Electronic Journal of Statistics
Vol. 11 (2017) 4515–4562
ISSN: 1935-7524
DOI: 10.1214/17-EJS1353

Sharp minimax adaptation over Sobolev ellipsoids in nonparametric testing

Pengsheng Ji

Department of Statistics
University of Georgia
Athens GA 30605
e-mail: psji@uga.edu

and

Michael Nussbaum*

Department of Mathematics
Cornell University
Ithaca NY 14853
e-mail: nussbaum@math.cornell.edu

Abstract: In the problem of testing for signal in Gaussian white noise, over a smoothness class with an L_2-ball removed, minimax rates of convergences (separation rates) are well known (Ingster [24]); they are expressed in the rate of the ball radius tending to zero along with noise intensity, such that a nontrivial asymptotic power is possible. It is also known that, if the smoothness class is a Sobolev type ellipsoid of degree β and size M, the optimal rate result can be sharpened towards a Pinsker type asymptotics for the critical radius (Ermakov [9]). The minimax optimal tests in that setting depend on β and M; but whereas in nonparametric estimation with squared L_2-loss, adaptive estimators attaining the Pinsker constant are known, the analogous problem in testing is open. First, for adaptation to M only, we establish that it is not possible at the critical separation rate, but is possible in the sense of the asymptotics of tail error probabilities at slightly slower rates. For full adaptation to (β, M), it is well known that a $\log\log n$-penalty over the separation rate is incurred. We extend a preliminary result of Ingster and Suslina [25] relating to fixed M and unknown β, and establish that sharp minimax adaptation to both parameters is possible. Thus a complete solution is obtained, in the basic L_2-case, to the problem of adaptive nonparametric testing at the level of asymptotic minimax constants.

MSC 2010 subject classifications: 62G10, 62G20.
Keywords and phrases: Minimax hypothesis testing, nonparametric signal detection, sharp asymptotic adaptivity, moderate deviation.

Received May 2016.

Contents

*Supported in part by NSF Grant DMS-1407600.

1. Introduction and main result

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

$$Y_j = f_j + n^{-1/2}\xi_j, \qquad j = 1, 2, ..., \tag{1.1}$$

with unknown, nonrandom signal $f = (f_j)_{j=1}^{\infty}$, and noise variables ξ_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

$$\Sigma(\beta, M) = \{f : \sum_{j=1}^{\infty} j^{2\beta} f_j^2 \le M\};$$

this might be called a Sobolev type ellipsoid with smoothness parameter β 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

$$B_\rho = \{f \in l_2 : \|f\|_2^2 \ge \rho\}.$$

Here $\rho^{1/2}$ is the radius of the open ball; for brevity we call ρ 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 ρ, consistent α-testing in that setting is possible. More precisely, there exist α-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 α-tests will have the trivial asymptotic (worst case) power α. According to a fundamental result of Ingster [24] there is a critical rate for ρ_n, the so-called *separation rate*

$$\rho_n \asymp n^{-4\beta/(4\beta+1)} \tag{1.2}$$

at which the transition in the power behaviour occurs. More precisely, consider a (possibly randomized) α-test ϕ in the model (1.1) for null hypothesis $H_0 : f = 0$, that is, a test fulfilling $E_{n,0}\phi \le \alpha$ where $E_{n,f}(\cdot)$ denotes expectation in the model (1.1). For given ϕ, we define the worst case type II error over the alternative $f \in \Sigma(\beta, M) \cap B_\rho$ as

$$\Psi(\phi, \rho, \beta, M) := \sup_{f \in \Sigma(\beta,M) \cap B_\rho} (1 - E_{n,f}\phi) . \tag{1.3}$$

The search for a best α-test in this sense leads to the minimax type II error

$$\pi_n(\alpha, \rho, \beta, M) := \inf_{\phi: E_{n,0}\phi \le \alpha} \Psi(\phi, \rho, \beta, M). \tag{1.4}$$

An α-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

$$0 < \liminf_n \pi_n(\alpha, \rho_n, \beta, M) \text{ and } \limsup_n \pi_n(\alpha, \rho_n, \beta, M) < 1 - \alpha.$$

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 β 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 (1.2), 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 the problem of nonparametric estimation of the signal f over $f \in \Sigma(\beta, M)$ with l_2-loss, where the risk asymptotics is given by the Pinsker constant, there is an array of results showing that adaptation is possible with neither a penalty in the rate nor in the constant, cf. Efromovich and Pinsker [7], Golubev [16], [17], 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 Φ be the distribution function of the standard normal, and for $\alpha \in (0, 1)$ let z_α be the upper α-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 $a_n \sim b_n$ if $a_n = b_n (1 + o(1))$. Furthermore, we write $a_n \asymp b_n$ if both $a_n = O(b_n)$ and $b_n = O(a_n)$ hold.

Proposition 1.1 (Ermakov [9]). *Suppose $\alpha \in (0,1)$ and that the radius ρ_n tends to zero at the separation rate, more precisely*

$$\rho_n \sim c\, n^{-4\beta/(4\beta+1)} \tag{1.5}$$

for some constant $c > 0$.

(i) For any sequence of tests ϕ_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

$$A(c,\beta,M) = A_0(\beta) M^{-1/2\beta} c^{2+1/2\beta} \tag{1.6}$$

and $A_0(\beta)$ is Ermakov's constant

$$A_0(\beta) = \frac{2(2\beta+1)}{(4\beta+1)^{1+1/2\beta}}. \tag{1.7}$$

(ii) For every $M > 0$ there exists a sequence of tests ϕ_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 β and M. Concerning adaptivity in both of these parameters, the following result is known.

Proposition 1.2 (Spokoiny [35]). *Let J be a subset of $(0,\infty) \times (0,\infty)$ such that there exist $M > 0$, $\beta_2 > \beta_1 > 0$ and*

$$[\beta_1, \beta_2] \times \{M\} \subseteq J.$$

(i) If $t_n \ll (\log\log n)^{1/2}$ and $\rho_n \sim c \cdot (t_n n^{-1})^{4\beta/(4\beta+1)}$ for some $c > 0$, then for any sequence of tests ϕ_n satisfying $E_{n,0}\phi_n \leq \alpha + o(1)$ we have

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

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

$$J = (1/2, \beta^*] \times [M_1, M_2].$$

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

$$\rho_n \sim c_1 \left((\log\log n)^{1/2} n^{-1}\right)^{4\beta/(4\beta+1)} \tag{1.8}$$

then

$$\sup_{(\beta,M)\in J} \Psi(\phi_n, \rho_n, \beta, M) = o(1). \tag{1.9}$$

Here the criterion to evaluate a test sequence does now include the worst case type II error over a whole range of β, M. Hence the critical radius rate (1.8) 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 (1.2); 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 ϕ_n of tests fulfilling (1.9) should be seen as *adaptive rate optimal*, comparable to rate optimal tests in the nonadaptive case (that is, tests fulfilling $\limsup_n \Psi(\phi_n, \rho_n, \beta, M) < 1-\alpha$ at ρ_n given by (1.2)). In Ingster and Suslina [25], chap. 7, the worst case adaptive error (1.9) is further analyzed, with a view to a sharp asymptotics; essentially a test is developed there which is sharp minimax adaptive over β for known M. We address this subject in our Sections 1.2 and 4 where the results of [25] are extended towards full minimax sharp adaptivity over β and M.

1.1. Adaptation over M only

Initially we now assume that β is fixed while we aim for adaptation over the ellipsoid size M. First, we present a negative result for adaptation at the classical separation rate (1.2).

Theorem 1.1. *Suppose $c > 0$, $0 < M_1 < M_2$ and (1.5). Then there is no test ϕ_n satisfying $E_{n,0}\phi_n \le \alpha + o(1)$ and both relations*

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

This result states that adaptation just over M is impossible at the separation rate.

We now modify the criterion, by enlarging the radius slightly and examining how the minimax error approaches zero. To be specific, we replace the constant c in (1.5) by a sequence c_n tending to infinity slowly. In that case the minimax type II error bound of Proposition 1.1, namely $\Phi(z_\alpha - \sqrt{A(c,\beta,M)/2})$ will tend to zero (since $A(c,\beta,M)$ as defined in (1.6) 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 1.2. *Assume $c_n \to \infty$ but $c_n = o(n^K)$ for every $K > 0$, and that $\rho_n = c_n\, n^{-4\beta/(4\beta+1)}$. For given $0 < M_1 < M_2$, there exists a test ϕ_n fulfilling*

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

and

$$\limsup_n \frac{1}{c_n^{2+1/2\beta}} \sup_{M_1 \le M \le M_2} M^{1/2\beta} \log \Psi(\phi_n, \rho_n, \beta, M) \le -\frac{A_0(\beta)}{4}.$$

To complement this result, a formal argument is needed that no α-test can be better in the sense of the log-asymptotics over radii ρ_n for the error of second kind. In Ermakov [12] the nonadaptive sharp asymptotics is studied in the above setting where type II error probability tends to zero.

Proposition 1.3. *Under the assumptions of Theorem 1.2, any test ϕ_n (possibly depending on M) satisfying $E_{n,0}\phi_n \leq \alpha + o(1)$ also fulfills*

$$\liminf_n \frac{M^{1/2\beta}}{c_n^{2+1/2\beta}} \log \Psi(\phi_n, \rho_n, \beta, M) \geq -\frac{A_0(\beta)}{4}. \tag{1.10}$$

This result is implied by Theorem 3 in [12], and hence the proof is omitted. In conjunction with Theorem 1.2, this proposition implies that if one switches to an error criterion expressed in the rate exponent of a slowly decaying error probability, then there is no penalty for adaptation. It is obvious from [12] that the bound (1.10) is nontrivial, in the sense that it specifies as optimal the quadratic tests using the optimal filtering weights found by Ermakov [9], and is not attained e.g. by tests with projection weights (i.e. weights from $\{0,1\}$).

It is of interest to consider a certain dual formulation of Theorems 1.1 and 1.2, where the radius ρ_n is allowed to depend on the ellipsoid parameters β and M, and a certain prescribed type II error level is to be attained, such as $\Phi(z_\alpha - d)$ for fixed $d > 0$. This formulation might be called the *variable radius approach.* A test is then optimal if it attains a given type II error level over the complement of sufficiently small balls. The "variable radius aproach" has been crucially used in [35] for expressing the rate penalty for adaptation (cf Proposition 1.2); we will also adopt it here for our sharp adaptation results. Note that in this setting, the sharp type II error asymptotics is encoded in the radius ρ_n.

Consider first the nonadaptive setting of Proposition 1.1. In this case, the connection between type II error level and optimal radius can easily be obtained by rescaling from Proposition 1.1: if for a $d > 0$, the constant c is determined by

$$c^{(4\beta+1)/4\beta} = A_1(\beta) M^{1/4\beta} d, \\ A_1(\beta) := (A_0(\beta)/2)^{-1/2}, \tag{1.11}$$

with $A_0(\beta)$ given by (1.7), or equivalently, the radius ρ_n is determined by

$$\rho_{n,M}^{(4\beta+1)/4\beta} = n^{-1} A_1(\beta) M^{1/4\beta} d, \tag{1.12}$$

one obtains that $A(c, \beta, M)$ given by (1.6) equals $2d^2$, and hence over a radius $\rho_{n,M}$ as in (1.12), the type II error level $\Phi(z_\alpha - d)$ is unimprovable and is attainable by a test depending on M.

For the adaptation problem in the variable radius setting, the two parts of the result below state the analogs of Theorems 1.1 and 1.2 respectively.

Theorem 1.3. *(i) Suppose $d > 0$, $0 < M_1 < M_2$, and also that $\rho_{n,M}$ is given by (1.12). Then there is no test ϕ_n satisfying $E_{n,0}\phi_n \leq \alpha + o(1)$ and both relations*

$$\Psi_n(\phi_n, \rho_{n,M_i}, \beta, M_i) \leq \Phi(z_\alpha - d) + o(1), \quad i = 1, 2.$$

(ii) Assume $d_n \to \infty$ but $d_n = o(n^K)$ for every $K > 0$, and that $\rho_{n,M}$ is given by

$$\rho_{n,M}^{(4\beta+1)/4\beta} = n^{-1} A_1(\beta) M^{1/4\beta} d_n. \tag{1.13}$$

For given $0 < M_1 < M_2$, there exists a test ϕ_n such that

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

and

$$\limsup_n \frac{1}{d_n^2} \log \sup_{M_1 \leq M \leq M_2} \Psi(\phi_n, \rho_{n,M}, \beta, M) \leq -\frac{1}{2}.$$

As with Theorem 1.2, to complement Theorem 1.3 a formal argument can be given that no α-test, possibly depending on M, can be better in the sense of the log-asymptotics over radii $\rho_{n,M}$ for the error of second kind. Such a result is analogous to Proposition 1.3 and is implicit in [12].

1.2. Adaptation over β and M

The adaptivity result of Spokoiny [35], discussed in Proposition 1.2, 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 1.1 or our Theorem 1.3. Some important results in this direction however are presented in section 7.1.3 of Ingster and Suslina [25]. Indeed in [25] the solution is presented for unknown $\beta \in [\beta_1, \beta_2]$ but fixed M. It should be noted that adaptation to β only, with M assumed known, does not have a practical interpretation in the context of smooth functions. We will address here the problem of a sharp risk bound for adaptation to the full parameter (β, M). For the analogous problem in the estimation case (regarding the Pinsker bound), solutions have been presented by Golubev [17] and Tsybakov [36], sec 3.7.

Our result can be summarized as follows: the lower asymptotic risk bound for known M, unknown $\beta \in [\beta_1, \beta_2]$ of [25] is achievable even for unknown M, by a refinement of the Bonferroni-type tests used to treat adaptation to β. Thus there is no further penalty for adaptation to M, in addition to the $\log\log n$-type penalty already incurred by adaptation to β.

We begin by stating the lower asymptotic risk bound for known M, unknown $\beta \in [\beta_1, \beta_2]$, a variation of Theorem 7.1 in [25]. Assume that $0 < \beta_1 < \beta_2$ are given as well as some $M > 0$. Let $D \in \mathbb{R}$ be arbitrary and define a radius sequence $\rho_{n,\beta,M}$ by

$$(\rho_{n,\beta,M})^{(4\beta+1)/4\beta} = n^{-1} A_1(\beta) M^{1/4\beta} \left((2\log\log n)^{1/2} + D \right). \tag{1.14}$$

Proposition 1.4. *Any sequence of tests ϕ_n satisfying $E_{n,0}\phi_n \leq \alpha + o(1)$ also fulfills*

$$\sup_{\beta \in [\beta_1, \beta_2]} \Psi(\phi_n, \rho_{n,\beta,M}, \beta, M) \geq (1-\alpha)\,\Phi(-D) + o(1). \tag{1.15}$$

In [25] the l_2-Sobolev ellipsoids represent a boundary case and are therefore not covered, but the above bound can be proved by very similar methods (cf. Section 4.2 below). Note that part (i) of Proposition 1.2 is implied by (1.15) by letting $D \to -\infty$.

As to the attainability of this bound, the test provided in section 7.3 of [25] depends on M. Indeed in [25] observations are assumed to be $X_j = v_j + \xi_j$, where ξ_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 [25], where $p = q = 2, s = \beta$; also r is ρ in [25]). This observation model is equivalent to ours upon setting $R^2 = nM$, $r^2 = n\rho$, and then $Y_j = n^{-1/2} X_j$, $f_j = n^{-1/2} v_j$. The reasoning provided in section 7.3.2 of [25] 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 (1.14) and β is from a certain grid of values in (β_1, β_2). Since in particular $R = n^{1/2} M^{1/2}$, it turns out that the test depends on M, though it has been made independent of $\beta \in (\beta_1, \beta_2)$. A version of such results for α_n-tests with $\alpha_n \to 0$ is given in [26].

The following theorem extends the result of [25] about attainability of the bound (1.15) for fixed M towards full adaptivity over (β, M).

Theorem 1.4. *Let $D \in \mathbb{R}$ be arbitrary and define a radius sequence $\rho_{n,\beta,M}$ by (1.14). Assume a nonempty interval $J = [\beta_1, \beta_2] \times [M_1, M_2] \subset \mathbb{R}_+^2$ is fixed. There exists a test ϕ_n such that*

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

and

$$\limsup_n \sup_{(\beta,M) \in J} \Psi(\phi_n, \rho_{n,\beta,M}, \beta, M) \leq (1 - \alpha) \, \Phi(-D) \, .$$

1.3. Further discussion

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

*Logarithmic vs. strong asymptotics***.** In [12] it is also shown that, for nonadaptive testing where $\rho_n = c_n \; n^{-4\beta/(4\beta+1)}$, $c_n \to \infty$, the lower bound (1.10) is attainable, so that the minimax type II error defined by (1.4) satisfies

$$\log \pi_n(\alpha, \rho_n, \beta, M) \sim -\frac{1}{4} A(c_n, \beta, M). \tag{1.16}$$

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 (1.16) can be strengthened to

$$\pi_n(\alpha, \rho_n, \beta, M) \sim \Phi(z_\alpha - \sqrt{A(c_n, \beta, M)/2}). \tag{1.17}$$

Results (1.16) and (1.17) have been obtained within a framework of efficient inference for moderate deviation probabilities, cf. Ermakov [11], [8]. Recall that in our setting $c_n = o(n^K)$ for every $K > 0$, so that the strong asymptotics (1.17) holds in the nonadaptive setting. It is an open question whether an adaptive analog of (1.17) holds.

For standardized sums S_n of independent random variables, if $\{S_n > x_n\}$ is a large or moderate deviation event, theorems on the relative error caused by replacing the exact distribution of S_n by its limiting distribution are sometimes called strong large or moderate deviation theorems to distinguish them from first order results on $\log P(S_n > x_n)$. For a background cf. [32], [22], [2], chap. 11.

*The detection problem***.** Instead of focussing on the worst case type II error $\Psi(\phi, \rho, \beta, M)$ (1.3) of α-tests ϕ, one may consider minimization of the sum of errors, that is of $E_{n,0}\phi + \Psi(\phi, \rho, \beta, M)$, over all tests ϕ. That has been called the detection problem in the literature; in [25] this problem is largely treated in parallel to the one for α-tests. There and in [23] one finds the analog of the nonadaptive sharp asymptotics of Proposition 1.1. It may be conjectured that analogs of our Theorems 1.1–1.4 concerning adaptivity hold there as well.

*The sup-norm problem***.** Lepski and Tsybakov [29] proved a sharp minimax result in testing when the alternative is a Hölder class (denoted $H(\beta, M)$, 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 (β, M) in the sup-norm case cf. [19]; for the testing case where β is given, Dümbgen and Spokoiny [6] established a sharp adaptivity result with respect to the size parameter M only. The result in Theorem 2.2. of [6] can be seen as a analog of our Theorem 1.3, 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 (β, M) seems to be an open problem in the sup-norm testing case, with regard to sharp minimaxity, although in [6] 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 [6] with ideas related to rank tests.

*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 in [35], has also been established in a discrete regression model [15], and in density models with direct and indirect observations [13], [1]. Testing in a white noise model with composite hypotheses derived from a shifted curve model has been treated in [3]. In a regression context, composite null hypotheses given by a parametric family have been considered in [21]. For a review of adaptive separation rates and further results in a Poisson process model cf. [14]. For sharp minimax testing in nongaussian models (the nonadaptive theory) cf. Ermakov [10] and references therein; for the analogous topic in estimation cf. [18], [31]. An interesting connection to random matrix theory has recently been made in [5] by establishing a Pinsker type constant for estimation in high dimensional regression models.

The structure of the paper is as follows. In Section 2 we prove the negative result of Theorems 1.1 and 1.3 (i) that adaptation over M fails at the separation radius (at rate $n^{-4\beta/(4\beta+1)}$). Section 3 presents the proofs that adaptation over M is possible if the radius is slightly enlarged, i.e. the proofs of Theorem 1.2 and its dual version Theorem 1.3 (ii). The proof of Theorem 1.4 about existence of adaptive tests in the two parameter framework (β, M) is presented in Section 4; for completeness a proof of the lower bound of Proposition 1.4 is also included. In an Appendix section some technical auxiliary results are collected.

2. Proof of the negative result at separation rate

The following lemma will serve to prove the result of Theorem 1.1 and its version in the variable radius setting (Theorem 1.3 (i)) in a unified way.

Lemma 2.1. *Let $c_i > 0$, $M_i > 0$, $i = 1, 2$ be constants such that $0 < M_1/c_1 \leq M_2/c_2$, and define sequences*

$$\rho_{n,i} = c_i n^{-4\beta/(4\beta+1)},\ i = 1, 2.$$

Assume there exists a test sequence ϕ_n satisfying, for some $\alpha > 0$

$$E_{n,0}\phi_n \leq \alpha + o(1) \tag{2.1}$$

and both relations

$$\Psi(\phi_n, \rho_{n,i}, \beta, M_i) \leq \Phi\left(z_\alpha - \sqrt{A(c_i, \beta, M_i)/2}\right) + o(1),\ i = 1, 2. \tag{2.2}$$

Then

$$\frac{M_1}{c_1} = \frac{M_2}{c_2}. \tag{2.3}$$

The proof will be carried out in several steps. For brevity we write $A_i = A(c_i, \beta, M_i), i = 1, 2$ in this section (cp. (1.6)). Let $\lambda(M, \rho)$, $\mu(M, \rho)$ be the solutions of (A.1) provided by Lemma A.1, and for some $\varepsilon \in (0, 1)$ set $\lambda_i = \lambda(\rho_{n,i}(1+\varepsilon), M_i(1-\varepsilon))$, $\mu_i = \mu(\rho_{n,i}(1+\varepsilon), M_i(1-\varepsilon))$, $i = 1, 2$. Define

$$f_{0,j,i}^2 = (\lambda_i - \mu_i j^{2\beta})_+, \quad j = 1, 2, \ldots,\ i = 1, 2.$$

Then according to (A.1)

$$\sum_{j=1}^{\infty} j^{2\beta} f_{0,j,i}^2 = M_i(1-\varepsilon),\ \sum_{j=1}^{\infty} f_{0,j,i}^2 = \rho_{n,i}(1+\varepsilon),\ i = 1, 2. \tag{2.4}$$

Define $N_i = (\mu_i/\lambda_i)^{1/2\beta}$ in agreement with (A.2) and note that $f_{0,j,i}^2 = 0$ if and only if $j \geq N_i$. Let $Q_{n,i}$ be the Gaussian prior for f where $f_j \sim N(0, f_{0,j,i}^2)$ independently if $f_{0,j,i}^2 > 0$, and $f_j = 0$ otherwise.

Lemma 2.2. *For all $\varepsilon \in (0, 1)$, the prior measures $Q_{n,i}$ satisfy*

$$Q_{n,i}(\Sigma(\beta, M_i) \cap B(\rho_{n,i})) = 1 + o(1),\ i = 1, 2.$$

Proof. It suffices to show that

$$Q_{n,i}\left(\Sigma(\beta, M_i)^c\right) = o\left(1\right), \tag{2.5}$$

$$Q_{n,i}\left(B\left(\rho_{n,i}\right)^c\right) = o\left(1\right), \; i = 1, 2. \tag{2.6}$$

We have, by the first relation of (2.4)

$$Q_{n,i}\left(\Sigma(\beta, M_i)^c\right) = P\left(\sum_{j\geq 1} j^{2\beta} f_j^2 > M_i\right)$$

$$= P\left(\sum_{j\geq 1} j^{2\beta}\left(f_j^2 - f_{0,j,i}^2\right) > \varepsilon M_i\right) \leq \left(\varepsilon M_i\right)^{-2} 2 \sum_{j\geq 1} j^{4\beta} f_{0,j,i}^4.$$

According to Lemma A.1, relation (A.5), the latter quantity is $O\left(\rho_{n,i}^{2+3/2\beta}\right) = o\left(1\right)$, which establishes (2.5). Furthermore, by the second relation of (2.4)

$$Q_{n,i}\left(B\left(\rho_{n,i}\right)^c\right) = P\left(\sum_{j\geq 1} f_j^2 < \rho_{n,i}\right)$$

$$= P\left(\sum_{j\geq 1}\left(f_j^2 - f_{0,j,i}^2\right) < -\rho_{n,i}\varepsilon\right) \leq 2\varepsilon^{-2}\rho_{n,i}^{-2}\sum_{j\geq 1} f_{0,j,i}^4 = 2\varepsilon^{-2} O\left(\rho_{n,i}^{2+3/2\beta}\right)$$

according to Lemma A.1, relation (A.6), which establishes (2.6). □

As a consequence, for $B_i = B_{\rho_{n,i}}$

$$\sup_{f\in\Sigma(\beta,M_i)\cap B_i} E_{f,n}(1-\phi_n) \geq \int E_{n,f}(1-\phi_n)\, Q_{n,i}(df) + o\left(1\right). \tag{2.7}$$

Recall $Y_j = f_j + n^{-1/2}\xi_j$. Let $Q_{n,0}$ be the prior distribution where $f = 0$ a.s. and consider the resulting joint distribution of $(Y_j)_{j=1}^{\infty}$ under the priors $Q_{n,i}$ $i = 0, 1, 2$, that is

$$\pi_{n,0} : Y_j \sim N(0, n^{-1}), \; j = 1, 2, \ldots$$
$$\pi_{n,i} : Y_j \sim N(0, n^{-1} + f_{0,j,i}^2), \; j = 1, 2, \ldots, i = 1, 2$$

with corresponding expectations $E_{n,i}^{\pi}$. Then we obtain

$$E_{n,0}^{\pi}\phi_n = E_{n,0}\phi_n,$$
$$E_{n,i}^{\pi}(1-\phi_n) = \int E_{n,f}(1-\phi_n)\, Q_{n,i}(df), \quad i = 1, 2.$$

Combining this with (2.2), (2.1) and (2.7) gives

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

$$E_{n,i}^{\pi}(1-\phi_n) \leq \Phi\left(z_\alpha - \sqrt{A_i/2}\right) + o(1), \; i = 1, 2.$$

The likelihood ratio of $\pi_{n,i}$ against $\pi_{n,0}$ is, using notation $\gamma_{j,i}^2 := nf_{0,j,i}^2$,

$$\begin{aligned}\frac{d\pi_{n,i}}{d\pi_{n,0}} &= \prod_j \left(\frac{n^{-1}}{n^{-1}+f_{0,j,i}^2}\right)^{1/2} \exp\left(-\frac{1}{2}\sum_j \left(\frac{Y_j^2}{n^{-1}+f_{0,j,i}^2} - \frac{Y_j^2}{n^{-1}}\right)\right) \\ &= \prod_j \left(\frac{1}{1+\gamma_{j,i}^2}\right)^{1/2} \exp\left(\frac{1}{2}\sum_j \frac{\gamma_{j,i}^2}{1+\gamma_{j,i}^2} nY_j^2\right).\end{aligned}$$

Set $\tilde{g}_{j,i} := \gamma_{j,i}^2 / \left(1+\gamma_{j,i}^2\right)$, then by the factorization theorem, it is seen that the bivariate statistic

$$\left(\sum_j \tilde{g}_{j,1}\left(nY_j^2 - 1\right), \sum_j \tilde{g}_{j,2}\left(nY_j^2-1\right)\right) \tag{2.8}$$

is sufficient for the family of distributions $\{\pi_{n,i}, i=0,1,2\}$. To simplify notation, set $z_j = 2^{-1/2}\left(nY_j^2-1\right)$ and $\tilde{g}_i := (\tilde{g}_{j,i})_{j=1}^{\infty}$. Since only finitely many $\tilde{g}_{j,i}$ are nonzero, the scalar product $\langle \tilde{g}_i, z\rangle$ and the euclidean norm $\|\tilde{g}_i\|$ are well defined. Set $g = \tilde{g}_i / \|\tilde{g}_i\|$ and define the bivariate statistic

$$T_n = (\langle g_1, z\rangle, \langle g_2, z\rangle)'$$

which is equivalent to (2.8) and thus sufficient in $\{\pi_{n,i}, i=0,1,2\}$. Write the induced family for T_n as $\{\pi_{n,i}^T, i=0,1,2\}$ with corresponding expectations $E_{n,i}^{\pi,T}$ and take the conditional expectation $\phi_n^*(T_n) = E_{n,\cdot}^{\pi}(\phi_n|T_n)$. By sufficiency the (possibly randomized) test $\phi_n^*(T_n)$ for null hypothesis $\left\{\pi_{n,0}^T\right\}$ against alternative $\left\{\pi_{n,i}^T, i=1,2\right\}$ is as good as ϕ_n (cf. e.g. Theorem 4.66 in [30]), that is

$$\begin{aligned}&E_{n,0}^{\pi,T}\phi_n^* \le E_{n,0}\phi_n \le \alpha + o(1),\\ &E_{n,i}^{\pi,T}(1-\phi_n^*) \le E_{n,i}^{\pi}\phi_n \le \Phi(z_\alpha - \sqrt{A_i/2}) + o(1), \quad i=1,2.\end{aligned}$$

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

Lemma 2.3. *As $n\to\infty$, for fixed $\varepsilon>0$, each distribution $\pi_{n,i}^T, i=0,1,2$ converges in total variation to $\pi_{0,i}^T := N(\mu_i,\Sigma)$, $i=0,1,2$ respectively, where $\mu_0 = (0,0)'$ and for $\delta_\varepsilon := (1+\varepsilon)\left[(1+\varepsilon)/(1-\varepsilon)\right]^{1/4\beta}$*

$$\mu_1 = (A_1/2)^{1/2}\,\delta_\varepsilon(1,r)', \tag{2.9}$$

$$\mu_2 = (A_2/2)^{1/2}\,\delta_\varepsilon(r,1)', \tag{2.10}$$

$$\Sigma = \begin{pmatrix} 1 & r \\ r & 1\end{pmatrix} \text{ where} \tag{2.11}$$

$$r = \left(\frac{M_1c_2}{M_2c_1}\right)^{1/(4\beta)} \cdot \frac{4\beta+1-M_1c_2/M_2c_1}{4\beta}. \tag{2.12}$$

The proof follows below. Here both the families $\left\{\pi_{n,i}^{T}, i=0,1,2\right\}$ and their limit $\left\{\pi_{0,i}^{T}, i=0,1,2\right\}$ depend on ε. It is then obvious that there exists a sequence $\varepsilon_n \to 0$ such that $\left\{\pi_{n,i}^{T}, i=0,1,2\right\}$ converges in total variation to a limit family defined as in the Lemma above, with δ_ε replaced by 1, still denoted $\left\{\pi_{0,i}^{T}, i=0,1,2\right\}$. Then by the weak compactness theorem (cf. [28], A.5.1), there exist a test ϕ^* and a subsequence $\phi_{n_k}^*$ such that $\phi_{n_k}^*$ converges weakly to ϕ^*. Thus

$$E_{0,0}^{\pi,T}\phi^* \le \alpha, \tag{2.13}$$

$$E_{0,i}^{\pi,T}(1-\phi^*) \le \Phi(z_\alpha - \sqrt{A_i/2}), \quad i=1,2. \tag{2.14}$$

Consider now the Neyman-Pearson test for $N(0,\Sigma)$ against a simple hypothesis $N(\mu_i,\Sigma)$. Here the type II error is $\Phi(z_\alpha - \left\|\Sigma^{-1/2}\mu_i\right\|)$, and we find (from (2.9)–(2.11) for $\varepsilon = 0$)

$$\left\|\Sigma^{-1/2}\mu_1\right\|^2 = \mu_1'\Sigma^{-1}\mu_1 = \frac{A_1}{2\left(1-r^2\right)}(1,r)\begin{pmatrix} 1 & -r \\ -r & 1 \end{pmatrix}(1,r)' = \frac{A_1}{2},$$

$$\left\|\Sigma^{-1/2}\mu_2\right\|^2 = \mu_2'\Sigma^{-1}\mu_2 = \frac{A_2}{2\left(1-r^2\right)}(r,1)\begin{pmatrix} 1 & -r \\ -r & 1 \end{pmatrix}(r,1)' = \frac{A_2}{2}.$$

Therefore, by (2.13), (2.14) the α-test ϕ^* is uniformly most powerful (UMP) for $N(0,\Sigma)$ against a composite alternative $\{N(\mu_1,\Sigma), N(\mu_2,\Sigma)\}$. It can be checked that r in (2.12) is monotone increasing in M_1c_2/M_2c_1: indeed for $g\left(t\right) = t^{1/4\beta}\left(4\beta+1-t\right)$ we have

$$g'\left(t\right) = \frac{\left(4\beta+1\right)}{4\beta}t^{1/4\beta}\left(t^{-1}-1\right) > 0 \text{ for } 0 < t < 1.$$

Assume now that $M_1/c_1 < M_2/c_2$; then it follows that $0 < r < 1$. Since μ_1 is a multiple of the vector $(1,r)'$ and μ_2 is a multiple of the vector $(r,1)'$, it follows that μ_1, μ_2 and the origin are not on the same line. We shall show that in that situation, a UMP test for alternative $\{N(\mu_1,\Sigma), N(\mu_2,\Sigma)\}$ does not exist. Indeed, the log-likelihood ratio of $N(\mu_i,\Sigma)$ against $N(0,\Sigma)$ has the form $\mu_i'\Sigma^{-1}T - \mu_i'\Sigma^{-1}\mu_i/2$ (T representing observations), thus by the necessity part of the Neyman-Pearson lemma ([28], Theorem 3.2.1), the most powerful test for $N(0,\Sigma)$ against $N(\mu_i,\Sigma)$ has the form of $\mathbf{1}\{\mu_i'\Sigma^{-1}T > k_i\}$. But since these two types of tests can never coincide, for any choice of thresholds k_i, there is no UMP test for $N(0,\Sigma)$ against $\{N(\mu_1,\Sigma), N(\mu_2,\Sigma)\}$. By this contradiction, the claim $M_1/c_1 = M_2/c_2$ in Lemma 2.1 is proved.

*Proof of Lemma **2.3**.* Note that under $\pi_{n,0}$, the r.v.'s $z_j = 2^{-1/2}\left(nY_j^2 - 1\right)$ are i.i.d. with $Ez_j = 0$ and $\mathrm{Var}\left(z_j\right) = 1$, hence $E\left\langle g_i, z\right\rangle = 0$, $\mathrm{Var}\left(\left\langle g_i, z\right\rangle\right) = 1$, and from Lemma A.2, Appendix we find

$$\mathrm{Cov}\left(\left\langle g_1, z\right\rangle, \left\langle g_2, z\right\rangle\right) = \left\langle g_1, g_2\right\rangle \underset{n\to\infty}{\longrightarrow} r.$$

To check the CLT in distribution for T_n under $\pi_{n,0}$, it suffices to show that

$$\sup_j \tilde{g}_{j,i}/\left\|\tilde{g}_i\right\| = o\left(1\right) \text{ for } i=1,2.$$

This follows from relation (A.13) in Lemma A.2 in conjunction with $N_1, N_2 \to \infty$, cf. (A.2). It remains to check the CLT in total variation. Consider the first component of T_n, that is

$$\langle g_1, z \rangle = \|\tilde{g}_1\|^{-1} \sum_j \tilde{g}_{j,1} z_j.$$

Here at most $[N_1]$ of the $\tilde{g}_{j,1}$ are nonzero, so one may apply Lemma A.4, identifying the sample size m there with $[N_1]$. Then the condition

$$\sup_{1 \le j \le N_1} \tilde{g}_{j,1} / \|\tilde{g}_1\| = O\left(N^{-1/2}\right)$$

follows again from (A.13). To check the condition on the characteristic function of z_1, note that

$$\phi(t) = E \exp(\boldsymbol{i} t z_1) = \left(1 - 2^{1/2} \boldsymbol{i} t\right)^{-1/2} \exp\left(\boldsymbol{i} t 2^{-1/2}\right),$$
$$|\phi(t)|^2 = \left(1 + 2t^2\right)^{-1},$$

such that $|\phi|^2$ is integrable. Hence the two marginal distributions of T_n satisfy the CLT in total variation. A straighforward extension of Lemma A.4 to bivariate coefficients $c_{jn} = (c_{jn,1}, c_{jn,2})$ for which the limit of $\sum_j c_{jn,1} c_{jn,2}$ exists, establishes the result for the law of the vector T_n under $\pi_{n,0}$.

Under $\pi_{n,1}$, we have $E 2^{1/2} z_j = n f_{0,j,1}^2 = \tilde{g}_{j,1}(1 + o(1))$, hence $E 2^{1/2} \langle g_1, z \rangle \sim \langle g_1, \tilde{g}_1 \rangle = \|\tilde{g}_1\|$ and $E 2^{1/2} \langle g_2, z \rangle \sim \langle g_2, \tilde{g}_1 \rangle = \|\tilde{g}_1\| \langle g_2, g_1 \rangle$. From (A.12) in Lemma A.2 we find

$$\|\tilde{g}_1\|^2 \sim c_1^{2+1/2\beta} M_1^{-1/2\beta} A_0(\beta)\, \delta_\varepsilon^2 = A_1 \delta_\varepsilon^2$$

and since $\langle g_2, g_1 \rangle \sim r$, we obtain that under $\pi_{n,1}$

$$E T_n \sim 2^{-1/2} \|\tilde{g}_1\| (1, r)' \sim (A_1/2)^{1/2} \delta_\varepsilon (1, r)'.$$

The result $E T_n \sim (A_2/2)^{1/2} \delta_\varepsilon (r, 1)'$ under $\pi_{n,2}$ is shown analogously. □

Proof of Theorem 1.1. In Lemma 2.1, set $c_1 = c_2 = c$. If there exists a test as in the theorem then according to (2.3) it follows that $M_1 = M_2$, which contradicts the assumption $M_1 < M_2$ in the theorem. □

Proof of Theorem 1.3 (i). In accordance with (1.12) set $c_i = M_i^{1/(1+4\beta)} \times \left(n^{-1} A_1(\beta)\, d\right)^{4\beta/(4\beta+1)}$, $i = 1, 2$ in Lemma 2.1. If there exists a test as in the theorem then according to (2.3) it follows that $1 = M_1 c_2 / M_2 c_1 = (M_1/M_2)^{4\beta/(4\beta+1)}$, which contradicts the assumption $M_1 < M_2$. □

3. Proofs for adaptation over M

Assume $f \in \Sigma(\beta, M)$, M fixed and set for a fixed $\gamma \in \left[0, (4\beta+1)^{-1}\right]$

$$\rho_{n,M} = c_n M^\gamma n^{-4\beta/(4\beta+1)} \tag{3.1}$$

where c_n is a sequence fulfilling

$$c_n \to \infty,\ c_n = o\left(n^K\right) \text{ for all } K > 0, \tag{3.2}$$

such that $\rho_n = o(1)$. Define the test statistic

$$T_n = 2^{-1/2} \sum_{j=1}^{\infty} d_j \left(nY_j^2 - 1\right). \tag{3.3}$$

where the coefficients d_j, (only finitely many are nonzero) are determined as follows. Let $\lambda = \lambda(M, \rho_{n,M})$, $\mu = \mu(M, \rho_{n,M})$ be the solutions of (A.1) provided by Lemma A.1, set $N = (\lambda/\mu)^{1/2\beta}$ and

$$\tilde{d}_j = \left(1 - (j/N)^{2\beta}\right)_+,\ \ j = 1, 2, \ldots$$

Observe that $\tilde{d}_j = 0$ for $j \geq N$, so the vector $\tilde{d} = \left(\tilde{d}_j\right)_{j=1}^{\infty}$ has finite Euclidean norm $\left\|\tilde{d}\right\|$ and we set

$$d_j = \tilde{d}_j / \left\|\tilde{d}\right\|,\ \ j = 1, 2 \ldots. \tag{3.4}$$

Observe that the that test statistic T_n in (3.3) now depends on M and the choice of $\rho_{n,M}$; we write $T_n(M, \rho_{n,M})$ to indicate that dependence. Assume also that $\alpha_n \to 0$ is a sequence fulfilling

$$\left|\log \alpha_n\right| = O\left(c_n^2\right) \tag{3.5}$$

and let $\tilde{z}(\alpha_n) = \left(2 \log \alpha_n^{-1}\right)^{1/2}$. Define a test by

$$\phi_n(M, \rho_{n,M}, \alpha_n) = \mathbf{1}\{T_n(M, \rho_{n,M}) > \tilde{z}(\alpha_n)\}. \tag{3.6}$$

The following lemma concerning the nonadaptive test (3.6) is similar to Theorem 4 in [12]; for clarity of exposition we present a proof here.

Lemma 3.1. *Under assumptions (3.1), (3.2) and (3.5), the test $\phi_n = \phi_n(M, \rho_{n,M}, \alpha_n)$ fulfills*

$$E_{n,0}\phi_n \leq \alpha_n (1 + o(1)), \tag{3.7}$$

$$\frac{M^{1/2\beta}}{(M^\gamma c_n)^{2+1/2\beta}} \log \Psi(\phi_n, \rho_{n,M}, \beta, M) \leq -\frac{A_0(\beta)}{4} + o(1) \tag{3.8}$$

uniformly over any interval $M \in [M_1, M_2]$ where $M_1 > 0$.

Proof. Under H_0, we have $f = 0$. In this case $T_n = U_n$ where

$$U_n = \sum_{1 \leq j \leq N} d_j z_j,\ z_j = 2^{-1/2}\left(\xi_j^2 - 1\right). \tag{3.9}$$

In view of $\sum_{1\le j\le N} d_j^2 = 1$, $Ez_j = 0$ and $\mathrm{Var}\,(z_j) = 1$, for the convergence in law of U_n to $N(0,1)$ it suffices to show that $\max_{1\le j\le N} |d_j| = o\,(1)$. Recall that $d_j = \tilde{d}_j / \left\| \tilde{d} \right\|$, where $\left| \tilde{d}_j \right| \le 1$, and observe that according to Lemma A.1

$$\left\| \tilde{d} \right\|^2 = \sum_{1\le j\le N} \left(1 - (j/N)^{2\beta}\right)^2 = \lambda^{-2} T_{(1)} \asymp \rho^{-1/2\beta} \asymp N,$$

so that

$$\max_{1\le j\le N} |d_j| = O\left(N^{-1/2}\right) \tag{3.10}$$

and $U_n \rightsquigarrow N(0,1)$.

Consider now

$$E_{n,0}\phi_n = P_{n,0}\left(U_n \ge \tilde{z}\,(\alpha_n)\right).$$

for $\tilde{z}\,(\alpha_n) = \left(\log \alpha_n^{-2}\right)^{1/2}$. We will apply Lemma A.5, Appendix to estimate that probability. Setting $m = [N]$, $Y_j = z_j$ and $c_{jm} = d_j$ for $d_j > 0$ there, we have seen above that this set of coefficients $\{c_{jm}\}_{j=1}^m$ fulfills the conditions of Lemma A.5. Furthermore, the z_j are standardized χ_1^2-variables where Cramer's condition is fulfilled for any $H > 0$. Setting $a_m = \tilde{z}\,(\alpha_n)$, we find from (3.5)

$$\tilde{z}\,(\alpha_n) = \left(\log \alpha_n^{-2}\right)^{1/2} = O\,(c_n)\,. \tag{3.11}$$

Note that if (3.1) holds then according to (A.2) in Lemma A.1 (Appendix), we have

$$N \asymp \rho_{n,M}^{-1/2\beta} \asymp c_n^{-1/2\beta} n^{2/(4\beta+1)}. \tag{3.12}$$

The basic growth condition $c_n = o\left(n^K\right)$ for every $K > 0$ and (3.12) imply that $N \gg n^\eta$ for some $\eta > 0$, so that $a_m = o\left(N^{1/6}\right)$. Lemma A.5 now implies

$$P_{n,0}\left(U_n \ge \tilde{z}\,(\alpha_n)\right) \le \exp\left(-\left(\tilde{z}\,(\alpha_n)\right)^2/2\right)(1 + o\,(1)) = \alpha_n\,(1 + o\,(1))$$

which proves (3.7).

Consider now, for $\rho = \rho_{n,M}$,

$$\begin{aligned}\Psi(\phi_n, \rho, \beta, M) &= \sup_{f\in\Sigma(\beta,M)\cap B_\rho} \left(1 - E_{n,f}\phi_n\right) \\ &= \sup_{f\in\Sigma(\beta,M)\cap B_\rho} P_{n,f}\left(T_n \le \tilde{z}\,(\alpha_n)\right).\end{aligned}$$

We now make use of some further results collected in the Appendix. By Lemma A.6, the set B_ρ in the supremum may be replaced by B_ρ^0 defined in (A.16). Set

$$T_n^0(f) = 2^{1/2} \sum_{j=1}^\infty d_j n^{1/2} f_j \xi_j, \tag{3.13}$$

then

$$\begin{aligned}T_n &= 2^{-1/2} \sum_{j=1}^\infty d_j \left(n f_j^2 + 2n^{1/2} f_j \xi_j + \xi_j^2 - 1\right) \\ &= L_n\,(d, f) + T_n^0(f) + U_n\end{aligned}$$

where $L_n(d, f)$ denotes the functional (A.17) and U_n is defined by (3.9). Note that

$$
\begin{aligned}
&ET_n^0(f) = 0, \\
&\tau_n(f) := \mathrm{Var}\left(T_n^0(f)\right) = 2\sum_{j=1}^{\infty} d_j^2 n f_j^2 \le 2\left(\max_j d_j^2\right)\sum_{j=1}^{N} n f_j^2 \\
&\qquad \le n\rho O\left(N^{-1}\right) = n\rho O\left(\rho^{1/2\beta}\right)
\end{aligned}
$$

by using (3.10) and $f \in B_\rho^0$, and then (A.2). The current choice $\rho = \rho_{n,M} \asymp n^{-4\beta/(4\beta+1)} c_n$ implies

$$
\tau_n := \sup_{f \in B_\rho^0} \tau_n(f) = O\left(n^{-1/(4\beta+1)} c_n^{1+1/2\beta}\right). \tag{3.14}
$$

By splitting into events $\left\{-T_n^0(f) \le 1\right\}$ and its complement, we find

$$
\begin{aligned}
&P_{n,f}\left(L_n(d, f) + T_n^0(f) + U_n \le \tilde{z}(\alpha_n)\right) \\
&\le P_{n,f}\left(L_n(d, f) + U_n \le \tilde{z}(\alpha_n) + 1\right) + P_{n,f}\left(T_n^0(f) \le -1\right).
\end{aligned}
$$

Here the second term is

$$
P_{n,f}\left(T_n^0(f) \le -1\right) = \Phi\left(-\tau_n^{-1/2}(f)\right) \le \Phi\left(-\tau_n^{-1/2}\right).
$$

Defining $\check{z}_n = \tilde{z}(\alpha_n) + 1$, we now have

$$
\Psi(\phi_n, \rho, \beta, M) \le \sup_{f \in \Sigma(\beta, M) \cap B_\rho^0} P_n\left(L_n(d, f) + U_n \le \check{z}_n\right) + \Phi\left(-\tau_n^{-1/2}\right). \tag{3.15}
$$

Note that (3.12) implies $N \ll n^2$. Hence if we take $p = n^2$ in Lemma A.3 and $\rho = \rho_{n,M}$, then the first p coefficients of the test statistic T_n given by d_j in (3.4) are given by the saddlepoint solution d_0 of (A.18), and all other d_j are zero. Then relation (A.4) can be used to find the asymptotics of

$$
\inf_{f \in \Sigma(\beta, M) \cap B_\rho^0} L_n(d, f) = S_n = n\left(T_{(1)}/2\right)^{1/2} \tag{3.16}
$$

where S_n is given by (A.20) and $T_{(1)}$ by (A.4). This implies

$$
\Psi(\phi_n, \rho_{n,M}, \beta, M) \le P_n\left(S_n + U_n \le \check{z}_n\right) + \Phi\left(-\tau_n^{-1/2}\right). \tag{3.17}
$$

Combining relation (A.4) with the current choice $\rho_{n,M} = c_n M^\gamma n^{-4\beta/(4\beta+1)}$ one obtains

$$
S_n \sim \left(M^\gamma c_n\right)^{1+1/4\beta} M^{-1/4\beta}\left(\frac{A_0(\beta)}{2}\right)^{1/2} \tag{3.18}
$$

and thus by (3.11)

$$h_n := S_n - \check{z}_n \sim S_n. \tag{3.19}$$

Again invoking Lemma A.5, Appendix, with c_{jm} and m as above but now setting $Y_j = -z_j$ and $a_m = h_n$, we find

$$P_n\left(S_n + U_n \leq \check{z}_n\right) = P_n\left(-U_n \geq h_n\right) \leq \exp\left(h_n^2/2\right)\left(1 + o\left(1\right)\right).$$

From (3.17) we now have

$$\Psi(\phi_n, \rho_{n,M}, \beta, M) \leq \exp\left(-h_n^2/2\right)\left(1 + o\left(1\right)\right) + \Phi\left(-\tau_n^{-1/2}\right).$$

Note that by (3.14), (3.18) and (3.19) we have $\tau_n^{-1/2} \gg h_n$ so that $\Phi\left(-\tau_n^{-1/2}\right) \leq \Phi\left(-h_n\right)$, and the well known relation $\Phi\left(-h_n\right) \sim (2\pi)^{-1/2} h_n^{-2} \exp\left(-h_n^2/2\right)$ now implies

$$\Psi(\phi_n, \rho_{n,M}, \beta, M) \leq \exp\left(-h_n^2/2\right)\left(1 + o\left(1\right)\right).$$

With (3.18) and (3.19) this yields

$$\frac{M^{1/2\beta}}{\left(M^\gamma c_n\right)^{2+1/2\beta}} \log \Psi(\phi_n, \rho, \beta, M) \leq \frac{-M^{1/2\beta} h_n^2/2}{\left(M^\gamma c_n\right)^{2+1/2\beta}}\left(1 + o\left(1\right)\right) = -\frac{A_0(\beta)}{4} + o\left(1\right).$$

The uniformity claim can be checked using part (iii) of Lemma A.1 and the uniformity implicit in Lemma A.5. □

Lemma 3.2. *Under the assumptions of Lemma 3.1, for given* $\alpha > 0$ *and* $0 < M_1 < M_2$*, there exists a test* ϕ_n^0 *fulfilling*

$$E_{n,0}\phi_n^0 \leq \alpha + o(1) \tag{3.20}$$

and

$$\limsup_n \frac{1}{c_n^{2+1/2\beta}} \sup_{M_1 \leq M \leq M_2} M^{1/2\beta - \gamma(2+1/2\beta)} \log \Psi(\phi_n^0, \rho_{n,M}, \beta, M) \leq -\frac{A_0(\beta)}{4}. \tag{3.21}$$

Proof. Let $L_n \to \infty$ be a sequence satisfying $L_n = O\left(c_n\right)$. Using notation $M_{(i)}$ for the smoothness bounds M_i, $i = 1, 2$ in the lemma, define a grid of values

$$M_l = \left(1 - \frac{l}{L_n}\right) M_{(1)} + \frac{l}{L_n} M_{(2)},\ l = 0, \ldots, L_n.$$

Consider again the test statistic T_n given by (3.3) and observe that it depends on M and ρ (via (A.1)); we use notation $T_n\left(M, \rho\right)$ to indicate that dependence. Define the test statistics

$$T_{n,l} := T_n\left(M_l, \rho_{n,M_{l-1}}\right),\ 1 \leq l \leq L_n.$$

For fixed $\alpha > 0$, set $\alpha_n = \alpha L_n^{-1}$ and, referring to (3.6), define the family of tests

$$\psi_{n,l} := \phi_n\left(M_l, \rho_{n,M_{l-1}}, \alpha_n\right)$$

and also the test

$$\phi_n^0 = \max_{1\le l\le L_n} \psi_{n,l}.$$

First check that ϕ_n^0 is an α-test: by Bonferroni's inequality

$$P_{n,0}\left(\phi_n^0 = 1\right) \le \sum_{1\le l\le L_n} P_{n,0}\left(\psi_{n,l} = 1\right). \tag{3.22}$$

We claim that

$$P_{n,0}\left(\psi_{n,l} = 1\right) \le \alpha L_n^{-1}\left(1 + o(1)\right) \tag{3.23}$$

uniformly over $1 \le l \le L_n$. Indeed relation (3.5) now holds in view of

$$|\log \alpha_n| = \log\left(\alpha^{-1} L_n\right) = o\left(L_n^2\right) = O\left(c_n^2\right),$$

and then the property

$$P_{n,0}\left(\phi_n\left(M, \rho_{n,M}, \alpha L_n^{-1}\right) = 1\right) \le \alpha L_n^{-1}\left(1 + o(1)\right) \tag{3.24}$$

has been shown for fixed M in (3.7). The uniformity statement in Lemma 3.1 now allows to claim (3.24) for $M = M_l$ uniformly over of $1 \le l \le L_n$. A minor modification of that proof, using the fact that $M_l/M_{l-1} = 1 + o(1)$ uniformly over $1 \le l \le L_n$, now allows to replace ρ_{n,M_l} by $\rho_{n,M_{l-1}}$ in (3.24) and thus to obtain (3.23). In conjunction with (3.22) this implies

$$P_{n,0}\left(\phi_n^0 = 1\right) \le \left(1 + o(1)\right) \sum_{1\le l\le L_n} \alpha L_n^{-1} = \alpha + o(1)$$

establishing (3.20).

For fixed $M > 0$, consider now the type II error probability (for $\rho = \rho_{n,M}$)

$$\begin{aligned}
\Psi(\phi_n^0, \rho_{n,M}, \beta, M) &= \sup_{f\in\Sigma(\beta,M)\cap B_\rho} P_{n,f}\left(\phi_n^0 = 0\right) \\
&\le \sup_{f\in\Sigma(\beta,M)\cap B_\rho} \min_{1\le l\le L_n} P_{n,f}\left(\psi_{n,l} = 0\right) \\
&\le \Psi(\psi_{n,l}, \rho_{n,M}, \beta, M) \text{ for } 1 \le l \le L_n. \\
&= \Psi(\phi_n\left(M_l, \rho_{n,M_{l-1}}, \alpha_n\right), \rho_{n,M}, \beta, M) \text{ for } 1 \le l \le L_n.
\end{aligned}$$

Let $\tilde{l} := \min\{l : 1 \le l \le L_n : M_l \ge M\}$; then in particular

$$\begin{aligned}
\Psi(\phi_n^0, \rho_{n,M}, \beta, M) &\le \Psi(\phi_n\left(M_{\tilde{l}}, \rho_{n,M_{\tilde{l}-1}}, \alpha_n\right), \rho_{n,M}, \beta, M) \\
&\le \Psi(\phi_n\left(M_{\tilde{l}}, \rho_{n,M_{\tilde{l}-1}}, \alpha_n\right), \rho_{n,M_{\tilde{l}-1}}, \beta, M_{\tilde{l}}).
\end{aligned} \tag{3.25}$$

The last relation holds in view of $\Sigma\left(\beta, M\right) \subset \Sigma\left(\beta, M_{\tilde{l}}\right)$ and $B_\rho \subset B_{\tilde{\rho}}$ for $\rho = \varrho_{n,M}$, $\tilde{\rho} = \varrho_{n,M_{\tilde{l}-1}}$. Note that (3.8) in Lemma 3.1 concerns the expression $\Psi(\phi_n\left(M, \rho_{n,M}, \alpha_n\right), \rho_{n,M}, \beta, M)$. By a slight extension of that lemma, noting that $M \sim M_{\tilde{l}} \sim M_{\tilde{l}-1}$ uniformly over $M \in \left[M_{(1)}, M_{(2)}\right]$, the pair $(M, \rho_{n,M})$

there can be replaced by $\left(M_{\tilde{l}}, \rho_{n,M_{\tilde{l}-1}}\right)$. In particular, the reasoning (3.16) on the basis of the saddlepoint property (A.19) still applies. In conjunction with (3.25) we thus obtain

$$\frac{M^{1/2\beta}}{\left(M^{\gamma}c_n\right)^{2+1/2\beta}} \log \Psi(\phi_n^0, \rho_{n,M}, \beta, M) \leq -\frac{A_0(\beta)}{4} + o\,(1)$$

uniformly over $M \in \left[M_{(1)}, M_{(2)}\right]$, which implies (3.21). □

Proof of Theorem 1.2. In (3.1) set $\gamma = 0$ and note that the test ϕ_n^0 found in Lemma 3.2 has the property claimed in the theorem. □

Proof of Theorem 1.3 (ii). In accordance with (1.13) set $\gamma = (4\beta+1)^{-1}$ in (3.1) and $c_n = (A_1\,(\beta)\,d_n)^{4\beta/(4\beta+1)}$ there. Then the test ϕ_n^0 found in Lemma 3.2 has the property claimed in the theorem, in view of

$$c_n^{2+1/2\beta} = d_n^2\,(A_1\,(\beta))^2 = d_n^2\;2/A_0\,(\beta)\,. \qquad \square$$

4. Proofs for adaptation over (β, M)

4.1. The adaptive test

We first consider adaptation over a two dimensional grid of the smoothness parameter (β, M) within an interval $J = \left[\beta_{(1)}, \beta_{(2)}\right] \times \left[M_{(1)}, M_{(2)}\right] \subset \mathbb{R}_+^2$. Define the integer sequences

$$L_{1,n} = [(\log\log n)\log n]\,,\; L_{2,n} = [\log\log n]\,,\; L_n = L_{1,n} \cdot L_{2,n}$$

and a set of multiindices

$$\Lambda_n = \left\{l \in \mathbb{Z}_+^2 : l = (l_1, l_2)\,, 1 \leq l_1 \leq L_{1,n}, 1 \leq l_2 \leq L_{2,n}\right\}.$$

For $l \in \Lambda_n$ define

$$\beta_l = \left(1 - \frac{l_1}{L_{1,n}}\right)\beta_{(1)} + \frac{l_1}{L_{1,n}}\beta_{(2)},\; M_l = \left(1 - \frac{l_2}{L_{2,n}}\right)M_{(1)} + \frac{l_2}{L_{2,n}}M_{(2)}$$

and consider the grid of values $\{(\beta_l, M_l)\,, l \in \Lambda_n\} \subset J$. Consider the test statistic T_n given by (3.3) with coefficients d_j determined by (3.4) where $\lambda = \lambda\,(\rho_n, M)$, $\mu = \mu\,(\rho_n, M)$ are the solutions of (A.1) provided by Lemma A.1, and $N = (\lambda/\mu)^{1/2\beta}$. Thus the test statistic T_n is determined by β, M, and $\rho = \rho_n$; we write $T_n\,(\beta, M, \rho)$ to indicate that dependence. Consider the radius $\rho_{n,\beta,M}$ given by (1.14), namely by

$$(\rho_{n,\beta,M})^{(4\beta+1)/4\beta} = n^{-1}A_1\,(\beta)\,M^{1/4\beta}\left((2\log\log n)^{1/2} + D\right) \tag{4.1}$$

(recall that $D \in \mathbb{R}$ is arbitrary; it will be fixed throughout this section). For the radius ρ associated to (β_l, M_l) according to (4.1), we introduce an abbreviation

$$\rho_{n,l} := \rho_{n,\beta_l,M_l}. \tag{4.2}$$

For any $l \in \Lambda_n$, consider the test statistic

$$T_{n,l} := T_n\left(\beta_l, M_l, \rho_{n,l}\right). \tag{4.3}$$

Furthermore, define z_n by

$$z_n = \left(2\log\left(L_n L_{2,n}\right)\right)^{1/2} \tag{4.4}$$

and define a test by (analogously to (3.6)) by

$$\psi_{n,l} := \mathbf{1}\left\{T_{n,l} > z_n\right\}.$$

Lemma 4.1. *The family of tests $\{\psi_{n,l}, l \in \Lambda_n\}$ fulfills*

$$E_{n,0}\psi_{n,l} \le \left(L_n L_{2,n}\right)^{-1}\left(1+o(1)\right), \tag{4.5}$$

$$\Psi(\psi_{n,l}, \rho_{n,l}, \beta_l, M_l) \le \Phi\left(-D\right) + o(1) \tag{4.6}$$

uniformly over $l \in \Lambda_n$.

The proof follows below; the lemma allows to prove a preliminary version of Theorem 1.4 where the interval J is replaced by the grid $J_n := \{(\beta_l, M_l) : l \in \Lambda_n\}$, using the standard Bonferroni approach. Define the tests

$$\phi_n^0 := \max_{l\in\Lambda_n}\psi_{n,l}; \qquad \phi_n = \alpha + (1-\alpha)\,\phi_n^0.$$

For the test ϕ_n^0 we obtain

$$\begin{aligned} E_{n,0}\phi_n^0 &\le E_{n,0}\sum_{l\in\Lambda_n}\psi_{n,l} \le \sum_{l\in\Lambda_n}\left(L_n L_{2,n}\right)^{-1}\left(1+o(1)\right) \\ &= \left(1+o(1)\right)L_{2,n}^{-1} = o(1) \end{aligned}$$

and for the type II error we have uniformly over $f \in \Sigma(\beta_l, M_l) \cap B_{\rho_{n.l}}$ and $l \in \Lambda_n$

$$\begin{aligned} E_{n,f}\left(1-\phi_n^0\right) &= E_{n,f}\min_{k\in\Lambda_n}\left(1-\psi_{n,k}\right) \le E_{n,f}\left(1-\psi_{n,l}\right) \\ &\le \Phi\left(-D\right) + o(1) \end{aligned} \tag{4.7}$$

in view of (4.6). The test ϕ_n therefore fulfills

$$\begin{aligned} E_{n,0}\phi_n &= \alpha + (1-\alpha)\,E_{n,0}\phi_n^0 = \alpha + o(1), \\ E_{n,f}\left(1-\phi_n\right) &= E_{n,f}\left(1-\alpha\right)\left(1-\phi_n^0\right) \\ &= (1-\alpha)\,\Phi\left(-D\right) + o(1) \end{aligned}$$

uniformly over $f \in \Sigma(\beta_l, M_l) \cap B_{\rho_{n.l}}$ and $l \in \Lambda_n$, in view of (4.7). For the supremal type II error $\Psi(\cdot)$ according to (1.3) we thus obtain

$$\sup_{(\beta,M)\in J_n} \Psi(\phi_n, \rho_{n,\beta,M}, \beta, M) = \sup_{l\in\Lambda_n} \Psi(\phi_n, \rho_{n,l}, \beta_l, M_l) \\ \leq (1-\alpha)\,\Phi(-D) + o(1) \tag{4.8}$$

which is the claimed preliminary version of Theorem 1.4 where the interval J is replaced by the grid J_n.

Proof of Lemma 4.1. Consider first (4.5); the argument here is similar to the proof of (3.7) in Lemma 3.1. Let $\lambda = \lambda(\rho, \beta, M)$, $\mu = \mu(\rho, \beta, M,)$ be the solutions of (A.1) provided by Lemma A.1, set $N(\rho, \beta, M) := (\lambda/\mu)^{1/2\beta}$ and define $N_l := N(\rho_{n,l}, \beta_l, M_l,)$. With a view to the condition $a_m = o\left(m^{1/6}\right)$ in Lemma A.5, we first establish that $z_n \ll N_l^{1/6}$ uniformly over $l \in \Lambda_n$. We denote by $\log^{(k)} n$ a k-fold logarithm iteration: $\log^{(k)} n := \log(\log(\ldots \log n))$. It suffices to show that

$$\log(L_n L_{2,n}) \ll N_l^{1/3} \tag{4.9}$$

uniformly. Here we have

$$\log(L_n L_{2,n}) = \log L_{1,n} + 2\log L_{2,n} \\ \asymp \log^{(2)} n + \log^{(3)} n + \log^{(3)} n \\ \asymp \log^{(2)} n; \tag{4.10}$$

On the other hand, the asymptotics of N_l is given by (A.2) where part (iii) of Lemma A.1 states uniformity of the $1+o(1)$ terms over $(\beta, M) \in J$. Then according to (A.2) we have

$$N_l = \left(\frac{(4\beta_l+1)\,M_l}{\rho_{n,l}}\right)^{1/2\beta_l} (1+o(1))$$

uniformly over $l \in \Lambda_n$. In conjunction with (4.1) this implies that a lower bound for the rate of $N_l \to \infty$ is given by $N_l \gg n^{2/(4\beta_l+1)}$, which implies

$$N_l \gg n^{2/\left(4\beta_{(2)}+1\right)} \left(\log^{(2)} n\right)^{-1/\left(4\beta_{(1)}+1\right)}.$$

This and (4.10) establish (4.9). As in the proof of Lemma 3.1, setting $T_{n,l} = U_{n,l}$ as in (3.9) for every $l \in \Lambda_n$, from Lemma A.5 we now obtain

$$P_{n,0}(U_{n,l} \geq z_n) \leq \exp\left(-z_n^2/2\right)(1+o(1))$$

uniformly over $l \in \Lambda_n$. The relation $\exp\left(-z_n^2/2\right) = (L_n L_{2,n})^{-1}$ now establishes (4.5).

To prove (4.6), we first note the inequality (3.15) obtained in the proof of Lemma 3.1, applied to the test $\psi_{n,l}$

$$\Psi(\psi_{n,l},\rho,\beta,M) \leq \sup_{f\in\Sigma(\beta,M)\cap B_\rho^0} P_{n,f}\left(L_n(d,f)+U_n \leq \check{z}_n\right) + \Phi\left(-t_n \tau_{n,l}^{-1/2}\right) \tag{4.11}$$

where t_n is to be chosen and $\check{z}_n = z_n + t_n$, and $(\rho,\beta,M) = (\rho_{n,l},\beta_l,M_l)$, and we note that $\tau_{n,l} = \sup_{f\in B_\rho^0} \mathrm{Var}\left(T_n^0(f)\right)$ with $T_n^0(f)$ defined by (3.13) now depend on l (via the coefficient vector d). As before we can claim

$$\inf_{f\in\Sigma(\beta,M)\cap B_\rho^0} L_n\left(d,f\right) = S_n = n\left(T_{(1)}/2\right)^{1/2}$$

where S_n is given by (A.20) and $T_{(1)}$ by (A.4) in Lemma A.1, for the pertaining values of ρ_l,β_l,M_l. This allows to infer from (4.11) relation (3.17) again, writing $S_{n,l} = S_n$:

$$\Psi(\psi_{n,l},\rho_{n,l},\beta_l,M_l) \leq P_n\left(S_{n,l}+U_{n,l} \leq \check{z}_n\right) + \Phi\left(-t_n \tau_{n,l}^{-1/2}\right). \tag{4.12}$$

Lemma A.1 now yields a relation, holding uniformly over $l \in \Lambda_n$

$$S_{n,l} = n\rho_{n,l}^{1+1/4\beta} M_l^{-1/4\beta}\left(\frac{A_0(\beta_l)}{2}\right)^{1/2}\left(1+o(1)\right). \tag{4.13}$$

Invoking (1.14) and (1.11) we find

$$\begin{aligned} S_{n,l} &= A_1\left(\beta_l\right)\left(\frac{A_0(\beta_l)}{2}\right)^{1/2}\left(\left(2\log\log n\right)^{1/2}+D\right)\left(1+o(1)\right) \\ &= \left(\left(2\log\log n\right)^{1/2}+D\right)\left(1+o(1)\right). \end{aligned} \tag{4.14}$$

As a consequence of Lemma A.1 (iii), the $o(1)$ term here is of algebraic rate, meaning there is $\gamma > 0$ such that this term is actually $o\left(n^{-\gamma}\right)$ uniformly over $l \in \Lambda_n$. This holds because $(\beta_l, M_l) \in J$ and because there exists $\gamma_1 > 0$ such that $\rho_{n,l}$ given by (4.2), (4.1) satisfies $\rho_{n,l} = o\left(n^{-\gamma_1}\right)$ uniformly over $l \in \Lambda_n$. This implies that (4.14) can be strengthened to

$$S_{n,l} = \left(2\log\log n\right)^{1/2} + D + o(1). \tag{4.15}$$

Furthermore, for $\tau_{n,l}$ we find for $\rho = \rho_{n,l}$

$$\begin{aligned} \tau_n &:= \sup_{f\in B_\rho^0} \mathrm{Var}\left(T_n^0(f)\right) = n\rho O\left(\rho^{1/2\beta}\right) \\ &= O\left(n^{-1/(4\beta+1)}\left(\log^{(2)} n\right)^{(2\beta+1)/(4\beta+1)}\right) \\ &= O\left(n^{-1/(4\beta_{(2)}+1)}\log n\right). \end{aligned} \tag{4.16}$$

For z_n we find in view of (4.4)

$$\begin{aligned} z_n &= (2\log(L_n L_{2,n}))^{1/2} \\ &= (2\log L_{1,n} + 4\log L_{2,n})^{1/2} \\ &= \left(2\log^{(2)} n + 2c\log^{(3)} n + 4\log^{(3)} n\right)^{1/2}. \end{aligned}$$

The inequality

$$(x+y)^{1/2} - x^{1/2} \le \frac{1}{2x^{1/2}} y \text{ for } x, y > 0$$

now implies

$$z_n = \left(2\log^{(2)} n\right)^{1/2} + o(1).$$

As a consequence, in conjunction with (4.15) we obtain

$$\begin{aligned} \check{z}_n - S_{n,l} &= z_n + t_n - S_{n,l} \\ &= t_n - D + o(1). \end{aligned}$$

It now suffices to choose $t_n = o(1)$ such that $t_n^{-2}\tau_n = o(1)$. The choice $t_n = (\log n)^{-1}$ clearly qualifies in view of (4.16). Now invoking (4.12) and the Lindeberg-Feller CLT for $U_{n,l}$ (with uniformity over $l \in \Lambda_n$, cp. Lemma A.4) concludes the proof. □

We now consider adaptation over the full interval $J = \left[\beta_{(1)}, \beta_{(2)}\right] \times \left[M_{(1)}, M_{(2)}\right]$ for the smoothness parameter (β, M). In terms of the grid $\{(\beta_l, M_l), l \in \Lambda_n\}$, define sets

$$V_l = [\beta_l, \beta_{l+1}] \times [M_{l-1}, M_l].$$

Note that $J \subset \bigcup_{l \in \Lambda_n} V_l$ and for $(\beta, M) \in V_l$ we have

$$\Sigma(\beta, M) \subset \Sigma(\beta_l, M_l). \tag{4.17}$$

We also set $\gamma_n := \left(\log^{(2)} n\right)$.

Lemma 4.2. *Suppose that for $(\beta, M) \in V_l$, the radius $\rho_{n,\beta,M}$ is given by (4.1). Then*

$$\sup_{l \in \Lambda_n} \sup_{(\beta,M) \in V_l} \left| \frac{\rho_{n,l}}{\rho_{n,\beta,M}} - 1 \right| = O(\gamma_n).$$

Proof. Write $\rho_n = \rho_{n,\beta,M}$ and note that

$$\log \rho_n = \frac{4\beta}{4\beta+1}\left(-\log n + \log A_1(\beta) + \frac{1}{4\beta}\log M + \log\left((2\log\log n)^{1/2} + D\right)\right) \tag{4.18}$$

and analogously for $\log \rho_{n,l}$. It suffices to prove that

$$|\log \rho_n - \log \rho_{n,l}| = O(\gamma_n)$$

uniformly over $(\beta, M) \in V_l$, $l \in \Lambda_n$. According to (4.18), the difference $\log \rho_n - \log \rho_{n,l}$ splits into four terms, say $T_j,\ j = 1, \ldots, 4$. Define a function $g(x) = 4x/(4x+1)$ for $x > 0$; we then have

$$T_1 = \left(g\left(\beta_l\right) - g\left(\beta\right)\right) \log n.$$

The function g satisfies $|g(x) - g(y)| \leq 4\,|x - y|$ for $x, y > 0$, so that

$$|T_1| \leq 4 L_{1,n}^{-1} \log n = \frac{4 \log n}{\left[\log n \left(\log^{(2)} n\right)^c\right]} = O\left(\gamma_n\right).$$

Next, we have

$$T_2 = g\left(\beta\right) \log A_1\left(\beta\right) - g\left(\beta_l\right) \log A_1\left(\beta_l\right).$$

Since $A_1\left(\beta\right) = \left(A_0\left(\beta\right)/2\right)^{-1/2}$ with $A_0\left(\beta\right)$ given by (1.7), we have

$$\log A_1\left(\beta\right) = \frac{1}{2} \log \frac{\left(4\beta + 1\right)^{1 + 1/2\beta}}{2\beta + 1} = \frac{2\beta + 1}{4\beta} \log \frac{4\beta + 1}{2\beta + 1}.$$

Clearly the function $\beta \mapsto g\left(\beta\right) \log A_1\left(\beta\right)$ has a bounded derivative over $\beta \in \left[\beta_{(1)}, \beta_{(2)}\right]$ so that

$$|T_2| = O\left(L_{1,n}^{-1}\right) = O\left(L_{1,n}^{-1}\right) = O\left(\frac{1}{\left[\log n \left(\log^{(2)} n\right)^c\right]}\right) = O\left(\gamma_n\right).$$

For the term T_3 we obtain

$$T_3 = \frac{1}{4\beta + 1} \log M - \frac{1}{4\beta_l + 1} \log M_l.$$

For this function of (β, M), the same Lipschitz type argument yields

$$|T_3| = O\left(L_{1,n}^{-1}\right) + O\left(L_{2,n}^{-1}\right) = O\left(\gamma_n\right) + O\left(\frac{1}{\log^{(2)} n}\right) = O\left(\gamma_n\right).$$

For the term T_4 we obtain

$$\begin{aligned} T_4 &= \left(g\left(\beta\right) - g\left(\beta_l\right)\right) \log\left(\left(2 \log \log n\right)^{1/2} + D\right) \\ &= O\left(L_{1,n}^{-1}\right) \log\left(\left(2 \log \log n\right)^{1/2} + D\right) = O\left(\frac{\log^{(3)} n}{L_{1,n}}\right) = O\left(\gamma_n\right). \end{aligned}$$

These bounds for $T_j,\ j = 1, \ldots, 4$ prove the lemma. □

Define also another radius sequence as

$$\tilde{\rho}_{n,l} := \rho_{n,l}\left(1 - \eta_n\right) \text{ where } \eta_n = \left(\log^{(2)} n\right)^{-c} \tag{4.19}$$

for a constant $c \in (1/2, 1)$. We thus have $\eta_n \gg \gamma_n$. For $(\beta, M) \in V_l$ we now have in view of Lemma 4.2

$$\frac{\tilde{\rho}_{n,l}}{\rho_{n,\beta,M}} = \rho_{n,l}\,(1-\eta_n)\,\rho_{n,l}^{-1}\,(1+O\,(\gamma_n)) = 1-\eta_n+O\,(\gamma_n)$$
$$= 1-\eta_n+o\,(\eta_n)\,.$$

As a consequence, there exists n_0 such that for $n \geq n_0$ we have

$$\rho_{n,\beta,M} \geq \tilde{\rho}_{n,l} \text{ for all } (\beta, M) \in V_l,\, l \in \Lambda_n.$$

In conjunction with (4.17) we obtain for $n \geq n_0$ and $(\beta, M) \in V_l$

$$\Sigma\,(\beta, M) \cap B\,(\rho_{n,\beta,M}) \subset \Sigma\,(\beta_l, M_l) \cap B\,(\tilde{\rho}_{n,l})\,. \tag{4.20}$$

In analogy to (4.3), for any $l \in \Lambda_n$ define the test statistic

$$\tilde{T}_{n,l} := T_n\,(\beta_l, M_l, \tilde{\rho}_{n,l})$$

and for z_n be defined by (4.4) define a test by

$$\tilde{\psi}_{n,l} := \mathbf{1}\left\{\tilde{T}_{n,l} > z_n\right\}.$$

Lemma 4.3. *The family of tests* $\left\{\tilde{\psi}_{n,l}, l \in \Lambda_n\right\}$ *fulfills*

$$E_{n,0}\tilde{\psi}_{n,l} = (L_n L_{2,n})^{-1}\,(1+o(1))\,, \tag{4.21}$$

$$\sup_{(\beta,M)\in V_l} \Psi(\tilde{\psi}_{n,l}, \rho_{n,\beta,M}, \beta, M) \leq \Phi\,(-D) + o(1) \tag{4.22}$$

uniformly over $l \in \Lambda_n$.

Proof. Consider first (4.21); the argument here follows verbatim the proof of (4.5) in Lemma 4.1, based on the fact $\tilde{\rho}_{n,l} = \rho_{n,l}\,(1+o\,(1))$.

To prove (4.22), observe that, in view of (4.20)

$$\Psi(\tilde{\psi}_{n,l}, \rho_{n,\beta,M}, \beta, M) = \sup_{f\in\Sigma(\beta,M)\cap B(\rho_{n,\beta,M})} \left(1 - E_{n,f}\tilde{\psi}_{n,l}\right)$$
$$\leq \Psi(\tilde{\psi}_{n,l}, \tilde{\rho}_{n,l}, \beta_l, M_l).$$

Note that the quantity $\Psi(\tilde{\psi}_{n,l}, \tilde{\rho}_{n,l}, \beta_l, M_l)$ is an exact analog of $\Psi(\psi_{n,l}, \rho_{n,l}, \beta_l, M_l)$ considered in (4.11): in both cases, the test ψ is defined by the statistic $T_n\,(\beta, M, \rho)$ (via (3.3), (A.1), (3.4)) with parameters (β_l, M_l) and radius ρ specified either as $\tilde{\rho}_{n,l}$ or as $\rho_{n,l}$, and with the same critical value z_n. The proof of (4.22) is therefore entirely analogous to that of (4.6), where the change from $\rho_{n,l}$ to $\tilde{\rho}_{n,l} = \rho_{n,l}\,(1-\eta_n)$ has to be taken into account. Let $\tilde{S}_{n,l}$ for $l \in \Lambda_n$ be the appropriate modification of the saddlepoint value $S_{n,l}$, i.e. $\tilde{S}_{n,l}$ is given by

(A.20) for the pertaining values $(\rho,\beta,M)=(\tilde{\rho}_{n,l},\beta_l,M_l)$. Analogously to (4.13) we obtain

$$\tilde{S}_{n,l}=n\tilde{\rho}_{n,l}^{1+1/4\beta}M_l^{-1/4\beta}\left(\frac{A_0(\beta_l)}{2}\right)^{1/2}(1+o(1)).$$

where $o(1)$ is uniform over $l\in\Lambda_n$. Analogously to (4.14) we find

$$\tilde{S}_{n,l}=(1-\eta_n)^{1+1/4\beta}\left((2\log\log n)^{1/2}+D\right)(1+o(1)). \tag{4.23}$$

Here $(1-\eta_n)^{1+1/4\beta}=1+O(\eta_n)$ uniformly over $\beta\in J$, and the choice $\eta_n=\left(\log^{(2)}n\right)^{-c}$, $c>1/2$ implies

$$\eta_n(\log\log n)^{1/2}=o(1).$$

This, in conjunction with same argument about the $o(1)$ term in (4.23) as used previously for (4.14) allows to obtain the analog of (4.15), that is

$$\tilde{S}_{n,l}=(2\log\log n)^{1/2}+D+o(1). \tag{4.24}$$

The rest of the proof follows verbatim that of Lemma 4.1 after (4.15). □

The proof of Theorem 1.4 is now concluded in the same way which led up to (4.8) for the preliminary version over the grid $J_n:=\{(\beta_l,M_l):l\in\Lambda_n\}$. Define the test $\tilde{\phi}_n^0$ and the randomized test $\tilde{\phi}_n$ by

$$\tilde{\phi}_n^0:=\max_{l\in\Lambda_n}\tilde{\psi}_{n,l};\qquad \tilde{\phi}_n=\alpha+(1-\alpha)\,\tilde{\phi}_n^0.$$

For $\tilde{\phi}_n^0$ we obtain

$$E_{n,0}\tilde{\phi}_n^0\le E_{n,0}\sum_{l\in\Lambda_n}\psi_{n,l}\le(1+o(1))\,L_{2,n}^{-1}=o(1)$$

and for the type II error we have uniformly over $f\in\Sigma(\beta,M)\cap B_{\rho_{n,\beta,M}}$ and $(\beta,M)\in V_l$, $l\in\Lambda_n$

$$\begin{aligned}
E_{n,f}\left(1-\tilde{\phi}_n^0\right)&=E_{n,f}\min_{k\in\Lambda_n}\left(1-\tilde{\psi}_{n,k}\right)\le E_{n,f}\left(1-\tilde{\psi}_{n,l}\right)\\
&\le\sup_{f\in\Sigma(\beta,M)\cap B(\rho_{n,\beta,M})}\left(1-E_{n,f}\tilde{\psi}_{n,l}\right)=\Psi(\tilde{\psi}_{n,l},\rho_{n,\beta,M},\beta,M)\\
&\le\Phi(-D)+o(1)
\end{aligned}$$

in view of (4.6). The test $\tilde{\phi}_n$ therefore fulfills

$$\begin{aligned}
E_{n,0}\tilde{\phi}_n&=\alpha+(1-\alpha)\,E_{n,0}\tilde{\phi}_n^0=\alpha+o(1),\\
E_{n,f}\left(1-\tilde{\phi}_n\right)&=E_{n,f}\,(1-\alpha)\left(1-\tilde{\phi}_n^0\right)\\
&=(1-\alpha)\,\Phi(-D)+o(1)
\end{aligned}$$

uniformly over $f\in\Sigma(\beta,M)\cap B_{\rho_{n,\beta,M}}$ and $(\beta,M)\in J$. This implies Theorem 1.4.

4.2. Lower risk bound

The proof of lower risk bound of Proposition 1.4 in [25] does not actually cover the present case of a Sobolev ellipsoid. Indeed in [25] the problem of adaptive testing is considered for parameters spaces given by a quadruplet (p, q, r, s), where the corresponding space is an l_p-ellipsoid of smoothness r with an l_q -ellipsoids of smoothness s removed. In this generality, the prior measures required for the lower risk bound have to be non-Gaussian, but the model of the present paper, which corresponds to the case $p = q = 2$, $r = \beta$ and $s = 0$, calls for Gaussian prior measures on the ellipsoids $\Sigma(\beta, M)$. In [25] the two sets of quadruplets $\Xi_{G_{01}}$ and $\Xi_{G_{02}}$ treated ([25], p. 278) exclude the present case which is on the boundary between these. In this section we will provide the necessary details for this boundary case in an abbreviated fashion.

As in [25], in the ellipsoid $\Sigma(\beta, M)$ the parameter $M = M_0$ will be considered fixed ($M_0 \in \left(M_{(1)}, M_{(2)}\right)$ here) and $\beta \in \left[\beta_{(1)}, \beta_{(2)}\right]$. Set $L_n = \log n / \log^{(2)} n$ and consider a grid of values of the smoothness parameter β as

$$\beta_l = \left(1 - \frac{l}{L_n}\right)\beta_{(1)} + \frac{l}{L_n}\beta_{(2)},\ l = 1, \ldots, L_n.$$

For proving Proposition 1.4, it suffices to prove the inequality (1.15) with the supremum over $\beta \in \left[\beta_{(1)}, \beta_{(2)}\right]$ replaced by a supremum over β_l, $l = 1, \ldots, L_n$. Introduce notation, similar to (4.2), but with $M = M_0$ now fixed,

$$\rho_{n,l} := \rho_{n,\beta_l,M_0} = \left(n^{-1} A_1(\beta_l) M_0^{1/4\beta_l}\left((2\log\log n)^{1/2} + D\right)\right)^{4\beta_l/(4\beta_l+1)}. \tag{4.25}$$

The first step is to find prior probability measures $Q_{n,l}$ such that

$$Q_{n,l}\left(\Sigma(\beta_l, M_0) \cap B(\rho_{n,l})\right) = 1 + o(1) \tag{4.26}$$

uniformly in l. In the prior measure $Q_{n,l}$, the f_j are independent Gaussian

$$f_j \sim N\left(0, f_{0,j,l}^2\right),\ j = 1, \ldots, n$$

and $f_{0,j,l}^2 = \left(\lambda - \mu j^{2\beta_l}\right)_+$ (cp. (A.18)), where λ, μ are the unique positive solutions of (A.1) where the triplet (ρ, β, M) is set as follows. Define $M_1 = M_0 - \varepsilon$ for some $\varepsilon > 0$ and $\tilde{\rho}_{n,l} := \rho_{n,l}(1 + \eta_n)$ with $\eta_n = \left(\log^{(2)} n\right)^{-c}$ for some $c \in (1/2, 1)$ (cp. 4.19). Then $(\rho, \beta, M) = (\tilde{\rho}_{n,l}, \beta_l, M_1)$ in (A.1), and then (4.26) follows by standard arguments similar to Lemma 2.2. After switching to the model of random f_j, the testing problem becomes the one of testing between H_0: $y_j \sim N\left(0, n^{-1}\right)$, and H_0: $y_j \sim N\left(0, f_{0,j,l}^2 + n^{-1}\right)$, or equivalently between H_0: $y_j \sim N(0, 1)$, and H_0: $y_j \sim N\left(0, n f_{0,j,l}^2 + 1\right)$. Denote $\gamma_{j,l}^2 = n f_{0,j,l}^2$, and let $\Lambda_{n,l}$ be the corresponding log-likelihood ratio:

$$\Lambda_{n,l} = \log \prod_{j=1}^{n} \frac{dN\left(0, \gamma_{j,l}^2 + 1\right)}{dN(0, 1)}. \tag{4.27}$$

Note that setting $N_l = (\lambda/\mu)^{1/2\beta_l}$, we have $\gamma_{j,l}^2 = 0$ for $j > N_l$. Observe that as $h \to 0$

$$\log \frac{dN(0,1+h)}{dN(0,1)}(y) = -\frac{1}{2}\log(1+h) + \frac{y^2}{2}\left(1 - \frac{1}{1+h}\right)$$
$$= -\frac{1}{2}\left(h - \frac{h^2}{2} + O\left(h^3\right)\right) + \frac{y^2}{2}\left(h - h^2 + O\left(h^3\right)\right). \tag{4.28}$$

To use that expansion with (4.27) we have to show

$$\sup_{j=1,\dots,n,\ l=1,\dots,L_n} \gamma_{j,l}^2 \to 0 \text{ uniformly in } j \text{ and } l. \tag{4.29}$$

Indeed

$$\gamma_{j,l}^2 = n f_{0,j,l}^2 = n\left(\lambda - \mu j^{2\beta_l}\right)_+$$
$$= n\lambda\left(1 - \left(j/N_l\right)^{2\beta_l}\right)_+ \leq n\lambda$$

Furthermore according to (A.3) we have, for C not depending on n and l

$$n\lambda = n\tilde{\rho}_{n,l}^{1+1/2\beta_l} M_1^{-1} \frac{2\beta_l + 1}{\left(4\beta_l+1\right)^{1/2\beta} 2\beta_l}\left(1 + o(1)\right) \text{ uniformly in } l$$
$$= n\left(n^{-1}\left(\left(2\log\log n\right)^{1/2} + D\right)\right)^{(4\beta_l+2)/(4\beta_l+1)} C$$
$$= n^{-1/(4\beta_l+1)}\left(\left(2\log\log n\right)^{1/2} + D\right)^{(4\beta_l+2)/(4\beta_l+1)} C$$
$$= o(1) \text{ uniformly in } l$$

which implies (4.29). We now expand (4.27), for $y_j \sim N(0,1)$

$$\log \prod_{j=1}^n \frac{dN\left(0, \gamma_{j,l}^2 + 1\right)}{dN(0,1)}(y_j)$$
$$= \sum_{j=1}^n \left\{-\frac{1}{2}\left(\gamma_{j,l}^2 - \frac{\gamma_{j,l}^4}{2} + O\left(\gamma_{j,l}^6\right)\right) + \frac{y_j^2}{2}\left(\gamma_{j,l}^2 - \gamma_{j,l}^4 + O\left(\gamma_{j,l}^6\right)\right)\right\}$$
$$= \sum_{j=1}^n \left\{\frac{\gamma_{j,l}^2}{2}\left(y_j^2 - 1\right) - \frac{\gamma_{j,l}^4}{4}\left(2y_j^2 - 1\right)\right\} + \sum_{j=1}^n \left(y_j^2 + 1\right) O\left(\gamma_{j,l}^6\right).$$

Setting $z_j = \left(y_j^2 - 1\right)/\sqrt{2}$, such that $E z_j = 0$, $\operatorname{Var}(z_j) = 1$, this can be written

$$\Lambda_{n,l} = \sum_{j=1}^n \left\{\frac{\gamma_{j,l}^2}{\sqrt{2}} z_j - \frac{1}{2}\frac{\gamma_{j,l}^4}{2}\right\} - \sum_{j=1}^n \frac{\gamma_{j,l}^4}{\sqrt{2}} z_j + \sum_{j=1}^n \left(z_j + 2\right) O\left(\gamma_{j,l}^6\right) \tag{4.30}$$
$$= \left\{\sum_{j=1}^n \frac{\gamma_{j,l}^2}{\sqrt{2}} z_j - \frac{1}{2}\sum_{j=1}^n \frac{\gamma_{j,l}^4}{2}\right\} + R_{n,l} \tag{4.31}$$

with $R_{n,l}$ being the last two terms on the r.h.s. of (4.30). Denote

$$\tilde{S}_{n,l} := \left(\sum_{j=1}^{n} \frac{\gamma_{j,l}^4}{2} \right)^{1/2} = \frac{n}{\sqrt{2}} \left(\sum_{j=1}^{n} f_{0,j,l}^4 \right)^{1/2}. \tag{4.32}$$

This coincides with the saddlepoint value S_n defined in (A.20) for the current values of β, M and ρ_n (i.e. $\beta = \beta_l, M = M_1$ and $\rho_n = \tilde{\rho}_{n,l}$). The approximation (4.31) suggests a normal approximation $\Lambda_{n,l} \approx N\left(-\frac{1}{2}\tilde{S}_{n,l}^2, \tilde{S}_{n,l}^2\right)$ but from Lemma A.1 or (4.15) it can be seen that $\tilde{S}_{n,l} \asymp \left(\log^{(2)} n\right)^{1/2}$, thus the normal approximation will be established in the sense

$$\tilde{\lambda}_{n,l} := \left(\Lambda_{n,l} + \frac{1}{2}\tilde{S}_{n,l}^2 \right) / \tilde{S}_{n,l} \rightsquigarrow N(0,1) \tag{4.33}$$

(convergence in distribution, uniform in l).

We formulate a version of the lower bound for type II error proved in Corollary 7.2 of [25]. Introduce the following notation: let $\Phi(x, y; r)$ be the two-dimensional distribution function of jointly normal random variables Z_1, Z_2, each having marginal distribution $N(0,1)$ and correlation r, and let $\Phi(x)$ be the distribution function of $N(0,1)$. Let π_l^n, $l = 0, \ldots, L_n$ be a set of probability measures on a sample space $(\Omega_n, \mathfrak{A}_n)$, where $\pi_l^n \ll \pi_0^n$ and define $\Lambda_{n,l} = \log \frac{d\pi_l^n}{d\pi_0^n}$, $l = 1, \ldots, L_n$. Define the mixture distribution

$$\pi^n := L_n^{-1} \sum_{1 \le l \le L_n} \pi_l^n.$$

Proposition 4.1. *[25] Let there be given a family $\{u_{n,l}, l = 1, \ldots, L_n\}$ such that for some $D > 0$*

$$\max_{1 \le l \le L_n} \left| u_{n,l} - (2 \log L_n)^{1/2} - D \right| = o(1) \tag{4.34}$$

and a family $\{\rho_{n,kl}, k, l = 1, \ldots, L_n\}$ satisfying $\rho_{n,kl} = \rho_{n,lk}$ such that

$$\max_{1 \le l < k \le L_n} \rho_{n,kl} u_{n,k} u_{n,l} = o(1). \tag{4.35}$$

Furthermore, assume that the family of random variables

$$\tilde{\lambda}_{n,l} := \left(\Lambda_{n,l} + \frac{1}{2} u_{n,l}^2 \right) / u_{n,l}$$

fulfills

$$\sum_{1 \le l \le L_n} \sup_{x \in \mathbb{R}} \left| P\left(\tilde{\lambda}_{n,l} < x | \pi_0^n \right) - \Phi(x) \right| = o(1), \tag{4.36}$$

$$\sum_{1 \le l < k \le L_n} \sup_{x,y \in \mathbb{R}} \left| P\left(\tilde{\lambda}_{n,l} < x, \tilde{\lambda}_{n,k} < y | \pi_0^n \right) - \Phi(x, y; \rho_{n,kl}) \right| = o(1). \tag{4.37}$$

Assume $\alpha \in (0,1)$. *Then for any sequence of tests* ϕ_n *satisfying* $E[\phi_n|\pi_0^n] \leq \alpha + o(1)$ *one has*

$$E[1-\phi_n|\pi^n] \geq (1-\alpha)\,\Phi(-D) + o(1).$$

The measures π_0^n and π_l^n are those where $y_j \sim N(0,1)$ and $y_j \sim N\left(0, f_{0,j,l}^2 + n^{-1}\right)$ (independent) respectively. Setting $u_{n,l} = \tilde{S}_{n,l}$ from (4.32), we find analogously to (4.24)

$$\tilde{S}_{n,l} = (2\log\log n)^{1/2} + D + o(1) \tag{4.38}$$

uniformly over $l = 1,\dots,L_n$ (note that in (4.24) the radius is given by $\rho_{n,l}(1-\eta_n)$ whereas now we use $\tilde{\rho}_{n,l} = \rho_{n,l}(1+\eta_n)$ with $\rho_{n,l}$ given by (4.25), but it can be checked that (4.38) still holds true). Furthermore we have

$$\begin{aligned}(2\log L_n)^{1/2} &= \left(2\log\left(\log n/\log^{(2)} n\right)\right)^{1/2}\\ &= 2^{1/2}\left(\log^{(2)} n - \log^{(3)} n\right)^{1/2}.\end{aligned}$$

Set $t_n = \log^{(2)} n$ and note that

$$0 \leq t_n^{1/2} - (t_n - \log t_n)^{1/2} \leq (\log t_n)\,\frac{1}{2}\,(t_n - \log t_n)^{-1/2} = o(1).$$

Hence

$$\begin{aligned}(2\log L_n)^{1/2} &= (2\log\log n)^{1/2} + o(1),\\ S_{n,l} - (2\log L_n)^{1/2} &\to D\end{aligned}$$

which establishes (4.34) for this choice of $u_{n,l}$.

To determine the family $\rho_{n,kl}$, consider the expansion (4.31) of the log-likelihood ratios $\Lambda_{n,l}$. Define $\rho_{n,kl}$ through

$$\rho_{n,kl}u_{n,l}u_{n,k} = \mathrm{Cov}\left(\Lambda_{n,l} - R_{n,l}, \Lambda_{n,k} - R_{n,k}\right).$$

We then have

$$\begin{aligned}\rho_{n,kl}u_{n,l}u_{n,k} &= E\left(\sum_{j=1}^{n}\frac{\gamma_{j,l}^2}{\sqrt{2}}z_j\right)\left(\sum_{j=1}^{n}\frac{\gamma_{j,k}^2}{\sqrt{2}}z_j\right)\\ &= \frac{1}{2}\sum_{j=1}^{n}\gamma_{j,k}^2\gamma_{j,l}^2 = \frac{n^2}{2}\sum_{j=1}^{n} f_{0,j,k}^2 f_{0,j,l}^2.\end{aligned}$$

Lemma 4.4. *The current choice* $L_n = \log n/\log^{(2)} n$ *implies that condition (4.35) is fulfilled.*

Proof. We first show that

$$\frac{N_l}{N_k} = o(1) \tag{4.39}$$

uniformly over $k < l$. Indeed, since $\beta_k < \beta_l$ and since N is related to the bandwidth parameter in a smoothing problem, we expect $N_k > N_l$. Note that according to (A.2), we have

$$\begin{aligned} \log N_k &= \left(-\frac{1}{2\beta_k}\log\rho_{n,k}\right)(1+o(1)) \\ &\geq \frac{2}{4\beta_k+1}\log n - C_1\log^{(3)} n + C_2 \end{aligned}$$

uniformly over $k = 1, \ldots, L$ and for N_l

$$\log N_l \leq \frac{2}{4\beta_l+1}\log n + C_3.$$

Accordingly

$$\begin{aligned} \frac{N_l}{N_k} &\leq \exp\left(\log N_l - \log N_k\right) \\ &\leq \exp\left(\left(\frac{1}{4\beta_l+1} - \frac{1}{4\beta_k+1}\right)2\log n + C_1\log^{(3)} n + C_4\right). \end{aligned}$$

The function $g(t) = (4t+1)^{-1}$ and its derivative are monotone decreasing for $t > 0$, hence

$$\begin{aligned} \frac{1}{4\beta_k+1} - \frac{1}{4\beta_l+1} &> |\beta_l - \beta_k|\,|g'(\beta_l)| = L_n^{-1}\frac{4}{(4\beta_l+1)^2} \\ &> L_n^{-1}C, \end{aligned}$$

where C does not depend on n and k, l hence

$$\begin{aligned} \frac{N_l}{N_k} &\leq \exp\left(-L_n^{-1}C\log n + C_1\log^{(3)} n + C_4\right) \\ &= \exp\left(-C\log^{(2)} n + C_1\log^{(3)} n + C_4\right) \end{aligned}$$

Clearly the exponent tends to $-\infty$, so (4.39) is proved. Now we have

$$\begin{aligned} \rho_{n,kl}u_{n,k}u_{n,l} &= \frac{n^2}{2}\sum_{j=1}^{n} f_{0,j,k}^2 f_{0,j,l}^2 \\ &= \frac{n^2}{2}\lambda_l\lambda_k\sum_{j=1}^{n}\left(1-(j/N_k)^{2\beta}\right)_+\left(1-(j/N_l)^{2\beta}\right)_+ \\ &\leq \frac{n^2}{2}\lambda_l\lambda_k\min(N_k, N_l). \end{aligned}$$

The asymptotics for λ_l, λ_k shown in (A.3) gives $\lambda_l \asymp N_l^{-(2\beta_l+1)}$. Observe that for $l = k$ the above term would be of the order

$$n^2 N_l^{-2(2\beta_l+1)} N_l = n^2 N_l^{-(4\beta_l+1)} \asymp 1.$$

We can use (4.39) to show it is $o(1)$ for $k < l$. Equivalently, it can be shown, analogously to (4.39), that

$$\frac{\lambda_k}{\lambda_l} = o\left(1\right)$$

and then it follows

$$\frac{n^2}{2}\lambda_l\lambda_k \min\left(N_k, N_l\right) \le o\left(1\right)\frac{n^2}{2}\lambda_l^2 N_l \asymp o\left(1\right). \qquad \square$$

We still need to establish the asymptotic normalities required, namely (4.36) and (4.37). In the sequel we use notation C for generic constant which does not depend on n, j and l, the value of which can change, even on the same line.

Proof of (4.36). Recall that $\beta \in \left[\beta_{(1)}, \beta_{(2)}\right]$ and define $\delta_n^2 := n^{-1/\left(8\beta_{(2)}+1\right)}$ so that for some $\varepsilon > 0$ we have

$$n^{-1/(4\beta_l+1)} \le n^{-\varepsilon}\delta_n^2 \tag{4.40}$$

for all $l = 1, \dots, L_n$. It suffices to show that

$$\sup_{l=1,\dots,L_n}\sup_{x\in\mathbb{R}}\left|P\left(\tilde{\lambda}_{n,l} < x|\pi_0^n\right) - \Phi\left(x\right)\right| = O\left(\delta_n\right) \tag{4.41}$$

since $L_n\delta_n = o\left(1\right)$. We have in view of (4.31)

$$\tilde{\lambda}_{n,l} = S_{n,l}^{-1}\left(\sum_{j=1}^{n}\frac{\gamma_{j,l}^2}{\sqrt{2}}z_j + R_{n,l}\right),$$

$$R_{n,l} = -\sum_{j=1}^{n}\frac{\gamma_{j,l}^4}{\sqrt{2}}z_j + \sum_{j=1}^{n}\left(z_j+2\right)O\left(\gamma_{j,l}^6\right). \tag{4.42}$$

Note that by (4.28) and (4.29) the $O\left(\gamma_{j,l}^6\right)$ term is uniform, that is $\left|O\left(\gamma_{j,l}^6\right)\right| \le C\gamma_{j,l}^6$ where C does not depend on j, l and n. To establish (4.41), it suffices to show

$$\sup_{x\in\mathbb{R}}\left|P\left(S_{n,l}^{-1}\sum_{j=1}^{n}\frac{\gamma_{j,l}^2}{\sqrt{2}}z_j < x\right) - \Phi\left(x\right)\right| = O\left(\delta_n\right), \tag{4.43}$$

uniformly in l, and in addition, using standard arguments, that

$$P\left(\left|S_{n,l}^{-1}R_{n,l}\right| \ge \delta_n\right) = O\left(\delta_n\right) \tag{4.44}$$

uniformly in l. To establish the latter, denote $NR_{n,l}$ the nonrandom term of $S_{n,l}^{-1}R_{n,l}$: its absolute value is bounded by

$$|NR_{n,l}| \leq CS_{n,l}^{-1}\left|\sum_{j=1}^{n}\gamma_{j,l}^{6}\right| \leq CS_{n,l}\sup_{j=1,\dots,n}\gamma_{j,l}^{2}.$$

In conjunction with (4.29) the stronger relation was shown

$$\sup_{j=1,\dots,n}\gamma_{j,l}^{2} \leq Cn^{-1/(4\beta_l+1)}\left((2\log\log n)^{1/2}+D\right)^{(4\beta_l+2)/(4\beta_l+1)}. \tag{4.45}$$

Furthermore, by (4.38) the term $S_{n,l}$ grows like a power of $\log^{(2)} n$. In view of $n^{-1/(4\beta_l+1)} \leq n^{-\varepsilon}\delta_n^2$ the last two facts imply

$$|NR_{n,l}| = o\left(\delta_n^2\right)$$

uniformly in l. To consider the random term of $S_{n,l}^{-1}R_{n,l}$, apply Chebyshev's inequality to its first component:

$$\begin{aligned}P\left(S_{n,l}^{-1}\left|\sum_{j=1}^{n}\frac{\gamma_{j,l}^{4}}{\sqrt{2}}z_j\right| \geq \delta_n\right) &\leq \frac{\sum_{j=1}^{n}\gamma_{j,l}^{8}}{2S_{n,l}^{2}\delta_n^{2}}\operatorname{Var}(z_1)\\ &\leq C\delta_n^{-2}\sup_{j=1,\dots,n}\gamma_{j,l}^{4}.\end{aligned}$$

In view of (4.45) and (4.40) the latter term is $o\left(\delta_n^2\right)$. The other random term $S_{n,l}^{-1}R_{n,l}$ (namely $\sum_{j=1}^{n} z_j O\left(\gamma_{j,l}^{6}\right)$) is treated in a similar fashion, which establishes (4.44). For (4.43) we use the Berry-Esseen type result of [20]: suppose v_j, $j = 1,\dots,n$ are independent random vectors in $\mathbb{R}^k$ with zero expectation and normalized in such a way that $\sum_{j=1}^{n} v_j$ has unit covariance matrix. If $\mathfrak{C}$ is the set of convex subsets of $\mathbb{R}^k$ and Φ is the standard normal measure then

$$\sup_{A\in\mathfrak{C}}\left|P\left(\sum_{j=1}^{n}v_j \in A\right) - \Phi(A)\right| \leq C\sum_{j=1}^{n}E\left\|v_j\right\|^{3} \tag{4.46}$$

where $\|\cdot\|$ is the Euclidean norm, and the constant C depends only on k. We use this result for $k = 1$, $A = (-\infty, x]$ and set $v_j = S_{n,l}^{-1}\gamma_{j,l}^{2}z_j/\sqrt{2}$. Then

$$\begin{aligned}\sum_{j=1}^{n}E\left\|v_j\right\|^{3} = \sum_{j=1}^{n}E\left|S_{n,l}^{-1}\gamma_{j,l}^{2}z_j/\sqrt{2}\right|^{3} &\leq CS_{n,l}^{-3/2}\sum_{j=1}^{n}\gamma_{j,l}^{6}\\ &\leq CS_{n,l}^{-1/2}\sup_{j=1,\dots,n}\gamma_{j,l}^{2} = o\left(\delta_n^2\right) \qquad (4.47)\end{aligned}$$

which establishes (4.43). ∎

Proof of (4.37). We will apply (4.46) for $k = 2$ and random vectors

$$v_j = \Sigma^{-1/2}\tilde{v}_j,\ \tilde{v}_j := \left(S_{n,l}^{-1}\gamma_{j,l}^2 z_j/\sqrt{2}, S_{n,l}^{-1}\gamma_{j,k}^2 z_j/\sqrt{2}\right)^\top,$$

where Σ is the covariance matrix

$$\Sigma = \operatorname{Cov}\left(\sum_{j=1}^n \tilde{v}_j\right) = \begin{pmatrix} 1 & \rho_{n,kl} \\ \rho_{n,kl} & 1 \end{pmatrix}.$$

Let ξ be a 2-vector of independent standard normals; then (4.46) implies that for any convex set A

$$\left| P\left(\sum_{j=1}^n v_j \in A\right) - P\left(\xi \in A\right)\right| \le C \sum_{j=1}^n E\left\|\Sigma^{-1/2}\tilde{v}_j\right\|^3. \tag{4.48}$$

Here $\Sigma = I_2 + o(1)$ uniformly over k, l (cp. (4.35) and (4.38)), so that it suffices to estimate

$$\sum_{j=1}^n E\left\|\tilde{v}_j\right\|^3 \le C \sum_{j=1}^n \left(\left|S_{n,l}^{-1}\gamma_{j,l}^2 z_j\right|^3 + \left|S_{n,k}^{-1}\gamma_{j,k}^2 z_j\right|^3\right).$$

By (4.47) this quantity is $o\left(\delta_n^2\right)$. Set $A = \Sigma^{-1/2}B$ in (4.48) where B is convex; then uniformly over k, l and over convex sets B

$$\left| P\left(\sum_{j=1}^n \tilde{v}_j \in B\right) - P\left(\Sigma^{1/2}\xi \in B\right)\right| = o\left(\delta_n^2\right). \tag{4.49}$$

Now set $B = B_{x,y} := (-\infty, x) \times (-\infty, y)$ and for the error terms $R_{n,l}$ defined by (4.42) consider random vectors

$$\check{v}_j := \tilde{v}_j + \left(S_{n,l}^{-1}R_{n,l}, S_{n,k}^{-1}R_{n,k}\right)' = \left(\tilde{\lambda}_{n,l}, \tilde{\lambda}_{n,k}\right)'.$$

By a standard reasoning, the results (4.44) and (4.49) for rectangles $B_{x,y}$ together imply

$$\left| P\left(\sum_{j=1}^n \check{v}_j \in B_{x,y}\right) - P\left(\Sigma^{1/2}\xi \in B_{x,y}\right)\right| = o\left(\delta_n\right).$$

Comparing notation with (4.37), we find

$$P\left(\Sigma^{1/2}\xi \in B_{x,y}\right) = \Phi\left(x, y; \rho_{n,kl}\right),$$

$$P\left(\sum_{j=1}^n \check{v}_j \in B_{x,y}\right) = P\left(\tilde{\lambda}_{n,l} < x, \tilde{\lambda}_{n,k} < y | \pi_0^n\right).$$

The claim (4.37) follows since $L_n^2\delta_n = o(1)$. □

Appendix

Lemma A.1. *(i) For every $\beta > 0$, $M > 0$, $\rho > 0$ such that $M/\rho > 1$ there are unique positive solutions λ, μ of the equations*

$$\sum_{j=1}^{\infty} j^{2\beta}\left(\lambda - \mu j^{2\beta}\right)_+ = M, \; \sum_{j=1}^{\infty}\left(\lambda - \mu j^{2\beta}\right)_+ = \rho. \tag{A.1}$$

(ii) Assume $\rho \to 0$ while (β, M) is fixed. Then

$$N := (\lambda/\mu)^{1/2\beta} = \left(\frac{(4\beta+1)M}{\rho}\right)^{1/2\beta}(1 + o(1)), \tag{A.2}$$

$$\lambda = \rho^{1+1/2\beta} M^{-1} \frac{2\beta+1}{(4\beta+1)^{1/2\beta}\, 2\beta}(1 + o(1)) \tag{A.3}$$

$$T_{(1)} := \sum_{j=1}^{\infty}\left(\lambda - \mu j^{2\beta}\right)_+^2 = \rho^{2+1/2\beta} M^{-1/2\beta} A_0(\beta)(1 + o(1)) \tag{A.4}$$

where $A_0(\beta)$ is given by (1.7). Furthermore, there exist $C > 0$ such that for $\rho \le C^{-1}$

$$T_{(2)} := \sum_{j=1}^{\infty} j^{4\beta}\left(\lambda - \mu j^{2\beta}\right)_+^4 \le C\rho^{2+3/2\beta}, \tag{A.5}$$

$$T_{(3)} := \sum_{j=1}^{\infty}\left(\lambda - \mu j^{2\beta}\right)_+^4 \le C\rho^{4+3/2\beta}. \tag{A.6}$$

(iii) Suppose that J_1, J_2 are finite closed subintervals of $(0,\infty)$. Then there exists $\gamma > 0$ such that the $o(1)$ terms occurring in (A.2)–(A.4) are $O(\rho^\gamma)$ uniformly over $(\beta, M) \in J_1 \times J_2$. Furthermore, the constant C can be chosen such that (A.5), (A.6) hold uniformly over $(\beta, M) \in J_1 \times J_2$ and $\rho \le C^{-1}$.

Proof. (i) Note that if $\lambda, \mu > 0$ then $\left(\lambda - \mu j^{2\beta}\right)_+ > 0$ only for finitely many j. Set $t = \lambda/\mu$, then it suffices to prove that there are unique positive solutions λ, t of

$$\lambda\sum_{j=1}^{\infty} j^{2\beta}\left(1 - t^{-1} j^{2\beta}\right)_+ = M, \; \lambda\sum_{j=1}^{\infty}\left(1 - t^{-1} j^{2\beta}\right)_+ = \rho. \tag{A.7}$$

Define for $t > 0$

$$g_1(t) = \sum_{j=1}^{\infty} j^{2\beta}\left(1 - t^{-1} j^{2\beta}\right)_+, \; g_2(t) = \sum_{j=1}^{\infty}\left(1 - t^{-1} j^{2\beta}\right)_+.$$

Both functions g_i are continuous on $\mathbb{R}_+$, fulfill $g_i(t) = 0$ for $0 < t \le 1$, are strictly monotone increasing for $t > 1$ and piecewise differentiable on intervals

$\left[j^{2\beta},(j+1)^{2\beta}\right]$ for $j \geq 1$. Solving for λ in (A.7), one sees that it suffices to prove that for the function $g_3 = g_1/g_2$ there is a unique solution $t > 1$ of

$$g_3(t) = M/\rho.$$

Suppose that $t > 1$; for $k_t = \left[t^{2\beta}\right]$ we have

$$g_3(t) = \frac{\sum_{j=1}^{k_t} j^{2\beta}\left(1 - t^{-1}j^{2\beta}\right)}{\sum_{j=1}^{k_t}\left(1 - t^{-1}j^{2\beta}\right)}$$

Thus if $1 < t < 2^{2\beta}$ then $k_t = 1$ and $g_3(t) = 1$. To show that $g_3(t)$ is strictly monotone increasing for $t > 2^{2\beta}$, it suffices to show that on each interval $\left(j^{2\beta},(j+1)^{2\beta}\right)$, $j \geq 2$ we have $g_3'(t) > 0$. Now $g_2^2(t) > 0$ and

$$g_2^2(t)\,g_3'(t) = g_1'(t)\,g_2(t) - g_2'(t)\,g_1(t)$$

$$= \left(t^{-2}\sum_{j=1}^{k_t} j^{4\beta}\right)\left(\sum_{j=1}^{k_t}\left(1 - t^{-1}j^{2\beta}\right)\right) - \left(t^{-2}\sum_{j=1}^{k_t} j^{2\beta}\right)\left(\sum_{j=1}^{k_t} j^{2\beta}\left(1 - t^{-1}j^{2\beta}\right)\right).$$

Let Y be a random variable with discrete uniform distribution on $\{1,\ldots,k_t\}$ and write $k_t^{-1}\sum_{j=1}^{k_t} j^{\alpha} = EY^{\alpha}$ for $\alpha > 0$. Then the above expression can be written

$$\begin{aligned} &t^{-2}k_t^2\left\{EY^{4\beta}\left(1 - t^{-1}EY^{2\beta}\right) - EY^{2\beta}\left(EY^{2\beta} - t^{-1}EY^{4\beta}\right)\right\} \\ &= t^{-2}k_t^2\left\{EY^{4\beta} - \left(EY^{2\beta}\right)^2\right\} = t^{-2}k_t^2\operatorname{Var}\left(Y^{2\beta}\right). \end{aligned}$$

For $t > 2^{2\beta}$ we have $k_t > 1$, and $\operatorname{Var}\left(Y^{2\beta}\right) > 0$. This shows that $g_3(t)$ is strictly monotone increasing for $t > 2^{2\beta}$. It is easy to see that $g_3(t) \nearrow \infty$ as $t \to \infty$, so that $g_3(t) = M/\rho$ has a unique solution.

For (ii), set $N := (\lambda/\mu)^{1/2\beta}$ and rewrite (A.7) as

$$\lambda \sum_{1\leq j\leq N} j^{2\beta}\left(1 - (j/N)^{2\beta}\right)_+ = M,\ \lambda \sum_{1\leq j\leq N}\left(1 - (j/N)^{2\beta}\right)_+ = \rho.$$

If $N = O(1)$ then $\lambda \to 0$ by the second equation, which contradicts the first. Hence $N \to \infty$; then replacing the sums by integrals one obtains

$$\lambda N^{2\beta+1}\kappa_1 \sim M \text{ where } \kappa_1 = \int_0^1 x^{2\beta}\left(1 - x^{2\beta}\right)dx,$$

$$\lambda N\kappa_2 \sim \rho \text{ where } \kappa_2 = \int_0^1 \left(1 - x^{2\beta}\right)dx.$$

Solving this for λ and N gives

$$N \sim M^{1/2\beta}\rho^{-1/2\beta}\left(\kappa_2/\kappa_1\right)^{1/2\beta}, \ \lambda \sim \rho^{1+1/2\beta}M^{-1}\kappa_2^{-1}\left(\kappa_1/\kappa_2\right)^{1/2\beta} \tag{A.8}$$

Computing the constants, we find

$$\kappa_1 = \frac{2\beta}{(2\beta+1)(4\beta+1)},$$
$$\kappa_2 = \frac{2\beta}{2\beta+1},$$

which proves (A.2) and (A.3).

Furthermore we find

$$\begin{aligned} T_{(1)} &= \lambda^2 \sum_{j=1}^{\infty}\left(1-(j/N)^{2\beta}\right)_+^2 \\ &\sim \lambda^2 N\kappa_3 \text{ where } \kappa_3 = \int_0^1 \left(1-x^{2\beta}\right)^2 dx \\ &\sim M^{-1/2\beta}\rho_n^{2+1/2\beta}\left(\frac{\kappa_1}{\kappa_2}\right)^{1/2\beta}\frac{\kappa_3}{\kappa_2^2} \text{ (using (A.8))} \end{aligned} \tag{A.9}$$

We find

$$\kappa_3 = \kappa_2 - \kappa_1 = \frac{8\beta^2}{(2\beta+1)(4\beta+1)},$$

so that

$$\left(\frac{\kappa_1}{\kappa_2}\right)^{1/2\beta}\frac{\kappa_3}{\kappa_2^2} = \frac{2(2\beta+1)}{(4\beta+1)^{1+1/2\beta}} = A_0(\beta)$$

which proves (A.4). To treat the expression $T_{(2)}$, note that

$$\begin{aligned} \sum_{j=1}^{\infty} j^{4\beta}\left(\lambda-\mu j^{2\beta}\right)_+^4 &= \lambda^4 N^{4\beta}\sum_{j=1}^{\infty}(j/N)^{4\beta}\left(1-(j/N)^{2\beta}\right)_+^4 \\ &\sim \lambda^4 N^{4\beta+1}\kappa_4 \text{ where } \kappa_4 = \int_0^1 x^{4\beta}\left(1-x^{2\beta}\right)^4 dx \\ &\sim \rho^{2+3/2\beta}M^{-2+1/2\beta}\kappa_1^{3/2\beta-2}\kappa_2^{-3/2\beta-2}\kappa_4 \text{ (using (A.8))} \end{aligned} \tag{A.10}$$

which proves (A.5). To treat the expression $T_{(3)}$, note that

$$\begin{aligned} \sum_{j=1}^{\infty}\left(\lambda-\mu j^{2\beta}\right)_+^4 &= \lambda^4\sum_{j=1}^{\infty}\left(1-(j/N)^{2\beta}\right)_+^4 \\ &\sim \lambda^4 N\kappa_5 \text{ where } \kappa_5 = \int_0^1 \left(1-x^{2\beta}\right)^4 dx \\ &\sim \rho^{4+3/2\beta}M^{-4+1/2\beta}\kappa_2^{-4-3/2\beta}\kappa_1^{3/2\beta}\kappa_5 \text{ (using (A.8))} \end{aligned} \tag{A.11}$$

which proves (A.6).

For (iii), note that the $o(1)$ terms arise from approximating the integrals κ_1, κ_2 by Riemann sums with N terms, and since the pertaining functions satisfy a Lipschitz condition uniformly over $\beta \in J_1$, the error there is of order N^{-1}. Since both β and M are bounded and bounded away from zero, by (A.2) N^{-1} is of order $\rho^{1/2\beta}$ uniformly, so for $\gamma = \min\{1/2\beta : \beta \in J_1\}$ the claim holds true. The second claim follows from this and the nature of the constants in (A.10), (A.11). □

Lemma A.2. *For given $M > \rho > 0$, let λ, μ be the solutions of (A.1) and define*

$$\tilde{g}_j = \frac{n\left(\lambda - \mu j^{2\beta}\right)_+}{1 + n\left(\lambda - \mu j^{2\beta}\right)_+}.$$

Assume that $\rho \to 0$ along with $n \to \infty$ such that $\rho \ll n^{-2\beta/(2\beta+1)}$.

(i) The vector $\tilde{g} = (\tilde{g}_j)_{j=1}^{\infty}$ has at most $[N]$ nonzero components, N being given by (A.2), fulfilling

$$\sup_j |\tilde{g}_j| = o(1),$$

and the squared norm $\|\tilde{g}\|^2 = \sum_{j=1}^{\infty} \tilde{g}_j^2$ fulfills

$$\|\tilde{g}\|^2 = n^2 \rho^{2+1/2\beta} M^{-1/2\beta} A_0(\beta)(1 + o(1)). \tag{A.12}$$

(ii) For fixed M and β

$$\sup_{1 \le j \le N} \frac{|\tilde{g}_j|}{\|\tilde{g}\|} = O\left(N^{-1/2}\right). \tag{A.13}$$

(iii) Let $c_i > 0$, $i = 1, 2$ be constants and define $\rho_i = c_i \rho$, $i = 1, 2$. For $M_2 > M_1$, let λ_i, μ_i be the solutions of (A.1) corresponding to a pair (M_i, ρ_i), let $\tilde{g}_{(i)}$ be the vector pertaining to i as above and define the unit vectors $g_{(i)} := \tilde{g}_{(i)} / \left\|\tilde{g}_{(i)}\right\|$, $i = 1, 2$. Then for the scalar product $\left\langle g_{(1)}, g_{(2)} \right\rangle = \sum_{j=1}^{\infty} g_{(1),j} g_{(2),j}$ one has

$$\left\langle g_{(1)}, g_{(2)} \right\rangle = r\,(1 + o(1)),$$
$$r = \left(\frac{\tilde{M}_1}{\tilde{M}_2}\right)^{1/(4\beta)} \cdot \frac{4\beta + 1 - \tilde{M}_1/\tilde{M}_2}{4\beta} \text{ for } \tilde{M}_i = M_i / c_i.$$

Proof. We first claim that

$$n\left(\lambda - \mu j^{2\beta}\right)_+ = o(1) \tag{A.14}$$

uniformly over $j = 1, 2, \ldots$. Indeed, with $N = (\lambda/\mu)^{1/2\beta}$

$$n\left(\lambda - \mu j^{2\beta}\right)_+ = n\lambda\left(1 - (j/N)^{2\beta}\right)_+ \le n\lambda$$

and using (A.3), for a constant C_β,

$$n\lambda \sim n\rho^{1+1/2\beta} M^{-1} C_\beta = o(1)$$

since $\rho \ll n^{-2\beta/(2\beta+1)}$, so (A.14) is shown. This implies, with $T_{(1)}$ from (A.4)

$$\|\tilde{g}\|^2 \sim n^2 T_{(1)} = n^2 \sum_{j=1}^{\infty} \left(\lambda - \mu j^{2\beta}\right)_+^2 \tag{A.15}$$

and relation (A.4) establishes (i).

For (ii), note that

$$\begin{aligned} \frac{|\tilde{g}_j|}{\|\tilde{g}\|} \leq \frac{n\lambda}{\|\tilde{g}\|} &\sim \frac{n\rho^{1+1/2\beta} M^{-1} C_\beta}{n T_{(1)}^{1/2}} \\ &\sim \frac{\rho^{1+1/2\beta} M^{-1} C_\beta}{\rho^{1+1/4\beta} M^{-1/4\beta} A_0^{1/2}(\beta)} = O\left(\rho^{1/4\beta}\right) \end{aligned}$$

and (A.2) establishes the claim.

For (iii), denote λ_i, N_i, $i = 1, 2$ the expressions λ,N from Lemma A.1 pertaining to (M_i, ρ_i) and note that

$$\left\langle g_{(1)}, g_{(2)} \right\rangle = \frac{\left\langle \tilde{g}_{(1)}, \tilde{g}_{(2)} \right\rangle}{\left\|\tilde{g}_{(1)}\right\| \left\|\tilde{g}_{(2)}\right\|},$$

and, in view of (A.14),

$$\begin{aligned} \left\langle \tilde{g}_{(1)}, \tilde{g}_{(2)} \right\rangle &\sim n^2 \lambda_1 \lambda_2 \sum_{j=1}^{\infty} \left(1 - (j/N_1)^{2\beta}\right)_+ \left(1 - (j/N_2)^{2\beta}\right)_+ \\ &= n^2 \lambda_1 \lambda_2 N_1^{1/2} N_2^{1/2} \cdot J_n \end{aligned}$$

where

$$J_n := N_1^{-1/2} N_2^{-1/2} \sum_{j=1}^{\infty} \left(1 - (j/N_1)^{2\beta}\right)_+ \left(1 - (j/N_2)^{2\beta}\right)_+ .$$

In (A.15) and (A.9) it has been shown that

$$\left\|\tilde{g}_{(i)}\right\| \sim n\lambda_i N_i^{1/2} \kappa_3^{1/2},\ i = 1, 2$$

so that now

$$\left\langle g_{(1)}, g_{(2)} \right\rangle \sim \kappa_3^{-1} J_n.$$

By (A.2) we have $N_2 \sim \left(\tilde{M}_2 / \tilde{M}_1\right)^{1/2\beta} N_1$ and thus

$$
\begin{aligned}
J_n &\sim \left(\tilde{M}_1/\tilde{M}_2\right)^{1/4\beta} N_1^{-1} \sum_{j=1}^{\infty} \left(1-(j/N_1)^{2\beta}\right)_+ \left(1-(N_1/N_2)^{2\beta}(j/N_1)^{2\beta}\right)_+ \\
&\sim \left(\tilde{M}_1/\tilde{M}_2\right)^{1/4\beta} \int_0^1 \left(1-x^{2\beta}\right)\left(1-x^{2\beta}\left(\tilde{M}_1/\tilde{M}_2\right)\right) dx \\
&= (M_1/M_2)^{1/4\beta}\left(\kappa_2-\kappa_1\left(M_1/M_2\right)\right).
\end{aligned}
$$

Using the values of the constants κ_i, $1,2,3$ found in the proof of Lemma A.1 we find

$$
\begin{aligned}
\left\langle g_{(1)}, g_{(2)}\right\rangle &\sim \left(\tilde{M}_1/\tilde{M}_2\right)^{1/4\beta}\left(\kappa_2\kappa_3^{-1}-\kappa_1\kappa_3^{-1}\left(\tilde{M}_1/\tilde{M}_2\right)\right) \\
&= \left(\tilde{M}_1/\tilde{M}_2\right)^{1/4\beta}\left(\frac{4\beta+1}{4\beta}-\frac{1}{4\beta}\left(\tilde{M}_1/\tilde{M}_2\right)\right) = r.
\end{aligned}
$$

□

Lemma A.3. *For some finite dimension $p>0$, define sets*

$$
\begin{aligned}
\mathcal{D}_0 &= \left\{d\in\mathbb{R}_+^p : \|d\|^2 = 1\right\}, \\
B_\rho^0 &= \{f\in\mathbb{R}^p : \|f\|^2=\rho\}.
\end{aligned}
\tag{A.16}
$$

For $f\in\mathbb{R}^p$ set $f^2 := \left(f_j^2\right)_{j=1}^p$, and for $d, f^2\in\mathbb{R}_+^p$ define the functional

$$
L(d,f) = \frac{n}{\sqrt{2}}\left\langle d, f^2\right\rangle = \frac{n}{\sqrt{2}}\sum_{j=1}^p d_j f_j^2. \tag{A.17}
$$

Consider the pair (d_0, f_0) given by

$$
d_0 = \frac{f_0^2}{\|f_0^2\|},\quad f_{0,j}^2 = \left(\lambda-\mu j^{2\beta}\right)_+,\ j=1,\dots,p \tag{A.18}
$$

where λ,μ are the solutions of (A.1) for the given β, M and ρ small enough.

The pair $d_0\in\mathcal{D}_0$, $f_0\in\Sigma(\beta,M)\cap B_\rho^0$ is a saddlepoint of the functional $L(d,f)$ such that

$$
L(d,f_0)\le L(d_0,f_0)\le L(d_0,f) \tag{A.19}
$$

for all $d\in\mathcal{D}_0$ and all $f\in\Sigma(\beta,M)\cap B_\rho^0$. The value of L at the saddlepoint is

$$
S_n := L(d_0,f_0) = \frac{n}{\sqrt{2}}\left\|f_0^2\right\|. \tag{A.20}
$$

Proof. It suffices to show that the pair (d_0, f_0) fulfills (A.18). Consider maximizing $L(d,f)$ in d for given f. Under the sole restriction $\|d\|=1$, by Cauchy-Schwartz the solution is found as

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

For $f = f_0$ we obtain the first inequality of (A.18). The second inequality is equivalent to

$$\lambda\rho - L(d_0, f) \leq \lambda\rho - L(d_0, f_0),\ f \in \Sigma(\beta, M) \cap B_\rho^0.$$

For this, in view of $\sum_{j=1}^p f_j^2 = \rho$, it is sufficient to show

$$\sum_{j=1}^{p} \left(\lambda - f_{0,j}^2\right) f_j^2 \leq \sum_{j=1}^{p} \left(\lambda - f_{0,j}^2\right) f_{0,j}^2,\ f \in \Sigma(\beta, M). \tag{A.21}$$

Note that if $f_{0,j}^2 > 0$ then $\lambda > \mu j^{2\beta}$, consequently $\lambda - f_{0,j}^2 = \mu j^{2\beta}$. Hence for $f \in \Sigma(\beta, M)$

$$\sum_{j=1}^{p} \left(\lambda - f_{0,j}^2\right) f_{0,j}^2 = \sum_{j=1}^{p} \mu j^{2\beta} f_{0,j}^2 = \mu M \geq \sum_{j=1}^{p} \mu j^{2\beta} f_{,j}^2.$$

Furthermore $\lambda - f_{0,j}^2 = \lambda - \left(\lambda - \mu j^{2\beta}\right)_+ = \min\left(\lambda, \mu j^{2\beta}\right)$ for $j = 1, \ldots, p$, hence

$$\sum_{j=1}^{p} \mu j^{2\beta} f_{,j}^2 \geq \sum_{j=1}^{p} \left(\lambda - f_{0,j}^2\right) f_{,j}^2$$

which establishes (A.21). □

The following is a slight extension of the CLT in total variation, given in Theorem 2.31 of [37] for i.i.d. random variables, to the case of linear combinations $\sum_{j=1}^n c_{jn} Y_j$ of i.i.d. random variables $Y_1, Y_2, \ldots$.

Lemma A.4. *Let $Y_1, Y_2, \ldots$ be i.i.d. random variables with $EY = 0$, $Var(Y) = 1$ and characteristic function ϕ such that $\int |\phi(t)|^2\, dt < \infty$. Let $\{c_{jm}\}_{j=1}^m$ be a double array of coefficients satisfying $\sum_{j=1}^m c_{jm}^2 = 1$ for all m and $\max_{1\leq j\leq m} |c_{jm}| = O\left(m^{-1/2}\right)$. Then the law of $S_m = \sum_{j=1}^m c_{jm} Y_j$ converges to $N(0, 1)$ in total variation.*

The corresponding CLT in distribution (a consequence of the Lindeberg-Feller theorem), requiring only $\max_{1\leq j\leq m} |c_{jm}| = o(1)$ and no conditions on ϕ, is known as the Hajek-Šidak theorem (Theorem 5.3 in [4]).

Proof. We can assume the c_{jm} are ordered: $c_{1m} \geq c_{2m} \geq \ldots \geq c_{mm}$. We claim there exists $\delta > 0$ such that

$$\min_{1\leq j\leq \delta m} c_{jm}^2 \geq \frac{1}{4} m^{-1}. \tag{A.22}$$

Since $\min_{1\leq j\leq \delta m} c_{jm}^2 \geq \max_{\delta m < j \leq m} c_{jm}^2$, it suffices to show that

$$\max_{\delta m < j \leq m} c_{jm}^2 \geq \frac{1}{4} m^{-1}.$$

Let C be such that $\max_{1\le j\le m} c_{jm}^2 \le Cm^{-1}$; then

$$m \max_{\delta m<j\le m} c_{jm}^2 \ge m\,(m-[\delta m])^{-1} \sum_{\delta m<j\le m} c_{jm}^2 = (1-\delta)^{-1}\left(1-\sum_{1\le j\le \delta m} c_{jm}^2\right)$$
$$\ge (1-\delta)^{-1}\left(1-[\delta m]\,Cm^{-1}\right).$$

Then $[\delta m]\,Cm^{-1} \le \delta C \le 1/4$ for sufficiently small δ, which establishes (A.22).

Now the characteristic function of S_m is $\prod_{j=1}^{m}\phi\,(c_{jm}t)$ which is integrable for sufficiently large m, so S_m has a density p_m which by the inversion formula for characteristic functions is given by

$$p_m\,(x) = \frac{1}{2\pi}\int \exp\,(\boldsymbol{i}tx)\prod_{j=1}^{m}\phi\,(c_{jm}t)\,dt. \tag{A.23}$$

By the Hajek-Šidak CLT, $S_m \rightsquigarrow N\,(0,1)$ in distribution which implies that $\prod_{j=1}^{m}\phi\,(c_{jm}t) \to \exp\left(-t^2/2\right)$ pointwise in t. It will be shown that the integral above converges to

$$\frac{1}{2\pi}\int \exp\,(\boldsymbol{i}tx)\exp\left(-t^2/2\right)dt = \frac{1}{\sqrt{2\pi}}\exp\left(-x^2/2\right).$$

Then an application of Scheffé's theorem (2.30 in [37]) concludes the proof.

The integral (A.23) can be split in two parts. First for every $\varepsilon > 0$

$$\int_{|t|>\varepsilon m^{1/2}}\left|\exp\,(\boldsymbol{i}tx)\prod_{j=1}^{m}\phi\,(c_{jm}t)\right|dt \le \int_{|t|>\varepsilon m^{1/2}}\left|\prod_{1\le j\le\delta m}\phi\,(c_{jm}t)\right|dt$$
$$\le \sup_{|t|>\varepsilon m^{1/2}}\prod_{3\le j\le\delta m}|\phi\,(c_{jm}t)|\int_{|t|>\varepsilon m^{1/2}}|\phi\,(c_{1m}t)\,\phi\,(c_{2m}t)|\,dt$$
$$\le \left(\sup_{|u|>\varepsilon/2}\prod_{3\le j\le\delta m}|\phi\,(u)|\right)\prod_{j=1,2}\left(\int_{|t|>\varepsilon m^{1/2}}|\phi\,(c_{jm}t)|^2\,dt\right)^{1/2}$$
$$\le \sup_{|u|>\varepsilon/2}|\phi\,(u)|^{[\delta m]-2}\,c_{1m}^{-1/2}c_{2m}^{-1/2}\int|\phi\,(u)|^2\,du. \tag{A.24}$$

Here of $c_{1m}^{-1/2}c_{2m}^{-1/2} \le 2m^{1/2}$ by (A.22), and $\sup_{|u|>\varepsilon/2}|\phi\,(u)| < 1$ by the Riemann-Lebesgue lemma and because the distribution of Y is non-lattice, hence (A.24) converges to zero geometrically fast.

Second, a Taylor expansion yields that $\phi\,(t) = 1 - t^2/2 + o\left(t^2\right)$ as $t \to 0$, so that there exists $\varepsilon > 0$ such that $|\phi\,(t)| \le 1 - t^2/4$ for every $|t| < C\varepsilon$. It follows that

$$\left|\exp\left(\boldsymbol{i}tx\right)\prod_{j=1}^{m}\phi\left(c_{jm}t\right)\right|\mathbf{1}\left\{|t|<\varepsilon m^{1/2}\right\}\leq\prod_{j=1}^{m}\left(1-\frac{c_{jm}^{2}t^{2}}{4}\right)$$
$$\leq\prod_{j=1}^{m}\exp\left(-c_{jm}^{2}t^{2}/4\right)=\exp\left(-t^{2}/4\right).$$

The proof can be concluded by applying the dominated convergence theorem to the remaining part of the integral (A.23). □

In the situation of the previous lemma, for linear combinations $S_m = \sum_{j=1}^{m} c_{jm} Y_j$, we now aim at an auxiliary result on tail probabilities $P(S_m \geq a_m)$ in the moderate deviation range where $a_m = o\left(m^{1/2}\right)$. For our purposes, it suffices to assume the restrictive growth condition $a_m = o\left(m^{1/6}\right)$ and establish an asymptotic upper bound; the stronger result $P(S_m \geq a_m) \sim 1 - \Phi(a_m)$ could be deduced from the general i.n.i.d. case treated in Theorem 10, chap. 8 of [32].

Lemma A.5. *Let $Y_1, Y_2, \ldots$ be i.i.d. random variables with $EY = 0$, $Var(Y) = 1$ and such that Cramér's condition is fulfilled: there exists $H > 0$ such that the moment generating function $E\exp(tX) < \infty$ exists for all $t \in (-H, H)$. Let $\{c_{jm}\}_{j=1}^{m}$ be a double array of coefficients satisfying $\sum_{j=1}^{m} c_{jm}^2 = 1$ for all m and $\max_{1\leq j\leq n}|c_{jm}| = O\left(m^{-1/2}\right)$. Suppose that $a_m \to \infty$ and $a_m = o\left(m^{1/6}\right)$; then for $S_m = \sum_{j=1}^{m} c_{jm} Y_j$ one has*

$$P\left(S_m \geq a_m\right) \leq \exp\left(-a_m^2/2\right)\left(1+o\left(1\right)\right) \text{ as } m \to \infty.$$

Proof. Let

$$L(t) = \log E\exp\left(tX\right) = \sum_{l=1}^{\infty}\kappa_l\frac{t^l}{l!}$$

be the cumulant generating function of Y and κ_l be its cumulants. We have $\kappa_1 = 0$ and $\kappa_2 = 1$. Also, it is well known (cf. [32], chap. 8, (2.10)) that under Cramér's condition there exists a constant $C > 0$ such that

$$|\kappa_l| \leq C\frac{l!}{H^l},\; l = 1, 2, \ldots \tag{A.25}$$

The cumulant generating function of S_m is

$$\tilde{L}_m\left(t\right) := \log E\exp\left(tS_m\right) = \sum_{j=1}^{n}L(c_{jm}t)$$
$$= \sum_{l=1}^{\infty}\kappa_l\frac{t^l}{l!}\left(\sum_{j=1}^{n}c_{jm}^{l}\right).$$

Denoting $\bar{c}_{l,m} = m^{-1}\sum_{j=1}^{m}\left(n^{1/2}c_{jm}\right)^l$, we may write

$$\tilde{L}_m(t) = m\sum_{l=1}^{\infty}\kappa_l\frac{\left(tm^{-1/2}\right)^l}{l!}\bar{c}_{l,m}$$
$$= \frac{t^2}{2} + m\sum_{l=3}^{\infty}\kappa_l\frac{\left(tm^{-1/2}\right)^l}{l!}\bar{c}_{l,m}.$$

Now observe that by Markov's inequality for any $t > 0$

$$P(S_m \geq a_m) = P(\exp(tS_m - ta_m) \geq 1) \leq \exp(-ta_m)E\exp(tS_m)$$
$$= \exp\left(\tilde{L}_m(t) - ta_m\right).$$

The choice $t = a_m$ gives

$$P(S_m \geq a_m) \leq \exp\left(-a_m^2/2\right)\exp\left(m\sum_{l=3}^{\infty}\kappa_l\frac{\left(a_m m^{-1/2}\right)^l}{l!}\bar{c}_{l,m}\right)$$
$$= \exp\left(-a_m^2/2\right)\exp\left(a_m^2 R_m\right) \text{ where } R_m := \frac{m}{a_m^2}\sum_{l=3}^{\infty}\kappa_l\frac{\left(a_m m^{-1/2}\right)^l}{l!}\bar{c}_{l,m}. \tag{A.26}$$

Now R_n can be written

$$R_m = \sum_{l=3}^{\infty}\kappa_l\frac{\left(a_m m^{-1/2}\right)^{l-2}}{l!}\bar{c}_{l,m}.$$

By assumption there is a constant $K > 0$ such that $\left|m^{1/2}c_{jm}\right| \leq K$, hence $|\bar{c}_{l,m}| \leq K^l$, and in view of (A.25) one has for $u_n = a_n n^{-1/2}$

$$|R_m| \leq \sum_{l=3}^{\infty}|\kappa_l|\frac{u_m^{l-2}}{l!}K^l \leq \sum_{l=3}^{\infty}Cu_m^{l-2}\frac{K^l}{H^l} = C\frac{K^2}{H^2}\sum_{l=1}^{\infty}(u_m K/H)^l$$
$$= C\frac{K^2}{H^2}\left(\frac{1}{1-u_m K/H} - 1\right) = C\frac{K^2}{H^2}\left(\frac{u_m K/H}{1-u_m K/H}\right).$$

This implies, in view of $a_m = o\left(m^{1/6}\right)$,

$$a_m^2 R_m = O\left(a_m^2 u_m\right) = O\left(a_m^3 m^{-1/2}\right) = o(1)$$

so that (A.26) establishes the claim. □

Lemma A.6. *Let $f_j \geq 0$, $d_j > 0$, $j = 1, \ldots, N$ and let $\xi_j \sim N(0,1)$, independent. Then for every $z > 0$ the function*

$$g_0(c) = P\left(\sum_{j=1}^{N} d_j\left(cf_j + \xi_j\right)^2 \leq z\right)$$

is monotone decreasing in $c \geq 0$.

We note that the implied maximization at $c = 0$ is a special case of Anderson's Lemma ([30], Propos. 3.62).

Proof. For $N = 1$ we have, φ being the standard normal density,

$$g_0(c) = P\left(-zd_1^{-1/2} - cf_1 \leq \xi_1 \leq zd_1^{-1/2} - cf_1\right),$$
$$g_0'(c) = \varphi\left(zd_1^{-1/2} - cf_1\right) - \varphi\left(-zd_1^{-1/2} - cf_1\right)$$
$$= \varphi\left(\left|zd_1^{-1/2} - cf_1\right|\right) - \varphi\left(zd_1^{-1/2} + cf_1\right).$$

Since $zd_1^{-1/2} + cf_1 \geq \left|zd_1^{-1/2} - cf_1\right|$, we have $g_0'(c) \leq 0$. Now for $N \geq 2$ and $x \in \mathbb{R}^N$ define the function

$$g(x) = P\left(\sum_{j=1}^{N} d_j (x_j + \xi_j)^2 \leq z\right).$$

Then $g_0(c) = g(cf)$; define also for every $j \in \{1, \ldots, N\}$ and every $\xi \in \mathbb{R}^N$

$$\xi_{(-j)} = (x_i)_{i=1,\ldots,N,i\neq j},$$
$$h\left(\xi_{(-j)}, x_{(-j)}\right) = \sum_{i=1,\ldots,N,i\neq j} d_i (x_i + \xi_i)^2.$$

This allows to write for every $j \in \{1, \ldots, N\}$

$$g(x) = P\left(h\left(\xi_{(-j)}, x_{(-j)}\right) + d_j (x_j + \xi_j)^2 \leq z\right)$$
$$= E^{\xi_{(-j)}} \mathbf{1}_{\left\{h\left(\xi_{(-j)}, x_{(-j)}\right) \leq z\right\}} P\left(d_j (x_j + \xi_j)^2 \leq z - h\left(\xi_{(-j)}\right) | \xi_{(-j)}\right).$$

Then, setting $g_j(x) = \frac{\partial}{\partial x_j} g(x)$, we obtain

$$g_0'(c) = \sum_{j=1}^{N} f_j g_j(cf)$$
$$= \sum_{j=1}^{N} E^{\xi_{(-j)}} \mathbf{1}_{\left\{h\left(\xi_{(-j)}, cf_{(-j)}\right) \leq z\right\}} \frac{d}{dc} P\left(d_j (cf_j + \xi_j)^2 \leq z - h\left(\xi_{(-j)}\right) | \xi_{(-j)}\right). \tag{A.27}$$

Here, according to what has been shown in the case $N = 1$, for every $\xi_{(-j)}$ such that $z - h\left(\xi_{(-j)}\right) > 0$ we have

$$\frac{d}{dc} P\left(d_j (cf_j + \xi_j)^2 \leq z - h\left(\xi_{(-j)}\right) | \xi_{(-j)}\right) < 0.$$

This implies that (A.27) is ≤ 0. □

Acknowledgement

We are grateful to a referee who pointed out a major shortcut in the proof of Theorem 1.2.

References

[1] Butucea, C., Matias, C. and Pouet, C. (2009). Adaptive goodness-of-fit testing from indirect observations. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** 352–372. MR2521406

[2] Chen, L. H. Y., Goldstein, L. and Shao, Q.-M. (2011). *Normal Approximation by Stein's Method*. Springer, New York. MR2732624

[3] Collier, O. (2012). Minimax hypothesis testing for curve registration. *Electron. J. Stat.* **6** 1129–1154. MR2988441

[4] DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability. Springer Texts in Statistics*. Springer, New York. MR2664452

[5] Dicker, L. H. (2016). Ridge regression and asymptotic minimax estimation over spheres of growing dimension. *Bernoulli* **22** 1–37. MR3449775

[6] Dümbgen, L. and Spokoiny, V. G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** 124–152. MR1833961

[7] Efroimovich, S. Y. and Pinsker, M. S. (1984). A self-training algorithm for nonparametric filtering. *Avtomat. i Telemekh.* **11** 58–65. MR797991

[8] Ermakov, M. (2011). Nonparametric signal detection with small type I and type II error probabilities. *Stat. Inference Stoch. Process.* **14** 1–19. MR2780646

[9] Ermakov, M. S. (1990). Minimax detection of a signal in Gaussian white noise. *Teor. Veroyatnost. i Primenen.* **35** 704–715. MR1090496

[10] Ermakov, M. S. (1997). Asymptotic minimaxity of chi-squared tests. *Teor. Veroyatnost. i Primenen.* **42** 668–695. MR1618797

[11] Ermakov, M. S. (2003). Asymptotically efficient statistical inferences for moderate deviation probabilities. *Teor. Veroyatnost. i Primenen.* **48** 676–700. MR2143210

[12] Ermakov, M. S. (2008). Nonparametric hypothesis testing for small type I and type II error probabilities. *Problemy Peredachi Informatsii* **44** 54–74. MR2435240

[13] Fromont, M. and Laurent, B. (2006). Adaptive goodness-of-fit tests in a density model. *Ann. Statist.* **34** 680–720. MR2281881

[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** 176–213. MR2779402

[15] Gayraud, G. and Pouet, C. (2005). Adaptive minimax testing in the discrete regression scheme. *Probab. Theory Related Fields* **133** 531–558. MR2197113

[16] Golubev, G. K. (1987). Adaptive asymptotically minimax estimates for smooth signals. *Problemy Peredachi Informatsii* **23** 57–67. MR893970

[17] Golubev, G. K. (1990). Quasilinear estimates for signals in L_2. *Problemy Peredachi Informatsii* **26** 19–24. MR1051584

[18] Golubev, G. K. and Nussbaum, M. (1990). A risk bound in Sobolev class regression. *Ann. Statist.* **18** 758–778. MR1056335

[19] Golubev, Y., Lepski, O. and Levit, B. (2001). On adaptive estimation for the sup-norm losses. *Math. Methods Statist.* **10** 23–37. MR1841807

[20] Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724–739. MR1106283

[21] Horowitz, J. L. and Spokoiny, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69** 599–631. MR1828537

[22] Inglot, T., Kallenberg, W. C. M. and Ledwina, T. (1992). Strong moderate deviation theorems. *Ann. Probab.* **20** 987–1003. MR1159582

[23] Ingster, Y. and Stepanova, N. (2011). Estimation and detection of functions from anisotropic Sobolev classes. *Electron. J. Stat.* **5** 484–506. MR2813552

[24] Ingster, Y. I. (1982). Minimax nonparametric detection of signals in white Gaussian noise. *Problemy Peredachi Informatsii* **18** 61–73. MR689340

[25] Ingster, Y. I. and Suslina, I. A. (2003). *Nonparametric Goodness-of-Fit Testing under Gaussian Models*. *Lecture Notes in Statistics* **169**. Springer-Verlag, New York. MR1991446

[26] Ingster, Y. I. and Suslina, I. A. (2005). Nonparametric hypothesis testing for small type I errors. II. *Math. Methods Statist.* **14** 28–52. MR2158070

[27] Korostelëv, A. P. (1993). An asymptotically minimax regression estimator in the uniform norm up to a constant. *Teor. Veroyatnost. i Primenen.* **38** 875–882. MR1318004

[28] Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*, Third ed. Springer, New York. MR2135927

[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** 17–48. MR1759508

[30] Liese, F. and Miescke, K.-J. (2008). *Statistical Decision Theory*. Springer, New York. MR2421720

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

[32] Petrov, V. V. (1995). *Limit Theorems of Probability Theory*. The Clarendon Press, Oxford University Press, New York. MR1353441

[33] Pinsker, M. S. (1980). Optimal filtration of square-integrable signals in Gaussian noise. *Problemy Peredachi Informatsii* **16** 52–68. MR624591

[34] Rohde, A. (2008). Adaptive goodness-of-fit tests based on signed ranks. *Ann. Statist.* **36** 1346–1374. MR2418660

[35] Spokoiny, V. G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24** 2477–2498. MR1425962

[36] Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York. MR2724359

[37] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge. MR1652247