The Annals of Statistics
1996, Vol. 24, No. 6, 2477–2498

ADAPTIVE HYPOTHESIS TESTING USING WAVELETS

BY V. G. SPOKOINY

Weierstrass Institute for Applied Analysis and Stochastics and Institute for Information Transmission Problems

Let a function f be observed with a noise. We wish to test the null hypothesis that the function is identically zero, against a composite nonparametric alternative: functions from the alternative set are separated away from zero in an integral (e.g., L_2) norm and also possess some smoothness properties. The minimax rate of testing for this problem was evaluated in earlier papers by Ingster and by Lepski and Spokoiny under different kinds of smoothness assumptions. It was shown that both the optimal rate of testing and the structure of optimal (in rate) tests depend on smoothness parameters which are usually unknown in practical applications. In this paper the problem of adaptive (assumption free) testing is considered. It is shown that adaptive testing without loss of efficiency is impossible. An extra log log-factor is inessential but unavoidable payment for the adaptation. A simple adaptive test based on wavelet technique is constructed which is nearly minimax for a wide range of Besov classes.

1. Introduction. Suppose we are given data

$$dX(t) = f(t)\,dt + \varepsilon dW(t), \qquad 0 \le t \le 1,$$

where f is an unknown function and W is a standard Wiener process. We wish to test the null hypothesis H_0: $f \equiv 0$ against the composite nonparametric alternative that the function f is separated away from zero in L_2-norm, $\|f\| \ge \varrho(\varepsilon)$, and also f possesses some smoothness properties. The problem is to describe the minimal (optimal) rate for the distance $\varrho(\varepsilon)$ for which testing with prescribed error probabilities is still possible. The result depends heavily on what kind of smoothness assumptions are imposed. For the cases of Hölder or Hilbert–Sobolev functional classes, this problem was exhaustively solved by Ingster (1982, 1993) and Ermakov (1990). It turned out that the optimal rate $\varrho(\varepsilon)$ for testing differs from the optimal rate for the problem of estimation of a function: if s is the smoothness parameter, then

$$\varrho(\varepsilon) = \varepsilon^{4s/(4s+1)}.$$

The case of Besov functional classes $B_{s,p,q}$ with $p < 2$ was considered in Lepski and Spokoiny (1995b). This case is not only of theoretical interest. It corresponds to the situation when functions from the alternative set are of

Received May 1995; revised April 1996.
AMS 1991 *subject classifications*. Primary 62G10; secondary 62G20.
Key words and phrases. Adaptive testing, signal detection, minimax hypothesis testing, nonparametric alternative, thresholding, wavelet decomposition.

inhomogeneous smoothness properties. The optimal rate was proved to be

$$r(\varepsilon) = \varepsilon^{4s''/(4s''+1)},$$

where $s'' = s - 1/(2p) + 1/4$. The rate-optimal test constructed in that paper makes heavy use of the pointwise adaptive procedure proposed in Lepski, Mammen and Spokoiny (1997) and developed in Lepski and Spokoiny (1995a).

However, the practical applications of this test or of that proposed by Ingster meet the crucial problem: the structure of the test uses knowledge of the smoothness parameters s, p, which are typically unknown. The present paper treats the problem of adaptive (assumption free) testing. The goal is to proposed a test which does not use any information about smoothness properties of the function f but which is at least nearly optimal.

The theory of adaptive nonparametric estimation is now well developed. We mention here the papers by Efroimovich and Pinsker (1984), Poljak and Tsybakov (1990), Golubev (1990) and Lepski (1991). The key point of the relevant results can be asserted as follows: an adaptive estimation of the function f for integrated losses is possible without loss of efficiency and can be performed even in an optimal way (up to the exact constant). The reader is referred to Donoho and Johnstone (1993) and Marron (1987) for further discussion of this problem.

In this context, it is worth mentioning an interesting phenomenon discovered by Lepski (1990) and then Brown and Low (1992): for some statistical estimation problems, an adaptive estimation without loss of efficiency is impossible. A typical example of this sort is the problem of estimation of a function f at a given point t_0. It was shown in Lepski (1990) that adaptive pointwise estimation leads to a nearly minimax rate, which is worse than a minimax one within an extra log-factor.

In the present paper it is shown that adaptive testing also leads to some loss of efficiency but in this case within an extra log log-factor. The difference from the preceding case is explained mostly by the structure of the loss function (it is bounded in the hypothesis testing problem). However, the related consideration seems to be more involved.

The rate-optimal adaptive test is also presented. We use the wavelet technique for the construction, which provides very useful tools for studying the problem under consideration.

The paper is organized as follows. In Section 2 we state the testing problem and formulate the main results. In Section 3 we explain the proposed adaptive test procedure, which makes use of wavelet decomposition. The proofs are postponed to the last section.

2. Main results. In this section we formulate the problem of minimax and adaptive minimax hypothesis testing and state the results.

2.1. *Model and hypothesis testing problem.* Assume we are given the data $X(t)$, $0 \le t \le 1$, obeying the following stochastic differential equation:

$$dX(t) = f(t)\,dt + \varepsilon\,dW(t), \qquad 0 \le t \le 1,\ X(0) = 0. \tag{2.1}$$

The function $f(\cdot)$ is unknown and the following statistical problem is considered: to test the null hypothesis H_0 that the function f is identically zero,

$$H_0 \colon f \equiv 0.$$

We wish to test this hypothesis against as large a class of alternatives as possible. That is why we do not assume any special (parametric) structure for the alternative set. This leads to considering a nonparametric alternative set. Following Ingster (1982, 1984a, b, 1993) and Lepski and Spokoiny (1995b), we assume only that the function f obeys some smoothness conditions. More precisely, the function f is supposed to lie in some Besov ball $B_{s,p,q}(M)$,

$$B_{s,p,q}(M) = \{f \colon \|f\|_{B_{s,p,q}} \le M\}.$$

The definition of the Besov norm $\|\cdot\|_{B_{s,p,q}}$ can be found, for example, in Triebel (1992). For the discussion of this notion in a statistical context, see Donoho and Johnstone (1995) or Donoho, Johnstone, Kerkyyacharian and Picard (1994). For the case of an integer s and $p = q$, one may apply Sobolev's type of seminorm $\|f\|_{H_{s,p}} = (\int |f^{(s)}(t)|^p \, dt)^{1/p}$ instead of the mentioned Besov norm. For this case, the parameter s might be viewed as the number of derivatives of the function f bounded in L_p-norm. Note that the definition of a Besov space can be done also in terms of the wavelet decomposition; see the property ISO2 in Section 3.1.

To be able to test the null against the alternative, we assume also that the alternative set is separated away from the null in L_2-norm. Hence we arrive at the following alternative:

$$H_1 \colon \mathscr{F}_\sigma(\varrho) = \{f \in B_{s,p,q}(M) \colon \|f\| \ge \varrho\}.$$

Now we define the hypothesis testing problem. A (nonrandomized) test ϕ is a measurable function of the observation $X(\cdot)$ with two values $\{0, 1\}$. As usual, the event $\{\phi = 0\}$ is treated as accepting the null hypothesis, and $\{\phi = 1\}$ means that the null is rejected. To simplify the exposition, we do not consider randomized tests. All the results can be extended to the case of randomized tests in a standard way; see, for example, Lehmann (1959) or Ingster (1993).

Let $\mathbf{P}_0$ be the distribution of the process $X(\cdot)$ under the null, that is, if we observe pure noise, and let $\mathbf{P}_f$ mean the distribution of the process X under f due to (2.1), $\mathbf{P}_f = \mathscr{L}(X|f)$.

The quality of any test ϕ is measured by the corresponding error probabilities of the first and second kinds. For the case under consideration with a simple hypothesis, the error probability of the first kind is

$$\alpha(\phi) = \mathbf{P}_0(\phi = 1).$$

If f is a point from the alternative set, $f \in \mathscr{F}_\sigma(\varrho)$, then the error probability of the second kind at f is defined as usual by $\beta(f) = \mathbf{P}_f(\phi = 0)$. The value $1 - \beta(f)$ is called the *power* of the test ϕ at f.

We consider further the minimax set-up, which leads to the following criterion:

$$\beta_\sigma(\phi, \varrho) = \sup_{f \in \mathscr{F}_\sigma(\varrho)} \mathbf{P}_f(\phi = 0). \tag{2.2}$$

2.2. *Minimax rate of testing*. Here we focus on the asymptotic hypothesis testing problem as the noise level tends to zero ($\varepsilon \to 0$). We are interested in evaluating the optimal (fastest) rate of decay to zero of the radius ϱ as a function of ε as $\varepsilon \to 0$, for which testing with prescribed error probabilities is still possible. The following definition of the minimax rate $\varrho(\varepsilon)$ was proposed in Ingster (1993).

DEFINITION 2.1. A sequence $\varrho(\varepsilon)$ is called the *minimax rate of testing* if $\varrho(\varepsilon) \to 0$ as $\varepsilon \to \infty$ and the following two conditions hold.

(i) For any sequence $\varrho'(\varepsilon)$ such that

$$\varrho'(\varepsilon)/\varrho(\varepsilon) = o_\varepsilon(1),$$

one has

$$\inf_{\phi_\varepsilon}\left[\mathbf{P}_0(\phi_\varepsilon) + \beta_\sigma(\phi_\varepsilon, \varrho'(\varepsilon))\right] = 1 - o_\varepsilon(1).$$

(ii) For any $\alpha, \beta > 0$, there exist a constant $C > 0$ and test ϕ_ε^* such that

$$\mathbf{P}_0(\phi_\varepsilon^*) \le \alpha + o_\varepsilon(1)$$
$$\beta_\sigma(\phi_\varepsilon^*, C\varrho(\varepsilon)) \le \beta + o_\varepsilon(1).$$

Here and below we denote by $o_\varepsilon(1)$ any sequence tending to zero as $\varepsilon \to 0$.

REMARK 2.1. The first condition of the above definition means that testing with a rate faster than $\varrho(\varepsilon)$ is impossible; if the distance between the null and the alternative set is less in order than $\varrho(\varepsilon)$, then any test has asymptotically trivial power in the sense that the sum of the error probabilities of the first and second kinds is close to 1. The second condition means roughly that, on the contrary, if the distance is of the order $\varrho(\varepsilon)$, then testing can be done with prescribed error probabilities.

It turns out that the rate $\varrho(\varepsilon)$ depends critically on the smoothness parameters $\sigma = (s, p, q, M)$.

THEOREM 2.1 [Lepski and Spokoiny (1995b)]. *Given* $\sigma = (s, p, q, M)$ *with* $sp > 1$, *let*

$$\varrho_\sigma(\varepsilon) = M^{1/(4s''+1)}\varepsilon^{4s''/(4s''+1)}, \tag{2.3}$$

where

$$s'' = s - \left(\frac{1}{2p} - \frac{1}{4}\right)_+ = \min\left\{s, s - \frac{1}{2p} + \frac{1}{4}\right\}.$$

Then $\varrho_\sigma(\varepsilon)$ *is the minimax rate of testing in the sense of Definition* 2.1.

The structure of rate optimal tests ϕ_ε^* is described in the next section.

2.3. *Adaptive testing*. Now we turn to the problem of adaptive testing when the parameters $\sigma = (s, p, q, M)$ are unknown. First we state the phenomenon of "lack of adaptability" for this problem, that is, we show that adaptive testing with the same rate is impossible. Then we describe the optimal adaptive rate of testing. For this we use the notion of adaptive factor.

We start with the definition of the problem of adaptive testing. Let again the alternative set H_1 be described as before, but let the parameter σ be unknown. We assume only that σ belongs to some set $\mathcal{T}$. For each $\sigma \in \mathcal{T}$, the optimal rate of testing $\varrho(\varepsilon) = \varrho_\sigma(\varepsilon)$ is from (2.3). Due to Definition 2.1, given σ, α_0, β_0, there are a constant c_1 and tests $\phi_{\sigma,\varepsilon}$ such that $\sigma(\phi_{\sigma,\varepsilon}) = P_0(\phi_{\sigma,\varepsilon} = 1) \le \alpha_0 + o_\varepsilon(1)$ and $\beta_\sigma(\phi_\sigma, c_1\varrho_\sigma(\varepsilon)) \le \beta_0 + o_\varepsilon(1)$. But now, for the problem of adaptive testing, we search for a universal test ϕ_ε such that $\sigma(\phi_\varepsilon) \le \alpha_0 + o_\varepsilon(1)$ and $\beta_\sigma(\phi_\varepsilon, c\varrho_\sigma(\varepsilon)) \le \beta_0 + o_\varepsilon(1)$ for some $c > 0$ and all $\sigma \in \mathcal{T}$.

We say that a set $\mathcal{T}$ is *nontrivial* if there are such p, q, M and $s_* < s^*$ that

$$(s, p, q, M) \in \mathcal{T}, \qquad \forall\, s \in [s_*, s^*].$$

The first result shows that adaptive testing (without loss of power) is impossible for any nontrivial set $\mathcal{T}$.

THEOREM 2.2. *Let $\mathcal{T}$ be nontrivial. Then for any $c > 0$ and any test ϕ,*

$$P_0(\phi = 1) + \sup_{\sigma\in\mathcal{T}} \beta_\sigma(\phi, c\varrho_\sigma(\varepsilon)) \ge 1 - o_\varepsilon(1).$$

The next question is how one should define the optimal adaptive rate. One way to do this was proposed by Lepski (1990) for the problem of adaptive estimation of a function at one point where the phenomenon of lack of adaptability appeared for the first time. We use another approach based on the notion of *adaptive factor*. Namely, we search for a sequence $t_\varepsilon \to \infty$ such that testing with the rate $\varrho_\sigma(\varepsilon t_\varepsilon)$ will be possible adaptively in $\sigma \in \mathcal{T}$. The next results show that for the problem under consideration the minimal adaptive factor is $(\ln\ln \varepsilon^{-2})^{1/4}$.

THEOREM 2.3. *Let*

$$t_\varepsilon = (\ln\ln \varepsilon^{-2})^{1/4}. \tag{2.4}$$

If $\mathcal{T}$ is a nontrivial set and if t'_ε is such that $t'_\varepsilon/t_\varepsilon = o_\varepsilon(1)$, then for any $c > 0$ and any test ϕ_ε,

$$P_0(\phi_\varepsilon = 1) + \sup_{\sigma\in\mathcal{T}} \beta_\sigma(\phi_\varepsilon, c\varrho_\sigma(\varepsilon t'_\varepsilon)) \ge 1 - o_\varepsilon(1).$$

THEOREM 2.4. *Let t_ε be as above and let a set $\mathscr{T}$ be of the form*

$$\mathscr{T} = \{\sigma = (s, p, q, M): s \le s^*, 1 \le p \le p^*, M_* \le M \le M^*, sp > 1\}$$

with some prescribed positive $s^*, p^*, M_* \le M^*$. Then there exist a constant $c_1 = c_1(s^*, p^*, M_*, M^*)$ and a test ϕ_ε such that

$$P_0(\phi_\varepsilon = 1) = o_\varepsilon(1)$$

$$\sup_{\sigma \in \mathscr{T}} \beta_\sigma(\phi_\varepsilon, c_1 \varrho_\sigma(\varepsilon t_\varepsilon)) = o_\varepsilon(1).$$

REMARK 2.2. Here we meet the degenerate behavior of the error probabilities for the adaptive test. The similar degenerate behavior of the losses appeared in the problem of adaptive estimation at a point; see Lepski and Spokoiny (1995a).

2.4. *Results for other nonparametric statistical models.* In the present paper we focus on the ideal "signal + white noise" model. Of course, the statistical practice needs to consider more realistic models such as density or spectral density function models, regression models with heteroskedastic non-Gaussian errors and so on. We believe that the ideas proposed are well applicable to the models mentioned above, but the exact theoretical study lies beyond the scope of the present paper. We cite only a few papers which can be helpful for these developments. Brown and Low (1996) proved the equivalence in the Le Cam sense of the "white noise" model and Gaussian regression model. Nussbaum (1993) stated a similar result for density models. Neumann and Spokoiny (1995) showed the equivalence in the estimation problem between the regression model with heteroskedastic non-Gaussian errors and the white noise model. Ingster (1984a, b, 1993) explored the hypothesis testing problem for the density and spectral density models. Kerkyacharian and Picard (1993) studied the optimal properties of the wavelet shrinkage procedure for the density model. Härdle and Mammen (1993) studied the problem of testing parametric versus nonparametric regression fit for the case of heteroskedastic errors.

3. Test procedure.

The construction of the test makes heavy use of the wavelet decomposition.

3.1. *Wavelet transform.* Assume we are given an orthonormal basis of compactly supported wavelets of $L_2[0,1]$. One may use the construction from Meyer (1990) or Cohen, Daubechies, Jawerth and Vail (1993). Let $\phi_{j,k}, \psi_{j,k}$ be a system of compactly supported orthogonal wavelets (supp $\phi \subseteq [-0, A]$ and supp $\psi \subseteq [-0, A]$). We suppose that ϕ and $\psi \in C^m$, where m is the maximal integer smaller than $s_{\max}$. This implies [cf. Daubechies (1992), Chapter 7] that $\psi(x)$ has at least m vanishing moments.

Let j_0 be such that $2^{j_0} > A + 1$. It has been shown in Cohen, Daubechies, Jawerth and Vail (1993) and Cohen, Daubechies and Vail (1993b) that an orthogonal wavelet basis on $[0,1]$ can be constructed by retaining $\psi_{j,k}$ and

$\phi_{j,k}$ as the interior wavelets and scaling functions and adding adapted edge wavelets and scaling functions. These edge elements are tailored so that the total number is exactly 2^j at resolution j. For the sake of simplicity, we use the same notation for the edge corrected and original functions. This construction provides an unconditional basis for the $B_{s,p,q}[0,1]$ space if $sp > 1$.

It is helpful to use also for $\phi_{j_0,k}$ the notation ψ_k, $k = 1, \dots, 2^{j_0}$. Denote also by $\mathscr{J}$ the set of resolution levels for the considered wavelet basis,

$$\mathscr{J} = \{j \geq j_0\}$$

and let $\mathscr{I}_j$ be the index set for the jth level,

$$\mathscr{I}_{j_0} = \{k: k = 1, \dots, 2^{j_0}\} \cup \{(u_0, k): k = 1, \dots, 2^{j_0}\},$$
$$\mathscr{I}_j = \{(j,k): k = 1, \dots, 2^j\}.$$

By $\mathscr{I}$ we denote the global index set for the considered basis, $\mathscr{I} = \{\mathscr{I}_j, j \in \mathscr{J}\}$. Now the wavelet decomposition of a function f can be represented in the form

$$f(t) = \sum_{I \in \mathscr{I}} \theta_I \psi_I(t) = \sum_{j \in \mathscr{J}} \sum_{I \in \mathscr{I}_j} \theta_I \psi_I(t),$$

where θ_I is the Ith wavelet coefficient,

$$\theta_I = \int_0^1 f(t) \psi_I(t)\, dt, \qquad I \in \mathscr{I}.$$

Let now X_I, $I \in \mathscr{I}$ be empirical wavelet coefficients for the model (2.1),

$$X_I = \int_0^1 \psi_I(t)\, dX(t).$$

The model equation (2.1) yields

$$X_I = \theta_I + \varepsilon \int_0^1 \psi_I(t)\, dW(t)$$

and the original functional model (2.1) is translated into the sequence space model

$$X_I = \theta_I + \varepsilon \xi_I, \qquad I \in \mathscr{I}, \tag{3.1}$$

where $\xi_I = \int \psi_I\, dW$ are standard normal and independent for different I. The wavelet transform is justified by the following (isometric) properties [cf. Triebel (1992), page 240].

(ISO1) For any function $f \in L_2[0,1]$,

$$\|f\|^2 = \|\theta\|^2 := \sum_{\mathscr{I}} \theta_I^2. \tag{3.2}$$

(ISO2) There are two constants C_1 and C_2 such that

$$C_1 \|f\|_{B_{s,p,q}} \leq \|\theta\|_{b_{s,p,q}} \leq C_2 \|f\|_{B_{s,p,q}},$$

where

$$\|\theta\|_{b_{s,p,q}} = \begin{cases} \left\{ \sum_{j \ge j_0} \left[2^{j(s+1/2-1/p)} \left(\sum_{\mathcal{I}_j} |\theta_I|^p \right)^{1/p} \right]^2 \right\}^{1/q}, & q < \infty, \\ \sup_{j \ge j_0} \left\{ 2^{j(s+1/2-1/p)} \left(\sum_{\mathcal{I}_j} |\theta_I|^p \right)^{1/p} \right\}, & q = \infty. \end{cases} \tag{3.3}$$

3.2. *A minimax test.* First we restrict the considered set of wavelet coefficients $\mathcal{I}$ by some subset $\mathcal{I}_\varepsilon$. This procedure is typical for statistical analysis based on wavelet technique; see for example Donoho, Johnstone, Kerkyacharian and Picard (1994).

Define the level j_ε as the minimal integer with

$$2^{j_\varepsilon} \ge \varepsilon^{-2}.$$

Set now

$$\mathcal{J}_\varepsilon = \{ j \in \mathcal{J} : j \le j_\varepsilon \},$$
$$\mathcal{I}_\varepsilon = \bigcup_{j \in \mathcal{J}_\varepsilon} \mathcal{I}_j.$$

It is convenient to introduce also the "normalized" observations $Y_I = \varepsilon^{-1} X_I$, that is due to (3.1),

$$Y_I = \varepsilon^{-1}\theta_I + \xi_I.$$

Denote for each $j \in \mathcal{J}$,

$$S_j = \varepsilon^{-2} \sum_{\mathcal{I}_j} \left(X_I^2 - \varepsilon^2 \right) = \sum_{\mathcal{I}_j} \left(Y_I^2 - 1 \right). \tag{3.4}$$

Given $\lambda > 0$, set also

$$\begin{aligned} S_j(\lambda) &= \varepsilon^{-2} \sum_{\mathcal{I}_j} \left[X_I^2 \mathbf{1}(|X_I| > \varepsilon\lambda) - \varepsilon^2 b(\lambda) \right] \\ &= \sum_{\mathcal{I}_j} \left[Y_I^2 \mathbf{1}(|Y_I| > \lambda) - b(\lambda) \right]. \end{aligned} \tag{3.5}$$

Here

$$b(\lambda) = E\left[\xi^2 \mathbf{1}(|\xi| > \lambda) \right]$$

and ξ means the standard normal variable.

Given $\sigma = (s, p, q, M)$, define the level $J = J(\sigma) \in \mathcal{J}$ by

$$2^{-J} = \left(\frac{\varepsilon}{M} \right)^{4/(4s''+1)}, \tag{3.6}$$

that is,

$$J = (s'' + 1/4)^{-1} \log_2(M/\varepsilon).$$

We assume without loss of generality that the right-hand side of this equality is an integer. Otherwise one can take its integer part. Obviously J depends on ε and J tends to infinity as ε tends to zero. In what follows we assume ε to be small enough and $J > j_0$.

Let $\mathscr{J}_+$ and $\mathscr{J}_-$ be the partition of the level set $\mathscr{J}_\varepsilon$ into two parts: above and below J:

$$\mathscr{J}_+ = \{j: j_0 \le j < J\}, \qquad \mathscr{J}_- = \{j \in \mathscr{J}_\varepsilon: j \ge J\}.$$

Now put for $j \in \mathscr{J}_-$,

$$\lambda_j = 4\sqrt{(j - J + 8)\ln 2}, \qquad j \ge J$$

and introduce the test statistics $T(J)$ given by

$$T(J) = 2^{-J/2}\Bigg[\sum_{j\in\mathscr{J}_+} S_j + \sum_{j\in\mathscr{J}_-} S_j(\lambda_j)\Bigg]. \tag{3.7}$$

The test ϕ^* is defined by

$$\phi^* = \mathbf{1}\big(T(J) > v(J)\chi_{\alpha_0}\big), \tag{3.8}$$

where χ_α is the $(1-\alpha)$-quantile of the standard normal law,

$$v^2(J) = 2^{-J+1}\Bigg[2^{j_0} + \sum_{j\in\mathscr{J}_+} 2^j + \sum_{j\in\mathscr{J}_-} 2^j d(\lambda_j)\Bigg] \tag{3.9}$$

and

$$d(\lambda) = \tfrac{1}{2}E\big[\xi^2\mathbf{1}(|\xi| > \lambda) - b(\lambda)\big]^2.$$

We finish describing the test ϕ^* by a few remarks.

REMARK 3.1. The test ϕ^* depends on $\sigma = (s, p, q, M)$ and ε, but this dependence is only through the value J.

REMARK 3.2. It is easy to check that $v(J)$ converges as $\varepsilon \to 0$ to the value v with

$$v^2 = 2 + \sum_{k=0}^{\infty} 2^k d\big(4\sqrt{(k+8)\ln 2}\big).$$

Hence this universal constant v can be used in place of $v(J)$ for the test ϕ^*.

REMARK 3.3. The choice of the thresholds λ_j of the form $G\sqrt{j-J}$ was proposed for the estimation problem in Delyon and Juditsky (1995).

3.3. *An adaptive test*. Now we describe the structure of the test ϕ_ε from Theorem 2.4.

The idea of the test is quite clear. For each set $\sigma = \{s, p, q, M\}$, one may determine the level $J(\sigma)$ and the corresponding test procedure ϕ^* from the above. Therefore, the range of adaptation $\mathscr{T}$ can be translated into a range $\mathscr{J}'_\varepsilon$ of the form $\mathscr{J}'_\varepsilon = \{J: J_{\min} \le J \le J_{\max}\}$ and for each $J \in \mathscr{J}'_\varepsilon$, we are given the test procedure $\phi(J)$. Our adaptive method can be viewed as follows: each test

$\phi(J)$ is to be applied independently and the whole procedure rejects the null hypothesis if at least one test does. The problem here is that each test has a finite error probability of the first kind, and the corresponding error probability of this composite procedure is too large. To cope with this, we take the threshold value for each test with an extra growth factor.

More precisely, let $J_{\min}, J_{\max}$ be taken by

$$J_{\min} = (s_{\max} + 1)^{-1} \log_2 \varepsilon^{-2},$$
$$J_{\max} = \log_2 \varepsilon^{-2}.$$

and

$$\mathscr{J}'_{\varepsilon} = \{J\colon J_{\min} \le J \le J_{\max}\}. \tag{3.10}$$

It is easy to see that $J(\sigma) \in \mathscr{J}'_{\varepsilon}$ for any $\sigma \in \mathscr{T}$. Obviously

$$m_{\varepsilon} = \#(\mathscr{J}'_{\varepsilon}) \le \log_2 \varepsilon^{-2}. \tag{3.11}$$

Let also $T(J)$ and $v(J)$ be defined by (3.7) and (3.9), respectively. Define the following test:

$$\phi_{\varepsilon} = \mathbf{1}\left(\sup_{J \in \mathscr{J}^*} T(J) v^{-1}(J) > 2\sqrt{\ln \ln \varepsilon^{-2}} \right). \tag{3.12}$$

REMARK 3.4. Now we are in a position to explain the nature of the log log-factor entering in the adaptive rate of testing. Later we will see that $T(J)/v(J)$ are under the null, asymptotically standard normal and, moreover, they are weakly dependent for different J. Hence our test statistic in (3.12) is the supremum of m_{ε} weakly dependent asymptotically Gaussian random variables and its distribution is degenerate around

$$\sqrt{2 \ln m_{\varepsilon}} \approx \sqrt{2 \ln \ln \varepsilon^{-2}}\,.$$

This explains the choice of the testing level in (3.12).

4. Proofs. In this section we prove Theorems 2.3 and 2.4. The result of Theorem 2.1 for the proposed test ϕ^*_{ε} can be easily deduced from the proof of Theorem 2.4.

Throughout this section, we identify the function f with the set of the corresponding wavelet coefficients $\theta = \{\theta_I,\ I \in \mathscr{I}\}$. Due to ISO2, one may translate the smoothness condition of the form $\|f\|_{B_{s,p,q}} \le M$ into the condition

$$\theta \in \Theta_{\sigma} = \left\{\theta\colon \|\theta\|_{b_{s,p,q}} \le M\right\}. \tag{4.1}$$

4.1. *Proof of Theorem* 2.4. First we study the behavior of the test ϕ_{ε} under H_0, that is for $\theta = 0$.

Let the level sets $\mathscr{J}_{\varepsilon} = \{j\colon j_0 \le j \le j_{\varepsilon}\}$ and $\mathscr{J}'_{\varepsilon} = \{J\colon J_{\min} \le J \le J_{\max}\}$ be as introduced in Section 3.3. In Lemma 4.1, we identify S_j from (3.4) with $S_j(\lambda)$ from (3.5) for $\lambda = 0$.

LEMMA 4.1. *The following conditions hold true under* H_0:

(i) *For any* $\lambda \geq 0$ *and each* $j \in \mathscr{J}_\varepsilon$,

$$ES_j(\lambda) = 0,$$
$$ES_j^2(\lambda) = 2^j d(\lambda),$$

where $d(\lambda)$ *is from* (3.9) *and particularly* $d(0) = 1$.

(ii) *The random variables* $S_j(\lambda_j)$ *are independent for different* j *and any* λ_j.

(iii) *Uniformly in* $j \in \mathscr{J}'_\varepsilon$ *and* $|t| \leq 2\ln \varepsilon^{-2}$,

$$\frac{P_0\left(2^{-j/2}S_j > t\right)}{1 - \Phi(t)} \to 1, \qquad \varepsilon \to 0.$$

PROOF. The first two statements follow directly from the definition (3.4). The last statement is an easy consequence of the general results on the rate of convergence in the central limit theorem for i.i.d. random variables [see, e.g., Amosova (1972)]. The only important fact here is that $2^j \to \infty$ uniformly in $j \in \mathscr{J}'_\varepsilon$ and each summand in S_j has finite moments. □

The next technical result describes the behavior of the test statistics $T(J)$ under H_0.

LEMMA 4.2. *The following statements are fulfilled uniformly in* $J \in \mathscr{J}'_\varepsilon$.

(i)
$$ET(J) = 0,$$
$$ET^2(J) = v^2(J);$$

(ii) *Uniformly in* $|t| \leq 2\ln \varepsilon^{-2}$,

$$\frac{P_0\left(v^{-1}(J)T(J) > t\right)}{1 - \Phi(t)} \to 1, \qquad \varepsilon \to 0.$$

PROOF. The first statement of the lemma can be readily checked using (i) and (ii) of Lemma 4.1. The second statement is again an application of general results on the rate of convergence in the central limit theorem [see Petrov (1975)]. □

The last lemma yields the desirable property of the test ϕ_ε under H_0. In fact, by (ii),

$$\begin{aligned} P_0(\phi_\varepsilon = 1) &\leq \sum_{J \in \mathscr{J}'_\varepsilon} P\left(T(J) > 2v(J)\sqrt{\ln\ln \varepsilon^{-2}}\right) \\ &\leq \sum_{J \in \mathscr{J}'_\varepsilon} \exp\left\{-\frac{1}{2} 4 \ln\ln \varepsilon^2\right\} \\ &= \frac{\#(\mathscr{J}'_\varepsilon)}{\left(\ln \varepsilon^{-2}\right)^2} \leq \frac{\ln \varepsilon^{-2}}{\left(\ln \varepsilon^{-2}\right)^2} \to 0, \qquad \varepsilon \to 0. \end{aligned}$$

Now we turn to studying the power of the test ϕ_ε. Denote

$$(4.2)\qquad \tilde{\varrho}_\sigma(\varepsilon) = \begin{cases} M^{1/(4s+1)}(\varepsilon t_\varepsilon)^{4s/(4s+1)}, & \text{if } p \ge 2,\\ M^{1/(4s''+1)}\varepsilon^{4s''/(4s''+1)} t_\varepsilon^{1-8/(p(4s''+1))}, & \text{if } p < 2,\end{cases}$$

where, recall, $s'' = s + 1/4 - 1/(2p)$. Obviously $\tilde{\varrho}(\varepsilon) \le \varrho_\sigma(\varepsilon t_\varepsilon)$ and it suffices to check that for some $c > 0$ and any $\sigma \in \mathscr{T}$,

$$(4.3)\qquad \beta_\sigma(\phi_\varepsilon, c\tilde{\varrho}_\sigma(\varepsilon)) = o_\varepsilon(1).$$

Let us fix some $\sigma = (s, p, q, M) \in \mathscr{T}$ and some $\theta \in \Theta_\sigma$, that is, $\|\theta\|_{b_{s,q,q}} \le M$. Define the level $J = J(\sigma)$ by the equality

$$(4.4)\qquad 2^{-J} = \begin{cases} (\varepsilon t_\varepsilon/M)^{4/(4s+1)}, & \text{if } p \ge 2,\\ \left(\varepsilon t_\varepsilon^{2/p}/M\right)^{2/(2s+1-1/p)}, & \text{if } 1 \le p < 2.\end{cases}$$

We will examine the behavior of the statistic $T(J)$ under P_θ. The goal is to show that for θ from the alternative set, one has with a large P_θ-probability $T(J) > 2v(J)t_\varepsilon^2$ that obviously yields the desired assertion.

For the proof, we use the following decomposition:

$$T(J) = E_\theta T(J) + T(J) - E_\theta T(J).$$

Denote

$$\gamma = \frac{1}{\tilde{\varrho}_\sigma(\varepsilon)}\theta.$$

The condition $\|\theta\|^2 \ge c|\tilde{\varrho}_\sigma(\varepsilon)|^2$ can be rewritten as

$$\|\gamma\|^2 \ge c.$$

We will show that for $\theta \in \Theta_\sigma$ one has

$$(4.5)\qquad E_\theta T(J) \ge \left[\tfrac{1}{2}\|\gamma\|^2 - c_1(\sigma)\right] t_\varepsilon^2$$

with some constant $c_1(\sigma)$ depending only on σ and uniformly bounded for $\sigma \in \mathscr{T}$. We will also prove that for ε small enough,

$$(4.6)\qquad \mathbf{D}_\theta T(J) := E_\theta[T(J) - E_\theta T(J)]^2 \le 4 + \|\gamma\|^2.$$

Finally we prove that $T(J)$, being centered and normalized, is asymptotically normal under P_θ. Namely, if

$$\zeta(J) = \frac{T(J) - E_\theta T(J)}{\sqrt{\mathbf{D}_\theta T(J)}},$$

then uniformly in $|t| < \ln \varepsilon^{-2}$,

$$(4.7)\qquad \frac{P(-\zeta(J) > t)}{1 - \Phi(t)} = 1 - o_\varepsilon(1).$$

These statements will be proved later on. Now we explain how they imply the assertion of the theorem. Indeed

$$\begin{aligned} P_\theta(\phi_\varepsilon = 0) &\le P_\theta\big(T(J) < 2v(J)t_\varepsilon^2\big) \\ &\le P_\theta\Big(E_\theta T(J) + \zeta(J)\sqrt{\mathbf{D}_\theta T(J)} < 2v(J)t_\varepsilon^2\Big) \\ &\le P_\theta\left(-\zeta(J) > \frac{E_\theta T(J) - 2v(J)t_\varepsilon^2}{\sqrt{\mathbf{D}_\theta T(J)}}\right). \end{aligned}$$

To prove our assertion, by (4.7), it suffices to check that

$$\frac{E_\theta T(J) - 2v(J)t_\varepsilon^2}{\sqrt{\mathbf{D}_\theta T(J)}} \to \infty, \qquad \varepsilon \to 0.$$

But if $\theta \in \Theta_\sigma$ is such that

$$\|\gamma\|^2 = \frac{\|\theta\|^2}{\tilde{\varrho}_\sigma^2(\varepsilon)} \ge 3c_1(\sigma) + 6v(J),$$

then by (4.5) and (4.6),

$$\frac{E_\theta T(J) - 2v(J)t_\varepsilon^2}{\sqrt{\mathbf{D}_\theta T(J)}} \ge \frac{t_\varepsilon^2\big(\|\gamma\|^2/2 - c_1(\sigma) - 2v(J)\big)}{2 + \|\gamma\|} \to \infty, \qquad \varepsilon \to 0.$$

To check (4.5) and (4.6) we use the following consequence of the smoothness condition $\|\theta\|_{b_{s,p,q}} \le M$.

LEMMA 4.3. *Let $\theta \in \Theta_\sigma$ and let λ_j be defined by* (3.2), $j \in \mathscr{J}$. *Then the following conditions hold*:

(i) $$2^{-J/2} \sum_{j \in \mathscr{J}_-} \sum_{\mathscr{I}_j} \varepsilon^{-2}\theta_I^2 \mathbf{1}\big(|\theta_I| \le \lambda_j \varepsilon\big) \le c_2(\sigma)t_\varepsilon^2;$$

(ii) $$2^{-J/2} \sum_{j \in \mathscr{J}_-} \sum_{\mathscr{I}_J} \mathbf{1}\big(|\theta_I| \ge \lambda_j \varepsilon\big) \le c_3(\sigma)t_\varepsilon^2;$$

where $c_3(\sigma) \le 2$.

(iii) *Uniformly in* $J \in \mathscr{J}'_\varepsilon$,

$$\varepsilon^{-2}2^{-J/2}\left[\|\theta\|^2 - \sum_{j \in \mathscr{J}_\varepsilon} \sum_{\mathscr{I}_j} \theta_I^2\right] \le 2M^2 2^{-J/2} \to 0.$$

PROOF. Consider first the case $p < 2$. The condition $\theta \in \Theta_\sigma$ yields for each $j \in \mathscr{J}$ [see (3.3)]:

$$\sum_{\mathscr{I}_j} |\theta_I|^p \le 2^{-js'p}M^p, \tag{4.8}$$

s' being $s + 1/2 - 1/p$.

Now

$$2^{-J/2}\sum_{j\in\mathscr{J}_-}\sum_{\mathscr{I}_J}\varepsilon^{-2}\theta_I^2\mathbf{1}\big(|\theta_I|\le\lambda_j\varepsilon\big)\le 2^{-J/2}\sum_{j\in\mathscr{J}_-}\sum_{\mathscr{I}_j}\varepsilon^{-p}|\theta_I|^p\lambda_j^{2-p}\mathbf{1}\big(|\theta_I|\le\lambda_j\varepsilon\big)$$
$$\le 2^{-J/2}\sum_{j\in\mathscr{J}_-}\varepsilon^{-p}\lambda_j^{2-p}\sum_{\mathscr{I}_j}|\theta_I|^p$$
$$\le \varepsilon^{-p}2^{-J/2}\sum_{j\in\mathscr{J}_-}\lambda_j^{2-p}2^{-js'p}.$$

Note that

$$\sum_{j\in\mathscr{J}_-}\lambda_j^{2-p}2^{-js'p}\le 2^{-Js'p}\sum_{k=0}^{\infty}\big(4\sqrt{k+8}\big)^{2-p}2^{-ks'p}\le c_2(\sigma)2^{-Js'p}.$$

Here $c_2(\sigma)$ is the latest sum and for $s'\ge 1/2$, one gets very roughly $c_2(\sigma)\le 48$.

Next, using the definition (4.4) of J and the equality $s'+1/(2p)=s+1/2+1/(2p)=s''+1/4$, one gets

$$(M/\varepsilon)^p 2^{-J/2}2^{-Js'p}=(M/\varepsilon)^p\left(\frac{\varepsilon t_\varepsilon^{2/p}}{M}\right)^{(s'p+1/2)/(s''+1/4)}=t_\varepsilon^2$$

and (i) is proved for $p<2$.

The case $p\ge 2$ can be considered in the same way, substituting everywhere 2 in place of p.

To check (ii) we note that for each j by (4.8),

$$\sum_{\mathscr{I}_j}\mathbf{1}\big(|\theta_I|\ge\lambda_j\varepsilon\big)\le\sum_{\mathscr{I}_j}\big(\lambda_j\varepsilon\big)^{-p}|\theta_I|^p\le\big(\lambda_j\varepsilon\big)^{-p}M^p2^{-js'p}.$$

We proceed further as before and, moreover, one can easily estimate $c_3(\sigma)\le 2$.

It remains to check (iii). Let j_ε be the latest resolution level in $\mathscr{J}_\varepsilon$. Using again (4.8) we obtain for any $j>j_\varepsilon$,

$$\sum_{\mathscr{I}_j}\theta_I^2\le\left[\sum_{\mathscr{I}_j}|\theta_I|^p\right]^{2/p}\le M^2 2^{-2js'}.$$

Recall that by definition $2^{-j_\varepsilon}\le\varepsilon^2$ and also the condition $sp>1$ gives $s'>1/2$. Hence

$$\varepsilon^{-2}2^{-J/2}\left[\|\theta\|^2-\sum_{j\in\mathscr{J}_\varepsilon}\sum_{\mathscr{I}_j}\theta_I^2\right]\le M^2 2^{-J/2}\varepsilon^{-2}\sum_{j>j_\varepsilon}2^{-2js'}\tag{4.9}$$
$$\le 2M^2 2^{-J/2}\to 0.$$

This completes the proof of the lemma. □

Now we are ready to show (4.5). One has

$$E_\theta T(J) = 2^{-J/2}\left[\sum_{j\in\mathscr{J}_+}\sum_{\mathscr{I}_J} E_\theta\left(Y_I^2 - 1\right) + \sum_{j\in\mathscr{J}_-}\sum_{\mathscr{I}_j} E_\theta\left[Y_1^2 \mathbf{1}\left(|Y_I| > \lambda_j\right) - b\left(\lambda_j\right)\right]\right],$$

where $Y_I = \varepsilon^{-1}\theta_I + \xi_I$.

The random errors ξ_I are standard normal and obviously

$$E_\theta\left(Y_I^2 - 1\right) = \varepsilon^{-2}\theta_I^2.$$

To estimate the second sum in (4.1), we use the following property of the standard normal law.

LEMMA 4.4. *For any $\lambda > 0$ and each y,*

$$B(y, \lambda) := E(y + \xi)^2 \mathbf{1}(|y + \xi| > \lambda) - E\xi^2 \mathbf{1}(|\xi| > \lambda) \geq \tfrac{1}{2} y^2 \mathbf{1}(|y| > \lambda).$$

PROOF. We assume without loss of generality that $y \geq 0$. It is easy to see that

$$yE\xi \mathbf{1}(|y + \xi| > \lambda) \geq 0$$

and

$$E\xi^2 \mathbf{1}(|y + \xi| > \lambda) - E\xi^2 \mathbf{1}(|\xi| > \lambda) \geq 0.$$

This yields

$$B(y, \lambda) \geq y^2 P(|y + \xi| > \lambda) \geq \tfrac{1}{2} y^2 \mathbf{1}(|y| > \lambda). \qquad \square$$

By this lemma for each $j \in \mathscr{J}_-$,

$$\begin{aligned}\sum_{\mathscr{I}_j} E_\theta\left[Y_I^2 \mathbf{1}\left(|Y_I| > \lambda_j\right) - b\left(\lambda_j\right)\right] &\geq \tfrac{1}{2}\sum_{\mathscr{I}_j} \varepsilon^{-2}\theta_I^2 \mathbf{1}\left(|\theta_I| > \lambda_j \varepsilon\right) \\ &= \frac{\varepsilon^{-2}}{2}\sum_{\mathscr{I}_j}\theta_I^2 - \frac{\varepsilon^{-2}}{2}\sum_{\mathscr{I}_j}\theta_I^2 \mathbf{1}\left(|\theta_I| < \lambda_j \varepsilon\right)\end{aligned}$$

Now applying Lemma 4.3 we obtain

$$\begin{aligned}ET(J) &\geq 2^{-J/2}\varepsilon^{-2}\left[\sum_{j\in\mathscr{J}_+}\sum_{\mathscr{I}_j}\theta_I^2 + \tfrac{1}{2}\sum_{j\in\mathscr{J}_-}\sum_{\mathscr{I}_j}\theta_I^2 - \tfrac{1}{2}\sum_{j\in\mathscr{J}_-}\sum_{\mathscr{I}_J}\theta_I^2 \mathbf{1}\left(|\theta_I| < \lambda_j \varepsilon\right)\right] \\ &\geq \tfrac{1}{2}\left[2^{-J/2}\varepsilon^{-2}\|\theta\|^2 - M^2 2^{-J/2} - c_2(\sigma) t_\varepsilon^2\right].\end{aligned}$$

The definition (4.4) of J gives by (4.2) for $p \geq 2$,

$$2^{-J/2}\varepsilon^{-2} = \varepsilon^{-2}\left(\frac{\varepsilon t_\varepsilon}{M}\right)^{2/(4s+1)} = \left(\varepsilon t_\varepsilon\right)^{-8s/(4s+1)} M^{2/(4s+1)} t_\varepsilon^2 = |\varrho_\sigma(\varepsilon t_\varepsilon)|^{-2} t_\varepsilon^2.$$

For $p < 2$, one has similarly

$$2^{-J/2}\varepsilon^{-2} = \varepsilon^{-2}\left(\varepsilon t_\varepsilon^{2/p}/M\right)^{2(4s''+1)} = |\tilde{\varrho}_\sigma(\varepsilon)|^{-2} t_\varepsilon^2$$

that completes the proof of (4.5).

The next step is in estimating $\mathbf{D}_\theta T(J)$.
Since ξ_I and hence Y_I are independent for different I one gets

$$\mathbf{D}_\theta T(J) = 2^{-J}\left[\sum_{j\in\mathcal{J}_+}\sum_{\mathcal{I}_j}\mathbf{D}_\theta Y_I^2 + \sum_{j\in\mathcal{J}_-}\sum_{\mathcal{I}_j}\mathbf{D}_\theta\left(Y_I^2\mathbf{1}(|Y_I| \ge \lambda_j)\right)\right].$$

Obviously

$$\begin{aligned}\mathbf{D}_\theta Y_I^2 &= E\left(|\varepsilon^{-1}\theta_I + \xi_I|^2 - E|\varepsilon^{-1}\theta_I + \xi_I|^2\right)^2\\ &= E\left(2\varepsilon^{-1}\theta_I\xi_I + \xi_I^2 - 1\right)^2\\ &= 4\varepsilon^{-2}\theta_I^2 + 2.\end{aligned}$$

To estimate the value $\mathbf{D}_\theta(Y_I^2\mathbf{1}(|Y_I| > \lambda_j))$ we use the following technical assertion.

LEMMA 4.5. *For each y and any $\lambda \ge 2$,*

$$\mathbf{D}\left(|y + \xi|^2\mathbf{1}(|y + \xi| > \lambda)\right) \le 4y^2 + 2\mathbf{1}(|y| > \lambda/2) + \lambda^2 e^{-\lambda^2/8}.$$

PROOF. First we note that for any y, λ,

$$\mathbf{D}\left(|y + \xi|^2\mathbf{1}(|y + \xi| > \lambda)\right) \le \mathbf{D}|y + \xi|^2 = 4y^2 + 2.$$

Next, one has readily for $\lambda \ge 2$ and $|y| < \lambda/2$,

$$\begin{aligned}\mathbf{D}\left(|y + \xi|^2\mathbf{1}(|y + \xi| > \lambda)\right) &\le E|y + \xi|^4\mathbf{1}(|y + \xi| > \lambda)\\ &\le E|\lambda/2 + \xi|^4\mathbf{1}(|\lambda/2 + \xi| > \lambda)\\ &\le \lambda^4 e^{-\lambda^2/8}\end{aligned}$$

and the lemma follows. □

Applying this result, we get

$$\begin{aligned}\mathbf{D}_\theta T(J) &\le 2^{-J}\Bigg[\sum_{j\in\mathcal{J}_+}\sum_{\mathcal{I}_j}\left(4\varepsilon^{-2}\theta_I^2 + 2\right)\\ &\qquad + \sum_{j\in\mathcal{J}_-}\sum_{\mathcal{I}_j}\left(4\varepsilon^{-2}\theta_I^2 + 2\mathbf{1}\left(|\theta_I| > \lambda_j\varepsilon/2\right) + \lambda_j^4 e^{-\lambda_j^2/8}\right)\Bigg]\\ &\le 4\varepsilon^{-2}2^{-J}\|\theta\|^2 + 2^{-J+j_0} + 2^{-J}\sum_{j\in\mathcal{J}_+}2^{j+1}\\ &\qquad + 2^{-J+1}\left[\sum_{j\in\mathcal{J}_-}\sum_{\mathcal{I}_j}\mathbf{1}\left(|\theta_I| > \lambda_j\varepsilon/2\right) + \sum_{j\in\mathcal{J}_-}2^j\lambda_j^4 e^{-\lambda_j^2/8}\right].\end{aligned} \tag{4.10}$$

Obviously

$$2\sum_{j\in\mathscr{J}_+} 2^{j-J} \le 2\sum_{k=1}^{J} 2^{-k} < 2,$$

$$\sum_{j\in\mathscr{J}_-} 2^{j-J}\lambda_j^4 e^{-\lambda_j^2/8} \le \sum_{j=J}^{\infty} 2^{j-J}\big(4\sqrt{j-J+8}\big)^4 2^{-2(j-J+8)} \le \sum_{k=0}^{\infty} 2^{-k} = 2.$$

Also, by (ii) of Lemma 4.3

$$2^{-J/2}\sum_{j\in\mathscr{J}_-}\sum_{\mathscr{I}_j}\mathbf{1}\big(|\theta_I| \ge \lambda_j\varepsilon/2\big) \le 4t_\varepsilon^2$$

and similarly to the above

$$\varepsilon^{-2}2^{-J/2}\|\theta\|^2 = \frac{\|\theta\|^2}{\tilde{\varrho}_\sigma^2(\varepsilon)}t_\varepsilon^2 = \|\gamma\|^2 t_\varepsilon^2.$$

Combining all these estimates, we conclude for ε small enough,

$$\mathbf{D}_\theta T(J) \le 4\cdot 2^{-J/2}\|\gamma\|^2 t_\varepsilon^2 + 4 + 4\cdot 2^{-J/2}t_\varepsilon^2 \le 4 + \|\gamma\|^2$$

since $2^{-J/2}t_\varepsilon^2 \to 0$ as $\varepsilon \to 0$ uniformly in $J \in \mathscr{J}_\varepsilon$. Assertion (4.6) follows.

It remains to establish asymptotic normality of $\zeta(J)$ in the sense of (4.7). To this end, we note that $\zeta(J)$ is a centered and normalized sum of independent random variables having arbitrary number of moments. Moreover, it is not difficult to check that the third or fourth absolute moment of $\zeta(J)$ is bounded uniformly on ε, and the desirable asymptotic normality can be proved by application, for instance, the general results by Amosova (1972).

4.2. *Proof of Theorem* 2.3. To prove the lower bound from Theorem 2.3, we apply the Bayes approach which is usual for such statements; see Ingster (1993). We restrict ourselves to the case with $p \ge 2$.

We proceed as follows. First we change a given nontrivial parameter set $\mathscr{T}$ by a finite subset $\mathscr{T}_\varepsilon$ with the cardinality $N_\varepsilon = \#(\mathscr{T}_\varepsilon) \asymp \ln\varepsilon^{-2}$. Then for each $\sigma \in \mathscr{T}_\varepsilon$, we construct a prior measure π_σ which concentrates on the corresponding alternative set $\mathscr{A}_\sigma = \{\theta\colon \|\theta\|_{b_{s,p,q}} \le M, \|\theta\| \ge c\varrho_\sigma(\varepsilon t_\varepsilon)\}$,

$$\pi_\sigma(\mathscr{A}_\sigma) = 1. \tag{4.11}$$

The choice of the constant c here will be made precise below.

The whole prior π_ε is taken of the form

$$\pi_\varepsilon = \frac{1}{N_\varepsilon}\sum_{\sigma\in\mathscr{T}_\varepsilon}\pi_\sigma.$$

Let P_{π_ε} denote the Bayes measure for the prior π_ε. Obviously for any test ϕ,

$$\sup_{\sigma\in\mathscr{T}}\sup_{\theta\in\mathscr{F}_\sigma} P_\theta(\phi = 0) \ge P_{\pi_\varepsilon}(\phi = 0).$$

We will show that for a special choice of the set $\mathscr{T}_\varepsilon$ and the priors π_σ, $\sigma \in \mathscr{T}_\varepsilon$, one has for c small enough

$$Z_{\pi_\varepsilon} := \frac{dP_{\pi_\varepsilon}}{dP_0} \to 1 \tag{4.12}$$

under P_0-probability as $\varepsilon \to 0$. This yields for any test ϕ [see Lehmann (1959)],

$$P_0(\phi = 1) + P_{\pi_\varepsilon}(\phi = 0) \geq 1 - \varrho_\varepsilon(1)$$

and hence the result of the theorem.

Now we present the construction of the set $\mathscr{T}_\varepsilon$ and the priors π_ε satisfying (4.11) and (4.12). Let $\mathscr{T}$ be a nontrivial parameter set with the corresponding s_*, s^*, p, q, M. To be more definite and to simplify calculation we assume that $M = 1$.

Recall that in the case of $p \geq 2$ the adaptive rate is defined as

$$\tilde{\varrho}_\sigma(\varepsilon) = \varrho_\sigma(\varepsilon t_\varepsilon) = (\varepsilon t_\varepsilon)^{4s/(4s+1)}.$$

Let, given $\sigma = (s, p, q) \in \mathscr{T}$, the level $j(\sigma)$ be defined by the equation

$$2^{-j} = (c\varepsilon t_\varepsilon)^{4(4s+1)} \tag{4.13}$$

or

$$j(\sigma) = \frac{4}{4s+1} \log_2 (c\varepsilon t_\varepsilon)^{-1} \tag{4.14}$$

with some $c \in (0, 1)$.

As usual, if this expression is not an integer, we assume its integer part. Since $j(\sigma)$ depends on σ only through s, we will use also the notation $j(s)$.

Denote

$$\begin{aligned} j_* &= j(s_*), \\ j^* &= j(s^*), \\ \mathscr{J}(\mathscr{T}) &= \{ j \in \mathscr{J}: j_* \leq j \leq j^* \} \end{aligned}$$

and define for each $j \in \mathscr{J}(\mathscr{T})$ the value $\sigma_j = (s_j, p, q)$ by the equality $j = j(s_j)$ or

$$2^{-j} = (c\varepsilon t_\varepsilon)^{4/(4s_j+1)}. \tag{4.15}$$

The set $\mathscr{T}_\varepsilon$ consists of σ_j, $j \in \mathscr{J}(\mathscr{T})$. Now we define for each j a prior π_j which is concentrated on the level j. Namely, let $\vartheta = (\vartheta_I, I \in \mathscr{I})$ be a random signal (vector) with $\vartheta_I = 0$ for $I \notin \mathscr{I}_j$ and ϑ_I are i.i.d. within $\mathscr{I}_j$ with the Bernoulli distribution of the form

$$\pi_j(\vartheta_I = \pm u_\varepsilon) = 1/2,$$

where

$$u_\varepsilon = (c\varepsilon t_\varepsilon)^{(4s+2)/(4s+1)}. \tag{4.16}$$

First we check the condition (4.11) for these priors. One has obviously

$$\|\vartheta\|^2 = \sum_{\mathscr{I}_j} u_\varepsilon^2 = 2^j u_\varepsilon^2$$

and by (4.13) with $s = s_j$ and by (4.15)

$$2^j u_\varepsilon^2 = (c\varepsilon t_\varepsilon)^{-4/(4s+1)+8s/(4s+1)} = (c\varepsilon t_\varepsilon)^{8s/(4s+1)} = c' r_\sigma^2(\varepsilon t_\varepsilon)$$

with $\sigma = \sigma_j = (s_j, p, q)$ and $c' = c^{8s/(4s+1)}$.

Next, in the same way

$$\begin{aligned}\|\vartheta\|_{b_{s,p,q}}^p &= 2^{js'p}\sum_{\mathscr{I}_j} u_\varepsilon^p \\ &= 2^{(s+1/2-1/p)jp} 2^j u_\varepsilon^p \\ &= (c\varepsilon t_\varepsilon)^{-(4p(s+1/2))/4s+1}(c\varepsilon t_\varepsilon)^{(4s+2)/(4s+1)p} = 1.\end{aligned}$$

This means that $\pi_j(\vartheta \in \mathscr{F}_{\sigma_j}) = 1$ and (4.11) is proved.

At the next step we evaluate the asymptotic expansion of the log-likelihood $\ln(dP_{\pi_j}/dP_0)$ for each $j \in \mathscr{J}(\mathscr{T})$. Denote

$$l_\varepsilon = c^2 t_\varepsilon^2.$$

LEMMA 4.6. *The following expansion holds true uniformly in* $j \in \mathscr{J}(\mathscr{T})$ *under the measure* P_0:

$$\ln\frac{dP_{\pi_j}}{dP_0} - l_\varepsilon S_j + l_\varepsilon^2/2 \to 0, \qquad \varepsilon \to 0, \tag{4.17}$$

and

$$\sup_{|t| \le \ln \varepsilon^{-2}} \frac{P_0(S_j > t)}{1 - \Phi(t)} \to 1, \qquad \varepsilon \to 0.$$

Here $S_j = 2^{-j/2}\sum_{\mathscr{I}_j}(\xi_I^2 - 1)$.

PROOF. A similar expansion can be found in Ingster (1993) and we give only a sketch of the proof.

One has easily for the model (3.1) and the prior π_j,

$$L_j := \ln\frac{dP_{\pi_j}}{dP_0} = \sum_{\mathscr{I}_j} \ln\left(\frac{1}{2}\exp\{\varepsilon^{-1}u_\varepsilon \xi_I - \varepsilon^{-2}u_\varepsilon^2\} + \frac{1}{2}\exp\{-\varepsilon^{-1}u_\varepsilon \xi_I - \varepsilon^{-2}u_\varepsilon^2\}\right).$$

Using the Taylor expansion, one has readily

$$L_j = \sum_{\mathscr{I}_j}\left[\tfrac{1}{2}\varepsilon^{-2}u_\varepsilon^2(\xi_I^2 - 1) - \tfrac{1}{12}\varepsilon^{-4}u_\varepsilon^4\xi_I^4 + O(\varepsilon^{-6}u_\varepsilon^6\xi_I^6)\right]$$

Notice now that by the definitions (4.16) and (4.13),

$$\varepsilon^{-2}u_\varepsilon^2 = \varepsilon^{-2}(c\varepsilon t_\varepsilon)^{2(4s+2)/(4s+1)} = 2^{-j/2}c^2 t_\varepsilon^2 = 2^{-j/2} l_\varepsilon.$$

Then, uniformly in $j \in \mathscr{J}(\mathscr{T})$ by the law of large numbers

$$\varepsilon^{-4}u_\varepsilon^4 \sum_{\mathscr{I}_j}(\xi_I^4 - 3) = l_\varepsilon^2 2^{-j}\sum_{\mathscr{I}_j}(\xi_I^4 - 3) \to 0$$

and

$$\varepsilon^{-6} u_\varepsilon^6 \sum_{\mathscr{I}_j} \xi_I^6 = l_\varepsilon^3 2^{-3j/2} \sum_{\mathscr{I}_j} \xi_I^6 \to 0$$

as $\varepsilon \to 0$ under the measure P_0.

Finally we remark that $\varepsilon^{-2} u_\varepsilon^2 \sum_{\mathscr{I}_j} (\xi_I^2 - 1) = l_\varepsilon S_j$ and the lemma follows. □

Now we check (4.12). The definition of π_ε yields

$$Z_{\pi_\varepsilon} = \frac{1}{N_\varepsilon} \sum_{j \in \mathscr{J}(\mathscr{T})} Z_{\pi_j},$$

where

$$N_\varepsilon = \#(\mathscr{J}(\mathscr{T})) \approx \left(\frac{4}{4s_* + 1} - \frac{4}{4s^* + 1} \right) \log_2 (c \varepsilon t_\varepsilon)^{-1}$$

and for $c < 1$

$$\frac{1}{N_\varepsilon} \exp(l_\varepsilon^2) = \frac{1}{N_\varepsilon} \exp(c^4 \ln\ln \varepsilon^{-1}) = \frac{1}{N_\varepsilon} (\ln \varepsilon^{-2})^{c^4} \to 0, \qquad \varepsilon \to 0.$$

Now the statement (4.12) follows from the next general assertion.

LEMMA 4.7. *Let* $(\zeta_{in}, i, n \ge 1)$ *be a triangle array of independent random variables on a probability space* $(\Omega, \mathscr{F}, P)$ *such that*

$$\sup_{i \le n} \sup_{|t| \le 2\sqrt{\ln n}} \left| \frac{P(\zeta_{in} > t)}{1 - \Phi(t)} - 1 \right| \to 0, \qquad n \to \infty. \tag{4.18}$$

If the sequence l_n *be such that*

$$\frac{1}{n} \exp(l_n^2) \to 0, \qquad n \to \infty,$$

then the following convergence holds under the measure P:

$$\frac{1}{n} \sum_{i=1}^{n} \exp\{l_n \zeta_{in} - l_n^2/2\} \to 1.$$

PROOF. The statement of the lemma means the law of large numbers for the random variables

$$Z_{in} = \exp\{l_n \zeta_{in} - l_n^2/2\}.$$

For this, it suffices to check [see Petrov (1975)] that

$$E Z_{in} \mathbf{1}(|\zeta_{in}| \le 2 l_n) \to 1, \qquad n \to \infty,$$

and

$$\frac{1}{n^2} \sum_{i=1}^{n} \mathbf{D}(Z_{in} \mathbf{1}(|\zeta_{in}| \le 2 l_n)) \to 0, \qquad n \to \infty.$$

Using the condition (4.18), one may replace ζ_{in} in these statements by a standard normal ζ and Z_{in} by $Z = \exp\{l_n \zeta - l_n^2/2\}$. To complete the proof it remains to note that

$$E \exp\{l_n \zeta - l_n^2/2\} \mathbf{1}(|\zeta| \le 2 l_n) \to 1, \qquad n \to \infty$$

and

$$n^{-1} \mathbf{D} Z \mathbf{1}(|\zeta| \le 2 l_n) \le n^{-1} \mathbf{D} Z \le n^{-1} e^{l_n^2} \to 0, \qquad n \to \infty. \qquad \square$$

Acknowledgments. The author thanks O. Lepski, D. Picard, G. Kerkyacharian, A. Juditsky and M. Neumann for helpful remarks and discussion.

REFERENCES

AMOSOVA, N. N. (1972). On limit theorem for probabilities of moderate deviations. *Vestnik Leningrad. Univ.* **13** 5–14. (In Russian.)

BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398.

BROWN, L. D. and LOW, M. G. (1992). Superefficiency and lack of adaptability in functional estimation. Technical report, Cornell Univ.

COHEN, A., DAUBECHIES, I., JAWERTH, B. and VIAL, P. (1993). Multiresolution analysis, wavelets, and fast algorithms on an interval. *C. R. Acad. Sci. Paris Sér. I Math.* **316** 417–421.

COHEN, A., DAUBECHIES, B. and VIAL, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis Sér. I Math.* **1** 54–81.

DAUBECHIES, I. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia.

DELYON, B. and JUDITSKY, A. (1995). Wavelet estimators, global error measures: revisited. *J. Applied Comp. and Harmonic Analysis*. To appear.

DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455.

DONOHO, D. L. and JOHNSTONE, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90** 1200–1204.

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1994). Wavelet shrinkage: asymptopia? (with discussion). *J. Roy. Statist. Soc. Ser. B* **57** 301–369.

EFROMOVICH, S. and PINSKER, M. S. (1984). Learning algorithm for nonparametric filtering. *Automat. Remote Control* **11** 1434–1440.

ERMAKOV, M. S. (1990). Minimax detection of a signal in a white Gaussian noise. *Theory Probab. Appl.* **35** 667–679.

GOLUBEV, G. K. (1990). Quasilinear estimates of signal in L_2. *Problems Inform. Transmission* **26** 15–20.

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

INGSTER, YU. I. (1982). Minimax nonparametric detection of signals in white Gaussian noise. *Problems Inform. Transmission* **18** 130–140.

INGSTER, YU. I. (1984a). Asymptotic minimax testing of nonparametric hypothesis on the distribution density of an independent sample. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov* **136** 74–96. (In Russian.)

INGSTER, YU. I. (1984b). An asymptotic minimax test of nonparametric hypothesis about spectral density. *Theory Probab. Appl.* **29** 846–847.

INGSTER, YU. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives I–III. *Math. Methods Statist.* **2** 85–114; **3** 171–189; **4** 249–268.

KERKYACHARIAN, G. and PICARD D. (1993). Density estimation by kernel and wavelet method, optimality in Besov space. *Statist. Probab. Lett.* **18** 327–336.

LEHMANN, E. L. (1959). *Testing Statistical Hypothesis*. Wiley, New York.

LEPSKI, O. V. (1990). One problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 459–470.

LEPSKI, O. V. (1991). Asymptotic minimax adaptive estimation. 1. Upper bounds. *Theory Probab. Appl.* **36** 645–659.

LEPSKI, O. V., MAMMEN, E. and SPOKOINY, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* To appear.

LEPSKI, O. V. and SPOKOINY, V. G. (1995a). Optimal pointwise adaptive methods in nonparametric estimation. Unpublished manuscript.

LEPSKI, O. V. and SPOKOINY, V. G. (1995b). Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. *Bernoulli*. To appear.

MARRON, J. S. (1988). Automatic smoothing parameter selection: a survey. *Empir. Econ.* **13** 187–208.

MEYER, Y. (1990). *Ondelettes*. Herrmann, Paris.

NEUMANN, M. and SPOKOINY, V. (1995). On the efficiency of wavelet estimators under arbitrary error distributions. *Math. Methods Statist.* **4** 137–166.

NUSSBAUM, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430.

PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, New York.

POLJAK, B. T. and TSYBAKOV, A. B. (1990). Asymptotic optimality of C_p-test for the orthogonal series estimation of regression. *Theory Probab. Appl.* **35** 293–306.

TRIEBEL, H. (1992). *Theory of Function Spaces II*. Birkhäuser, Basel.

INSTITUTE FOR APPLIED ANALYSIS AND STOCHASTICS
MOHRENSTRASSE 39
BERLIN 10117
GERMANY
E-MAIL: spokoiny@wias-berlin.de