On Testing Positivity/Monotonicity/Convexity of nonparametric Signals

Anatoli Iouditski and Arkadi Nemirovski

A.Iouditski@inrialpes.fr
INRIA Rhone-Alpes
655 avenue de l'Europe
38330 Montbonnot St. Matrin, France

Nemirovs@ie.technion.ac.il
Faculty of Industrial Engineering and Management
Technion – Israel Institute of Technology
Technion City, Haifa 32000, Israel

September 10, 1999

1 Introduction

The problem. The problem addressed in this paper is as follows: let a nonparametric signal $f : [0,1] \to \mathbf{R}$ be observed according to the standard "signal + white noise" model, so that the observation is a realization of the Gaussian random process $X_n^f(\cdot)$ on $[0,1]$:

$$dX_n^f = f(t)dt + n^{-1/2}dW(t),\ 0 \le t \le 1, \quad X_n^f(0) = 0; \tag{1}$$

here $W(\cdot)$ is the standard Wiener process and the parameter n plays the role of "volume of observations". We know that f is regular (belongs to a given Hölder ball Σ, see below) and are interested

(I) either to estimate the distance

$$\Phi_r[f] = \inf\{\| f - g \|_r |\ g \in \mathcal{M}\}$$

from f to a given cone $\mathcal{M}$ in the space of functions on $[0,1]$; here r, $1 \le r < \infty$, is fixed, and $\| \cdot \|_r$ is the standard L_r-norm on the unit segment. Basically, we are interested in the particular cases when $\mathcal{M}$ is the cone of (i) nonpositive, or (ii) monotone, or (iii) concave functions,

(II) or to decide whether or not the signal belongs to the cone $\mathcal{M}$, more precisely, to distinguish between the hypotheses

H_0: $f \in \mathcal{M}$

and

H_ϵ: $\Phi_r[f] \ge \epsilon$.

We quantify our abilities to handle **(I)** and **(II)** by the minimax risks defined as follows.

• For **(I)**, the minimax risk is

$$\mathcal{R}^*_{\text{est}}(n) = \inf_{\widehat{F}} \sup_{f\in\Sigma} E\{|\widehat{F}[X_n^f] - \Phi_r[f]|\}$$

where the infimum is taken over all estimates (Borel, w.r.t. the uniform metric, functionals of the observation (1)) and E stands for expectation over the noise affecting the observation;

• For **(II)**, the minimax risk ("resolution") is

$$\mathcal{R}^*(n) = \inf\{\epsilon \mid H_0 \text{ and } H_\epsilon \text{ are } \tfrac{1}{8}\text{-testable via observation (1)}\},$$

where p-testability of the pair H_0, H_ϵ via observation (1) means that there exists a test – a Borel functional $T[\cdot]$ taking values in $\{0, 1\}$ – such that

$$\sup_{f\in\Sigma} P_{f \text{ meets } H_0}\{T[X_n^f] = 1\} \; + \; \sup_{f\in\Sigma} P_{f \text{ meets } H_\epsilon}\{T[X_n^f] = 0\} \leq p, \tag{2}$$

the probability being taken w.r.t. the distribution of noise affecting the observation.

As always in nonparametric statistics, we are interested in the asymptotic as $n \to \infty$ behaviour of the minimax risks.

We focus on the case when the set Σ of signals in question is the Hölder ball $\Sigma_\rho(\beta, L)$ specified by a the triple of parameters $\beta > 0$, $L > 0$ and $\rho \geq 2L$. It is defined as the set of all continuous functions $f : [0,1] \to [-\rho, \rho]$ which are $m = \lfloor \beta - 0 \rfloor = \max\{m \in \mathbf{Z} : m < \beta\}$ times continuously differentiable and such that the m-th derivative is Hölder continuous, with the exponent $\gamma = \beta - m$ and constant $2L$:

$$|f^{(m)}(t) - f^{(m)}(t')| \leq 2L|t - t'|^{\beta - m} \quad \forall t, t' \in [0,1].$$

It is clear that the minimax risk of recovering $\Phi_r[f]$, $f \in \Sigma$, cannot be worse than the minimax risk

$$\mathcal{R}(n) = \inf_{\widehat{f}} \sup_{f\in\Sigma} E\{\| \widehat{f}(X_n^f) - f \|_r\} = O\left(L^{1/(2\beta+1)} n^{-\beta/(2\beta+1)}\right) \tag{3}$$

of recovering signals $f \in \Sigma$, the estimation error being measured in the $\| \cdot \|_r$-norm. This observation is an immediate consequence of the fact that $\Phi_r[\cdot]$ is Lipschitz continuous with the Lipschitz constant equal to 1 in the $\| \cdot \|_r$-norm. So the worst-case over $f \in \Sigma$ risk of estimating $\Phi_r[f]$ by the "plug-in" estimate

$$\widehat{F}(Y) = \Phi_r[\widehat{f}(Y)],$$

associated with an estimate $\widehat{f}(\cdot)$ of f, is not worse than the worst-case $\| \cdot \|_r$-risk of $\widehat{f}$ itself:

$$\sup_{f\in\Sigma} E\{|\Phi_r[\widehat{f}(X_n^f)] - \Phi[f]|\} \leq \sup_{f\in\Sigma} E\{\| \widehat{f}(X_n^f) - f \|_r\}.$$

Taking inf over $\widehat{f}$ we come to

$$\mathcal{R}^*_{\text{est}}(n) \leq \mathcal{R}(n),$$

as claimed.

Now, the minimax risk $\mathcal{R}^*(n)$ in the hypotheses testing problem **(II)** cannot be "essentially worse" than $\mathcal{R}^*_{\text{est}}(n)$ and thusit cannot be essentially larger than $\mathcal{R}(n)$:

$$\mathcal{R}^*(n) \leq 32\mathcal{R}^*_{\text{est}}(n) \quad [\leq 32\mathcal{R}(n)]. \tag{4}$$

Indeed, given an estimate $\widehat{F}$ of $\Phi_r[f]$ with certain worst-case, w.r.t. $f \in \Sigma$, risk R, we can convert it into the following test for distinguishing between H_0 and H_{32R}: given X_n^f, set $T = \{\widehat{F}(X_n^f) \leq 16R\}$ and claim that the true hypothesis is H_0 if $T = 1$ or H_{32R} if $T = 0$. It follows immediately from the Tschebyshev inequality that the resulting test satisfies (2) with $p = \frac{1}{8}$, $\epsilon = 32R$, thus $\mathcal{R}^*(n) \leq 32R$. Since R can be chosen arbitrarily close to $\mathcal{R}^*_{\text{est}}(n)$ it implies the inequality (4).

Outline of results. The first result of the present paper states that the "plug-in" estimate of $\Phi_r[\cdot]$ and the associated test for distinguishing between H_0 and H_ϵ are "nearly optimal" – their performance cannot be improved by more than logarithmic factors. Specifically, we prove (Theorem 2.1 below) that for any large enough observation sample length n one has

$$\mathcal{R}^*(n) \geq C(\beta,r)\mathcal{R}(n)\left(\ln \mathcal{R}(n)\right)^{-\theta(\beta)}, \quad C(\beta,r) > 0, \tag{5}$$

provided that the "degree of regularity" β of our signals "is coherent" with the hypothesis H_0 (namely, $\beta \geq 1$ when $\mathcal{M}$ is the cone of nonincreasing, and $\beta \geq 2$ when $\mathcal{M}$ is the cone of concave functions). The latter assumption, roughly speaking, says that the degree of regularity of functions from $\mathcal{M}$ is not better than the a priori known degree of regularity of our signals. Note that (5) combines with (4) to imply that all three quantities $\mathcal{R}^*(n)$, $\mathcal{R}^*_{\rm est}(n)$, $\mathcal{R}(n)$ are "nearly" (up to logarithmic in these quantities factors) the same:

$$O\left(\frac{L^{1/(2\beta+1)}n^{-\beta/(2\beta+1)}}{\ln^\theta(n)}\right) \leq \mathcal{R}^*(n) \leq O(1)\mathcal{R}^*_{\rm est}(n) \leq O\left(L^{1/(2\beta+1)}n^{-\beta/(2\beta+1)}\right). \tag{6}$$

The outlined result deserves some comments. From large literature on estimating functionals of nonparametric signals and nonparametric hypotheses testing (see. e.g. [1-29] and references therein) it is known that

- Typically, the minimax risk associated with a nonparametric hypotheses testing problem "decide whether the signal underlying observations belongs to a given convex set X or is at a "large" $\|\cdot\|_r$-distance from the set" is essentially less than the risk at which one can estimate the $\|\cdot\|_r$-distance from the signal to the set X. E.g., when $X = \{0\}$ ("decide whether the signal f underlying the observations is nontrivial", the minimax, on $\Sigma_\rho(\beta, L)$, risk in the testing problem is $O\left(n^{-2\beta/(4\beta+1)}\right)$ for $1 \leq r \leq 2$ and $O\left(n^{-\beta/(2\beta+1-1/p)}\right)$ for $r > 2$ [24]. However, the minimax on the same set $\Sigma_\rho(\beta, L)$ risk of recovering $\| f \|_r$ is, up to logarithmic in n factors, the plug-in risk $O(n^{-\beta/(2\beta+1)})$, except for the case when r is an even integer [25]. To the best of our knowledge, the phenomenon observed in the current paper, which we refere to as the "near optimality" of the simplest plug-in schemes, both in hypotheses testing and estimating the distance from f to X is new.
- The minimax, w.r.t. $\Sigma_\rho(\beta, L)$, risk of recovering a "good" functional $F[f]$ is essentially better than the risk $\mathcal{R}(n) = O(n^{-\beta/(2\beta+1)})$ of recovering the signal f itself. E.g., when F is smooth (Frechet differentiable on L_2 with Hölder continuous, with exponent γ, gradient), F can be recovered (cf. [10, 11, 13]) on $\Sigma_\rho(\beta, L)$ with "parametric" worst-case risk $O(n^{-1/2})$. Even a non-smooth (although otherwise "simple") functional $F[f] = \| f \|_2$ can be recovered on $\Sigma_\rho(\beta, L)$ with the risk $O(n^{-4\beta/(4\beta+1)})$ [17, 24], which still is much better than the risk $\mathcal{R}(n)$ of a nonparametric estimate $\widehat{f}$ of $f \in \Sigma_\rho(\beta, L)$ in the $\|\cdot\|_2$-norm. Relation (6) says that this nice phenomenon disappears when instead of estimating the $\|\cdot\|_2$-distance from f to 0[1)] we want to estimate the $\|\cdot\|_2$-distance from f to a less trivial (although quite natural) convex cone, say, the one of nonpositive functions.

The outlined result on "near optimality" of the plug-in estimates/tests in the context of measuring distances to the cones of nonpositive/nonincreasing/concave functions raises several additional questions:

[1)] Or from f to a linear subspace in L_2 – the latter problem turns out to be essentially as good as the one of estimating $\| f \|_2$

1. When speaking about the cones of (ii) nonincreasing and (iii) concave functions, we have assumed that the regularity of f is compatible with the regularity of functions from the cone in question (i.e., $\beta \geq 1$ in the case of (ii) and $\beta \geq 2$ in the case of (iii)). What happens when this assumption is violated? It turns out that here the plug-in estimates/tests become "essentially non-optimal". Say, when $\beta < 1$, the minimax, on $\Sigma_\rho(\beta, L)$, risk of recovering the $\|\cdot\|_r$-distance from f to the cone of nonincreasing functions is $O\left(n^{-(\frac{1}{3}\wedge\frac{2\beta}{(4\beta+1)})}\right)$. This phenomenon is natural – since the degree of regularity of a monotone function is "nearly 1", to measure the distance from a "poorly regular" signal f to the cone of monotone functions is basically the same as to measure the distance from f to 0, and the minimax risk in this latter problem, as we just have mentioned, is significantly less than $\mathcal{R}(n)$.

2. Our result states that the plug-in estimate/test related to $\|\cdot\|_r$-distances from the cones of nonpositive/nonincreasing/concave functions are optimal up to logarithmic in n factors. A natural question is whether one can replace here "logarithmic in n" by "constant", i.e., whether there is a possibility to measure the distances from nonparametric signals to the cones in question asymptotically better than we can recover the signals themselves. We demonstrate that such a possibility does exist in the case of (i), i.e., when the associated cone is the cone of nonpositive functions. However, we do not know neither whether such a possibility exists in the cases (ii), (iii), nor the answer to the following, seemingly interesting, question:

   (?) *Let $F[f]$ be a convex functional on the space $C[0,1]$ of real-valued continuous functions on $[0,1]$ such that F is Lipschitz continuous, with constant 1, w.r.t. the norm $\|\cdot\|_r$ (variant: let $F[f]$ be the $\|\cdot\|_r$-distance from f to a given convex closed, in the uniform norm, subset X of $C[0,1]$). Is it true that $F[f]$ can be estimated via observation (1) with the worst-case, over $\Sigma_\rho(\beta, L)$, risk which is "better in order than the plug-in risk", i.e., is $o(\mathcal{R}(n))$ as $n \to \infty$?*

   In the case when F is the $\|\cdot\|_r$-distance from f to a closed convex set X, we may pose a similar question about the risk of testing the hypothesis "$f \in X$" vs. the alternative "the $\|\cdot\|_r$-distance from f to X is $\geq \epsilon$".

   In all particular cases when the answer to (?) is known, it is affirmative (e.g., the $\|\cdot\|_r$-norm of $f \in \Sigma_\rho(\beta, L)$ indeed can be recovered asymptotically better than f [16, 6, 17, 24, 30]). Is it a "law of nature" or merely a consequence of "simplicity" of the functionals F/sets X which have been studied?

The contents. The rest of the paper is organized as follows. Relation (5) is proved in Section 2; the technique we use here is quite close to the one developed in [25] and is not that standard. A "better-than-plug-in" estimate of the distance from a signal to the cone of nonpositive functions (item 2 above) is built in Section 3. The proofs of main results are placed in Appendices.

2 Lower bound on $\mathcal{R}^*(n)$

Let us fix a class $\Sigma_\rho(\beta, L)$ with $\rho \geq 2L$, an $r \in [1, \infty)$ and a nonnegative integer $q \leq \beta$, and let $\mathcal{M}_q$ be the closure, in the uniform norm on $[0,1]$, of the set of all C^∞ functions with nonpositive everywhere q-th derivative. Let, further,

$$\Phi_r[f] = \inf\{\| f - g \|_r |\ g \in \mathcal{M}_q\}$$

be the $\| \cdot \|_r$-distance from a continuous function $f : [0, 1] \to \mathbf{R}$ to $\mathcal{M}_q$. Note that for $q = 0, 1$ and 2, $\Phi_r[f]$ is the $\| \cdot \|_r$-distance from f to the cones of nonpositive, nonincreasing, and concave functions, respectively – the cones we indeed are interested in.

Given observation $X_n^f(t)$ in (1) of a signal $f \in \Sigma_\rho(\beta, L)$ we are interested to distinguish between the following two hypotheses:

H_0: $f \in \mathcal{M}_q$;

H_ϵ: $\Phi_r[f] \geq \epsilon$.

Recall that we call the hypotheses H_0, H_ϵ are *p-testable*, m being the size of the observations sample, if there exists a test $T[\cdot]$ (a functional of observation (1) taking values 0 or 1) satisfying (2), and we denote by $\mathcal{R}^*(n)$ the infimum of those $\epsilon > 0$ for which the hypotheses H_0 and H_ϵ are $\frac{1}{8}$-testable.

Our goal is to establish the following

Theorem 2.1 *Let $\beta \geq q$. There exists $C = C(\beta, r) > 0$, $\vartheta = \vartheta(\beta, r)$ and $M = M(\beta, \rho, L, r, q)$ such that for all $n \geq M$ it holds*

$$L^{-1/(2\beta+1)} n^{\beta/(2\beta+1)} \mathcal{R}^*(n) \geq C \left(\frac{1}{\ln n}\right)^{\vartheta}. \qquad (7)$$

The proof of the theorem is placed in Appendix 1.

3 Distance to the cone of nonpositive functions

Let $r \in [1, \infty)$, and let

$$\Phi_r[f] = \| f_+ \|_r = \left(\int_0^1 (f_+(t))^r dt \right)^{1/r}, \quad f_+ = f 1_{f>0}$$

be the $\| \cdot \|_r$-distance from f to the cone of nonpositive functions. In the previous section we have demonstrated that the minimax, over $f \in \Sigma_\rho(\beta, L)$, risk $\mathcal{R}^*_{\rm est}(n)$ of recovering $\Phi_r[f]$ from observations (1) cannot be more than by logarithmic in n factor less than the minimax risk $\mathcal{R}(n)$ of recovering f. Our goal now is to demonstrate that the ratio $\frac{\mathcal{R}^*_{\rm est}(n)}{\mathcal{R}(n)}$ indeed converges to 0 at a logarithmic rate as $n \to \infty$:

Theorem 3.1 *Given $\beta > 0, L > 0, \rho > 0$ and $r \in [1, \infty)$, for every n there exists an estimator $\widehat{\Phi}_n$ of $\Phi_r[f]$ via observation (1) such that the worst-case, w.r.t. the Hölder ball $\Sigma = \Sigma_\rho(\beta, L)$, risk*

$$\mathcal{R}^*_{est}(\widehat{\Phi}_n) = \sup_{f \in \Sigma} E\{|\widehat{\Phi}_n(X_n^f) - \Phi_r[f]|\}$$

for all $n \geq n_0(L, \beta, r, \rho)$ admits the upper bound

$$\mathcal{R}^*_{\rm est}(\widehat{\Phi}_n) \leq C L^{1/(2\beta+1)} (n \ln n)^{-2\beta/(2\beta+1)} \leq C' (\ln n)^{-2\beta/(2\beta+1)} \mathcal{R}(n), \qquad (8)$$

with $C, C' < \infty$ depending on β, r only.

The theorem is proved in Appendix 2. Here we present the construction of the estimator $\widehat{\Phi}_n$.

- Given observation X_n^f, we start with building the usual kernel smoothing

$$\widehat{f}_h(t) = \frac{1}{h}\int_0^1 K_t\left(\frac{x-t}{h}\right) dX_t^f(x) \tag{9}$$

of the observation. Here

$$K_t(s) = \begin{cases} K(s), & t \le 1/2 \\ K(-s), & t < 1/2 \end{cases}$$

and the kernel $K(s)$ satisfies the usual restrictions: it is a C^∞ function which vanishes outside of $[0,1]$ and is such that

$$\int K(s)ds = 1,$$

$$\int s^\ell K(s)ds = 0 \text{ for } \ell = 1, ..., m,$$

where m is the largest integer which is $< \beta$. The "bandwidth" h of the smoothing is chosen as a function of n, namely, as

$$h = (L^2 n \ln n)^{-1/(2\beta+1)}, \tag{10}$$

Note that the bandwidth we use is by logarithmic in n factor less than the standard bandwidth $(L^2 n)^{-1/(2\beta+1)}$ yielding an optimal in order non-parametric estimate of f. As a result, the expectation

$$f_h(t) = \frac{1}{h}\int_0^1 K_t\left(\frac{x-t}{h}\right) f(x)dx \tag{11}$$

of $\widehat{f}_h(t)$ is "very close to f" – its uniform distance from f is $\epsilon(n) = O((n \ln n)^{-2\beta/(2\beta+1)})$. Consequently, $\Phi_r[f]$ and $\Phi_r[f_h]$ coincide with each other within accuracy $\epsilon(n)$, which is the "target quality" of the estimator we intend to build. Thus, all we need is to recover from our observations $\Phi_r[f_h]$ with a quality of order of $\epsilon(n)$.

- To build an $O(\epsilon(n))$-accurate estimate of $\Phi_r[f_h]$, we use the technique proposed in [25]. Namely, we approximate the function $t^r 1_{t>0}$ by a trigonometric polynomial T_N of an appropriately high order N and construct an unbiased estimate $T_N^*(\widehat{f}_h(t))$ of $T_N(f_h(t))$; here $T_N^*(\cdot)$ is an appropriately chosen trigonometric polynomial of degree N. The estimate $\widehat{\Phi}_n$ of $\Phi_r[f] \approx \Phi_r[f_h]$ is nothing but

$$\widehat{\Phi}_n = \left(\int_0^1 T_N^*(\widehat{f}_h(t))dt\right)^{1/r}. \tag{12}$$

The outlined scheme defines $\widehat{\Phi}_n$ up to the choice of the trigonometric polynomial $T_N^*(\cdot)$. This "missing element" of the construction is specified as follows (to simplify notation, we assume from now on that the situation is normalized by the requirement $\rho = 1$):

1. By construction,

$$\begin{array}{lrl} (a) & \widehat{f}_h(t) = f_h(t) + \lambda_h \xi_h(t), & \\ (b) & \lambda_h = \left[E\left\{ \left(\int\limits_0^1 K_t\left(\frac{t-x}{h}\right) n^{-1/2} dW(x) \right)^2 \right\} \right]^{1/2} = \frac{\|K\|_2}{\sqrt{nh}}, & \\ (c) & \xi_h(t) = \frac{1}{h\lambda_h} \int\limits_0^1 K\left(\frac{t-u}{h}\right) n^{-1/2} dW(u) = \frac{1}{\|K\|_2\sqrt{h}} \int\limits_0^1 K\left(\frac{t-u}{h}\right) dW(u). & \end{array} \tag{13}$$

Note that $\xi_h(t) \sim \mathcal{N}(0,1)$. We have therefore

$$\begin{array}{rcl} E\{\widehat{f}_h(t)\} & = & f_h(t), \\ \mathrm{Var}\widehat{f}_h(t) & = & E\left\{ \left(\widehat{f}_h(t) - f_h(t) \right)^2 \right\} = \lambda_h^2. \end{array}$$

2. Let $g(t)$ be a once for ever fixed nonnegative function on the axis which is equal to 1 when $0 \le t \le 1$, never exceeds 1 and vanishes when $t \ge 2$. We set

$$q(t) = (\max[0,t])^r \, g(t).$$

Let α_k, $k = 0, 1, ...$ be Fourier coefficients of $q(\cdot)$ on $[-2, 2]$:

$$\alpha_k = \int\limits_{-2}^{2} q(t)\psi_k(t)dt = \int\limits_{0}^{2} q(t)\psi_k(t)dt,$$

where $\psi_0(t) = \frac{1}{2}$ and

$$\psi_{2j-1}(t) = 2^{-1/2}\sin(\pi j t/2), \quad \psi_{2j}(t) = 2^{-1/2}\cos(\pi j t/2),$$

for $j \ge 1$. We set

$$\kappa = \frac{4}{\pi\|K\|_2\sqrt{r(2\beta+1)}} \text{ for } r > 2, \text{ and } \kappa = \frac{4}{\pi\|K\|_2\sqrt{2(2\beta+1)}} \text{ for } 1 \le r \le 2, \quad \theta = \kappa/2 \tag{14}$$

and set

$$N = \lfloor \theta L^{-1/(2\beta+1)} (n \ln n)^{\beta/(2\beta+1)} \rfloor. \tag{15}$$

3. For $k = 0, ..., N$ we define the functions $\phi_{k,\lambda}(t)$ as follows: $\phi_{0,\lambda}(t) = \psi_0(t) = \frac{1}{2}$, and

$$\phi_{2j-1,\lambda}(t) = \exp(\pi^2 j^2 \lambda^2/8)\psi_{2j-1}(t), \quad \phi_{2j,\lambda}(t) = \exp(\pi^2 j^2 \lambda^2/8)\psi_{2j}(t)$$

for $j \ge 1$. We set

$$T^*_{N,\lambda}(t) = \sum_{k=0}^{N} \alpha_k \phi_{k,\lambda}(t). \tag{16}$$

The trigonometric polynomial $T^*_N(u)$ underlying the estimate $\widehat{\Phi}_n$ (see (12)) is

$$T^*_N(u) = T^*_{N,\lambda_h}(u).$$

Appendix 1. Proof of Theorem 2.1

4.1 Setup

Function $G_u(\cdot)$. Let us partition the segment $[0,1]$ into $q+1$ equal segments Δ_ℓ, $\ell = 0,1,...,q$, and let t_ℓ be the left endpoints of these segments. Given a C^∞ function g which vanishes outside Δ_0, consider $(q+1)\times(q+1)$ matrix $A[g]$ with the entries

$$a_{i\ell}[g] = \int\limits_{\Delta_\ell} t^i g(t-t_\ell)dt, \quad \ell, i = 0,1,...,q.$$

Of course, one can choose g to be positive on the interior of Δ_0 and such that the matrix $A[g]$ is nonsingular (look what happens if g is a probability density close, in the weak sense, to the unit mass placed at the midpoint of Δ_0). Let us once for ever fix a nonnegative C^∞ function g which vanishes outside of Δ_0, is positive on the interior of the segment and is such that the matrix $A[g]$ is nonsingular. Since $A[g]$ is nonsingular, there exist reals ω_ℓ, $\ell = 0,1,...,q$, such that

$$\sum_{\ell=0}^{q} \omega_\ell a_{i\ell}[g] = \begin{cases} 0, & i < q \\ 1, & i = q \end{cases}. \tag{17}$$

It is clear that among the reals ω_ℓ, $\ell = 0,...,q$, there are positive ones (otherwise $\sum\limits_\ell \omega_\ell a_{q\ell}[g]$ were nonpositive). Let $\ell_* \in \{0,1,...,q\}$ be such that ω_{ℓ_*} is positive. Let also $\omega = \max_\ell |\omega_\ell|$. We define function $G_u(t)$ (t is real argument, u is real parameter) as follows:

- When $q = 0$, we set
$$G_u(t) = ug(t);$$
- When $q > 0$, we set $G_u(t) = 0$ for $t \le 0$ and
$$G_u(t) = \int\limits_0^t \frac{(t-\tau)^{q-1}}{(q-1)!}\left(\sum_{\ell=0}^{q}[\omega_\ell u - \widehat{\omega}_\ell]g(\tau - t_\ell)\right)d\tau, \quad \widehat{\omega}_\ell = \begin{cases} 0, & \ell = \ell_* \\ \omega, & \text{otherwise} \end{cases}$$

for $t \ge 0$. In other words, $G_u(\cdot)$ vanishes on the nonpositive ray; when $t \ge 0$, $G_u(\cdot)$ satisfies the relation

$$\frac{d^q}{dt^q}G_u(t) = \sum_{\ell=0}^{q}[\omega_\ell u - \widehat{\omega}_\ell]g(t-t_\ell); \quad \left.\frac{d^i}{dt^i}\right|_{t=0} G_u(t) = 0,\ i = 0,1,...,q-1. \tag{18}$$

Note that by construction

$$G_u(t) = G(t) + uD(t),$$

where both G and D are C^∞ functions vanishing on the nonpositive ray.

We claim that

G.1. $\frac{d^q}{dt^q}G_u(t)$ and $D(\cdot)$ vanish outside of $[0,1]$.

> **G.1** is evident when $q = 0$; let us verify this claim for $q \ge 1$. The fact that $\frac{d^q}{dt^q}G_u(t)$ vanishes when $t \ge 1$ is readily given by (18), so that all we need is to verify that $D(t)$ vanishes outside of $[0,1]$. We already know that $D(t) = 0$ when $t \le 0$. For

$t \geq 1$ we have

$$\begin{array}{rcl} D(t) & = & \int\limits_0^t \frac{(t-\tau)^{q-1}}{(q-1)!} \underbrace{\left[\sum_{\ell=0}^{q} \omega_\ell g(\tau - t_\ell)\right]}_{h(\tau)} d\tau \\ & = & \int\limits_0^1 \frac{(t-\tau)^{q-1}}{(q-1)!} h(\tau) d\tau \\ & & \text{[since } h(\cdot) \text{ vanishes outside of } [0,1]] \\ & = & \sum\limits_{i=0}^{q-1} c_i(t) \int\limits_0^1 \tau^i h(\tau) d\tau \\ & = & \sum\limits_{i=0}^{q-1} c_i(t) \sum\limits_{\ell=0}^{q} \omega_\ell a_{i\ell}[g] \\ & = & 0 \\ & & \text{[see (17)]} \end{array}$$

G.2. When $-1 \leq u \leq 0$, $G_u(t) \in \mathcal{M}_q$, and when $0 \leq u \leq 1$, we have

$$\Phi_r[G_u(\cdot)] \geq \kappa_1 u \tag{19}$$

(from now on, $\kappa_i > 0$ depend on β, r only).

Indeed, when $-1 \leq u \leq 0$, the right hand side in (18) is nonpositive (since g is nonnegative, $\omega_{\ell_*} > 0$ and $\widehat{\omega}_\ell \geq |\omega_\ell|$ when $\ell \neq \ell_*$), i.e., $g_u(\cdot) \in \mathcal{M}_q$ for the indicated u. Now let $0 \leq u \leq 1$. By (18), q-th derivative of $G_u(\cdot)$ is nonnegative on Δ_{ℓ_*} and is $\geq \kappa_{1,1} u$ on a once for ever fixed segment $\Delta \subset \Delta_{\ell_*}$. Let ψ be once for ever fixed nontrivial nonnegative C^∞ function which vanishes outside of Δ, and let $\theta(t) = (-1)^q \frac{d^q}{dt^q}\psi(t)$. Whenever $h(\cdot) \in \mathcal{M}_q$, we have $\frac{d^q}{dt^q}[G_u(t) - h(t)] \geq \kappa_{1,1} u$ on Δ, whence

$$\int [G_u(t) - h(t)]\theta(t) dt = \int\limits_\Delta \left(\frac{d^q}{dt^q}[G_u(t) - h(t)]\right)\psi(t) dt \geq \kappa_{1,2} u,$$

so that $\| G_u(\cdot) - h(\cdot) \|_r \geq \frac{\kappa_{1,2} u}{\|\theta\|_{r/(r-1)}}$, and (19) follows.

We set

$$\begin{array}{ll} (a) & \kappa_0^{-1} = 1 + \max\limits_{0 \leq k \leq \beta+1} \max\limits_{\substack{|u| \leq 1 \\ 0 \leq t \leq 1}} \left| \frac{d^k}{dt^k} G_u(t) \right|, \\ (b) & \kappa_* = \kappa_0 \| D(\cdot) \|_2 . \end{array} \tag{20}$$

Parameters N, h and $\alpha(N)$. Given n, let us set

$$\begin{array}{lrcl} (a) & N & = & \lfloor (200L\kappa_*)^{2/(2\beta+1)} (n \ln n)^{1/(2\beta+1)} \rfloor, \\ & h & = & N^{-1}; \\ (b) & \alpha(N) & = & L\kappa_* n^{1/2} N^{-\beta-1/2}. \end{array} \tag{21}$$

Note that for all large enough values of n it holds

$$\frac{0.001}{\sqrt{\ln n}} \leq \alpha(N) \leq \frac{0.01}{\sqrt{\ln N}}. \tag{22}$$

4.2 Translation to the space of sequences

We are about to translate our problem of hypotheses testing to the space of sequences. Let $I_1, ..., I_N$ be the partition of the segment $[0,1]$ into N segments of the length h each, and let t_i be the left endpoint of I_i. We associate with a point $\theta = (\theta_1, ..., \theta_N)$ from the unit cube $B_N = [-1,1]^N$ the function

$$\begin{array}{rcl} f_\theta(t) & = & L\kappa_0 h^\beta \sum\limits_{i=1}^N G_{\theta_i}(h^{-1}(t-t_i)) \\ & = & L\kappa_0 h^\beta \Big[\underbrace{\sum\limits_{i=1}^N G(h^{-1}(t-t_i))}_{F(t)} + \sum\limits_{i=1}^N \theta_i D(h^{-1}(t-t_i)) \Big]. \end{array} \tag{23}$$

and claim first of all that $f_\theta \in \Sigma_\rho(\beta, L)$.

Indeed, by construction f_θ is a C^∞ function on the axis which vanishes on the nonpositive ray. For $\beta + 1 \geq p \geq q$ we have

$$f_\theta^{(p)}(t) = L\kappa_0 h^{\beta-p} \sum_{i=1}^N H_i^p(h^{-1}(t-t_i)), \quad H_i^p(t) = \frac{d^p}{dt^p} G_{\theta_i}(t).$$

Since $p \geq q$, the functions H_i^p vanish outside the respective segments I_i (by **G.1**). Taking into account the origin of κ_0, we conclude that whenever $\beta + 1 \geq p \geq q$ it holds

$$f_\theta^{(p)}(t_i) = 0, \ i = 1, ..., N; \quad |f_\theta^{(p)}(t)| \leq L h^{\beta - p}, \ 0 \leq t \leq 1. \tag{24}$$

Now consider two cases: (a) β is not integer; (b) β is integer.

In the case of (a), the largest integer, m, which is $\leq \beta$, is $\geq q$, so that (24) is valid for $p = m$ and $p = m+1$. Let $t \leq t'$ be two points of $[0,1]$. If t, t' are from the same segment of the partition $[0,1] = I_1 \cup ... \cup I_N$, then from (24) as applied to $p = m+1$ it follows that

$$|f_\theta^{(m)}(t) - f_\theta^{(m)}(t')| \leq L h^{\beta-m-1}|t'-t| \leq L h^{\beta-m-1}|t'-t|^{\beta-m} h^{1-\beta+m} = L|t-t'|^{\beta-m}.$$

If the points $t < t'$ belong to two distinct segments of the partition, then let $t_+ \leq t_-$ be, respectively, the right endpoint of the segment containing t and the left endpoint of the segment containing t'. Taking into account that $f_\theta^{(m)}$ vanishes at $t_\pm$ by (24) and applying the above computation, we get

$$\begin{array}{c} |f_\theta^{(m)}(t)| = |f_\theta^{(m)}(t) - f_\theta^{(m)}(t_+)| \leq L|t-t_+|^{\beta-m} \leq L|t-t'|^{\beta-m}, \\ |f_\theta^{(m)}(t')| = |f_\theta^{(m)}(t') - f_\theta^{(m)}(t_-)| \leq L|t'-t_-|^{\beta-m} \leq L|t-t'|^{\beta-m}, \end{array}$$

whence

$$|f_\theta^{(m)}(t) - f_\theta^{(m)}(t')| \leq 2L|t-t'|^{\beta-m} \quad \forall t, t' \in [0,1],$$

so that $f_\theta \in \mathcal{H}(\beta, L)$.

In the case of (b), (24) is valid for $p = \beta$ (recall that $q \leq \beta$). It follows that $|f_\theta^{(\beta)}(t)| \leq L$, i.e., $f_\theta^{(m)} = f_\theta^{(\beta-1)}$ is Lipschitz continuous with constant L, whence $f_\theta \in \mathcal{H}(\beta, L)$.

We have seen that $f_\theta \in \mathcal{H}(\beta, L)$. This inclusion combines with the fact that f_θ is C^∞ function vanishing on the nonpositive ray to yield that $|f_\theta(t)| \leq 2L \leq \rho$ on $[0,1]$. Thus, $f_\theta \in \Sigma_\rho(\beta, L)$, as claimed.

We further claim that

$$\begin{array}{ll} (a) & \theta \le 0 \Rightarrow f_\theta \in \mathcal{M}_q; \\ (b) \quad \forall \theta \in B_N: & \Phi_r[f_\theta] \ge \kappa_2 L h^\beta \Psi_r(\theta), \end{array} \tag{25}$$

where

$$\Psi_r(\theta) = \left(\frac{1}{N} \sum_{i=1}^{N} ([\theta_i]_+)^r \right)^{1/r} \qquad [a_+ = \max[0,a]] \tag{26}$$

This is an immediate consequence of **G.1-2**.

For $i = 1, ..., N$, let

$$Y_i = Y_i^\theta = \frac{\sqrt{n}}{\| D(\cdot) \|_2 \sqrt{h}} \int_{(i-1)h}^{ih} D(h^{-1}(t - t_i))[dX_n^{f_\theta}(t) - L\kappa_0 h^\beta F(t)dt]. \tag{27}$$

From (23), (20) it immediately follows that

$$Y_i = \alpha(N)\theta_i + \xi_i, \qquad i = 1, ..., N, \tag{28}$$

where $\alpha(N)$ is given by (21.b) and

$$\xi_i = \frac{1}{\| D(\cdot) \|_2 \sqrt{h}} \int_{(i-1)h}^{ih} D(h^{-1}(t - t_i))dW(t),$$

so that ξ_i are independent $\mathcal{N}(0,1)$ random variables. It is straightforward to see that the set of statistics Y_i, $i = 1, ..., N$, is sufficient for the parametric submodel (with $f \in \Sigma^N = \{f_\theta\}_{\theta \in B_N}$). Therefore, when restricting f to belong to Σ^N and setting $s_i = \alpha(N)\theta_i$, $i = 1, ..., N$, the original "signal + white noise" model (1) becomes the "sequence space" model

$$Y_i = s_i + \xi_i, \qquad i = 1, ..., N, \tag{29}$$

with $s = (s_1, ..., s_N)$ from the cube $S_N = [-\alpha(N), \alpha(N)]^N$. With this transformation, the original testing problem (reduced to Σ^N) becomes the problem of testing, via observations (29) of an $s \in S_N$, the hypothesis

H_0: $f^s \equiv f_{\alpha^{-1}(N)s} \in \mathcal{M}_q$

versus the alternative

$H_\epsilon : \Phi_r[f^s] \ge \epsilon.$

Now consider the problem of testing, via observations (29) of an $s \in S_N$, the hypothesis

H_0^N: $s \le 0$

versus the alternative

H_δ^N: $\Psi_r(s) \ge \delta$

with

$$\delta = \delta(\epsilon) = \frac{\alpha(N)}{L\kappa_2 h^\beta}\epsilon$$

(see (25)).

We claim that if the pair of hypotheses H_0, H_ϵ is $\frac{1}{8}$-testable, then so is the pair $H_0^N, H_{\delta(\epsilon)}^N$. Indeed, whenever $s \in S_n$ meets H_0^N, f^s meets H_0 (see (25.a)), and whenever $s \in S_N$ meets $H_{\delta(\epsilon)}^N$, f^s meets H_ϵ (see (25.b)). Thus, denoting by $\mathcal{R}_s(N)$ the infimum of those $\delta > 0$ for which the hypotheses H_0^N and H_δ^N are $\frac{1}{8}$-testable, we get the relation

$$\mathcal{R}^*(n) \geq L\kappa_2 h^\beta \alpha^{-1}(N)\mathcal{R}_s(N) = \kappa_3\sqrt{\frac{N}{n}}\mathcal{R}_s(N). \tag{30}$$

4.3 In the sequence space

We intend to establish the following

Proposition 4.1 *For all large enough values of N one has*

$$\mathcal{R}_s(N) \geq \kappa_4(\ln N)^{-2}\alpha(N). \tag{31}$$

From Proposition 4.1 to Theorem 2.1. Postponing for the moment the proof of Proposition 4.1, let us derive from it the statement of Theorem 2.1. We have:

$$\begin{array}{rcll}
\mathcal{R}^*(n) & \geq & \kappa_3\sqrt{\frac{N}{n}}\mathcal{R}_s(N) & \text{[see (30)]}\\
& \geq & \kappa_{4,1}\sqrt{\frac{N}{n}}\alpha(N)(\ln N)^{-2} & \text{[see (31)]}\\
& = & \kappa_{4,2}L^{1/(2\beta+1)}(n\ln n)^{-\beta/(2\beta+1)}(\ln N)^{-2} & \text{[by (21)]}
\end{array}$$

Taking into account (21), we conclude that for all large enough values of n one has

$$\mathcal{R}^*(n) \geq \kappa_{4,3}L^{1/(2\beta+1)}n^{-\beta/(2\beta+1)}(\ln n)^{-(5\beta+2)/(2\beta+1)},$$

as required in (7).

4.4 Proof of Proposition 4.1

Our proof can be outlined as follows: we define two prior distributions, $\mu_{N,0}$ and $\mu_{N,1}$, on S_N, in such a way that the random sequences s generated from the distribution $\mu_{N,0}$ are nonpositive, while those generated from $\mu_{N,1}$ "typically" have a "significant" positive part: $\Phi_r[s] > \delta_*$ with certain $\delta_* > 0$. Assuming that the hypotheses H_0^N, $H_{\delta_*}^N$ are testable, it should be possible to distinguish well between the two hypotheses on observations (29) saying that the observations are associated with random $s \in S_N$ distributed according to priors $\mu_{N,0}$, $\mu_{N,1}$, respectively. On the other hand, we will show that our priors result in the distributions of observations (29) too close to each other to allow for reliable identification. Thus, the assumption that it is possible to distinguish well between the hypotheses H_0^N and $H_{\delta_*}^N$ leads to a contradiction, whence $\mathcal{R}_s(N) \geq \delta_*$.

4.4.1 Preliminaries

Let $\mu_{N,0}$ and $\mu_{N,1}$ be two probability measures on the parameter set S_N, and let $P_{N,0}$ $P_{N,1}$ be the corresponding distributions of observations (29):

$$P_{N,j} = \mu_{N,j} * \mathcal{L}, \qquad j = 0, 1,$$

where $\mathcal{L}$ is the distribution of observation noises ξ in (29) (i.e., $\mathcal{L}$ is the N-dimensional Gaussian distribution with zero mean and unit covariance matrix). Let also

$$\mathcal{K}(P_{N,0}, P_{N,1}) = \int \ln\left(\frac{dP_{N,1}}{dP_{N,0}}\right) dP_{N,1} \tag{32}$$

be the Kullback distance between $P_{N,0}$ and $P_{N,1}$. We need the following statement (which can be obtained from the Fano inequality; we, however, prefer to present a direct proof).

Lemma 4.1 *One has*

$$R(N) \equiv \inf_T \left(P_{N,0}\{T=1\} + P_{N,1}\{T=0\}\right) \geq \exp\{-e^{-1} - \mathcal{K}(P_{N,0}, P_{N,1})\}. \tag{33}$$

the infimum being taken over all tests (functions of observations (29) *taking values* $0, 1$*).*

Proof. Consider a test $T(\cdot)$ for distinguishing between two hypotheses, H_0 and H_1, on the distribution of observations (29), saying, respectively, that the distribution is $P_{N,0}$ and $P_{N,1}$. Let $A = \{Y \in \mathbf{R}^N : T(Y) = 1\}$, $B = \mathbf{R}^N \backslash A$, $\phi = P_{N,0}(A)$, $\psi = P_{N,1}(B)$, $p = \phi + \psi$. We have

$$\begin{array}{rcl} \mathcal{K}(P_{N,0}, P_{N,1}) & = & -\int\limits_A \ln\left(\frac{dP_{N,0}}{dP_{N,1}}\right) dP_{N,1} - \int\limits_B \ln\left(\frac{dP_{N,0}}{dP_{N,1}}\right) dP_{N,1} \\ & \geq & -P_{N,1}(A) \ln\left(\frac{P_{N,0}(A)}{P_{N,1}(A)}\right) - P_{N,1}(B) \ln\left(\frac{P_{N,0}(B)}{P_{N,1}(B)}\right) \\ & & \qquad\qquad\qquad\qquad \text{[Jensen's inequality]} \\ & = & (1-\psi)\ln\left(\frac{1-\psi}{\phi}\right) + \psi \ln\left(\frac{\psi}{1-\phi}\right) \\ & = & \underbrace{(1-\psi)\ln\left(\frac{1-\psi}{p-\psi}\right)}_{g_1(\psi)} + \underbrace{\psi \ln\left(\frac{\psi}{1-p+\psi}\right)}_{g_2(\psi)}. \end{array}$$

We claim that

$$g_1(\psi) + g_2(\psi) \geq (p+1)\ln\frac{1}{p}, \;\; 0 \leq \psi \leq p, \tag{34}$$

whence, by the preceding computation,

$$\mathcal{K}(P_{N,0}, P_{N,1}) \geq (1+p)\ln\frac{1}{p}. \tag{35}$$

To justify (34), observe that for $0 \leq \psi \leq p$ it holds

$$\begin{array}{rcl}
g_1'(\psi) & = & \ln\big(\underbrace{\frac{p-\psi}{1-\psi}}_{h}\big) + \underbrace{\frac{1-p}{p-\psi}}_{\frac{1-h}{h}} \\
& \geq & \ln(1) = 0 \\
& & \qquad\qquad [\text{concavity of } \ln(\cdot)] \\
\Rightarrow & & \\
g_1(\psi) & \geq & g(0) = \ln\frac{1}{p}, \quad 0 \leq \psi \leq p; \\
g_2'(\psi) & = & \ln\big(\underbrace{\frac{\psi}{1-p+\psi}}_{1-h}\big) + \underbrace{\frac{1-p}{1-p+\psi}}_{h} \\
& \leq & \ln(1) = 0 \\
& & \qquad\qquad [\text{concavity of } \ln(\cdot)] \\
\Rightarrow & & \\
g_2(\psi) & \geq & g_2(p) = -p\ln\frac{1}{p}, \ 0 \leq \psi \leq p,
\end{array}$$

and (34) follows.

Since $p\ln p \geq \exp\{-1\}$, (35) implies that $p \geq \exp\{-e^{-1} - \mathcal{K}(P_{N,0}, P_{N,1})\}$, as required in (33). ∎

4.4.2 Applying Lemma 4.1

We are about to use Lemma 4.1 in the situation where the priors $\mu_{N,j}$ are of product structure:

$$\mu_{N,j} = \mu_j^N, \ j = 0, 1;$$

here μ_0, μ_1 are probability measures on $[-\alpha(N), \alpha(N)]$. We assume that

$$\begin{array}{lrcl}
(a) & \operatorname{supp}\mu_0 & \subset & [-\alpha(N), 0], \\
(b) & \int t_+\mu_1(dt) & = & v_1 > 0.
\end{array} \tag{36}$$

We claim that the following implication holds true:

$$\left.\begin{array}{ll}
(a) & \frac{4\alpha^{2r}(N)}{Nv_1^{2r}} \leq \frac{1}{4} \\
(b) & \mathcal{K}(P_{N,0}, P_{N,1}) \leq 0.3
\end{array}\right\} \Rightarrow \mathcal{R}_s(N) \geq \frac{1}{2}v_1. \tag{37}$$

Indeed, assume, on contrary to what should be proved, that the premise in (37) is satisfied and that there exists a test T, based on observations (29), such that

$$\sup_{\substack{s\in S_N \\ s\leq 0}} \mathrm{Prob}_s\{T = 1\} \ + \sup_{\substack{s\in S_N \\ \Psi_r(s) > \frac{1}{2}v_1}} \mathrm{Prob}_s\{T = 0\} \leq \frac{1}{8}, \tag{38}$$

where Prob_s stands for the probability w.r.t. the distribution of observations (29) associated with s. Let us look how well the test T distiguishes between the distributions $P_{N,0}$, $P_{N,1}$. In view of (36.a), the prior $\mu_{N,0}$ "sits" on the set of nonpositive $s \in S_N$; applying (38), we therefore get

$$P_{N,0}\{T = 1\} = \int \mathrm{Prob}_s\{T = 1\}\mu_{N,0}(ds) \leq \frac{1}{8}. \tag{39}$$

Now let $A = \{s \in S_N : \Psi_r(s) \le \frac{1}{2}v_1\}$. Applying (38), we get

$$P_{N,1}\{T=0\} = \int \mathrm{Prob}_s\{T=0\}\mu_{N,1}(ds) \le \frac{1}{8} + \int\limits_A \mu_{N,1}(ds) \equiv \frac{1}{8} + p. \tag{40}$$

We are about to prove that $p \le \frac{1}{4}$. To this end, we first observe that

$$E_{\mu_{N,1}}\{\Psi_r^r(s)\} = \int (t_+)^r \mu_1(dt) \equiv v^r \ge v_1^r. \tag{41}$$

Setting $w = \frac{1}{2}v_1$, we have

$$\begin{array}{rcl} p &=& \mu_{N,1}\left\{s : \left(\frac{1}{N}\sum\limits_{i=1}^N ([s_i]_+)^r\right)^{1/r} \le w\right\} \\ &=& \mu_{N,1}\left\{s : \frac{1}{N}\sum\limits_{i=1}^N ([s_i]_+)^r \le w^r\right\} \\ &\le& \mu_{N,1}\left\{s : \frac{1}{N}\sum\limits_{i=1}^N ([s_i]_+)^r \le \frac{1}{2^r}v^r\right\} \\ &\le& \frac{N^{-1}\int [(t_+)^{2r} - v^{2r}]\mu_1(dt)}{(v^r - 2^{-r}v^r)^2} \\ && \qquad \text{[Tschebyshev inequality]} \\ &\le& \frac{4\alpha^{2r}(N)}{Nv^{2r}} \\ && \qquad \text{[since } \mathrm{supp}\,\mu_1 \subset [-\alpha(N), \alpha(N)]] \\ &\le& \frac{1}{4} \\ && \qquad \text{[by (37.}a\text{) and in view of } v \ge v_1] \end{array}$$

as claimed.

Since $p \le \frac{1}{4}$, (39), (40) imply that

$$P_{N,0}\{T=1\} + P_{N,1}\{T=0\} \le \frac{1}{2},$$

whence, by (33),

$$\exp\{-e^{-1} - \mathcal{K}(P_{N,0}, P_{N,1})\} \le 0.5,$$

or, which is the same,

$$\mathcal{K}(P_{N,0}, P_{N,1}) \ge \ln 2 - e^{-1} = 0.325...$$

which contradicts (37.b). Implication (37) is proved.

Note that the Kullback distance between the marginal measures $P_{N,0}$ and $P_{N,1}$, due to the product structure of model (29) and of the priors $\mu_{N,0}, \mu_{N,1}$, can be written down as

$$\begin{array}{c} \mathcal{K}(P_{N,0}, P_{N,1}) = N\mathcal{K}(p_{\mu_0}, p_{\mu_1}), \\ \mathcal{K}(p_{\mu_0}, p_{\mu_1}) = \int \ln\left(\frac{p_{\mu_1}(y)}{p_{\mu_0}(y)}\right) p_{\mu_1}(y)dy, \end{array} \tag{42}$$

where, for a finitely supported measure μ on the axis,

$$\begin{array}{c} p_\mu(y) = \int \varphi(y-t)\mu(dt), \\ \varphi(y) = \frac{1}{\sqrt{2\pi}}\exp\{-y^2/2\} \end{array}$$

(φ is the standard Gaussian density on the axis). Thus, (37) can be rewritten as

$$\left.\begin{array}{ll}(a) & \frac{4\alpha^{2r}(N)}{Nv_1^{2r}} \leq \frac{1}{4} \\ (b) & N\mathcal{K}(p_{\mu_0}, p_{\mu_1}) \leq 0.3\end{array}\right\} \Rightarrow \mathcal{R}_s(N) \geq \frac{1}{2}v_1. \tag{43}$$

4.4.3 Specifying μ_0, μ_1

It is time now to specify our choice of the measures μ_j, $j = 0, 1$. These measures will be "$\alpha(N)$-scalings" of probability measures ν_j on $[-1, 1]$:

$$\mu_j(A) = \nu_j(\alpha^{-1}(N)A) \quad [\lambda A = \{t = \lambda a \mid a \in A\}] \tag{44}$$

The measures ν_j, $j = 0, 1$, are determined by two parameters: a positive integer m and a real $\sigma \in (0, 1]$; these parameters will be specified later on.

Given m, σ, we define μ_j, $j = 0, 1$, as follows. Consider the Banach space $C[-1, 0]$ of continuous functions on $[-1, 0]$ equipped with the uniform norm, and let $\mathcal{P}_m$ be the subspace of this space comprised of all polynomials of degree $\leq m$. The value $p(\sigma)$ of a polynomial $p \in \mathcal{P}_m$ at the point σ is a linear functional on the finite-dimensional subspace $\mathcal{P}_m$ of $C[-1, 0]$; let ν_* be the norm of this linear functional on $\mathcal{P}_m$. By Hahn-Banach Theorem, we can extend the functional in question from $\mathcal{P}_m$ to the entire $C[-1, 0]$, not increasing the norm of the functional. Taking into account that every continuous linear functional on $C[-1, 0]$ is an integral over a measure (not necessarily nonnegative) of bounded variation, we see that there exists a measure ν on $[-1, 0]$ such that

$$\begin{array}{ll}(a) & \int p(t)\nu(dt) = p(\sigma) \quad \forall p \in \mathcal{P}_m; \\ (b) & \mathrm{Var}(\nu) \equiv \int |\nu(dt)| = \nu_*.\end{array} \tag{45}$$

The quantity ν_* can be computed explicitly. Indeed, by its origin, ν_* is the maximum of $p(\sigma)$, the maximum being taken over all polynomials $p(\cdot)$ with $\deg p \leq m$ and $\max\limits_{-1 \leq t \leq 0} |p(t)| \leq 1$. The corresponding extremal polynomial $\pi_m(t)$ is known (Markov's Theorem); it is obtained from the Tschebyshev polynomial $T_m(t) = \cos(m \arccos(t))$ by linear substitution of argument which maps the segment $[-1, 0]$ onto the segment $[-1, 1]$:

$$\pi_m(t) = T_m(2t + 1).$$

Thus, $\nu_* = \pi_m(\sigma) = T_m(1 + 2\sigma)$. Taking into account that $T_m(h) = \mathrm{ch}\,(m\,\mathrm{acosh}(h))$ for $h > 1$, we get

$$\nu_* = \mathrm{ch}\,(m\,\mathrm{acosh}(1 + 2\sigma)). \tag{46}$$

Let now $\nu = \nu_+ - \nu_-$ be the decomposition of ν into its positive and negative components, and let δ_σ be the unit mass placed at the point σ. We set

$$\nu_0 = \mathrm{Var}^{-1}(\nu_+)\nu_+; \quad \nu_1 = [\mathrm{Var}(\nu_-) + 1]^{-1}[\nu_- + \delta_\sigma]. \tag{47}$$

By construction, $\nu_\pm$ are probability measures on $[-1, 1]$. We claim that

$$\begin{array}{ll}(a) & \mathrm{supp}\,\nu_0 \subset [-1, 0]; \\ (b) & \mathrm{supp}\,\nu_1 \subset [-1, 0] \cup \{\sigma\}, \\ & \nu_1(\{\sigma\}) = \frac{2}{1+\mathrm{ch}\,(m\,\mathrm{acosh}(1+2\sigma))}; \\ (c) & \int t^i \nu_0(dt) = \int t^i \nu_1(dt), \quad i = 0, 1, ..., m.\end{array} \tag{48}$$

Indeed, (a) is evident. To prove (b,c), observe first that $(45.a)$ applied with $p(t)\equiv 1$ implies that

$$\mathrm{Var}(\nu_+) = \mathrm{Var}(\nu_-) + 1, \tag{49}$$

whence, in view of $(45.b)$,

$$\mathrm{Var}(\nu_+) = \mathrm{Var}(\nu_-) + 1 = \frac{\nu_* + 1}{2} = \frac{\mathrm{ch}\,(m\,\mathrm{acosh}(1+2\sigma)) + 1}{2};$$

in particular, $(48.b)$ indeed takes place. Besides this, (49) ensures that

$$\int t^i[\nu_0(dt) - \nu_1(dt)] = \frac{1}{\mathrm{Var}(\nu_+)}\int t^i[\nu_+(dt) - \nu_-(dt) - \delta_\sigma(dt)] = \frac{1}{\mathrm{Var}(\nu_+)}[\int t^i\nu(dt) - \sigma^i],$$

and the latter quantity is 0 for $i \le m$ due to $(45.a)$. We have proved $(48.c)$.

Since the measures μ_j, $j=0,1$, are obtained from the measures ν_j by scaling (44), relation (48) implies that

$$\begin{array}{ll}(a) & \mathrm{supp}\,\mu_0 \subset [-\alpha(N), 0];\\ (b) & \mathrm{supp}\,\mu_1 \subset [-\alpha(N), 0] \cup \{\sigma\alpha(N)\} \subset [-\alpha(N), \alpha(N)],\\ & \mu_1(\{\sigma\alpha(N)\}) = \frac{2}{1+\mathrm{ch}\,(m\,\mathrm{acosh}(1+2\sigma))};\\ (c) & \int t^i\mu_0(dt) = \int t^i\mu_1(dt), \quad i=0,1,...,m.\end{array} \tag{50}$$

4.4.4 Bounding $\mathcal{K}(p_{\mu_0}, p_{\mu_1})$

Now let us prove the following

Lemma 4.2 *Let $\alpha\in(0,1]$, $m>0$ be an integer, and let ϕ,ψ be two probability measures on $[-1,1]$ such that*

$$\int t^i\phi(dt) = \int t^i\psi(dt),\ i=0,1,...,m. \tag{51}$$

For a probability measure γ on $[-1,1]$, let

$$p_{\gamma,\alpha}(y) = \int\varphi(y-\alpha t)\gamma(dt) = \varphi(y)\int\exp\{\alpha t y - \alpha^2t^2/2\}\gamma(dt)$$
$$[\varphi(y) = \tfrac{1}{\sqrt{2\pi}}\exp\{-y^2/2\}]$$

Then for every $T\ge 2$ and all $\alpha\le\frac{1}{10T}$ one has

$$\mathcal{K}(p_{\phi,\alpha}, p_{\psi,\alpha}) \equiv \int\ln\left(p_{\psi,\alpha}(y)/p_{\phi,\alpha}(y)\right)p_{\psi,\alpha}(y)dy \le \frac{3}{2}(10\alpha T)^{m+1} + 12\alpha\exp\{-(T-1)^2/2\}. \tag{52}$$

Proof. 1^0. We have

$$\begin{array}{rcl}\mathcal{K}(\alpha) & \equiv & \mathcal{K}(p_{\phi,\alpha}, p_{\psi,\alpha}) = \int\ln\left(\frac{g_\psi(\alpha,y)}{g_\phi(\alpha,y)}\right)g_\psi(\alpha,y)\varphi(y)dy,\\ g_\gamma(\alpha,y) & = & \int\exp\{\alpha t y - \alpha^2t^2/2\}\gamma(dy).\end{array} \tag{53}$$

For every $T \geq 2$, we have

$$\begin{array}{rcl}\mathcal{K}(\alpha) & = & \mathcal{K}_T(\alpha) + H_T(\alpha),\\ \mathcal{K}_T(\alpha) & = & \int\limits_{-T}^{T} \ln\left(\frac{g_\psi(\alpha,y)}{g_\phi(\alpha,y)}\right) g_\psi(\alpha,y)\varphi(y)dy,\\ H_T(\alpha) & = & \int\limits_{|y|\geq T} \ln\left(\frac{g_\psi(\alpha,y)}{g_\phi(\alpha,y)}\right) g_\psi(\alpha,y)\varphi(y)dy.\end{array} \tag{54}$$

2^0. We claim that

2^0(i): One has

$$H_T(\alpha) \leq 12\alpha\exp\{-(T-1)^2/2\}. \tag{55}$$

Indeed, whatewer is a probabilistic measure γ on $[-1,1]$, we clearly have

$$\exp\{-\alpha^2/2 - \alpha|y|\} \leq g_\gamma(\alpha,y) \leq \exp\{\alpha|y|\},$$

whence

$$\begin{array}{rcl}H_T(\alpha) & \leq & \int\limits_{|y|\geq T} \ln\left(\frac{\exp\{\alpha|y|\}}{\exp\{-\alpha^2/2-\alpha|y|\}}\right)\exp\{\alpha|y|\}\varphi(y)dy\\ & = & 2\int\limits_{y\geq T}\left(2\alpha y + \alpha^2/2\right)\exp\{\alpha y - y^2/2\}\frac{dy}{\sqrt{2\pi}}\\ & = & \sqrt{\frac{2}{\pi}}\int\limits_{z\geq T-\alpha}\left(2\alpha z + 5\alpha^2/2\right)\exp\{-z^2/2\}\exp\{\alpha^2/2\}dz\\ & & \text{[substitution } z = y-\alpha]\\ & \leq & 6\exp\{\alpha^2/2\}\alpha\int\limits_{z\geq T-1} z\exp\{-z^2/2\}dz\\ & & \text{[since } T-\alpha \geq T-1 \geq 1 \geq \alpha, \text{ whence } 2\alpha z + 5\alpha^2/2 \leq 6\alpha z]\\ & \leq & 12\alpha\exp\{-(T-1)^2/2\}.\end{array}$$

2^0(ii) The function $\mathcal{K}_T(\cdot)$ can be extended, as an analytic function, to the circle

$$D_T = \{z \in \mathbf{C} : |z| \leq d_t \equiv \frac{1}{10T}\},$$

and

$$z \in D_T \Rightarrow |\mathcal{K}_T(z)| \leq \frac{3}{2}. \tag{56}$$

Indeed, if γ is a probability measure on $[-1,1]$ and $|y| \leq T$, then the function of α

$$h_y(\alpha) = g_\gamma(\alpha,y) = \int \exp\{\alpha t y - \alpha^2 t^2/2\}\gamma(dt)$$

is analytic, and the modulus of its derivative in the circle $|\alpha| \leq a \leq 1$ does not exceed $(T+1)\exp\{aT + a^2/2\}$, whence

$$|\alpha| \leq a \leq 1 \Rightarrow |h_y(\alpha) - h_y(0)| = |h_y(\alpha) - 1| \leq a(T+1)\exp\{aT + a^2/2\}.$$

It follows that

$$\forall (y \in \mathbf{R} : |y| \leq T, \alpha \in \mathbf{C} : |\alpha| \leq d_T) : \quad |g_\gamma(\alpha,y) - 1| \leq \frac{T+1}{10T}\exp\{0.105\} \leq \frac{1}{5}\exp\{0.105\} \leq \frac{1}{4}. \tag{57}$$

As a result,

$$\forall (y \in \mathbf{R} : |y| \leq T, \alpha \in \mathbf{C} : |\alpha| \leq d_T) : \quad \left| \frac{g_\psi(\alpha, y)}{g_\phi(\alpha, y)} - 1 \right| \leq \frac{1 + \frac{1}{4}}{1 - \frac{1}{4}} - 1 = \frac{2}{3}.$$

We see that the function $\ln\left(\frac{g_\psi(\alpha,y)}{g_\phi(\alpha,y)}\right)$, regarded as function of α, can be extended analytically from the real axis to the circle D_T on the complex plane, an the modulus of the extension in this circle does not exceed the quantity $\sum_{j=1}^{\infty} \frac{1}{j}\left(\frac{2}{3}\right)^j = -\ln\left(1 - \frac{2}{3}\right) = \ln 3$:

$$\forall (y \in \mathbf{R} : |y| \leq T, \alpha \in \mathbf{C} : |\alpha| \leq d_T) : \quad \left| \ln\left(\frac{g_\psi(\alpha, y)}{g_\phi(\alpha, y)}\right) \right| \leq \ln 3. \tag{58}$$

Combining (57) as applied to $\gamma = \psi$ with (58), we see that the function $\mathcal{K}_T(\alpha)$ indeed can be extended, as an analytic function of α, to the circle D_T, and the modulus of the extension in this circle does not exceed the quantity

$$\int\limits_{|y| \leq T} (\ln 3)\left(1 + \frac{1}{4}\right)\varphi(y)dy \leq \frac{5 \ln 3}{4} \leq \frac{3}{2},$$

as claimed.

2^0(iii) The function $\mathcal{K}_T(\alpha)$ has zero of order at least $m+1$ at the point $\alpha = 0$.

Indeed, let

$$f(\alpha, y) = g_\psi(\alpha, y) - g_\phi(\alpha, y) = \int \exp\{\alpha t y - \alpha^2 t^2/2\}[\psi(dt) - \phi(dt)].$$

For $\ell = 0, 1, ..., m$ we have

$$\begin{aligned} \left.\frac{\partial^\ell}{\partial \alpha^\ell}\right|_{\alpha=0} f(\alpha, y) &= \int \left.\frac{\partial^\ell}{\partial \alpha^\ell}\right|_{\alpha=0} \exp\{\alpha t y - \alpha^2 t^2\}[\psi(dt) - \phi(dt)] \\ &= \int t^\ell f_\ell(y)[\psi(dt) - \phi(dt)] \\ &= 0 \\ & \qquad\qquad [\text{see (51)}] \end{aligned}$$

It follows that there exist $C < \infty$, $c > 0$ such that

$$\forall (\alpha, y \in \mathbf{R} : |y| \leq T, |\alpha| \leq c) : \quad |f(\alpha, y)| \leq C|\alpha|^{m+1}.$$

Since

$$\ln\left(\frac{g_\psi(\alpha, y)}{g_\phi(\alpha, y)}\right) = \ln\left(1 + \frac{f(\alpha, y)}{g_\phi(\alpha, y)}\right),$$

we conclude that there exist $C' < \infty$, $c' > 0$ such that

$$\forall (\alpha, y \in \mathbf{R} : |y| \leq T, |\alpha| \leq c') : \quad \left| \ln\left(\frac{g_\psi(\alpha, y)}{g_\phi(\alpha, y)}\right) \right| \leq C'|\alpha|^{m+1},$$

whence

$$|\mathcal{K}_T(\alpha)| \leq \int\limits_{|y| \leq T} \left| \ln\left(\frac{g_\psi(\alpha, y)}{g_\phi(\alpha, y)}\right) \right| g_\psi(\alpha, y)\varphi(y)dy \leq o(|\alpha^m|), \quad \alpha \to 0.$$

as claimed in (iii).

3^0. By 2^0(ii-iii), the function $\mathcal{K}_T$ in the circle D_T satisfies the bound

$$|\mathcal{K}_T(\alpha)| \leq \frac{3}{2}\left(\frac{|\alpha|}{d_T}\right)^{m+1}. \tag{59}$$

Combining (59) and 2^0(i), we come to (52). ∎

4.5 Concluding the proof of Proposition 4.1

Now let us specify the parameters m, σ underlying the construction of the measures ν_j, μ_j, $j = 0, 1$, as

$$\begin{array}{rcl} m & = & \rfloor \ln N \lfloor, \\ \sigma & = & \frac{1}{4m^2}. \end{array} \tag{60}$$

1^0. Observe that the data $(\alpha = \alpha(N), m, \psi = \nu_1, \phi = \nu_0)$ satisfy the premise of Lemma 4.2, and that the functions $p_{\phi,\alpha}(\cdot)$, $p_{\psi,\alpha}(\cdot)$ associated with this data are $p_{\mu_0}(y)$, $p_{\mu_1}(y)$, respectively. Setting

$$T = \sqrt{2 \ln N} + 1$$

and taking into account (22), we see that $10\alpha(N)T \leq 0.2$, provided that N is large enough. Thus, we may use (52) to get the estimate

$$\begin{array}{rcl} N\mathcal{K}(p_{\mu_0}, p_{\mu_1}) & \leq & \frac{3}{2}N(10\alpha(N)T)^{m+1} + 12N\alpha(N)\exp\{-(T-1)^2/2\} \\ & \leq & \frac{3}{2}N(0.2)^{m+1} + 12N\alpha(N)\exp\{-\ln N\} \\ & = & \frac{3}{2}N(0.2)^{m+1} + 12\alpha(N), \end{array} \tag{61}$$

and the concluding quantity, in view of (60) and (22), is ≤ 0.3 for all large enough values of N. Thus, (43.a) is valid for all large enough values of N.

2^0. The measure μ_0 clearly satisfies (36.a). In view of (50), μ_2 satisfies (36.b) with

$$v_1 = \frac{2\sigma\alpha(N)}{1 + \mathrm{ch}\,(m\,\mathrm{acosh}(1 + 2\sigma))}.$$

We have $\mathrm{ch}\,(2\sqrt{\sigma}) \geq 1 + \frac{1}{2}(2\sqrt{\sigma})^2 = 1 + 2\sigma$, whence $\mathrm{acosh}(1 + 2\sigma) \leq 2\sqrt{\sigma} = m^{-1}$. It follows that for all large enough values of N it holds

$$v_1 \geq \frac{2\sigma\alpha(N)}{1 + \mathrm{ch}\,(1)} \geq 0.75\sigma\alpha(N) \geq \kappa_5(\ln N)^{-2}\alpha(N). \tag{62}$$

In view of (22), (62), (21) the condition in (43.a) is satisfied for all large enough values of N.

3^0. We see that when n (or, which is the same, N) is large enough, the measures μ_j, $j = 0, 1$, satisfy the premise in (43), and the corresponding v_1 satisfies (62). Applying (43), we get the bound

$$\mathcal{R}_s(N) \geq \kappa_6(\ln N)^{-2}\alpha(N),$$

as required in Proposition 4.1. ■

Appendix 2. Proof of Theorem 3.1

We start with a simple technical lemma.

Lemma 5.3 *Let*

$$T_N(z) = \sum_{k=0}^{N} \alpha_k \psi_k(z) \tag{63}$$

be the Fourier polynomial of order $N > 1$ of the function q. Let j be an integer. Then there exists C_0 (from now on, all C_i depend on r only) such that for every $z \in [-2, 2]$ it holds

$$|q^{(j)}(z) - T_N^{(j)}(z)| \leq \frac{C_0(\pi/4)^j}{N^{r-j}(r-j)} \quad \textit{for} \quad 0 \leq j < r \tag{64}$$

Further, if r is integer, $|T_N^{(r)}(z)| \le C_0(\pi/4)^r(1+\ln N)$, and

$$\left|T_N^{(j)}(z)\right| \le C_0(\pi/4)^j N^{j-r} \quad for \quad j > r. \tag{65}$$

Proof: It can be easily verified that the Fourier coefficients α_k of q satisfy $|\alpha_k| \le C_0 k^{-(r+1)}$ for some $C_0 < \infty$. Then for any $z \in [-2,2]$ and every $j < r$ one has

$$|q^{(j)}(z) - T_N^{(j)}(z)| \le \sum_{k=N+1}^{\infty} \left(\frac{\pi k}{4}\right)^j |\alpha_k| \le C_0 \left(\frac{\pi}{4}\right)^j \sum_{k=N+1}^{\infty} k^{j-r-1} = C_0 \left(\frac{\pi}{4}\right)^j N^{j-r}.$$

On the other hand, for $j \ge r$

$$\left|T_N^{(j)}(z)\right| \le \sum_{k=1}^{N} \left(\frac{\pi k}{4}\right)^j |\alpha_k| \le C_0 \left(\frac{\pi}{4}\right)^j \sum_{k=1}^{N} k^{j-r-1}.$$

The latter sum can be bounded by $(1+\ln N)$ for $j = r$ and by N^{j-r} for $j > r$. ∎

Lemma 5.4 *Let the trigonometric polynomial $T^*_{N,\lambda}(\cdot)$ be defined by (16). Suppose that ξ is an $\mathcal{N}(0,1)$ random variable. Then $T^*_{N,\lambda}(z+\lambda\xi)$ is an unbiased estimate of $T_N(z)$, i.e.*

$$E\{T^*_{N,\lambda}(z+\lambda\xi)\} = T_N(z), \tag{66}$$

and there exists C_1 such that for all $z \in [-1,1]$ and all $\lambda \in [0,1]$ the variance of this estimate can be bounded as follows:

$$E\{(T^*_{N,\lambda}(z+\lambda\xi) - T_N(z))^2\} \quad \le \quad C_1 \left[\lambda^2 (z_+)^{2r-2} + \lambda^{2r} \ln^2 N + N^{-2r} \exp\left\{\frac{\pi^2\lambda^2 N^2}{16}\right\}\right], \tag{67}$$

where $z_+ = \max[0,z]$.

Proof: To prove (66), it suffices to verify that $E\{\phi_{k,\lambda}(z+\lambda\xi)\} = \psi_k(z)$, $k = 0,...,N$. The relation is evident for $k = 0$. For $j > 0$ we have

$$\begin{aligned}
E\{\phi_{2j-1,\lambda}(z+\lambda\xi)\} &= 2^{-1/2}(2\pi)^{-1/2}\exp\{\pi^2 j^2\lambda^2/8\} \int_{-\infty}^{\infty} \sin(\pi j(z+\lambda x)/2)\exp\{-x^2/2\}dx \\
&= 2^{-1/2}(2\pi)^{-1/2}\mathrm{Im}\left(\int_{-\infty}^{\infty} \exp\left\{\pi^2 j^2\lambda^2/8 + i\pi j(z+\lambda x)/2 - x^2/2\right\} dx\right) \\
&= \mathrm{Im}\left(2^{-1/2}\exp\{i\pi jz/2\}\,(2\pi)^{-1/2}\int_{-\infty}^{\infty} \exp\left\{-(x - i\pi j\lambda/2)^2/2\right\} dx\right) \\
&= 2^{-1/2}\sin(\pi jz/2) = \psi_{2j-1}(z).
\end{aligned}$$

Similar computation yields the equality $E\{\phi_{2j,\lambda}(z+\lambda\xi)\} = \psi_{2j}(z)$. (66) is proved.

Let us prove (67). According to Theorem 1.1 in [13], for every real z one has

$$E\{(T^*_{N,\lambda}(z+\lambda\xi) - T_N(z))^2\} = \sum_{j=1}^{\infty} \frac{\lambda^{2j}|T_N^{(j)}(z)|^2}{j!} \equiv \sum_{j=1}^{\infty} \frac{I_j}{j!}. \tag{68}$$

Let $|z| \le 1$, $0 \le \lambda \le 1$. The associated quantities I_j can be bounded as follows:

1. When $1 \le j < r$, we have (all $c_i(r)$ depend on r only): $|q^{(j)}(z)| \le c_0(r)(z_+)^{r-j}$, whence, by Lemma 5.3, $|T_N^{(j)}(z)| \le c_1(r)\left[(z_+)^{r-j} + N^{-(r-j)}\right]$. Thus,

$$1 \le j < r \Rightarrow I_j \le c_2(r)\lambda^{2j}\left[(z_+)^{2(r-j)} + N^{-2(r-j)}\right] \le c_3(r)\left[\lambda^{2j}(z_+)^{2(r-j)} + (\lambda^2N^2\pi^2/16)^jN^{-2r}\right]. \quad (69)$$

2. When $j = r$ (which is possible only when r is integer), Lemma 5.3 says that

$$I_r \le c_4(r)\lambda^{2r}\ln^2(N). \quad (70)$$

3. Finally, for $j > r$ Lemma 5.3 implies that

$$I_j \le c_6(r)(\lambda^2\pi^2N^2/16)^jN^{-2r}. \quad (71)$$

Thus,

$$\begin{aligned}\sum_{j=1}^{\infty}\frac{I_j}{j!} &\le c_7(r)\left[\sum_{1\le j<r}\lambda^{2j}(z_+)^{2r-2j} + \lambda^{2r}\ln^2(N) + N^{-2r}\sum_{j=1}^{\infty}\frac{1}{j!}(\lambda^2\pi^2N^2/16)^j\right] \\ &\le c_8(r)\left[\lambda^2(z_+)^{2r-2} + \lambda^{2r}\ln^2(N) + N^{-2r}\exp\{\pi^2\lambda^2N^2/16\}\right].\end{aligned}$$

∎

Let

$$\Psi_r[f_h] = \int_0^1 (f_{h,+})^r(t)dt \quad [f_{h,+} = (f_h)_+ = \max[f_h, 0]],$$
$$\gamma_n(t) = T^*_{N,\lambda_h}(\widehat{f}_h(t)),$$
$$\widehat{\Psi}_n = \int_0^1 \gamma_n(t)dt,$$

so that

$$\Phi_r[f_h] = (\Psi_r[f_h])^{1/r}, \quad \widehat{\Phi}_n = \left(\widehat{\Psi}_n\right)^{1/r}.$$

Lemma 5.5 *There exists C_3 such that*

$$E\{|\widehat{\Psi}_n - E\{\widehat{\Psi}_n\}|\} \le C_3\left(N^{-1}\Phi_r^{r-1}[f_h] + N^{-r}\right)$$

Proof: Since the support of the kernel K is contained in $[0, 1]$, we conclude from the definition of $\widehat{f}_h(t)$ that $\widehat{f}_h(t)$ and $\widehat{f}_h(t')$ are independent random variables when $|t' - t| > 2h$. Consider the partition of the interval $[0, 1]$ into the bins of length $2h$. For the sake of simplicity, we suppose that $h^{-1} = 4m$ for some integer m, so that there are exactly $2m$ bins. Let us denote

$$\xi_j = \int_{4jh}^{2(2j+1)h}(\gamma_n(t) - E\{\gamma_n(t)\})dt, \quad \eta_j = \int_{2(2j+1)h}^{4(j+1)h}(\gamma_n(t) - E\{\gamma_n(t)\})dt.$$

Then we have for any p, $1 \le p \le 2$:

$$\begin{aligned}&\left[E\{|\widehat{\Psi}_n - E\{\widehat{\Psi}_n\}|^p\}\right]^{1/p} = \left[E\left\{\left|\int_0^1(\gamma_n(t) - E\{\gamma_n(t)\})dt\right|^p\right\}\right]^{1/p} \\ &\le \left[E\left\{|\sum_{j=0}^{m-1}\xi_j|^p\right\}\right]^{1/p} + \left[E\left\{|\sum_{j=0}^{m-1}\eta_j|^p\right\}\right]^{1/p} \le \left[\sum_{j=0}^{m-1}E\{|\xi_j|^p\}\right]^{1/p} + \left[\sum_{j=0}^{m-1}E\{|\eta_j|^p\}\right]^{1/p} \quad (72)\end{aligned}$$

the last $\le$ in this chain is given by the Marcinkiewicz-Zygmund inequality as applied to the collections $\{\xi_j\}$ and $\{\eta_j\}$ of independent (see above) random variables. By Jensen's inequality (67) implies that

$$\begin{aligned} E\{|\xi_j|^p\} &\le (2h)^{p-1} \int_{4jh}^{2(2j+1)h} E\left\{|\gamma_n(t) - E\{\gamma_n(t)\}|^p\right\} dt \\ &\le (2h)^{p-1} \int_{4jh}^{2(2j+1)h} \left(E\left\{(\gamma_n(t) - E\{\gamma_n(t)\})^2\right\}\right)^{p/2} \\ &\le C_4 h^{p-1}\left(\lambda_h^p \int_{4jh}^{2(2j+1)h} (f_{h,+}(t))^{p(r-1)}dt + \lambda_h^{rp}(\ln N)^p + N^{-rp}\exp\left\{\frac{p\pi^2N^2\lambda_h^2}{32}\right\}\right). \end{aligned}$$

The same bound holds for $E\{|\eta_j|^p\}$. Let us set $p = r/(r-1)$ for $r > 2$ and $p = 2$ for $1 \le r \le 2$. Substituting the associated bounds for $E\{|\xi_j|^p\}$, $E\{|\eta_j|^p\}$ into (72), we get

$$\begin{aligned} E\{|\widehat{\Psi}_n - E\{\widehat{\Psi}_n\}|\} &\le \left[E\left\{|\widehat{\Psi}_n - E\{\widehat{\Psi}_n\}|^p\right\}\right]^{1/p} \\ &\le C_5 h^{\frac{1}{r}\wedge\frac{1}{2}}\left(\lambda_h \parallel f_{h,+} \parallel_r^{r-1} + \lambda_h^r \ln N + N^{-r}\exp\left\{\tfrac{\pi^2N^2\lambda_h^2}{32}\right\}\right) \end{aligned} \tag{73}$$

(from now on, $a\wedge b = \min[a,b]$). Now recall that λ_h and N are such that $\lambda_h^2N^2 = \theta^2 \parallel K \parallel_2^2 \ln n$, while $\frac{\theta^2\pi^2\|K\|_2^2}{32} \le \frac{1}{2r(2\beta+1)}$ for $r > 2$ and $\frac{\theta^2\pi^2\|K\|_2^2}{32} \le \frac{1}{4(2\beta+1)}$ for $1 \le r \le 2$ (see (14), (15)). Thus,

$$h^{\frac{1}{r}\wedge\frac{1}{2}}\exp\left\{\frac{\pi^2N^2\lambda_h^2}{32}\right\} \le n^{-\left(\frac{1}{2r(2\beta+1)}\wedge\frac{1}{4(2\beta+1)}\right)}. \tag{74}$$

Furthermore, from (10), (14), (15) it follows that $h^{\frac{1}{r}\wedge\frac{1}{2}}\lambda_h^r \ln N \le N^{-r}$ and $h^{\frac{1}{r}\wedge\frac{1}{2}}\lambda_h \le N^{-1}$, provided that n is large enough. Combining this observation, (73) and (74), we get the desired bound

$$E|\widehat{\Psi}_n - E\widehat{\Psi}_n| \le C_3(N^{-r} + N^{-1} \parallel f_{h,+} \parallel_r^{r-1}). \qquad \blacksquare$$

We can now finish the proof of the theorem. We have

$$\begin{aligned} E\{|\widehat{\Phi}_n - \Phi_r[f]|\} &= E\{|\widehat{\Psi}_n^{1/r} - \Psi_r^{1/r}[f]|\} \\ &\le E\{|(\widehat{\Psi}_n)^{1/r} - (E\{\widehat{\Psi}_n\})^{1/r}|\} + |(E\{\widehat{\Psi}_n\})^{1/r} - \Psi_r^{1/r}[f_h]| + |\Phi_r[f_h] - \Phi_r[f]| \\ &= \delta_1 + \delta_2 + \delta_3. \end{aligned} \tag{75}$$

Recall that $\Phi_r(f)$ is Lipschitz continuous, with constant 1, in the norm $\|\cdot\|_r$. Therefore for all large enough values of n it holds

$$\delta_3 = |\Phi_r[f_h] - \Phi_r[f]| \le \parallel f_h(t) - f(t) \parallel_r \le Lh^\beta \le AL^{1/(2\beta+1)}(n\ln n)^{-\beta/(2\beta+1)},$$

where A depends on β only (the concluding inequality is a well-known consequence of the inclusion $f \in \Sigma_\rho(\beta, L)$). Further, the bound (64) for $j = 0$ gives us

$$\delta_2 \le |E\{\widehat{\Psi}_n\} - \Psi_r[f_h]|^{1/r} \le \max_{z\in[-1,1]} |(z_+)^r - T_N(z)|^{1/r} \le C_6N^{-1}. \tag{76}$$

We also have

$$\begin{aligned}\delta_1 &= E\{|\widehat{\Psi}_n^{1/r} - (E\{\widehat{\Psi}_n\})^{1/r}|\} \\ &\leq \left(E\left\{|\widehat{\Psi}_n - E\{\widehat{\Psi}_n\}\right\}|\right)^{1/r} 1_{\Phi_r[f_h]\leq 2C_6N^{-1}} + \frac{E\{|\widehat{\Psi}_n - E\{\widehat{\Psi}_n\}|}{(E\{\widehat{\Psi}_n)^{(r-1)/r}} 1_{\Phi_r[f_h]>2C_6N^{-1}}.\end{aligned}$$

We conclude from (76) that in the case of $\Phi_r[f_h] > 2C_6N^{-1}$ it holds

$$(E\{\widehat{\Psi}_n\})^{1/r} \geq \Phi_r[f_h] - C_6N^{-1} \geq \Phi_r[f_h]/2. \tag{77}$$

Finally, by Lemma 5.5 δ_1 admits the upper bound

$$\delta_1 \leq C_7 \left(N^{-1} + \frac{N^{-1}(\Phi_r[f_h])^{r-1} + N^{-r}}{(\Phi_r[f_h])^{r-1}} 1_{\Phi_r[f_h]>2C_6N^{-1}}\right) \leq C_8N^{-1},$$

the concluding inequality being readily give by (15). Combining the latter bound and (77), (76) and (75), we arrive at the result announced in Theorem. ∎

References

[1] Birgé, L. and Massart, P. (1995). Estimation of integral functionals of a density. *Ann. Statist.* **23**, 11–29.

[2] Bickel, P.J. and Ritov, Y. (1988) Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhya*, **50**, 381–393.

[3] Donoho, D.L. and Nussbaum, M. (1990). Minimax quadratic estimation of a quadratic functional. *J. Complexity*, **6**, 290–323.

[4] Donoho, D.L. and Liu, R.C. (1991). Geometrizing rate of convergence, III. *Ann. Statist.* **19**, 668–701.

[5] Efroimovich, S. and Low, M. (1996). On Bickel and Ritov's conjecture about adaptive estimation of the integral of the square of density derivative. *Ann. Statist.* **24**, 682–686.

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

[7] Fan, J. (1991). On the estimation of quadratic functionals. *Ann. Statist.* **19**, 1273–1294.

[8] Hall, P. and Marron, J.S. (1987). Estimation of integrated squared density derivatives. *Statist. and Prob. Letters*, **6**, 109–115.

[9] Khasminski, R.Z. and Ibragimov, I.A. (1979). On the nonparametric estimation of functionals. In *Proceedings of the Second Prague Symp. on Asymptotic Statistics* (P.Mandl and M.Huskova eds.) North-Holland, Amsterdam. 41–51.

[10] Ibragimov, I.A. and Khasminski, R.Z. (1979). *Asymptotic Estimation Theory.* (in Russian). Nauka. Moscow. English translation: *Statistical Estimation: Asymptotic Theory.* (1981) Springer. Berlin, Heidelberg, New York.

[11] Khasminski, R.Z. and Ibragimov, I.A. (1980). Some estimation problems for stochastic differential equations. *Lecture Notes Control Inform. Sci.* Springer, New York. **25**, 1–12.

[12] Ibragimov, I.A. and Khasminski, R.Z. (1980). On the estimation of distribution density. *Zap. Nauch. Sem. LOMI*, **98**, 61–85.

[13] Ibragimov, I.A. and Nemirovski and Khasminski, R.Z. (1986). Some problems on nonparametric estimation in Gaussian white noise. *Theory Probab. Appl.*, **31**, 391–406.

[14] Ibragimov, I.A. and Khasminski, R.Z. (1987). Estimation of linear functionals in Gaussian noise. *Theory Probab. Appl.*, **32**, 30–39.

[15] Ibragimov, I.A. and Khasminski, R.Z. (1991). Asymptotic normal families of distributions and effective estimation. *Ann. Statist.*, **19**, 1681–1724.

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

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

[18] Kerkyacharian, J. and Picard, D. (1996). Estimating nonquadratic functionals of a density using Haar basis. *Ann: Statist.*, **24**, 485–508.

[19] Korostelev, A.P. (1990). On the accuracy of estimation of non-smooth functionals of regression. *Theory Probab. Appl.*, **35**, 768–770.

[20] Korostelev, A.P. and Tsybakov, A.B. (1994). *Minimax Theory of Image Reconstruction.* Lecture Notes in Statist. Springer, New York.

[21] Koshevnik, Yu. and Levit, B. Ya. (1976). On a nonparametric analogue of the information matrix. *Theory Probab. Appl.*, **21**, 738–753.

[22] Laurent, B. (1996) Efficient estimation of integral functionals of a density. *Ann: Statist.*, **24**, 659–682.

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

[24] Lepski, O. and Spokoiny, V. (1995). Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. *Bernoulli*, to appear.

[25] Lepski, O., Nemirovski, A., and Spokoiny, V. (1999). On estimation of the L_r norm of a regression function. *Probability Theory and Related Fields* **v.** 113, 221-253.

[26] Levit, B.Ya. (1974). On optimality of some statistical estimates. In *Proceedings of the Prague Symp. on Asymptotic Statistics* (J. Hajek, ed.) Univ. Karlova, Prague. **2**, 215–238.

[27] Levit, B.Ya. (1975). Efficiency of a class of nonparametric estimates. *Theory Probab. Appl.*, **20**, 738–754.

[28] Levit, B.Ya. (1978). Asymptotically efficient estimation of nonlinear functionals. *Problems info. Transmission*, **14**, 65–72.

[29] Rudin, W. (1966) *Real and Complex Analysis* McGraw-Hill.

[30] Spokoiny, V. (1996). Adaptive hypothesis testing using wavelets. *Annals of Statistics*, **26**.