Dümbgen, Lutz; Spokoiny, Vladimir G.

Working Paper

Multiscale testing of qualitative hypotheses

SFB 373 Discussion Paper, No. 1999,99

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Dümbgen, Lutz; Spokoiny, Vladimir G. (1999) : Multiscale testing of qualitative hypotheses, SFB 373 Discussion Paper, No. 1999,99, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, https://nbn-resolving.de/urn:nbn:de:kobv:11-10046865

This Version is available at:
https://hdl.handle.net/10419/61743

Multiscale Testing of Qualitative Hypotheses

Lutz Dümbgen and Vladimir G. Spokoiny

Medizinische Universität zu Lübeck
and
Weierstraß-Institut für Angewandte Analysis und Stochastik, Berlin

October 1999

Abstract. 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.

Keywords and phrases. adaptivity, concavity, Lévy's modulus of continuity, monotonicity, multiple test, nonparametric, positivity.

AMS subject classification. 62G10; Secondary: 62G20.

The research was carried out within SFB 373 at Humboldt University, Berlin and was printed using funds made available by the Deutsche Forschungsgemeinschaft

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) use the so-called dip or excess mass functional. Further results for these procedures are given by Mammen et al. (1992) and Cheng and Hall (1999). But the available distribution theory relies on additional smoothness constraints on f. Statistical tests that are valid under the qualitative assumption only, have some optimality properties and are computationally feasible are yet unknown. There is another aspect of testing qualitative assumptions: 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.

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:

$$\begin{aligned} \mathcal{H}_{\leq 0} &:= \Big\{ f : f \leq 0 \Big\}, \\ \mathcal{H}_{\downarrow} &:= \Big\{ f : f \text{ is non-increasing} \Big\}, \\ \mathcal{H}_{\mathrm{conc}} &:= \Big\{ f : f \text{ is concave} \Big\}. \end{aligned}$$

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 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) containing many basic results and further references. Under the nonparametric approach it is typically assumed that f belongs to a certain class $\mathcal{F}$ of smooth functions, and its distance to the null hypothesis $\{0\}$ is quantified by some semi-norm $\|f\|$. For a given level $\alpha \in]0,1[$ and some number $\delta > 0$ the goal is to find a statistical test $\phi : \mathcal{C}(I) \to I$ whose minimal power

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

is as large as possible under the constraint $\mathbb{E}_0 \, \phi(Y) \leq \alpha$. 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 $\mathcal{F}$ 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 (1997) 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 (1996) treated Hölder smoothness classes and the supremum norm. The latter case is of special importance for us since the sup-norm seems to be the most suitable for describing the alternative set for the considered qualitative null hypothesis, see Section 3.2 for a discussion in terms of test signals. The tests of Lepski and Tsybakov (1996) are based on a kernel estimator of f with a kernel function and bandwidth depending on $\mathcal{F}$. It is a general problem that the available optimal tests ϕ depend explicitly on the class $\mathcal{F}$ and may fail if the latter is altered. With this problem in mind we review some results of Lepski and Tsybakov (1996) 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) choice of a smoothing parameter for testing a simple hypothesis, where deviation from the null hypothesis is measured by some integral norm, was considered in Ledwina and Kallenberg (1995), Fan (1996), Spokoiny (1996), Hart (1997) among others. The main message of Spokoiny (1996) is that the adaptive approach leads necessarily to suboptimal rates by a (log log)-factor. This issue differs drastically from what we have for the sup-norm: Adaptive testing is possible without any loss of efficiency!

In Section 3 we introduce tests for the three nonparametric hypotheses $\mathcal{H}_{\leq 0}$, $\mathcal{H}_{\downarrow}$ and $\mathcal{H}_{\mathrm{conc}}$. Given any of these composite hypotheses, say $\mathcal{H}_o$, we introduce two different func-

tionals $\Delta(f)$ measuring the distance of f to $\mathcal{H}_o$ and show how to maximize approximately

$$\inf_{g \in \mathcal{F}:\, \Delta(g) \geq \delta} \mathbb{E}_g\, \phi(Y)$$

over all tests ϕ satisfying

$$\sup_{f \in \mathcal{H}_o} \mathbb{E}_f\, \phi(Y) \;\leq\; \alpha.$$

Again the proposed tests are based on the multiscale idea as introduced in Section 2 and adaptive in a certain sense. Moreover, whenever the hypothesis $\mathcal{H}_o$ 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 Lévy's modulus of continuity which is of independent interest.

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

$$\langle g, g_o\rangle_{2,I} \;\geq\; \|g_o\|_{2,I}^2 \quad \text{for all } g \in \mathcal{G}.$$

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

For $\beta, L > 0$ the Hölder smoothness class $\mathcal{F}(\beta, L)$ is defined as follows: In case of $0 < \beta \leq 1$ let

$$\mathcal{F}(\beta, L) \ := \ \Big\{ f : |f(x) - f(y)| \leq L|x - y|^\beta \text{ for all } x, y \Big\}.$$

For $k < \beta \leq k + 1$ with an integer $k > 0$ let $\mathcal{F}(\beta, L)$ be the set of functions that are k times differentiable and whose k–th derivative belongs to $\mathcal{F}(\beta - k, L)$. Suppose that we want to test $\{0\}$ versus

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

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

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

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

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

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

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 case of $0 < \beta \leq 1$ one can easily show that

$$\psi(x) \ = \ 1\{|x| \leq 0\}(1 - |x|^\beta).$$

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

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

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

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

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

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

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

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

almost surely, 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 $\lim_{h\to 0} D(h) = 0$.

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

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

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

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

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*

$$\rho_n = \rho_n(\beta) \;:=\; \Big(\frac{\log n}{n}\Big)^{\beta/(2\beta+1)} \quad \text{and} \quad c_* = c_*(\beta, L) \;:=\; \Big(\frac{2L^{1/\beta}}{(2\beta+1)\|\psi\|_2^2}\Big)^{\beta/(2\beta+1)}.$$

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

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

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

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

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

Remark 1 The result of Theorem 2.2 can be treated in the following way. If a function f deviates from the null with a distance greater $L^{1/3}\rho_n$ then the test rejects the null with probability close to one. This deviation bound cannot be significantly improved in the sense that for every test ϕ_n of level α, there exists an alternative function g with deviation $(1-\epsilon_n)L^{1/3}\rho_n$ which will not be rejected with probability $1 - \alpha - o_n(1)$ or larger.

Adaptivity. Part (a) of Theorem 2.2 is a modification of Lepski and Tsybakov's (1996) lower bound. Part (b) is novel in that one test, $1\{T(\cdot) > \kappa_\alpha\}$, is asymptotically optimal for all Hölder smoothness classes $\mathcal{F}(\beta, L)$, $L > 0$. In other words, it is adaptive with respect to the second parameter of $\mathcal{F}(\beta, L)$.

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

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

Thus our test with the triangular kernel ψ is *rate optimal* over arbitrary Hölder classes with respect to 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

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

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

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

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 similarly as Lepski and Tsybakov (1996).

3 Testing the qualitative assumptions

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

3.1 Lipschitz alternatives and sup-norm distance

In this subsection let $\mathcal{H}_o$ be either $\mathcal{H}_{\le 0}$ or $\mathcal{H}_{\downarrow}$. We assume that under the alternative f belongs to the class $\mathcal{F}(1,L)$ for some unknown parameter $L>0$ and measure its distance to $\mathcal{H}_o$ by

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

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

$$\Delta_J(f) \;=\; \begin{cases} \sup\limits_{t\in J} f(t) & \text{if } \mathcal{H}_o=\mathcal{H}_{\le 0}, \\ \sup\limits_{s,t\in J\,:\,s<t} \dfrac{f(t)-f(s)}{2} & \text{if } \mathcal{H}_o=\mathcal{H}_{\downarrow}. \end{cases}$$

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

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

for some $h\in]0,1/2]$, where ψ is the triangular kernel given by $\psi(x) := 1\{|x|\le 1\}(1-|x|)$ with $\|\psi\|_2^2 = 2/3$. If we had one specific Lipschitz class $\mathcal{F}(1,L)$ in mind, it would be indeed sufficient to use the test statistic $\Delta_J(\widehat{f}_{n,h})$ 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

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

One can show that

$$T_o(Y) \;\le\; T_o(W) \quad \text{if } f \in \mathcal{H}_o \tag{3.1}$$

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

$$\kappa_{o,\alpha} \;:=\; \min\Big\{r \in \mathbf{R} : \mathbb{P}\{T_o(W) \le r\} \ge 1-\alpha\Big\}$$

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

Theorem 3.1 *Let*

$$\rho_n \;:=\; \Big(\frac{\log n}{n}\Big)^{1/3},$$

and let $(\epsilon_n)_{n\ge1}$ be as described in Theorem 2.2.

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

$$\limsup_{n\to\infty} \inf_{g\in\mathcal{F}(\beta,L)\,:\,\Delta_J(g)\ge(1-\epsilon_n)L^{1/3}\rho_n} \mathbb{E}_g\,\phi_n(Y) \;\le\; \alpha.$$

(b) *Let $J = J_n(L) := [L^{-2/3}\rho_n, 1 - L^{-2/3}\rho_n]$. Then*

$$\lim_{n\to\infty} \inf_{g\in\mathcal{F}(\beta,L)\,:\,\Delta_J(g)\ge(1+\epsilon_n)L^{1/3}\rho_n} \mathbb{P}_g\{T_o(Y) \ge \kappa_{o,\alpha}\} \;=\; 1.$$

3.2 Test signals and derivatives

In this subsection we consider the null hypotheses $\mathcal{H}_\downarrow$ and $\mathcal{H}_{\mathrm{conc}}$ and describe a second class of tests in terms of *test signals*. Let us first illustrate this approach for the hypothesis $\mathcal{H}_\downarrow$: Figure 1 shows a smooth function $g \notin \mathcal{H}_\downarrow$ together with the the unique function $f_o \in \mathcal{H}_\downarrow$ 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 $[-R,R]$, e.g. $\psi(x) = 1\{|x| \le 1\}\,x(1-|x|)$, 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 $0 < h \le 1/(2R)$ and $Rh \le t \le 1 - Rh$ might be an appropriate test statistic for the null hypothesis $\mathcal{H}_\downarrow$.

Figure 1 around here.

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

$$\langle f, \psi \rangle \ \le \ 0 \quad \text{for all } f \in \mathcal{H}_o. \tag{3.2}$$

Lemma 6.2 in Section 6 provides sufficient conditions for this requirement. Then we propose the test statistic

$$\widetilde{T}(Y) = \widetilde{T}(Y, \psi) \ := \sup_{h \in]0, 1/(2R)]} \ \sup_{t \in [Rh, 1-Rh]} \Big(\widehat{\Psi}(t,h) - C(2Rh) \Big) \tag{3.3}$$

which is just a onesided version of (2.4). Requirement (3.2) on ψ implies that

$$\widetilde{T}(Y) \ \le \ \widetilde{T}(W) \quad \text{whenever } f \in \mathcal{H}_o.$$

Equality holds if

$$\begin{cases} f \text{ is constant and } \mathcal{H}_o = \mathcal{H}_\downarrow, \\ f \text{ is linear and } \mathcal{H}_o = \mathcal{H}_{\mathrm{conc}}. \end{cases}$$

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

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

Multiple tests. Our method can be viewed as a multiple test procedure. Let $\widetilde{\mathcal{D}}_\alpha$ be the random family of intervals defined as

$$\tilde{D}_\alpha = \Big\{ [t - Rh, t + Rh] : Rh \le t \le 1 - Rh \text{ and } \widehat{\Psi}(t,h) > C(2Rh) + \widetilde{\kappa}_\alpha \Big\}.$$

Then $\widetilde{T}(Y) > \widetilde{\kappa}_\alpha$ if, and only if, $\widetilde{\mathcal{D}}_\alpha$ is nonempty. One may claim with confidence $1-\alpha$ that the unknown regression function f violates the qualitative assumption, i.e. antitonicity or concavity, on every interval $J \in \widetilde{\mathcal{D}}_\alpha$. Consequently, when the null hypothesis $\mathcal{H}_o$ is rejected, we do 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 (3.2), note that for smooth functions g antitonicity is equivalent to $g^{(1)} \le 0$ while concavity is

equivalent to $g^{(2)} \le 0$. Here $g^{(j)}$ denotes the j–th derivative of g. Now we want to find an optimal test signal ψ for testing $\mathcal{H}_o$ versus all alternatives of the form $\Big\{g \in \mathcal{F}(k+1, L) : \widetilde{\Delta}_J(g) \ge \delta\Big\}$, where

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

This leads to the following optimization problem:

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

Note that the set $\Big\{g \in \mathcal{F}(k+1, 1) : g^{(k)}(0) \ge 1\Big\}$ is convex, while $\mathcal{H}_o$ is even a convex cone. Thus a pair (g_o, f_o) solves problem (3.4) if, and only if, the difference

$$\psi \;:=\; g_o - f_o$$

satisfies

$$\langle f, \psi \rangle \;\le\; \langle f_o, \psi \rangle \;=\; 0 \quad \text{for all } f \in \mathcal{H}_o \tag{3.5}$$

and

$$\langle g, \psi \rangle \;\ge\; \|\psi\|_2^2 \quad \text{for all } g \in \mathcal{F}(k+1, 1) \text{ with } g^{(k)}(0) \ge 1. \tag{3.6}$$

These inequalities imply that $\|(g - f) - \psi\|_2^2 \le \|g - f\|_2^2 - \|\psi\|_2^2$ for any pair (g, f) as in (3.4). Therefore the difference ψ is unique and satisfies (3.2).

Lemma 3.2 **(a)** *In case of $\mathcal{H}_o = \mathcal{H}_{\downarrow}$ a solution (g_o, f_o) of problem (3.4) is given by*

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

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

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

(b) *In case of $\mathcal{H}_o = \mathcal{H}_{\text{conc}}$ a solution (g_o, f_o) of problem (3.4) is given by*

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

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

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

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

Figure 2 around here.

Theorem 3.3 *Let $\widetilde{T}$ be defined with $\psi = g_o - f_o$, where (g_o, f_o) solves the optimization problem (3.4). For $L > 0$ let*

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

With $J = J_n(L) = [Rc_*\rho_n/L, 1 - Rc_*\rho_n/L]$,

$$\lim_{n\to\infty} \inf_{g\in\mathcal{F}(k+1,L)\,:\,\widetilde{\Delta}_J(g)\geq(1+\epsilon_n)c_*\rho_n} \mathbb{P}_g\{\widetilde{T}(Y) > \widetilde{\kappa}_\alpha\} \;=\; 1,$$

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

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

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

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

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

Therefore our test identifies pairs (t,h) such that $\widehat{f}_h^{(k)}(t)$ is significantly greater than zero. This shows that our methods are related and have potential applications to Chaudhuri and Marron's (1997) method. Translated into the present set-up, the latter authors use test statistics such as

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

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

Rate optimality. The rate ρ_n shown in Theorem 3.3 coincides with the optimal rates for estimating the k–th derivative of functions in $\mathcal{F}(k+1,L)$ in the sup-norm, see Ibragimov and Khasminskii (1980). Moreover, our optimization problem (3.4) is closely

related (but does not coincide) with the optimal recovery problem from Donoho (1994a, 1994b) arising in estimation of the function and its derivatives in sup-norm. Similarly to the estimation problem, the case of a smoothness degree which differs from $k+1$ would require different test signals. At the same time, one can easily see that the application of the proposed test leads to the optimal rate of testing for an arbitrary Hölder class $\mathcal{F}(\beta, L)$ with $\beta > k$.

4 Modifications and further developments

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

$$Y_i \;=\; f(x_i) + \epsilon_i \quad \text{for } i = 1, 2, \ldots, n, \tag{4.1}$$

where $x_i := (i - 1/2)/n$, and the random errors ϵ_i are independent with Gaussian distribution $\mathcal{N}(0, \sigma^2)$. One can show that Theorems 2.2, 3.1 and 3.3 remain valid with σc_* in place of c_*, provided that we replace $\widehat{\Psi}(t,h)$ with

$$\widehat{\Psi}_n(t,h) \;:=\; \sigma^{-1}\Big(\sum_{i=1}^n \psi_{t,h}(x_i)^2\Big)^{-1/2} \sum_{i=1}^n \psi_{t,h}(x_i) Y_i. \tag{4.2}$$

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

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

$$\Big|\widehat{\sigma}_n/\sigma - 1\Big| \;=\; o_p\Big((\log n)^{1/2}\Big). \tag{4.3}$$

For instance, if $\widehat{\sigma}_n^2$ equals

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

see Rice (1984) for the first and Gasser et al. (1986) for the second proposal, then (4.3) holds whenever f has bounded total variation $\mathrm{TV}(f)$. Indeed, elementary calculations show that

$$\mathbb{E}((\widehat{\sigma}_n^2/\sigma^2 - 1)^2) \;\le\; \text{const.}(1 + \mathrm{TV}(f)^2)/n$$

for all $n \ge 3$.

General regression models. If one observes $Y_i = f(x_i) + E_i$ for $i = 1, 2, \ldots, n$ with arbitrary fixed numbers x_i and independent, identically distributed random errors E_i, one can modify the multiscale tests of $\mathcal{H}_\downarrow$ in Section 3.2 using linear rank statistics instead of linear statistics such as $\sum_i \psi_{t,h}(i/n)Y_i$; see Dümbgen (1998). There the aspect of localizing interesting features such as modes is discussed in more detail.

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 (1999).

5 Numerical examples

In this section we illustrate the tests of Section 3.2 for $\mathcal{H}_\downarrow$ and $\mathcal{H}_{\rm conc}$ within the Gaussian regression model (4.1) with sample size $n = 700$ and standard deviation $\sigma = 1$. For notational convenience the test signals $\psi_\downarrow$ and $\psi_{\rm conc}$ are rescaled to have support $[-1, 1]$, namely $\psi_\downarrow(x) := 1\{|x| \le 1\}\, x(1 - |x|)$ and $\psi_{\rm conc} := 1\{|x| \le 1\}(-1/8 + 9x^2/8 - |x|^3)$.

As for $\mathcal{H}_\downarrow$, Figure 3 shows a Monte Carlo estimator of the distribution function of

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

in case of $f \equiv 0$. Here S_n denotes $\{d/n : d = 1, 2, \ldots, \lfloor n/2 \rfloor\}$, $L_n(d/n)$ stands for $\{j/n : j = d, d+1, \ldots, n-d\}$, and $\widehat{\Psi}_n(t, h)$ is the linear filter defined in (4.2) with the test signal $\psi_\downarrow$. The Monte-Carlo estimator is based upon 9999 simulations, and the vertical lines indicate the (estimated) critical values $\widetilde{\kappa}_{n,\alpha}$ for $\alpha \in \{0.50, 0.10, 0.05\}$.

Figure 3 around here.

Figure 4 shows four realizations of the random function $\widetilde{T}_n(\vec{Y}, \cdot)$ on S_n, again in case of $f \equiv 0$. The lower dashed line depicts the additive correction term $-C(2h)$, while the upper horizontal line shows the critical value $\widetilde{\kappa}_{n,0.95} = 2.018$.

Figure 4 around here.

The process $\widetilde{T}_n(\vec{Y}, \cdot)$ behaves differently if, for example, f is the function depicted in Figure 1. Figure 5 shows observations Y_i together with this regression function f (left plot) and the corresponding process $\widetilde{T}_n(\vec{Y}, \cdot)$ (right plot).

Figure 5 around here.

We see that the critical value $\widetilde{\kappa}$ is exceeded for bandwidths h in two disjoint regions. For two of these bandwidths Figure 6 shows the process $\widehat{\Psi}_n(t,\cdot)$ on $L_n(h)$ (upper row). In addition, for both bandwidths a location parameter t with $\widehat{\Psi}_n(t,h) > C(2h) + \widetilde{\kappa}_{n,0.05}$ was picked. Each plot in the lower row shows the data vector $\vec{Y}$ together with its orthogonal projection onto the linear span of

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

Note that the smaller bandwidth is appropriate for detecting and localizing a sharp increasing trend of f over a small interval on the right hand side, while the larger bandwidth enables us to find a moderate increasing trend over a larger interval on the left hand side. These pictures illustrate the benefits of using several bandwidths simultaneously.

Figure 6 around here.

Now we show analogous plots for a function $f \notin \mathcal{H}_{\mathrm{conc}}$ and the multiscale statistic $\widetilde{T}_n$ based on the test signal ψ_{conc}. More precisely, we define $\widehat{\Psi}_n(t,h)$ as in (4.2) with

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

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 $g \in \mathcal{H}_{\mathrm{conc}}$, a consequence of Lemma 6.2. Figure 7 shows simulated data and the process $\widetilde{T}_n(\vec{Y},\cdot)$. Figure 8 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

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

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 14.1.1). For if we apply Theorem 2.1 to $\psi(x) := 1\{|x| \le 1\}$, then

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

so that

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

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

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

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

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

Moreover,

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

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

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

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

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

Then the random variable

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

is finite almost surely. More precisely, $\mathbb{P}\{S(X) > r\} \leq p(r)$ 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, $X(a)^2/\sigma(a)^2 \leq 2V \log(1/\sigma(a)^2) + S(X) \log\log(e^e/\sigma(a)^2)$ for arbitrary $a \in \mathcal{T}$. Since $(x + y)^{1/2} \leq x^{1/2} + x^{-1/2} y/2$ for arbitrary positive numbers x and y, Theorem 6.1 implies that

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

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

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

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

$$\mathbb{P}\Big\{\omega(X, v) > \eta\Big\} \ \le \ C \exp\Big(-\frac{\eta^2}{Cv^2 \log(e/v)}\Big) \quad \text{for } 0 < \epsilon \le 1, \eta > 0. \tag{6.1}$$

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

$$\sigma(a)^2 \ \ge \ \sigma(\widetilde{a})^2 - u\delta \ \ge \ \sigma(\widetilde{a})^2(1 - 2u)$$

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

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

implies that either

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

or

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

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

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

according to assumptions (i) and (iii) and inequality (6.1). 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

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

Finally let

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

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

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

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

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

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

$$\begin{aligned}
\rho\Big((t,h),(t',h')\Big)^2 &:= \mathrm{Leb}\Big([t-h,t+h] \triangle [t'-h',t'+h']\Big), \\
\sigma(t,h)^2 &:= \mathrm{Leb}([t-h,t+h]) = 2h.
\end{aligned}$$

Then $\sigma(b)^2 \leq \sigma(a)^2 + \rho(a,b)^2$ for all $a, b \in \mathcal{T}$, 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 $\mathcal{T}$ with $\mathrm{Var}(X(t,h)) = \sigma(t,h)^2$. It suffices to show that this process X and the triple $(\mathcal{T}, \rho, \sigma)$ 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,

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

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

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

For let $\mathcal{T}''$ be any maximal subset of $\{a \in \mathcal{T} : \sigma(a)^2 \leq \delta\}$ 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 $0 < 2h \leq \delta$ imply that

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

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 $(1 + 2/(u\delta))(2 + 2/u) \leq 12u^{-2}\delta^{-1}$. □

6.2 Basic properties of the tests and test signals

Proof of inequality (3.1). 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:

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

The assertion about $\mathcal{H}_{\le 0}$ is obvious, because $\psi_{t,h} \ge 0$. The assertion about $\mathcal{H}_{\downarrow}$ follows from Lemma 6.2 below, because $\int (\psi_{u,h} - \psi_{s,h})(x)\,dx = 0$ and

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

Lemma 6.2 *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*

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

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

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

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

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

Proof of Lemma 6.2. As for part (a), let $f \in \mathcal{H}_{\downarrow}$ and $\widetilde{f} := f - f(a)$. Then $\widetilde{f} \in \mathcal{H}_{\downarrow}$, and our assumptions on ψ imply that $\int \psi f\,d\mu = \int \psi \widetilde{f}\,d\mu \le 0$, because $\psi \widetilde{f} \le 0$.

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

$$\widetilde{f}(x) \;:=\; f(x) - \frac{x-b}{c-b}\,(f(c) - f(b)).$$

If $f \in \mathcal{H}_{\mathrm{conc}}$, then $\widetilde{f}$ belongs to $\mathcal{H}_{\mathrm{conc}}$, too, and $\int \psi f\,d\mu = \int \psi \widetilde{f}\,d\mu \le 0$, because $\psi \widetilde{f} \le 0$. $\square$

Proof of Lemma 3.2. The functions $\psi_{\downarrow}$ and ψ_{conc} are constructed such that they satisfy the conditions of Lemma 6.2 (a) and (b), respectively, where μ is Lebesgue measure on the line. Moreover, in both cases, $f_o \psi \equiv 0$. Thus condition (3.5) is satisfied. It remains to verify condition (3.6).

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

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

so that

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

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

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

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

$$\begin{aligned} \widetilde{g} \;&\leq\; \psi_{\mathrm{conc}} \;\leq\; 0 \quad \text{on } [0,a], \\ \widetilde{g} \;&\geq\; \psi_{\mathrm{conc}} \;\geq\; 0 \quad \text{on } [a,b]. \end{aligned}$$

Consequently

$$\langle \widetilde{g}, \psi_{\mathrm{conc}} \rangle \;=\; 2\langle \widetilde{g}, \psi_{\mathrm{conc}} \rangle_{[0,b]} \;\geq\; 2\|\psi_{\mathrm{conc}}\|^2_{2,[0,b]} \;=\; \|\psi_{\mathrm{conc}}\|_2^2. \qquad \square$$

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 1996).

Lemma 6.3 *Let $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ be independent random variables with standard Gaussian distribution. Then*

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

provided that $w_m = (2\log m)^{1/2}(1-\epsilon_m)$ with $\lim_{m\to\infty} \epsilon_m = 0$ and $\lim_{m\to\infty} (\log m)^{1/2} \epsilon_m = \infty$.

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

Proof of Lemma 6.3. 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

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

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

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

for any $\delta > 0$. If $\delta = (\log\eta + 2\epsilon_m \log m)/w_m^2$, then the latter bound equals

$$\exp\Big(-\frac{(\log\eta + 2\epsilon_m \log m)^2}{2w_m^2}\Big) \;=\; \exp\Big(-\epsilon_m^2 \log m + o(1)\Big) \;\to\; 0 \quad \text{as } m \to \infty. \qquad \square$$

Proof of Theorem 3.1 (a). Let ψ be the triangular kernel with $\psi(x) = 1\{|x| \leq 1\}(1 - |x|)$. 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 $\mathcal{F}(1, L)$. Now let $[a, a + 2b] \subset J \subset \,]0, 1[$ for some $b > 0$. For $\ell = 1, 2$ define

$$\mathcal{J}_\ell \;:=\; \Big\{\text{integers } j : (2j-1)h \in \,]a + (\ell-1)b, a + \ell b]\Big\}.$$

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

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

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 $\mathbb{E}_0\, \phi(Y) \leq \alpha$,

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

Now we want to determine $h = h_n$ such that the right hand side tends to zero 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 $n^{1/2} L h^{3/2} \|\psi\|_2 \widehat{\Psi}((2j-1)h, h)$. With $\Gamma_i := (-1)^\ell \widehat{\Psi}((2i-1)h, h)$ for $i \in \mathcal{J}_\ell$, 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 $w := n^{1/2} L h^{3/2} \|\psi\|_2$ and the random variable $Z_i := \exp(w\Gamma_i - w^2/2)$, then one can write

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

for $j \in \mathcal{J}_1, k \in \mathcal{J}_2$. Consequently, $(\#\mathcal{G}_o)^{-1} \sum_{g \in \mathcal{G}_o} (d\mathbb{P}_g / d\mathbb{P}_0)(Y) - 1$ equals

$$\begin{cases} S_1 & \text{if } \mathcal{H}_o = \mathcal{H}_{\leq 0}, \\ S_1 S_2 + S_1 + S_2 & \text{if } \mathcal{H}_o = \mathcal{H}_{\downarrow}, \end{cases}$$

where $S_\ell := (\#\mathcal{J}_\ell)^{-1} \sum_{i \in \mathcal{J}_\ell} Z_i - 1$. Therefore, since S_1 and S_2 are independent, the left hand side of (6.2) tends to zero if

$$\mathbb{E}_0 |S_\ell| \;\to\; 0 \quad \text{for } \ell = 1, 2.$$

According to Lemma 6.3, the latter condition holds as $n \to \infty$, provided that $h_n \to 0$ and the corresponding $w = w_n$ satisfies

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

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

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

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

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

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

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

$$\inf_{g\in\mathcal{F}(\beta,L):\,\|g\|_J\geq(1+\epsilon_n)\delta}\ \max_{t\in J}(n/h)^{1/2}\|\psi\|_2^{-1}|\langle g,\psi_{t,h}\rangle| - C(2h)\ \to\ \infty.$$

By construction of ψ and definition of h, the function $\delta\psi_{t,h}$ belongs to $\mathcal{F}(\beta,L)$, and for $g\in\mathcal{F}(\beta,L)$ with $|g(t)|\geq\delta$,

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

Thus

$$\begin{aligned}
&\inf_{g\in\mathcal{F}(\beta,L):\,\|g\|_J\geq(1+\epsilon_n)\delta}\ \max_{t\in J}(n/h)^{1/2}\|\psi\|_2^{-1}|\langle g,\psi_{t,h}\rangle| - C(2h)\\
&\quad\geq\ (1+\epsilon_n)\inf_{g\in\mathcal{F}(\beta,L):\,\|g\|_J\geq\delta}\ \max_{t\in J}(n/h)^{1/2}\|\psi\|_2^{-1}|\langle g,\psi_{t,h}\rangle| - C(2h)\\
&\quad\geq\ (1+\epsilon_n)\|\psi\|_2 n^{1/2}h^{1/2}\delta - C(2h)\\
&\quad=\ \epsilon_n(2/(2\beta+1))^{1/2}(\log n)^{1/2} + o(1)\ \to\ \infty. \qquad \square
\end{aligned}$$

Proof of Theorem 3.1 (b). In case of $\mathcal{H}_o=\mathcal{H}_{\leq 0}$ the proof is almost identical to the proof of Theorem 2.2 (b). Thus we focus on $\mathcal{H}_o=\mathcal{H}_{\downarrow}$. Let $\delta=\delta_n=L^{1/3}\rho_n$, $h=h_n=\delta/L$ and $J=[h,1-h]$. For $h\leq s<t\leq 1-h$, the probability $\mathbb{P}_g\{T_o(Y)>\kappa_\alpha\}$ is not greater than

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

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 one. Thus it suffices to show that

$$\inf_{g\in\mathcal{F}(1,L):\,\Delta_J(g)\geq(1+\epsilon_n)\delta}\ \max_{h\leq s<t\leq 1-h}(n/h)^{1/2}\|\psi\|_2^{-1}\langle g,\psi_{t,h}-\psi_{s,h}\rangle/2 - C(2h)\ \to\ \infty.$$

For $g\in\mathcal{F}(1,L)$ with $\Delta_J(g)\geq\delta$ let $h\leq s<t\leq 1-h$ such that $g(t)-g(s)\geq 2\delta$. Letting $\gamma:=(g(s)+g(t))/2$,

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

Thus

$$\begin{aligned}
&\inf_{g\in\mathcal{F}(1,L):\,\Delta_J(g)\geq(1+\epsilon_n)\delta}\ \max_{h\leq s<t\leq 1-h}(n/h)^{1/2}\|\psi\|_2^{-1}\langle g,\psi_{t,h}-\psi_{s,h}\rangle/2-C(2h)\\
&\quad\geq\ (1+\epsilon_n)\inf_{g\in\mathcal{F}(1,L):\,\Delta_J(g)\geq\delta}\ \max_{h\leq s<t\leq 1-h}(n/h)^{1/2}\|\psi\|_2^{-1}\langle g,\psi_{t,h}-\psi_{s,h}\rangle/2-C(2h)\\
&\quad\geq\ (1+\epsilon_n)L\|\psi\|_2 n^{1/2}h^{3/2}-C(2h)\\
&\quad=\ (2/3)^{1/2}(\log n)^{1/2}\epsilon_n+o(1)\ \to\ \infty. \qquad \square
\end{aligned}$$

Proof of Theorem 3.3. Let $h=h_n\in\,]0,1/(2R)]$ and $\delta=\delta_n>0$ such that $\lim_{n\to\infty}h_n=\lim_{n\to\infty}\delta_n=0$, and set $J=[Rh,1-Rh]$. For any $t\in J$, the probability $\mathbb{P}_g\{\widetilde{T}(Y)>\widetilde{\kappa}_\alpha\}$ is not smaller than

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

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

$$\inf_{g\in\mathcal{F}(k+1,L):\,\widetilde{\Delta}_J(g)\geq(1+\epsilon_n)\delta}\ \max_{t\in J}(n/h)^{1/2}\|\psi\|_2^{-1}\langle g,\psi_{t,h}\rangle-C(2Rh)\ \to\ \infty.$$

If $g\in\mathcal{F}(k+1,L)$ and $t\in J$ 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

$$\begin{aligned}
ag(t+h\cdot)\ &\in\ \mathcal{F}(k+1,ah^{k+1}L),\\
(ag(t+h\cdot))^{(k)}(0)\ &=\ ah^kg^{(k)}(t)\ \geq\ \delta ah^k.
\end{aligned}$$

If $h:=\delta/L$ and $a:=L^k\delta^{-(k+1)}$, then $ah^{k+1}L=\delta ah^k=1$. Consequently, by (3.2)

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

so that

$$\begin{aligned}
&\inf_{g\in\mathcal{F}(k+1,L):\,\widetilde{\Delta}_J(g)\geq(1+\epsilon)\delta}\ \max_{t\in J}(n/h)^{1/2}\|\psi\|_2^{-1}\langle g,\psi_{t,h}\rangle-C(2Rh)\\
&\quad\geq\ (1+\epsilon_n)\inf_{g\in\mathcal{F}(k+1,L):\,\widetilde{\Delta}_J(g)\geq\delta}\ \max_{t\in J}(n/h)^{1/2}\|\psi\|_2^{-1}\langle g,\psi_{t,h}\rangle-C(2Rh)\\
&\quad\geq\ (1+\epsilon_n)L^{-(2k+1)/2}\|\psi\|_2 n^{1/2}\delta^{(2k+3)/2}-C(2R\delta/L).
\end{aligned}$$

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

Acknowledgement. The first author is grateful to the Weierstraß Institute, Berlin, for its hospitality and support.

References

[1] Brown, L.D. and M.G. Low (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24**, 2384-2398.

[2] Chaudhuri, P. and J.S. Marron (1997). SiZer for exploration of structures in curves. Preprint

[3] Cheng, M.-Y. and P. Hall (1999). Mode testing in difficult cases. to appear in *Ann. Statist.*

[4] Donoho, D. and M. Low (1992). Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.* **20**, 944-970

[5] Donoho, D. (1994a). Statistical estimation and optimal recovery. *Ann. Statist.* **22** 238–270.

[6] Donoho, D. (1994b). Asymptotic minimax risk for sup-norm: Solution via optimal recovery. *Probab. Theory and Rel. Fields* **99** 145–170.

[7] Dümbgen, L. (1998). Application of local rank tests to nonparametric regression. Preprint

[8] Dümbgen, L. (1999). Optimal confidence bands for shape-restricted curves. Preprint

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

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

[11] Gasser, T. and L. Sroka and C. Jennen-Steinmetz (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73** 625–633.

[12] Grama, I. and M. Nussbaum (1998). Asymptotic equivalence for nonparametric generalized linear models. *Prob. Theory and Related Fields.* **111** No. 2, 167–214.

[13] Hart, J. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests* New York, Berlin, Heidelberg: Springer.

[14] Hartigan, J.A. and P.M. Hartigan (1985). The dip test of unimodality. *Ann. Statist.* **13**, 70-84.

[15] Ibragimov, I.A. and R.Z. Khasminskii (1980). Estimates of signal, its derivatives, and point of maximum for Gaussian observations. *Theory Probab. Appl.* **25** No.4, 703-716.

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

[17] Ingster, Y.I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives, I-III. *Math. Methods Statist.* **2**; 85-114, 171-189, 249-268.

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

[19] LEONOV, S.L. (1999). Remarks on extremal problems in nonparametric curve estimation. *Stat. Prob. Letters* **43** No. 2, 169–178.

[20] LEPSKI, O.V. AND A.B. TSYBAKOV (1996). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. Preprint. Humboldt University, Berlin.

[21] MAMMEN, E., J.S. MARRON AND N.I. FISHER (1992). Some asymptotics for multimodality tests based on kernel density estimates. *Prob. Theory Rel. Fields* **91**, 115-132.

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

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

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

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

[26] SILVERMAN, B.W. (1981) Using kernel density estimates to investigate multimodality. *J. Roy. Statist. Soc. B* **43**, 97-99.

[27] SPOKOINY, V. (1996) Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24**, 2477-2498.

[28] SPOKOINY, V. (1998) Adaptive and spatially adaptive testing a nonparametric hypothesis. *Math. Methods of Statistics*, **7**, no. 3, 245–273.

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

Lutz Dümbgen
Institut für Mathematik
Medizinische Universität zu Lübeck
Wallstraße 40
D-23560 Lübeck
Germany
E-mail: duembgen@math.mu-luebeck.de
URL: http://www.math.mu-luebeck.de/

Vladimir G. Spokoiny
Weierstraß-Institut für Angewandte Analysis und Stochastik
Mohrenstraße 39
D-10117 Berlin
Germany
E-mail: spokoiny@wias-berlin.de

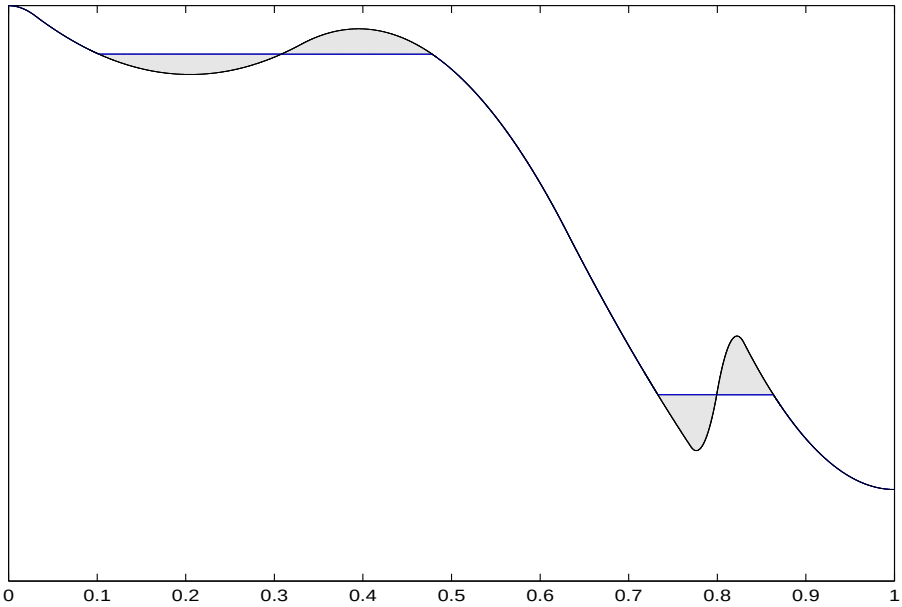

Figure 1: A function $g \notin \mathcal{H}_{\downarrow}$ and its projection f_o onto $\mathcal{H}_{\downarrow}$

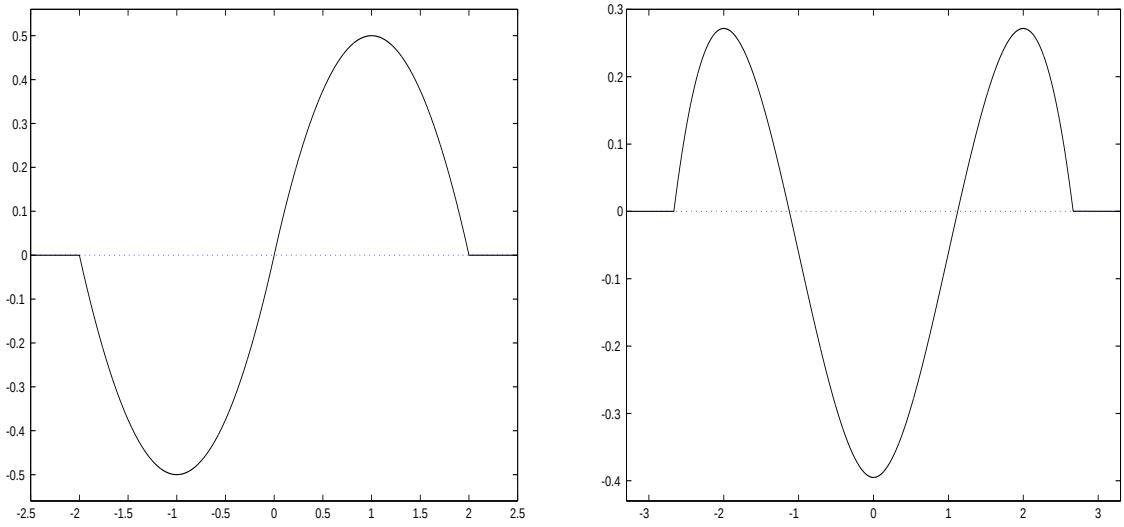

Figure 2: The test signals $\psi_{\downarrow}$ and ψ_{conc}

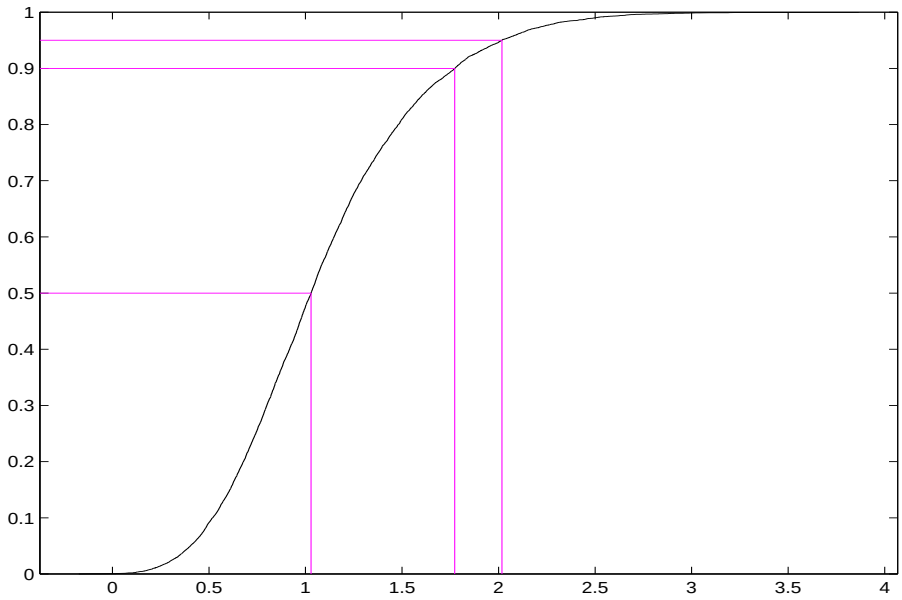

Figure 3: Estimated distribution function of $\widetilde{T}_n(\vec{Y})$ under $f \equiv 0$.

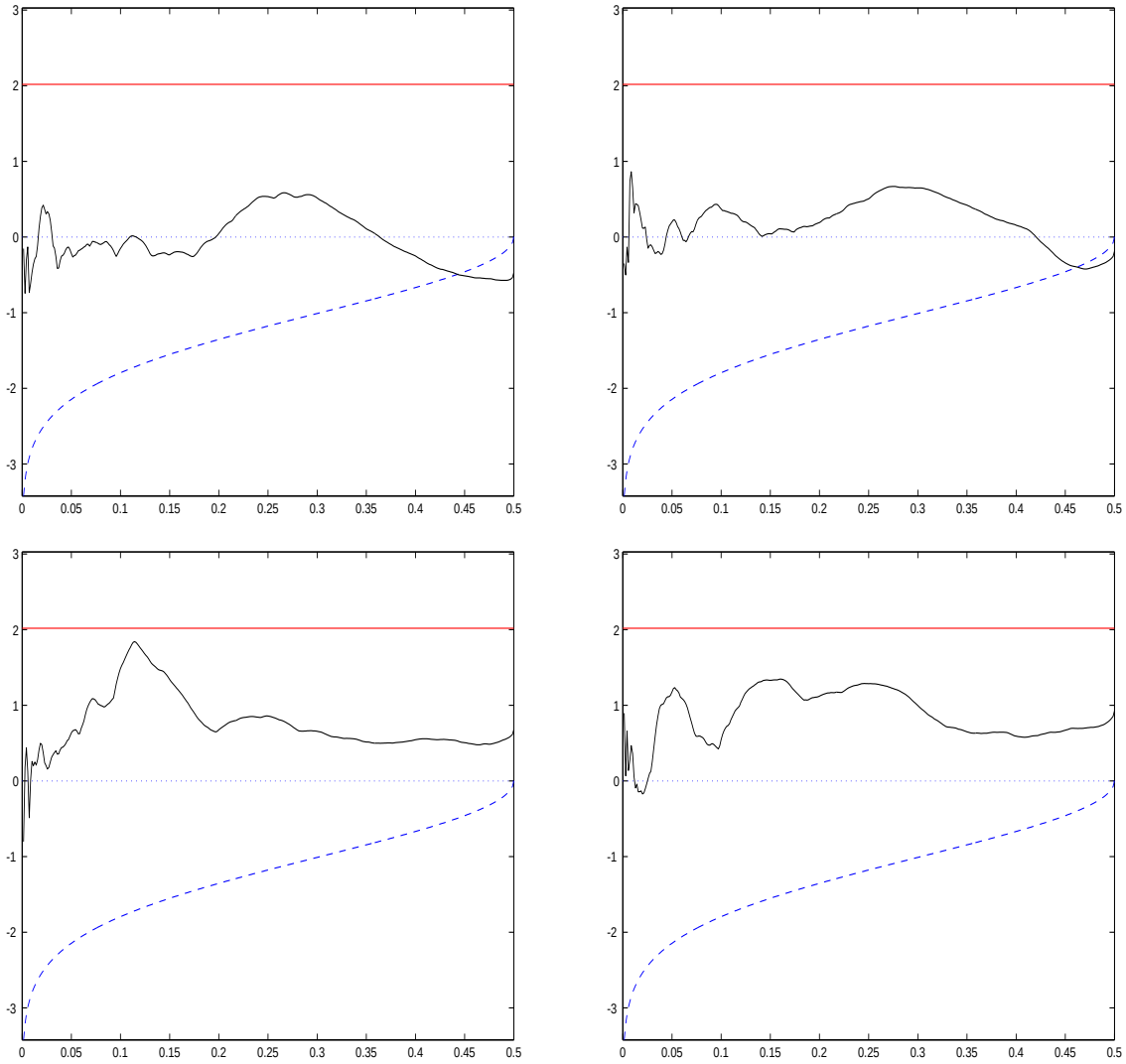

Figure 4: Four realizations of $\widetilde{T}_n(\vec{Y}, \cdot)$ in case of $f \equiv 0$

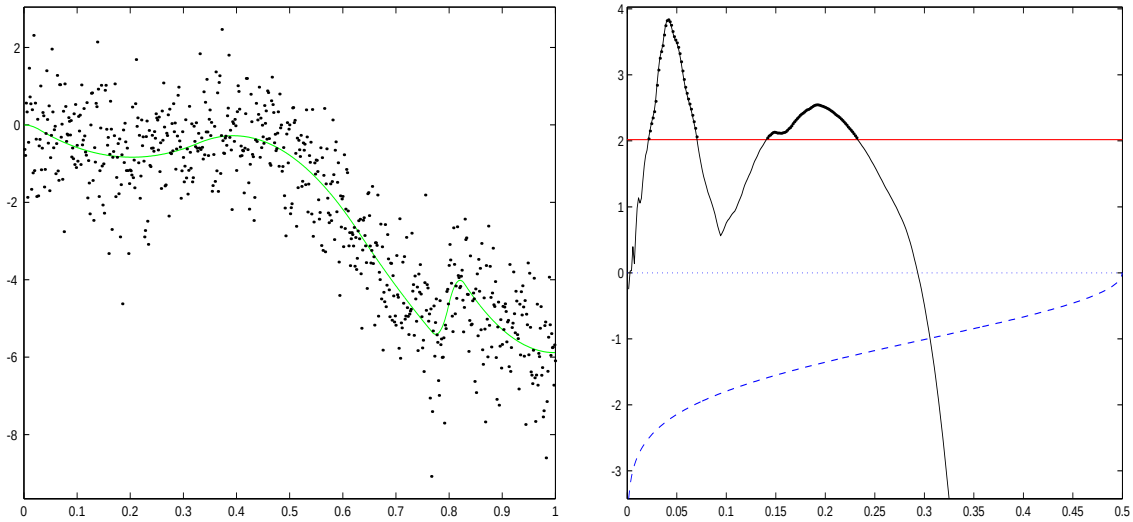

Figure 5: Simulated data with $f \notin \mathcal{H}_{\downarrow}$ and $\widetilde{T}_n(\vec{Y}, \cdot)$

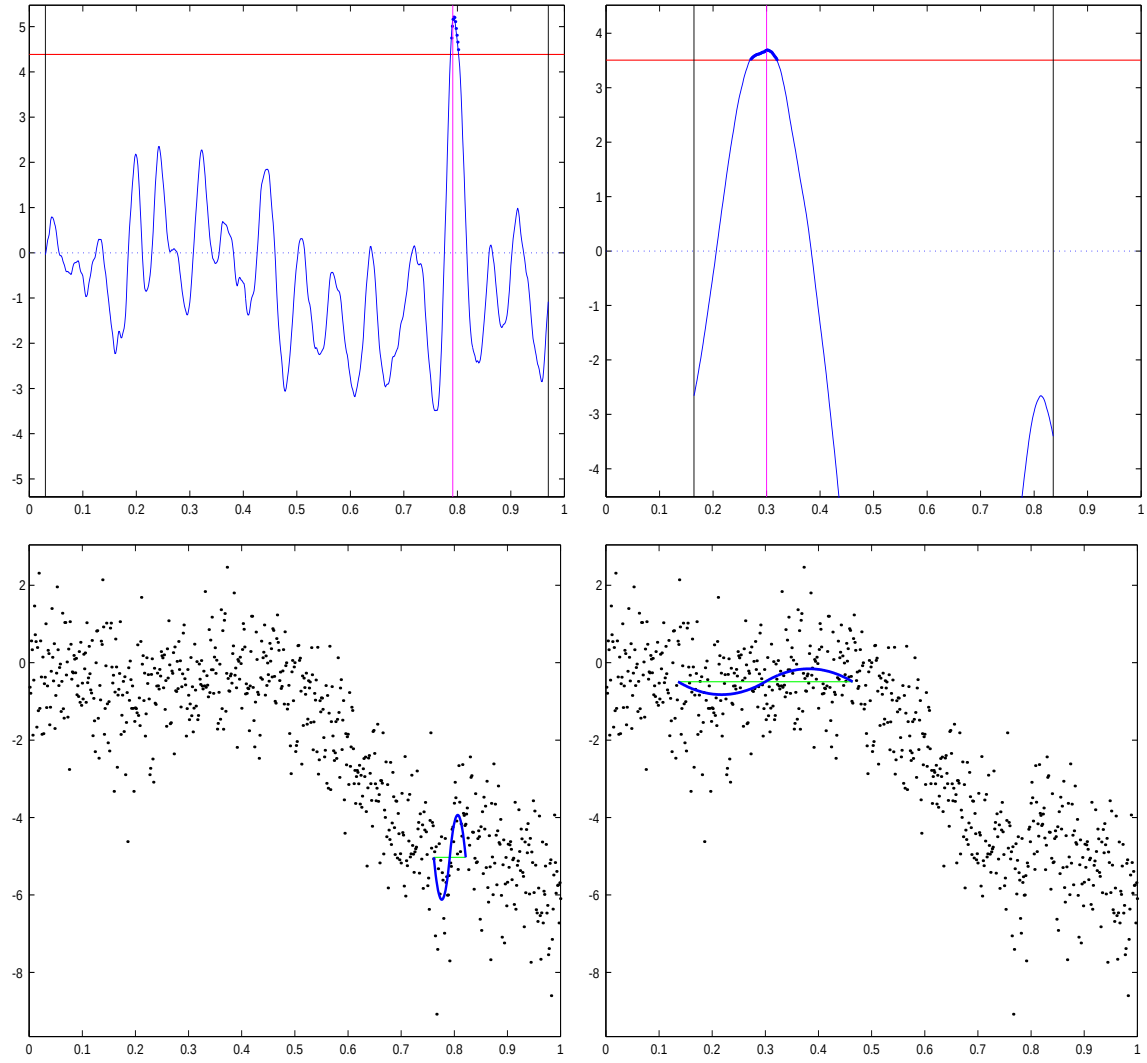

Figure 6: The process $\widehat{\Psi}_n(\cdot, h)$ for $h = 0.029$ and $h = 0.170$

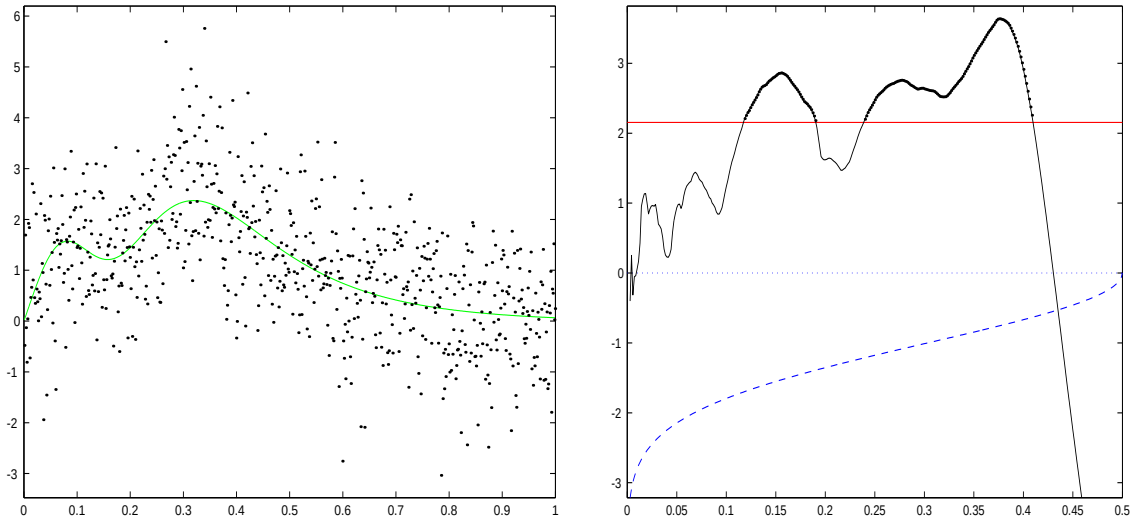

Figure 7: Simulated data with $f \notin \mathcal{H}_{\text{conc}}$ and $\widetilde{T}_n(\vec{Y}, \cdot)$

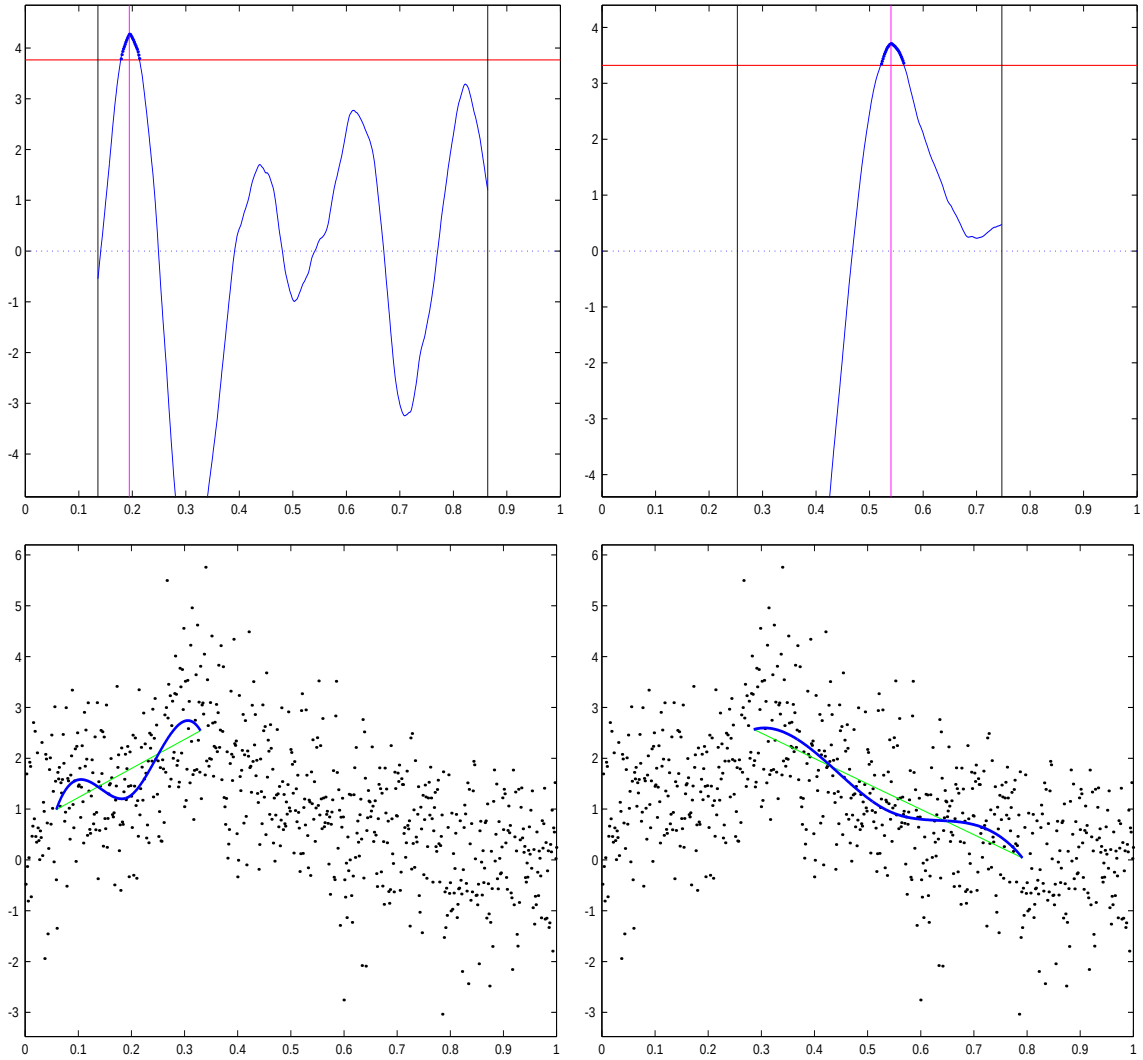

Figure 8: The processes $\widehat{\Psi}_n(\cdot, h)$ for $h = 0.136$ and $h = 0.253$