The Annals of Statistics
2001, Vol. 29, No. 1, 124–152

MULTISCALE TESTING OF QUALITATIVE HYPOTHESES[1]

BY LUTZ DÜMBGEN AND VLADIMIR G. SPOKOINY

Medizinische Universität zu Lübeck and Weierstrass Institute for Applied Analysis and Stochastics

Suppose that one observes a process Y on the unit interval, where $dY(t) = n^{1/2}f(t)\,dt + dW(t)$ with an unknown function parameter f, given scale parameter $n \geq 1$ ("sample size") and standard Brownian motion W. We propose two classes of tests of qualitative nonparametric hypotheses about f such as monotonicity or concavity. These tests are asymptotically optimal and adaptive in a certain sense. They are constructed via a new class of multiscale statistics and an extension of Lévy's modulus of continuity of Brownian motion.

1. Introduction. Many nonparametric statistical models involve some unknown function f on the real line. For instance, f might be the probability density of some distribution or a regression function. In many applications qualitative assumptions about f such as monotonicity, unimodality or concavity are plausible, though not necessarily satisfied. A natural question is how to test such assumptions. In the context of density estimation there exist various proposals for testing unimodality versus multimodality of f. Silverman (1981) developed a test based on critical bandwidths of kernel density estimators, whereas Hartigan and Hartigan (1985) and Müller and Sawitzki (1991) used the so-called dip or excess mass functional. Further results for these procedures are given by Mammen, Marron and Fisher (1992) and Cheng and Hall (1999). But the available distribution theory relies on additional smoothness constraints on f.

Koul and Schick (1997) considered the problem of comparing two nonparametric regression curves against a one-sided alternative, which, in the case of a common design, reduces to testing the hypothesis of positivity. The authors, however, discussed only asymptotic power against single directional alternatives, which reduces the problem to the classical setup.

The aim of this paper is to propose a test for a qualitative hypothesis against a general smooth alternative with unknown degree of smoothness. There is another aspect of testing qualitative assumptions which we are interested in: if there is evidence that such an assumption is violated one would often like to identify, with a certain confidence, regions where this violation occurs.

Received January 1998; revised August 2000.

[1]Supported by the Deutsche Forschungsgemeinschaft, SFB 373, "Simulation and Quantification of Economic Processes," at Humboldt University.

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

Key words and phrases. Adaptivity, concavity, Lévy's modulus of continuity, monotonicity, multiple test, nonparametric, positivity.

In the present paper we study such problems in detail within the (continuous) white noise model. Suppose that one observes a stochastic process Y on the unit interval $I := [0, 1]$, where

$$Y(t) = n^{1/2} \int_{[0,t]} f(x)\,dx + W(t).$$

Here f is an unknown function in $L^1(I)$, $n \geq 1$ is a given scale parameter and W is standard Brownian motion. We consider the following hypotheses:

$$\mathscr{H}_{\leq 0} := \{f : f \leq 0\},$$
$$\mathscr{H}_{\downarrow} := \{f : f \text{ is nonincreasing}\},$$
$$\mathscr{H}_{\text{conc}} := \{f : f \text{ is concave}\}.$$

Note that these are nonparametric rather than finite or finite-dimensional hypotheses. The ideal white noise model serves as a prototype for various statistical models involving regression functions or distribution densities. The results of Brown and Low (1996), Nussbaum (1996) and Grama and Nussbaum (1998) on the asymptotic equivalence of these models can be used to transfer the lower bounds of the present paper to other models. The main benefit of the white noise model is the applicability of rescaling arguments as, for instance, in Donoho and Low (1992).

There is an extensive literature on nonparametric testing of the simple hypothesis $\{0\}$. As a starting point we recommend the survey of Ingster (1993), which contains many basic results and additional references. Under the nonparametric approach it is typically assumed that f belongs to a certain class $\mathscr{F}$ of smooth functions, and its distance to the null hypothesis $\{0\}$ is quantified by some seminorm $\|f\|$. For a given level $\alpha \in]0, 1[$ and some number $\delta > 0$ the goal is to find a statistical test $\phi : \mathscr{C}(I) \to I$ whose minimal power

$$\inf_{g \in \mathscr{F} : \|g\| \geq \delta} \mathbb{E}_g \phi(Y)$$

is as large as possible under the constraint that $\mathbb{E}_0 \phi(Y) \leq \alpha$. Here and subsequently the dependency of probabilities and expected values on the functional parameter f is indicated by a subscript. Approximate solutions, as $n \to \infty$, for this testing problem are known for various classes $\mathscr{F}$ and seminorms $\|\cdot\|$. Ingster (1986, 1993) described the case of L_p-norm, $1 \leq p \leq \infty$, and Hölder and Sobolev smoothness classes. Spokoiny (1998) extended the results to the case of arbitrary Besov classes. Sharp optimal asymptotic results are known for a few cases: Ermakov (1990) found the sharp asymptotics for Sobolev balls and L_2-distance, while Lepski and Tsybakov (2000) also treated Hölder smoothness classes and the supremum norm. The latter case is of special importance for us since the sup-norm seems to be most suitable in order to describe the alternative set for our qualitative null hypotheses; see Section 3.2 for a discussion in terms of test signals. The tests of Lepski and Tsybakov (2000) are based on a kernel estimator of f with a kernel function and bandwidth depending on $\mathscr{F}$. It is a general problem that the available optimal tests

ϕ depend explicitly on the class $\mathscr{F}$ and may fail if the latter is altered. With this problem in mind we review some results of Lepski and Tsybakov (2000) in Section 2 and introduce a new class of multiscale statistics combining kernel estimators of various bandwidths. These statistics lead to adaptive tests in the sense that they are asymptotically optimal for many Hölder classes simultaneously.

The problem of adaptive (data-driven) testing a simple or parametric hypothesis, where deviation from the null hypothesis is measured by some integral norm, was considered in Eubank and Hart (1992), Ledwina (1994), Ledwina and Kallenberg (1995), Fan (1996), Spokoiny (1996) and Hart (1997), among others. The underlying idea is to consider simultaneously a family of test statistics corresponding to different values of the smoothing parameters which leads to multiple testing. Eubank and Hart (1992), Ledwina (1994), Ledwina and Kallenberg (1995) and Hart (1997) discussed the so-called "order selection test" or "data-driven Neyman test": the deviation of the underlying function f from the null hypothesis is estimated via an orthogonal series expansion, and a modified Mallows' criterion is used for selecting the number of random coefficients to be included. This method allows one to combine testing and model selection within one approach and provides root-n consistency against any directional alternative, but it has no power against a general smooth alternative. The tests in Fan (1996) and Spokoiny (1996) are based on the maximum of centered and standardized statistics and are rate-optimal adaptive against a smooth alternative. The main message of Spokoiny (1996) is that the adaptive approach leads necessarily to suboptimal rates by a factor $\log(\log(n))$. By way of contrast, the present paper shows that adaptive testing with respect to the sup-norm is possible without essential loss of efficiency. The reason is that a sup-norm test is based on the maximum of nonparametric estimates $\hat{f}_h(x)$ of the model function $f(x)$ for different locations x. Even with a fixed value of the smoothing parameter h this requires an adjustment for multiple testing, while an additional adjustment for using different h turns out to be unncessary.

In Section 3 we introduce tests for the three nonparametric hypotheses $\mathscr{H}_{\le 0}$, $\mathscr{H}_{\downarrow}$ and $\mathscr{H}_{\mathrm{conc}}$. Given any of these composite hypotheses, say $\mathscr{H}_o$, we introduce two different functionals $\Delta(f)$ measuring the distance of f to $\mathscr{H}_o$ and show how to maximize approximately

$$\inf_{g\in\mathscr{F}:\Delta(g)\ge\delta} \mathbb{E}_g\phi(Y)$$

over all tests ϕ satisfying

$$\sup_{f\in\mathscr{H}_o} \mathbb{E}_f\phi(Y) \le \alpha.$$

Again the proposed tests are based on the multiscale idea as introduced in Section 2 and are adaptive in a certain sense. Moreover, whenever the hypothesis $\mathscr{H}_o$ is rejected we can identify with confidence $1-\alpha$ one or several intervals $J \subset [0, 1]$ on which the qualitative assumption about f is violated. Thus our procedures may be interpreted as multiple tests and lead automatically to nonparametric confidence sets.

Section 4 describes some possible extensions and modifications for other, more traditional statistical models. Some numerical examples for regression with Gaussian errors are presented in Section 5. All proofs are deferred to Section 6. There we present an extension of Levý's modulus of continuity which is of independent interest.

Two recent papers offering somewhat similar procedures are Chaudhuri and Marron (1999) and Hall and Heckman (2000). The former paper treats statistical inference about the modality of an unknown regression curve, using Gaussian kernel estimators; see also the discussion in Section 3.2. The latter paper is about testing monotonicity based on local linear smoothing with a variable "locality" parameter. In both cases the theoretical results are quite different from the ones presented here, and the issue of optimal testing is not discussed.

2. Multiscale tests of the hypothesis "$f = 0$." Let us first introduce some notation. For measurable functions f, g on the real line let $\langle f, g\rangle := \int f(x)g(x)\,dx$ and $\|f\|_2 := \langle f, f\rangle^{1/2}$. When the integrals are restricted to some interval $J \subset \mathbf{R}$ we use an additional subscript J and write $\langle f, g\rangle_J$, $\|f\|_{2,J}$. Moreover, let $\|f\|_J$ denote the supremum norm $\sup_{x\in J}|f(x)|$.

Suppose that we want to test the null hypothesis $\{0\}$ versus a simple alternative $\{g\}$ with $g \in L^2(I)$. Since $\log(d\mathbb{P}_g/d\mathbb{P}_0)(Y) = n^{1/2}\int_I g\,dY - n\|g\|^2/2$, the Neyman–Pearson test rejects the null hypothesis at level α if the linear test statistic

$$\|g\|_{2,I}^{-1}\int_I g(x)\,dY(x)$$

exceeds the $(1-\alpha)$ quantile of the standard Gaussian distribution. For $\int_I g\,dY$ is normally distributed with mean $n^{1/2}\langle f, g\rangle_I$ and variance $\|g\|_{2,I}^2$. Therefore the power of this test is an increasing function of $n^{1/2}\|g\|_{2,I}$.

In the case of a closed and convex alternative $\mathscr{G} \subset L^2(I)\backslash\{0\}$ let g_o be the unique point in $\mathscr{G}$ minimizing $\|g_o\|_{2,I}$. It is well known from convex analysis that g_o is uniquely determined by

$$\langle g, g_o\rangle_{2,I} \ge \|g_o\|_{2,I}^2 \text{ for all } g \in \mathscr{G}.$$

Therefore a Neyman–Pearson test of $\{0\}$ versus $\{g_o\}$ is automatically an optimal test of $\{0\}$ versus $\mathscr{G}$. Its minimal power over $\mathscr{G}$ is attained at the least favorable parameter g_o.

For $\beta, L > 0$ and an interval $J \subset \mathbf{R}$ the Hölder smoothness class $\mathscr{F}_J(\beta, L)$ is defined as follows. In the case $0 < \beta \le 1$ let

$$\mathscr{F}_J(\beta, L) := \{f : |f(x) - f(y)| \le L|x-y|^\beta \text{ for all } x, y \in J\}.$$

For $k < \beta \le k+1$ with an integer $k > 0$ let $\mathscr{F}_J(\beta, L)$ be the set of functions that are k times differentiable on J and whose kth derivative belongs to $\mathscr{F}_J(\beta - k, L)$. We also write $\mathscr{F}(\beta, L)$ instead of $\mathscr{F}_{\mathbf{R}}(\beta, L)$.

Suppose that we want to test $\{0\}$ versus

$$\{g \in \mathcal{F}(\beta, L) : \|g\|_J \geq \delta\}$$

for some $\delta > 0$ and some interval $J \subset I$. This alternative is not convex but is the union of the closed convex sets

$$\{g \in \mathcal{F}(\beta, L) : g(t) \geq \delta\} \text{ and } \{g \in \mathcal{F}(\beta, L) : -g(t) \geq \delta\}$$

over all $t \in J$. Thus we look first for the least favorable points within these sets.

Let $\psi = \psi(\cdot, \beta)$ be the unique solution of the following optimization problem:

$$\text{Minimize } \|\psi\|_2 \text{ over all } \psi \in \mathcal{F}(\beta, 1) \text{ with } \psi(0) \geq 1. \tag{2.1}$$

It is known that ψ is an even function with compact support, say, $[-R, R]$, and $\psi(0) = 1 > |\psi(x)|$ for $x \neq 0$. For instance, in the case $0 < \beta \leq 1$ one can easily show that

$$\psi(x) = 1\{|x| \leq 1\}(1 - |x|^\beta).$$

For the case $\beta > 1$ an explicit solution is known only for $\beta = 2$; see, for example, Leonov (1999). Donoho (1994a) and Leonov (1999) contain some useful properties of ψ and advice on how this function can be constructed numerically. For any scale parameter $h > 0$ and any location parameter $t \in \mathbf{R}$ let

$$\psi_{t,h}(x) := \psi\left(\frac{x-t}{h}\right). \tag{2.2}$$

A simple rescaling argument shows that for $\delta > 0$ the function $\tilde{\psi} := \pm\delta\psi_{t,h}$ belongs to $\mathcal{F}(\beta, \delta h^{-\beta})$ and minimizes $\|\tilde{\psi}\|_2$ under the additional constraint $\pm\tilde{\psi}(t) \geq \delta$. In the case $Rh \leq t \leq 1 - Rh$ this function is supported by I and thus minimizes $\|\tilde{\psi}\|_{2,I}$ as well. Then with

$$\widehat{\Psi}(t, h) := h^{-1/2}\|\psi\|_2^{-1} \int_I \psi_{t,h}(x)\, dY(x) \tag{2.3}$$

the test statistic $\pm\widehat{\Psi}(t, h)$ is optimal for testing $\{0\}$ versus $\{g \in \mathcal{F}(\beta, \delta h^{-\beta}) : \pm g(t) \geq \delta\}$. Note that

$$\operatorname{Var}(\widehat{\Psi}(t, h)) = 1 \quad \text{and} \quad \mathbb{E}\widehat{\Psi}(t, h) = (n/h)^{1/2}\|\psi\|_2^{-1}\langle f, \psi_{t,h}\rangle.$$

The following theorem implies that all these test statistics $\widehat{\Psi}(t, h)$ can be combined in a specific way.

THEOREM 2.1. *Let ψ be any function in $L^2(\mathbf{R})$ with bounded total variation and compact support $[-R, R]$. For real numbers $h > 0$ and $t \in [Rh, 1 - Rh]$ let $\psi_{t,h}$ and $\widehat{\Psi}(t, h)$ be defined as in* (2.2) *and* (2.3). *Then, almost surely,*

$$\sup_{h \in]0, R^{-1}/2]} \sup_{t \in [Rh, 1-Rh]} (|\widehat{\Psi}(t, h) - \mathbb{E}\widehat{\Psi}(t, h)| - C(2Rh))/D(2Rh) < \infty,$$

where $C(r) := (2\log(1/r))^{1/2}$ and $D(r) := (\log(e/r))^{-1/2} \log\log(e^e/r)$.

REMARKS. The rationale behind the additive correction term $C(2Rh)$ is that the random variables $\widehat{\Psi}((2j-1)Rh, h) - \mathbb{E}\widehat{\Psi}((2j-1)Rh, h)$, $j = 1, 2, \ldots, \lfloor (2Rh)^{-1} \rfloor$, are independent with standard normal distribution. The maximum of these variables is known to be $C(2Rh) + o_p(1)$ as $h \to 0$. Note further that $D(\cdot)$ is bounded and strictly positive on $]0, 1]$ with $\lim_{r \to 0} D(r) = 0$.

MULTISCALE TEST. For any function ψ as in Theorem 2.1 we define the global test statistic

$$T(Y) = T(Y, \psi) := \sup_{h \in]0, R^{-1}/2]} \sup_{t \in [Rh, 1-Rh]} (|\widehat{\Psi}(t, h)| - C(2Rh)). \tag{2.4}$$

In the case $f = 0$ this test statistic equals $T(W)$ and is finite, by Theorem 2.1. Therefore the critical value

$$\kappa_\alpha = \kappa_\alpha(\psi) := \min\{r \in \mathbf{R} : \mathbb{P}\{T(W) \le r\} \ge 1 - \alpha\} \tag{2.5}$$

is well defined for any $\alpha \in]0, 1[$. Then $1\{T(\cdot) > \kappa_\alpha\}$ defines a test of $\{0\}$ at level α which is asymptotically optimal in the following sense.

THEOREM 2.2. *Let the test statistic $T(Y)$ be defined as in* (2.4) *with the solution $\psi = \psi(\cdot, \beta)$ of* (2.1). *We define*

$$\rho_n = \rho_n(\beta) := \left(\frac{\log n}{n}\right)^{\beta/(2\beta+1)}$$

and

$$c_* = c_*(\beta, L) := \left(\frac{2L^{1/\beta}}{(2\beta+1)\|\psi\|_2^2}\right)^{\beta/(2\beta+1)}.$$

Then for arbitrary numbers $\varepsilon_n > 0$ with $\lim_{n\to\infty} \varepsilon_n = 0$ and $\lim_{n\to\infty} (\log n)^{1/2} \times \varepsilon_n = \infty$ the following two conclusions hold:

(a) *For any fixed nondegenerate interval $J \subset I$ and arbitrary tests ϕ_n with $\mathbb{E}_0 \phi_n(Y) \le \alpha$,*

$$\limsup_{n\to\infty} \inf_{g \in \mathcal{F}(\beta, L): \|g\|_J \ge (1-\varepsilon_n) c_* \rho_n} \mathbb{E}_g \phi_n(Y) \le \alpha.$$

(b) *Let $J = J_n(\beta, L) := [R(c_* \rho_n / L)^{1/\beta}, 1 - R(c_* \rho_n / L)^{1/\beta}]$. Then*

$$\lim_{n\to\infty} \inf_{g \in \mathcal{F}(\beta, L): \|g\|_J \ge (1+\varepsilon_n) c_* \rho_n} \mathbb{P}_g\{T(Y) \ge \kappa_\alpha\} = 1.$$

The result of Theorem 2.2 may be read as follows. If the underlying function f deviates from the null hypothesis by at least $(1 + \varepsilon_n) c_* \rho_n$, then the test rejects the null with probability close to 1. This deviation bound cannot be significantly improved in the sense that, for every test ϕ_n of $\{0\}$ at level α, there exists an alternative function g with deviation $(1 - \varepsilon_n) c_* \rho_n$ which will not be detected with probability $1 - \alpha - o(1)$ or larger.

ADAPTIVITY. Part (a) of Theorem 2.2 is a modification of Lepski and Tsybakov's (2000) lower bound. Part (b) is novel in that one test, $1\{T(\cdot) > \kappa_\alpha\}$, is asymptotically optimal for all Hölder smoothness classes $\mathscr{F}(\beta, L)$, $L > 0$. In other words, it is adaptive with respect to the second parameter of $\mathscr{F}(\beta, L)$.

Adaptivity with respect to both parameters, β and L, is still an open problem. However, suppose that we use the test statistic T corresponding to the, say, triangular kernel $\psi(\cdot, 1)$. Then it follows from Ingster (1986) that for arbitrary $\beta > 0$ there is a constant $c(\beta, L) \geq c_*(\beta, L)$ such that

$$\lim_{n\to\infty} \inf_{g\in\mathscr{F}(\beta, L):\, \|g\|_I \geq c(\beta, L)\rho_n} \mathbb{P}_g\{T(Y) \geq \kappa_\alpha\} = 1.$$

Thus our test with the triangular kernel ψ is *rate optimal* over arbitrary Hölder classes with respect to the supremum norm over the whole unit interval I.

KERNEL ESTIMATORS OF f. If ψ is viewed as a kernel function it leads to the kernel estimator

$$\hat{f}_{n,h}(t) := \frac{\int_I \psi_{t,h}\, dY}{n^{1/2}\langle 1, \psi_{t,h}\rangle} = c_{n,h}\widehat{\Psi}(t, h)$$

of $f(t)$, where $c_{n,h} := (nh)^{-1/2}\|\psi\|_2/\langle 1, \psi\rangle$ is the standard deviation of $\hat{f}_h(t)$. Then our test statistic $T(Y)$ may be written as

$$T(Y) = \sup_{h\in]0, R^{-1}/2]} \left(c_{n,h}^{-1}\|\hat{f}_{n,h}\|_{[Rh, 1-Rh]} - C(2Rh)\right).$$

Thus we combine kernel estimators with arbitrary bandwidths in a specific way.

BOUNDARY EFFECTS. For the sake of simplicity we restricted our attention to the supremum norm on compact subintervals of $]0, 1[$ instead of the whole interval I. This restriction can be avoided by using suitable boundary kernels similar to those used by Lepski and Tsybakov (2000).

3. Testing the qualitative assumptions.

We propose two classes of tests corresponding to different notions of distance from the composite null hypothesis $\mathscr{H}_{\leq 0}$, $\mathscr{H}_{\downarrow}$ or $\mathscr{H}_{\text{conc}}$.

3.1. *Lipschitz alternatives and sup-norm distance.* In this section let $\mathscr{H}_o$ be either $\mathscr{H}_{\leq 0}$ or $\mathscr{H}_{\downarrow}$. We assume that under the alternative f belongs to the class $\mathscr{F}(1, L)$ for some unknown parameter $L > 0$ and measure its distance to $\mathscr{H}_o$ by

$$\Delta_J(f) := \inf_{f_o\in\mathscr{H}_o} \|f - f_o\|_J$$

for some interval $J \subset I$. Elementary calculus shows that, in the case $f \notin \mathscr{H}_o$,

$$\Delta_J(f) = \begin{cases} \sup_{t \in J} f(t), & \text{if } \mathscr{H}_o = \mathscr{H}_{\leq 0}, \\ \sup_{s, t \in J : s < t} \dfrac{f(t) - f(s)}{2}, & \text{if } \mathscr{H}_o = \mathscr{H}_{\downarrow}. \end{cases}$$

A natural test statistic might be $\Delta_J(\hat{f})$, where $\hat{f}$ is some estimator of f. Specifically let

$$\hat{f}_{n,h}(t) := n^{-1/2} h^{-1} \int \psi_{t,h} \, dY = (3nh/2)^{-1/2} \widehat{\Psi}(t, h)$$

for some $h \in]0, 1/2]$, where ψ is the triangular kernel given by $\psi(x) := 1\{|x| \leq 1\}(1 - |x|)$ with $\|\psi\|_2^2 = 2/3$. If we had one specific Lipschitz class $\mathscr{F}(1, L)$ in mind, it would indeed be sufficient to use the test statistic $\Delta_J(\hat{f}_{n,h})$ with a suitable bandwidth $h = h_n(L)$. But in order to achieve adaptivity with respect to L we combine all bandwidths and use the test statistic

$$\begin{aligned} T_o(Y) &:= \sup_{h \in]0, 1/2]} \left(\Delta_{[h, 1-h]}(\widehat{\Psi}(\cdot, h)) - C(2h)\right) \\ &= \sup_{h \in]0, 1/2]} \left((3nh/2)^{1/2} \Delta_{[h, 1-h]}(\hat{f}_{n,h}) - C(2h)\right). \end{aligned}$$

One can show that

$$T_o(Y) \leq T_o(W) \quad \text{if } f \in \mathscr{H}_o, \tag{3.6}$$

with equality if $f = 0$. Moreover, $T_o(W)$ is finite, according to Theorem 2.1. Thus the critical value

$$\kappa_{o,\alpha} := \min\left\{r \in \mathbf{R} : \mathbb{P}\{T_o(W) \leq r\} \geq 1 - \alpha\right\}$$

is well defined, and we reject the null hypothesis $\mathscr{H}_o$ at level α if $T_o(Y) > \kappa_{o,\alpha}$. This test is asymptotically optimal for any Lipschitz class $\mathscr{F}(1, L)$, $L > 0$.

THEOREM 3.1. *Let $(\varepsilon_n)_{n \geq 1}$ be as described in Theorem 2.2 and*

$$\rho_n := \left(\frac{\log n}{n}\right)^{1/3}.$$

(a) *For any fixed nondegenerate interval $J \subset I$ and arbitrary tests ϕ_n with $\mathbb{E}_0 \phi_n(Y) \leq \alpha$,*

$$\limsup_{n \to \infty} \inf_{g \in \mathscr{F}(1, L) : \Delta_J(g) \geq (1 - \varepsilon_n) L^{1/3} \rho_n} \mathbb{E}_g \phi_n(Y) \leq \alpha.$$

(b) *Let $J = J_n$ be any nonvoid subinterval of* $[L^{-2/3}\rho_n, 1 - L^{-2/3}\rho_n]$, *and let $J' = J'_n$ be its neighborhood* $[\min(J) - L^{-2/3}\rho_n, \max(J) + L^{-2/3}\rho_n]$. *Then*

$$\lim_{n \to \infty} \inf_{g \in \mathscr{F}_{J'}(1, L) : \Delta_J(g) \geq (1 + \varepsilon_n) L^{1/3} \rho_n} \mathbb{P}_g\{T_o(Y) \geq \kappa_{o,\alpha}\} = 1.$$

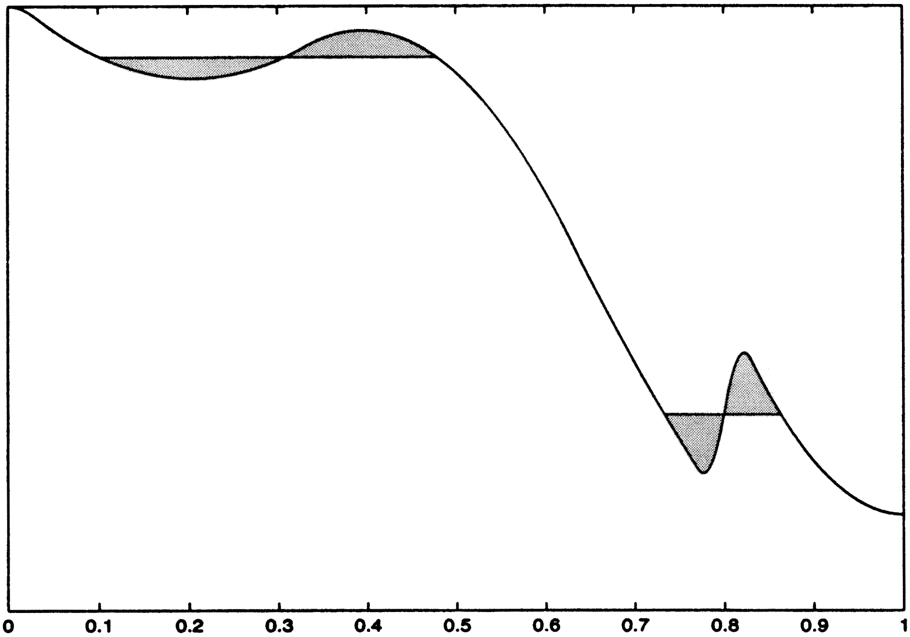

FIG. 1. *A function $g \notin \mathscr{H}_\downarrow$ and its projection f_o onto $\mathscr{H}_\downarrow$.*

SPATIAL ADAPTIVITY. Statement (b) of Theorem 3.1 shows that our test is *spatially adaptive*: if $\Delta_J(f) \geq (1 + \varepsilon_n)L^{1/3}\rho_n$ on some interval J and f is locally Lipschitz with constant L on a slightly larger interval J', then the null hypothesis will be rejected with probability close to 1.

3.2. *Test signals and derivatives.* In this section we consider the null hypotheses $\mathscr{H}_\downarrow$ and $\mathscr{H}_{\text{conc}}$ and describe a second class of tests in terms of *test signals*. Let us first illustrate this approach for the hypothesis $\mathscr{H}_\downarrow$: Figure 1 shows a smooth function $g \notin \mathscr{H}_\downarrow$ together with the unique function $f_o \in \mathscr{H}_\downarrow$ minimizing $\|g - f_o\|_{2,I}$. The shaded region shows the difference $g - f_o$. This difference is similar to the sum of two functions with disjoint support but *similar shape*. More precisely, for a suitable odd function ψ with compact support $[-R, R]$, for example, $\psi(x) = 1\{|x| \leq 1\}x(1 - |x|)$, the difference $g - f_o$ is similar to $a\psi_{t,h} + a'\psi_{t',h'}$, where $0 < a < a'$, $h > h' > 0$ and $t + Rh < t' - Rh'$. Therefore a suitably weighted maximum of all statistics $\widehat{\Psi}(t, h)$ with $0 < h \leq R^{-1}/2$ and $Rh \leq t \leq 1 - Rh$ should be an appropriate test statistic for the null hupothesis $\mathscr{H}_\downarrow$.

Generally let $\mathscr{H}_o = \mathscr{H}_\downarrow$ or $\mathscr{H}_o = \mathscr{H}_{\text{conc}}$, and let ψ be a test signal in $L^2(\mathbf{R})$ with compact support $[-R, R]$ and bounded total variation such that

$$\langle f, \psi \rangle \leq 0 \quad \text{for all } f \in \mathscr{H}_o. \tag{3.7}$$

Lemma 6.1 provides sufficient conditions for this requirement. Then we propose the test statistic

$$(3.8)\qquad \widetilde{T}(Y) = \widetilde{T}(Y, \psi) := \sup_{h\in]0,R^{-1}/2]} \sup_{t\in[Rh,1-Rh]} (\widehat{\Psi}(t, h) - C(2Rh)),$$

which is just a one-sided version of (2.4). Requirement (3.7) on ψ implies that

$$\widetilde{T}(Y) \begin{cases} \le \widetilde{T}(W), & \text{if } f \in \mathscr{H}_o, \\ = \widetilde{T}(W), & \text{if } f \text{ is constant and } \mathscr{H}_o = \mathscr{H}_\downarrow, \\ = \widetilde{T}(W), & \text{if } f \text{ is linear and } \mathscr{H}_o = \mathscr{H}_{\text{conc}}. \end{cases}$$

Thus with the $(1-\alpha)$ quantile $\tilde{\kappa}_\alpha = \tilde{\kappa}_\alpha(\psi)$ of $\widetilde{T}(W)$,

$$\max_{f\in\mathscr{H}_o} \mathbb{P}_f\{\widetilde{T}(Y) > \tilde{\kappa}_\alpha\} = \mathbb{P}\{\widetilde{T}(W) > \tilde{\kappa}_\alpha\} \le \alpha.$$

MULTIPLE TESTS. Our method can be viewed as a multiple test procedure. Let $\widetilde{\mathscr{D}}_\alpha$ be the random family of all intervals $[t - Rh, t + Rh]$ with $h > 0$ and $Rh \le t \le 1-Rh$ such that $\widehat{\Psi}(t, h) > C(2Rh)+\tilde{\kappa}_\alpha$. Then $\widetilde{T}(Y) > \tilde{\kappa}_\alpha$ if, and only if, $\widetilde{\mathscr{D}}_\alpha$ is nonempty. One may claim with confidence $1 - \alpha$ that the unknown regression function f violates the qualitative assumption, that is, being non-increasing and concave, on every interval $J \in \widetilde{\mathscr{D}}_\alpha$. Consequently, whenever the null hypothesis $\mathscr{H}_o$ is rejected, we have some information about *where* this violation occurs. Analogous considerations apply to the other multiscale tests of this paper.

OPTIMAL TEST SIGNALS. In order to identify a "good" test signal ψ satisfying (3.7), note that a smooth function g is non-increasing if and only if $g^{(1)} \le 0$ while concavity of g is equivalent to $g^{(2)} \le 0$. Here $g^{(k)}$ denotes the kth derivative of g. Now we want to find an optimal test signal ψ for testing $\mathscr{H}_o$ versus all alternatives of the form $\{g \in \mathscr{F}(k+1, L) : \widetilde{\Delta}_J(g) \ge \delta\}$, where

$$\widetilde{\Delta}_J(g) := \sup_{t\in J} g^{(k)}(t) \quad \text{and} \quad k := \begin{cases} 1, & \text{if } \mathscr{H}_o = \mathscr{H}_\downarrow, \\ 2, & \text{if } \mathscr{H}_o = \mathscr{H}_{\text{conc}}. \end{cases}$$

This leads to the following optimization problem:

$$(3.9)\quad \text{Minimize } \|g - f\|_2 \text{ over all } (g, f) \in \mathscr{F}(k+1, 1) \times \mathscr{H}_o \text{ with } g^{(k)}(0) \ge 1.$$

Note that the set $\{g \in \mathscr{F}(k+1, 1) : g^{(k)}(0) \ge 1\}$ is convex, while $\mathscr{H}_o$ is even a convex cone. Thus a pair (g_o, f_o) solves problem (3.9) if, and only if, the difference

$$\psi := g_o - f_o$$

satisfies

$$(3.10)\qquad \langle f, \psi\rangle \le \langle f_o, \psi\rangle = 0 \qquad \text{for all } f \in \mathscr{H}_o$$

and

$$\langle g, \psi \rangle \geq \|\psi\|_2^2 \text{ for all } g \in \mathscr{F}(k+1, 1) \text{ with } g^{(k)}(0) \geq 1. \tag{3.11}$$

These inequalities imply that $\|(g-f)-\psi\|_2^2 \leq \|g-f\|_2^2 - \|\psi\|_2^2$ for any pair (g, f) as in (3.9). Therefore the difference ψ is unique and satisfies (3.7).

LEMMA 3.1. (a) *In the case* $\mathscr{H}_o = \mathscr{H}_\downarrow$ *a solution* (g_o, f_o) *of problem* (3.9) *is given by*

$$g_o(x) := x(1 - |x|/2) \quad \textit{and} \quad f_o(x) := 1\{|x| \geq 2\} g_o(x).$$

For the corresponding test signal $\psi_\downarrow := g_o - f_o$,

$$\|\psi_\downarrow\|_2^2 = 8/15 = 0.53\bar{3}.$$

(b) *In the case* $\mathscr{H}_o = \mathscr{H}_{\mathrm{conc}}$ *a solution* (g_o, f_o) *of problem* (3.9) *is given by*

$$g_o(x) := -32/81 + x^2/2 - |x|^3/6 \quad \textit{and} \quad f_o(x) := 1\{|x| \geq 8/3\} g_o(x).$$

For the corresponding test signal $\psi_{\mathrm{conc}} := g_o - f_o$,

$$\|\psi_{\mathrm{conc}}\|_2^2 = 2^{16}/(3^8 \cdot 5 \cdot 7) \approx 0.2854.$$

The optimal test signals $\psi_\downarrow$ and ψ_{conc} are depicted in Figure 2.

THEOREM 3.2. *Let* $\widetilde{T}$ *be defined with* $\psi = g_o - f_o$, *where* (g_o, f_o) *solves the optimization problem* (3.9). *For* $L > 0$ *let*

$$\rho_n := \left(\frac{\log n}{n}\right)^{1/(2k+3)} \quad \textit{and} \quad c_* = c_*(L) := \left(\frac{2L^{2k+1}}{(2k+3)\|\psi\|_2^2}\right)^{1/(2k+3)}.$$

Let $J = J_n$ *be any nonvoid subinterval of* $[RL^{-1}c_*\rho_n, 1 - RL^{-1}c_*\rho_n]$, *and let* $J' = J_n'$ *be its neighborhood* $[\min(J) - RL^{-1}c_*\rho_n, \max(J) + RL^{-1}c_*\rho_n]$. *Then*

$$\lim_{n\to\infty} \inf_{g \in \mathscr{F}_{J'}(k+1, L) : \widetilde{\Delta}_J(g) \geq (1+\varepsilon_n) c_* \rho_n} \mathbb{P}_g\{T(Y) > \tilde{\kappa}_\alpha\} = 1,$$

provided that $\lim_{n\to\infty} (\log n)^{1/2} \varepsilon_n = \infty$.

KERNEL ESTIMATORS OF $f^{(k)}$. Another interpretation of our test is in terms of the kernel estimator

$$\hat{f}_h^{(k)}(t) := \frac{\int_I \psi_{t,h}(x)\, dY(x)}{n^{1/2} \int (x-t)^k \psi_{t,h}(x)\, dx} = c_{n,h} \widehat{\Psi}(t, h)$$

of $f^{(k)}(t)$, where $c_{n,h} := n^{-1/2} h^{-k-1/2} \|\psi\|_2 / (\int x^k \psi(x)\, dx)$. Then the test statistic $T(Y)$ may be written as

$$T(Y) = \sup_{h \in]0, R^{-1}/2]} \left(c_{n,h}^{-1} \sup_{t \in [Rh, 1-Rh]} \hat{f}_h^{(k)}(t) - C(2Rh) \right).$$

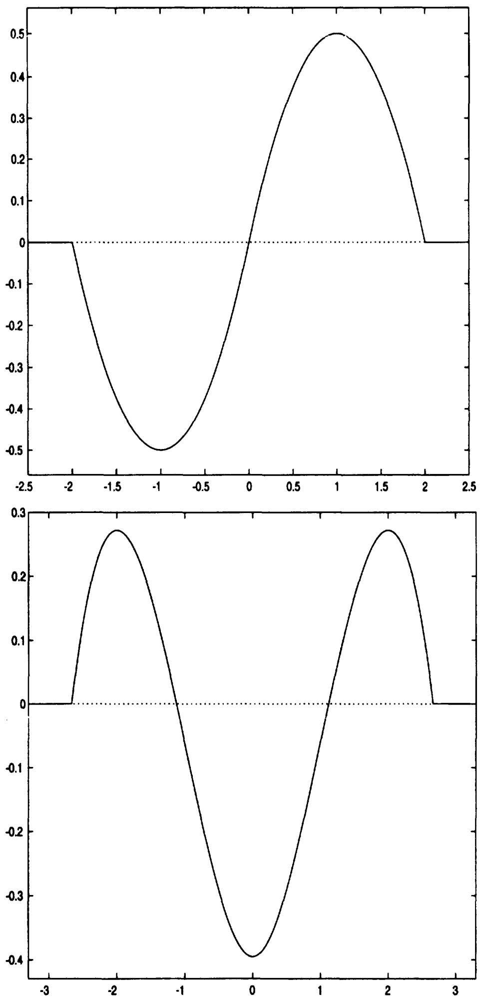

FIG. 2. *The test signals $\psi_{\downarrow}$ and ψ_{conc}.*

Therefore our test identifies pairs (t, h) such that $\hat{f}_h^{(k)}(t)$ is significantly greater than 0. This shows that our methods are related and have potential applications to Chaudhuri and Marron's (1999) method. Translated into the present setup, the latter authors use test statistics such as

$$\sup_{h \in [a, b]} \sup_{t \in [Rh, 1-Rh]} (nh)^{1/2} |\hat{f}_h^{(1)}(t)|,$$

with fixed $[a, b] \subset\,]0, 1[$ in order to identify a set of pairs (t, h) such that $\mathbb{E}\hat{f}_h^{(1)}(t) \neq 0$ (with a certain confidence).

RATE OPTIMALITY. The rate ρ_n appearing in Theorem 3.2 coincides with the optimal rate for estimating the kth derivative of a function in $\mathscr{F}(k+1, L)$ with respect to the sup-norm; see Ibragimov and Khasminskii (1980). Moreover, our optimization problem (3.9) is closely related to (but does not coincide with) the optimal recovery problem from Donoho (1994a, b) arising in the estimation of a function and its derivatives in sup-norm. Similarly to the estimation problem, the case of a smoothness degree differing from $k+1$ would require different test signals. At the same time, one can easily verify that the test proposed here yields the optimal rate of testing for an arbitrary Hölder class $\mathscr{F}(\beta, L)$ with $\beta > k$.

4. Modifications and further developments.

4.1. *Gaussian regression.* Suppose that instead of the process Y on I we observe a random vector $\vec{Y} \in \mathbf{R}^n$ with components

$$Y_i = f(x_i) + \varepsilon_i \quad \text{for } i = 1, 2, \dots, n, \tag{4.12}$$

where $x_i := (i - 1/2)/n$, and the random errors ε_i are independent with Gaussian distribution $\mathscr{N}(0, \sigma^2)$. One can show that Theorems 2.2, 3.1 and 3.2 remain valid with σc_* in place of c_*, provided that we replace $\widehat{\Psi}(t, h)$ with

$$\widehat{\Psi}_n(t, h) := \sigma^{-1}\Big(\sum_{i=1}^n \psi_{t,h}(x_i)^2\Big)^{-1/2} \sum_{i=1}^n \psi_{t,h}(x_i) Y_i. \tag{4.13}$$

Moreover, it suffices to consider pairs (t, h) such that $t = j/n$ and $h = R^{-1}d/n$ for integers $d \in [1, n/2]$ and $j \in [d, n-d]$.

Suppose that σ is unknown and replaced with an estimator $\hat{\sigma}_n$. Then our tests are asymptotically valid and keep their optimality properties provided that

$$|\hat{\sigma}_n/\sigma - 1| = o_p((\log n)^{-1/2}). \tag{4.14}$$

For instance, let $\hat{\sigma}_n^2$ be defined as

$$(2(n-1))^{-1} \sum_{i=1}^n (Y_i - Y_{i-1})^2 \text{ or } (6(n-2))^{-1} \sum_{i=1}^{n-1} (2Y_i - Y_{i-1} - Y_{i+1})^2;$$

see Rice (1984) for the first and Gasser, Sroka and Jennen-Steinmetz (1986) for the second proposal. Then (4.14) holds whenever f has bounded total variation $\mathrm{TV}(f)$. Indeed, elementary calculations show that

$$\mathbb{E}((\hat{\sigma}_n^2/\sigma^2 - 1)^2) = O((1 + TV(f)^2)/n).$$

4.2. *General regression models.* If one observes $Y_i = f(x_i) + E_i$ for $i = 1, 2, \dots, n$ with arbitrary fixed numbers x_i and independent, identically distributed random errors E_i, one can modify the multiscale tests of $\mathscr{H}_\downarrow$ in Section 3.2 using linear rank statistics in place of the linear statistics in (4.13); see Dümbgen (1998). In that paper the aspect of localizing interesting features such as modes is discussed in more detail.

4.3. *Other testing problems.* If a qualitative property of f is plausible one can construct a confidence set for f under this assumption only. There are asymptotically optimal and adaptive confidence bands for monotone or concave functions f based on appropriate multiscale statistics; see Dümbgen (2000).

5. Numerical examples. In this section we illustrate the tests of Section 3.2 for $\mathscr{H}_\downarrow$ and $\mathscr{H}_{\mathrm{conc}}$ within the Gaussian regression model (4.12) with sample size $n = 700$ and standard deviation $\sigma = 1$. For notational convenience the test signals $\psi_\downarrow$ and ψ_{conc} are rescaled to have support $[-1, 1]$, namely, $\psi_\downarrow(x) := 1\{|x| \le 1\}x(1 - |x|)$ and $\psi_{\mathrm{conc}} := 1\{|x| \le 1\}(-1/8 + 9x^2/8 - |x|^3)$.

As for $\mathscr{H}_\downarrow$, we estimated the distribution function of

$$\widetilde{T}_n(\vec{Y}) := \max_{h \in S_n} \widetilde{T}_n(\vec{Y}, h), \text{ with } \widetilde{T}_n(\vec{Y}, h) := \max_{t \in L_n(h)} (\widehat{\Psi}_n(t, h) - C(2h))$$

in the case $f \equiv 0$, using 9999 Monto Carlo simulations. Here S_n denotes the set of scale parameters $1/n, 2/n, \ldots, \lfloor n/2 \rfloor$, and $L_n(h)$ stands for the set of location parameters $h, 2h, \ldots, 1 - h$. Further $\widehat{\Psi}_n(t, h)$ is the linear filter defined in (4.13) with the test signal $\psi_\downarrow$. Here are some estimated $(1 - \alpha)$ quantiles $\tilde{\kappa}_{n,\alpha}$:

α	0.50	0.10	0.05
$\tilde{\kappa}_{n,\alpha}$	1.029	1.773	2.018

Figure 3 shows four realizations of the random function $\widetilde{T}_n(\vec{Y}, \cdot)$ on S_n, again in the case $f \equiv 0$. The lower dashed line depicts the additive correction term, $h \mapsto -C(2h)$, while the upper horizontal line shows the critical value $\tilde{\kappa}_{n,0.05} = 2.018$.

The process $\widetilde{T}_n(\vec{Y}, \cdot)$ behaves differently if, for example, f is the function depicted in Figure 1. Figure 4 shows observations Y_i together with this regression function f (left plot) and the corresponding process $\widetilde{T}_n(\vec{Y}, \cdot)$ (right plot). We see the critical value $\tilde{\kappa}_{n,0.05}$ is exceeded for bandwidths h in two disjoint regions. For two of these bandwidths Figure 5 shows the process $\widehat{\Psi}_n(t, \cdot)$ on $L_n(h)$ (upper row). In addition, for both bandwidths a location parameter t with $\widehat{\Psi}_n(t, h) > C(2h) + \tilde{\kappa}_{n,0.05}$ was picked. Each plot in the lower row shows the data vector $\vec{Y}$ together with its orthogonal projection onto the linear span of

$$(1\{|x_i - t| < h\})_{i=1}^n \quad \text{and} \quad (\psi_{t,h}(x_i))_{i=1}^n.$$

Note that the larger bandwidth enables us to find a moderate increasing trend over a large interval on the left-hand side, while the smaller bandwidth is appropriate for detecting and localizing a sharp increasing trend of f over a smaller interval on the right-hand side. Indeed the underlying function f is a quadratic spline (i.e., $f^{(2)}$ is piecewise constant) such that

$$\widetilde{\Delta}_{[0,0.05]}(f) = f^{(1)}(0.33) = 5.8 \quad \text{and} \quad \widetilde{\Delta}_{[0.5,1]}(f) = f^{(1)}(0.80) = 60.7,$$

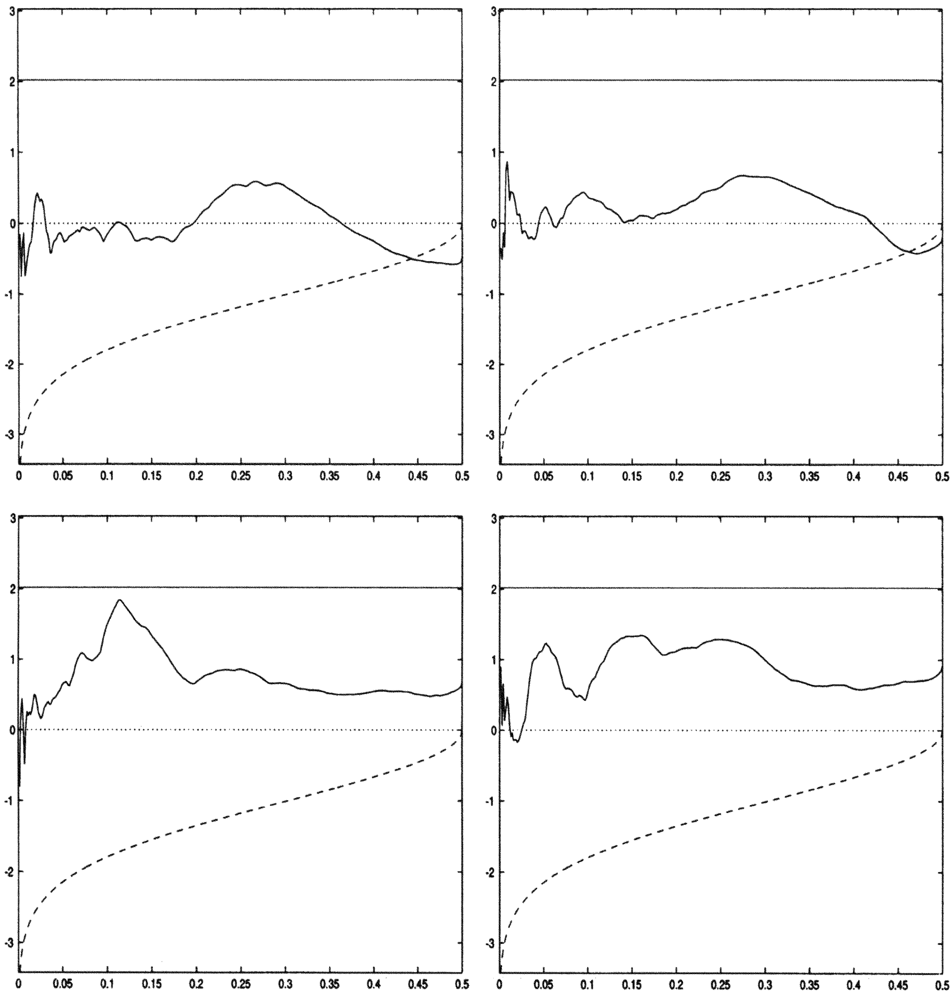

FIG. 3. *Four realizations of* $\widetilde{T}_n(\vec{Y}, \cdot)$ *in the case* $f \equiv 0$.

whereas

$$\sup_{t \in [0.1, 0.5]} |f^{(2)}| = 90 \quad \text{and} \quad \sup_{t \in [0.77, 0.83]} |f^{(2)}| = 2720.$$

For $n = 700$ and $k = 1$ the number $\rho_n c_*(90)$ in Theorem 3.2 equals 5.518, which is slightly smaller than $f^{(1)}(0.33)$. The number $\rho_n c_*(2720)$ equals 42.66, which is about 0.71 times $f^{(1)}(0.80)$. These pictures and numbers illustrate the benefits of using several bandwidths simultaneously, which yields spatial adaptivity as stated in Theorem 3.2.

Now we show analogous plots for a function $f \notin \mathscr{H}_{\mathrm{conc}}$ and the multiscale statistic $\widetilde{T}_n$ based on the test signal ψ_{conc}. More precisely, we define $\widehat{\Psi}_n(t, h)$

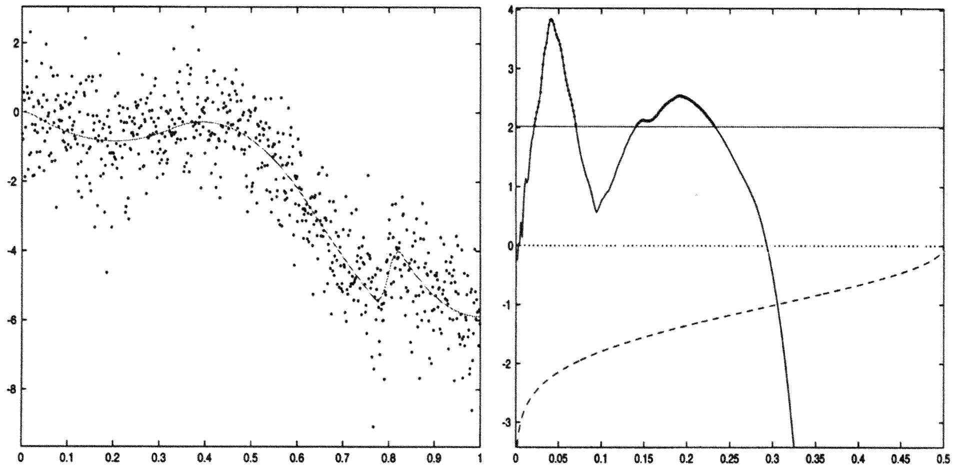

FIG. 4. *Simulated data with $f \notin \mathscr{H}_{\downarrow}$ and $\widetilde{T}_n(\vec{Y}, \cdot)$.*

as in (4.13) with

$$\psi(x, nh) := 1\{|x| \le 1\}(-a(nh) + (1 + a(nh))x^2 - |x|^3)$$

in place of $\psi(x)$, where $a(d) := (1 + d^{-2}/2)/(8 + d^{-2})$. For then $\mathbb{E}_g\widehat{\Psi}(t, h) \le 0$ for all $g \in \mathscr{H}_{\text{conc}}$, a consequence of Lemma 6.1. Figure 6 shows simulated data and the process $\widetilde{T}_n(\vec{Y}, \cdot)$. Figure 7 shows the process $\widehat{\Psi}_n(\cdot, h)$ for two different bandwidths together with "convex features" of the data. The latter are orthogonal projections of $\vec{Y}$ onto the linear span of

$$(1\{|x_i - t| < h\})_{i=1}^n, \quad (1\{|x_i - t| < h\}(x_i - t))_{i=1}^n \quad \text{and} \quad (\psi_{t,h}(x_i, nh))_{i=1}^n.$$

6. Proofs.

6.1. *An extension of Lévy's modulus of continuity.* Theorem 2.1 may be seen as a generalization of Lévy's modulus of continuity for Brownian motion [cf. Shorack and Wellner (1986), Theorem 4.1.1]. For if we apply Theorem 2.1 to $\psi(x) := 1\{|x| \le 1\}$, then

$$\widehat{\Psi}(t, h) - \mathbb{E}\widehat{\Psi}(t, h) = (2h)^{-1/2}(W(t + h) - W(t - h)),$$

so that

$$\sup_{s,t \in I: s<t} \left(\frac{|W(t) - W(s)|}{(t - s)^{1/2}} - C(t - s) \right) \Big/ D(t - s) < \infty \quad \text{almost surely.}$$

Theorem 2.1 itself follows from a general theorem about stochastic processes with sub-Gaussian increments on some pseudometric space $(\mathscr{T}, \rho)$. For any subset $\mathscr{T}'$ of $\mathscr{T}$ and $\varepsilon > 0$ the capacity number (covering number) $N(\varepsilon, \mathscr{T}')$ is defined as the supremum of $\#\mathscr{T}''$ over all $\mathscr{T}'' \subset \mathscr{T}'$ such that $\rho(a, b) > \varepsilon$ for arbitrary different points $a, b \in \mathscr{T}''$.

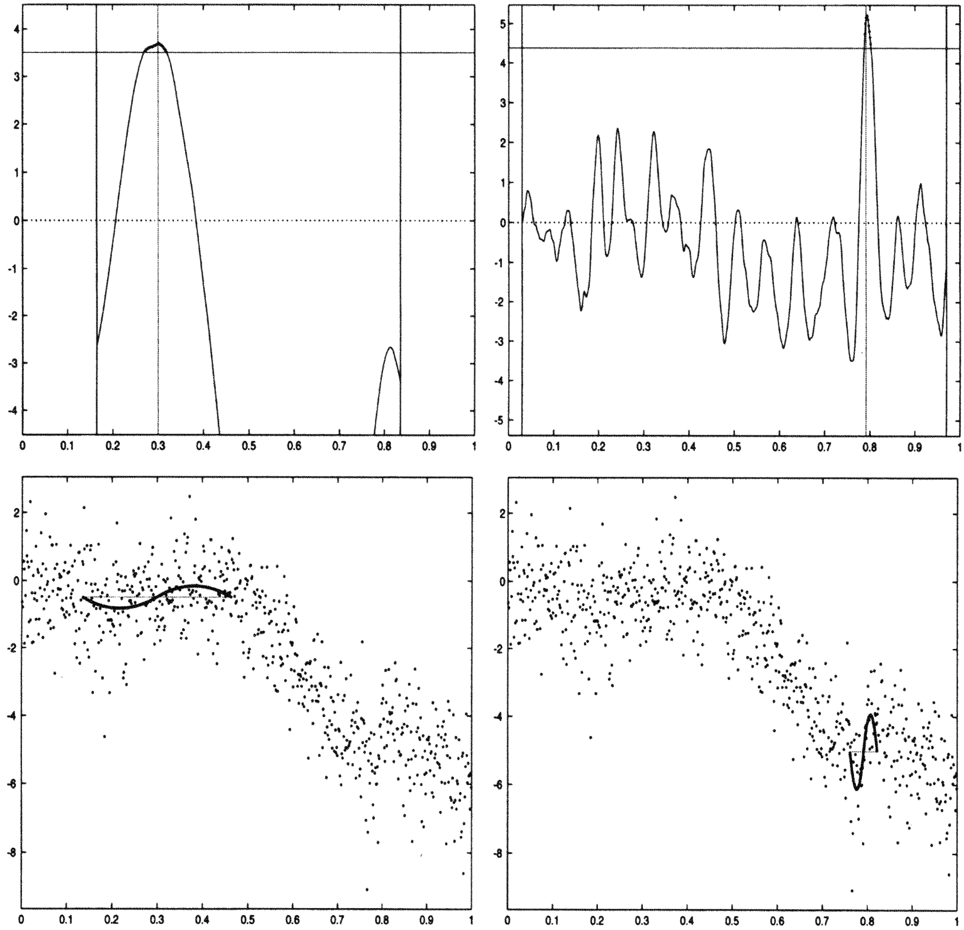

FIG. 5. *The process* $\widehat{\Psi}_n(\cdot, h)$ *for* $h = 0.170$ *and* $h = 0.029$.

THEOREM 6.1. *Let* X *be a stochastic process on a pseudometric space* $(\mathscr{T}, \rho)$ *with continuous sample paths. Suppose that the following three conditions hold*:

(i) *There is a function* $\sigma : \mathscr{T} \to]0, 1]$ *and a constant* $K \geq 1$ *such that*

$$\mathbb{P}\{X(a) > \sigma(a)\eta\} \leq K \exp(-\eta^2/2) \quad \textit{for all } \eta > 0 \textit{ and } a \in \mathscr{T}.$$

Moreover,

$$\sigma(b)^2 \leq \sigma(a)^2 + \rho(a, b)^2 \quad \textit{for all } a, b \in \mathscr{T}.$$

(ii) *For some constants* $L, M \geq 1$,

$$\mathbb{P}\{|X(a) - X(b)| > \rho(a, b)\eta\} \leq L \exp(-\eta^2/M) \quad \textit{for all } \eta > 0 \textit{ and } a, b \in \mathscr{T}.$$

(iii) *For some constants* $A, B, V > 0$,

$$N((\delta u)^{1/2}, \{a \in \mathscr{T} : \sigma(a)^2 \leq \delta\}) \leq A u^{-B} \delta^{-V} \quad \textit{for all } u, \delta \in]0, 1].$$

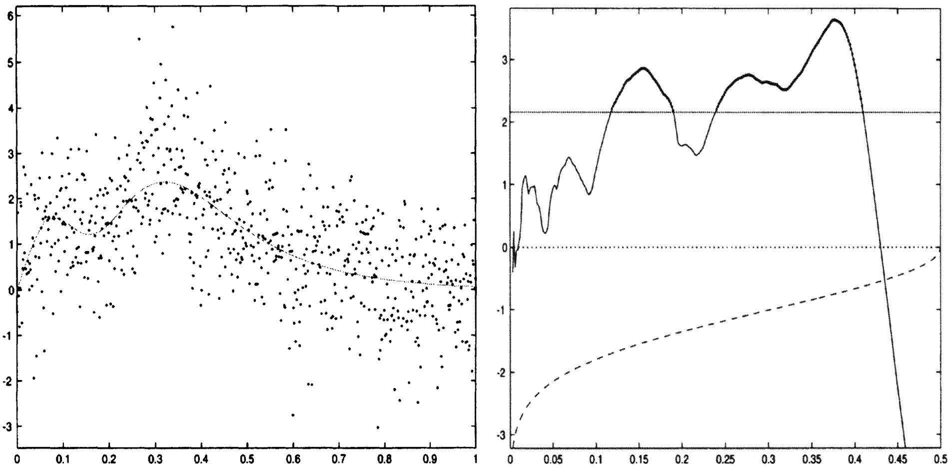

FIG. 6. *Simulated data with $f \notin \mathscr{H}_{\text{conc}}$ and $\widetilde{T}_n(\vec{Y}, \cdot)$.*

Then the random variable

$$S(X) := \sup_{a \in \mathscr{T}} \frac{X(a)^2/\sigma(a)^2 - 2V \log(1/\sigma(a)^2)}{\log\log(e^e/\sigma(a)^2)}$$

is finite almost surely. More precisely, $\mathbb{P}\{S(X) > r\} \le p(r)$ *for some function* p *depending only on the constants* K, L, M, A, B, V *such that* $\lim_{r\to\infty} p(r) = 0$.

REMARK 1. By definition of $S(X)$, the ratio $X(a)^2/\sigma(a)^2$ is not greater than $2V\log(\sigma(a)^{-2}) + S(X)\log\log(e^e\sigma(a)^{-2})$ for arbitrary $a \in \mathscr{T}$. Since $(x+y)^{1/2} \le x^{1/2} + x^{-1/2}y/2$ for arbitrary positive numbers x and y, Theorem 6.1 implies that

$$\sup_{a\in\mathscr{T}}(|X(a)|/\sigma(a) - C(\sigma(a)^2))/D(\sigma(a)^2) < \infty \quad \text{almost surely,}$$

with $C(\cdot)$ and $D(\cdot)$ as defined in Theorem 2.1.

REMARK 2. Theorem 6.1 can be applied, for instance, to stochastic processes whose index set is the family of all quadrangles in $[0, 1]^d$ or the family of all Euclidean balls on the unit sphere in $\mathbf{R}^d$. Thus it has potential applications to multiscale tests for image analysis and for directional data.

PROOF OF THEOREM 6.1. For positive numbers v let

$$\omega(X, v) := \sup_{a, b\in\mathscr{T}:\rho(a, b)\le v} |X(a) - X(b)|.$$

It follows from assumptions (ii) and (iii) with $\delta = 1$, Theorem 2.2.4 of van der Vaart and Wellner (1996) and elementary calculations that

$$\mathbb{P}\{\omega(X, v) > \eta\} \le C\exp\left(-\frac{\eta^2}{CV^2\log(e/v)}\right) \quad \text{for } 0 < v \le 1, \eta > 0. \tag{6.15}$$

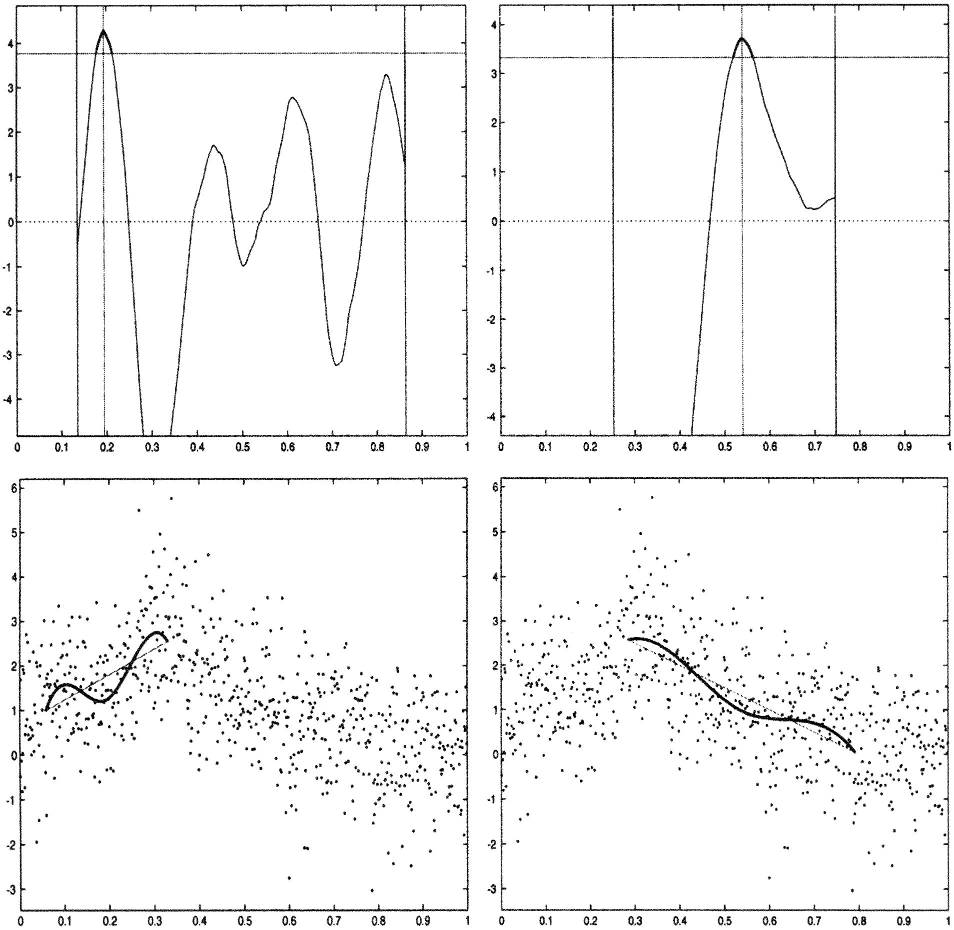

FIG. 7. *The process* $\widehat{\Psi}_n(\cdot, h)$ *for* $h = 0.136$ *and* $h = 0.253$.

Here and throughout the sequel C denotes a generic positive constant depending only on K, L, M, A, B, V. Its value may differ from place to place.

For $0 < \delta \leq 1$ let $\mathscr{T}(\delta) := \{a \in \mathscr{T} : \delta/2 < \sigma(a)^2 \leq \delta\}$. Now fix some $u \leq 1/2$, and let $\mathscr{T}(\delta, u)$ be a maximal subset of $\mathscr{T}(\delta)$ such that $\rho(a, b)^2 > u\delta$ for arbitrary different $a, b \in \mathscr{T}(\delta, u)$. For each $a \in \mathscr{T}(\delta)$ there exists a point $\tilde{a} \in \mathscr{T}(\delta, u)$ such that $\rho(a, \tilde{a})^2 \leq u\delta$. In particular,

$$\sigma(a)^2 \geq \sigma(\tilde{a})^2 - u\delta \geq \sigma(\tilde{a})^2(1 - 2u)$$

by assumption (ii) and the definition of $\mathscr{T}(\delta)$. For $0 < \lambda < 1$ and $r > 0$, the inequality

$$X(a)^2 > \sigma(a)^2 r$$

implies that either

$$\omega\big(X, (u\delta)^{1/2}\big)^2 \geq |X(a) - X(\tilde{a})|^2 > \lambda^2 X(a)^2 \geq \lambda^2 \delta r/2$$

or

$$X(\tilde{a})^2 \geq (1-\lambda)^2 X(a)^2 > (1-\lambda)^2 \sigma(a)^2 r \\ \geq (1-\lambda)^2 (1-2u)\sigma(\tilde{a})^2 r.$$

Thus, for any nonincreasing function $r :]0, 1] \to]0, \infty[$,

$$\begin{aligned}
\Pi(\delta) &:= \mathbb{P}\{X(a)^2/\sigma(a)^2 > r(a) \text{ for some } a \in \mathscr{T}(\delta)\} \\
&\leq \mathbb{P}\Big\{\omega(X, (u\delta)^{1/2})^2 > \lambda^2 \delta r(\delta)/2\Big\} \\
&\quad + \sum_{b \in \mathscr{T}(\delta, u)} \mathbb{P}\{X(b)^2 > (1-\lambda)^2(1-2u)\sigma(b)^2 r(\delta)\} \\
&\leq C \exp\left(-\frac{\lambda^2 r(\delta)}{Cu \log(e/(u\delta))}\right) \\
&\quad + Cu^{-B}\delta^{-V} \exp\left(-\frac{(1-\lambda)^2(1-2u)r(\delta)}{2}\right) \\
&\leq C \exp\left(-\frac{\lambda^2 r(\delta)}{Cu \log(e/(u\delta))}\right) \\
&\quad + C \exp\big(B \log(1/u) + V \log(1/\delta) + ur(\delta) - (1/2 - \lambda) r(\delta)\big)
\end{aligned}$$

according to assumptions (i) and (iii) and inequality (6.15). Specifically let

$$r(\delta) := 2V \log(1/\delta) + S \log \log(e^e/\delta)$$

for some constant $S \geq 1$. If we set

$$\lambda = \lambda(\delta) := (S/4) \log \log(e^e/\delta)/r(\delta),$$

then $(1/2 - \lambda) r(\delta) = V \log(1/\delta) + (S/4) \log \log(e^e/\delta)$, whence $\Pi(\delta)$ is not greater than

$$C \exp\left(-\frac{S^2 (\log \log(e^e/\delta))^2}{Cur(\delta) \log(e/(u\delta))}\right) \\ + C \exp(B \log(1/u) + ur(\delta) - (S/4) \log \log(e^e/\delta)).$$

Finally, let

$$u = u(\delta) := (r(\delta) \log(e/\delta))^{-1},$$

which is less than $1/2$ if $S \geq 2$. Then $1/u \leq C(\log(e/\delta))^2$, so that

$$\Pi(\delta) \leq C \exp((C - S/C) \log \log(e^e/\delta)).$$

Now we apply this bound to $\delta = 2^{-k}$, $k \geq 0$. This yields

$$\begin{aligned}
&\mathbb{P}\Big\{X(a)^2/\sigma(a)^2 > 2V\,\log(1/\sigma(a)^2) + S\log\log(e^e/\sigma(a)^2) \text{ for some } a \in \tau\Big\}\\
&\quad \leq \sum_{k=0}^{\infty} \Pi(2^{-k})\\
&\quad \leq C\sum_{k-0}^{\infty} \exp(-(S/C - C)\log\log(e^e 2^k))\\
&\quad = C\sum_{k=0}^{\infty} (e + k\log 2)^{-(S/C-C)}\\
&\quad \to 0 \quad \text{as } S \to \infty.
\end{aligned}$$

□

PROOF OF THEOREM 2.1. Without loss of generality let $f = 0$, $R = 1$ and $\|\psi\|_2 = 1$. Let $\mathscr{T}$ be the set of all pairs (t, h) with $0 < h \leq 1/2$, $h \leq t \leq 1-h$, and define

$$\rho((t, h), (t', h'))^2 := \mathrm{Leb}([t-h, t+h] \,\triangle\, [t'-h', t'+h']),$$

$$\sigma(t, h)^2 := \mathrm{Leb}([t-h, t+h]) = 2h.$$

Then $\sigma(b)^2 \leq \sigma(a)^2 + \rho(a, b)^2$ for all $a, b \in \mathscr{T}$, and

$$X(t, h) := (2h)^{1/2}\widehat{\Psi}(t, h) = 2^{1/2}\int_I \psi_{t,h}\,dW$$

defines a centered Gaussian process on $\mathscr{T}$ with $\mathrm{Var}(X(t, h)) = \sigma(t, h)^2$. It suffices to show that this process X and the triple $(\mathscr{T}, \rho, \sigma)$ satisfies the assumptions of Theorem 6.1 with $V = 1$; see also Remark 1 on Theorem 6.1.

Since $\mathbb{P}\{|Z| \geq \eta\} \leq \exp(-\eta^2/2)$ for standard Gaussian random variables Z, our process X satisfies condition (i) with $K = 1$. As for the continuity of its sample paths, the assumptions about ψ imply that

$$\psi(x) = \int_{[-1,x]} g\,dP$$

for all but at most countably many numbers $x \in [-1, 1]$, where P is some probability measure on $[-1, 1]$, and g is some measurable function with $|g| \leq \mathrm{TV}(\psi)$, $\int g\,dP = 0$. Integration by parts shows that

$$X(t, h) = -2^{1/2}\int g(x)W(t+hx)P(dx),$$

which is continuous in (t, h) by continuity of W and dominated convergence. Moreover,

$$\begin{aligned}
&\operatorname{Var}(X(t, h) - X(t', h')) \\
&\quad = 2\operatorname{Var}\Big(\int g(x)\big(W(t+hx) - W(t'+h'x)\big)P(dx)\Big) \\
&\quad \le 2\Big(\int |g(x)|\,|t+hx-t'-h'x|P(dx)\Big)^2 \\
&\quad \le 2\mathrm{TV}(\psi)^2\rho\big((t,h),(t',h')\big)^2.
\end{aligned}$$

Hence condition (ii) of Theorem 6.1 holds with $L = 1$ and $M = 4\mathrm{TV}(\psi)^2$. Finally,

$$N\Big((u\delta)^{1/2},\ \big\{a \in \mathcal{T} : \sigma(a)^2 \le \delta\big\}\Big) \le 12u^{-2}\delta^{-1} \quad \text{for all } u, \delta \in\,]0,1].$$

For let $\mathcal{T}''$ be any maximal subset of $\{a \in \mathcal{T} : \sigma(a)^2 \le \delta\}$ such that $\rho(a,b)^2 > u\delta$ for arbitrary different points $a, b \in \mathcal{T}''$. With $m := \lfloor 2/(u\delta)\rfloor$ define $M_j := [(j-1)u\delta/2, ju\delta/2[$ for $j = 1, 2, \ldots, m$ and $M_{m+1} := [mu\delta/2, 1]$. For any $(t,h) \in \mathcal{T}''$ let $t - h \in M_j$ and $t + h \in M_k$. The inequalities $0 < 2h \le \delta$ imply that

$$0 \le k - j \le 1 + 2/u,$$

and there are at most $(1+2/(u\delta))(2+2/u)$ pairs (j, k) with these properties. Moreover, since all sets M_l have length at most $u\delta/2$, for any pair (j,k) of integers there is at most one point $(t,h) \in \mathcal{T}''$ such that $t - h \in M_j$ and $t + h \in M_k$. Thus the cardinality of $\mathcal{T}''$ is not greater than $(1+2/(u\delta))(2+2/u) \le 12u^{-2}\delta^{-1}$. □

6.2. *Basic properties of the test signals.* Here we collect some useful statements about test signals ψ.

PROOF OF INEQUALITY (3.6). Since $\widehat{\Psi}(t,h) = \widehat{\Psi}(t,h,Y)$ can be written as $\widehat{\Psi}(t,h,W) + n^{1/2}\langle f, \psi_{t,h}\rangle$, it suffices to show that for $h \in\,]0,1/2]$ and $h \le s \le t \le 1-h$ the following inequalities hold:

$$\begin{aligned}
\langle \psi_{t,h} f\rangle &\le 0 \quad \text{for } f \in \mathcal{H}_{\le 0},\\
\langle \psi_{t,h} - \psi_{s,h,} f\rangle &\le 0 \quad \text{for } f \in \mathcal{H}_{\downarrow}.
\end{aligned}$$

The assertion about $\mathcal{H}_{\le 0}$ is obvious, because $\psi_{t,h} \ge 0$. The assertion about $\mathcal{H}_{\downarrow}$ follows from Lemma 6.1, because $\int(\psi_{t,h} - \psi_{s,h})(x)\,dx = 0$ and

$$\psi_{t,h} - \psi_{s,h} \begin{cases} \ge 0 & \text{on} \quad [(s+t)/2,\ \infty[, \\ \le 0 & \text{on} \quad]-\infty,\ (s+t)/2]. \end{cases}$$

□

LEMMA 6.1. *Let μ be some measure on the line, and let $\psi \in L^2(\mu)$.*

(a) *Suppose that $\int \psi(x)\mu(dx) = 0$ and*

$$\psi \begin{cases} \leq 0 & \text{on }]-\infty, a[, \\ \geq 0 & \text{on }]a, \infty[\end{cases}$$

for some real number a. Then $\int \psi(x) f(x)\mu(dx) \leq 0$ for all $f \in \mathscr{H}_\downarrow$.

(b) *Suppose that $\int \psi(x)\mu(dx) = \int \psi(x) x\mu(dx) = 0$ and*

$$\psi \begin{cases} \geq 0 & \text{on }]-\infty, b[, \\ \leq 0 & \text{on }]b, c[, \\ \geq 0 & \text{on }]c, \infty[\end{cases}$$

for some real numbers b, c with $b < c$. Then $\int \psi(x) f(x)\mu(dx) \leq 0$ for all $f \in \mathscr{H}_{\text{conc}}$.

PROOF OF LEMMA 6.1. As for part (a), let $f \in \mathscr{H}_\downarrow$ and $\tilde{f} := f - f(a)$. Then $\tilde{f} \in \mathscr{H}_\downarrow$, and our assumptions on ψ imply that $\int \psi f \, d\mu = \int \psi \tilde{f} \, d\mu \leq 0$, because $\psi \tilde{f} \leq 0$.

Part (b) follows similarly, this time with the auxiliary function

$$\tilde{f}(x) := f(x) - \frac{x - b}{c - b}\big(f(c) - f(b)\big) - f(b).$$

If $f \in \mathscr{H}_{\text{conc}}$, then $\tilde{f}$ belongs to $\mathscr{H}_{\text{conc}}$, too, and $\int \psi f \, d\mu = \int \psi \tilde{f} \, d\mu \leq 0$, because $\psi \tilde{f} \leq 0$. $\square$

PROOF OF LEMMA 3.1. The functions $\psi_\downarrow$ and ψ_{conc} are constructed such that they satisfy the conditions of Lemma 6.1(a) and (b), respectively, where μ is Lebesgue measure on the line. Moreover, in both cases, $f_o \psi \equiv 0$. Thus condition (3.10) is satisfied. It remains to verify condition (3.11).

For $g \in \mathscr{F}(2, 1)$ with $f^{(1)}(0) \geq 1$ the inner product $\langle g, \psi_\downarrow \rangle$ equals $\langle \tilde{g}, \psi_\downarrow \rangle$, where $\tilde{g}(x) := (g(x) - g(-x))/2$, because $\psi_\downarrow$ is an odd function. Since $\tilde{g}$ is an odd function in $\mathscr{F}(2, 1)$ with $\tilde{g}^{(1)}(0) = g^{(1)}(0) \geq 1$,

$$\tilde{g}(x) = \int_0^x \tilde{g}^{(1)}(s) \, ds \geq \int_0^x (1 - s) \, ds = \psi_\downarrow(x) \geq 0 \quad \text{for } x \in [0, 2],$$

so that

$$\langle \tilde{g}, \psi_\downarrow \rangle = 2\langle \tilde{g}, \psi_\downarrow \rangle_{[0, 2]} \geq 2\|\psi_\downarrow\|^2_{2, [0, 2]} = \|\psi_\downarrow\|^2_2.$$

Let $b = 8/3$, and let a be the unique point in $]0, b[$ with $\psi_{\text{conc}}(a) = 0 = \psi_{\text{conc}}(b)$. For $g \in \mathscr{F}(3, 1)$ with $g^{(2)}(0) \geq 1$ the inner product $\langle g, \psi_{\text{conc}} \rangle$ equals $\langle \tilde{g}, \psi_{\text{conc}} \rangle$, where $\tilde{g}(x) := (g(x) + g(-x))/2 - (g(a) + g(-a))/2$, because ψ_{conc} is an even function with $\langle 1, \psi_{\text{conc}} \rangle = 0$. Since $\tilde{g}$ is an even function in $\mathscr{F}(3, 1)$ with $\tilde{g}^{(2)}(0) = g^{(2)}(0) \geq 1$,

$$\tilde{g}^{(1)}(x) = \int_0^x \tilde{g}^{(2)}(s) \, ds \geq \int_0^x (1 - s) \, ds = \psi^{(1)}_{\text{conc}}(x) \quad \text{for } x \in [0, b[.$$

This, together with $\tilde{g}(a) = \psi_{\text{conc}}(a) = 0$, implies that

$$\tilde{g} \le \psi_{\text{conc}} \le 0 \quad \text{on } [0, a],$$
$$\tilde{g} \ge \psi_{\text{conc}} \ge 0 \quad \text{on } [a, b].$$

Consequently,

$$\langle \tilde{g},\ \psi_{\text{conc}} \rangle = 2\langle \tilde{g},\ \psi_{\text{conc}} \rangle_{[0,b]} \ge 2\|\psi_{\text{conc}}\|^2_{2,[0,b]} = \|\psi_{\text{conc}}\|_2^2. \qquad \square$$

6.3. *Minimax optimality.* The proofs of Theorems 2.2(a) and 3.1(a) rely on the following result about Gaussian likelihood ratios [cf. Ingster (1993) or Lepski and Tsybakov (2000)].

LEMMA 6.2. *Let $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ be independent random variables with standard Gaussian distribution. If $w_m = (2\log m)^{1/2}(1 - \varepsilon_m)$ with $\lim_{m\to\infty} \varepsilon_m = 0$ and $\lim_{m\to\infty} (\log m)^{1/2}\varepsilon_m = \infty$, then*

$$\lim_{m\to\infty} \mathbb{E}\left| m^{-1} \sum_{i=1}^{m} \exp(w_m \Gamma_i - w_m^2/2) - 1 \right| = 0.$$

For the reader's convenience a proof is given here.

PROOF OF LEMMA 6.2. Let $Z_m := \exp(w_m\Gamma_1 - w_m^2/2)$. Since $\mathbb{E}Z_m = 1$, the assertion follows from the weak law of large numbers for triangular arrays, provided that

$$\lim_{m\to\infty} \mathbb{E}1\{|Z_m - 1| \ge \eta m\}|Z_m - 1| = 0 \quad \text{for any } \eta > 0.$$

But for $m \ge 1/\eta$, the expectation of $1\{|Z_m - 1| \ge \eta m\}|Z_m - 1|$ is not greater than

$$\begin{aligned} \mathbb{E}1\{Z_m \ge \eta m\}Z_m &\le \mathbb{E}Z_m^{1+\delta}(\eta m)^{-\delta} \quad \text{for any } \delta > 0 \\ &= \exp(\delta(1+\delta)w_m^2/2 - \delta \log(\eta m)) \\ &= \exp(\delta(1+\delta)(1-\varepsilon_m)^2 \log m - \delta \log m - \delta \log \eta) \end{aligned}$$

for any $\delta > 0$. In the case $\delta = \varepsilon_m$ the latter bound equals

$$\exp\Big(-\Big(\varepsilon_m^2 + O(\varepsilon_m^3)\Big)\log m + o(1)\Big) \to 0 \quad \text{as } m \to \infty. \qquad \square$$

PROOF OF THEOREM 3.1(a). Let ψ be the triangular kernel with $\psi(x) = 1\{|x| \le 1\}(1 - |x|)$. For a given bandwidth $h \in\,]0, 1/2]$ and any integer j let

$$g_j := Lh\psi_{(2j-1)h,\,h}.$$

All these functions g_j belong to $\mathscr{F}(1, L)$. Now let $[a, a+2b] \subset J \subset\,]0, 1[$ for some $b > 0$. For $l = 1, 2$ define

$$\mathscr{J}_l := \{\text{integers } j : (2j-1)h \in\,]a + (l-1)b, a + lb]\}.$$

These sets $\mathcal{J}_l$ contain at least $b/(2h) - 1$ indices, and

$$\{g \in \mathcal{F}(1, L) : \Delta_J(g) \geq Lh\} \supset \begin{cases} \{g_k : k \in \mathcal{J}_2\}, & \text{if } \mathcal{H}_o = \mathcal{H}_{\leq 0}, \\ \{g_k - g_j : (j, k) \in \mathcal{J}_1 \times \mathcal{J}_2\}, & \text{if } \mathcal{H}_o = \mathcal{H}_{\downarrow}. \end{cases}$$

Let $\mathcal{G}_o$ denote the finite set on the right-hand side, depending on $\mathcal{H}_o$. Then for any test $\phi : \mathcal{C}[0, 1] \to [0, 1]$ with $\mathbb{E}_0\phi(Y) \leq \alpha$,

$$\begin{aligned} &\inf_{g \in \mathcal{F}(1,L):\Delta_J(g) \geq Lh} \mathbb{E}_g\phi(Y) - \alpha \\ &\quad \leq \min_{g \in \mathcal{G}_o} \mathbb{E}_g\phi(Y) - \mathbb{E}_0\phi(Y) \\ &\quad \leq (\#\mathcal{G}_o)^{-1} \sum_{g \in \mathcal{G}_o} \mathbb{E}_g\phi(Y) - \mathbb{E}_0\phi(Y) \\ &\quad \leq \mathbb{E}_0\left((\#\mathcal{G}_o)^{-1} \sum_{g \in \mathbb{G}_o} \frac{d\mathbb{P}_g}{d\mathbb{P}_0}(Y) - 1 \right)\phi(Y) \\ &\quad \leq \mathbb{E}_0\left| (\#\mathcal{G}_o)^{-1} \sum_{g \in \mathcal{G}_o} \frac{d\mathbb{P}_g}{d\mathbb{P}_0}(Y) - 1 \right|. \end{aligned} \tag{6.16}$$

Now we want to determine $h = h_n$ such that the right-hand side tends to 0 as $n \to \infty$.

Recall that $\log(d\mathbb{P}_g/d\mathbb{P}_0)(Y) = n^{1/2} \int_I g\, dY - n\|g\|_2^2/2$. If $g = Lh\psi_{(2j-1)h,\, h}$, the stochastic integral $n^{1/2} \int g\, dY$ is equal to $n^{1/2}Lh^{3/2}\|\psi\|_2\widehat{\Psi}((2j-1)h,\, h)$. With $\Gamma_i := (-1)^l\widehat{\Psi}((2i-1)h,\, h)$ for $i \in \mathcal{J}_l$, the random variables Γ_i, $i \in \mathcal{J}_1 \cup \mathcal{J}_2$, are independent and standard normally distributed under $\mathbb{P}_0$. If we define the constant $w := n^{1/2}Lh^{3/2}\|\psi\|_2$ and the random variable $Z_i := \exp(w\Gamma_i - w^2/2)$, Then we can write

$$\begin{aligned} \frac{d\mathbb{P}_{gk}}{d\mathbb{P}_0}(Y) - 1 &= Z_k - 1, \\ \frac{d\mathbb{P}_{gk-gj}}{d\mathbb{P}_0}(Y) - 1 &= Z_jZ_k - 1 = (Z_j - 1)(Z_k - 1) + (Z_j - 1) + (Z_k - 1) \end{aligned}$$

for $j \in \mathcal{J}_1$, $k \in \mathcal{J}_2$. Consequently,

$$(\#\mathcal{G}_o)^{-1} \sum_{g \in \mathcal{G}_o} (d\mathbb{P}_g/d\mathbb{P}_0)(Y) - 1 = \begin{cases} S_2, & \text{if } \mathcal{H}_o = \mathcal{H}_{\leq 0}, \\ S_1S_2 + S_1 + S_2, & \text{if } \mathcal{H}_o = \mathcal{H}_{\downarrow}, \end{cases}$$

where $S_l := (\#\mathcal{J}_l)^{-1} \sum_{i \in \mathcal{J}_l} Z_i - 1$. Therefore, since S_1 and S_2 are independent, the expected value (6.16) tends to 0 if

$$\mathbb{E}_0|S_l| \to 0 \quad \text{for } l = 1, 2.$$

According to Lemma 6.2, the latter condition holds as $n \to \infty$, provided that $h_n \to 0$ and the corresponding $w = w_n$ satisfies

$$(\log n)^{1/2}\left(1 - \frac{w_n^2}{2\log(b/(2h_n) - 1)}\right)$$
$$= (\log n)^{1/2}\left(1 - \frac{L^2 n h_n^3/3}{\log(b/(2h_n) - 1)}\right) \to \infty.$$

If $h_n = L^{-2/3}(1 - \varepsilon_n)\rho_n$, where $\rho_n = (\log(n)/n)^{1/3}$, then

$$(\log n)^{1/2}\left(1 - \frac{w_n^2}{2\log(b/(2h_n) - 1)}\right) = (\log n)^{1/2}(1 - (1 - \varepsilon_n)^3) + o(1) \to \infty.$$

The corresponding lower bound Lh_n for $\Delta_J(g)$ equals $(1 - \varepsilon_n)L^{1/3}\rho n$, as desired. □

PROOF OF THEOREM 2.2(b). Let $\delta = \delta_n := c_*\rho n$, $h = h_n = (\delta/L)^{1/\beta}$ and $J = J_n = [Rh, 1 - Rh]$. For any $t \in J$, the probability of rejecting the null hypothesis, $\mathbb{P}_g\{T(Y) > \kappa_\alpha\}$, is bounded from below by

$$\begin{aligned}
&\mathbb{P}_g\{|\widehat{\Psi}(t,h)| > C(2h) + \kappa_\alpha\} \\
&= \mathbb{P}_0\Big\{|\widehat{\Psi}(t,h) + (n/h)^{1/2}\|\psi\|_2^{-1}\langle g, \psi_{t,h}\rangle| > C(2h) + \kappa_\alpha\Big\} \\
&\geq \mathbb{P}_0\Big\{-\text{sign}\,(\langle g, \psi_{t,h}\rangle)\widehat{\Psi}(t,h) < (n/h)^{1/2}\|\psi\|_2^{-1}|\langle g, \psi_{t,h}\rangle| - C(2h) - \kappa_\alpha\Big\} \\
&= \Phi\Big((n/h)^{1/2}\|\psi\|_2^{-1}|\langle g, \psi_{t,h}\rangle| - C(2h) - \kappa_\alpha\Big),
\end{aligned}$$

where Φ denotes the standard Gaussian distribution function. Thus it suffices to show that

$$(1 + \varepsilon_n)\max_{t\in J}(n/h)^{1/2}\|\psi\|_2^{-1}|\langle g, \psi_{t,h}\rangle| - C(2h) \to \infty$$

uniformly for all $g \in \mathscr{F}(\beta, L)$ such that $\|g\|_J \geq \delta$. Let g be any such function, and let $t \in J$ with $|g(t)| \geq \delta$. By construction of ψ and definition of h, the function $\delta\psi_{t,h}$ belongs to $\mathscr{F}(\beta, L)$, and the considerations following (2.2) show that

$$|\langle g, \psi_{t,h}\rangle| = \delta^{-1}|\langle g, \delta\psi_{t,h}\rangle| \geq \delta^{-1}\|\delta\psi_{t,h}\|_2^2 = h\delta\|\psi\|_2^2.$$

Thus

$$\begin{aligned}
&(1 + \varepsilon_n)\max_{t\in J}(n/h)^{1/2}\|\psi\|_2^{-1}|\langle g, \psi_{t,h}\rangle| - C(2h) \\
&\geq (1 + \varepsilon_n)\|\psi\|_2 n^{1/2}h^{1/2}\delta - C(2h) \\
&= \varepsilon_n(2/(2\beta + 1))^{1/2}(\log n)^{1/2} + o(1) \to \infty.
\end{aligned}$$ □

PROOF OF THEOREM 3.1(b). In the case $\mathscr{H}_o = \mathscr{H}_{\le 0}$ the proof is almost identical to the proof of Theorem 2.2(b). Thus we focus on $\mathscr{H}_o = \mathscr{H}_{\downarrow}$. Let $\delta = \delta_n = L^{1/3}\rho_n$ and $h = h_n = \delta/L$. Further let $J = J_n$ be any nonvoid subinterval of $[h, 1-h]$ and $J' = J'_n$ its h-neighborhood. For $s, t \in J$ with $s < t$ the probability $\mathbb{P}_g\{T_o(Y) > \kappa_{o,\alpha}\}$ is not smaller than

$$\begin{aligned}
&\mathbb{P}_g\{(\widehat{\Psi}(t,h) - \widehat{\Psi}(s,h))/2 > C(2h) + \kappa_{0,\alpha}\}\\
&\quad = \mathbb{P}_0\{(\widehat{\Psi}(s,h) - \widehat{\Psi}(t,h))/2\\
&\qquad\qquad < (n/h)^{1/2}\|\psi\|_2^{-1}\langle g, \psi_{t,h} - \psi_{s,h}\rangle/2 - C(2h) - \kappa_{o,\alpha}\}\\
&\quad \ge \Phi((n/h)^{1/2}\|\psi\|_2^{-1}\langle g, \psi_{t,h} - \psi_{s,h}\rangle/2 - C(2h) - \kappa_{o,\alpha}),
\end{aligned}$$

provided that the argument of $\Phi(\cdot)$ is positive, because the variance of $(\widehat{\Psi}(s,h) - \widehat{\Psi}(t,h))/2$ is not greater than 1. Thus it suffices to show that

$$(1+\varepsilon_n) \max_{s,t\in J\,:\,s<t} (n/h)^{1/2}\|\psi\|_2^{-1}\langle g, \psi_{t,h} - \psi_{s,h}\rangle/2 - C(2h) \to \infty$$

uniformly for all $g \in \mathscr{F}_{J'}(1, L)$ with $\Delta_J(g) \ge \delta$. For any such function g we pick two points $s, t \in J$ with $s < t$ and $g(t) - g(s) \ge 2\delta$. Letting $\gamma := (g(s)+g(t))/2$,

$$\begin{aligned}
\langle g, \psi_{t,h} - \psi_{s,h}\rangle/2 &= 2^{-1}\int (g(x) - \gamma)(\psi_{t,h} - \psi_{s,h})(x)\,dx\\
&\ge 2^{-1}\int ((\delta - L|x-t|)\psi_{t,h}(x) - (-\delta + L|x-s|)\psi_{s,h}(x))\,dx\\
&= h\int_{-1}^{1} (\delta - Lh|x|)(1-|x|)\,dx\\
&= Lh^2\|\psi\|_2^2.
\end{aligned}$$

Thus

$$\begin{aligned}
&(1+\varepsilon_n) \max_{s,t\in J\,:\,s<t} (n/h)^{1/2}\|\psi\|_2^{-1}\langle \tilde{g}, \psi_{t,h} - \psi_{s,h}\rangle/2 - C(2h)\\
&\quad \ge (1+\varepsilon_n)L\|\psi\|_2 n^{1/2}h^{3/2} - C(2h)\\
&\quad = (2/3)^{1/2}(\log n)^{1/2}\varepsilon_n + o(1) \to \infty. \qquad \square
\end{aligned}$$

PROOF OF THEOREM 3.2. Let $h = h_n \in]0, R^{-1}/2]$ and $\delta = \delta_n > 0$ such that $\lim_{n\to\infty} h_n = \lim_{n\to\infty} \delta_n = 0$. Further let $J = J_n$ be any nonvoid subinterval of $[Rh, 1-Rh]$ and $J' = J'_n$ its Rh-neighborhood. For any $t \in J$, the probability $\mathbb{P}\{\widetilde{T}(Y) > \tilde{\kappa}_\alpha\}$ is not smaller than

$$\begin{aligned}
&\mathbb{P}_g\{\widehat{\Psi}(t,h) > C(2Rh) + \tilde{\kappa}_\alpha\}\\
&\quad = \mathbb{P}_0\{\widehat{\Psi}(t,h) + (n/h)^{1/2}\|\psi\|_2^{-1}\langle g, \psi_{t,h}\rangle > C(2Rh) + \tilde{\kappa}_\alpha\}\\
&\quad \ge \Phi((n/h)^{1/2}\|\psi\|_2^{-1}\langle g, \psi_{t,h}\rangle - C(2Rh) - \tilde{\kappa}_\alpha).
\end{aligned}$$

Now the question is how to choose $h = h_n$ and $\delta = \delta_n$ such that

$$(1+\varepsilon_n)\max_{t\in J}(n/h)^{1/2}\|\psi\|_2^{-1}\langle g, \psi_{t,h}\rangle - C(2Rh) \to \infty$$

uniformly for all $g \in \mathscr{F}_{J'}(k+1, L)$ with $\widetilde{\Delta}_J(g) \geq \delta$. For any such function g we pick some point $t \in J$ with $g^{(k)}(t) \geq \delta$. Then

$$\langle g, \psi_{t,h}\rangle = h\langle g(t+h\cdot), \psi\rangle = ha^{-1}\langle ag(t+h\cdot), \psi\rangle$$

for any $a > 0$. Note that $ag(t+h\cdot)$ belongs to the Hölder class $\mathscr{F}_{[-R,R]}(k+1, ah^{k+1}L)$ and

$$(ag(t+h\cdot))^{(k)}(0) = ah^k g^{(k)}(t) \geq \delta a h^k.$$

Specifically let $h := \delta/L$ and $a := L^k\delta^{-(k+1)}$, so that $ah^{k+1}L = \delta a h^k = 1$. Then, by (3.11),

$$\langle g, \psi_{t,h}\rangle \geq ha^{-1}\|\psi\|_2^2 = L^{-(k+1)}\delta^{k+2}\|\psi\|_2^2,$$

whence

$$\begin{aligned}(1+\varepsilon_n)&\max_{t\in J}(n/h)^{1/2}\|\psi\|_2^{-1}\langle g, \psi_{t,h}\rangle - C(2Rh)\\ &\geq (1+\varepsilon_n)L^{-(2k+1)/2}\|\psi\|_2 n^{1/2}\delta^{(2k+3)/2} - C(2R\delta/L).\end{aligned}$$

The right-hand side equals $\varepsilon_n(2/(2k+3))^{1/2}(\log n)^{1/2} + o(1)$ and tends to ∞, provided that $\delta = \delta_n(L)$ as stated in the theorem. □

Acknowledgments. The authors thank two reviewers and an associated editor for helpful comments and suggestions on the paper. The first author is grateful to the Weierstrass Institute, Berlin, for its hospitality and support.

REFERENCES

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

CHAUDHURI, P. and MARRON, J. S. (1999). SiZer for exploration of structures in curves. *J. Amer. Statist. Assoc.* **94** 807–823.

CHENG, M.-Y. and HALL, P. (1999). Mode testing in difficult cases. *Ann. Statist.* **27** 1294–1315.

DONOHO, D. (1994a). Statistical estimation and optimal recovery. *Ann. Statist.* **22** 238–270.

DONOHO, D. (1994b). Asymptotic minimax risk for sup-norm: solution via optimal recovery. *Probab. Theory Related Fields* **99** 145–170.

DONOHO, D. and LOW, M. (1992). Ronormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.* **20** 944–970.

DÜMBGEN, L. (1998). Application of local rank tests to nonparametric regression. Preprint.

DÜMBGEN, L. (2000). Optimal confidence bands for shape-restricted curves. Preprint.

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

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

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

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

GRAMA, I. and NUSSBAUM, M. (1998). Asymptotic equivalence for nonparametric generalized linear models. *Probab. Theory Related Fields* **111** 167–214.

HALL, P. and HECKMAN, N. E. (2000). Testing for monotonicity of a regression mean by calibrating for linear functions. *Ann. Statist.* **28** 20–39.

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

HARTIGAN, J. A. and HARTIGAN, P. M. (1985). The dip test of unimodality. *Ann. Statist.* **13** 70–84.

IBRAGIMOV, I. A. and KHASMINSKII, R. Z. (1980). Estimates of signal, its derivatives, and point of maximum for Gaussian observations. *Theory Probab. Appl.* **25** 703–716.

INGSTER, Y. I. (1986). Minimax testing of nonparametric hypothesis on a distribution density in L_p-metrics. *Theory Probab. Appl.* **32** 333–337.

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

KOUL, H. L. and SCHICK, A. (1997). Testing for the equality of two nonparametric regression curves. *J. Statist. Plann. Inference* **65** 293–314.

LEDWINA, T. (1994). Data-driven version of Neyman's smooth test of fit. *J. Amer. Statist. Assoc.* **89** 1000–1005.

LEDWINA, T. and KALLENBERG, W. C. M. (1995). Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests. *Ann. Statist.* **23** 1594–1608.

LEONOV, S. L. (1999). Remarks on extremal problems in nonparametric curve estimation. *Statist. Probab. Lett.* **43** 169–178.

LEPSKI, O. V. and TSYBAKOV, A. B. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probab. Theory Related Fields* **117** 17–48.

MAMMEN, E., MARRON, J. S. and FISHER, N. I. (1992). Some asymptotics for multimodality tests based on kernel density estimates. *Probab. Theory Related Fields* **91** 115–132.

MÜLLER, D. W. and SAWITZKI, G. (1991). Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.* **86** 738–746.

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

RICE, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12** 1215–1230.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

SILVERMAN, B. W. (1981). Using kernel density estimates to investigate multimodality. *J. Roy. Statist. Soc. Ser. B* **43** 97–99.

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

SPOKOINY, V. (1998). Adaptive and spatially adaptive testing a nonparametric hypothesis. *Math. Methods Statist.* **7** 245–273.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer, New York.

INSTITUT FÜR MATHEMATIK
MEDIZINISCHE UNIVERSITÄT ZU LÜBECK
WALLSTRASSE 40
LÜBECK D-23560
GERMANY
E-MAIL: duembgen@math.mu-luebeck.de

WEIERSTRASS INSTITUTE
MOHRENSTRASSE 39
BERLIN 10117
GERMANY
E-MAIL: spokoiny@wias-berlin.de