Sharp Adaptive Nonparametric Testing for Sobolev Ellipsoids

PENGSHENG JI AND MICHAEL NUSSBAUM

University of Georgia and Cornell University

Abstract

We consider testing for presence of a signal in Gaussian white noise with intensity \( n^{-1/2} \), when the alternatives are given by smoothness ellipsoids with an \( L_2 \)-ball of (squared) radius \( \rho \) removed. It is known that, for a fixed Sobolev type ellipsoid \( \Sigma(\beta, M) \) of smoothness \( \beta \) and size \( M \), a squared radius \( \rho \approx n^{-4\beta/(4\beta+1)} \) is the critical separation rate, in the sense that the minimax error of second kind over \( \alpha \)-tests stays asymptotically between 0 and 1 strictly (Ingster \[22\]). In addition, Ermakov \[9\] found the sharp asymptotics of the minimax error of second kind at the separation rate. For adaptation over both \( \beta \) and \( M \) in that context, it is known that a log log-penalty over the separation rate for \( \rho \) is necessary for a nonzero asymptotic power. Here, following an example in nonparametric estimation related to the Pinsker constant, we investigate the adaptation problem over the ellipsoid size \( M \) only, for fixed smoothness degree \( \beta \). It is established that the sharp risk asymptotics can be replicated in that adaptive setting, if \( \rho \to 0 \) more slowly than the separation rate. The penalty for adaptation here turns out to be a sequence tending to infinity arbitrarily slowly.

1 Introduction and main result

Consider the Gaussian white noise model in sequence space, where observations are

\[ Y_j = f_j + n^{-1/2} \xi_j, \quad j = 1, 2, ..., \tag{1} \]

with unknown, nonrandom signal \( f = (f_j)_{j=1}^\infty \), and noise variables \( \xi_j \) which are i.i.d. \( N(0,1) \). We intend to test the null hypothesis of “no signal” against nonparametric alternatives described as follows. For some \( \beta > 0 \) and \( M > 0 \), let \( \Sigma(\beta, M) \) be the set of sequences

\[ \Sigma(\beta, M) = \{ f = (f_j)_{j=1}^\infty : \sum_{j=1}^\infty j^{2\beta} f_j^2 \leq M \}; \]

Supported in part by NSF Grant DMS-08-05632
2000 Mathematics Subject Classification: 62G10, 62G20
Key words and phrases: minimax hypothesis testing, nonparametric signal detection, sharp asymptotic adaptivity, moderate deviation
this might be called a Sobolev type ellipsoid with smoothness parameter \( \beta \) and size parameter \( M \). Consider further the complement of an open ball in the sequence space \( l_2 \): if \( \| f \|_2^2 = \sum_{j=1}^{\infty} f_j^2 \) is the squared norm then

\[
B_\rho = \{ f \in l_2 : \| f \|_2^2 \geq \rho \}.
\]

Here \( \rho^{1/2} \) is the radius of the open ball; by an abuse of language we call \( \rho \) itself the “radius”. We study the hypothesis testing problem

\[ H_0 : f = 0 \quad \text{against} \quad H_a : f \in \Sigma(\beta, M) \cap B_\rho. \]

Assuming that \( n \to \infty \), implying that the noise size \( n^{-1/2} \) tends to zero, we expect that for a fixed radius \( \rho \), consistent \( \alpha \)-testing in that setting is possible. More precisely, there exist \( \alpha \)-tests with type II error tending to zero uniformly over the nonparametric alternative \( f \in \Sigma(\beta, M) \cap B_\rho \). If now the radius \( \rho = \rho_n \) tends to zero as \( n \to \infty \), the problem becomes more difficult and if \( \rho_n \to 0 \) too quickly, all \( \alpha \)-tests will have the trivial asymptotic (worst case) power \( \alpha \). According to a fundamental result of Ingster [22] there is a critical rate for \( \rho_n \), the so-called separation rate

\[
\rho_n \asymp n^{-4\beta/(4\beta+1)} \quad (2)
\]

at which the transition in the power behaviour occurs. More precisely, consider a (possibly randomized) \( \alpha \)-test \( \phi_n \) in the model (1) for null hypothesis \( H_0 : f = 0 \), that is, a test fulfilling

\[
E_{n, \phi_n} \leq \alpha \quad \text{where} \quad E_{n,f}(\cdot) \text{ denotes expectation in the model (1)}.
\]

For given \( \phi_n \), we define the worst case type II error over the alternative \( f \in \Sigma(\beta, M) \cap B_\rho \) as

\[
\Psi(\phi_n, \rho, \beta, M) := \sup_{f \in \Sigma(\beta, M) \cap B_\rho} (1 - E_{n,f} \phi_n). \quad (3)
\]

The search for a best \( \alpha \)-test in this sense leads to the minimax type II error

\[
\pi_n(\alpha, \rho, \beta, M) := \inf_{\phi_n : E_{n,\phi_n} \leq \alpha} \Psi(\phi_n, \rho, \beta, M). \quad (4)
\]

An \( \alpha \)-test which attains the infimum above for a given \( n \) is minimax with respect to type II error. Ingster’s separation rate result can now be formulated as follows: if \( \rho_n \asymp n^{-4\beta/(4\beta+1)} \) and \( 0 < \alpha < 1 \) then

\[
0 < \lim \inf_n \pi_n(\alpha, \rho_n, \beta, M) \quad \text{and} \quad \lim \sup_n \pi_n(\alpha, \rho_n, \beta, M) < 1 - \alpha.
\]

Moreover, if \( \rho_n \gg n^{-4\beta/(4\beta+1)} \) then \( \pi_n(\alpha, \rho_n, \beta, M) \to 0 \), and if \( \rho_n \ll n^{-4\beta/(4\beta+1)} \) then \( \pi_n(\alpha, \rho_n, \beta, M) \to 1 - \alpha \).

These minimax rates in nonparametric testing, presented here in the simplest case of an \( l_2 \)-setting, have been extended in two ways. In the first of these, Ermakov [9] found the exact asymptotics of the minimax type II error \( \pi_n(\alpha, \rho, \beta, M) \) (equivalently, of the maximin power) at the separation rate. The shape of that result and its derivation from an underlying Bayes-minimax theorem on ellipsoids exhibit an analogy to the Pinsker constant in nonparametric estimation. In another direction, Spokoiny [35] considered the adaptive version of the minimax nonparametric testing problem, where both \( \beta \) and \( M \) are unknown, and showed that the rate at which \( \rho_n \to 0 \) has to be slowed down by a \( \log \log n \)-factor if nontrivial asymptotic
power is to be achieved. Thus an “adaptive minimax rate” was specified, analogous to Ingster’s nonadaptive separation rate \(2\), where the additional \(\log \log n\) factor is interpreted as a penalty for adaptation. However this result did not involve a sharp asymptotics of type II error in the sense of \(9\).

It is noteworthy that in nonparametric estimation over \(f \in \Sigma(\beta, M)\) with \(l_2\)-loss (as opposed to testing), where the risk asymptotics is given by the Pinsker constant, there is a multitude of results showing that adaptation is possible with neither a penalty in the rate nor in the constant, cf. Efromovich and Pinsker \(8\), Golubev \(17\), \(18\), Tsybakov \(36\). The present paper deals with the question of whether the sharp risk asymptotics for testing in the sense of \(9\) can be reproduced in an adaptive setting, in the context of a possible rate penalty for adaptation.

Let us present the well known result on sharp risk asymptotics for testing in the nonadaptive setting. Let \(\Phi\) be the distribution function of the standard normal, and for \(\alpha \in (0, 1)\) let \(z_\alpha\) be the upper \(\alpha\)-quantile, such that \(\Phi(z_\alpha) = 1 - \alpha\). Write \(a_n \gg b_n\) (or \(b_n \ll a_n\)) iff \(b_n = o(a_n)\), and \(a_n \sim b_n\) iff \(\lim_n a_n/b_n = 1\).

**Proposition 1** (Ermakov \(9\)) Suppose \(\alpha \in (0, 1)\) and that the radius \(\rho_n\) tends to zero at the separation rate, more precisely
\[
\rho_n \sim c \cdot n^{-\beta/(4\beta+1)}
\]
for some constant \(c > 0\).

(i) For any sequence of tests \(\phi_n\) satisfying \(E_{n,0}\phi_n \leq \alpha + o(1)\) we have
\[
\Psi(\phi_n, \rho_n, \beta, M) \geq \Phi(z_\alpha - \sqrt{A(c, \beta, M)/2}) + o(1) \text{ as } n \to \infty,
\]
where
\[
A(c, \beta, M) = A_0(\beta)M^{-1/(2\beta)}c^{2+1/2\beta}
\]
and \(A_0(\beta)\) is Ermakov’s constant
\[
A_0(\beta) = \frac{2(2\beta + 1)}{(4\beta + 1)^{1+1/2\beta}}
\]
(ii) For every \(M > 0\) there exists a sequence of tests \(\phi_n\) satisfying \(E_{n,0}\phi_n \leq \alpha + o(1)\) such that
\[
\Psi(\phi_n, \rho_n, \beta, M) \leq \Phi(z_\alpha - \sqrt{A(c, \beta, M)/2}) + o(1).
\]

This gives the sharp asymptotics for the minimax type II error at the separation rate, analogous to the Pinsker constant \(33\) for nonparametric estimation. The optimal test attaining the bound of (ii) above, as given in \(9\), depends on \(\beta\) and \(M\). Concerning adaptivity in both of these parameters, the following result is known.
Proposition 2 (Spokoiny [35]). Let $T$ be a subset of $(0, \infty) \times (0, \infty)$ such that there exist $M > 0$, $\beta_2 > \beta_1 > 0$ and

$$T \supseteq \{ (\beta, M) : \beta_1 \leq \beta \leq \beta_2 \}.$$ 

(i) If $t_n \ll (\log \log n)^{1/2}$ and $\rho_n \sim c \cdot (n/t_n)^{-4\beta/(4\beta+1)}$, then for any $c > 0$ and any sequence of tests $\phi_n$ satisfying $E_{n,0}\phi_n \leq \alpha + o(1)$, and not depending on $\beta$ or $M$, we have

$$\sup_{(\beta, M) \in T} \Psi(\phi_n, \rho_n, \beta, M) \geq 1 - \alpha + o(1).$$

(ii) For any $\beta^* > 1/2$ and $0 < M_1 \leq M_2$, let

$$T = \{ (\beta, M) : 1/2 < \beta \leq \beta^*, M_1 \leq M \leq M_2 \}.$$ 

Then there exist a constant $c_1 = c_1(\beta^*, M_1, M_2)$ and a sequence of tests $\phi_n$ satisfying $E_{n,0}\phi_n = o(1)$ such that, if

$$\rho_n \sim c_1 \left( \frac{n}{(\log \log n)^{1/2}} \right)^{-4\beta/(4\beta+1)} \tag{7}$$

then

$$\sup_{(\beta, M) \in T} \Psi(\phi_n, \rho_n, \beta, M) = o(1). \tag{8}$$

Here the criterion to evaluate a test sequence has changed, to include the worst case type II error over a whole range of $\beta, M$. Hence the critical radius rate (7) has to be interpreted as an adaptive separation rate. It differs by a factor $(\log \log n)^{2\beta/(4\beta+1)}$ from the nonadaptive separation rate (2); this factor is an example of the well-known phenomenon of a penalty for adaptation. Furthermore, as noted in [35], a degenerate behaviour occurs here, in that both error probabilities at the critical rate tend to zero. Thus any sequence $\phi_n$ of tests fulfilling (8) should be seen as adaptive rate optimal, comparable to rate optimal tests in the nonadaptive case (that is, tests fulfilling $\lim \sup \Psi(\phi_n, \rho_n, \beta, M) < 1 - \alpha$ at $\rho_n$ given by (2)). In Ingster and Suslina [23], chap. 7, the worst case adaptive error (8) is further analyzed, with a view to a sharp asymptotics; cf. Remark 2 below for a discussion in relation to our results.

In this paper we address the question of whether an exact type II error asymptotics in the sense of [9] is possible in an adaptive setting. In our approach $\beta$ is kept fixed, while we aim for adaptation over the ellipsoid size $M$. First, we present a negative result for adaptation at the classical separation rate (2).

Theorem 1 Suppose $c > 0$, $0 < M_1 < M_2 < \infty$ and $\rho_n \sim c \cdot n^{-4\beta/(4\beta+1)}$. Then there is no test $\phi_n$ satisfying $E_{n,0}\phi_n \leq \alpha + o(1)$, not depending on $i = 1, 2$ but satisfying both relations

$$\Psi_n(\phi_n, \rho_n, \beta, M_i) \leq \Phi(z_\alpha - \sqrt{A(c, \beta, M_i)/2}) + o(1), \quad i = 1, 2.$$ 

This result states that adaptation even just over $M$ is impossible at the separation rate. Instead, we enlarge the radius slightly and examine how the minimax error approaches zero.


To be specific, we replace the constant $c$ in $\rho_n \sim c \cdot n^{-4\beta/(4\beta+1)}$ by a sequence $c_n$ tending to infinity slowly. In that case the minimax type II error bound of Proposition 1, namely $\Phi(z_\alpha - \sqrt{A(c, \beta, M)/2})$ will tend to zero (since $A(c, \beta, M)$ as defined in (5) contains a factor $c^{2+1/(2\beta)}$). When the log-asymptotics of this error probability is considered, as in moderate and large deviation theory, it turns out that adaptation to Ermakov’s constant is possible.

**Theorem 2** Assume $c_n \to \infty$ and $c_n = o(n^K)$ for every $K > 0$. If $\rho_n = c_n \cdot n^{-4\beta/(4\beta+1)}$ then there exists a test $\phi_n$ not depending on $M$ such that

$$E_{n,0}\phi_n \leq \alpha + o(1),$$

and for all $M > 0$

$$\limsup_n \frac{1}{c_n^{2+1/(2\beta)}} \log \Psi(\phi_n, \rho_n, \beta, M) \leq -\frac{A_0(\beta)M^{-1/(2\beta)}}{4}. $$

However now, since the optimality criterion has been changed, a formal argument is needed that no $\alpha$-test can be better in the sense of the log-asymptotics for the error of second kind. Such a result is implied by Theorem 3 in Ermakov [11], where the nonadaptive sharp asymptotics is studied in a setting where $\rho_n = c_n \cdot n^{-4\beta/(4\beta+1)}$ with $c_n \to \infty$, hence type II error probability tends to zero.

**Proposition 3** Under the assumptions of the previous theorem, any test $\phi_n$ (possibly depending on $M$) satisfying $E_{n,0}\phi_n \leq \alpha + o(1)$ also fulfills

$$\liminf_n \frac{1}{c_n^{2+1/(2\beta)}} \log \Psi(\phi_n, \rho_n, \beta, M) \geq -\frac{A_0(\beta)M^{-1/(2\beta)}}{4} \quad (9)$$

This result is implied by Theorem 3 in [11], and hence the proof is omitted.

To further discuss the context of the main results, we note the following points.

**Remark 1** Logarithmic vs. strong asymptotics. In [11] it is also shown that, for nonadaptive testing where $\rho_n = c_n \cdot n^{-4\beta/(4\beta+1)}$, $c_n \to \infty$, the lower bound (9) is attainable, so that the minimax type II error defined by (4) satisfies

$$\log \pi_n(\alpha, \rho_n, \beta, M) \sim -\frac{1}{4}A(c_n, \beta, M). \quad (10)$$

This holds as long as $\rho_n \ll n^{-2\beta/(2\beta+1)}$. Moreover if additionally $\rho_n \ll n^{-3\beta/(3\beta+1)}$ then the log-asymptotics (10) can be strengthened to

$$\pi_n(\alpha, \rho_n, \beta, M) \sim \Phi(z_\alpha - \sqrt{A(c_n, \beta, M)/2}). \quad (11)$$

Results (10) and (11) have been obtained within a framework of efficient inference for moderate deviation probabilities, cf. Ermakov [10], [12]. Recall that in our setting $c_n = o(n^K)$.
for every $K > 0$, so that the strong asymptotics (11) holds in the nonadaptive setting. It is an open question whether an adaptive analog of (11) holds.

For standardized sums $T_n$ of independent random variables, if $\{T_n > x_n\}$ is a large or moderate deviation event, theorems on the relative error caused by replacing the exact distribution of $T_n$ by its limiting distribution are sometimes called strong large or moderate deviation theorems to distinguish them from first order results on log

Remark 2 Sharp asymptotics with both $\beta, M$ unknown. The adaptivity result of Spokoiny [33], discussed in Proposition 2 about the rate penalty for adaptation $(\log \log n)^{2\beta/(4\beta+1)}$, does not provide a sharp risk asymptotics in the sense of either Proposition 1 or our Theorems 1 and 2. Some results in this direction are presented in section 7.1.3 of Ingster and Suslina [23]. To clarify the relation to our setting where $\beta$ is fixed and adaptivity refers to the size parameter $M$, let us discuss these results here.

Let us first reformulate the result of Proposition 1 (that is [9]) for known $\beta, M$ in a certain dual way, where a given type II error is prescribed and it is shown to be attainable on a radius sequence $\rho_n$ which then varies with $\beta, M$. Suppose $\alpha \in (0, 1)$ and $d > 0$ are given, and suppose the radius $\rho_n$ satisfies

$$\rho_n^{(4\beta+1)/4\beta} \sim n^{-1} A_1(\beta) M^{1/4\beta} d$$

where $A_1(\beta) = (A_0(\beta)/2)^{-1/2}$, and $A_0(\beta)$ is given by (6). Then for any sequence of tests $\phi_n$ satisfying $E_{n,0}\phi_n \leq \alpha + o(1)$ we have

$$\Psi(\phi_n, \rho_n, \beta, M) \geq \Phi(z_\alpha - d) + o(1) \text{ as } n \to \infty,$$

and there is a sequence $\phi_n$ (depending on $\beta, M$) attaining this lower bound. This follows directly from Proposition 1 by setting $d = \sqrt{A(c, \beta, M)/2}$ and solving for $c$.

In the setting of [23], the smoothness parameter $\beta$ varies over a range $[\beta_1, \beta_2]$, as in Proposition 2. To state the lower asymptotic risk bound, assume that $0 < \beta_1 < \beta_2$, that $M > 0$ is fixed and define

$$\mathcal{T} = \{((\beta, M) : \beta_1 \leq \beta \leq \beta_2\}.$$

Let $D \in R$ be arbitrary and define a radius sequence $\rho_{n,\beta,M}$ by

$$\rho_{n,\beta,M}^{(4\beta+1)/4\beta} = n^{-1} A_1(\beta) M^{1/4\beta} \left((2 \log \log n)^{1/2} + D\right). \tag{12}$$

The lower asymptotic risk bound (a variation of Theorem 7.1 in [23]) can then be formulated as follows. For any sequence of tests $\phi_n$ satisfying $E_{n,0}\phi_n \leq \alpha + o(1)$ we have

$$\sup_{(\beta, M) \in \mathcal{T}} \Psi(\phi_n, \rho_{n,\beta,M}, \beta, M) \geq (1 - \alpha) \Phi(-D) + o(1). \tag{13}$$

Note in this setting, the test sequences $\phi_n$ are assumed not to depend on $\beta$ but the radius $\rho_{n,\beta,M}$ does. Note that part (i) of Proposition 2 is implied by (13) by letting $D \to -\infty$.

As to the attainability of this bound, the test provided in section 7.3 of [23] depends on $M$. Indeed in [24] observations are assumed to be $X_j = v_j + \xi_j$, where $\xi_j$ are i.i.d. standard normal and $v = (v_j)_{j=1}^\infty$ satisfies restrictions $\sum_j v_j^2 \geq r^2$, $\sum_j j^{2\beta} v_j^2 \leq R^2$ where $R \to \infty$ and $r/R \to 0$ (the "power norm" case in the book, where $p = q = 2, s = \beta$; also $r$ is $\rho$ in
This observation model is equivalent to ours upon setting \( R^2 = nM \), \( r^2 = np \), and then \( Y_j = n^{-1/2}X_j \), \( f_j = n^{-1/2}u_j \). The reasoning provided in section 7.3.2 of \[23\] makes it clear that the test constructed uses solutions of an extremal problem under restrictions 
\[
\left\{ v : \sum_j v_j^2 \geq r^2, \sum_j j^{2\beta}v_j^2 \leq R^2 \right\}
\]
where \( r^2 = n\rho_n,\beta,M \) with \( \rho_n,\beta,M \) from \[12\] and \( \beta \) is from a certain grid of values in \((\beta_1,\beta_2)\). Since in particular \( R = n^{1/2}M^{1/2} \), it turns out that the estimator depends on \( M \), though it has been made independent of \( \beta \in (\beta_1,\beta_2) \). A version of such results for \( \alpha_n \)-tests with \( \alpha_n \to 0 \) is given in \[24\].

It should be noted that adaptation to \( \beta \) only, with \( M \) remaining fixed, does not have a practical interpretation in the context of smooth functions. Thus the problem of a sharp risk bound for adaptation to \((\beta,M)\) remains open in nonparametric testing; for the analogous problem in the estimation case (regarding the Pinsker bound), solutions have been presented by Golubev \[18\] and Tsybakov \[36\], sec 3.7.

**Remark 3** *The detection problem.* Instead of focussing on the worst case type II error \( \Psi(\phi_n,\rho,\beta,M) \) of \( \alpha \)-tests \( \phi_n \), one may consider minimization of the sum of errors, that is of \( E_n,\rho \phi_n + \Psi(\phi_n,\rho,\beta,M) \), over all tests \( \phi_n \). That has been called the detection problem in the literature; in \[23\] this problem is largely treated in parallel to the one for \( \alpha \)-tests. There and in \[25\] one finds the analog of the nonadaptive sharp asymptotics of Proposition \[1\]. It may be conjectured that analogs of our Theorems \[1\] and \[2\] concerning adaptivity hold there as well.

**Remark 4** *The plug-in method.* In the present setting, where the degree of smoothness \( \beta \) is fixed but the ellipsoid size \( M \) is unknown, a natural approach to adaptivity is to try to estimate \( M \) and use a plug-in method. However uniformly consistent estimators of \( M \) do not exist (since the unit ball in \( L_2 \) is not compact), hence for minimax optimality, such a straightforward argument fails. In the estimation setting, the solution found by Golubev \[17\] is to apply, for a biased estimator of \( M \), the same saddle point reasoning which lies at the heart of the Pinsker \[23\] result about minimax optimal estimation. The paper \[17\] concerns the continuous white noise model indexed by \( t \in [0,1] \), and the adaptivity there incorporates two local aspects: one with respect to time \( t \in [0,1] \) and the other with respect to a local variant of Sobolev smoothness classes. For more discussion cf. \[20\]. Our result here is the analog of the one by Golubev \[17\] for estimation, but in testing it turns out that adaptivity is possible only in conjunction with a tail probability (moderate deviation) approach. To further clarify the connection to adaptive estimation, in section \[5.1\] we present a short outline of the result of \[17\] in a simplified setting.

**Remark 5** *Quadratic functionals.* In the literature it has been noted that the nonparametric testing problem with an \( l_2 \)-ball removed is related to the estimation problem of the quadratic functional \( Q(f) = \|f\|_2^2 \). In particular, it is known that the optimal separation rate for testing \( \rho_n^{1/2} \asymp n^{-2\beta/(4\beta+1)} \) (comp. \[2\]) and the minimax optimal rate for estimating \( Q(f) \) over \( \Sigma(\beta,M) \) coincide if \( 0 < \beta < 1/4 \), but if \( \beta \geq 1/4 \) then the latter rate becomes \( n^{-1/2} \) (the so-called elbow effect; cf. Klemelä \[29\] and references therein). Butucea \[2\] gave a unified argument for lower bounds in the estimation and testing cases when rates coincide. As far as adaptive estimation rates for \( Q(f) \) are concerned, the logarithmic penalty factor in the ”irregular” case \( 0 < \beta < 1/4 \) has been established in \[7\]. In \[6\] it has been shown that at the
point $\beta = 1/4$ the optimal adaptive rate is $n^{-1/2}c_n$ where $c_n \to \infty$ slower than any power function of $n$, and for $\beta > 1/4$, there is no adaptation penalty on the optimal rate $n^{-1/2}$. In the case $0 < \beta < 1/4$, the only sharp adaptive minimaxity result for estimation of $Q(f)$ we are aware of is in [26]; it concerns a case where the $l_2$-Sobolev class $\Sigma(\beta, M)$ is replaced by an $l_p$-smoothness body with $p = 4$.

**Remark 6** *The sup-norm problem.* Lepski and Tsybakov [29] proved a sharp minimax result in testing when the alternative is a Hölder class (denoted $H(\beta, L)$, say) with an sup-norm ball removed, which is a testing analog of the minimax estimation result of Korostelev [27] and also a sup-norm analog of Ermakov [9]. For adaptive minimax estimation with unknown $(\beta, L)$ in the sup-norm case cf. [19]; for the testing case where $\beta$ is given, Dümbgen and Spokoiny [5] established a sharp adaptivity result with respect to the size parameter $L$ only. The result in Theorem 2.2. of [5] can be seen as a analog of the one given here, although the methodology in the sup-norm case is much different due to the connection to deterministic optimal recovery, cf. [29]. The case of unknown $(\beta, L)$ seems to be an open problem in the sup-norm testing case, with regard to sharp minimaxity, although in [5] a test is given which is adaptive rate optimal without a log log $n$-type penalty. Rohde [34] discusses the sup-norm case for regression with nongaussian errors, combining methods of [5] with ideas related to rank tests.

**Remark 7** *Density, regression and other models.* The phenomenon of the log log $n$-type penalty in the rate for adaptation when an $L_2$-ball is removed, as found by [35], has also been established in a discrete regression model [15], and in density models with direct and indirect observations [13], [3]. For a review of adaptive separation rates and further results in a Poisson process model cf. [14].

The structure of the paper is as follows. In Section 2, we discuss the background, for the nonadaptive setting, of the sharp asymptotic minimaxity result for testing of Ermakov [9] and its analogy to the Pinsker [33] constant. In Section 3 we present the proof of Theorem 1 about the lower bound (the necessary penalty) for adaptation and in Section 4 Theorem 2 concerning attainability is proved. In an appendix (Section 5.1), we present some more background for the reader, by giving a brief sketch of the estimation analog of our nonparametric testing result (Golubev [17]). Finally, Section 5.2 contains some proofs for the background Section 2.

### 2 The Bayes-minimax problem for nonparametric testing

The purpose of this expository section is to elucidate the analogy between the Pinsker constant [33] for $l_2$-estimation over ellipsoids and the constant found by Ermakov [9] for nonparametric testing over ellipsoids with an $l_2$-ball removed. We draw on the background explanation given in [23], sec. 4.1, but we focus specifically on the fact that very similar Bayes-minimax problems are at the root of the estimation and testing variants. For the theory underlying the Pinsker constant cf. [11], [31], [36].
For this exposition, we shall assume that observations are for \( j = 1, \ldots, n \); we will thus assume \( f \in \mathbb{R}^n \) and understand the sets \( \Sigma(\beta, M) \) and \( B_\rho \) accordingly, i.e. they refer only to the first \( n \) coefficients of \( f \). By \( ||\cdot|| \) and \( \langle \cdot, \cdot \rangle \) we denote euclidean norm and inner product in \( \mathbb{R}^n \). Since most expressions will depend on \( n \), for this discussion we shall often suppress dependence on \( n \) in the notation. Assume that the radius \( \rho \) tends to zero at the critical rate, that is \( \rho \sim n^{-4\beta/(4\beta+1)} \). Let \( \mathbb{R}^n_+ = [0, \infty)^n \); for a certain \( d \in \mathbb{R}^n_+ \), consider a quadratic statistic of the form \( \tilde{T} = n \sum_{j=1}^n d_j Y_j^2 \). Under \( H_0 \), we have \( E_{0,n} \tilde{T} = \sum_{j=1}^n d_j \) and \( \text{Var}_{0,n} \tilde{T} = 2 \|d\|^2 \). Since we will work with the normalized test statistic, obtained by centering and dividing by the standard deviation, it is obvious that we need only consider coefficients \( d \) fulfilling \( \|d\|^2 = 1 \). Accordingly define, for such coefficients \( d \), the statistic

\[
T = \frac{1}{\sqrt{2}} \left( \tilde{T} - \sum_{j=1}^n d_j \right). \tag{14}
\]

Under \( H_0 \), we now have \( E_{0}T = 0 \) and \( \text{Var}_{0}T = 1 \). We will consider quadratic tests

\[
\psi_d = \mathbf{1}\{T > z_\alpha\}. \tag{15}
\]

A further condition on \( d \) is imposed by requiring \( d \in D \), a set which is defined for a given sequence \( \delta = (\log n)^{-1} \) as

\[
D = \{d \in \mathbb{R}^n_+ : \|d\|^2 = 1 \text{ and } \sup_j d_j^2 \leq \delta/n\rho\}. \tag{16}
\]

For any test, we are interested in the worst case type II error under the constraint \( f \in \Sigma(\beta, M) \cap B_\rho \). A monotonicity argument shows that for every \( \psi_d \), this is attained when \( \|f\|^2 \) is minimal, i.e. at \( \|f\|^2 = \rho \). It follows that for quadratic tests \( \psi_d \), we may replace the restriction \( f \in B_\rho \) by \( f \in B'_\rho \) where

\[
B'_\rho = \{f \in \mathbb{R}^n : \rho \leq \|f\|^2 \leq 2\rho\}.
\]

For \( f \in \mathbb{R}^n \) we set \( f^2 := (f_j^2)_{j=1}^n \). For \( d \in D \) and \( g \in \mathbb{R}^n_+ \) define the functional

\[
L(d, g) = \frac{n}{\sqrt{2}} \langle d, g \rangle.
\]

**Lemma 1**  
(a) Under \( H_0 \), we have \( T \sim N(0, 1) \) uniformly over \( d \in D \).
(b) The statistic \( T \) given by \( (14) \) fulfills

\[
T - L(d, f^2) \sim N(0, 1)
\]

uniformly over \( d \in D \) and \( f \in B'_\rho \).
(c) Suppose \( f \) is random such that \( f_j \sim N(0, \sigma_j^2) \) for a certain \( \sigma \in \mathbb{R}^n \). Then the statistic \( T \) given by \( (14) \) fulfills

\[
T - L(d, \sigma^2) \sim N(0, 1)
\]

uniformly over \( d \in D \) and \( \sigma \in B'_\rho \).
Denote the expectation under the model of (c) by $E^*_σ$. The lemma implies that for uniformly over $d \in D$ and $f \in \{0\} \cup (\Sigma(\beta, M) \cap B'_ρ)$

$$E_f(1 - ψ_d) = \Phi(z_α - L(d, f^2)) + o(1) \quad (17)$$

$$E^*_f(1 - ψ_d) + o(1). \quad (18)$$

In particular, all quadratic tests $ψ_d$ with $d \in D$ are asymptotic $α$-tests under $H_0 : f = 0$. To characterize the worst case error under the alternative $H_a : f \in Σ(β, M) \cap B_ρ$, we use (17) and the strict monotonicity of $Φ$ and look for a saddlepoint of the functional $L(d, f^2)$.

**Lemma 2** For $n$ large enough, there exists a saddlepoint $d_0 \in D, f_0 \in Σ(β, M) \cap B'_ρ$ of the functional $L(d, f^2)$ such that

$$L(d, f_0^2) \leq L(d_0, f_0^2) \leq L(d_0, f^2)$$

for all $d \in D$ and all $f \in Σ(β, M) \cap B'_ρ$.

The normal distribution on the signal $f$ postulated in (c) will be interpreted as a prior distribution. The next result shows that the Bayesian tests in this context are quadratic tests $ψ_d$, and in particular, if the $σ^2$ is taken at the saddlepoint ($σ^2_0 = f^2_0$) then $d \in D$, i.e. it fulfills the infinitesimality condition $d_j^2 \leq δ/np$.

**Lemma 3** (a) For any $σ^2 \in \mathbb{R}_+^n$, the Neyman-Pearson $α$-test for simple hypotheses

$$H_0 : Y_j \sim N(0, n^{-1}), j = 1, \ldots, n \quad \text{vs.}$$

$$H^*_a : Y_j \sim N(0, σ^2_j + n^{-1}), j = 1, \ldots, n$$

is equivalent to a quadratic test of form $ψ_d = 1 \{T > t\}$ where $T = \sum_{j=1}^n d_j Y_j^2, d \in \mathbb{R}_+^n, \|d\| = 1$.

(b) If $σ^2 = f^2_0$ then the pertaining $d$ is in $D$ for $n$ large enough, and $t \to z_α$.

Part (b) implies that

$$\inf_{φ : E_φ \leq α} E^*_f(1 - φ) = \inf_{d \in D} E^*_f(1 - ψ_d) + o(1). \quad (19)$$

We are now ready to present the essence of the argument underlying the result of Ermakov [9]. Recall that $π_n(α, ρ, β, M)$ denotes the minimax type II error over all $α$-tests. Denote the value of $L(d, f^2)$ at the saddlepoint

$$L_0 := L(d_0, f^2_0) = \sup_{d \in D} \inf_{f \in Σ(β, M) \cap B'_ρ} L_n(d, f^2) = \inf_{f \in Σ(β, M) \cap B'_ρ} \sup_{d \in D} L_n(d, f^2). \quad (20)$$
We begin with an $\alpha' > \alpha$ such that asymptotic $\alpha$-tests are $\alpha'$-tests for $n$ large enough. Then

$$
\pi_n(\alpha', \rho, \beta, M) = \inf_{\phi: E_0\phi \leq \alpha'} \sup_{f \in \Sigma(\beta, M) \cap B'_\rho} E_f (1 - \phi)
$$

(21)

$$
\leq \inf_{d \in D} \sup_{f \in \Sigma(\beta, M) \cap B'_\rho} E_f (1 - \psi_d)
$$

$$
= \inf_{d \in D} \sup_{f \in \Sigma(\beta, M) \cap B'_\rho} E_f (1 - \psi_d)
$$

$$
= \inf_{d \in D} \sup_{f \in \Sigma(\beta, M) \cap B'_\rho} \Phi(z_{\alpha} - L_n(d, f^2)) + o(1) \text{ [relation (17)]}
$$

$$
= \Phi(z_{\alpha} - L_n(d_0, f^2_{0j})) + o(1) \text{ [monotonicity of $\Phi$ and (20)]}
$$

$$
= \inf_{d \in D} \sup_{f \in \Sigma(\beta, M) \cap B'_\rho} \Phi(z_{\alpha} - L_n(d_0, f^2_{0j})) + o(1)
$$

(18)

The main term of the last expression is the Bayes risk for a prior distribution $f_j \sim N(0, f^2_{0j})$ in the original model $Y_j \sim N(f_j, n^{-1})$. Since $f_0 \in \Sigma(\beta, M) \cap B'_\rho$ and is extremal there, it fulfills

$$
\sum_{j=1}^n f^2_{0j} j^{2\beta} = M, \quad \sum_{j=1}^n f^2_{0j} = \rho
$$

(see the precise description of the saddlepoint $(d_0, f_0)$ in Lemma 7 below). It can therefore be shown that (as in the original Pinsker [33] result) that this prior distribution asymptotically concentrates on every set of the form $\Sigma(\beta, M(1+\varepsilon)) \cap B'_\rho(1-\varepsilon)$ for $\varepsilon > 0$. A standard reasoning by truncation shows that in this case, for a certain probability measure $G$ strictly concentrated on $\Sigma(\beta, M(1+\varepsilon)) \cap B'_\rho(1-\varepsilon)$

$$
\inf_{\phi: E_0\phi \leq \alpha} E_{f_0}^* (1 - \phi) \leq \inf_{\phi: E_0\phi \leq \alpha} \int E_f (1 - \phi) dG(f) + o(1).
$$

However, by the relation between Bayes and minimax risk

$$
\inf_{\phi: E_0\phi \leq \alpha} \int E_f (1 - \phi) dG(f) \leq \pi_n(\alpha, \rho(1-\varepsilon), \beta, M(1+\varepsilon)).
$$

(22)

Summarizing (21)-(22) we have obtained for every $\varepsilon > 0$

$$
\pi_n(\alpha(1+\varepsilon), \rho, \beta, M) \leq \Phi(z_{\alpha} - L_n(d_0, f^2_{0j})) + o(1) \leq \pi_n(\alpha, \rho(1-\varepsilon), \beta, M(1+\varepsilon)) + o(1)
$$

Below in Lemma 8 it is shown that if $\rho = c \cdot n^{-4\beta/(4\beta+1)}$, $c$ constant then

$$
L(d_0, f^2_{0j}) \sim \sqrt{A_0 M^{-1/(2\beta)} c^{2+1/(2\beta)}/2}.
$$

Since the right side is continuous in $M$ and $c$, the result of Proposition 4 follows.
3 Proof of Theorem 1

For brevity we write \( A_i = A(c, \beta, M_i), i = 1, 2 \) in this section. Assume there exists a test \( \phi_n \) not depending on on \( M \) such that

\[
E_{0,n} \phi_n \leq \alpha + o(1), \tag{23}
\]

\[
\sup_{f \in \Sigma(\beta, M_i) \cap B_\rho} E_{f,n}(1 - \phi_n) \leq \Phi(z_\alpha - \sqrt{A_i/2}) + o(1), \tag{24}
\]

for \( i = 1 \) or \( 2 \). Let \( G_{n,M_i} \) be the Gaussian prior for \( f \) with \( f_j \sim N(0, \sigma_j^2) \) independently, where

\[
\sigma_j^2(M_i) = (\lambda - \mu_j)^2 \beta, \quad j = 1, 2, \ldots
\]

and where \( \lambda \) and \( \mu \) are determined by

\[
\sum j^2 \sigma_j^2 = M_i \quad \text{and} \quad \sum \sigma_j^2 = \rho.
\]

It can be shown that \( G_{n,M_i} \) asymptotically concentrates on \( \Sigma(\beta, M_i (1 + \varepsilon)) \cap B_\rho(1 - \varepsilon) \) for any small \( \varepsilon > 0 \). Then

\[
\sup_{\Sigma(\beta, M_i (1+\varepsilon)) \cap B_\rho(1-\varepsilon)} E_{f,n}(1 - \phi_n) \geq (1 + o(1)) \cdot \int E_{f,n}(1 - \phi_n) G_{n,M_i}(df).
\]

Recall \( Y_j = f_j + n^{-1/2} \xi_j \). Let the joint distributions of \( (Y_j)_{0}^{\infty} \) under the priors \( G_{n,0}, G_{n,M_1} \) and \( G_{n,M_2} \) be \( Q_{0,n}, Q_{1,n} \) and \( Q_{2,n} \), respectively, i.e.,

\[
Q_{0,n} : Y_j \sim N(0, n^{-1}), \quad j = 1, 2, \ldots
\]

\[
Q_{1,n} : Y_j \sim N(0, n^{-1} + \sigma_j^2(M_1)), \quad j = 1, 2, \ldots
\]

\[
Q_{2,n} : Y_j \sim N(0, n^{-1} + \sigma_j^2(M_2)), \quad j = 1, 2, \ldots
\]

Therefore,

\[
E_{Q_{0,n}} \phi_n = E_{0,n} \phi_n.
\]

\[
E_{Q_{i,n}}(1 - \phi_n) = \int E_{f,n}(1 - \phi_n) G_{n,M_i}(df), \quad i = 1, 2.
\]

Combining these with (24) and (23) gives

\[
E_{Q_{0,n}} \phi_n \leq \alpha + o(1),
\]

\[
E_{Q_{i,n}}(1 - \phi_n) \leq \Phi \left( z_\alpha - \sqrt{A_i/2} \right) + \sup_{f \in \Sigma(\beta, M_i (1+\varepsilon)) \cap B_\rho(1-\varepsilon)} E_{f,n}(1 - \phi_n) - \sup_{f \in \Sigma(\beta, M_i) \cap B_\rho} E_{f,n}(1 - \phi_n) + o(1).
\]

Note that \( E_{f,n}(1 - \phi_n) \) is continuous in \( f \). Since \( \varepsilon \) can be arbitrarily small, we have

\[
E_{Q_{i,n}}(1 - \phi_n) \leq \Phi \left( z_\alpha - \sqrt{A_i/2} \right) + o(1), \quad i = 1, 2.
\]
The likelihood ratio of \( Q_{i,n} \) against \( Q_{0,n} \) is

\[
\frac{dQ_{i,n}}{dQ_{0,n}} = \exp \left( -\frac{1}{2} \sum_j \left( \frac{Y_j^2}{n^{-1} + \sigma_j^{2}(M_i)} - \frac{Y_j^2}{n^{-1}} \right) \right) \cdot \prod_j n^{-1} \left( \frac{n^{-1} + \sigma_j^{2}(M_i)}{n^{-1}} \right)^{1/2} \\
= \exp \left( \frac{1}{2} \sum_j \frac{n^2 \sigma_j^{2}(M_i)}{1 + n \sigma_j^{2}(M_i)} Y_j^2 \right) \cdot \prod_j \left( \frac{n^{-1} + \sigma_j^{2}(M_i)}{n^{-1}} \right)^{1/2}.
\]

Therefore, by the factorization theorem, it is seen that the bivariate vector

\[
T_n = \left( n^2 \sigma_j^{2}(M_i) (Y_j^2 - n^{-1}) \right) / \left( 1 + n \sigma_j^{2}(M_i) \right) \sqrt{2n^2 \sum_k \sigma_k^{4}(M_1)} \prod_j \frac{n^2 \sigma_j^{2}(M_2) (Y_j^2 - n^{-1})}{(1 + n \sigma_j^{2}(M_2)) \sqrt{2n^2 \sum_k \sigma_k^{4}(M_2)}}
\]

is a sufficient statistic for the family of distributions \( \{Q_{0,n}, Q_{1,n}, Q_{2,n}\} \). Write the induced family for \( T_n \) as \( \{Q_{0,n}^{T}, Q_{1,n}^{T}, Q_{2,n}^{T}\} \) and take the conditional expectation \( \phi_n^{\ast}(T_n) = E_{Q_{i,n}}(\phi_n | T_n) \).

By sufficiency the (possibly randomized) test \( \phi_n^{\ast}(T_n) \) for \( \{Q_{0,n}^{T}, Q_{1,n}^{T}, Q_{2,n}^{T}\} \) is as good as \( \phi_n \) (cf. for instance Theorem 4.66 in [30]), that is

\[
E_{Q_{0,n}^{T}} \phi_n^{\ast} = E_{0,n} \phi_n \leq \alpha + o(1),
\]

(25)

\[
E_{Q_{i,n}^{T}} (1 - \phi_n^{\ast}) = E_{Q_{i,n}} \phi_n \leq \Phi(z_\alpha - \sqrt{A_i/2}) + o(1), \quad i = 1, 2.
\]

(26)

Then we have the following lemma, which is proved later.

**Lemma 4** Under \( \{Q_{0,n}, Q_{1,n}, Q_{2,n}\} \), the law of the statistic \( T_n \) converges in total variation to \( N(0, \Sigma) \), \( N(\mu_1, \Sigma) \) and \( N(\mu_2, \Sigma) \) respectively, where

\[
\mu_1 = \left( \sqrt{A_1/2}, r \sqrt{A_1/2} \right)', \\
\mu_2 = (r \sqrt{A_2/2}, \sqrt{A_2/2})', \\
\Sigma = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}, \\
r = \left( \frac{M_1}{M_2} \right)^{1/(4\beta)} \frac{4\beta + 1 - M_1/M_2}{4\beta}.
\]

(27)

Then by the weak compactness theorem (c.f. [28], A.5.1 ), there exists a test \( \phi^{\ast} \) and a subsequence \( \phi_{n_k}^{\ast} \) such that \( \phi_{n_k}^{\ast} \) converges weakly to \( \phi^{\ast} \). Thus

\[
E_{Q_{0,n}^{T}} \phi^{\ast} \leq \alpha, \\
E_{Q_{i,n}^{T}} (1 - \phi^{\ast}) \leq \Phi(z_\alpha - \sqrt{A_i/2}), \quad i = 1, 2.
\]

For \( i = 1, 2 \) respectively, by the Neyman-Pearson lemma and some direct calculations, the right hand side of the previous inequality is the type II error of the uniformly most powerful test for \( N(0, \Sigma) \) against \( N(\mu_i, \Sigma) \). Therefore, \( \phi^{\ast} \) is a uniformly most powerful test for \( N(0, \Sigma) \) against \( \{N(\mu_1, \Sigma), N(\mu_2, \Sigma)\} \).
Note that $r$ in Lemma 4 is monotone increasing with respect to $M_1/M_2$, and then $0 < r < 1$ for $M_2 > M_1 > 0$. Thus, $\mu_1$, $\mu_2$ and the origin are not on the same line. For $i = 1, 2$ respectively, the log-likelihood ratio for $N(\mu_i, \Sigma)$ against $N(0, \Sigma)$ is $T_i^{n-1} \mu_i = T_i \cdot A_i$. Then by the necessity part of the Neyman-Pearson lemma ([28], Theorem 3.2.1), the uniformly most powerful test for $N(0, \Sigma)$ against $N(\mu_1, \Sigma)$ is $T_1$. For simplicity, we only show the result for the first coordinate of $T_n$. The proof can be extended to $T_n$ naturally. Under $Q_{0,n}$, the characteristic function of $\frac{n(Y_1^2 - 1/n)^{1/2}}{\sqrt{2}} \sim N(0,1)$ is $g(t) = \exp(-t^2/2)$. Note $g(t) = 1 - \frac{1}{2}t^2 + o(t^2)$, as $t \to 0$ and $\int |g(t)| < \infty$. The density of $T_{n,1}$ can be written as

$$p_{n}(x) = \frac{1}{2\pi} \int e^{-itx} \prod g \left( \frac{\sigma_j^2(M_1) \cdot t}{(1 + n \sigma_j^2(M_1)) \sqrt{\sum \sigma_k^4(M_1)}} \right),$$

where, by Levy’s continuity theorem, the integrand converges to $e^{-itx} \exp\{-t^2/2\}$. By splitting the integral into two parts and using dominated convergence, it can be shown that the integral converges to

$$\frac{1}{2\pi} \int e^{-itx} e^{-t^2/2} dt = \frac{e^{-x^2/2}}{\sqrt{2\pi}}.$$

Then an application of Scheffé’s theorem (cf. [37], 2.30) establishes convergence in total variation. The correlation $r$ can be calculated directly. ■

4 Proof of Theorem 2

Choose $\tilde{N}$ and $\gamma_n = o(1)$ such that

$$\gamma_n^{1/2\beta} \cdot n^{2/(4\beta+1)} \gg \tilde{N} \gg c_n^{-1/2\beta} \cdot n^{2/(4\beta+1)},$$

e.g. $\gamma_n = c_n^{-1/2}$, $\tilde{N} = c_n^{-1/3\beta} \cdot n^{2/(4\beta+1)}$. Define

$$M_0 = M_0(f) = \sum_{j=1}^{\tilde{N}} j^{2\beta} f_j^2 + \gamma_n,$$

$$N = N(M_0) = \left( \frac{(4\beta + 1) M_0}{\rho} \right)^{1/2\beta},$$

$$\tilde{\lambda} = \tilde{\lambda}(M_0) = \frac{2\beta + 1}{2\beta} \left( \frac{1}{M_0(4\beta + 1)} \right)^{1/(2\beta)} \rho^{(2\beta+1)/2\beta},$$

$$\tilde{d}_j = \tilde{d}_j(M_0) = \tilde{\lambda}[1 - (j/N)^{2\beta}]_+,$$

which all depend on the unknown $f$. Define the oracle statistic

$$T_n^* = \frac{n^2 \sum j \tilde{d}_j(M_0) Y_j^2 - n \sum j \tilde{d}_j(M_0)}{\sqrt{2n^2 \sum j \tilde{d}_j^2(M_0)}},$$
and the oracle test \( \phi_n^* = 1\{T_n^* > z_\alpha\} \). The following lemma holds; it is proved later.

**Lemma 5** Under the assumptions of Theorem 2, the oracle test \( \phi_n^* \) is an asymptotic \( \alpha \)-test and

\[
\limsup_n \frac{1}{\epsilon_n^{2+1/(2\beta)}} \log \Psi(\phi_n^*, \rho_n, \beta, M) \leq - \frac{A_0(\beta)M^{-1/2\beta}}{4}
\]

Define

\[
\hat{M} = \sum_{j=1}^{\hat{N}} (Y_j^2 - 1/n)j^{2\beta} + \gamma_n
\]

and introduce the statistic

\[
T_n = \frac{n^2 \sum \tilde{d}_j(\hat{M})Y_j^2 - n \sum \tilde{d}_j(\hat{M})}{\sqrt{2n^2 \sum \tilde{d}_j^2(\hat{M})}}
\]

and also the test

\[
\phi_n = 1\{T_n > z_\alpha\}.
\]

For \( \hat{M} \), we have the following lemma, which is proved later.

**Lemma 6** Under the assumptions of Theorem 2 we have

\[
\frac{\hat{M}}{M_0(f)} - 1 = o_p(1),
\]

uniformly for \( f \in \Sigma(\beta, M) \cap B_\rho \).

Now rewrite

\[
T_n = \sum_j \frac{\tilde{d}_j(\hat{M})}{\sqrt{\sum \tilde{d}_j^2(\hat{M})}} \cdot \frac{Y_j^2 - 1/n}{\sqrt{2n^{-2}}},
\]

where \( \tilde{d}_j(\hat{M}) = \tilde{\lambda}(1 - (j/N(\hat{M}))^{2\beta})_+ \). Since \( \tilde{\lambda} \) in the last display can be canceled, for simplicity we write \( \tilde{d}_j(\hat{M}) = (1 - (j/N(\hat{M}))^{2\beta})_+ \) from now on in this section. First, since \( N(\hat{M}) \geq N(\gamma_n) \), we have

\[
\sum \tilde{d}_j^2(\hat{M}) = \sum \left(1 - \left(\frac{j}{N(\hat{M})}\right)^{2\beta}\right)^2 + \sim N(\hat{M}) \int_0^1 (1 - t^{2\beta})_+^2 dt 
= N(\hat{M})K(\beta).
\]
Therefore,

\[ T_n = (1 + o(1)) \sum_j \frac{\hat{d}_j(\hat{M})}{\sqrt{N(M)K(\beta)}} \cdot \frac{Y_j^2 - 1/n}{\sqrt{2n^{-2}}}. \]

By Lemma 6

\[ T_n = (1 + o(1)) \sum_j \frac{\hat{d}_j(\hat{M})}{\sqrt{N(M_0(f))K(\beta)}} \cdot \frac{Y_j^2 - 1/n}{\sqrt{2n^{-2}}}. \]

At this point, make \( \hat{M} \) independent of \( Y_j^2 \) by sample splitting. Set \( n = \tau n + (1 - \tau)n \), where \( \tau \) is close to 1 but fixed, and \( n_1 = \tau n, n_2 = (1 - \tau)n \). Assume two sets of observations

\[ Y_{1j} = f_j + n_1^{-1/2} \xi_{1j}, j = 1, 2, \ldots \]  
(29)

\[ Y_{2j} = f_j + n_2^{-1/2} \xi_{2j}, j = 1, 2, \ldots \]  
(30)

Use \( \{Y_{2j}\} \) to obtain \( \hat{M} \), and now replace \( T_n \) by

\[ T_n^s = (1 + o(1)) \sum_j \frac{\hat{d}(\hat{M})}{\sqrt{N(M_0(f))K(\beta)}} \cdot \frac{Y_{2j}^2 - n^{-1}}{\sqrt{2n^{-1}}}. \]

Denote the difference of coefficients by \( \Delta_j = \hat{d}_j(\hat{M}) - \hat{d}_j(M_0(f)) \). Note the largest difference is obtained at \( j \approx \min\{N(M), N(M_0(f))\} \). Then

\[ |\Delta_j| \leq \frac{|\hat{M} - M_0(f)|}{\gamma_n} \]

uniformly for all \( j \). Note in \( T_1 \) there are at most \( C_2 \gamma_n^{-1/(2\beta)} n^{2/(4\beta + 1)} \) nonzero coefficients. Then

\[ T_n^s = (1 + o(1)) \sum_{j=1}^{C_2 \gamma_n^{-1/(2\beta)} n^{2/(4\beta + 1)}} \frac{\hat{d}_j(M_0(f))}{\sqrt{N(M_0(f))K(\beta)}} \eta_j + r_n \]

where \( \eta_j = \frac{Y_{2j}^2 - n^{-1}}{\sqrt{2n^{-1}}} \), and

\[ r_n = \sum_{j=1}^{C_2 \gamma_n^{-1/(2\beta)} n^{2/(4\beta + 1)}} \frac{\Delta_j \eta_j}{\sqrt{N(M_0(f))K(\beta)}}. \]

Under \( H_0 \), the r.v.'s \( \eta_j \) are independent of \( \hat{M} \) and \( E\eta_j = 0 \), \( \text{Var}(\eta_j) = 1 \). Thus \( \text{Var}(r_n) = ER_n^2 = EE(r_n^2|\{Y_{2j}\}) \) and

\[ E(r_n^2|\{Y_{2j}\}) = E \sum_{j=1}^{C_2 \gamma_n^{-1/(2\beta)} n^{2/(4\beta + 1)}} \frac{\Delta_j^2}{N(M_0(f))K(\beta)} \leq \frac{|\hat{M} - M_0(f)|^2}{\gamma_n^{2+1/(2\beta)}}. \]

Therefore, by the result for \( \text{Var}(\hat{M}) \) in the proof of Lemma 6

\[ \text{Var}(r_n) \leq \frac{E|\hat{M} - M_0(f)|^2}{\gamma_n^{2+1/(2\beta)}} = \frac{\text{Var}(\hat{M})}{\gamma_n^{2+1/(2\beta)}} \leq \frac{2K(\beta)\bar{N}_{1/(2\beta)}^{4\beta + 1}}{n^{2+1/(2\beta)}} + \frac{4\bar{N}_{1/(2\beta)}^{2\beta}M}{n^{2+1/(2\beta)}}. \]
where the last two terms converge to 0 by the first inequality in (28). Hence, under $H_0$, the r.v.'s $T_n$ and $T_n^*$ converge to $N(0, 1)$ in law.

Next, we consider $T_n$ or $T_n^*$ under the alternative. The worst case type II error is determined by the following quantity

$$L_n = \frac{n}{\sqrt{2}} \inf_{f \in \Sigma(\beta, M) \cap B_\rho} \frac{\sum_{j=1}^{\tilde{N}} f_j^2 \tilde{d}_j(\hat{M})}{\left(\sum \tilde{d}_j(\hat{M})\right)^{1/2}}.$$

First, since $N(\hat{M}) \geq \left(\frac{7n}{c_n}\right)^{1/(2\beta)} \cdot n^{2/(4\beta+1)} \to \infty,$

$$\tilde{d}_j^2 = \sum_{j=1}^{\tilde{N}} \left(1 - \left(\frac{j}{N}\right)^{2\beta}\right)^2 + \left(1 + o(1)\right)N \int_0^1 (1 - t^{2\beta})^2 dt = (1 + o(1))N \cdot \frac{8\beta^2}{(2\beta + 1)(4\beta + 1)}.$$  \hspace{1cm} (31)

Second, consider

$$\sum_{j=1}^{\tilde{N}} f_j^2 \tilde{d}_j(\hat{M}) = \sum_{j=1}^{\tilde{N}} f_j^2 (1 - (j/N)^{2\beta})_+.$$

Note

$$\sum_{j=1}^{\tilde{N}} f_j^2 = \sum_{j=1}^{\infty} f_j^2 - \sum_{j=\tilde{N}+1}^{\infty} f_j^2 \geq \rho - \tilde{N}^{-2\beta} M \tilde{N} \rho$$

$$= \rho \left(1 - \frac{M}{\rho \tilde{N}^{2\beta}}\right)$$

$$= \rho(1 + o(1)),$$ \hspace{1cm} (32)

where the last step is refers to the second inequality of (28). On the other hand, since $\tilde{N} \gg N$ and $N(\hat{M}) = [(4\beta+1)M_0^{-1}]^{1/(2\beta)},$

$$\sum_{j=1}^{N} f_j^2 (j/N)^{2\beta} + \sum_{j=\tilde{N}+1}^{\tilde{N}} f_j^2 \leq \sum_{j=1}^{\tilde{N}} f_j^2 (j/N)^{2\beta} \leq N^{-2\beta} M_0(\hat{f}) \leq \rho(1 + 4\beta)^{-1}.$$  \hspace{1cm} (33)

Combining (35)-(37) gives

$$\sum_{j=1}^{\tilde{N}} f_j^2 \tilde{d}_j \geq (1 + o(1))\hat{\lambda}_\rho \cdot \frac{4\beta}{4\beta + 1}.$$
Combining this with (34) gives
\[ n \sum_{j=1}^{\tilde{N}} f_j^2 \tilde{d}_j \geq (1 + o(1)) \frac{n}{\sqrt{2}} \sqrt{\frac{2(2\beta + 1)}{4\beta + 1}} \rho^2 / N \]
\[ \geq (1 + o(1)) \sqrt{\frac{(2\beta + 1)c_n^{2+1/(2\beta)}}{(4\beta + 1)^{1+1/(2\beta)}(M + \gamma_n)^{1/(2\beta)}}} \]
\[ \geq (1 + o(1)) \sqrt{\frac{1}{2} A_0(\beta)c_n^{2+1/(2\beta)} M^{-1/(2\beta)}} \]

Theorem 2 is proved.

**Proof of Lemma 5.** Rewrite
\[ T_n^* = \sum_{j} \frac{\tilde{d}_j(M_0(f))}{\sqrt{\sum \tilde{d}_j^2(M_0(f))}} \cdot \frac{Y_j^2 - 1/n}{\sqrt{2n^{-2}}} \]

Under \( H_0 \), we have \( f = 0 \), and \( M_0(f) = \gamma_n \). Since
\[ \sum [1 - (j/N)^{2\beta}]_+^2 \sim N \cdot \int_0^1 (1 - t^{2\beta})^2 dt = K(\beta) \cdot (\gamma_n/c_n)^{1/2} n^{2/(4\beta + 1)} \]
then
\[ \left| \frac{\tilde{d}_j(M_0(f))}{\sqrt{\sum \tilde{d}_j^2(M_0(f))}} \right| \leq \frac{1}{\sqrt{K(\beta) \cdot (\gamma_n/c_n)^{1/2} n^{2/(4\beta + 1)}}} = o(1) \]
uniformly for all \( j \). It can be shown that \( T_n^* \) converges to \( N(0, 1) \) in law.

By similar arguments, the worst type II error is \((1 + o(1))\Phi(z - L_n)\) where
\[ L_n = \inf_{f \in \Sigma(\beta, M) \cap B_0} \frac{n \sum f_j^2 \tilde{d}_j}{(2 \sum \tilde{d}_j^2)^{1/2}} \]

Note \( \tilde{d}_j = \tilde{d}_j(M_0(f)) \) depending on \( f \). By the second inequality of (28), we have \( \tilde{N} \gg N(M_0(f)) \) and \( \tilde{d}_j = 0 \) for \( j \geq \tilde{N} \),
\[ L_n = \frac{n}{\sqrt{2}} \inf_{f \in \Sigma(\beta, M) \cap B_0} \frac{\sum_{j=1}^{\tilde{N}} f_j^2 \tilde{d}_j}{(\sum \tilde{d}_j^2)^{1/2}}. \]

First, since \( N(M_0(f)) \geq \left( \frac{2a}{c_n} \right)^{1/(2\beta)} \cdot n^{2/(4\beta + 1)} \to \infty \) uniformly for \( f \in \Sigma(\beta, M) \cap B_0 \),
\[ \tilde{d}_j^2 = \tilde{\lambda}^2 \sum_{j=1}^{\tilde{N}} \left( 1 - (j/N)^{2\beta} \right)_+^2 \]
\[ = (1 + o(1)) \tilde{\lambda}^2 N \int_0^1 (1 - t^{2\beta})^2 dt \]
\[ = (1 + o(1)) \tilde{\lambda}^2 N \cdot \frac{8\beta^2}{(2\beta + 1)(4\beta + 1)}, \quad (34) \]
uniformly for \( f \in \Sigma(\beta, M) \cap B_\rho \). Second, consider

\[
\hat{N} \sum_{j=1}^N f_j^2 d_j = \lambda \sum_{j=1}^N f_j^2 (1 - (j/N)^{2\beta})_+ = \tilde{\lambda} \left[ \tilde{N} \sum_{j=1}^N f_j^2 - \left( \sum_{j=1}^N f_j^2 (j/N)^{2\beta} + \sum_{j=N+1}^{\tilde{N}} f_j^2 \right) \right]. \tag{35}
\]

Note

\[
\sum_{j=1}^{\tilde{N}} f_j^2 = \sum_{j=1}^{\infty} f_j^2 - \sum_{j=\tilde{N}+1}^{\infty} f_j^2 \geq \rho - \tilde{N}^{-2\beta} M \\
= \rho \left( 1 - \frac{M}{\rho N^{2\beta}} \right) = \rho(1 + o(1)), \tag{36}
\]

where the last step is due to the second inequality of (28). On the other hand, since \( \tilde{N} \gg N \) and \( N = \lfloor \rho^{-1}(4\beta + 1)M_0(f) \rfloor^{1/(2\beta)} \),

\[
\sum_{j=1}^{\tilde{N}} f_j^2 (j/N)^{2\beta} + \sum_{j=\tilde{N}+1}^{N} f_j^2 \leq \sum_{j=1}^{\tilde{N}} f_j^2 (j/N)^{2\beta} \leq N^{-2\beta} M_0(f) \\
= \rho(1 + 4\beta)^{-1} \tag{37}
\]

Combining (35)-(37) gives

\[
\sum_{j=1}^{\tilde{N}} f_j^2 d_j \geq (1 + o(1))\tilde{\lambda} \rho \cdot \frac{4\beta}{4\beta + 1}
\]

uniformly for \( f \in \Sigma(\beta, M) \cap B_\rho \). Combining this with (31) gives

\[
\frac{n}{\sqrt{2}} \sum_{j=1}^{\tilde{N}} f_j^2 d_j \geq (1 + o(1))n \left( \sqrt{2/\beta + 1} \right) \rho^2 / N \geq (1 + o(1)) \sqrt{\frac{(2\beta + 1)c_n^{2+1/(2\beta)}}{(4\beta + 1)^{1+1/(2\beta)}(M + \gamma_n)^{1/(2\beta)}}} \geq (1 + o(1)) \sqrt{\frac{(2\beta + 1)c_n^{2+1/(2\beta)}}{(4\beta + 1)^{1+1/(2\beta)}M^{1/(2\beta)}}},
\]

uniformly for \( f \in \Sigma(\beta, M) \cap B_\rho \). Therefore,

\[
L_n \geq (1 + o(1)) \sqrt{\frac{1}{2} A_0(\beta)c_n^{2+1/(2\beta)} M^{-1/(2\beta)}},
\]

and the result follows.
Proof of Lemma 6. Since
\[ \text{Var}(\hat{M}) = \sum_{j=1}^{N} \left( \frac{2}{n^2} + \frac{4f_j^2}{n} \right) j^{4\beta} \leq (1 + o(1)) \frac{2K(\beta)N^{4\beta+1}}{n^2} + \frac{4N^{2\beta}M}{n}, \]
by the first inequality of [28],
\[ \frac{\text{Var}(\hat{M})}{\gamma_n^2} = o(1) \]
uniformly for \( f \in \Sigma \cap V_\rho \). Combining with \( E\hat{M} = M_0(f) \) and using Chebyshev’s inequality give
\[ \frac{\left| \hat{M} - M_0(f) \right|}{\gamma_n} = o_p(1), \]
and then
\[ \left| \frac{\hat{M}}{M_0(f)} - 1 \right| \leq \frac{\left| \hat{M} - M_0(f) \right|}{\gamma_n} = o_p(1), \]
uniformly for \( f \in \Sigma \cap V_\rho \).

5 Appendix

5.1 Adaptive minimax estimation with known \( \beta \)

For the convenience of the reader, we sketch the modified plug-in method of Golubev [17] allowing to attain the Pinsker bound for known smoothness \( \beta \) and unknown bound \( M \), in the framework of Sobolev ellipsoids. For more comprehensive results, allowing also for unknown \( \beta \), cf. [18], [36]. Consider the estimation problem for \( f = (f_j)_{j=1}^{\infty} \), with squared \( l_2 \)-loss, in the Gaussian sequence model
\[ Y_j = f_j + n^{-1/2}\xi_j \]
with \( f \in \Sigma(\beta, M) \). With known \( \beta \) and unknown \( M \), the aim is to find an estimator which is asymptotically minimax in the sense of Pinsker [33]. For known \( M \), the optimal filter coefficients are \( (1 - \mu j^\beta)_+ \), where \( \mu \) is determined by
\[ \frac{1}{n} \sum_j j^\beta (1 - \mu j^\beta)_+ = \mu M. \]
Since
\[ \mu \sim \left( \frac{\beta \cdot n^{-1}}{M(\beta + 1)(2\beta + 1)} \right)^{\beta/(2\beta+1)}, \]
the optimal truncation index (or bandwidth) is of the order \( n^{1/(2\beta+1)} \).
Choose \( n^{1/(2\beta+1/2)} \gg \tilde{N} \gg n^{1/(2\beta+1)} \) and \( 1 \gg \gamma_n \gg \tilde{N}^{2\beta+1/2}/n \), and define
\[
M_{0,f} = \sum_{j=1}^{\tilde{N}} j^{2\beta} f_j^2 + \gamma_n.
\]
Define \( N = N(M_{0,f}) = \alpha \cdot n^{1/(2\beta+1)} M_{0,f}^{1/(2\beta+1)} \), where \( \alpha \) is a constant to be chosen. Define "oracle" filter coefficients, depending on \( f \), as
\[
d_j = d(j/N), \text{ where } d(t) = \left(1 - t^\beta\right)_+.
\]
Consider the oracle estimator \((d_j Y_j)_{1}^\infty\). Its risk is
\[
\sum (1 - d_j)^2 f_j^2 + \frac{1}{n} \sum d_j^2 = \sum_{j=1}^{\tilde{N}} (1 - d_j)^2 f_j^2 + \sum_{j > \tilde{N}} (1 - d_j)^2 f_j^2 + \frac{1}{n} \sum d_j^2 := A_1 + A_2 + A_3.
\]
To bound the terms \( A_i \), note first
\[
A_1 \leq \sup_{j \leq \tilde{N}} (1 - d_j)^2 j^{-2\beta} M_{0,f} \leq N^{-2\beta} M_{0,f} = \alpha^{-2\beta} n^{-2\beta/(2\beta+1)} (M + \gamma_n)^{1/(2\beta+1)}.
\]
Second, \( A_2 \leq \sum_{j \geq \tilde{N}} f_j^2 \leq N^{-2\beta} M = o(n^{-2\beta/(2\beta+1)}) \). Furthermore,
\[
A_3 = \frac{N}{N} \sum_{j \geq \tilde{N}} (1 - (j/N)^\beta)_+^2 = \alpha n^{-2\beta/(2\beta+1)} M_{0,f}^{1/(2\beta+1)} \int_0^\infty (1 - t^\beta)_+^2 dt (1 + o(1)) \text{ uniformly over } f \in \Sigma (\beta, M)
\]
\[
\leq \alpha n^{-2\beta/(2\beta+1)} M_{0,f}^{1/(2\beta+1)} \cdot \frac{2\beta^2}{(\beta + 1)(2\beta + 1)} (1 + o(1)).
\]
Combine these and choose \( \alpha = \left(\frac{(\beta+1)(2\beta+1)}{\beta}\right)^{1/(2\beta+1)} \), and we find that the supremal risk, over \( f \in \Sigma (\beta, M) \), of the oracle estimator is at most
\[
c(\beta) \cdot n^{-2\beta/(2\beta+1)} M_{0,f}^{1/(2\beta+1)} (1 + o(1)), \quad (38)
\]
where
\[
c(\beta) = \left(\frac{\beta}{\beta + 1}\right)^{2\beta/(2\beta+1)} \cdot (1 + 2\beta)^{1/(2\beta+1)}
\]
is the Pinsker constant.

The next step is to show that the risk (38) is also attained when the unknown \( M_{0,f} \) is replaced by an unbiased estimator. The latter is \( \hat{M}_n = \sum_{j=1}^{\tilde{N}_n} j^{2\beta} \hat{f}_j^2 + \gamma_n \), where \( \hat{f}_j^2 = y_j^2 - n^{-1} \). Then
\[
E(\hat{M}_n) = \sum_{j=1}^{\tilde{N}_n} j^{2\beta} \hat{f}_j^2 + \gamma_n = M_{0,f} \leq M + \gamma_n.
\]
\[
\text{Var}(\hat{M}) = \sum_{j=1}^{\tilde{N}} j^{4\beta} \text{Var}(Y_j^2) \\
= \sum_{j=1}^{\tilde{N}} j^{4\beta} n^{-2}(2 + 4nf_j^2) \\
= 2n^{-2} \sum_{j=1}^{\tilde{N}} j^{4\beta} + 4n^{-1} \sum_{j=1}^{\tilde{N}} j^{3\beta} f_j^2 \\
= J_1 + J_2,
\]

where the first term
\[
J_1 = 2n^{-2} \tilde{N}^{4\beta+1} \cdot \frac{1}{N} \sum_{j=1}^{\tilde{N}} \left( j/\tilde{N} \right)^{4\beta} \sim 2n^{-2} \tilde{N}^{4\beta+1} \cdot \int_0^1 x^{4\beta} dx = o(1)
\]
since \(\tilde{N} = o(n^{1/(2\beta+1/2)})\), and the second term
\[
J_2 \leq 4n^{-1} \tilde{N}^{2\beta} \sum_{j=1}^{\tilde{N}} j^{2\beta} f_j^2 \leq 4n^{-1} \tilde{N}^{2\beta} M = 4M \cdot n^{-2} \tilde{N}^{4\beta+1} = o(J_1)
\]
uniformly for \(f \in \Sigma(\beta, M)\) since \(\tilde{N} \gg n^{1/(2\beta+1)}\). Combining these gives \(\text{Var}(\hat{M}) = o(1)\) uniformly for \(f \in \Sigma(\beta, M)\). Recalling \(\gamma_n \gg \tilde{N}^{2\beta+1/2}/n\) gives
\[
\text{Var}\left( \frac{\hat{M} - M_{0,f}}{\gamma_n} \right) \sim \frac{2K n^{-2} \tilde{N}^{4\beta+1}}{\gamma_n^2} = o(1),
\]
and then
\[
\left| \frac{\hat{M}}{M_{0,f}} - 1 \right| \leq \left| \frac{\hat{M} - M_{0,f}}{\gamma_n} \right| = o_p(1)
\]
uniformly.

Finally, it can be shown that the difference between the oracle estimator \((d_jY_j)^\infty_i\) and the estimator \(\left( d(j/N(\hat{M}))Y_j \right)^\infty_i \) is negligible, i.e.
\[
E \sum_{j=1}^{\infty} \left( d(j/N(M_{0,f})) - d(j/N(\hat{M})) \right)^2 Y_j^2 = o(n^{-2\beta/(2\beta+1)}).
\]

### 5.2 Proofs for Section 2

**Proof of Lemma** 1. (a) Under the null hypothesis we have \(Y_j^2 = n^{-1}\xi_j^2\), hence \(T = \sum d_j \left( \xi_j^2 - 1 \right) / \sqrt{\mathcal{N}}\). Then it follows from (16) and \(np \to \infty\) that the CLT infinitesimal condition
\[
\sup_j d_j^2 = o(1)
\]
holds uniformly over $d \in \mathcal{D}$, proving the assertion.

(b) Since $Y_j^2 = f_j^2 + 2n^{-1/2} f_j \xi_j + n^{-1} \xi_j^2$, we have

$$T = \frac{1}{\sqrt{2}} \sum d_j \left(n f_j^2 + 2n^{1/2} f_j \xi_j + (\xi_j^2 - 1)\right), \quad (39)$$

$$T - L(d, f) = \frac{1}{\sqrt{2}} \sum d_j \left(2n^{1/2} f_j \xi_j + (\xi_j^2 - 1)\right). \quad (40)$$

An easy calculation gives

$$\text{Var}_f T = \frac{1}{2} \sum d_j^2 (4n f_j^2 + 2) = 1 + 2n \sum d_j f_j^2$$

where in view of (16) we have for $f \in B'_\rho$

$$n \sum d_j^2 f_j^2 \leq \delta \rho^{-1} \sum f_j^2 \leq 2 \delta = o(1).$$

Consequently, $\text{Var}_f T \to 1$ uniformly. Now the CLT infinitesimality condition on the sum (40) amounts to

$$\sup_j d_j^2 (n f_j^2 + 1) = o(1). \quad (41)$$

For $f \in B'_\rho$ we have $f_j^2 \leq 2 \rho$, hence in view of (16)

$$d_j^2 (n f_j^2 + 1) \leq d_j^2 (2n \rho + 1) \leq 2 \delta$$

for $n$ sufficiently large. Hence (41) is fulfilled uniformly over $d \in \mathcal{D}$ and $f \in B'_\rho$, and the claim follows.

(c) Set $f_j \sim N(0, \sigma_j^2)$; then in view of (39)

$$T - L(d, \sigma) = \frac{1}{\sqrt{2}} \sum d_j \left(2n^{1/2} f_j \xi_j + (\xi_j^2 - 1)\right) + \frac{n}{\sqrt{2}} \sum d_j \left(f_j^2 - \sigma_j^2\right). \quad (42)$$

An easy calculation gives

$$\text{Var}_f T = \frac{1}{2} \sum d_j^2 (4n \sigma_j^2 + 2) + n \sum d_j^2$$

$$= 1 + n \sum d_j^2 (2\sigma_j^2 + \sigma_j^4)$$

where in view of (16) we have for $\sigma \in B'_\rho$

$$n \sum d_j^2 \sigma_j^2 \leq \delta \rho^{-1} \sum \sigma_j^2 \leq 2 \delta = o(1),$$

$$n \sum d_j^2 \sigma_j^4 \leq 2 \rho n \sum d_j^2 \sigma_j^2 \leq 4 \rho \delta = o(1).$$

Consequently, $\text{Var}_f T \to 1$ uniformly. Now the infinitesimality condition on the sum (42) amounts to

$$\sup_j d_j^2 (1 + n \sigma_j^2 + n \sigma_j^4) = o(1). \quad (43)$$
For $\sigma \in B'_\rho$, we have $\sigma^2 \leq 2\rho$, hence in view of (16)

$$d_j^2 (1 + n\sigma^2_j + n\sigma^4_j) \leq d_j^2 (1 + n\rho + n\rho^2) \leq 3\delta$$

for $n$ sufficiently large. Hence (43) is fulfilled uniformly over $d \in D$ and $\sigma \in B'_\rho$, and the claim follows.  

**Proof of Lemma 2.** Let $\tilde{D}$ be defined as $D$ in (16) but with condition $\|d\|^2 = 1$ replaced by $\|d\|^2 \leq 1$. Then, since $L(d, f)$ is linear in $d$, for every $\tilde{d} \in \tilde{D}$ there is a $d \in D$ such that $L(\tilde{d}, f^2) \leq L(d, f^2)$ for every $f$. Hence it suffices to prove the claim for $D$ replaced by the compact convex set $\tilde{D}$. The restriction $f \in \Sigma(\beta, M) \cap B'_\rho$ is equivalent to $f^2$ being in the set

$$\left\{ g \in \mathbb{R}_+^n : \sum g_j j^{2\beta} \leq M, \rho \leq \sum g_j \leq 2\rho \right\}$$

which is convex and compact (and nonempty for large enough $n$ since $\rho \to 0$). The functional $L$ is bilinear in $d$ and $f^2$; the standard minimax theorem now furnishes the result.  

**Lemma 7** For $n$ large enough, the saddlepoint $d_0, f_0$ of Lemma 2 is given by

$$d_0 = \frac{f_0^2}{\|f_0\|}, \quad f_{0,j}^2 = \left( \lambda - \mu j^{2\beta} \right)_+, \quad j = 1, \ldots, n$$

where $\lambda, \mu$ are the unique positive solutions of the equations

$$\sum_{j=1}^n j^{2\beta} \left( \lambda - \mu j^{2\beta} \right)_+ = M, \quad \sum_{j=1}^n \left( \lambda - \mu j^{2\beta} \right)_+ = \rho. \quad (45)$$

The value of $L$ at the saddlepoint is

$$L_0 = L(d_0, f_0) = \frac{n}{\sqrt{2}} \|f_0^2\|. \quad (46)$$

**Proof.** Ignore initially the restriction $\sup_j d_j^2 \leq \delta/n\rho$ and consider maximizing $L(d, f^2)$ in $d$ for given $f$. Under the sole restriction $\|d\| = 1$, by Cauchy-Schwartz the solution is found as

$$d(f) = \frac{f^2}{\|f^2\|}.$$  

It remains to minimize $L(d(f), f) = n \|f^2\| / \sqrt{2}$ under the restrictions on $f^2$. Setting $g_j = f_j^2$, one has to minimize $\|g\|$ on the convex set (44). This is solved using Lagrange multipliers $\lambda, \mu$.  

To show that the solution $d_0$ fulfills the restriction $\sup_j d_j^2 \leq \delta/n\rho$, we note that

$$f_{0,j}^2 = \left( \lambda - \mu j^{2\beta} \right)_+ \leq \lambda \left( 1 - \mu \lambda^{-1} j^{2\beta} \right)_+ \leq \lambda; \quad (47)$$

24
below (cf. [54], Lemma 8) it is shown that \( \lambda \asymp n^{-1-1/(4\beta+1)} \) and \( n \| f_0^2 \| \asymp L_{n,0} \asymp 1 \). This implies

\[
\begin{align*}
n \rho d_{0,n,j}^2 &= n \rho \cdot O \left( n^2 \lambda^2 \right), \\
n \rho \lambda^2 &\asymp n \cdot n^{-4\beta/(4\beta+1)} \cdot n^{-2/(4\beta+1)} = n^{-1/(4\beta+1)};
\end{align*}
\]

(48)

thus for \( \delta = (\log n)^{-1} \) we have that \( d_0 \in \mathcal{D} \) for \( n \) large enough. ■

**Proof of Lemma 8.** The log-likelihood ratio is

\[
\log \frac{(n^{-1})^{n/2}}{(\sigma_n^2 + n^{-1})^{n/2}} \exp \left( -\frac{1}{2} \sum_{j=1}^{n} \left( \frac{Y_j^2}{\sigma_j^2 + n^{-1}} - \frac{Y_j^2}{n^{-1}} \right) \right)
\]

\[
= \frac{1}{2} \sum_{j=1}^{n} n \sigma_j^2 \left( \frac{\sigma_j^2}{n \sigma_j^2 + 1} \right) - \frac{n}{2} \sum_{j=1}^{n} \log (n \sigma_j^2 + 1).
\]

This shows (a) by setting \( d = \tilde{d}/\| \tilde{d} \| \) for \( \tilde{d}_j = \frac{n \sigma_j^2}{n \sigma_j^2 + 1} \). Now for \( \sigma_j^2 = f_{0j}^2 \), we have, as \( \lambda \asymp n^{-1-1/(4\beta+1)} \),

\[
n f_{0j}^2 = n \lambda \left( 1 - \lambda^{-1} \mu j^{2\beta} \right) \leq n \lambda \asymp n \cdot n^{-1-1/(4\beta+1)} = n^{-1/(4\beta+1)} = o(1),
\]

hence \( \tilde{d}_j \asymp n f_{0j}^2 \) uniformly over \( j = 1, \ldots, n \). This implies \( \| \tilde{d} \| \sim n \| f_0^2 \| \asymp n \) and

\[
d_j = \frac{\tilde{d}_j}{\| \tilde{d} \|} \asymp f_{0j}^2
\]

uniformly in \( j \leq n \). The proof of \( n \rho d_{0,n,j}^2 \leq \delta \) now exactly follows (17), (48). The convergence \( t \to z_\alpha \) now is a consequence of Lemma 11(a). ■

**Lemma 8** Suppose \( \rho = c \cdot n^{-4\beta/(4\beta+1)} \), \( c \) constant. Then the saddlepoint value \( L_0 \) of (20) fulfills

\[
L_0 = L(d_0, f_0^2) \sim \sqrt{A_0 \cdot M^{-1/(2\beta)} \cdot c^{2+1/(2\beta)} / 2}.
\]

**Proof.** The proof of Lemma 7 shows that \( L(d_0, f_0^2) \) is also the saddlepoint value under the weaker restrictions \( \| d \|^2 \leq 1 \), \( f \in \Sigma(\beta, M) \cap B_\rho \). Let us sketch a derivation of the asymptotics by a renormalization technique. Suppose that \( d_j = h_{j}^{1/2} \cdot d(j) \), \( j \leq n \) where \( h \) is a bandwidth parameter tending to 0, and the continuous function \( d : [0, \infty) \to [0, \infty) \) satisfies

\[
\int_0^\infty d^2(x) \, dx \leq 1.
\]

(49)

Consider another continuous function \( \sigma : [0, \infty) \to [0, \infty) \) satisfying

\[
\int_0^\infty x^{2\beta} \sigma^2(x) \, dx \leq 1 \quad \text{and} \quad \int_0^\infty \sigma^2(x) \, dx \geq 1
\]

(50)
and set $\sigma^2_j = Mh^{2\beta+1}\sigma^2(hj)$, $j \leq n$. Choose $h = (\rho/M)^{1/(2\beta)}$. The coefficient vector $d = (d_j)_{j=1}^n$ satisfies

$$\|d\|^2 = h \sum_{j=1}^n d(hj) \to \int_0^\infty d(x)dx \leq 1.$$ 

Identifying $f^2 \in \mathbb{R}_+^n$ with $(\sigma^2_j)_{j=1}^n$, the restriction $f \in \Sigma(\beta, M)$ is asymptotically satisfied since

$$\sum_{j=1}^\infty j^{2\beta} \sigma^2_j = Mh \sum_{j=1}^\infty (jh)^{2\beta} \sigma^2(jh) \to M \int_0^\infty x^{2\beta} \sigma^2(x)dx \leq M, \quad h \to 0.$$ 

The restriction $f \in B_\rho$ is also asymptotically satisfied since

$$\sum_{j=1}^\infty \sigma^2_j = Mh \sum_{j=1}^\infty \sigma^2(jh) = \rho h \sum_{j=1}^\infty \sigma^2(jh) \sim \rho \int_0^\infty \sigma^2(x)dx \geq \rho.$$ 

Therefore,

$$\frac{n}{\sqrt{2}} \sum_{j=1}^n d_j \sigma^2_j = \frac{n}{\sqrt{2}} Mh^{2\beta+1/2} h \sum_{j=1}^\infty d(hj) \sigma^2(jh) \sim \frac{e^{1+1/(4\beta)} M^{-1/(4\beta)}}{\sqrt{2}} \int_0^\infty d(x)\sigma^2(x)dx.$$ 

The saddle point problem (20) for each $n$ is thus asymptotically expressed in terms of a fixed continuous problem with constraints (49) and (50). There is unique positive solution $(\lambda^*, \mu^*)$ for the equations (cp. [16]),

$$\int_0^\infty x^{2\beta}(\lambda - \mu x^{2\beta})dx = 1,$$

$$\int_0^\infty (\lambda - \mu x^{2\beta})dx = 1.$$ 

Let $\|\cdot\|_2$ and $\langle \cdot, \cdot \rangle_2$ denote norm and scalar product in $L_2(\mathbb{R}_+)$. Then the saddle point $(d^*, \sigma^*)$ is given by

$$d^* = \frac{\sigma^*}{\|\sigma^*\|_2}, \quad \sigma^*(x) = (\lambda^* - \mu^* x^{2\beta})_+.$$ 

Then the value of the game is

$$\sup_{d \text{ in } (49)} \inf_{\sigma \text{ in } (60)} \langle d, \sigma^2 \rangle_2 = \inf_{\sigma \text{ in } (60)} \sup_{d \text{ in } (49)} \langle d, \sigma^2 \rangle_2 = \langle d^*, \sigma^* \rangle_2 = \|\sigma^*\|_2 = \sqrt{A_0(\beta)},$$

where the sup is taken for $d$ satisfying (49), the inf is taken for $\sigma$ satisfying (50), and $A_0(\beta)$ is Ermakov’s constant in (6). The continuous saddlepoint problem arises naturally in a continuous Gaussian white noise setting and a parameter space described by the continuous Fourier transformation, e.g. a Sobolev class of functions on the whole real line (cf. [16], [17]).
The above argument provides the guideline for a more rigorous proof, based on calculating the sharp asymptotics of \( \lambda \) and \( \mu \) directly from (45). The rough order of \( \lambda \) can be found as follows. By equating \( f_0^2 = \sigma_j^2 \), we find
\[
\left( \lambda - \mu j^{2\beta} \right)_+ = M h^{2\beta+1} \sigma^2(h_j),
\]
\[
= \lambda \left( 1 - \left( \frac{\mu}{\lambda} \right)^{1/2} j \right)^{2\beta}_+.
\]
we find \( \lambda \sim h^{2\beta+1} \), \( h \sim (\mu/\lambda)^{1/2\beta} \) and thus
\[
\lambda \sim h^{2\beta+1} \approx \rho^{(2\beta+1)/(2\beta)} \propto n^{-1/(4\beta+1)}.
\]

Remark 8 The paper of Ermakov [9], when calculating the asymptotics of \( \lambda, \mu \) in (45) and of \( A = 2L_0^2 \) (in a more general framework where \( \sum a_j f_j^2 \leq P_0, \sum b_j f_j^2 \geq \rho \)), contains an error for \( \lambda \). Here is the correction using the notations therein. Let \( a_j = L_j^{2\gamma} \), \( b_j = M_j^{2\nu} \), where \( \gamma > \nu \geq 0 \), \( L \) and \( M \) are positive constants, and set \( \epsilon = n^{-1/2} \). Then as \( \epsilon \to 0 \) we have that
\[
\lambda \sim \frac{(2\gamma + 2\nu + 1)}{2(\gamma - \nu)} \left( \frac{L}{P_0(4\gamma + 1)} \right)^{\frac{4\nu + 1}{2(\gamma - \nu)}} \left( \frac{1}{M} \right)^{\frac{4\nu + 1}{2(\gamma - \nu)}} \left[ \rho(4\nu + 1) \right]^{\frac{2(\nu + \nu) + 1}{2(\gamma - \nu)}},
\]
\[
\mu \sim \frac{(4\nu + 1)\rho \lambda}{P_0(4\gamma + 1)}, \quad A \sim \epsilon^{-4} \rho \lambda \frac{4\gamma - 4\nu}{4(\gamma + 1)}.
\]

References

[1] Belitser, E. and Levit, B., (1995). On minimax filtering on ellipsoids. Math. Meth. Statist. 4 259–273

[2] Butucea, C. (2007). Goodness-of-fit testing and quadratic functional estimation from indirect observations. Ann. Statist. 35 1907–1930.

[3] Butucea, C., Matias, C. and Pouet, C. (2009). Adaptive goodness-of-fit testing from indirect observations. Ann. Inst. Henri Poincaré Probab. Stat. 45 no. 2, 352–372.

[4] Chen, L. H. Y., Goldstein, L. and Shao, Q.-M. (2011). Normal Approximation by Stein’s method. Springer, Heidelberg, 2011.

[5] Dumbgen, L. and Spokoiny, V. G. (2001). Multiscale testing of qualitative hypotheses. Ann. Statist. 29, no. 1, 124–152.

[6] Efromovich, S. (1994). On adaptive estimation of nonlinear functionals. Statist. Probab. Lett. 19 , no. 1, 57–63.

[7] Efromovich, S. and Low, M. (1996). On optimal adaptive estimation of a quadratic functional. Ann. Statist. 24 , no. 3, 1106–1125.
[8] Efroimovich, S. and Pinsker, M.S. (1984). A learning algorithm for nonparametric filtering. *Automat. Remote Control* **11** 1434–1440.

[9] Ermakov, M. S. (1990). Minimax detection of a signal in Gaussian white noise. (Russian) *Teor. Veroyatnost. i Primenen.* **35** no. 4, 704–715; translation in *Theory Probab. Appl.* **35** no. 4, 667–679.

[10] Ermakov, M. S. (2003). Asymptotically efficient statistical inferences for moderate deviation probabilities. (Russian) *Teor. Veroyatnost. i Primenen.* **48** no. 4, 676–700; translation in *Theory Probab. Appl.* **48** no. 4, 622–641.

[11] Ermakov, M. S. (2008). Nonparametric hypothesis testing for small type I and type II error probabilities. (Russian) *Problemry Peredachi Informatsii* **44**, no. 2, 54–74; translation in *Probl. Inf. Transm.* **44**, no. 2, 119–137.

[12] Ermakov, M. S. (2011). Nonparametric signal detection with small type I and type II error probabilities. *Stat. Inference Stoch. Process.* **14** no. 1, 1–19.

[13] Fromont, M., and Laurent, B. (2006). Adaptive goodness-of-fit tests in a density model. *Ann. Statist.* **34** no. 2, 680–720.

[14] Fromont, M., Laurent, B., and Reynaud-Bouret, P. (2011). Adaptive tests of homogeneity for a Poisson process. *Ann. Inst. Henri Poincaré Probab. Stat.* **47**, no. 1, 176–213.

[15] Gayraud, G. and Pouet, C. (2005). Adaptive minimax testing in the discrete regression scheme. *Probab. Theory Related Fields* **133** no. 4, 531–558.

[16] Golubev, G. K. (1982). On minimax filtering of functions in $L_2$ (in Russian). *Problems Inform. Transmission* **18**, No. 4, 67–75.

[17] Golubev, G. K. (1987). Adaptive asymptotically minimax estimates of smooth signals. *Probl. Inform. Transmission* **23** 57–67.

[18] Golubev, G. K. (1990). Quasilinear estimates for signals in $L_2$. (Russian) *Problemy Peredachi Informatsii* **26**, no. 1, 19–24; translation in *Problems Inform. Transmission* **26** (1990), no. 1, 15–20.

[19] Golubev, G. K., Lepski, O. and Levit, B. (2001). On adaptive estimation for the sup-norm losses. *Math. Methods Statist.* **10**, no. 1, 23–37.

[20] Golubev, G. K. and Nussbaum, M. (1990). A risk bound in Sobolev class regression, *Ann. Statist.* **18** 758–778.

[21] Inglot, T. Kallenberg, W. C. M. and Ledwina, T. (1992). Strong moderate deviation theorems. *Ann. Probab.* **20** (1992), no. 2, 987–1003.

[22] Ingster, Yu. I. (1982). Minimax nonparametric detection of signals in white Gaussian noise. *Probl. Inform. Transmission* **18**, no. 2, p. 61.

[23] Ingster, Yu. I. and Suslina, I. A. (2003). *Nonparametric Goodness-of-fit Testing under Gaussian models. Lecture Notes in Statistics, 169.* Springer-Verlag, New York.
[24] Ingster, Yu. I and Suslina, I. A. (2005). Nonparametric hypothesis testing for small type I errors. II. Math. Methods Statist. 14 no. 1, 28–52.

[25] Ingster, Y. and Stepanova, N. (2011). Estimation and detection of functions from anisotropic Sobolev classes. Electron. J. Stat. 5 484–506.

[26] Klemelä, J. (2006). Sharp adaptive estimation of quadratic functionals. Probab. Theory Relat. Fields 134 539-64

[27] Korostelev, A. P. (1993). An asymptotically minimax regression estimator in the uniform norm up to a constant. (Russian) Teor. Veroyatnost. i Primenen. 38 , no. 4, 875–882; translation in Theory Probab. Appl. 38, no. 4, 737–743

[28] Lehmann, E. L. and Romano, J. P. (2005). Testing Statistical Hypotheses. Springer, NY.

[29] Lepski, O. V. and Tsybakov, A. B. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. Probab. Theory Related Fields 117 , no. 1, 17–48.

[30] Liese, F. and Miescke, K-J. (2008). Statistical Decision Theory. Estimation, Testing, and Selection. Springer Series in Statistics. Springer, New York.

[31] Nussbaum, M., (1999). Minimax risk: Pinsker bound. In: Encyclopedia of Statistical Sciences, Update Volume 3, 451-460 (S. Kotz, Ed.). Wiley, New York.

[32] Petrov, V. V. (1995). Limit Theorems of Probability Theory. Sequences of Independent Random Variables. Oxford University Press, New York.

[33] Pinsker, M. S. (1980). Optimal filtration of square-integrable signals in Gaussian noise. (Russian) Problems Inform. Transmission 16 , no. 2, 52–68

[34] Rohde, A. (2008). Adaptive goodness-of-fit tests based on signed ranks. Ann. Statist. 36 1346–1374.

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

[36] Tsybakov, A. B. (2009). Introduction to Nonparametric Estimation. Springer, NY

[37] van der Vaart, A. W. (1998). Asymptotic Statistics. Cambridge Univ. Press