Volume 14, No. 3 (2005), pp. Allerton Press, Inc.

MATHEMATICAL METHODS OF STATISTICS

# LOCAL FUNCTIONAL HYPOTHESIS TESTING

F. Abramovich and R. Heller

Dept. Statistics and Operations Research, Tel Aviv University
Tel Aviv 69978, Israel
E-mail: felix@post.tau.ac.il, rheller@post.tau.ac.il

*We consider a standard "signal+white noise" model on the unit interval and want to test whether the signal is present on a subinterval $\Omega_\Delta \subseteq [0,1]$ of length $\Delta$. The composite alternative is that the unknown signal $f$ is separated away from zero in terms of its average power $\gamma(f) = \|f\|^2_\Delta/\Delta$ on $\Omega_\Delta$ and also possesses some regularity properties. We evaluate the asymptotically optimal (minimax) rates for testing the presence of a signal on $\Omega_\Delta$, where both the noise level and the interval length tend to zero. We derive corresponding rate-optimal tests for local signal detection.*

*Key words: adaptive testing, Besov spaces, functional hypothesis testing, local signal detection, minimax hypothesis testing, wavelets.*

*2000 Mathematics Subject Classification: primary 62G10, secondary 62G20.*

## 1. Introduction

Consider a standard "signal + white noise" model, where the data is a sample path of a stochastic process

$$dY(t) = f(t)\,dt + \varepsilon\,dW(t), \qquad t \in [0,1], \tag{1}$$

$f$ is an unknown signal and $W$ is a standard Wiener process. We wish to verify whether the data contains a signal in addition to the noise.

¿From a statistical perspective, detection of signal's presence is a functional hypothesis testing problem. The standard nonparametric functional hypothesis testing setting considers the *global* testing of the null hypothesis $H_0\colon f(t) \equiv 0$ on the whole unit interval against the composite nonparametric alternative that $f$ is separated away from zero in $L_2[0,1]$ norm, $\|f\|_{[0,1]} \geq \rho(\varepsilon)$, and also possesses some smoothness properties. For the prescribed error probabilities of Type I (erroneous rejection of $H_0$) and Type II (erroneous acceptance of $H_0$), the rate of decay of

$\rho(\varepsilon)$ as $\varepsilon \to 0$ is traditionally viewed as a natural measure of goodness of a test (see Ingster [9], [10]). The goal then is to find the minimal (optimal) $\rho(\varepsilon)$ for which such testing is still possible and to construct the rate-optimal test.

The corresponding global optimal (in the minimax sense) functional hypothesis testing procedures were first studied by Ingster [9] and further developed in Ermakov [6], Ingster [10], Spokoiny [16, 17], Lepski & Spokoiny [14], Ingster & Suslina [11] and Horowitz & Spokoiny [8] for various separation distances between the two hypotheses and different smoothness assumptions under the alternative. See Ingster & Suslina [12] for a comprehensive review. The equivalent non-asymptotic problem is addressed in Baraud [4].

In particular, for the Besov classes $B^s_{p,q}$ with $sp > 1$, Lepski & Spokoiny [14] showed that the optimal testing rate for this setting is

$$\rho^2(\varepsilon) = \varepsilon^{8s''/(4s''+1)}, \tag{2}$$

where $s'' = \min(s, s - 1/(2p) + 1/4)$, and derived the corresponding rate-optimal test. However, the proposed test required the knowledge of the parameters of the Besov class, which are typically unknown in practice. Spokoiny in [16] considered the problem of *adaptive* minimax testing, where the above parameters are unknown *a priori* but are assumed to lie within a given range. He showed that there is an unavoidable price to pay for adaptivity although the price is remarkably low. The adaptive testing rate is

$$\rho^2(\varepsilon) = \varepsilon^{8s''/(4s''+1)} \cdot (\log\log \varepsilon^{-2})^{2s'/(4s''+1)}, \tag{3}$$

where $s' = \min(s, s - 1/p + 1/2)$, which is only within a log-log factor of (2). Spokoiny constructed also the adaptive test that achieves the optimal rate (3).

Abramovich *et al.* [2] and Abramovich & Angelini [1] adapted the results for testing in the "signal+white noise" model (1) for detecting differences among signals or groups of signals within functional analysis of variance (FANOVA) framework.

However, in a variety of applications, the true signal $f$ in (1) has a local nature and one is interested in its localized detection, for example, within a local neighborhood of some specific point of interest. Similar problems often arise in FANOVA settings in detecting local differences between signals. For example, in seismic signal processing researchers try to determine the source of the signal (explosion or earthquake) by analysing its local behavior right after the onset time. In this paper we extend the results of Spokoiny [16] and Lepski & Spokoiny [14] mentioned above for the local testing when both the noise level $\varepsilon$ and the interval length $\Delta$ monotonically tend to zero.

Consider local functional hypothesis testing $H_0\colon f \equiv 0$ on a subinterval $\Omega_\Delta \subseteq [0, 1]$ of length $\Delta$, where *both* $\Delta$ and $\varepsilon$ tend to zero. As the interval gets shorter, more localized signals can be detected. However, when it becomes "too short", accurate detection is impossible. The detection ability depends obviously also on the local smoothness properties of $f$ on $\Omega_\Delta$. We assume that under the alternative $f$ belongs to a Besov ball of radius $M$ on $\Omega_\Delta$, $B^s_{p,q}(\Omega_\Delta, M)$, where $s > 0$, $1 \le p, q < \infty$, $sp > 1$ and $s > 1/2$ for $p \ge 2$.

For local testing it is more appropriate to define the separation of the alternative set from the null hypothesis in terms of the *average power* of a signal on $\Omega_\Delta$,

$\gamma(f) = \|f\|^2_{\Omega_\Delta}/\Delta$, which is also invariant under rescaling of $t$ in (1), rather than in terms of its $L_2(\Omega_\Delta)$ norm. A function $f$ from the alternative set satisfies then $\gamma(f) > \gamma(\varepsilon, \Delta)$ and for prescribed error probabilities of both types, the optimal rates of local testing procedures are measured in terms of $\gamma(\varepsilon, \Delta)$. The goal is to find the minimal $\gamma(\varepsilon, \Delta)$ for which testing is still possible and construct the optimal test.

Testing the null hypothesis against a constant alternative $H_1\colon f(t) = c$, $t \in \Omega_\Delta$, implies an obvious lower bound for $\gamma(\varepsilon, \Delta)$, which is the classical rate $\varepsilon^2/\Delta$. If $\varepsilon\Delta^{-1/2}$ does not tend to zero, the two hypotheses are asymptotically undistinguishable in the sense that for any test the sum of probabilities of its Type I and Type II errors tends to 1. The interval is "too short" to detect signals under the given noise level or, equivalently, the noise is "too strong" to detect signals on an interval of a given length.

For $\varepsilon\Delta^{-1/2} \to 0$ we show that the asymptotic minimax rate of local testing the null hypothesis against a general nonparametric alternative described above is

$$\gamma(\varepsilon, \Delta) = \begin{cases} \varepsilon^{8s''/(4s''+1)}\Delta^{-(2s+1)/(4s''+1)} & \text{if} \quad \varepsilon\Delta^{-s'} \to 0, \\ \varepsilon^2\Delta^{-1} & \text{otherwise}, \end{cases} \tag{4}$$

where $s' > 1/2$ was defined before. We consider also the *adaptive* local testing, where the parameters of the Besov ball under the alternative are not known *a priori*. As in global testing, the resulting adaptive test yields an additional log-log factor in the rate.

At first sight, it might seem somewhat paradoxical that when $\varepsilon\Delta^{-1/2}$ tends to zero slowly, so that $\varepsilon\Delta^{-s'} \not\to 0$, one obtains the classical rate but the rate surprisingly slows down as $\varepsilon\Delta^{-s'} \to 0$. We come back to this phenomenon with more rigorous arguments in Section 2.3 and provide here a somewhat intuitive explanation. Note that the geometry of the alternative set varies with both $\varepsilon$ and $\Delta$. For a fixed $\Delta$, when $\varepsilon > \Delta^{1/2}$, the whole alternative set is strongly compressed and covered by heavy noise, so that nothing can be extracted from the chaos. As the chaos disperses ($\varepsilon$ decreases), the alternative set spreads out and first distinguishable signals start to appear. When $\varepsilon$ is still fairly large, these signals are so simple that they can be detected at the classical rate. As dispersion continues, less and less primitive signals come out of the noise. At a certain critical point depending on the "complexity" of the alternative set (via $s'$) they already have such a complicated structure that cannot be detected at the classical rate with given accuracy any more. This explains slower rates in this case. These half-heuristic considerations certainly suffer from several drawbacks. For example, they treat $\Delta$ as fixed, while the results in (4) are asymptotic with respect to both $\varepsilon$ and $\Delta$. Still we find them useful to give one some insight into the above phenomenon.

The paper is organized as follows. Section 2 contains the main results including the definition of the local testing problem, its optimal testing rate $\gamma(\varepsilon, \Delta)$, and the non-adaptive and adaptive rate-optimal test procedures. The resulting tests are based on the empirical wavelet coefficients of the data and, in a way, can be viewed as the local versions of the corresponding global tests of Spokoiny [16]. Concluding remarks and discussion are made in Section 3. All the proofs are given in the Appendix.

## 2. Local Functional Testing

2.1. Local hypothesis testing problem. Consider again the white noise model (1). Let $\Omega_\Delta \subseteq [0,1]$ be a subinterval of length $\Delta$. We want to test the null hypothesis that there is no signal on $\Omega_\Delta$ :

$$H_0 \colon f(t) = 0, \quad \forall t \in \Omega_\Delta. \tag{5}$$

We do not specify any parametric form for $f$ under the alternative hypothesis and wish to test the null hypothesis (5) against as large a class of alternatives as possible. As in the estimation problem, we have to assume some regularity conditions on $f$ to distinguish it from completely irregular white noise. In particular, we assume that $f$ belongs to a Besov ball, $B^s_{p,q}(\Omega_\Delta, M)$, of radius $M$ on $\Omega_\Delta$. Besov classes are known to have exceptional expressive power: for particular choices of the parameters $s$, $p$, and $q$ they include the Hölder ($p = q = \infty$) and Sobolev ($p = q = 2$) classes of smooth functions, and functions of bounded variation sandwiched between $B^1_{1,1}$ and $B^1_{1,\infty}$. We refer to Meyer [15] for rigorous definitions and a detailed study of Besov spaces.

As we have mentioned in the Introduction, to be able to distinguish between the two hypotheses, the set of alternatives should also be separated away from zero in terms of the average power on $\Omega_\Delta$. We consider then the nonparametric alternative hypothesis of the form

$$H_1 \colon f \in \mathcal{F}(\gamma(\varepsilon, \Delta)) = \left\{ f \colon f \in B^s_{p,q}(\Omega_\Delta, M), \gamma(f) \geq \gamma(\varepsilon, \Delta) \right\}. \tag{6}$$

The first (regularity) constraint bounds the set of possible alternatives, while the second one cuts out the alternatives "too close" to the null hypothesis.

2.2. Minimax local testing rate. A (non-randomized) test $\phi$ is defined as a measurable function of the data with two values 0 and 1 that correspond to accepting and rejecting the null hypothesis respectively. As usual, the quality of the test $\phi$ is measured by the Type I and Type II errors. The probability of the Type I error is defined as

$$\alpha(\phi) = P_{f \equiv 0}(\phi = 1),$$

while the probability of the Type II error for the composite nonparametric alternative hypothesis $H_1$ is defined as

$$\beta(\phi, \gamma(\varepsilon, \Delta)) = \sup_{f \in \mathcal{F}(\gamma(\varepsilon, \Delta))} P_f(\phi = 0).$$

We focus on the asymptotic hypothesis testing problem as the noise level $\varepsilon$ and the length of the interval $\Delta$ monotonically tend to zero. Our aim is to evaluate the fastest rate of decay to zero of $\gamma(\varepsilon, \Delta)$ for which testing with prescribed $\alpha$ and $\beta$ is still possible.

Consider first testing the null hypothesis $H_0 : f(t) = 0$, $t \in \Omega_\Delta$, against a constant alternative $H_1 \colon f(t) = c$, where obviously $\gamma(f) = c^2$. Simple calculus shows that for the prescribed $\alpha$ and $\beta$ the testing is possible if $c > (Z_{1-\alpha} + Z_{1-\beta})\varepsilon\Delta^{-1/2}$, where $Z_{1-\alpha}$ and $Z_{1-\beta}$ are the corresponding quantiles of the standard normal distribution. The necessary condition for asymptotic (as $\varepsilon \to 0$ and $\Delta \to 0$) distinguishability (non-triviality) of these and the more general hypotheses (5) and (6) is, therefore, the requirement $\varepsilon\Delta^{-1/2} \to 0$.

To define the minimax rate $\gamma(\varepsilon, \Delta)$ for local hypothesis testing (5)–(6), consider the new variables $u = \varepsilon\Delta^{-1/2}$ and any other variable $v$, so that the corresponding Jacobian is not zero. For any two functions $g(\varepsilon, \Delta)$ and $g'(\varepsilon, \Delta)$ we say that $g'(\varepsilon, \Delta) = o_{\varepsilon\Delta^{-1/2}}(g(\varepsilon, \Delta))$ if $g'(u, v)/g(u, v) \to 0$ as $u \to 0$ for any fixed $v$.

**Definition 2.1.** A sequence $\gamma(\varepsilon, \Delta)$ is called the *minimax rate of local testing* if $\gamma(\varepsilon, \Delta) = o_{\varepsilon\Delta^{-1/2}}(1)$ as $\varepsilon \to 0$ and $\Delta \to 0$ monotonically and the following two conditions hold:

(i) for any $\gamma'(\varepsilon, \Delta) = o_{\varepsilon\Delta^{-1/2}}(\gamma(\varepsilon, \Delta))$, one has

$$\inf_{\phi_{\varepsilon,\Delta}} \left[\alpha(\phi_{\varepsilon,\Delta}) + \beta(\phi_{\varepsilon,\Delta}, \gamma'(\varepsilon, \Delta))\right] = 1 - o_{\varepsilon\Delta^{-1/2}}(1),$$

(ii) for any $\alpha > 0$ and $\beta > 0$ there exists a constant $c > 0$ and a test $\phi^*_{\varepsilon,\Delta}$ such that

$$\alpha(\phi^*_{\varepsilon,\Delta}) \le \alpha + o_{\varepsilon\Delta^{-1/2}}(1), \qquad \beta(\phi^*_{\varepsilon,\Delta}, c\gamma(\varepsilon, \Delta)) \le \beta + o_{\varepsilon\Delta^{-1/2}}(1).$$

The first condition states that local testing with a rate faster than $\gamma(\varepsilon, \Delta)$ is impossible, while the second one guarantees the existence of a test with the rate $\gamma(\varepsilon, \Delta)$.

Define $p' = \min(p, 2)$, $s' = s - 1/p' + 1/2$, and $s'' = s - 1/(2p') + 1/4$. To derive the minimax rate for the local testing on $\Omega_\Delta$, note that by time rescaling and normalization $\tilde{f}(u) = \sqrt{\Delta} f(\Delta u)$, $0 \le u \le 1$, the problem can be transformed to the global testing on the whole unit interval, where $\|\tilde{f}\|_{[0,1]} = \|f\|_{\Omega_\Delta}$. The Besov norm (the $L_p$-norm + the Besov semi-norm) is not homogeneous under rescaling and, as a result, a Besov ball $B^s_{p,q}(\Omega_\Delta, M)$ on $\Omega_\Delta$ does not transform to a Besov *ball* on $[0, 1]$. However, the semi-norm itself is homogenous and only it is, in fact, essential in nonparametric settings. In particular, all the existing results for global testing remain true when the Besov norms are replaced by the corresponding semi-norms. Furthermore, it is easy to show that under such rescaling, the resulting Besov radius will be $M\Delta^{s'}$ which, unlike the standard global testing setting with a *fixed* Besov radius, will tend to zero as $\Delta \to 0$. However, changing, in addition, the noise level $\varepsilon$ to $\tilde{\varepsilon} = \varepsilon\Delta^{-s'}$ implies a Besov ball of a fixed radius $M$. Thus, for $\varepsilon\Delta^{-s'} \to 0$, after the above transformations one can apply the corresponding results of Lepski & Spokoiny [14] for global testing (see (2)) to get the minimax rate $\gamma(\varepsilon, \Delta)$ in this case:

$$\gamma(\varepsilon, \Delta) = \Delta^{2s'-1}\gamma(\tilde{\varepsilon}, [0, 1]) = \Delta^{2s'-1}\tilde{\varepsilon}^{8s''/(4s''+1)} = \varepsilon^{8s''/(4s''+1)}\Delta^{-(2s+1)/(4s''+1)}.$$

In the following Section 2.3 we show that the case $\varepsilon\Delta^{-1/2} \to 0$ but $\varepsilon\Delta^{-s'} \not\to 0$ is similar to a paramteric one and propose a test that achieves the classical rate $\varepsilon^2\Delta^{-1}$.

Summarizing all the above arguments leads to the following Theorem 2.1 that establishes the minimax rate $\gamma(\varepsilon, \Delta)$ for the local hypothesis testing (5)–(6):

**Theorem 2.1.** *Given* $\theta = (s, p, q, M)$, *where* $sp > 1$, $p, q \geq 1$, *and* $s > 1/2$ *for* $p \geq 2$, *the minimax rate of local testing* (5)–(6) *as* $\varepsilon \to 0$, $\Delta \to 0$ *monotonically and* $\varepsilon \Delta^{-1/2} \to 0$ *is*

$$\gamma(\varepsilon, \Delta) = \begin{cases} \varepsilon^{8s''/(4s''+1)} \Delta^{-(2s+1)/(4s''+1)} & \text{if} \quad \varepsilon \Delta^{-s'} \to 0, \\ \varepsilon^2 \Delta^{-1} & \text{otherwise.} \end{cases} \tag{7}$$

2.3. Nonadaptive minimax local test. In this section we construct the rate-optimal test $\phi^*_{\varepsilon,\Delta}$ under the distinguishability condition $\varepsilon \Delta^{-1/2} \to 0$. Similarly to the rate-optimal test of Spokoiny [16] for global testing, the resulting test statistics will be based on the empirical wavelet coefficients of the data on $\Omega_\Delta$.

Given a compactly supported scaling function $\varphi$ of regularity $r > s$ and the corresponding mother wavelet $\psi$, one can generate an orthonormal wavelet basis $\{\varphi_{j_\Delta k},\, \psi_{jk};\ j \geq j_\Delta\}$ on the interval $\Omega_\Delta$, where $j_\Delta$ is the minimal integer such that $2^{-j_\Delta} |\operatorname{supp} \varphi| < \Delta$. In fact, asymptotically we may assume that $j_\Delta = \log_2 \Delta^{-1}$. The number of scaling coefficients, $C_\varphi$, is finite and depends on the support of the scaling function, while the number of wavelet coefficients on each resolution level $j$ is $[2^j \Delta] \sim 2^{j - j_\Delta}$ (see Anderson *et al.* [3] for details).

For clarity of exposition we use the same notation for interior and edge wavelets and in what follows denote $\varphi_{j_\Delta k}$ by $\psi_{j_\Delta - 1, k}$. Let $\mathcal{J} = \{j \geq j_\Delta - 1\}$ be the set of resolution levels for the considered wavelet basis and let $\mathcal{J}_j$ be the index set for the $j$th level:

$$\begin{aligned} \mathcal{J}_{j_\Delta - 1} &= \{(j_\Delta - 1, k) \colon k = 0, \ldots, C_\varphi - 1\}, \\ \mathcal{J}_j &= \{(j, k) \colon k = 0, \ldots, 2^{j - j_\Delta} - 1\}, \qquad j \geq j_\Delta. \end{aligned}$$

Then $f$ is expanded in the orthonormal wavelet series on $\Omega_\Delta$ as

$$f(t) = \sum_{j \in \mathcal{J}} \sum_{I \in \mathcal{J}_j} w_I \psi_I(t), \tag{8}$$

where $w_I = \int_{\Omega_\Delta} f(t) \psi_I(t)\, dt$.

¿From Parseval's identity,

$$\|f\|^2_{\Omega_\Delta} = \sum_{j \in \mathcal{J}} \sum_{I \in \mathcal{J}_j} w_I^2. \tag{9}$$

Moreover, wavelet series constitute unconditional bases for Besov spaces $B^s_{p,q}(\Omega_\Delta)$, $\max(0, 1/p - 1/2) < s < r$; $p, q \geq 1$, and the Besov norm of $f$ is equivalent to the corresponding sequence space norm of its wavelet coefficients:

$$\begin{aligned} \|f\|_{B^s_{p,q}} \asymp \|w\|_{b^s_{p,q}} = \Bigg( \sum_{I \in \mathcal{J}_{j_\Delta - 1}} |w_I|^p \Bigg)^{1/p} \\ + \begin{cases} \left( \sum_{j=j_\Delta}^\infty 2^{j(s + \frac{1}{2} - \frac{1}{p}) q} \left( \sum_{I \in \mathcal{J}_j} |w_I|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}}, & 1 \leq q < \infty, \\ \sup_{j \geq j_\Delta} 2^{j(s + \frac{1}{2} - \frac{1}{p})} \left( \sum_{I \in \mathcal{J}_j} |w_I|^p \right)^{\frac{1}{p}}, & q = \infty \end{cases} \end{aligned}$$

(e.g., Section 6.10 in Meyer [15], Anderson *et al.* [3]). The original testing setting (5)–(6) can therefore be equivalently reformulated in the wavelet domain, where the sparseness of wavelet bases over Besov spaces is exploited to significantly reduce the dimensionality of the problem.

Performing the wavelet transform of (1) on the interval $\Omega_\Delta$ one has

$$Y_I = w_I + \varepsilon\xi_I, \qquad I \in \mathcal{J}_j, \quad j \in \mathcal{J},$$

where the empirical wavelet coefficients of the data are $Y_I = \int_{\Omega_\Delta} \psi_I(t)\, dY(t)$ and $\xi_I \overset{\text{iid}}{\sim} N(0,1)$.

Let $j_\varepsilon$ be the maximal integer such that $j_\varepsilon < \log_2 \varepsilon^{-2}$. Asymptotically we can assume again that $j_\varepsilon = \log_2 \varepsilon^{-2}$ and that $j_\Delta < j_\varepsilon$ for sufficiently small $\varepsilon\Delta^{-1/2}$. Set $\mathcal{J}_\varepsilon = \{j \in \mathcal{J} : j \leq j_\varepsilon\}$. Let $j_\theta$ be the resolution level defined as

$$j_\theta = \frac{4}{4s'' + 1}\left( \log_2(M/\varepsilon) + \frac{1}{2p'} \log_2 \Delta^{-1} \right). \tag{10}$$

Since $sp > 1$ and $s > 1/2$ for $p \geq 2$, one can easily verify that $j_\theta < j_\varepsilon$ for sufficiently small $\varepsilon\Delta^{-1/2}$.

Consider first the case $j_\theta > j_\Delta$. Let $\mathcal{J}_\varepsilon = \mathcal{J}_- \cup \mathcal{J}_+$, where $\mathcal{J}_- = \{j_\Delta - 1, \ldots, j_\theta - 1\}$ and $\mathcal{J}_+ = \{j_\theta, \ldots, j_\varepsilon - 1\}$. As in Spokoiny [16], for each $j \in \mathcal{J}_-$, define $S_j$ to be

$$S_j = \sum_{I \in \mathcal{J}_j} (Y_I^2 - \varepsilon^2), \tag{11}$$

while, for each $j \in \mathcal{J}_+$ and for a given threshold $\lambda > 0$, define $S_j(\lambda)$ to be

$$S_j(\lambda) = \sum_{I \in \mathcal{J}_j} \left[ (Y_I^2 \mathbf{1}(|Y_I| > \varepsilon\lambda) - \varepsilon^2 b(\lambda) \right], \tag{12}$$

where $\mathbf{1}(A)$ is the indicator function of the set $A$, $b(\lambda) = E\left[\xi^2 \mathbf{1}(|\xi| > \lambda)\right] = 2(\Phi(-\lambda) + \lambda\phi(\lambda))$, $\xi$ is a $N(0,1)$ random variable, and $\Phi$ and $\phi$ are its probability and density function respectively.

With the above notation, introduce the following test statistics:

$$T(j_\theta) = \sum_{j \in \mathcal{J}_-} S_j \tag{13}$$

and

$$Q(j_\theta) = \sum_{j \in \mathcal{J}_+} S_j(\lambda_j), \tag{14}$$

where

$$\lambda_j = 4\sqrt{(j - j_\theta + 8)\log 2}. \tag{15}$$

Let $V_0^2(j_\theta)$ and $W_0^2(j_\theta)$ be the variances of $T(j_\theta)$ and $Q(j_\theta)$, respectively, under $H_0$. It is easy to see that

$$V_0^2(j_\theta) = 2\varepsilon^4 2^{j_\theta - j_\Delta} \qquad \text{and} \qquad W_0^2(j_\theta) = \varepsilon^4 \sum_{j \in \mathcal{J}_+} 2^{j-j_\Delta} d(\lambda_j),$$

where $d(\lambda) = E\big[\xi^4 \mathbf{1}(|\xi| > \lambda)\big] - b(\lambda)^2$ and

$$E\big[\xi^4 \mathbf{1}(|\xi| > \lambda)\big] = 6\Phi(-\lambda) + 2\lambda(3+\lambda^2)\phi(\lambda).$$

Finally, for a given significance level $\alpha \in (0,1)$, define the following test:

$$\phi^*_{\varepsilon,\Delta} = \mathbf{1}\left\{ \frac{T(j_\theta) + Q(j_\theta)}{\sqrt{V_0^2(j_\theta) + W_0^2(j_\theta)}} > Z_{1-\alpha} \right\}. \tag{16}$$

The resulting test statistic has a clear intuitive meaning and is essentially the standardized sum of squares of the thresholded empirical wavelet coefficients $Y_I$ with the properly chosen level-dependent thresholds. The coefficients on the coarse levels $j \in \mathcal{J}_-$ are not thresholded. The resulting coefficients are centered to yield $ES_j = 0$ and $ES_j(\lambda) = 0$ under $H_0$. The null hypothesis is rejected when the above test statistic is large.

For $j_\theta \le j_\Delta$, one cannot extract any information from the wavelet coefficients to distinguish between the two hypotheses, so testing is entirely based on the sum of squares of the scaling coefficients by performing the standard $\chi^2$-test.

The following theorem establishes the asymptotic optimality of the proposed test procedure:

**Theorem 2.2.** *Let the mother wavelet $\psi$ be of regularity $r > s$, and let the parameters $\theta = (s,p,q,M)$ of the Besov ball $B^s_{p,q}(\Omega_\Delta, M)$ be known, where $1 \le p,q \le \infty$, $sp > 1$, and $s > 1/2$ for $p \ge 2$. For local functional hypothesis testing* (5)–(6) *at a given significance level $\alpha \in (0,1)$ define the following test*:

$$\phi^*_{\varepsilon,\Delta} = \begin{cases} \mathbf{1}\left\{ \frac{T(j_\theta)+Q(j_\theta)}{\sqrt{V_0^2(j_\theta)+W_0^2(j_\theta)}} \ge Z_{1-\alpha} \right\} & \text{if} \quad j_\theta > j_\Delta, \\ \mathbf{1}\left\{ \sum_{I \in \mathcal{J}_\Delta - 1} Y_I^2/\varepsilon^2 \ge \chi^2_{C_\varphi, 1-\alpha} \right\} & \text{if} \quad j_\theta \le j_\Delta, \end{cases} \tag{17}$$

*where $j_\theta$ is given in* (10). *Then, for any $\beta \in (0,1)$, $\phi^*_{\varepsilon,\Delta}$ is a level $\alpha$ asymptotically rate-optimal test as $\varepsilon \to 0$, $\Delta \to 0$ monotonically and $\varepsilon\Delta^{-1/2} \to 0$.*

The construction of the test (17) helps in better understanding the classical rate when $\varepsilon\Delta^{-s'} \not\to 0$. Using the wavelet expansion (8), the function $f$ can be represented as a sum of its initial gross approximation $f_0$ generated by a finite linear combination of scaling functions and complementary details given by an infinite number of wavelet terms. When $\varepsilon\Delta^{-1/2} \not\to 0$, $j_\Delta > j_\varepsilon$ for sufficiently small $\varepsilon$ and $\Delta$, both of these components are covered by a strong noise and no accurate detection is available. For $\varepsilon\Delta^{-1/2} \to 0$ but $\varepsilon\Delta^{-s'} \not\to 0$ it becomes possible to detect $f_0$, while the details are still non-distinguishable from the noise. Detection of $f_0$ is, in a way, a *parametric* testing problem that under general regularity assumptions has a classical rate (e.g., Section 2.6.1 in Ingster & Suslina [12]). Finally, when $\varepsilon\Delta^{-s'} \to 0$, for sufficiently small $\varepsilon$ and $\Delta$, one has $j_\Delta < j_\theta < j_\varepsilon$ and it is now possible to detect the detailed wavelet terms. However, the problem in this case becomes nonparametric and, therefore, results in slower rates.

We finish this section by the following two remarks.

**Remark 2.1.** For $p \geq 2$ corresponding to "spatially" homogeneous signals whose wavelet coefficients are concentrated on coarse resolution levels, the above optimal test (17) can be simplified by truncating the wavelet series at level $j_\theta - 1$ (see also Abramovich *et al.* [2]). The resulting test $\phi^*_{\varepsilon,\Delta}$ becomes

$$\phi^*_{\varepsilon,\Delta} = \begin{cases} \mathbf{1}\left\{\frac{T(j_\theta)}{V_0(j_\theta)} > Z_{1-\alpha}\right\} & \text{if} \quad j_\theta > j_\Delta, \\ \mathbf{1}\left\{\sum_{I\in\mathcal{J}_{j_\Delta-1}} Y_I^2/\varepsilon^2 > \chi^2_{C_\varphi,1-\alpha}\right\} & \text{if} \quad j_\theta \leq j_\Delta. \end{cases}$$

**Remark 2.2.** Using the results of Fan [7] one can get asymptotic approximations for $b(\lambda_j)$ and $d(\lambda_j)$ :

$$E\left[\xi^{2k}\mathbf{1}(|\xi| > \lambda_j)\right] = \sqrt{2/\pi}\lambda_j^{2k-1}2^{-8(j-j_\theta+8)} + \mathcal{O}(\lambda_j^{2k-3}2^{-8(j-j_\theta+8)}), \quad k = 1, 2, \ldots$$

2.4. Adaptive local minimax test. The rate-optimal test derived in the previous section relies on the knowledge of the parameters of the Besov ball $\theta = (s, p, q, M)$. However, they are typically unknown in practice. In this section we consider the *adaptive* local testing problem where the above parameters are not specified *a priori* but are assumed to lie within a given range, and extend the corresponding results of Spokoiny [16] for adaptive global testing. We first construct the adaptive test and then show its asymptotic optimality.

Assume now that $\theta = (s, p, q, M)$ is unknown, but $1/2 < s \leq s_{\max}$, $1 \leq p \leq p_{\max}$, $1 \leq q < \infty$, $sp > 1$, and $0 < M_{\min} \leq M \leq M_{\max}$. Denote such a range of $\theta$ by $\mathcal{T}$. For each given set of parameters $\theta$ one may determine $j_\theta$ from (10). In fact, the range $\mathcal{T}$ determines essentially a range of admissible levels of the form $j_{\min} \leq j_\theta \leq j_{\max}$. One performs a series of tests of type (17) for each admissible level and rejects the null hypothesis if it is rejected at least for one of them.

More precisely, let $j_{\min} = \frac{4}{4s''_{\max}+1}(\log_2(M_{\min}/\varepsilon) + \frac{1}{2p'_{\max}}\log_2\Delta^{-1})$, $j_{\max} = j_\varepsilon - 1$, where $s''_{\max} = s_{\max} - 1/(2p'_{\max}) + 1/4$. Choose a mother wavelet of regularity $r > s_{\max}$. Since the number of admissible levels is $\mathcal{O}(\log(\varepsilon^{-2}\Delta))$, a Bonferroni type correction for multiple testing leads to the following asymptotic adaptive test, where we distinguish between two possible cases depending on whether $j_\Delta$ lies below or within the admissible range for $j_\theta$ (cf. (17)):

$$\phi^a_{\varepsilon,\Delta} = \begin{cases} \max_{j_{\min}\leq j_\theta\leq j_{\max}} \phi^a_{\varepsilon,\Delta,j_\theta}, & j_{\min} > j_\Delta, \\ \max_{j_\Delta\leq j_\theta\leq j_{\max}} \phi^a_{\varepsilon,\Delta,j_\theta}, & j_\Delta \in [j_{\min}, j_{\max}], \end{cases} \tag{18}$$

where $\phi^a_{\varepsilon,\Delta,j_\theta} = \mathbf{1}\left[\frac{T(j_\theta)+Q(j_\theta)}{\sqrt{V_0^2(j_\theta)+W_0^2(j_\theta)}} > \sqrt{2\log\log(\varepsilon^{-2}\Delta)}\right]$ if $j_\theta > j_\Delta$, and $\phi^a_{\varepsilon,\Delta,j_\Delta} = \mathbf{1}\left[\sum_{I\in\mathcal{J}_{j_\Delta-1}} Y_I^2/\varepsilon^2 > \chi^2_{C_\varphi,1-\alpha}\right]$.

Note that $j_{\min} > j_\Delta$ if $\varepsilon\Delta^{-s'_{\max}} \to 0$ and $j_{\max} > j_\Delta$ if $\varepsilon\Delta^{-1/2} \to 0$. The rate of testing is given in the following theorem:

**Theorem 2.3.** *Suppose $\varepsilon\Delta^{-1/2} \to 0$ as $\varepsilon \to 0$ and $\Delta \to 0$ monotonically. Then the rate $\gamma(\varepsilon, \Delta)$ of the adaptive test* (18) *for testing* (5)–(6) *is*

$$\gamma(\varepsilon,\Delta) = \begin{cases} \varepsilon^{8s''/(4s''+1)}\Delta^{-(2s+1)/(4s''+1)} \cdot (\log\log(\varepsilon^{-2}\Delta))^{2s'/(4s''+1)} & \\ & \text{if} \quad \varepsilon\Delta^{-s'} \to 0, \\ \varepsilon^2\Delta^{-1} & \text{otherwise.} \end{cases}$$

*Moreover, if $\varepsilon\Delta^{-s'_{\max}} \to 0$, then*

$$\alpha(\phi^a_{\varepsilon,\Delta}) = o_{\varepsilon\Delta^{-1/2}}(1),$$
$$\sup_{\theta\in\mathcal{T}} \beta(\phi^a_{\varepsilon,\Delta}, c\gamma(\varepsilon,\Delta)) = o_{\varepsilon\Delta^{-1/2}}(1) \qquad \text{for some} \quad c > 0.$$

Theorem 2.3 establishes that if $\varepsilon\Delta^{-s'} \to 0$, the adaptive test (18) is *nearly* rate-optimal (up to an additional $\log\log(\varepsilon^{-2}\Delta)$-factor). The results of Spokoiny [16] imply that there is no adaptive testing without loss of efficiency and such an extra log-log factor is an unavoidable (though inexpensive) price to pay for adaptivity in nonparametric testing. In addition, the above theorem demonstrates the degenerate behavior of the error probabilities for $\phi^a_{\varepsilon,\Delta}$ when $\varepsilon\Delta^{-s'_{\max}} \to 0$, which is also typical for adaptive global testing (see Ingster & Suslina [12]).

## 3. Concluding Remarks and Discussion

We considered the problem of testing the presence of a signal on an interval when both the noise level and the interval length tend to zero against a nonparametric alternative. We derived optimal (minimax) nonadaptive and adaptive tests extending the analogous results of Spokoiny [16] for fixed-length intervals.

Although in the paper we consider only one-dimensional signals, the extensions of the obtained results to two-dimensional signals (e.g., image analysis) and, in general, to $d$-dimensional signals are straightforward using $d$-dimensional wavelet transforms. See Horowitz & Spokoiny [8] for the corresponding $d$-dimensional global testing problem.

As in Abramovich *et al.* [2] and Abramovich & Angelini [1], the results of this paper can be directly applied in FANOVA models for testing *local* differences among groups of signals or their contrasts.

In practice one observes a *discrete* data sample of size $n$ with noise variance $\sigma^2$. Therefore, the sampled versions of the derived tests should be applied with empirical wavelet coefficients obtained by the *discrete* wavelet transform. Well-known results (see Brown & Low [5]) show the asymptotic equivalence (under some mild conditions) as $n \to \infty$, of the discrete model to the continuous "signal + white noise" model (1) with $\varepsilon = \sigma/\sqrt{n}$. The distinguishability condition $\varepsilon\Delta^{-1/2} \to 0$ for discrete data naturally corresponds to the usual requirement on the number of data points $n\Delta$ sampled on $\Omega_\Delta$ to tend to infinity.

## 4. Appendix

4.1. Proof of Theorem 2.2. Consider first the case $\varepsilon\Delta^{-s'} \to c^*$, where $0 \leq c^* < M$. In this case for sufficiently small $\varepsilon$ and $\Delta$ we can assume that $j_\Delta < j_\theta < j_\varepsilon$ and the proof is somewhat similar to that of Spokoiny [16] for global testing with necessary modifications. The statistics $T(j_\theta)$ and $Q(j_\theta)$ are the sums of $j_\theta - j_\Delta$ and $j_\varepsilon - j_\theta$ independent, squared integrable random variables respectively. Moreover, under the null hypothesis, they have zero means and variances $V_0^2(j_\theta)$ and $W_0^2(j_\theta)$. By the central limit theorem, the resulting standardized test statistic in (16) is then asymptotically normal $N(0,1)$ and the significance level of $\phi^*_{\varepsilon,\Delta}$ is asymptotically $\alpha$.

Consider now the Type II error of the test $\phi^*_{\varepsilon,\Delta}$. It is straightforward to see that for any specific $f \in \mathcal{F}(\gamma(\varepsilon,\Delta))$, one has asymptotically

$$\beta(\phi^*_{\varepsilon,\Delta}, f) = \Phi\left(R(\theta)^{\frac{1}{2}} Z_{1-\alpha} - \frac{E_f(T(j_\theta) + Q(j_\theta))}{\{\mathrm{Var}_f(T(j_\theta)) + Q(j_\theta))\}^{\frac{1}{2}}}\right) + o_{\varepsilon\Delta^{-1/2}}(1),$$

where $R(\theta) = (V_0^2(j_\theta) + W_0^2(j_\theta))/(\mathrm{Var}_f(T(j_\theta) + Q(j_\theta)))$. Since $R(\theta)$ is bounded from above by one, the asymptotic behavior of $\beta(\phi^*_{\varepsilon,\Delta}, f)$ depends only on the ratio $E_f(T(j_\theta) + Q(j_\theta))/(\mathrm{Var}_f(T(j_\theta) + Q(j_\theta))^{1/2}$.

The following lemmas provide the necessary bounds for $E_f(T(j_\theta) + Q(j_\theta))$ and $\mathrm{Var}_f(T(j_\theta)) + Q(j_\theta))$. Their technical proofs essentially repeat the analogous arguments of Spokoiny [16] (pp. 2491–2493) up to somewhat different notation and the initial resolution level being $j_\Delta$.

**Lemma 4.1.** *For any $f \in \mathcal{F}(\gamma(\varepsilon,\Delta))$,*

$$E_f\left(T(j_\theta) + Q(j_\theta)\right) \geq \frac{1}{2}\|f\|^2_{\Omega_\Delta} - M^2\varepsilon^{4s'} - \frac{1}{2}M^{p'}\varepsilon^{-p'+2}2^{-j_\theta s' p'}.$$

**Lemma 4.2.** *For any $f \in \mathcal{F}(\gamma(\varepsilon,\Delta))$, there exist positive constants $c_1$, $c_2$ and $c_3$ such that*

$$\mathrm{Var}_f\left(T(j_\theta) + Q(j_\theta)\right) \leq c_1\varepsilon^2\|f\|^2_{\Omega_\Delta} + c_2 \cdot 2^{j_\theta - j_\Delta}\varepsilon^4 + c_3\varepsilon^{4-p'}2^{-j_\theta s' p'}.$$

Recall that $\|f\|^2_{\Omega_\Delta}/\Delta \geq \gamma(\varepsilon,\Delta)$. Substituting $j_\theta$ from (10) and $\gamma(\varepsilon,\Delta) = \varepsilon^{8s''/(4s''+1)}\Delta^{-(2s+1)/(4s''+1)}$, Lemma 4.1 and Lemma 4.2 imply that there exists a constant $\tilde{c}_\beta$ such that

$$\inf_{f\in\mathcal{F}(\gamma(\varepsilon,\Delta))} \frac{E_f(T(j_\theta) + Q(j_\theta))}{\left\{\mathrm{Var}_f(T(j_\theta)) + Q(j_\theta))\right\}^{\frac{1}{2}}} > \tilde{c}_\beta,$$

where $\tilde{c}_\beta > 0$ satisfies $\Phi(Z_{1-\alpha} - \tilde{c}_\beta) = \beta$ (in fact, $\tilde{c}_\beta = Z_{1-\alpha} + Z_{1-\beta}$). This shows that the test $\phi^*_{\varepsilon,\Delta}$ achieves the asymptotic minimax rate.

Consider now the case $M \leq c^* \leq \infty$. Then, asymptotically $j_\theta \leq j_\Delta$ and the test (16) involves only the scaling coefficients. Under the null hypothesis

$\sum_{I\in\mathcal{J}_{j_\Delta-1}} Y_I^2/\varepsilon^2$ has a $\chi^2$ distribution with $C_\varphi$ degrees of freedom and the significance level of the test is clearly $\alpha$. The Type II error for any $f \in \mathcal{F}(\gamma(\varepsilon,\Delta))$ is

$$\beta(\phi^*_{\varepsilon,\Delta}, f) = P_f\bigg(\sum_{I\in\mathcal{J}_{j_\Delta-1}} Y_I^2/\varepsilon^2 < \chi^2_{C_\varphi,1-\alpha}\bigg).$$

Let $f_0$ be the projection of $f$ on the finite-dimensional linear span of $\varphi_{j_\Delta,k}$, $k = 0,\dots,C_\varphi - 1$. For any $f \in \mathcal{F}(\gamma(\varepsilon,\Delta))$ we have $\|f_0\|^2 = \sum_{\mathcal{I}\in\mathcal{J}_{j_\Delta-1}} w_I^2$ and

$$\beta(\phi^*_{\varepsilon,\Delta}, f) = X^2_{C_\varphi,\varepsilon^{-2}\|f_0\|^2}(\chi^2_{C_\varphi,1-\alpha}), \tag{19}$$

where $X^2_{\nu,d}(\cdot)$ is the probability function of the noncentral $\chi^2$ distribution with $\nu$ degrees of freedom and non-centrality parameter $d$. Since $f \in B^s_{p,q}(\Omega_\Delta, M)$ and $\varepsilon\Delta^{-s'} \to c^* \geq M$,

$$\sum_{j\geq j_\Delta}\sum_{\mathcal{I}\in\mathcal{J}_j} w_\mathcal{I}^2 \leq 2M^2\Delta^{2s'} \leq 2\varepsilon^2$$

(e.g., Meyer [15] and Spokoiny [16]). On the other hand, for any $c$ and any $f \in \mathcal{F}(c\varepsilon^2\Delta^{-1})$

$$\|f\|^2_{\Omega_\Delta} = \|f_0\|^2 + \sum_{j\geq j_\Delta}\sum_{\mathcal{I}\in\mathcal{J}_j} w_\mathcal{I}^2 \geq c\varepsilon^2.$$

Hence, for $c > 2$, $\|f_0\|^2 \geq (c-2)\varepsilon^2$ and (19) yields

$$\beta(\phi^*_{\varepsilon,\Delta}, f) \leq X^2_{C_\varphi,c-2}(\chi^2_{C_\varphi,1-\alpha}),$$

where one can always find a $c = c(\beta) > 2$ such that the above expression will be less than any fixed $\beta$.

Finally, note that although the proof showed that the test (16) achieves the rate $\varepsilon^{8s''/4s''+1}\Delta^{-(2s+1)/(4s''+1)}$ for any $0 \leq c^* < M$, this rate in fact coincides with the classical rate $\varepsilon^2\Delta^{-1}$ for non-zero $c^*$. □

4.2. Proof of Theorem 2.3. For convenience define the following two tests:

$$\phi^{(1)}_{\varepsilon,\Delta} = \mathbf{1}\left[\max_{\max\{j_\Delta,j_{\min}\}<j_\theta\leq j_{\max}} \frac{T(j_\theta)+Q(j_\theta)}{\sqrt{V_0^2(j_\theta)+W_0^2(j_\theta)}} > \sqrt{2\log\log(\varepsilon^{-2}\Delta)}\right],$$
$$\phi^{(2)}_{\varepsilon,\Delta} = \mathbf{1}\left[\sum_{I\in\mathcal{J}_{j_\Delta-1}} Y_I^2/\varepsilon^2 > \chi^2_{C_\varphi,1-\alpha}\right].$$

Obviously, $\phi^a_{\varepsilon,\Delta} = \begin{cases} \phi^{(1)}_{\varepsilon,\Delta} & \text{if} \quad j_{\min} > j_\Delta, \\ \max(\phi^{(1)}_{\varepsilon,\Delta}, \phi^{(2)}_{\varepsilon,\Delta}) & \text{otherwise.}\end{cases}$

Under the null hypothesis,

$$\left\{\frac{T(j_\theta)+Q(j_\theta)}{\sqrt{V_0^2(j_\theta)+W_0^2(j_\theta)}}, \quad j_\theta \in [\max\{j_\Delta, j_{\min}\}, j_\varepsilon - 1]\right\}$$

is a sequence of $\mathcal{O}(\log(\varepsilon^{-2}\Delta))$ weakly dependent, asymptotically $N(0,1)$ random variables. Applying the well-known extreme value results for Gaussian random variables (e.g., Leadbetter *et al.* [13]) one has

$$P_0(\phi^{(1)}_{\varepsilon,\Delta}=0)\to 0 \qquad \text{as} \quad \varepsilon\Delta^{-1/2}\to 0.$$

The significance level of the test $\phi^{(2)}_{\varepsilon,\Delta}$ is clearly $\alpha$ and therefore

$$\alpha(\phi^{a}_{\varepsilon,\Delta})\le P_0(\phi^{(1)}_{\varepsilon,\Delta}=1)+P_0(\phi^{(2)}_{\varepsilon,\Delta}=1)\le \alpha+o_{\varepsilon\Delta^{-1/2}}(1).$$

Choose now any set of parameters $\theta=(s,p,q,M)\in\mathcal{T}$. Consider first the case $\varepsilon\Delta^{-s'}\to 0$ and $\varepsilon\Delta^{-s'_{\max}}\to c^*$, $0\le c^*<M_{\min}$. Then, for sufficiently small $\varepsilon$ and $\Delta$, $j_{\min}>j_\Delta$. Define $j^*_\theta$ by

$$j^*_\theta=\frac{4}{4s''+1}\left(\log_2\left(\frac{M}{\varepsilon}\right)+\frac{1}{2p'}\log_2\frac{\Delta^{-1}}{2\log\log(\varepsilon^{-2}\Delta)}\right).$$

Given $\theta$, for any $f$ from the alternative

$$\begin{aligned}(20)\quad P_f(\phi^{(1)}_{\varepsilon,\Delta}=0)&\le P_f\left(\frac{T(j^*_\theta)+Q(j^*_\theta)}{(V_0^2(j^*_\theta)+W_0^2(j^*_\theta))^{\frac12}}\le\sqrt{2\log\log(\varepsilon^{-2}\Delta)}\right)\\ &\le\Phi\left(\sqrt{2\log\log(\varepsilon^{-2}\Delta)}-\frac{E_f(T(j^*_\theta)+Q(j^*_\theta))}{(\mathrm{Var}_f(T(j^*_\theta)+Q(j^*_\theta)))^{\frac12}}\right)+o_{\varepsilon\Delta^{-1/2}}(1).\end{aligned}$$

Substituting $j^*_\theta$ and $\gamma(\varepsilon,\Delta)=\varepsilon^{\frac{8s''}{4s''+1}}\Delta^{-\frac{2s+1}{4s''+1}}(2\log\log(\varepsilon^{-2}\Delta))^{\frac{2s'}{4s''+1}}$ in (20) and repeating the arguments in the proof of Theorem 2.2, a straightforward calculus implies that it is always possible to find a constant $c$ such that for any $f\in\mathcal{F}(c\gamma(\varepsilon,\Delta))$

$$(21)\qquad \frac{E_f(T(j^*_\theta)+Q(j^*_\theta))}{(\mathrm{Var}_f(T(j^*_\theta)+Q(j^*_\theta)))^{\frac12}}>\sqrt{2\log\log(\varepsilon^{-2}\Delta)},$$

and hence

$$\sup_{f\in\mathcal{F}(c\gamma(\varepsilon,\Delta))}P_f\big(\phi^{(1)}_{\varepsilon,\Delta}=0\big)=o_{\varepsilon\Delta^{-1/2}}(1).$$

Consider next the case $\varepsilon\Delta^{-s'}\to 0$ and $\varepsilon\Delta^{-s'_{\max}}\to c^*$, $M_{\min}\le c^*\le\infty$. Asymptotically, $j_\Delta\in[j_{\min},j_{\max}]$, but

$$P_f\big(\phi^{a}_{\varepsilon,\Delta}=0\big)\le P_f\big(\phi^{(1)}_{\varepsilon,\Delta}=0\big),$$

and the previous arguments remain valid for this case as well. Summarizing, for $\varepsilon\Delta^{-s'}\to 0$ the adaptive test $\phi^{a}_{\varepsilon,\Delta}$ always achieves the optimal rate $\gamma(\varepsilon,\Delta)$ up to an additional $\log\log(\varepsilon^{-2}\Delta)$ factor.

Finally, suppose $\varepsilon\Delta^{-s'}\not\to 0$. Asymptotically, $j_\Delta\in[j_{\min},j_{\max}]$ again, but now use the fact that

$$P_f\big(\phi^{a}_{\varepsilon,\Delta}=0\big)\le P_f\big(\phi^{(2)}_{\varepsilon,\Delta}=0\big).$$

The test $\phi^{(2)}_{\varepsilon,\Delta}$ does not depend on $\theta$ and involves only the scaling coefficients. Hence, for $\varepsilon\Delta^{-s'} \not\to 0$ the adaptive test obviously achieves the same classical rate $\varepsilon^2\Delta^{-1}$ as the nonadaptive test in Section 2.2 without any price for adaptivity. $\square$

**Acknowledgment.** We are delighted to thank Vadim Grinshtein and Yoav Benjamini for fruitful discussions and Vladimir Spokoiny for valuable remarks. Helpful and constructive comments of the reviewers which have substantially improved the paper are gratefully acknowledged.

## References

[1] F. Abramovich and C. Angelini, *Testing in mixed-effects FANOVA models*, J. Statist. Plann. Infer. (2006), to appear.

[2] F. Abramovich, A. Antoniadis, T. Sapatinas, and B. Vidakovic, *Optimal testing in a fixed-effects functional analysis of variance models*, Int. J. Wavelets, Multiresolution and Inform. Processing, **2** (2004), 323–349.

[3] L. Andersson, N. Hall, B. Jawerth, and G. Peters, *Wavelets on closed subsets of the real line.* In: *Topics in the Theory and Applications of Wavelets*, L. Schumaker and G. Webb (eds.), Academic Press, 1993.

[4] Y. Baraud, *Non-asymptotic minimax rates of testing in signal detection*, Bernoulli, **8** (2002), 577–606.

[5] L. D. Brown and M. G. Low, *Asymptotic equivalence of nonparametric regression and white noise*, Ann. Statist. Assoc., **93** (1996), 961–983.

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

[7] J. Fan, *Test of significance based on wavelet thresholding and Neyman's truncation*, J. Amer. Statist. Assoc., **91** (1996), 674–688.

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

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

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

[11] Yu. I. Ingster and I. A. Suslina, *Minimax nonparametric hypothesis testing for ellipsoids and Besov bodies*, ESAIM: Probab. Statist., **4** (2000), 53–135.

[12] Yu. I. Ingster and I. A. Suslina, *Nonparametric Goodness-of-Fit Testing under Gaussian Models*, Lect. Notes Statist., **169**, Springer, New York, 2003.

[13] M. R. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York, 1986.

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

[15] Y. Meyer, *Wavelets and Operators.* Cambridge University Press, Cambridge, 1992.

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

[17] V. G. Spokoiny, *Adaptive and spatially adaptive testing of nonparametric hypothesis*, Math. Methods Statist., **7** (1998), 245–273.

[*Received March* 2005; *revised August* 2005]