$T_j^* \leq 2\lambda_n, \qquad 0 \leq j \leq j(n), \tag{4.1}$$

$$T_{j,\infty}^* \leq 2\lambda_n, \qquad 0 \leq j \leq j(n). \tag{4.2}$$

It remains to understand what follows for the function f from these inequalities. For this we need to impose some regularity conditions on the design and smoothness conditions on the function f.

Regularity (or smoothness) conditions on a function f defined on the interval $[0,1]$ can be formulated in a different forms. We choose a way based on the accuracy of approximation of this function by piecewise polynomials of certain degree s. Given $j \leq j(n)$, denote by $\{A_I,\ I \in \mathcal{I}_j\}$ the partition of the interval $[0,1]$ into the intervals of the length 2^{-j}: if $I = (j,k)$ then $A_I = [k2^{-j}, (k+1)2^{-j})$. Next, for an integer s, define $\mathcal{P}_s(j)$ as the set of piecewise polynomials of degree $s-1$ on the partition $\{A_I\}$ i.e. every function g from $\mathcal{P}_s(j)$ coincides on each A_I with a polynomial $a_0 + a_1 x + \ldots + a_{s-1}x^{s-1}$ where the coefficients $a_0, \ldots, a_{s-1}$ may depend on I. Now the condition that a function f has regularity s can be understood in the sense that this function is approximated by functions from $\mathcal{P}_s(j)$ with the rate 2^{-js}, or, more precisely,

$$\inf_{g \in \mathcal{P}_s(j)} \left[\int_0^1 |f(t) - g(t)|^2 dt \right]^{1/2} \leq C 2^{-js}$$

where a positive constant C depends on s only.

In our conditions we change the integral by summation over observation points. This helps to present the results in a more readable form without changing the sense of required conditions. It can be easily seen that if the design is regular, then the both forms are equivalent up to a constant.

Let now a function f be fixed. Let also j_0 be such that $2^{j_0 - 1} \geq s$. Set for $j \geq j_0$

$$r_s(j) = \inf_{g \in \mathcal{P}_s(j - j_0)} \|f - g\|_n = \inf_{g \in \mathcal{P}_s(j-j_0)} \left[\sum_{i=1}^n |f(X_i) - g(X_i)|^2 \right]^{1/2}. \tag{4.3}$$

The quantity $r_s(j)$ characterizes the accuracy of approximation of f by piecewise polynomials. In our procedure, we use the Haar approximation which corresponds to the case with $s = 1$.

Next we quantify the design regularity. Set

$$u_*(j) = \inf_{I \in \mathcal{I}_j} 2^j M_I / n, \tag{4.4}$$

$$u^*(j) = \sup_{I \in \mathcal{I}_j} 2^j M_I / n. \tag{4.5}$$

Here $M_I = \#\{i : X_{i,1} \in A_I\}$. Design regularity means particularly that $u_*(j)$ is bounded away from zero i.e. each interval A_I contains enough design points X_i.

Our design condition will be formulated in term of the functions $r_s(j), u_*(j), u^*(j)$ and of the matrix V_j which is submatrix of $V(j) = (\Psi^T(j)\Psi(j))^{-1}$, see Subsection

3.2, $V_j = (v_{I,I'}\,,\ I, I' \in \mathcal{I}_j)$. Clearly V_j is a $N_j \times N_j$-matrix, $N_j = 2^j$. Set

$$v^*(j) \quad = \quad \|V_j\|, \tag{4.6}$$

Here the norm $\|A\|$ of a matrix A is understood as the maximal eigenvalue of this matrix or equivalently, $\|A\| = \sup_{\gamma:\|\gamma\|=1} \|A\gamma\|$ where sup is taken over $\gamma \in R^{N_j}$ and $\|\gamma\|^2 = \gamma_1^2 + \ldots + \gamma_{N(j)}^2$. One may define $v^*(j)$ as the maximal eigenvalues of V_j. We will understand design regularity in the sense that V_j is non degenerate and $v^*(j)$ are bounded for large enough j. Note that the values $v^*(j)$, $u_*(j)$ and $u^*(j)$ are closely related to each other, namely, the regularity condition in term of $v^*(j)$ is stronger than in term of $u_*(j)$ and $u^*(j)$. Indeed, $u_*(j)$ and $u^*(j)$ characterize only the properties of the projection of the design on intervals A_I corresponding to j th level whenever $v^*(j)$ speak also about identifiability of the coefficients θ_I^* from this level.

Theorem 4.2. *Let condition (D) hold. There exist constants C_1 and C_2 such that if, for some $j \le j(n)$, the model function f satisfies the following inequality*

$$\inf_{a,b} \|f - a - b\psi_1\|_n^2 \ge C_1 r_s^2(j) + C_2 \frac{u^*(j)}{u_*(j)} v^*(j) 2^{j/2} \lambda_n \tag{4.7}$$

with $\psi_1(x) = x$, then

$$\boldsymbol{P}_f(\phi^* = 0) \le \delta_3(n) \to 0, \qquad n \to \infty,$$

where $\delta(n)$ is shown in Proposition 4.1.

Remark 4.1. It is of interest to compare this result with more standard results on the rate of hypothesis testing. For instance, it was shown in Ingster (1982) that if f belongs to a Sobolev ball $W_s(1)$ with

$$W_s(1) = \left\{ f : \int_0^1 |f^{(s)}(x)|^2 dx \le 1 \right\},$$

$f^{(s)}$ being s th derivative of f, then the optimal rate of testing is $n^{-2s/(4s+1)}$. But the corresponding testing procedure makes heavily use of knowledge of s. Concerning our procedure, it is adaptive i.e. we do not need to know s. Next, the condition $f \in W_s(1)$ yields $n^{-1/2} r_s(j) \le C 2^{-js}$ and, if the design is regular (that is all $v^*(j)$ are bounded), then the optimization over j in the right hand-side of (4.7) gives the rate $(n/\lambda_n)^{-2s/(4s+1)}$ for the deviation of the function f from the space of linear functions. Therefore, our procedure provides with the near optimal rate of testing by some logarithmic factor which can be viewed as the price for the adaptation.

Remark 4.2. The result of Theorem 4.2 helps to understand what happens in the case when our design is not regular and, for instance, $u_*(j) = 0$ for all large j. It was already mentioned that the procedure can be applied in this situation too and the error probability of the first kind is very small. Concerning the error probability of the second kind, the inspection of the proof shows that design irregularity decreases the sensitivity of our procedure in the following sense: there exist smooth alternatives with probably large L_2-norm which are not detected. But this case corresponds to the situation when f is deviated from the best linear approximation only in the domain where are very few design points. It seems that there is no

testing method which could help to detect such an alternative. For the remaining alternatives our testing method is still sensitive.

The result of Theorem 4.2 is formulated for the case when smoothness properties of the function f are measured in L_2-norm. It can be seen from the proof that for this situation it suffices to apply only the test ϕ_{detail} which just corresponds to testing in L_2-norm. At the same time, it was shown in Lepski and Spokoiny (1995) and Spokoiny (1996b) that the case when we measure smoothness properties in some norm L_p with $p < 2$ leads to modification of the testing procedure to attain the optimal rate of testing. The latter choice with $p < 2$ corresponds to situation of a function f with inhomogeneous smoothness properties, particularly when this function has jumps or jumps of derivatives. Different testing procedures which allow rate optimal and even exact asymptotically optimal testing in this situation, can be found in Lepski and Spokoiny (1995), Ingster (1993, 1996), but all of them requires knowledge of p. Another approach was proposed in Spokoiny (1995) with the aim to construct an universal testing procedure which is near optimal for each p. This approach is based on combination of two testing methods one of them corresponds to testing in L_2-norm and another one corresponds to testing in L_∞-norm. The above proposed testing procedure exploit just this idea. We conclude by stating one more result in this spirit.

Given $j \le j(n)$, let $\boldsymbol{\theta}^*(j) = (\theta^*_I,\, I \in \mathcal{I}(j))$ be due to (3.11) and let $\boldsymbol{\theta}^*_j = (\theta^*_I,\, I \in \mathcal{I}_j)$ be the subvector tested at j th step. The test $\phi_{\mathrm{detail}}(j)$ is sensitive when $\|\boldsymbol{\theta}^*_j\|^2 \ge C\lambda_n 2^j$, see Proposition 4.1 and Lemma 5.1 below. At the same time, the test $\phi_{\mathrm{loc}}(j)$ is sensitive in the situation when at least one coefficient θ^*_I is greater than $C'\lambda_n$. This means that it is reasonable to apply the test ϕ_{loc} when the most of coefficients from j level are small and a few of them are of order λ_n. This just corresponds to the case of a function with inhomogeneous smoothness properties, particularly to a function with jumps.

Set

$$w(j,t) = \sum_{I \in \mathcal{I}_j} |\theta^*_I|^2 \mathbf{1}(|\theta^*_I| \le t). \tag{4.8}$$

We exploit the fact that under some regularity condition on f, the value $w(j,t)$ is small for j large enough.

Theorem 4.3. *Let also some s be fixed. There are constants C_1, C_2 and C_3 such that if, for some $j_1 \le j_2 \le j(n)$, the model function f satisfies the following inequality*

$$\inf_{a,b} \|f - a - b\psi_1\|_n^2 \ge C_1 r_s^2(j_2) + \frac{v^*(j_2)u^*(j_2)}{u_*(j_2)} \left[C_2 2^{j_1/2}\lambda_n + C_3 \sum_{j=j_1}^{j_2} w(j, t_j) \right]$$

with $t_j = 2\lambda_n \sqrt{v^*(j)}$, then

$$\boldsymbol{P}_f(\phi^* = 0) \le \delta(n) \to 0, \qquad n \to \infty,$$

where $\delta(n)$ is the same as in Proposition 4.1.

As a corollary of the last result, we show that our testing procedure provides with the near optimal rate of testing over Sobolev balls $W_{s,p}(1)$ with $p < 2$ and $s \le 1$,

$$W_{s,p}(1) = \left\{ f : \int_0^1 |f^{(s)}(x)|^p dx \le 1 \right\}.$$

Corollary 4.1. *There is a constant $C_4 > 0$ depending on n and design $X_1, \dots, X_n$ only such that if $f \in W_{s,p}(1)$ with $s \le 1$ and $sp - 1 + p/2 > 0$, and if*

$$\inf_{a,b} \|f - a - b\psi_1\|_n^2 \ge C_4 n^{-(2sp-1+p/2)/(2sp-1+p)} \lambda_n^{2(p-1)/(2sp-1+p)}$$

then

$$\boldsymbol{P}_f(\phi^* = 0) \le \delta(n) \to 0, \qquad n \to \infty,$$

where $\delta(n)$ is from Proposition 4.1.

By comparison with the results from Lepski and Spokoiny (1995) or Spokoiny (1996a) we see that the rate shown in this corollary is near optimal by the logarithmic factor $\lambda_n^{2(p-1)/(2sp-1+p)}$.

5. Proofs

In this section we prove Theorems 4.1 through 4.3 and other statements from the above.

5.1. Proof of Theorem 4.1

At the first step, we reduce the case of a linear null hypothesis about f to the case with the simple null $f \equiv 0$.

Under the null hypothesis, the function f is linear, $f = \theta_0 + \theta_1\psi_1$. Then clearly all the remaining coefficients of the vector $\boldsymbol{\theta}(j_1) = (\theta_I,\, I \in \mathcal{I}(j_1))$ are zero. Taking into account the model equation (1.1) we conclude that when considering test statistics based on $\hat{\boldsymbol{\theta}}_{j_1}$ one may change θ_0, and θ_1 by zero without any influence on their behavior, i.e. we reduce the problem to the simple zero null.

At the next step, we evaluate the error probabilities of the first kind for the tests ϕ_{loc} and ϕ_{detail}.

Given $j \le j(n)$, let $\hat{\theta}_I$, $I \in \mathcal{I}_j$, be the elements of the vector $\hat{\boldsymbol{\theta}}_j$ and let $V_j = (v_{I,I'},\, I, I' \in \mathcal{I}_j)$ be the corresponding covariance matrix. The local test $\phi_{\mathrm{loc}}(j)$ is based on statistics $T_I = v_{I,I}^{-1/2}\hat{\theta}_I$, and

$$\boldsymbol{P}(\phi_{\mathrm{loc}}(j) = 1) \le \sum_{I \in \mathcal{I}_j} \boldsymbol{P}(|T_I| > \lambda_n).$$

Obviously one can represent T_I in the form $T_I = a_1Y_1 + \ldots + a_nY_n$ with some coefficients a_i depending on I and on the design $X_1, \ldots, X_n$. Using the model equation (1.1) with zero in place of f, we get

$$T_I = \sum_{i=1}^{n} a_i\xi_i = \zeta_I.$$

Recall that the choice of normalizer $v_{I,I}^{-1/2}$ for $\hat{\theta}_I$ was made to provide standard normality of the stochastic term $\zeta_I = a_1\xi_1 + \ldots + a_n\xi_n$. Hence

$$\boldsymbol{P}(|T_I| > \lambda_n) \;=\; \boldsymbol{P}(|\zeta_I| > \lambda_n) \le 2\exp\{-\lambda_n^2/2\}.$$

This yields

$$\begin{aligned} \boldsymbol{P}(\phi_{\mathrm{loc}}(j) = 1) \;&=\; \boldsymbol{P}\left(\max_{I \in \mathcal{I}_j} |T_I| > \lambda_n\right) \le \sum_{I \in \mathcal{I}_j} \boldsymbol{P}(|T_I| > \lambda_n) \\ &\le\; 2^{j+1}\exp\{-\lambda_n^2/2\} \end{aligned}$$

and

$$\boldsymbol{P}(\phi_{\mathrm{loc}} = 1) \le \sum_{j=0}^{j(n)} \boldsymbol{P}(\phi_{\mathrm{loc}}(j) = 1) \le 2^{j(n)+2}\exp\{-\lambda_n^2/2\}.$$

Recall that the definition of $j(n)$ implies $2^{j(n)+1} \le n$ and hence

$$\boldsymbol{P}(\phi_{\mathrm{loc}} = 1) \le 2n\exp\{-2\log n\} = o_n(1).$$

Next we consider the test ϕ_{detail}. Let us fix again some level $j \le j(n)$. We suppose for simplicity that the matrix V_j is of the full rank. The general case can be studied in the same way.

The subtest $\phi_{\mathrm{detail}}(j)$ is based on the statistic $S_j = \|\boldsymbol{\eta}_j\|^2 = \|V_j^{-1/2}\hat{\boldsymbol{\theta}}_j\|^2$. Again we can represent $\boldsymbol{\eta}_j = V_j^{-1/2}\hat{\boldsymbol{\theta}}_j$ in the form

$$\boldsymbol{\eta}_j = A(Y) = A(\boldsymbol{\xi}) = \boldsymbol{\zeta}_j$$

where A is a linear operator from $\mathbb{R}^n$ into $\mathbb{R}^{N_j}$ and $\boldsymbol{\zeta}_j$ is a standard normal vector in $\mathbb{R}^{N_j}$. Now one has

$$\begin{aligned}
\boldsymbol{P}(\phi_{\mathrm{detail}}(j) = 1) &= \boldsymbol{P}(|S_j - N_j| > \sqrt{2N_j}\lambda_n) \\
&\le \boldsymbol{P}\left(\left|\|\boldsymbol{\zeta}_j\|^2 - N_j\right| > \lambda_n\sqrt{2N_j}\right) \\
&\le \boldsymbol{P}\left(\left|\|\boldsymbol{\zeta}_j\|^2 - N_j\right| > \lambda_n\sqrt{2N_j}\right).
\end{aligned}$$

Next, see Petrov (1975),

$$\boldsymbol{P}\left(\frac{\left|\|\boldsymbol{\zeta}_j\|^2 - N_j\right|}{\sqrt{2N_j}} > 2\sqrt{\log n}\right) \le \exp\{-\varkappa \log n\} = n^{-\varkappa}$$

with $\varkappa \ge 2$. Therefore,

$$\boldsymbol{P}(\phi_{\mathrm{detail}}(j) = 1) \le n^{-\varkappa}.$$

Summing up over all j from zero to $j(n)$ we conclude that

$$\boldsymbol{P}(\phi_{\mathrm{detail}} = 1) \le \sum_{j=0}^{j(n)} \boldsymbol{P}(\phi_{\mathrm{detail}}(j) = 1) \le n^{-\varkappa}\log n \to 0,$$

as $n \to \infty$. This completes the proof of Theorem 4.1.

5.2. Proof of Proposition 4.1

Let, for some $j \le j(n)$ and some $I \in \mathcal{I}_j$,

$$|v_{I,I}^{-1/2}\theta_I^*| > 2\lambda_n.$$

We use the decomposition $v_{I,I}^{-1/2}\hat{\theta}_I = v_{I,I}^{-1/2}\theta_I^* + \zeta_I$ where ζ_I is standard normal. Clearly

$$\begin{aligned}
\boldsymbol{P}_f(\phi_{\mathrm{loc}}(j) = 0) &\le \boldsymbol{P}_f\left(|v_{I,I}^{-1/2}\hat{\theta}_I| < \lambda_n\right) \\
&= \boldsymbol{P}_f\left(|v_{I,I}^{-1/2}\theta_I^* + \zeta_I| < \lambda_n\right) \\
&\le \boldsymbol{P}_f\left(|\zeta_I| > \lambda_n\right) \le e^{-\lambda_n^2/2} \to 0, \qquad n \to \infty,
\end{aligned}$$

as required.

Next we consider the situation when

$$T_j^* = 2^{-(j+1)/2}\boldsymbol{\theta}_j^{*T}V_j^{-1}\boldsymbol{\theta}_j^* > 2\lambda_n. \tag{5.1}$$

For notational simplicity, we suppose that the matrix V_j is non-degenerate. We will show that under the above assumption,

$$\boldsymbol{P}_f(T_j < \lambda_n) \le \delta(n) \to n, \qquad n \to \infty, \tag{5.2}$$

that obviously implies the assertion.

Recall that in the case with $\det V_j \neq 0$, one has $T_j = 2^{-(j+1)/2}(S_j - 2^j)$ where $S_j = \|V_j^{-1/2}\hat{\boldsymbol{\theta}}_j\|^2$. By construction, we may represent the vector $V_j^{-1/2}\hat{\boldsymbol{\theta}}_j$ in the form

$$V_j^{-1/2}\hat{\boldsymbol{\theta}}_j = \boldsymbol{b}_j + \boldsymbol{\zeta}_j$$

where $\boldsymbol{b}_j = V_j^{-1/2}\boldsymbol{\theta}_j^*$ and $\boldsymbol{\zeta}_j$ is a standard Gaussian vector. Notice that

$$\|\boldsymbol{b}_j\|^2 = \boldsymbol{\theta}_j^{*T} V_j^{-1}\boldsymbol{\theta}_j^* = 2^{(j+1)/2}T_j^*.$$

Denote

$$\gamma_j = \|\boldsymbol{b}_j\|^{-1}\sum_{I\in\mathcal{I}_j} b_I\zeta_I.$$

Clearly γ_j is a standard Gaussian random variable and one can decompose

$$S_j = \|\boldsymbol{b}_j + \boldsymbol{\zeta}_j\|^2 = \|\boldsymbol{b}_j\|^2 + \|\boldsymbol{\zeta}_j\|^2 + 2\|\boldsymbol{b}_j\|\gamma_j.$$

Now one has

$$\begin{aligned}\boldsymbol{P}_f(|T_j| < \lambda_n) &= \boldsymbol{P}\left(\left|\|\boldsymbol{b}_j\|^2 + \|\boldsymbol{\zeta}_j\|^2 - 2^j + 2\|\boldsymbol{b}_j\|\gamma_j\right| > \lambda_n 2^{(j+1)/2}\right)\\ &\le \boldsymbol{P}\left(2^{-(j+1)/2}|\|\boldsymbol{\zeta}_j\|^2 - 2^j| > \tfrac{3}{4}T_j^* - \lambda_n\right) + \boldsymbol{P}(|\gamma_j| > \tfrac{1}{4}T_j^{*-1/2}).\end{aligned}$$

It remains to note that $\frac{3}{4}T_j^* - \lambda_n \ge \lambda_n/2$ in view of (5.1) and we end up using the arguments from the proof of Theorem 4.1.

5.3. **Proof of Theorem 4.2**

We begin again by reduction of the testing problem with a linear null to the problem with the simple zero null hypothesis. Define coefficients θ_0, θ_1 by

$$(\theta_0, \theta_1) = \operatorname*{arginf}_{(a,b)} \|f - a - b\psi_1\|_n = \operatorname*{arginf}_{(a,b)} \sum_{i=1}^n (f(X_i) - a - bX_i)^2.$$

We set

$$f' = f - \theta_0 - \theta_1\psi_1.$$

Similarly to the proof of Theorem 4.1, we change f by f'. With this change, the vectors $\boldsymbol{\theta}^*(j)$ will be transformed into $\boldsymbol{\theta}'(j)$, having the same subvectors $\boldsymbol{\theta}_j^*$, $j \ge 0$. At the same time, by the triangle inequality for all a, b,

$$\begin{aligned}\|f' - a - b\psi_1\|_n &\ge \|f - (a - \theta_0) - (b - \theta_1)\psi_1\|_n\\ &\ge \varrho(n).\end{aligned}$$

Here we set

$$\varrho(n) = \inf_{a,b} \|f - a - b\psi_1\|_n.$$

Obviously the function f' has the same smoothness proprieties as f

$$\inf_{g\in\mathcal{P}_s(j)} \|f'-g\|_n \le r_s(j).$$

From this point, we may change the original regression function f by f'. About this new function f we know that

$$\|f\|_n = \inf_{a,b} \|f-a-b\psi_1\|_n \;\ge\; \varrho(n), \tag{5.3}$$

$$\inf_{g\in\mathcal{P}_s(j)} \|f-g\|_n \;\le\; r_s(j), \tag{5.4}$$

$$\boldsymbol{\theta}^*(j) \;=\; \operatorname*{arginf}_{\boldsymbol{\theta}(j)} \|f-\Psi(j)\boldsymbol{\theta}(j)\|_n, \tag{5.5}$$

for all j from zero to $j(n)$.

Now we turn directly to the proof of the theorem using the result of Proposition 4.1. We show that the condition (4.7) of the theorem with C_1 and C_2 large enough along with (5.3) and (5.4) contradict to the constraints from (4.1).

First we rewrite the latter constraints in term of $\|\boldsymbol{\theta}_j^*\|$. Recall that $\boldsymbol{\theta}_j^*$ is the subvector of $\boldsymbol{\theta}^*(j)$ corresponding to j th level, and V_j is the corresponding covariance submatrix of $V(j)$.

Lemma 5.1. *If* $T_j^* = 2^{-(j+1)/2}\boldsymbol{\theta}_j^{*T}V_j^{-1}\boldsymbol{\theta}_j^* \le 2\lambda_n$, *then*

$$\|\boldsymbol{\theta}_j^*\|^2 \le 2^{(j+3)/2}\lambda_n v^*(j). \tag{5.6}$$

Similarly, the inequality $T_{j,\infty}^* = \max_{I\in\mathcal{I}_j} v_{I,I}^{-1/2}|\theta_I^*| \le 2\lambda_n$ *implies*

$$\max_{I\in\mathcal{I}_j} |\theta_I^*| \le 2\lambda_n\sqrt{v^*(j)}. \tag{5.7}$$

Proof. The both statements are the direct consequences of the definition of the norm of a matrix. Indeed, denote $\boldsymbol{\eta}_j = V_j^{-1/2}\boldsymbol{\theta}_j^*$. Then $T_j^* = \|\boldsymbol{\eta}_j\|^2$ and $\boldsymbol{\theta}_j^* = V_j^{1/2}\boldsymbol{\eta}_j$. Next, obviously $\|V_j^{1/2}\| = \sqrt{\|V_j\|} = \sqrt{v^*(j)}$. Particularly this yields that

$$\|\boldsymbol{\theta}_j^*\|^2 = \|V_j^{1/2}\boldsymbol{\eta}_j\|^2 \le \left(\|V_j^{1/2}\|\,\|\boldsymbol{\eta}_j\|\right)^2 \le v^*(j)T_j^*,$$

and (4.1) implies (5.6).

Similarly one has $v_{I,I} \le \|V_j\|$ for all $I\in\mathcal{I}_j$ and hence

$$|\theta_I^*| = v_{I,I}^{1/2}|v_{I,I}^{-1/2}\theta_I^*| \le \sqrt{v^*(j)}T_{j,\infty}^*.$$

□

Recall the notation $\mathcal{L}(j)$ for the linear space generated by functions ψ_I, $I\in\mathcal{I}_{j'}$, with $0\le j'<j$. We denote also by $\Pi(j)f$ the projection of f onto the space $\mathcal{L}(j)$ with respect to the norm $\|\cdot\|_n$,

$$\Pi(j)f = \operatorname*{arginf}_{h\in\mathcal{L}(j)} \|f-h\|_n.$$

Particularly, $\Pi(0)f$ denotes the projection of f on the space of linear functions and one has by (5.3) $\Pi(0)f = 0$.

Lemma 5.2. *For each* $j \le j(n)$,

$$\|\Pi(j+1)f\|_n \le \|\Pi(j)f\|_n + \|\boldsymbol{\theta}_j^*\|.$$

Proof. Since $\mathcal{L}(j-1) \subseteq \mathcal{L}(j)$, then

$$\Pi(j)f = \Pi(j)\Pi(j+1)f.$$

When denoting $f(j+1) = \Pi(j+1)f$, one has therefore $\Pi(j)f = \Pi(j)f(j+1)$ and we have to show that

$$\|\Pi(j)f(j+1)\|_n \geq \|f(j+1)\|_n - \|\boldsymbol{\theta}_j^*\|.$$

In view of (5.5)

$$f(j+1) = \sum_{I \in \mathcal{I}(j)} \theta_I^* \psi_I'.$$

Denote by f_j the part of this sum corresponding to the last level $\mathcal{I}_j$ in $\mathcal{I}(j)$,

$$f_j = \sum_{I \in \mathcal{I}_j} \theta_I^* \psi_I'.$$

By construction, the functions ψ_I', $I \in \mathcal{I}_j$, are ortonormal w.r.t. to the inner product $\|\cdot\|_n$ and particularly

$$\|f_j\|_n^2 = \sum_{I \in \mathcal{I}_j} |\theta_I^*|^2 = \|\boldsymbol{\theta}_j^*\|^2.$$

Next, obviously $f(j+1) - f_j \in \mathcal{L}(j)$, and by definition of $\Pi(j)$,

$$\|f(j+1) - \Pi(j)f(j+1)\|_n \leq \|f(j+1) - (f(j+1) - f_j)\|_n = \|f_j\|_n = \|\boldsymbol{\theta}_j^*\|$$

and the assertion follows by the triangle inequality. □

Lemma 5.3. *Given $j_1 \leq j(n)$, let (4.1) hold true for all $j \leq j_1$. Then*

$$\|\Pi(j_1)f\|_n^2 \leq \varkappa_1 2^{j_1/2} \lambda_n v^*(j_1) \tag{5.8}$$

with $\varkappa_1 = 2^{3/2}(2^{1/4}-1)^{-2}$.

Proof. Recursive application of Lemma 5.2 gives

$$\|\Pi(j_1)f\|_n \leq \sum_{j=0}^{j_1-1} \|\boldsymbol{\theta}_j^*\|.$$

Here we have used that $\Pi(0)f = 0$. Since the norm $v^*(j)$ obviously increases with j, then this result along with the bound (5.6) yields

$$\|\Pi(j_1)f\|_n \leq \sum_{0}^{j_1-1} \left(2^{(j+3)/2} \lambda_n v^*(j_1)\right)^{1/2} = \left(2^{3/2} \lambda_n v^*(j_1)\right)^{1/2} \sum_{j=0}^{j_1-1} 2^{j/4}$$

and the assertion follows by straightforward calculation. □

Let some $j_1 \leq j(n)$ be fixed and let $g \in \mathcal{P}_s(j_1 - j_0)$ be such that

$$\|f - g\|_n \leq r_s(j_1).$$

Lemma 5.4. *There is a constant $\varkappa_2 > 0$ depending on s only and such that for each $j \leq j(n)$*

$$\|\Pi(j)f\|_n \geq \varkappa_2 u^*(j)/u_*(j) \left(\|f\|_n - r_s(j)\right).$$

Proof. Let $g \in \mathcal{P}_s(j-j_0)$ be such that $\|f-g\|_n \le r_s(j)$. Then

$$\begin{aligned}\|\Pi(j)f\|_n &= \|\Pi(j)g + \Pi(j)(f-g)\|_n \ge \|\Pi(j)g\|_n - \|\Pi(j)(f-g)\|_n \\ &\ge \|\Pi(j)g\|_n - r_s(g).\end{aligned}$$

Recall that g is a piecewise polynomial function on the partition A_I, $I \in \mathcal{I}_{j-j_0}$ and the projection $\Pi(j)g$ means the approximation of each polynomial on interval A_I of length $2^{-(j-j_0)}$ by piecewise constant functions with piece length 2^{-j}. Therefore, it suffices to prove that for each piece A_I and every polynomial $P(x) = a_0 + a_1 x + \ldots + a_{s-1}x^{s-1}$, it holds

$$\sum_{A_I} [\Pi(j)P(X_i)]^2 \ge \varkappa u^*(j)/u_*(j) \sum_{A_I} P^2(X_i)$$

where the constant $\varkappa$ depends on s only. The similar fact with integration instead of summation over A_I was stated in Ingster (1993) and we present here only a sketch of the proof for our situation.

The key idea of the proof can be formulated as a separate statement.

Lemma 5.5. *Let $u_* \le 1$, $u^* \ge 1$ and let μ be a measure on the interval $[0,1]$ such that*

$$u_* 2^{-j_0} \le \mu(A_k) \le u^* 2^{-j_0} \tag{5.9}$$

for all intervals $A_k = [k2^{-j_0}, (k+1)2^{-j_0})$, $k = 0, 1, \ldots, 2^{j_0}$. Then there exists a constant $\varkappa$ depending on s only and such that for every polynomial $P(x) = a_0 + a_1 x + \ldots + a_{s-1}x^{s-1}$

$$\sum_{k=0}^{2^{j_0}-1} \left[\int_{A_k} P(x)\mu(dx)\right]^2 \ge \varkappa u_*/u^* \int_0^1 P^2(x)\mu(dx).$$

Proof. We begin by reducing the case of an arbitrary u_* and u^* to the case with $u_* = u^* = 1$. Define measure μ' on $[0,1]$ by $d\mu'/d\mu(x) = u_* 2^{-j_0}/\mu(A_k)$ if $x \in A_k$. Due to (5.9), $d\mu'/d\mu \le 1$ and obviously $\mu'(A_k) = u2^{-j_0}$. Next, similarly $d\mu/d\mu' = u_*^{-1}2^{j_0}\mu(A_k) \le u_*^{-1}u^*$. Now

$$\begin{aligned}\sum_{k=0}^{2^{j_0}-1} \left[\int_{A_k} P(x)\mu(dx)\right]^2 &\ge \sum_{k=0}^{2^{j_0}-1} \left[\int_{A_k} P(x)\mu'(dx)\right]^2, \\ \int_0^1 P^2(x)\mu'(dx) &\ge u_*^{-1}u^* \int_0^1 P^2(x)\mu(dx).\end{aligned}$$

Therefore, it suffices to show that

$$\sum_{k=0}^{2^{j_0}-1} \left[\int_{A_k} P(x)\mu'(dx)\right]^2 \ge \varkappa \int_0^1 P^2(x)\mu'(dx),$$

or, equivalently to consider the case when $\mu(A_k) = 2^{-j_0}$ for all $k = 0, \ldots, 2^{j_0}-1$.

Let $a = (a_0, \ldots, a_{s-1})$ be the vector of coefficients of P. Then obviously

$$\int_0^1 P^2(x)\mu(dx) \le C\|a\|^2$$

where $\|a\|^2 = a_0^2 + \ldots + a_{s-1}^2$. Introduce matrix M with elements $\mu_{k,l} = \int_{A_k} x^l \mu(dx)$, $k = 0, \ldots, 2^{j_0} - 1$, $l = 0, \ldots, s-1$). Then Ma is a vector in the space $\mathbb{R}^{2^{j_0}}$ and

$$\sum_{k=0}^{2^{j_0}-1} \left[\int_{A_k} P(x)\mu(dx) \right]^2 = \|Ma\|^2.$$

Now we use that $\|Ma\|^2 = a^T M^T M a \geq \|a\|^2 / \|(M^T M)^{-1}\|$. It remains to note that the conditions $s < 2^{j_0 - 1}$ and $\mu(A_k) = 2^{-j_0}$ yield that $\|(M^T M)^{-1}\| \leq C$ for some constant C depending on s only. □

Application of this result to each interval A_I, $I \in \mathcal{I}_{j-j_0}$ yields the desirable assertion. □

Summing up the results of Lemma 5.1 through 5.4 we resume that the inequality $\|f\|_n \geq r_s(j) + C\sqrt{2^{j/2}\lambda_n v^*(j) u^*(j)/u_*(j)}$ with C large enough contradicts to the constraints (4.1) and the theorem is proved.

5.4. **Proof of Theorem 4.3**

We proceed in the same line as in the proof of Theorem 4.2. The difference is only in evaluating the norm $\|\Pi(j)f\|_n$, see Lemmas 5.2 and 5.3. Similarly to Lemma 5.2 one can show that

$$\|\Pi(j+1)f\|_n \leq \|\Pi(j)f\|_n + \|\boldsymbol{\theta}_j^*\|.$$

(We use here the same notation as above.)

Next, in view of the constraints from (4.2) and Lemma 5.1, one has

$$\|\boldsymbol{\theta}_j^*\|^2 = \sum_{I \in \mathcal{I}_j} |\theta_I^*|^2 = \sum_{I \in \mathcal{I}_j} |\theta_I^*|^2 \mathbf{1}(|\theta_I^*| \leq t_j) = r(j, t_j)$$

where $t_j = 2\lambda_n \sqrt{v^*(j)}$. Using this bound for j between j_1 and j_2 and the bound from Lemma 5.2 for $j < j_1$, we estimate

$$\|\Pi(j_2)f\|_n \leq \varkappa_1 2^{j_1} \lambda_n v^*(j_1) + \sum_{j=j_1}^{j_2} r(j, t_j).$$

This allows to complete the proof by the same arguments as for Theorem 4.2.

5.5. **Proof of Corollary 4.1**

Let $j \leq j(n)$ and $\boldsymbol{\theta}_j^* = (\theta_I^*,\, I \in \mathcal{I}_j)$ is the subvector of $\boldsymbol{\theta}(j)$ corresponding to j th level. It is well known, see Triebel (1992) that if $f \in W_{s,p}(1)$ for $s \leq 1$ and if the design is regular, then the sum

$$f_j = \sum_{I \in \mathcal{I}_j} \theta_I^* \psi_I'$$

fulfill the condition

$$n^{-1} \sum_{i=1}^{n} |f_j(X_i)|^p \leq C 2^{-jsp}.$$

By definition, all the functions ψ'_I, $I \in \mathcal{I}_j$, have non intersecting supports and $|\psi'_I| \le \lambda_I^{-1} = M_I^{-1/2}$, where M_I is the number of design points in I th interval and due to (4.4) $M_I \ge n2^{-j}u_*(j)$. This yields

$$\sum_{i=1}^{n} |f_j(X_i)|^p = \sum_{I\in\mathcal{I}_j} |\theta^*_I|^p M_I \lambda_I^{-p} \ge (n2^{-j}u_*(j))^{1-p/2} \sum_{I\in\mathcal{I}_j} |\theta^*_I|^p.$$

Hence

$$\sum_{I\in\mathcal{I}_j} |\theta^*_I|^p \le Cn2^{-jsp}(n2^{-j}u_*(j))^{-1+p/2} \le C'n^{p/2}2^{-j(sp-1+p/2)}$$

where C' depends on design only (through $u_*(j)$). This gives

$$w(j,t) = \sum_{I\in\mathcal{I}_j} |\theta^*_I|^2 \mathbf{1}(|\theta^*_I| \le t) \le t^{2-p} \sum_{I\in\mathcal{I}_j} |\theta^*_I|^p \le t^{2-p}C'n^{p/2}2^{-j(sp-1+p/2)}.$$

Since $sp - 1 + p/2 > 0$ we obtain

$$\sum_{j=j_1}^{j_2} w(j,t_j) \le C''n^{p/2}\lambda_n^{2-p}2^{-j_1(sp-1+p/2)}.$$

Here $t_j = 2\lambda_n\sqrt{v^*(j)}$ and C'' depends on the design only. Now one selects j_1 to minimize the sum $C_2 2^{j_1/2}\lambda_n + C''n^{p/2}\lambda_n^{2-p}2^{-j_1(sp-1+p/2)}$ that leads to the rate $n^{-(2sp-1+p/2)/(2sp-1+p)}\lambda_n^{2(p-1)/(2sp-1+p)}$ shown in the corollary.

References

[1] Burnashev, M.V. (1979) On the minimax detection of an inaccurately known signal in a white Gaussian noise background. *Theory Probab. Appl.* **24** 107 – 119.

[2] Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistic* **4** 1926 – 1947.

[3] Härdle, W., Spokoiny, V. and Sperlich, S. (1995). Semiparametric single index versus fixed link function modeling. Preprint **190**, Weierstrass Institute, Berlin. (to appear in *Annals of Statistics*).

[4] Ibragimov,I.A. and Khasminskii,R.Z. (1977). One problem of statistical estimation in Gaussian white noise. *Soviet Math. Dokl.* **236** N.4 1351 – 1354.

[5] Ingster, Yu.I. (1982). Minimax nonparametric detection of signals in white Gaussian noise *Problems Inform. Transmission* **18** 130 – 140.

[6] Ingster, Yu.I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I–III. *Math. Methods of Statist.* **2**, 85 – 114, **3**, 171 – 189, **4**, 249 – 268.

[7] Lehmann, E.L. (1959) *Testing Statistical Hypothesis* Wiley, New York.

[8] Lepski, O.V. and Spokoiny, V.G. (1995). Minimax Nonparametric Hypothesis Testing: The Case of an Inhomogeneous Alternative. *Bernoulli*, tentatively accepted.

[9] Mann, H.B. and Wald, A.(1942) On the choice of the number of intervals in the application of the chi-square test. *Ann. Math. Stat.* **13** 306–317

[10] Neyman, J. (1937) "Smooth test" for goodness of fit. *Scand.Aktuarietidskr.* **20** : 149–199.

[11] Petrov, V.V. (1975) *Sums of Independent Random Variables*. Springer, New York.

[12] Spokoiny, V. (1996a) Adaptive hypothesis testing using wavelets. *Annals of Statistics*, to appear.

[13] Spokoiny, V. (1996b) Adaptive and spatially adaptive testing of a nonparametric hypothesis. *Probability Theory and Related Fields*, submitted.

[14] Triebel, H. (1992). *Theory of function spaces*. Birkhäuser, Basel.

Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin, Germany., E-mail: spokoiny@wias-berlin.de