Electronic Journal of Statistics
Vol. 7 (2013) 146–190
ISSN: 1935-7524
DOI: 10.1214/13-EJS766

Minimax testing of a composite null hypothesis defined via a quadratic functional in the model of regression

Laëtitia Comminges and Arnak S. Dalalyan

LIGM/IMAGINE and ENSAE/CREST/GENES
e-mail: laetitia.comminges@enpc.fr; dalalyan@imagine.enpc.fr

Abstract: We consider the problem of testing a particular type of composite null hypothesis under a nonparametric multivariate regression model. For a given quadratic functional Q, the null hypothesis states that the regression function f satisfies the constraint $Q[f] = 0$, while the alternative corresponds to the functions for which $Q[f]$ is bounded away from zero. On the one hand, we provide minimax rates of testing and the exact separation constants, along with a sharp-optimal testing procedure, for diagonal and nonnegative quadratic functionals. We consider smoothness classes of ellipsoidal form and check that our conditions are fulfilled in the particular case of ellipsoids corresponding to anisotropic Sobolev classes. In this case, we present a closed form of the minimax rate and the separation constant. On the other hand, minimax rates for quadratic functionals which are neither positive nor negative makes appear two different regimes: "regular" and "irregular". In the "regular" case, the minimax rate is equal to $n^{-1/4}$ while in the "irregular" case, the rate depends on the smoothness class and is slower than in the "regular" case. We apply this to the problem of testing the equality of Sobolev norms of two functions observed in noisy environments.

AMS 2000 subject classifications: Primary 62G08, 62G10; secondary 62G20.
Keywords and phrases: Nonparametric hypotheses testing, sharp asymptotics, separation rates, minimax approach, high-dimensional regression.

Received August 2012.

1. Introduction

1.1. Problem statement

Consider the nonparametric regression model with multi-dimensional random design: We observe $(x_i, \mathbf{t}_i)_{i=1,\dots,n}$ obeying the relation

$$x_i = f(\mathbf{t}_i) + \xi_i, \quad i = 1, \dots, n, \tag{1}$$

where $\mathbf{t}_i \in \Delta \subset \mathbb{R}^d$ are random design points, $1 \le d < \infty$, $f : \Delta \to \mathbb{R}$ is the unknown regression function and ξ_is represent observation noise. Throughout this work, we assume that the vectors $\mathbf{t}_i = (t_i^1, \dots, t_i^d)$, for $i = 1, \dots, n$, are independent and identically distributed with uniform distribution on $\Delta = [0, 1]^d$, which

is equivalent to $t_i^k \overset{\text{iid}}{\sim} \mathcal{U}(0,1)$. Furthermore, conditionally on $\mathcal{T}_n = \{\mathbf{t}_1, \ldots, \mathbf{t}_n\}$, the variables $\xi_1, \ldots, \xi_n$ are assumed i.i.d. with zero mean and variance τ^2, for some known $\tau \in (0, \infty)$.

Let $L_2(\Delta)$ denote the Hilbert space of all squared integrable functions defined on Δ. Assume that we are given two disjoint subsets $\mathcal{F}_0$ and $\mathcal{F}_1$ of $L_2(\Delta)$. We are interested in analyzing the problem of testing hypotheses:

$$H_0 : f \in \mathcal{F}_0 \qquad \text{against} \qquad H_1 : f \in \mathcal{F}_1. \tag{2}$$

To be more precise, let us set $\mathbf{z}_i = (x_i, \mathbf{t}_i)$ and denote by P_f be the probability distribution of the data vector $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ given by (1). The expectation with respect to P_f is denoted by E_f. The goal is to design a testing procedure $\phi_n : (\mathbb{R} \times \Delta)^n \to \{0, 1\}$ for which we are able to establish theoretical guarantees in terms of the cumulative error rate (the sum of the probabilities of type I and type II errors):

$$\gamma_n(\mathcal{F}_0, \mathcal{F}_1, \phi_n) = \sup_{f \in \mathcal{F}_0} P_f(\phi_n = 1) + \sup_{f \in \mathcal{F}_1} P_f(\phi_n = 0). \tag{3}$$

To measure the statistical complexity of this testing problem, it is relevant to analyze the minimax error rate

$$\gamma_n(\mathcal{F}_0, \mathcal{F}_1) = \inf_{\phi_n} \gamma_n(\mathcal{F}_0, \mathcal{F}_1, \phi_n), \tag{4}$$

where $\inf_{\phi_n}$ denotes the infimum over all testing procedures.

The focus in this paper is on a particular type of null hypotheses H_0 that can be defined as the set of functions lying in the kernel of some quadratic functional $Q : L_2(\Delta) \to \mathbb{R}$, *i.e.*, $\mathcal{F}_0 \subset \big\{f \in L_2(\Delta) : Q[f] = 0\big\}$. As described later in this section, this kind of null hypotheses naturally arises in several problems including variable selection, testing partial linearity of a regression function or the equality of norms of two signals. Then, it is appealing to define the alternative as the set of functions satisfying $|Q[f]| > \rho^2$ for some $\rho > 0$. However, without further assumptions on the nature of functions f, it is impossible to design consistent testing procedures for discriminating between $\mathcal{F}_0$ and $\mathcal{F}_1$. One approach to making the problem meaningful is to assume that the function f belongs to a smoothness class. Typical examples of smoothness classes are Sobolev and Hölder classes, Besov bodies or balls in reproducing kernel Hilbert spaces.

In the present work, we assume that the function f belongs to a smoothness class Σ that can be seen as an ellipsoid in the infinite-dimensional space $L_2(\Delta)$. Thus, the null and the alternative are defined by

$$\mathcal{F}_0 = \Big\{f \in \Sigma : Q[f] = 0\Big\}, \qquad \mathcal{F}_1 = \mathcal{F}_1(\rho) = \Big\{f \in \Sigma : |Q[f]| \geq \rho^2\Big\}. \tag{5}$$

One can take note that both hypotheses are composite and nonparametric.

1.2. Background on minimax rate- and sharp-optimality

Given the observations $(x_i, \mathbf{t}_i)_{i=1,\dots,n}$, we consider the problem of testing the composite hypothesis $\mathcal{F}_0$ against the nonparametric alternative $\mathcal{F}_1(\rho)$ defined by (5). The goal here is to obtain, if possible, both rate and sharp asymptotics for the cumulative error rate in the minimax setup. These notions are defined as follows. For a fixed small number $\gamma \in (0,1)$, the function r_n^* is called minimax rate of testing if:

- $\exists C' > 0$ such that $\forall\, C < C'$, we have $\liminf_{n\to\infty} \gamma_n(\mathcal{F}_0, \mathcal{F}_1(Cr_n^*)) \geq \gamma$,
- $\exists C'' > 0$ and a test ϕ_n s.t. $\forall\, C > C''$, $\limsup_{n\to\infty} \gamma_n(\mathcal{F}_0, \mathcal{F}_1(Cr_n^*), \phi_n) \leq \gamma$.

A testing procedure ϕ_n is called minimax rate-optimal if the condition $\limsup_{n\to\infty} \gamma_n(\mathcal{F}_0, \mathcal{F}_1(Cr_n^*), \phi_n) \leq \gamma$ holds for some $C > 0$. Note that the minimax rate and the rate-optimal test may depend on the prescribed significance level γ. However, in most situations this dependence cancels out from the rate and appears only in the constants. If the constants C' and C'' coincide, then their common value is called exact separation constant and any test satisfying the second condition is called minimax sharp optimal. The minimax rate r_n^* is actually not uniquely defined, but the product of the minimax rate with the exact separation constant is uniquely defined up to an asymptotic equivalence. For more details on minimax hypotheses testing we refer to (Ingster and Suslina, 2003).

While minimax rate-optimality is a desirable feature for a testing procedure, it may still lead to overly conservative tests. A (partial) remedy for this issue is to consider sharp asymptotics of the error rate. In fact, one can often prove that when $n \to \infty$,

$$\gamma_n(\mathcal{F}_0, \mathcal{F}_1(\rho)) = 2\Phi(-u_n(\rho)) + o(1), \tag{6}$$

where Φ is the c.d.f. of the standard Gaussian distribution, $u_n(\cdot)$ is some "simple" function from $\mathbb{R}_+$ to $\mathbb{R}$ and $o(1)$ is a term tending to zero uniformly in ρ as $n \to \infty$. This relation implies that by determining r_n^* as a solution with respect to ρ to the equation $u_n(\rho) = z_{1-\gamma/2}$—where z_α stands for the α-quantile of the standard Gaussian distribution—we get not only the minimax rate, but also the exact separation constant. When relation (6) is satisfied, we say that Gaussian asymptotics hold.

1.3. Overview of the main contributions

Our contributions focus on the case where the smoothness class Σ is an ellipsoid in $L_2(\Delta)$ and the quadratic functional Q admits a diagonal form in the orthonormal basis corresponding to the directions of the axes of the ellipsoid Σ. To be more precise, let $\mathcal{L}$ be a countable set and $\{\varphi_l\}_{l\in\mathcal{L}}$ be an orthonormal system in $L_2(\Delta)$. For a function $f \in L_2(\Delta)$, let $\boldsymbol{\theta}[f] = \{\theta_l[f]\}_{l\in\mathcal{L}}$ be the generalized Fourier coefficients with respect to this system, *i.e.*, $\theta_l[f] = \langle f, \varphi_l\rangle$, where $\langle\cdot,\cdot\rangle$ denotes the inner product in $L_2(\Delta)$. The functional sets $\Sigma \subset L_2(\Delta)$ under

consideration are subsets of ellipsoids with directions of axes $\{\varphi_l\}_{l\in\mathcal{L}}$ and with coefficients $\mathbf{c} = \{c_l\}_{l\in\mathcal{L}} \in \mathbb{R}_+^{\mathcal{L}}$:

$$\Sigma \subset \Big\{ f = \sum\nolimits_{l\in\mathcal{L}} \theta_l[f]\varphi_l : \ \sum\nolimits_{l\in\mathcal{L}} c_l\theta_l[f]^2 \le 1 \Big\}. \tag{7}$$

The diagonal quadratic functional is defined by a set of coefficients $\mathbf{q} = \{q_l\}_{l\in\mathcal{L}}$: $Q[f] = \sum_{l\in\mathcal{L}} q_l\theta_l[f]^2$. Note that if Q is definite positive, *i.e.*, $q_l > 0$ for all $l \in \mathcal{L}$, then the null hypothesis becomes $f = 0$ and the problem under consideration is known as detection problem. However, the goal of the present work is to consider more general types of diagonal quadratic functionals. Namely, two situations are examined: (a) all the coefficients q_l are nonnegative and (b) the two sets $\mathcal{L}_+ = \{l \in \mathcal{L} : q_l > 0\}$ and $\mathcal{L}_- = \{l \in \mathcal{L} : q_l < 0\}$ are nonempty.

In the first situation, we establish Gaussian asymptotics of the cumulative error rate and propose a minimax sharp-optimal test. Under some conditions, we show that the sequence[1]

$$r^*_{n,\gamma} = \min\Big\{ \rho > 0 : \inf_{\mathbf{v}\in\mathbb{R}_+^{\mathcal{L}}:\langle\mathbf{v},\mathbf{c}\rangle\le 1;\langle\mathbf{v},\mathbf{q}\rangle\ge\rho^2} \|\mathbf{v}\|_2^2 \ge 8n^{-2}z_{1-\gamma/2} \Big\} \tag{8}$$

provides the minimax rate of testing with constants $C' = C'' = 1$. This result is instantiated to some examples motivating our interest for testing the hypotheses (5). One example, closely related to the problem of variable selection (Comminges and Dalalyan, 2012), is testing the relevance of a particular covariate in high-dimensional regression. This problem is considered in a more general setup corresponding to testing that a partial derivative of order $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)$, denoted by $\partial^{\alpha_1+\dots+\alpha_d} f/\partial t_1^{\alpha_1} \dots \partial t_d^{\alpha_d}$, is identically equal to zero against the hypothesis that this derivative is significantly different from 0. As a consequence of our main result, we show that if f lies in the anisotropic Sobolev ball of smoothness $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_d)$, and we set $\delta = \sum_{i=1}^d \alpha_i/\sigma_i$, $\bar\sigma = \big(\frac{1}{d}\sum_{i=1}^d \sigma_i^{-1}\big)^{-1}$, then the minimax optimal-rate is $r_n^* = n^{-2\bar\sigma(1-\delta)/(4\bar\sigma+d)}$ provided that $\delta < 1$ and $\bar\sigma > d/4$. Furthermore, we derive Gaussian asymptotics and exhibit the exact separation constant in this problem.

The second situation we examine in this paper concerns the case where the cardinalities of both $\mathcal{L}_+$ and $\mathcal{L}_-$ are nonzero. A typical application of this kind of problem is testing the equality of the norms of two signals observed in noisy environments. In this set-up, we provide minimax rates of testing and exhibit the presence of two regimes that we call regular regime and irregular regime. In the regular regime, the minimax rate is $r_n^* = n^{-1/4}$, while in the irregular case it may be of the form n^{-a} with an $a < 1/4$ that depends on the degree of smoothness of the functional class.

Note that all our results are non-adaptive: our testing procedures make explicit use of the smoothness characteristics of the function f. Adaptation to the unknown smoothness for the problem we consider is an open question for which the works (Spokoiny, 1996, Gayraud and Pouet, 2005) may be of valuable guidance.

[1] We denote by $\|\cdot\|_2$ and by $\langle\cdot,\cdot\rangle$ the usual norm and the inner product in $\ell_2(\mathcal{L})$, the space of squared summable arrays indexed by $\mathcal{L}$.

1.4. Relation to previous work

Starting from the seminal papers by Ermakov (1990) and Ingster (1993a,b,c), minimax testing of nonparametric hypotheses received a great deal of attention. A detailed review of the literature on this topic being out of scope of this section, we only focus on discussing those previous results which are closely related to the present work. The goal here is to highlight the common points and the most striking differences with the existing literature. The major part of the statistical inference for nonparametric hypotheses testing was developed for the Gaussian white noise model (GWNM) and its equivalent formulation as Gaussian sequence model (GSM). As recent references for the problem of testing a simple hypothesis in these models, we cite (Ermakov, 2011, Ingster et al., 2012), where the reader may find further pointers to previous work. In the present work, the null hypothesis defined by (5) is composite and nonparametric. Early references for minimax results for composite null hypotheses include (Horowitz and Spokoiny, 2001, Pouet, 2001, Gayraud and Pouet, 2001, 2005), where the case of parametric null hypothesis is of main interest. These papers deal with the one-dimensional situation and provide only minimax rates of testing without attaining the exact separation constant. Furthermore, the alternative is defined as the set of functions that are at least at a Euclidean distance ρ from the null hypothesis, which is very different from the alternatives considered in this work.

More recently, the nonasymptotic approach to the minimax testing gained popularity (Baraud et al., 2003, 2005, Laurent et al., 2011, 2012). One of the advantages of the nonasymptotic approach is that it removes the frontier between the concepts of parametric and nonparametric hypotheses, while its limitation is that there is no result on sharp optimality (even the notion itself is not well defined). Note also that all these papers deal with the GSM considering as main application the case of one dimensional signals, as opposed to our set-up of regression with high-dimensional covariates.

Let us review in more details the papers (Ingster and Sapatinas, 2009) and (Laurent et al., 2011) that are very closely related to our work either by the methodology which is used or by the problem of interest. Ingster and Sapatinas (2009) extended some results on the goodness-of-fit testing for the d-dimensional GWNM to the goodness-of-fit testing for the multivariate nonparametric regression model. More precisely, they tested the null hypothesis $H_0 : f = f_0$, where f_0 is a known function, against the alternative $H_1 : f \in \Sigma, \int_\Delta (f - f_0)^2 \geq r_n^2$, where Σ is an ellipsoid in the Hilbert space $L_2(\Delta)$. They obtained both rate and sharp asymptotics for the error probabilities in the minimax setup. So the model they considered is the same as the one we are interested in here, but the hypotheses H_0 and H_1 are substantially different. As a consequence, the testing procedure we propose takes into account the general forms of H_0 and H_1 given by (5) and is different from the asymptotically minimax test of Ingster and Sapatinas (2009). Furthermore, we substantially relaxed the contraint on the noise distribution by replacing Gaussianity assumption by the condition of bounded 4th moment.

Laurent et al. (2011) considered the GWNM from the inverse problem point of view, *i.e.*, when the signal of interest g undergoes a linear transformation T before being observed in noisy environment. This corresponds to $f = T[g]$ with a compact injective operator T. Then the two assertions $g = 0$ and $T[g] = 0$ are equivalent. Consequently, if the goal is to detect the signal f, one can consider the two testing problems:

1. (inverse formulation) $H_0 : T^{-1}[f] = 0$ against $H_1 : \|T^{-1}[f]\|_2 \geq \rho$.
2. (direct formulation) $H_0 : f = 0$ against $H_1 : \|f\|_2 \geq \rho$.

The authors discussed advantages and limitations of each of these two formulations in terms of minimax rates. Depending on the complexity of the inverse problem and on the assumptions on the function to be detected (sparsity or smoothness), they proved that the specific treatment devoted to inverse problem which includes an underlying inversion of the operator, may worsen the detection accuracy. For each situation, they also highlighted the cases where the direct strategy fails while a specific test for inverse formulation works well. The inverse formulation is closely related to our definition (5) of the hypotheses H_0 and H_1, since $Q[f] = \|T^{-1}[f]\|_2^2$ is a quadratic functional. However, our setting is more general in that we consider functionals with non-trivial kernels and with possibly negative diagonal entries.

1.5. Organization

The rest of the paper is organized as follows. The results concerning sharp asymptotics for positive semi-definite diagonal functionals are provided in Section 2. In particular, the rates of separation for a general class of tests called linear U-tests are explored in Subsection 2.2. The asymptotically optimal linear U-test is provided in Subsection 2.3 along with its rate of separation, which is shown to coincide with the minimax exact rate in Subsection 2.4. Section 3 is devoted to a discussion of the assumptions and to the consequences of the main result for some relevant examples. The results for nonpositive and nonnegative diagonal quadratic functionals are stated in Section 4 along with an application to testing the equality of the norms of two signals. A summary and some perspectives are provided in Section 5. Finally, the proofs of the results are postponed to the Appendix.

2. Minimax testing for nonnegative quadratic functionals

2.1. Additional notation

In what follows, the notation $A_n = O(B_n)$ means that there exists a constant $c > 0$ such that $A_n \leq cB_n$ and the notation $A_n = o(B_n)$ means that the ratio A_n/B_n tends to zero. The relation $A_n \sim B_n$ means that A_n/B_n tends to 1, while the relation $A_n \asymp B_n$ means that there exist constants $0 < c_1 < c_2 < \infty$ and n_0 large enough such that $c_1 \leq A_n/B_n \leq c_2$ for $n \geq n_0$. For a real number

c, we denote by c_+ its positive part $\max(0, c)$ and by $\lfloor c \rfloor$ its integer part. For a set $\mathcal{A}$, $\mathbf{1}_A$ stands for its indicator function and $|\mathcal{A}|$ denotes its cardinality. Given a $q > 0$ and a function f, $\|f\|_q = \big(\int_\Delta |f(t)|^q dt\big)^{1/q}$ is the conventional ℓ_q-norm of f. Similarly, for a vector or an array $\mathbf{u}$ indexed by a countable set $\mathcal{L}$, $\|\mathbf{u}\|_q = (\sum_{l\in\mathcal{L}} |u_l|^q)^{1/q}$ is the ℓ_q-norm of $\mathbf{u}$. As usual, we also denote by $\|\mathbf{u}\|_0$ and $\|\mathbf{u}\|_\infty$, respectively, the number of nonzero entries and the magnitude of the largest entry of $\mathbf{u} \in \mathbb{R}^{\mathcal{L}}$.

In the sequel, without loss of generality, we assume that the standard deviation of the noise is equal to one: $\tau = 1$. The case of general but known τ can be deduced as a consequence of our results.

Recall that we consider quadratic functionals Q that are of the form $Q[f] = \sum_{l\in\mathcal{L}} q_l \theta_l[f]^2$, for some given array $\mathbf{q} = \{q_l\}_{l\in\mathcal{L}}$. The major difference between the functional $\sum_{l\in\mathcal{L}} \theta_l[f]^2$ that appears in the problem of detection (Ingster and Sapatinas, 2009, Ingster et al., 2012) and this general functional actually lies in the fact that the support of $\mathbf{q}$ defined by $S_F = \mathrm{supp}(\mathbf{q}) = \big\{l \in \mathcal{L} : q_l \neq 0\big\}$ is generally different from $\mathcal{L}$. Furthermore, large coefficients q_l amplify the error of estimating $Q[f]$ and, therefore, it becomes more difficult to distinguish H_0 from H_1. An interesting question, to which we answer in the next sections, is what is the interplay between $\mathbf{c}$ and $\mathbf{q}$ that makes it possible to distinguish between the null and the alternative.

Let S_F^c denote the complement of S_F and, for a set $L \subset \mathcal{L}$, $\mathrm{span}\big(\{\varphi_l\}_{l\in L}\big)$ be the closed linear subspace of $L_2(\Delta)$ spanned by the set $\{\varphi_l\}_{l\in L}$. Let $\Pi_{S_F} f$ and $\Pi_{S_F^c} f$ be the orthogonal projections of a function $f \in \Sigma$ on $\mathrm{span}\big(\{\varphi_l\}_{l\in S_F}\big)$ and $\mathrm{span}\big(\{\varphi_l\}_{l\in S_F^c}\big)$ respectively. To simplify notation, the subscript S_F^c is omitted in the rest of the paper, *i.e.*, $\Pi_{S_F^c} f$ is replaced by Πf. Finally, throughout this work we will assume that f is centered, *i.e.*, $\int_\Delta f(\mathbf{t})\, d\mathbf{t} = 0$, and that $\{\varphi_l\}$ is an orthonormal basis of the subspace of $L_2(\Delta)$ consisting of all centered functions. In other terms, all the functions φ_i are orthogonal to the constant function.

2.2. Linear U-tests and their error rate

We start by introducing a family of testing procedures that we call linear U-tests. To this end, we split the sample into two parts: a small part of the sample is used to build a pilot estimator $\widehat{\Pi f}_n$ of Πf, whereas the remaining observations are used for distinguishing between H_0 and H_1. Let us set $m = n - \lfloor \sqrt{n} \rfloor$ and call the two parts of the sample $\mathcal{D}_1 = \{(x_i, \mathbf{t}_i) : i = 1, \dots, m\}$ and $\mathcal{D}_2 = \{(x_i, \mathbf{t}_i) : i = m+1, \dots, n\}$. Using a pilot estimator $\widehat{\Pi f}_n$ of Πf, we define the adjusted observations $\tilde{x}_i = x_i - \widehat{\Pi f}_n(\mathbf{t}_i)$ and $\tilde{\mathbf{z}}_i = (\tilde{x}_i, \mathbf{t}_i)$.

Definition 1. Let $\mathbf{w}_n = \{w_{l,n}\}_{l\in S_F}$ be an array of real numbers containing a finite number of nonzero entries and such that $\|\mathbf{w}_n\|_2 = 1$. Let u be a real number. We call a linear U-test based on the array $\mathbf{w}_n$ the procedure $\phi_n^{\mathbf{w}} = \mathbf{1}_{\{U_n^{\mathbf{w}} > u\}}$, where U_n is the linear in $\mathbf{w}_n$ U-statistic defined by

$$U_n = \left(\frac{2}{m(m-1)}\right)^{1/2} \sum_{1\le i<j\le m} \tilde{x}_i \tilde{x}_j \sum_{l\in S_F} w_{l,n} \varphi_l(\mathbf{t}_i) \varphi_l(\mathbf{t}_j). \tag{9}$$

We shall prove that an appropriate choice of $\mathbf{w}_n$ and u leads to a linear U-test that is asymptotically sharp-optimal. The rationale behind this property relies on the by now well-understood principle of smoothing out high frequencies of a noisy signal. In fact, if we call $\{\theta_l[f]\}_{l\in S_F}$ the (relevant part of the) representation of f in the frequency domain, then $\{\frac{1}{m}\sum_{i=1}^m \tilde{x}_i\varphi_l(\mathbf{t}_i)\}_{l\in S_F}$ is a nearly unbiased estimator of this representation. Then, the array $\mathbf{w}_n$ acts as a low pass filter that shrinks to zero the coefficients corresponding to high frequencies in order to prevent over-fitting.

The first step in establishing theoretical guarantees on the error rate of a linear U-test consists in exploring the behavior of the statistic U_n under the null.

Proposition 1. *Let $w_{n,l} \geq 0$ for all $n \in \mathbb{N}$ and $l \in \mathcal{L}$. Assume that $E[\xi_1^4] < \infty$ and the following conditions are fulfilled:*

- *For some $C_w < \infty$, $\|\mathbf{w}_n\|_\infty^2 \|\mathbf{w}_n\|_0 \leq C_w$.*
- *As $n \to \infty$, $\|\mathbf{w}_n\|_0 \to \infty$ so that $\|\mathbf{w}_n\|_0 = o(n)$.*
- *For some $C_\varphi < \infty$, $\sup_{\mathbf{t}\in\Delta} \sum_{l:w_{l,n}\neq 0} \varphi_l^2(\mathbf{t}) \leq C_\varphi \|\mathbf{w}_n\|_0$.*
- *As $n \to \infty$, $\sup_{f\in\Sigma} E_f[\|\Pi f - \widehat{\Pi f}_n\|_4^4] = o(1)$.*

Then, uniformly in $f \in \mathcal{F}_0$, the U-statistic defined by (9) converges in distribution to the standard Gaussian distribution $\mathcal{N}(0,1)$.

In other terms, this proposition claims that under appropriate conditions, for every $u \in \mathbb{R}$, the sequence $\sup_{f\in\mathcal{F}_0} |P_f(U_n > u) - \Phi(u)|$ tends to zero, as n goes to infinity. This means that under the null, the distribution of the test statistic U_n is asymptotically parameter free. This is frequently referred to as Wilks' phenomenon.

To complete the investigation of the error rate of a linear U-test, we need to characterize the behavior of the test statistic U_n under the alternative. As usual, this step is more involved. Roughly speaking, we will show that under the alternative the test statistic U_n is close to a Gaussian random variable with mean $h_n[f, \mathbf{w}_n] = \left(\frac{m(m-1)}{2}\right)^{1/2} \sum_{l\in\mathcal{L}(\mathbf{w}_n)} w_{l,n}\theta_l^2[f]$ and variance 1. The rigorous statement is provided in the next proposition.

Proposition 2. *Let the assumptions of Proposition 1 be satisfied. Assume that in addition:*

- *There exists a sequence ζ_n such that $\zeta_n^{-1} = o(n)$ and $\sup_{l\in S_F: w_{l,n}<\zeta_n} c_l^{-1} = o(1)$.*
- *For some $p > 4$, we have $\sup_{f\in\Sigma} \|\Pi_{S_F} f\|_p < \infty$.*

Then, for every $\rho > 0$, the type II error of the linear U-test based on $\mathbf{w}_n$ satisfies:

$$\sup_{f\in\mathcal{F}_1(\rho)} P_f(\phi_n^{\mathbf{w}} = 0) \leq \sup_{f\in\mathcal{F}_1(\rho)} \Phi(u - h_n[f, \mathbf{w}_n]) + o(1), \qquad (10)$$

where the term $o(1)$ does not depend on ρ.

Let us provide an informal discussion of the assumptions introduced in the previous propositions. The first two assumptions in Proposition 1 mean that most nonzero entries of the array $\mathbf{w}_n$ should be of the same order. Arrays that

have a few spikes and many small entries are discarded by these assumptions. Furthermore, the number of samples in the frequency domain that are not annihilated by $\mathbf{w}_n$ should be small as compared to the sample size n. The third assumption of Proposition 1 is trivially satisfied for bases of bounded functions such as sine and cosine bases and their tensor products. For localized bases like wavelets, this assumption imposes a constraint on the size of the support of $\mathbf{w}_n$: it should not be too small. The last assumption of Proposition 1 will be discussed in more detail later. One should also take note that the only reason for requiring from the functions f to be smooth under the null is the need to be able to construct a uniformly consistent pilot estimator of Πf.

Concerning the assumptions imposed in Proposition 2, the first one means that only coefficients θ_l corresponding to high frequencies are strongly shrunk by $\mathbf{w}_n$. This is a kind of coherence assumption between the smoothing filter $\mathbf{w}_n$ and the coefficients $\mathbf{c} = \{c_l\}_{l\in\mathcal{L}}$ encoding the prior information on the signal smoothness. The second assumption of Proposition 2 is rather weak and usual in the context of regression with random design. It is only needed for getting uniform control of the error rate and the actual value of the norm $\|\Pi_{S_F} f\|_p$ does not enter in any manner in the definition of the testing procedure.

Let us draw now the consequences of the previous propositions on the cumulated error rate of a linear U-test. Using the monotonicity of the Gaussian c.d.f. Φ, under the assumptions of Proposition 2, we get

$$\gamma_n(\mathcal{F}_0, \mathcal{F}_1(\rho), \phi_n^{\mathbf{w}}) \le \Phi(-u) + \Phi\Big(u - \inf_{f\in\mathcal{F}_1(\rho)} h_n[f, \mathbf{w}_n]\Big) + o(1), \tag{11}$$

where the term $o(1)$ is uniform in $\rho > 0$. Using the symmetry of Φ and the monotonicity of Φ' on $\mathbb{R}_+$, one easily checks that the value of the threshold u minimizing the main term in the right-hand side of the last display is $u = \frac{1}{2}\inf_{f\in\mathcal{F}_1(\rho)} h_n[f, \mathbf{w}_n]$. This result provides a constructive tool for determining the rate of separation of a given linear U-test. In fact, one only needs to set $u = z_{1-\gamma/2}$ and find a sequence r_n such that $\inf_{f\in\mathcal{F}_1(r_n)} h_n[f, \mathbf{w}_n] \sim 2z_{1-\gamma/2}$, where z_α is the α-quantile of $\mathcal{N}(0,1)$.

Remark 1. We explain here the use of $\tilde{x}_i$ instead of x_i in our testing procedure. Actually if we were only interested in rate-optimality, this precaution would not have been necessary. The problem only arises when dealing with sharp-optimality and it concerns the variance of U_n. Indeed we need some terms that appear in the variance to tend to zero when $Q[f] = 0$ or $Q[f]$ is small (those terms only need to be bounded for the rate-optimality). If we had used x_i instead of $\tilde{x}_i$, we would have ended up with terms like $\|f\|_2$ in the variance. The information contained in the assertion "$Q[f]$ is small" concerns only the coefficients $\{\theta_l\}_{l\in S_F}$, thus it implies that $\|\Pi_{S_F} f\|_2$ is small but it does not say anything about $\|f\|_2$. We can also remark that this problem does not arise in the Gaussian sequence model as one estimates θ_l^2 by an unbiased estimator whose variance makes appear only θ_l.

Remark 2. We chose to consider only the criterion $\gamma_n(\mathcal{F}_0, \mathcal{F}_1(\rho), \phi_n^{\mathbf{w}})$ so as to simplify the exposition of our results. But we could have dealt with the classical

Neyman-Pearson criterion that we recall here. For a significance level $0 < \alpha < 1$ and a test ψ, we set

$$\alpha(\mathcal{F}_0, \psi) = \sup_{f \in \mathcal{F}_0} P_f(\psi = 1), \quad \beta(\mathcal{F}_1, \psi) = \inf_\psi \sup_{f \in \mathcal{F}_1} P_f(\psi = 0),$$

Instead of the minimax risk $\gamma_n(\mathcal{F}_0, \mathcal{F}_1(\rho))$ we could have considered the quantity $\beta_n(\mathcal{F}_0, \mathcal{F}_1(\rho)) = \inf_{\psi:\alpha(\mathcal{F}_0,\psi)\le\alpha} \beta(\mathcal{F}_1(\rho), \psi)$. This criterion is considered in Ingster and Sapatinas (2009) and more generally in Ingster and Suslina (2003). The transposition to our case is straightforward.

2.3. Minimax linear U-tests

The relation (11) being valid for a large variety of arrays $\mathbf{w}_n$, it is natural to look for a $\mathbf{w}_n$ minimizing the right-hand side of (11). This leads to the following saddle point problem:

$$\sup_{\substack{\mathbf{w}\in\mathbb{R}_+^{\mathcal{L}} \\ \|\mathbf{w}\|_2=1}} \inf_{f\in\mathcal{F}_1(\rho)} \sum_{l\in\mathcal{L}} w_l \theta_l[f]^2 = \sup_{\substack{\mathbf{w}\in\mathbb{R}_+^{\mathcal{L}} \\ \|\mathbf{w}\|_2=1}} \inf_{\substack{\mathbf{v}\in\mathbb{R}_+^{\mathcal{L}} \\ \langle \mathbf{v},\mathbf{c}\rangle\le 1, \langle \mathbf{v},\mathbf{q}\rangle\ge\rho^2}} \langle \mathbf{w}, \mathbf{v}\rangle. \tag{12}$$

It turns out that this saddle point problem can be solved with respect to $\mathbf{w}$ and leads to a one-parameter family of smoothing filters $\mathbf{w}$.

Proposition 3. *Assume that for every $T > 0$, the set $\mathcal{N}(T) = \{l \in S_F : c_l < Tq_l\}$ is finite. For a given $\rho > 0$, assume that the equation*

$$\frac{\sum_{l\in\mathcal{L}} q_l (Tq_l - c_l)_+}{\sum_{l\in\mathcal{L}} c_l (Tq_l - c_l)_+} = \rho^2 \tag{13}$$

has a solution and denote it by T_ρ. Then, the pair $(\mathbf{w}^*, \mathbf{v}^*)$ defined by

$$v_l^* = \frac{(T_\rho q_l - c_l)_+}{\sum_{l\in\mathcal{L}} c_l (T_\rho q_l - c_l)_+} \qquad w_l^* = \frac{v_l^*}{\|\mathbf{v}^*\|_2} \tag{14}$$

provides a solution to the saddle point problem (12), that is

$$\langle \mathbf{w}^*, \mathbf{v}^*\rangle = \sup_{\substack{\mathbf{w}\in\mathbb{R}_+^{\mathcal{L}} \\ \|\mathbf{w}\|_2=1}} \inf_{\substack{\mathbf{v}\in\mathbb{R}_+^{\mathcal{L}} \\ \langle \mathbf{v},\mathbf{c}\rangle\le 1, \langle \mathbf{v},\mathbf{q}\rangle\ge\rho^2}} \langle \mathbf{w}, \mathbf{v}\rangle = \inf_{\substack{\mathbf{v}\in\mathbb{R}_+^{\mathcal{L}} \\ \langle \mathbf{v},\mathbf{c}\rangle\le 1, \langle \mathbf{v},\mathbf{q}\rangle\ge\rho^2}} \langle \mathbf{w}^*, \mathbf{v}\rangle.$$

This result tells us that the "optimal" weights $\mathbf{w}_n$ for the linear U-test $\phi_n^{\mathbf{w}}$ should be of the form (14), which is particularly interesting because of its dependence on only one parameter $T > 0$. The next theorem provides a simple strategy for determining the minimax sharp-optimal test among linear U-tests satisfying some mild assumptions. We will show later in this section that this test is also minimax sharp-optimal among all possible tests.

Theorem 1. *Assume that $E[\xi_1^4] < \infty$ and for every $T > 0$, the set $\mathcal{N}(T) = \{l \in S_F : c_l < Tq_l\}$ is finite. For a prescribed significance level $\gamma \in (0, 1)$, let $T_{n,\gamma}$*

be a sequence of positive numbers such that the following relation holds true: as $n \to \infty$,

$$\left(\frac{m(m-1)}{2}\sum_{l\in\mathcal{L}}(T_{n,\gamma}q_l - c_l)_+^2\right)^{1/2} = \left(\sum_{l\in\mathcal{L}} c_l(T_{n,\gamma}q_l - c_l)_+\right)(2z_{1-\gamma/2} + o(1)). \tag{15}$$

Let us define

$$r^*_{n,\gamma} = \left\{\frac{\sum_{l\in\mathcal{L}} q_l(T_{n,\gamma}q_l - c_l)_+}{\sum_{l\in\mathcal{L}} c_l(T_{n,\gamma}q_l - c_l)_+}\right\}^{1/2}. \tag{16}$$

If the following conditions are fulfilled:

[C1] *For some* $C_1 > 0$, $|\mathcal{N}(T_{n,\gamma})| \max_{l\in\mathcal{N}(T_{n,\gamma})} q_l^2 \le C_1 \sum_{l\in\mathcal{N}(T_{n,\gamma})} \left(q_l - \frac{c_l}{T_{n,\gamma}}\right)^2$.

[C2] *As* $n \to \infty$, $\sum_{l\in\mathcal{N}(T_{n,\gamma})} q_l^2 = o(n^2 \min_{l\in\mathcal{N}(T_{n,\gamma})} q_l^2)$.

[C3] *For some constant* $C_3 > 0$, $\sup_{\mathbf{t}\in\Delta} \sum_{l\in\mathcal{N}(T_{n,\gamma})} \varphi_l^2(\mathbf{t}) \le C_3|\mathcal{N}(T_{n,\gamma})|$.

[C4] *As* $n \to \infty$, $|\mathcal{N}(T_{n,\gamma})| \to \infty$ *so that* $|\mathcal{N}(T_{n,\gamma})| = o(n)$.

[C5] *As* $n \to \infty$, $T_{n,\gamma} \inf_{l\in S_F} q_l$ *tends to* $+\infty$.

[C6] *As* $n \to \infty$, $\sup_{f\in\Sigma} E_f[\|\Pi f - \widehat{\Pi} f_n\|_4^4] = o(1)$.

[C7] *For some* $p > 4$, *it holds that* $\sup_{f\in\Sigma} \|\Pi_{S_F} f\|_p < \infty$.

then the linear U-test $\widehat{\phi}^*_n = \mathbf{1}_{\{U^{\widehat{\mathbf{w}}^*}_n > z_{1-\gamma/2}\}}$ *based on the array* $\widehat{\mathbf{w}}^*_n$ *defined by*

$$\widehat{w}^*_{l,n} = \frac{(T_{n,\gamma}q_l - c_l)_+}{\left[\sum_{l'\in\mathcal{L}}(T_{n,\gamma}q_{l'} - c_{l'})_+^2\right]^{1/2}}$$

satisfies

$$\gamma_n(\mathcal{F}_0, \mathcal{F}_1(r^*_{n,\gamma}), \widehat{\phi}^*_n) \le \gamma + o(1), \qquad as \qquad n \to \infty. \tag{17}$$

The proof of this result, provided in the Appendix, is a direct consequence of Proposition 1, 2 and 3. As we shall see below, the rate $r^*_{n,\gamma}$ defined in Theorem 1 is the minimax sharp-rate in the problem of testing hypotheses (5), provided that the assumptions of the theorem are fulfilled. As expected, getting such a strong result requires non-trivial assumptions on the nature of the functional class, that of the hypotheses to be tested, as well as the interplay between them. Some short comments on these assumptions are provided in the remark below, with a further development left to subsequent sections.

Remark 3. The very first assumption is that the set $\mathcal{N}(T)$ is finite. It is necessary for ensuring that the linear U-test we introduced is computable. This assumption is fulfilled when, roughly speaking, the coefficients which express the regularity, $\{c_l\}_{l\in\mathcal{L}}$, grow at a faster rate than the coefficients $\{q_l\}_{l\in\mathcal{L}}$ of the quadratic functional Q. Assumptions **[C1]**, **[C2]**, **[C4]** and **[C5]** are satisfied in most cases we are interested in. Two illustrative examples—concerning Sobolev ellipsoids with quadratic functionals related to partial derivatives—for which these hypotheses are satisfied are presented in Subsections 3.3 and 3.4. Assumption **[C3]** is essentially a constraint on the basis $\{\varphi_l\}$; we show in Subsection 3.1 that it is

satisfied by many bases commonly used in statistical literature. **[C6]** and **[C7]** are related to additional technicalities brought by the regression model, which force us to impose more regularity than in the Gaussian sequence model.

Remark 4. The result stated in Theorem 1 is in the spirit of the previous work on the sharp asymptotics in minimax testing, initiated by Ermakov (1990) in the problem of detection (*i.e.*, $Q[f] = \|f\|_2^2$) under Gaussian white noise. The explicit form[2] of the weights $\widehat{w}^*_{l,n}$ is obtained by solving a quadratic optimization problem called the extremal problem in a series of recent works (Ingster and Suslina, 2003, Ingster and Sapatinas, 2009, Ingster and Stepanova, 2011, Ingster et al., 2012), see also Ermakov (2004) for a similar result in the heteroscedastic GWNM. In the case $q_l = 1$, $\forall l \in \mathcal{L}$, the aforementioned extremal problem is equivalent to the saddle point problem (12). In a nutshell, the main differences of Theorem 1 as compared to the existing results is the extension to the case of general coefficients q_l and to non-Gaussian error distribution, as well as the use in the test statistic $U_n^{\mathbf{w}}$ of the adjusted responses $\{\tilde{x}_i\}$ instead of the raw data $\{x_i\}$.

2.4. Lower bound

We shall state in this section the result showing that the rate $r^*_{n,\gamma}$ introduced in Theorem 1 is the minimax rate of testing and the exact separation constant associated with this rate is equal to one. This also implies that the testing procedure proposed in previous subsection is not only minimax rate-optimal but also minimax sharp-optimal among all possible testing procedures. In this subsection, we consider the functional classes $\Sigma = \Sigma_{p,L}$ defined by

$$\Sigma_{p,L} = \Big\{ f = \sum\nolimits_{l\in\mathcal{L}} \theta_l[f]\varphi_l : \ \sum\nolimits_{l\in\mathcal{L}} c_l\theta_l[f]^2 \le 1, \ \|f\|_p \le L, \ \Pi_{S_F^c} f = 0 \Big\}.$$

Clearly, for $p > 4$, this functional class is smaller than those satisfying conditions of Theorem 1. Therefore, any lower bound proven for these functional classes will also be a lower bound for the functional classes for which Theorem 1 is applicable.

Theorem 2. *Assume that ξ_is are standard Gaussian random variables and that for every $T > 0$, the set $\mathcal{N}(T) = \{l \in S_F : c_l < Tq_l\}$ is finite. For a prescribed significance level $\gamma \in (0,1)$, let $T_{n,\gamma}$ and $r^*_{n,\gamma}$ be as in Theorem 1. If conditions* **[C1]**, **[C3]** *and*

[C8] *as $n \to \infty$, $|\mathcal{N}(T_{n,\gamma})| \to \infty$ so that $|\mathcal{N}(T_{n,\gamma})| \log(|\mathcal{N}(T_{n,\gamma})|) = o(n)$,*

[C9] *as $n \to \infty$, $\max_{l\in\mathcal{N}(T_{n,\gamma})} c_l = o(n|\mathcal{N}(T_{n,\gamma})|^{1/2})$,*

are fulfilled, then for every $C < 1$ the minimax risk satisfies

$$\gamma_n(\mathcal{F}_0, \mathcal{F}_1(Cr^*_{n,\gamma})) \ge \gamma + o(1), \qquad \text{as} \qquad n \to \infty. \tag{18}$$

[2] The first use of this type of weights for statistical purposes goes back to Pinsker (1980), who showed that these weights lead to asymptotically minimax nonparametric estimators of the signal observed in Gaussian white noise.

Although the main steps of the proof of this theorem, postponed to the Appendix, are close to those of (Ingster and Sapatinas, 2009), we have made several improvements which resulted in both shorter and more transparent proof and relaxed assumptions. The most notable improvement is perhaps the fact that in condition **[C3]** it is not necessary to have $C_3 = 1$. We will further discuss this point and the other assumptions in the next section.

Remark 5. If we were only interested in minimax rate-optimality, we could have used simpler prior in the proof of Theorem 2 which would also yield the desired lower bound under slightly weaker assumptions. One can also deduce from the proof that for a concrete pair $(\mathbf{c}, \mathbf{q})$, a simple way to deduce the minimax rate of separation consists in finding a sequence r_n such that $n(r_n)^2 \asymp M(r_n^{-2})^{1/2}$, where $M(T) = \sum_{l \in \mathcal{N}(T)} q_l^2$.

3. Examples

3.1. Bases satisfying assumption [C3]

First we give examples of orthonormal bases satisfying assumption **[C3]**, irrespectively of the nature of arrays $\mathbf{c}$ and $\mathbf{q}$ defining the smoothness class and the quadratic functional Q. One can take note that despite more general settings considered in the present work, our assumption **[C3]** is significantly weaker than the corresponding assumption in (Ingster and Sapatinas, 2009), which requires C_3 to be equal to one. In fact, in a remark, Ingster and Sapatinas (2009) suggest that their proof remains valid under our assumption **[C3]** if assumption **[C4]** is strengthened to $|\mathcal{N}(T_{n,\gamma})| = o(n^{2/3})$. Due to a better analysis, we succeeded to establish sharp asymptotics under the weak version of **[C3]** without any additional price (except that a logarithmic factor appears now in the corresponding condition in Theorem 2).

Fourier basis Let us consider first the following Fourier basis in dimension d for which $\mathcal{L} = \mathbb{Z}^d$ and

$$\varphi_{\boldsymbol{k}}(\mathbf{t}) = \begin{cases} 1, & \boldsymbol{k} = 0, \\ \sqrt{2}\cos(2\pi\, \boldsymbol{k} \cdot \mathbf{t}), & \boldsymbol{k} \in (\mathbb{Z}^d)_+, \\ \sqrt{2}\sin(2\pi\, \boldsymbol{k} \cdot \mathbf{t}), & -\boldsymbol{k} \in (\mathbb{Z}^d)_+, \end{cases} \tag{19}$$

where $(\mathbb{Z}^d)_+$ denotes the set of all $\boldsymbol{k} \in \mathbb{Z}^d \setminus \{0\}$ such that the first nonzero element of $\boldsymbol{k}$ is positive and $\boldsymbol{k} \cdot \mathbf{t}$ stands for the usual inner product in $\mathbb{R}^d$. Since all the basis functions are bounded by $\sqrt{2}$, **[C3]** is obviously satisfied with $C_3 = 2$. Furthermore, if the set $\mathcal{N}(T)$ is symmetric, *i.e.*, $\boldsymbol{k} \in \mathcal{N}(T)$ implies $-\boldsymbol{k} \in \mathcal{N}(T)$, then **[C3]** is fulfilled with $C_3 = 1$.

Tensor product Fourier basis We can also consider the traditional tensor product Fourier basis as in Ingster and Sapatinas (2009). **[C3]** is then obviously satisfied with $C_3 = 2^d$. Moreover, if the set $\mathcal{N}(T)$ is orthosymmetric, *i.e.*, $(k_1, \dots, k_d) \in \mathcal{N}(T)$ implies $(\pm k_1, \dots, \pm k_d) \in \mathcal{N}(T)$, then **[C3]** is fulfilled with $C_3 = 1$.

Haar basis Let $\{\varphi_{j,k}(\cdot), j \in \mathbb{N}, k \in \{1,\dots,2^j\}\}$, be the standard orthonormal Haar basis on $[0,1]$, where j is the scale parameter and k is the shift. The tensor product $(\varphi_{\boldsymbol{j},\boldsymbol{k}})_{\boldsymbol{j},\boldsymbol{k}}$ Haar basis is then

$$\varphi_{\boldsymbol{j},\boldsymbol{k}} = \prod_{i=1}^{d} \varphi_{j_i,k_i},$$

where $\boldsymbol{j} = (j_1,\dots,j_d)$ and $\boldsymbol{k} = (k_1,\dots,k_d)$. As shown in (Ingster and Sapatinas, 2009), under the extra assumption that the coefficients $c_l = c_{\boldsymbol{j},\boldsymbol{k}}$ and $q_l = q_{\boldsymbol{j},\boldsymbol{k}}$ depend only on the scale parameter, *i.e.*, $c_{\boldsymbol{j},\boldsymbol{k}} = c_{\boldsymbol{j}}$ and $q_{\boldsymbol{j},\boldsymbol{k}} = q_{\boldsymbol{j}}$, assumption [C3] is satisfied with $C_3 = 1$. Note that the same holds true for the multivariate Haar basis defined in the more commonly used way (see Cohen (2003), chapter 2): $\{\varphi_l(\mathbf{t}) = \prod_{i=1}^{d} \psi_{j,k_i}^{\omega_i}(t_i)\}$, where $l = (j, \boldsymbol{k}, \boldsymbol{\omega})$ such that $j \in \mathbb{N}$, $\boldsymbol{k} \in \{1,\dots,2^j\}^d$ and $\boldsymbol{\omega} \in \{0,1\}^d \setminus \{0\}$ with $\psi_{j,k}^0$ and $\psi_{j,k}^1$ being the scaled and shifted mother wavelet and father wavelet, respectively.

Compactly supported wavelet basis Since we are not limited to the case $C_3 = 1$, any orthonormal wavelet basis satisfies assumption **[C3]**, as long as the wavelets are compactly supported and provided that the coefficients c_l and q_l depend on the level of the resolution and not on the shift.

3.2. Examples of estimators satisfying [C6]

We present below pilot estimators that in two different contexts satisfy assumption **[C6]**.

Tensor-product Fourier basis For the first example, we assume that the orthonormal system $\{\varphi_l\}$ is the tensor product Fourier basis. Then we have $\sup_l \sup_{t\in\Delta} |\varphi_l(t)| \le 2^{d/2}$. The anisotropic Sobolev ball with radius R and smoothness $\boldsymbol{\sigma} = (\sigma_1,\dots,\sigma_d) \in (0,\infty)^d$ is defined by

$$W_2^{\boldsymbol{\sigma}}(R) = \Big\{ f : \sum\nolimits_{\boldsymbol{l}\in\mathbb{Z}^d} \sum\nolimits_{i=1}^{d} (2\pi l_i)^{2\sigma_i} \theta_{\boldsymbol{l}}[f]^2 \le R \Big\}.$$

The estimator we suggest to use is constructed as follows. We first estimate $\theta_l[f]$ by $\widehat{\theta}_l = \frac{1}{n}\sum_{i=1}^{n} x_i \varphi_l(\mathbf{t}_i)$. Then we choose a tuning parameter $T = T_n > 0$ and define the pilot estimator

$$\widehat{\Pi f}_n = \sum_{l\in S_F^{\mathbf{c}}: c_l < T} \widehat{\theta}_l \varphi_l. \tag{20}$$

To ease notation, we set $\mathcal{N}_1(T) = \{l \in S_F^{\mathbf{c}} : c_l < T\}$ and $\mathcal{N}_2(T) = S_F^{\mathbf{c}} \setminus \mathcal{N}_1(T)$.

Lemma 1. *Assume that either one of the following conditions is satisfied:*

- $\mathbf{c}$ *satisfies the condition* $\sum_l c_l^{-1} < \infty$,
- $\Sigma \subset W_2^{\boldsymbol{\sigma}}(R)$ *for some* $R > 0$ *and for some* $\boldsymbol{\sigma} \in (0,\infty)^d$ *such that* $\bar{\sigma} = (\frac{1}{d}\sum_i \frac{1}{\sigma_i})^{-1} > d/4$.

If $T = T_n \to \infty$ so that $|\mathcal{N}_1(T)| = o(n^{1/2})$, then $\widehat{\Pi f}_n$ defined by (20) satisfies condition **[C6]**.

Compactly supported orthonormal wavelet basis The same method can be applied in the case of an orthonormal basis of compactly supported wavelets of $L_2[0,1]^d$. We suppose that the coefficients $c_l = c_{j,\boldsymbol{k}}$ correspond to those of a Besov ball $B^s_{2,2}$, *i.e.*, $c_j = 2^{js}$, and that $\sigma = s - d/4 > 0$. Let us set, for $J \in \mathbb{N}$,

$$\widehat{\Pi f}_n = \sum\nolimits_{\boldsymbol{k}\in[1,2^J]^d} \widehat{\alpha}_{J,\boldsymbol{k}}\varphi_{J,\boldsymbol{k}} \quad \text{where} \quad \widehat{\alpha}_{J,\boldsymbol{k}} = \frac{1}{n}\sum\nolimits_{i=1}^n x_i\varphi_{J,\boldsymbol{k}}(\mathbf{t}_i).$$

Lemma 2. *If the sequence $J = J_n$ tends to infinity so that $2^{Jd} = o(n)$, then $\sup_{f\in\Sigma} E_f\|\Pi f - \widehat{\Pi f}_n\|^4 \to 0$ as $n \to \infty$.*

In the following two subsections, we apply the previous results to two examples of quadratic functionals involving derivatives. The orthonormal system we use is the tensor product Fourier basis.

3.3. Testing partial derivatives

We assume here that f belongs to a Sobolev class with anisotropic constraints and the quadratic functional Q corresponds, roughly speaking, to the squared L_2-norm of a partial derivative. More precisely, let $\boldsymbol{\alpha} \in \mathbb{R}^d_+$ and $\boldsymbol{\sigma} \in \mathbb{R}^d_+$ be two given vectors and define, for every $\boldsymbol{l} \in \mathcal{L} = \mathbb{Z}^d \setminus \{0\}$,

$$q_{\boldsymbol{l}} = \prod\nolimits_{j=1}^d (2\pi l_j)^{2\alpha_j}, \qquad \text{and} \qquad c_{\boldsymbol{l}} = \sum\nolimits_{j=1}^d (2\pi l_j)^{2\sigma_j}.$$

We will assume that $\sum_{j=1}^d (\alpha_j/\sigma_j) < 1$.

For a function $f = \sum_{l\in\mathcal{L}} \theta_l\varphi_l \in L_2(\Delta)$, we set $\|f\|^2_{2,c} = \sum_{l\in\mathcal{L}} c_l\theta_l^2$ and $\|f\|^2_{2,q} = \sum_{l\in\mathcal{L}} q_l\theta_l^2$. Then, for a 1-periodic function which is differentiable enough, and if the α_j and σ_j are integers, we have

$$\|f\|^2_{2,q} = \|\partial^{\sum_j \alpha_j} f/\partial t_1^{\alpha_1}\dots\partial t_d^{\alpha_d}\|_2^2, \quad \text{and} \quad \|f\|^2_{2,c} = \sum\nolimits_{j=1}^d \|\partial^{\sigma_j} f/\partial t_j^{\sigma_j}\|_2^2.$$

Proposition 4. *Let us define δ, $\bar\sigma$, (κ_j) and κ by $\delta = \sum_{j=1}^d \alpha_j/\sigma_j$, $\frac{1}{\bar\sigma} = \frac{1}{d}\sum_{j=1}^d \frac{1}{\sigma_j}$, $\kappa_j = \frac{1}{2\sigma_j} + \frac{\alpha_j}{\sigma_j}\frac{4\bar\sigma+d}{2\bar\sigma(1-\delta)}$ and $\kappa = \sum_{j=1}^d \kappa_j$. If $\delta < 1$ and $\bar\sigma > d/4$, then the exact minimax rate $r^*_{n,\gamma}$ is given by $r^*_{n,\gamma} = C^*_\gamma r^*_n(1+o(1))$, where the minimax rate r^*_n and the exact separation constant are*

$$r^*_n = n^{-\frac{2\bar\sigma(1-\delta)}{4\bar\sigma+d}}, \quad \text{and} \quad C^*_\gamma = \big(4z^2_{1-\gamma/2}\kappa C(d,\boldsymbol{\sigma},\boldsymbol{\alpha})\big)^{\frac{\bar\sigma(1-\delta)}{4\bar\sigma+d}}(1+2\kappa^{-1})^{\frac{2(1+\delta)\bar\sigma+d}{2(4\bar\sigma+d)}}$$

with

$$C(d,\boldsymbol{\sigma},\boldsymbol{\alpha}) = \pi^{-d}\frac{\prod_{i=1}^d \Gamma(\kappa_i)}{\big(\prod_{i=1}^d \sigma_i\big)(1-\delta)\Gamma(\kappa+2)}.$$

Furthermore, the sequence of linear U-tests ϕ_n of Theorem 1 is asymptotically minimax with $T_{n,\gamma} \sim (r^*_{n,\gamma})^{-2}(1+2\kappa^{-1})$.

Remark 6. The previous result can be used for performing dimensionality reduction through variable selection (Comminges and Dalalyan, 2011, 2012). Indeed, in a high-dimensional set-up it is of central interest to eliminate the irrelevant covariates. The coordinate t_i of $\mathbf{t}$ is irrelevant if f is constant on the line $\{\mathbf{t} \in \Delta : t_j = a_j \text{ for all } j \neq i\}$, whatever the vector $\mathbf{a} \in \Delta$ is. This implies that the i^{th} partial derivative of f is zero. Therefore, one can test the relevance of a variable, say t_1, by comparing $\|\partial f/\partial t_1\|_2$ with 0. In our notation, this amounts to testing hypotheses (5) with $Q[f] = \|f\|^2_{2,\mathbf{q}}$ such that $q_{\boldsymbol{l}} = (2\pi l_1)^2$. Combining Proposition 4 and Theorem 1, one can easily deduce a minimax sharp-optimal test and the minimax sharp-rates for this variable selection problem.

Remark 7. Another interesting particular case of the setting described in this subsection concerns the problem of component identification in partial linear models (Samarov et al., 2005). We say that f obeys a partial linear model if for some small subset J of indices $\{1, \dots, d\}$ and for a vector $\boldsymbol{\beta} \in \mathbb{R}^{|J^c|}$, one can write $f(\mathbf{t}) = g(\mathbf{t}_J) + \boldsymbol{\beta}^\top \mathbf{t}_{J^c}$ for every $\mathbf{t} \in \Delta$. The problem of component identification in this model is to determine for an index j whether $j \in J$ or not. One way of addressing this issue is to perform a test of hypothesis $Q[f] = \|f\|^2_{2,\mathbf{q}} = 0$, where $q_{\boldsymbol{l}} = (2\pi l_j)^4$. Roughly speaking, this corresponds to checking whether the second order partial derivative of f with respect to t_j is zero or not (if the null is not rejected, then $j \in J^c$). Once again, Proposition 4 and Theorem 1 provide a minimax sharp-optimal test for this problem along with the minimax rates and exact separation constants.

Remark 8. In the case where the covariates $\mathbf{t}_i$ are not observable and only x_i's are available, our model coincides with the convolution model, for which the minimax rates of testing were obtained by Butucea (2007) in the one-dimensional case with simple null hypothesis. It would be interesting to extend our results to such a model and to get minimax rates and, if possible, separation constants in the multidimensional convolution model.

3.4. Testing the relevance of a direction in a single-index model

Recall that a single-index model is a particular case of (1) corresponding to functions f that can be written in the form $f(\mathbf{t}) = g(\boldsymbol{\beta}_0^\top \mathbf{t})$ for some univariate function $g : \mathbb{R} \to \mathbb{R}$ and some vector $\boldsymbol{\beta}_0 \in \mathbb{R}^d$. Assume now that for a candidate vector $\boldsymbol{\beta} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$ we wish to test the goodness-of-fit of the single-index model (Dalalyan et al., 2008, Gaïffas and Lecué, 2007). This corresponds to testing the hypothesis

$$\exists g : \mathbb{R} \to \mathbb{R} \qquad \text{such that} \qquad f(\mathbf{t}) = g(\boldsymbol{\beta}^\top \mathbf{t}), \quad \forall \mathbf{t} \in \Delta.$$

This condition implies that $\frac{\partial f}{\partial t_i}(\mathbf{t}) \equiv \frac{\beta_i}{\|\boldsymbol{\beta}\|_2^2} \sum_{j=1}^d \beta_j \frac{\partial f}{\partial t_j}(\mathbf{t}) = \frac{\beta_i}{\|\boldsymbol{\beta}\|_2^2}\, \boldsymbol{\beta}^\top \nabla f(\mathbf{t})$, $\forall i \in \{1, \dots, d\}$, which in turn can be written as

$$\sum_{i=1}^d \Big(\frac{\partial f}{\partial t_i} - \frac{\beta_i}{\|\boldsymbol{\beta}\|_2^2}\, \boldsymbol{\beta}^\top \nabla f(\mathbf{t}) \Big)^2 \equiv 0.$$

Without loss of generality, we assume that $\|\boldsymbol{\beta}\|_2 = 1$. Let us set $q_{\boldsymbol{l}} = \sum_{i=1}^d (2\pi)^2 \times \big(l_i - (\boldsymbol{\beta}^\top \boldsymbol{l})\beta_i\big)^2 = (2\pi)^2\big(\|\boldsymbol{l}\|_2^2 - (\beta^\top \boldsymbol{l})^2\big)$. We consider homogeneous Sobolev smoothness classes, that is $c_{\boldsymbol{l}} = \sum_{i=1}^d (2\pi l_i)^{2\sigma}$, with $\sigma > d/4$. Then, when σ is an integer, for a 1-periodic function which is smooth enough,

$$\|f\|_{2,\mathbf{c}}^2 = \sum_{i=1}^d \Big\| \frac{\partial^\sigma f}{\partial t_i^\sigma} \Big\|_2^2 \quad \text{and} \quad \|f\|_{2,\mathbf{q}}^2 = \sum_{i=1}^d \Big\| \frac{\partial f}{\partial t_i} - \beta_i [\boldsymbol{\beta}^\top \nabla f] \Big\|^2.$$

To state the result providing the minimax rate and the exact constant in this problem, we introduce the constants

$$\bar{C}_0 = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \big[\|\mathbf{x}\|_2^2 - (\boldsymbol{\beta}^\top \mathbf{x})^2 - \|\mathbf{x}\|_{2\sigma}^{2\sigma}\big]_+^2 d\mathbf{x},$$
$$\bar{C}_1 = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \big(\|\mathbf{x}\|_2^2 - (\boldsymbol{\beta}^\top \mathbf{x})^2\big)\big(\|\mathbf{x}\|_2^2 - (\boldsymbol{\beta}^\top \mathbf{x})^2 - \|\mathbf{x}\|_{2\sigma}^{2\sigma}\big)_+ d\mathbf{x},$$

and $\bar{C}_2 = \bar{C}_1 - \bar{C}_0$.

Proposition 5. *In the setting described above, the exact minimax rate $r_{n,\gamma}^*$ is given by $r_{n,\gamma}^* = C_\gamma^* r_n^*(1 + o(1))$, where*

$$r_n^* = n^{-\frac{2(\sigma-1)}{4\sigma+d}} \qquad \text{and} \qquad C_\gamma^* = \Big(\frac{4 z_{1-\gamma/2} (\bar{C}_1/\bar{C}_2)^{\frac{d+4}{2(\sigma-1)}} \bar{C}_1^2}{\sigma^{d-1}(\sigma-1)\bar{C}_0}\Big)^{\frac{\sigma-1}{4\sigma+d}}.$$

The sequence of tests ϕ_n of Theorem 1 is minimax sharp-optimal if $T = T_{n,\gamma}$ is chosen as $T = (C_\gamma^* r_n^*)^{-2}\big(\bar{C}_1/\bar{C}_2\big)$.

Remark 9. The testing procedures provided in Propositions 4 and 5 require the precise knowledge of the smoothness parameter $\boldsymbol{\sigma}$, which may not be available in practice. Indeed, the parameter $\boldsymbol{\sigma}$ explicitly enters in the definition of the tuning parameter T_n. The adaptation to the unknown smoothness $\boldsymbol{\sigma}$ is an interesting problem for future research. We believe that rates of separation similar to those of Propositions 4 and 5 can be established for adaptive tests (up to logarithmic factors) using the Berry-Esseen type theorem for degenerate U-statistics of Butucea et al. (2009).

4. Nonpositive and nonnegative diagonal quadratic functionals

In this section we consider the more general setting obtained by abandoning the assumption that all the entries q_l of the array $\mathbf{q}$ have the same sign. That is, we still have $Q[f] = \sum_{l\in\mathcal{L}} q_l \theta_l^2$, but now

$$\mathcal{L}_+ = \{l : q_l > 0\} \neq \varnothing \qquad \text{and} \qquad \mathcal{L}_- = \{l : q_l < 0\} \neq \varnothing. \tag{21}$$

The sets $\mathcal{F}_0$ and $\mathcal{F}_1(r_n)$ are defined as before, cf. (5), and we use the same notation as in the positive case. Namely, for $T > 0$, we set $\mathcal{N}(T) = \big\{l \in S_F : c_l < T|q_l|\big\}$, $N(T) = |\mathcal{N}(T)|$ and $M(T) = \sum_{l\in\mathcal{N}(T)} q_l^2$.

We point out that, in the case considered in this section, a phenomenon of phase transition occurs: there is a regular case in which the rate is independent of the precise degree of smoothness, and an irregular case where the rate is smoothness-dependent. To be more precise, let $|Q|$ denote the diagonal positive quadratic functional whose coefficients are $|q_l|$ for every $l \in \mathcal{L}$. Let us recall that the minimax rate r_n^* in testing the significance of $|Q|[f]$ (see Remark 5) is determined by

$$n(r_n^*)^2 \asymp M(r_n^{*-2})^{1/2}.$$

In our context, this rate corresponds to the irregular case: if Σ contains functions that are not smooth enough (compared to the difficulty of the problem, that is to say if q_l's are "too large" compared to c_l's), the minimax rate corresponding to Q is the same as for $|Q|$ obtained in previous sections. By contrast, in the regular case, the minimax rate is smoothness-independent and equals $r_n^* = n^{-1/4}$.

4.1. Testing procedure and upper bound on the minimax rate

The testing procedure we use in the present context is of the same type as the one used for nonnegative quadratic functionals. More precisely, for a tuning parameter T_n and for a threshold u, we set $\phi_n(T) = \mathbf{1}_{|U_n(T)|>u}$, where the U-statistic $U_n(T)$ is defined by

$$U_n(T) = \binom{n}{2}^{-1/2} \sum_{1\le i<j\le n} x_i x_j G_T(\mathbf{t}_i, \mathbf{t}_j).$$

with $G_T(\mathbf{t}_1, \mathbf{t}_2) = M(T)^{-1/2} \sum_{l\in\mathcal{N}(T)} q_l \varphi_l(\mathbf{t}_1)\varphi_l(\mathbf{t}_2)$.

Theorem 3. *Let $\gamma \in (0,1)$ be a fixed significance level. Let us denote by $\mathtt{T}_Q[f]$ the linear functional $\mathtt{T}_Q[f] = \sum_{l\in\mathcal{N}(T)} q_l \theta_l[f]\varphi_l$. Assume that $T > 0$ is such that the assumptions*

- **[D1]** *there exists $D_1 > 0$ such that $|\mathcal{N}(T)| \max_{l\in\mathcal{N}(T)} q_l^2 \le D_1 \sum_{l\in\mathcal{N}(T)} q_l^2$,*
- **[D2]** *there exists $D_2 > 0$ such that $\sup_{\mathbf{t}\in\Delta} \sum_{l\in\mathcal{N}(T)} \varphi_l(\mathbf{t})^2 \le D_2|\mathcal{N}(T)|$,*
- **[D3]** *there exists $D_3 > 0$ such that $\sup_{f\in\Sigma} \|f\|_4 \le D_3$,*
- **[D4]** *there exists $D_4 > 0$ such that $\sup_{f\in\Sigma} \|f \cdot \mathtt{T}_Q[f]\|_2 \le D_4$,*

are fulfilled. Set $B_1 = 6 + 12D_1D_2D_3^2 + 6D_1D_2D_3^4$ and $B_2 = 4D_4$. Then, for every

$$u \ge \frac{n}{T\sqrt{2M(T)}} + \gamma^{-1/2}\big(B_1 + B_2 n M(T)^{-1}\big)^{1/2},$$

the type I error is bounded by $\gamma/2$: $\sup_{f\in\mathcal{F}_0} P_f(\phi_n(T) = 1) \le \frac{\gamma}{2}$.

If, in addition,

$$\rho^2 \ge \big[u + \gamma^{-1/2}\big(B_1 + B_2 n M(T)^{-1}\big)^{1/2}\big]\frac{\sqrt{2M(T)}}{n} + \frac{1}{T}$$

then the type II error is also bounded by $\gamma/2$: $\sup_{f\in\mathcal{F}_1(\rho)} P_f(\phi_n(T) = 0) \le \frac{\gamma}{2}$.

As a consequence, if we choose

$$u = (2M(T))^{-1/2}(n/T) + \gamma^{-1/2}\big(B_1 + B_2 n M(T)^{-1}\big)^{1/2}$$

then the cumulative error rate of the test $\phi_n(T)$ is bounded by γ for every alternative $\mathcal{F}_1(\rho)$ such that $\rho^2 \geq 2\sqrt{2}\gamma^{-1/2}n^{-1}\big(B_1 M(T) + B_2 n\big)^{1/2} + 2T^{-1}$.

This theorem provides a nonasymptotic evaluation of the cumulative error rate of the linear U-test based on the array $w_l \propto q_l$ truncated at the level T. In the cases where the constants B_1 and B_2 can be reliably estimated and the function $M(T)$ admits a simple form, it is reasonable to choose the truncation level T by minimizing the expression $2\sqrt{2}\gamma^{-1/2}n^{-1}\big(B_1 M(T) + B_2 n\big)^{1/2} + 2T^{-1}$. By choosing T in such a way, we try to enlarge the set of alternatives for which the cumulative error rate stays below the prescribed level γ. Therefore, the last theorem implies the following non-asymptotic upper bound on the minimax rate of separation:

$$(r^*_{n,\gamma})^2 \leq \inf_{T>0}\Big(\frac{2\sqrt{2}\big(B_1 M(T) + B_2 n\big)^{1/2}}{n\gamma^{1/2}} + \frac{2}{T}\Big). \tag{22}$$

This non-asymptotic bound clearly shows the presence of two asymptotic regimes. The first one corresponds to the case where n is much larger than $M(T^*)$, whereas the second regime corresponds to $n = o(M(T^*))$. Here, T^* is the minimizer of the bound on ρ^2 obtained in the theorem above. The next corollary exhibits the rates of separation in these two different regimes.

Corollary 1. *Assume that the arrays* **q** *and* **c** *are such that $M(\alpha T) \asymp_{T\to\infty} M(T)$ for every $\alpha > 0$. Let T_n^0 be any sequence of positive numbers satisfying $T_n^0\sqrt{M(T_n^0)} \asymp n$. If for the sequence $T_n = T_n^0 \wedge n^{1/2}$ all the assumptions of Theorem 3 are satisfied, then for some $C > 0$ the linear U-test $\phi_n(T)$ based on the threshold $T = T_n$ satisfies*

$$\gamma_n(\mathcal{F}_0, \mathcal{F}_1(CT_n^{-1/2}), \phi_n) \leq \gamma.$$

Thus, the rate of convergence is $r_n^* = (T_n^0)^{-1/2}$ if $T_n^0 = o(n^{1/2})$ and $r_n^* = n^{-1/4}$ otherwise.

Remark 10. Condition **[D4]** of Theorem 3 is more obscure than the other assumptions of theorem. Clearly, it imposes additional smoothness constraints on the function f. Using the Cauchy-Schwarz inequality, one can easily check that either one of the assumptions **[D4-1]** and **[D4-2]** below is sufficient for **[D4]**:

[D4-1] For some D_5 and D_6, $\sup_{f\in\Sigma}\|f\|_\infty \leq D_5$ and $\max_{l\in\mathcal{N}(T)}|q_l/c_l| \leq D_6$.
[D4-2] For some constant D_4', $\sup_{f\in\Sigma}\|\mathtt{T}_Q[f]\|_4 \leq D_4'$.

4.2. Lower bound on the minimax rate

We will show in this subsection that the asymptotic rate of separation provided by Corollary 1 is unimprovable, in the sense that there is no testing procedure

having a faster separation rate. To this end, for every $a \in \{-,+\}$ we set $M_a(T) = \sum_{l\in\mathcal{L}_a\cap\mathcal{N}(T)} q_l^2$, $N_a(T) = |\mathcal{L}_a \cap \mathcal{N}(T)|$, $M^*(T) = M_+(T) \vee M_-(T)$, and

$$N^*(T) = N_+(T)\mathbf{1}_{\{M_+(T)>M_-(T)\}} + N_-(T)\mathbf{1}_{\{M_+(T)\le M_-(T)\}}.$$

Theorem 4. *Let us consider the problem of testing $H_0 : f \in \mathcal{F}_0$ against $H_1 : f \in \mathcal{F}_1(\rho)$, where $\mathcal{F}_0$ and $\mathcal{F}_1$ are defined by (5) and*

$$\Sigma_L = \Big\{ f = \sum\nolimits_{l\in\mathcal{L}} \theta_l[f]\varphi_l : \ \sum\nolimits_{l\in\mathcal{L}} c_l\theta_l[f]^2 \le 1, \ \|f\|_4 \vee \|f\cdot \mathtt{T}_Q[f]\|_2 \le L \Big\}.$$

Assume that the sets $\mathcal{L}_+$ and $\mathcal{L}_-$ defined by (21) are both nonempty and that ξ_i's are Gaussian. The following assertions are true.

1. *For every $\gamma < 1/4$ there exists a constant $C > 0$ such that the inequality $\liminf_{n\to\infty} \gamma_n(\mathcal{F}_0, \mathcal{F}_1(Cn^{-1/4})) > \gamma$ is satisfied.*
2. *Let T_n^0 be a sequence of reals such that $4T_n^0\sqrt{M(T_n^0)} \ge nz_{1-\gamma/2}^{-1}$ as $n\to\infty$. If the assumptions* **[D1]** *(cf. Theorem 3) and*

   [D5] *$N^*(T_n^0)\to\infty$ so that $N^*(T_n^0)\log N^*(T_n^0) = o(n)$,*

   [D6] *$\exists D_6 > 0$ such that $\sup_{\mathbf{t}\in\Delta}\sum_{l\in\mathcal{N}^*(T_n^0)}\varphi_l(\mathbf{t})^2 \le D_6 N^*(T_n^0)$,*

   are fulfilled, then there exists a constant $C > 0$ such that the inequality $\liminf_{n\to\infty}\gamma_n\big(\mathcal{F}_0, \mathcal{F}_1\big(C(T_n^0)^{-1/2}\big)\big) \ge \gamma$ is satisfied.

Corollary 2. *Combining the two assertions of this theorem, we get that the minimax rate of separation r_n^* is lower bounded by $n^{-1/4} \vee (T_n^0)^{-1/2} = (n^{1/2} \wedge T_n^0)^{-1/2} = T_n^{-1/2}$. Thus, if the conditions of Theorems 3 and 4 are satisfied, then the minimax rate of separation is given by $r_n^* = T_n^{-1/2}$, where $T_n = n^{1/2} \wedge T_n^0$ and T_n^0 is determined from the relation $T_n^0 M(T_n^0)^{1/2} \asymp n$.*

4.3. Testing equality of norms

As an application of the testing methodology developed in this section, we consider the problem of testing the equality of norms of two functions observed in noisy environment. More precisely, let us consider the following two-sample problem: for $i = 1,\dots,n$ we observe $(x_{1,i}, \mathbf{t}_{1,i})$ and $(x_{2,i}, \mathbf{t}_{2,i})$ such that

$$x_{s,i} = g_s(\mathbf{t}_{s,i}) + \xi_{s,i}, \qquad i = 1,\dots,n; \quad s = 1,2,$$

where $\mathbf{t}_{s,i}$'s are independent random vectors drawn from the uniform distribution over $[0,1]^d$. Furthermore, we assume that $\xi_{s,i}$'s are i.i.d. such that $E(\xi_{s,i}|\{\mathbf{t}_{s,j}\}) = 0$, $E(\xi_{s,i}^2|\{\mathbf{t}_{s,j}\}) = 1$ and, for some $C_\xi < \infty$, $E(\xi_{s,i}^4|\{\mathbf{t}_{s,j}\}) \le C_\xi$ almost surely.

Assuming that both g_1 and g_2 belong to a smoothness class Σ, we wish to test the hypothesis

$$H_0 : \|g_1\|_{W_2^\alpha} = \|g_2\|_{W_2^\alpha}, \qquad \text{against} \qquad H_1 : \big|\|g_1\|_{W_2^\alpha}^2 - \|g_2\|_{W_2^\alpha}^2\big| \ge \rho^2,$$

where for any function g we denoted by $\|g\|_{W_2^{\boldsymbol{\alpha}}}$ the (anisotropic) Sobolev norm of order $\boldsymbol{\alpha} \in \mathbb{R}_+^d$ (the precise definition is given below). It can be useful to perform such a test prior to using a shifted curve model in the context of curve registration (Dalalyan and Collier, 2012, Collier, 2012). Indeed, if there exists $\boldsymbol{\tau} \in [0,1]^d$ such that $g_1(\mathbf{t}) = g_2(\mathbf{t} - \boldsymbol{\tau})$ for every $\mathbf{t} \in [0,1]^d$ and the function g_1 is one-periodic, then necessarily $\|g_1\|_{W_2^{\boldsymbol{\alpha}}} = \|g_2\|_{W_2^{\boldsymbol{\alpha}}}$ for any $\boldsymbol{\alpha}$. Thus, the rejection of the null hypothesis implies the inadequacy of the shifted curve model. In order to show how this type of test can be derived from the framework presented in the previous subsections, let us consider the case of a Sobolev ellipsoid Σ.

Let the family $\{\psi_m\}_{l \in \mathcal{M}}$ be an orthonormal basis of the subspace $L_{2,c}([0,1]^d)$ of $L_2([0,1]^d)$ consisting of all the functions orthogonal to the constant function. We will assume that both g_1 and g_2 are centered (this implies that they are orthogonal to the constant function as well). The Fourier coefficients of a function g w.r.t. a basis $\{\psi_m\}$ will be denoted by $\theta_m^{\psi}[g]$. We assume that for some array $\mathbf{c}$ and some constant $L > 0$ it holds that

$$g_s \in \Sigma_L^0 = \Big\{ g \in L_{2,c}([0,1]^d) : \sum_{m \in \mathcal{M}} c_m \theta_m^{\psi}[g]^2 \le 1,\ \|g\|_4 \le L \Big\},\ \forall s \in \{1,2\}.$$

Assume now that we wish to test H_0 against H_1, where

$$H_0 : \sum_{m \in \mathcal{M}} q_m \theta_m^{\psi}[g_1]^2 = \sum_{m \in \mathcal{M}} q_m \theta_m^{\psi}[g_2]^2,$$
$$H_1 : \Big| \sum_{m \in \mathcal{M}} q_m (\theta_m^{\psi}[g_1]^2 - \theta_m^{\psi}[g_2]^2) \Big| \ge \rho^2,$$

and $\mathbf{q} = \{q_m\}$ is a given array. In order to show that this problem can be solved within the framework of the previous subsections, we introduce the functional set

$$\Sigma_L = \Big\{ f : [0,1]^{2d} \to \mathbb{R} : f(\mathbf{t}) = g_1(\mathbf{t}_1) + g_2(\mathbf{t}_2) \text{ with } g_1, g_2 \in \Sigma_L^0 \Big\}.$$

Setting $\mathcal{L} = \mathcal{M} \times \{1,2\}$ and for $l = (m,s) \in \mathcal{M} \times \{1,2\}$

$$\varphi_l(\mathbf{t}_1, \mathbf{t}_2) = \psi_m(\mathbf{t}_s), \qquad \text{for all} \qquad \mathbf{t} = (\mathbf{t}_1, \mathbf{t}_2) \in [0,1]^d \times [0,1]^d,$$

we get an orthonormal basis of Σ_L. Clearly, for a function $f \in \Sigma_L$, we have $\theta_l^{\varphi}[f] = \theta_{m,s}^{\varphi}[f] = \theta_m^{\psi}[g_s]$. This implies that Σ_L is included in the set $\Sigma_L^2 = \{f : \sum_{(m,s)} c_m \theta_{m,s}^{\varphi}[f]^2 \le 2;\ \|f\|_4 \le 2L\}$ and contains the set $\Sigma_L^1 = \{f : \sum_{(m,s)} c_m \theta_{m,s}^{\varphi}[f]^2 \le 1, \|f\|_4 \le L\}$. Therefore, for studying the rate of separation of a testing procedure we can assume that $f \in \Sigma_L^2$, whereas for establishing lower bounds on the minimax rate of separation we can use the relation $\Sigma_L^1 \subset \Sigma_L$. In both cases, this perfectly matches the framework of the previous subsections.

We give a concrete example by setting $\mathcal{M} = \mathbb{Z}^d$ and choosing as $\{\psi_{\boldsymbol{m}}\}$ the Fourier basis in dimension d. Similarly to the example in Subsection 3.3, we focus on anisotropic Sobolev smoothness classes defined via coefficients

$$c_{\boldsymbol{m}} = \sum_{j=1}^{d} (2\pi m_j)^{2\sigma_j}, \qquad \boldsymbol{m} \in \mathbb{Z}^d,$$

for some $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_d) \in \mathbb{R}_+^d$. As it was done previously, $\delta = \sum_{j=1}^d \alpha_j/\sigma_j$ and $\bar{\sigma}$ stands for the harmonic mean of σ_j's: $\bar{\sigma} = \left(\frac{1}{d}\sum_{j=1}^d \sigma_j^{-1}\right)^{-1}$. We still assume that $\delta < 1$ and $\bar{\sigma} > d/4$. To test the equality of Sobolev norms, we introduce the coefficients $q_{\boldsymbol{l}}$, $\boldsymbol{l} = (\boldsymbol{m}, s) \in \mathbb{Z}^d \times \{1, 2\}$, of the quadratic functional Q:

$$q_{\boldsymbol{m},s} = (-1)^s \prod_{j=1}^d (2\pi m_j)^{2\alpha_j}, \qquad (\boldsymbol{m}, s) \in \mathbb{Z}^d \times \{1, 2\}.$$

Theorems 3 and 4, as well as the computations done in the proof of Proposition 4, imply that the minimax rate of separation in the problem described above is: $r_n^* = n^{-\frac{2(1-\delta)\bar{\sigma}}{4\bar{\sigma}+d} \wedge \frac{1}{4}}$. It is interesting to note that if $\delta \geq 1/2$ then we are in the irregular regime irrespectively of the value of $\bar{\sigma}$ and, therefore, the rate of separation is strictly slower than the rate $n^{-1/4}$.

5. Conclusion and outlook

We have presented a statistical analysis of the problem of testing the significance of the value $Q[f]$ for a quadratic functional Q of a regression function f. While the overwhelming majority of previous research focused on the case of a function f observed at any point in Gaussian white noise, we have considered here the more realistic setting when the observations are noisy values of f at a finite number of points uniformly randomly drawn from $[0, 1]^d$. Furthermore, we have explored not only the case of positive semi-definite functional Q but also the situation when Q is neither positive nor negative semi-definite. In the first situation we have established asymptotic results providing the minimax rates of separation along with the sharp constants. In the second case, the analysis we have carried out is nonasymptotic and leads to the asymptotically minimax rate of separation, which exhibits two different regimes: the regular and the irregular regimes. Another distinctive feature of our approach is that we have put the emphasis on the multidimensional setting $d > 1$, even if at this stage we have not tackled the problem of increasingly high dimensionality: $d = d_n \to \infty$ as the sample size n tends to infinity.

The results we have obtained are closely related to those of estimating quadratic functionals. While the presence of such a relation is not surprising in itself, the actual nature of the relation uncovers some interesting new phenomena. In fact, the test statistic used in our work is a properly normalized estimator of the quadratic functional $Q[f]$, which is constructed following the classical approach of weighted squared linear functional estimation (cf., for instance, Donoho and Nussbaum (1990)). Usually, the proper choice of the shrinkage weights and the resulting rates of convergence differ in the problem of hypothesis testing and in the problem of estimation. This is why the well-known "elbow" effect (phase transition) in estimating quadratic functionals disappears when the problem of hypotheses testing is considered for $Q[f] = \int_{[0,1]} f^2(t)\,dt$. Interestingly, the results of Section 4 show that this difference between the rates of convergence is erased when the quadratic functional $Q[f]$ is neither positive nor negative.

In fact, the rates of separation we have obtained in this case coincide with the square-root of the rates of estimation (Donoho and Nussbaum, 1990, Fan, 1991). Therefore, the "elbow" effect is present in this problem of hypotheses testing. More interestingly, the rates of separation we obtained in the case of positive semi-definite functionals Q coincide with the rates of estimation of the functional $\sqrt{Q[f]}$ in the case $Q[f] = \int_{[0,1]} f^2(t)\,dt$, at least in the Gaussian white noise model (Lepski et al., 1999). An intriguing question worth of being further explored is whether this analogy extends to the model of regression with random design and general positive semi-definite functionals $Q[f]$.

Several relevant problems remained out of scope of the present paper. Most important ones are the possibility of extending our results to the case of nondiagonal functionals $Q[f]$ and the attainability of the obtained rates of separation by adaptive tests. More specifically, in some applications such as in deconvolution it may be more realistic to assume that the functional basis in which the smoothness of f is expressed does not coincide with the basis of the singular vectors of (the bilinear operator underlying) Q. This means that $Q[f]$ will be of the form $Q[f] = \sum_{l,l' \in \mathcal{L}} q_{l,l'} \theta_l[f]\theta_{l'}[f]$ rather than $Q[f] = \sum_{l \in \mathcal{L}} q_l \theta_l[f]^2$. Furthermore, it would be more reasonable to replace the assumption $\sum_{l \in \mathcal{L}} c_l \theta_l[f]^2 \le 1$ with some *known* array $\mathbf{c} = \{c_l\}_{l \in \mathcal{L}}$ by the assumption $\sum_{l \in \mathcal{L}} c_l(\mu^*)\theta_l[f]^2 \le 1$, where $\mathbf{c}(\mu) = \{c_l(\mu)\}_{l \in \mathcal{L}}$ is a collection of arrays such that the mapping $\mu \mapsto \mathbf{c}(\mu)$ is known but the precise value μ^* for which the smoothness constraint is valid is unknown. In the light of the previous discussion, it seems natural to study these two extensions (nondiagonal Q and adaptation to the smoothness class) by considering the problem of testing and the problem of estimating functionals in a joint framework. In particular, any progress in establishing upper bounds for estimators of $Q[f]$ or $\sqrt{|Q[f]|}$ will straightforwardly lead to upper bounds for the rates of separation. Quite surprisingly, these problems of estimation received little attention in the context of nonparametric regression[3]. They constitute interesting avenues for future research.

Appendix A: Proofs of results stated in Section 2

A.1. Proof of Proposition 1

Throughout the proof, the terms $o(1)$, $O(1)$ and the equivalences are uniform over Σ. Let $\mathcal{L}(\mathbf{w}_n)$ be the support of $\mathbf{w}_n$. $E_f^{\mathcal{D}_2}$ will denote the conditional expectation with respect to $\mathcal{D}_2$. We define

$$h_n[f, \mathbf{w}_n] = \Big(\frac{m(m-1)}{2}\Big)^{1/2} \sum\nolimits_{l \in \mathcal{L}(\mathbf{w}_n)} w_{l,n}\theta_l^2[f], \tag{23}$$

$$G_n(\mathbf{t}_1, \mathbf{t}_2) = \sum_{l \in \mathcal{L}(\mathbf{w}_n)} w_{l,n}\varphi_l(\mathbf{t}_1)\varphi_l(\mathbf{t}_2). \tag{24}$$

[3] Minimax and adaptive estimation for quadratic functionals is well studied in the case of Gaussian white noise model. However, these results do not always carry over the regression model as noticed by Efromovich (2003).

This allows us to rewrite the U-statistic U_n in the form $U_n = U_{n,0} + U_{n,1} + U_{n,2}$ where

$$U_{n,k} = \Big(\frac{2}{m(m-1)}\Big)^{1/2} \sum_{1\le i<j\le m} K_{n,k}(\tilde{\mathbf{z}}_i, \tilde{\mathbf{z}}_j), \qquad k = 0, 1, 2,$$

are U-statistics with the kernels

$$K_{n,0}(\tilde{\mathbf{z}}_1, \tilde{\mathbf{z}}_2) = \xi_1\xi_2 G_n(\mathbf{t}_1, \mathbf{t}_2), \tag{25}$$

$$K_{n,1}(\tilde{\mathbf{z}}_1, \tilde{\mathbf{z}}_2) = \Big[\xi_1\big(f - \widehat{\Pi f}_n\big)(\mathbf{t}_2) + \xi_2\big(f - \widehat{\Pi f}_n\big)(\mathbf{t}_1)\Big] G_n(\mathbf{t}_1, \mathbf{t}_2), \tag{26}$$

$$K_{n,2}(\tilde{\mathbf{z}}_1, \tilde{\mathbf{z}}_2) = \big(f - \widehat{\Pi f}_n\big)(\mathbf{t}_1)\big(f - \widehat{\Pi f}_n\big)(\mathbf{t}_2) G_n(\mathbf{t}_1, \mathbf{t}_2). \tag{27}$$

To prove Proposition 1 and the subsequent results, we need two auxiliary lemmas. For an integer $p > 0$, let us define

$$\mathcal{M}_p = \|\Pi_{S_F} f\|_p^p + E_f\big[\|\Pi f - \widehat{\Pi f}_n\|_p^p$$

Lemma 3. *Let $\mathbf{w}_n = (w_{l,n})_{l\in\mathcal{L}}$ be a family of positive numbers containing only a finite number of nonzero entries and such that $\sum_{l\in\mathcal{L}} w_{l,n}^2 = 1$. Let $\mathcal{L}(\mathbf{w}_n)$ be the support of $\mathbf{w}_n$. Then the expectation of the U-statistic U_n is given by:*

$$E_f[U_n] = E_f[U_{n,2}] = h_n[f, \mathbf{w}_n],$$

whereas for the variances it holds

$$E_f[U_{n,0}^2] = 1,$$

$$E_f[U_{n,1}^2] \le 2\|\mathbf{w}_n\|_\infty^2 \Big(\sup_{\mathbf{t}\in\Delta} \sum_{l\in\mathcal{L}(\mathbf{w}_n)} \varphi_l^2(\mathbf{t})\Big)\mathcal{M}_2, \tag{28}$$

$$\begin{aligned} Var_f[U_{n,2}] \le{}& 8\|\mathbf{w}_n\|_\infty^2 \Big(\sup_{\mathbf{t}\in\Delta} \sum_{l\in\mathcal{L}(\mathbf{w}_n)} \varphi_l^2(\mathbf{t})\Big)\mathcal{M}_4 \\ &+ 8h_n[f, \mathbf{w}_n]\|\mathbf{w}_n\|_\infty \Big(\sup_{\mathbf{t}\in\Delta} \sum_{l\in\mathcal{L}(\mathbf{w}_n)} \varphi_l^2(\mathbf{t})\Big)^{1/2} \mathcal{M}_4^{1/2}. \end{aligned} \tag{29}$$

Proof. It is clear that $E_f U_{n,0} = E_f U_{n,1} = 0$, while

$$E_f[U_{n,2}] = \Big(\frac{m(m-1)}{2}\Big)^{1/2} E_f[K_{n,2}(\tilde{\mathbf{z}}_1, \tilde{\mathbf{z}}_2)]$$

with

$$E_f[K_{n,2}(\tilde{\mathbf{z}}_1, \tilde{\mathbf{z}}_2)] = E_f\Big[\sum\nolimits_{l\in\mathcal{L}(\mathbf{w}_n)} w_{l,n}\Big(\int \big(f(\mathbf{t}) - \widehat{\Pi f}_n(\mathbf{t})\big)\varphi_l(\mathbf{t})d\mathbf{t}\Big)^2\Big].$$

As $\widehat{\Pi f}_n \in \operatorname{span}\big(\{\varphi_l\}_{l\in S_F^c}\big)$, we have $\int \widehat{\Pi f}_n \varphi_l = 0$ for all $l \in S_F$. Therefore

$$E_f[U_{n,2}] = \Big(\frac{m(m-1)}{2}\Big)^{1/2} \sum\nolimits_{l\in\mathcal{L}(\mathbf{w}_n)} w_{l,n}\theta_l^2[f] = h_n[f, \mathbf{w}_n].$$

Now, let us evaluate the variances. Since ξ_is are non correlated zero-mean random variables with variance one, and φ_l's are orthonormal, it holds that $E_f[U_{n,0}^2] = E_f[G_n(\mathbf{t}_1,\mathbf{t}_2)^2] = \sum_l w_{l,n}^2 = 1$. For $U_{n,1}$, we have

$$\mathrm{Var}_f[U_{n,1}] = E_f[U_{n,1}^2] = E_f E_f^{\mathcal{D}_2}[K_{n,1}^2(\tilde{\mathbf{z}}_1,\tilde{\mathbf{z}}_2)].$$

Using the definition of $G_n(\mathbf{t}_1,\mathbf{t}_2)$, we get

$$\begin{aligned} E_f^{\mathcal{D}_2}[K_{n,1}^2(\tilde{\mathbf{z}}_1,\tilde{\mathbf{z}}_2)] &= 2\int_\Delta\int_\Delta \big(f-\widehat{\Pi f}_n\big)^2(\mathbf{t}_1)G_n^2(\mathbf{t}_1,\mathbf{t}_2)d\mathbf{t}_1 d\mathbf{t}_2 \\ &= 2\int_\Delta \big(f-\widehat{\Pi f}_n\big)^2(\mathbf{t}_1)\sum_{l\in\mathcal{L}(\mathbf{w}_n)} w_{l,n}^2\varphi_l^2(\mathbf{t}_1)d\mathbf{t}_1 \\ &\le 2\Big(\max_{l\in\mathcal{L}(\mathbf{w}_n)} w_{l,n}^2\Big)\Big(\sup_{\mathbf{t}\in\Delta}\sum\nolimits_{l\in\mathcal{L}(\mathbf{w}_n)}\varphi_l^2(\mathbf{t})\Big)\|f-\widehat{\Pi f}_n\|_2^2. \end{aligned}$$

Then, the Pythagoras theorem yields

$$E_f\|f-\widehat{\Pi f}_n\|_2^2 = \|f-\Pi f\|_2^2 + E_f\|\Pi f-\widehat{\Pi f}_n\|_2^2 = \mathcal{M}_2.$$

This completes the proof (28). As for the variance of $U_{n,2}$, we have

$$\mathrm{Var}_f[U_{n,2}] = E_f E_f^{\mathcal{D}_2}[U_{n,2}^2] - (E_f[U_{n,2}])^2 = A_{n,1}+A_{n,2}+A_{n,3},$$

where

$$\begin{aligned} A_{n,1} &= E_f\iint \big(f-\widehat{\Pi f}_n\big)^2(\mathbf{t}_1)\big(f-\widehat{\Pi f}_n\big)^2(\mathbf{t}_2)G_n^2(\mathbf{t}_1,\mathbf{t}_2)d\mathbf{t}_1 d\mathbf{t}_2, \\ A_{n,2} &= \frac{4\binom{m}{3}}{m(m-1)}E_f\iiint \big(f-\widehat{\Pi f}_n\big)^2(\mathbf{t}_1)\big(f-\widehat{\Pi f}_n\big)(\mathbf{t}_2)G_n(\mathbf{t}_1,\mathbf{t}_2) \\ &\qquad\qquad \times\big(f-\widehat{\Pi f}_n\big)(\mathbf{t}_3)G_n(\mathbf{t}_1,\mathbf{t}_3)d\mathbf{t}_1 d\mathbf{t}_2 d\mathbf{t}_3, \end{aligned}$$

and

$$A_{n,3} = \frac{4\binom{m}{4}}{m(m-1)}E_f\Big\{\iint f(\mathbf{t}_1)f(\mathbf{t}_2)G_n(\mathbf{t}_1,\mathbf{t}_2)d\mathbf{t}_1 d\mathbf{t}_2\Big\}^2 - (E_f U_{n,2})^2.$$

Let us bound the first term $A_{n,1}$:

$$A_{n,1} = E_f\sum_{l,l'\in\mathcal{L}(\mathbf{w}_n)} w_{l,n}w_{l',n}\Big(\int(f-\widehat{\Pi f}_n)^2(\mathbf{t})\varphi_l(\mathbf{t})\varphi_{l'}(\mathbf{t})\,d\mathbf{t}\Big)^2.$$

Now, in view of Bessel's inequality,

$$\begin{aligned} A_{n,1} &\le \max_{l\in\mathcal{L}(\mathbf{w}_n)} w_{l,n}^2 E_{f,B}\sum_{l\in\mathcal{L}(\mathbf{w}_n)}\int(f-\widehat{\Pi f}_n)^4(\mathbf{t})\varphi_l^2(\mathbf{t})\,d\mathbf{t} \\ &\le \Big(\max_{l\in\mathcal{L}(\mathbf{w}_n)} w_{l,n}^2\Big)\Big(\sup_{\mathbf{t}\in\Delta}\sum\nolimits_{l\in\mathcal{L}(\mathbf{w}_n)}\varphi_l^2(\mathbf{t})\Big)E_f\big[\big\|f-\widehat{\Pi f}_n\big\|_4^4\big], \end{aligned}$$

and the expression inside the last expectation can be bounded using the inequality $\|f-\widehat{\Pi f}_n\|_4^4 \le 8(\|\Pi_{S_F} f\|_4^4 + \|\Pi f - \widehat{\Pi f}_n\|_4^4)$.

The term $A_{n,2}$ can be dealt with similarly. Using the Cauchy-Schwarz inequality,

$$\begin{aligned} A_{n,2} &= \frac{4\binom{m}{3}}{m(m-1)} \sum_{l,l'\in\mathcal{L}(\mathbf{w}_n)} w_{l,n}w_{l',n}\theta_l[f]\theta_{l'}[f] E_f\Big\{\int (f-\widehat{\Pi f}_n)^2 \varphi_l\varphi_{l'}\Big\} \\ &\le \binom{m}{2}^{1/2}\Big(\sum_l w_{l,n}^2\theta_l[f]^2\Big)\Big(\sum_{l,l'\in\mathcal{L}(\mathbf{w}_n)}\Big\{\int E_f\big[(f-\widehat{\Pi f}_n)^2\big]\varphi_l\varphi_{l'}\Big\}^2\Big)^{1/2} \\ &\le \Big(\max_{l\in\mathcal{L}(\mathbf{w}_n)} w_{l,n}\Big) h_n[f,\mathbf{w}_n]\Big(\sum_{l,l'\in\mathcal{L}(\mathbf{w}_n)} E_f\Big\{\int (f-\widehat{\Pi f}_n)^2\varphi_l\varphi_{l'}\Big\}^2\Big)^{1/2}. \end{aligned}$$

By virtue of the Bessel inequality, it holds that

$$\begin{aligned} A_{n,2} &\le \Big(\max_{l\in\mathcal{L}(\mathbf{w}_n)} w_{l,n}\Big) h_n[f,\mathbf{w}_n]\Big(\sum_{l\in\mathcal{L}(\mathbf{w}_n)}\int E_f\big[(f-\widehat{\Pi f}_n)^4(\mathbf{t})\big]\varphi_l^2(\mathbf{t})\,d\mathbf{t}\Big)^{1/2} \\ &\le \Big(\max_{l\in\mathcal{L}(\mathbf{w}_n)} w_{l,n}\Big) h_n[f,\mathbf{w}_n]\Big(\sup_{\mathbf{t}\in\Delta}\sum_{l\in\mathcal{L}(\mathbf{w}_n)}\varphi_l^2(\mathbf{t})\Big)^{1/2}\big(E_f[\|f-\widehat{\Pi f}_n\|_4^4]\big)^{1/2}. \end{aligned}$$

The last expectation can be bounded in the same way as we did several lines above for the term $A_{n,1}$. The last term $A_{n,3}$ is actually negative

$$A_{n,3} = \frac{4\binom{m}{4}}{m(m-1)}\Big(\sum_{l\in\mathcal{L}(\mathbf{w}_n)} w_{l,n}\theta_l^2\Big)^2 - \frac{m(m-1)}{2}\Big(\sum_{l\in\mathcal{L}(\mathbf{w}_n)} w_{l,n}\theta_l^2\Big)^2 \le 0.$$

Combining all these estimates, we get (29). □

Lemma 4. *Let $\mathbf{w}_n = (w_{l,n})_{l\in\mathcal{L}}$ be a family of positive numbers containing only a finite number of nonzero entries and such that $\sum_{l\in\mathcal{L}} w_{l,n}^2 = 1$. Assume that the random variable ξ_1 has finite fourth moment: $E_f[\xi_1^4] < \infty$. If, as $n\to\infty$,*

$$\|\mathbf{w}_n\|_\infty = o(1) \qquad \text{and} \qquad \|\mathbf{w}_n\|_\infty^2\Big(\sup_{\mathbf{t}\in\Delta}\sum\nolimits_{l\in\mathcal{L}(\mathbf{w}_n)}\varphi_l(\mathbf{t})^2\Big)^2 = o(n), \tag{30}$$

then $U_{n,0}$ is asymptotically Gaussian $\mathcal{N}(0,1)$.

Proof. This result is an immediate consequence of (Hall, 1984, Theorem 1). □

With these tools at hand, we are now in a position to establish the asymptotic normality of the U-statistic U_n which leads to an evaluation of the type I error of the U-test. Let us recall that, for $f\in\mathcal{F}_0$, it holds $Q[f] = \sum q_l\theta_l[f]^2 = 0$ and, therefore, $\theta_l[f] = 0$ for all $l\in S_F = \{l : q_l \ne 0\}$. Hence, for every $f\in\mathcal{F}_0$, $h_n[f,\mathbf{w}_n] = 0$ and $\Pi_{S_F} f = 0$. So, it follows from Lemma 3 that under the assumptions of the proposition, the convergences $E_f[U_{n,1}^2]\to 0$ and $E_f[U_{n,2}^2]\to 0$ hold true uniformly in $f\in\mathcal{F}_0$. This implies that $U_{n,1}$ and $U_{n,2}$ tend to

zero in P_f-probability, uniformly in $f \in \mathcal{F}_0$. On the other hand, according to Lemma 4, $U_{n,0} \to \mathcal{N}(0,1)$ in distribution. The claim of the proposition follows from Slutsky's lemma.

A.2. Proof of Proposition 2

We first note that for every $\bar{h} > 0$ it holds

$$\sup_{f\in\mathcal{F}_1(\rho)} P_f(U_n \le u) = \Big(\sup_{\substack{f\in\mathcal{F}_1(\rho)\\ h_n[f,\mathbf{w}_n]>\bar{h}}} P_f(U_n \le u)\Big) \bigvee \Big(\sup_{\substack{f\in\mathcal{F}_1(\rho)\\ h_n[f,\mathbf{w}_n]\le\bar{h}}} P_f(U_n \le u)\Big). \tag{31}$$

The value of $\bar{h}$ will be made precise later in the proof. Assume merely by now that $\bar{h} > 2(1+u)$. Then,

$$\begin{aligned}\sup_{\substack{f\in\mathcal{F}_1(\rho);\\ h_n[f,\mathbf{w}_n]>\bar{h}}} P_f(U_n \le u) &\le \sup_{f\in\Sigma; h_n[f,\mathbf{w}_n]>\bar{h}} \frac{\mathrm{Var}_f[U_n]}{\big(E_f[U_n]-u\big)^2}\\ &= \sup_{f\in\Sigma; h_n[f,\mathbf{w}_n]>\bar{h}} \frac{\mathrm{Var}_f[U_n]}{\big(h_n[f,\mathbf{w}_n]-u\big)^2}.\end{aligned}$$

Using the conditions of the proposition and the inequalities of Lemma 3, we get that for some constants C, C' independent of $\bar{h}$,

$$\begin{aligned}\sup_{\substack{f\in\mathcal{F}_1(\rho)\\ h_n[f,\mathbf{w}_n]>\bar{h}}} P_f(U_n \le u) &\le \sup_{f\in\Sigma; h_n[f,\mathbf{w}_n]>\bar{h}} \frac{C(1+h_n[f,\mathbf{w}_n])}{\big(h_n[f,\mathbf{w}_n]-u\big)^2}\\ &\le C\frac{1+\bar{h}}{\big(\bar{h}-u\big)^2} \le C'\bar{h}^{-1}.\end{aligned} \tag{32}$$

Let us switch to the second sup in (31). Let $\delta_n > 0$ be a sequence tending to zero. One readily checks that

$$\begin{aligned}P_f(U_n \le u) &= P_f(h_n[f,\mathbf{w}_n] + U_{n,0} + U_{n,1} + (U_{n,2} - h_n[f,\mathbf{w}_n]) \le u)\\ &\le P_f(h_n[f,\mathbf{w}_n] + U_{n,0} \le u + \delta_n)\\ &\quad + P_f(-U_{n,1} - (U_{n,2} - h_n[f,\mathbf{w}_n]) \ge \delta_n)\\ &\le F_{U_{0,n}}(u - h_n[f,\mathbf{w}_n] + \delta_n) + \frac{2\mathrm{Var}_f(U_{n,1}) + 2\mathrm{Var}_f(U_{n,2})}{\delta_n^2},\end{aligned} \tag{33}$$

where $F_{U_{0,n}}(\cdot)$ is the c.d.f. of $U_{0,n}$. On the one hand, we know from Lemma 4 that $U_{n,0}$ converges in distribution to $\mathcal{N}(0,1)$. This entails that $F_{U_{0,n}}$ converges uniformly over $\mathbb{R}$ to Φ. Therefore,

$$\begin{aligned}F_{U_{0,n}}(u - h_n[f,\mathbf{w}_n] + \delta_n) &= \Phi(u - h_n[f,\mathbf{w}_n] + \delta_n) + o(1)\\ &= \Phi(u - h_n[f,\mathbf{w}_n]) + o(1) + \delta_n O(1).\end{aligned}$$

On the other hand, in view of Lemma 3, $\mathrm{Var}_f(U_{n,1}) + \mathrm{Var}_f(U_{n,2}) = O(\|\Pi_{S_F} f\|_4^4 + \|\Pi_{S_F} f\|_2^2)$.

Then we have,

$$\|\Pi_{S_F} f\|_2^2 = \sum_{l\in S_F} \theta_l^2 \le \frac{1}{\zeta_n} \sum_{w_{l,n}\ge\zeta_n} w_{l,n}\theta_l^2 + \sum_{w_{l,n}<\zeta_n} \theta_l^2$$
$$\le \frac{\sqrt{2}\, h_n[f,\mathbf{w}_n]}{\zeta_n(m-1)} + \sup_{l\in S_F: w_{l,n}<\zeta_n} c_l^{-1}.$$

Applying Hölder's inequality we get

$$\|\Pi_{S_F} f\|_4^4 \le \|\Pi_{S_F} f\|_2^{2(p-4)/(p-2)} \|\Pi_{S_F} f\|_p^{2p/(p-2)}.$$

Therefore, we have

$$\sup_{\substack{f\in\mathcal{F}_1(\rho)\\ h_n[f,\mathbf{w}_n]\le\bar{h}}} P_f(U_n \le u) \le \sup_{f\in\mathcal{F}_1(\rho)} \Phi(u - h_n[f,\mathbf{w}_n]) + o(1)$$
$$+ \delta_n O(1) + \frac{o(1)(\bar{h}^{(p-4)/(p-2)} + \bar{h})}{\delta_n^2}.$$

Choosing $\bar{h}$ large enough and then making δ_n tend to zero sufficiently slowly we get the desired result.

A.3. Proof of Proposition 3

Using Kneser's minimax theorem for bilinear forms (Kneser, 1952), we can interchange the sup and the inf as follows:

$$\sup_{\substack{\mathbf{w}\in\mathbb{R}_+^{\mathcal{L}}\\ \|\mathbf{w}\|_2=1}} \inf_{\substack{\mathbf{v}\in\mathbb{R}_+^{\mathcal{L}}\\ \langle\mathbf{v},\mathbf{c}\rangle\le 1, \langle\mathbf{v},\mathbf{q}\rangle\ge\rho^2}} \langle\mathbf{w},\mathbf{v}\rangle = \inf_{\substack{\mathbf{v}\in\mathbb{R}_+^{\mathcal{L}}\\ \langle\mathbf{v},\mathbf{c}\rangle\le 1, \langle\mathbf{v},\mathbf{q}\rangle\ge\rho^2}} \sup_{\substack{\mathbf{w}\in\mathbb{R}_+^{\mathcal{L}}\\ \|\mathbf{w}\|_2=1}} \langle\mathbf{w},\mathbf{v}\rangle$$
$$= \inf_{\substack{\mathbf{v}\in\mathbb{R}_+^{\mathcal{L}}\\ \langle\mathbf{v},\mathbf{c}\rangle\le 1, \langle\mathbf{v},\mathbf{q}\rangle\ge\rho^2}} \|\mathbf{v}\|_2, \tag{34}$$

Furthermore, the array $\mathbf{w}^*$ attaining the sup is given by $w_l^* = v_l/\|\mathbf{v}\|_2$. Now, the minimization at the right-hand side of (34) involves a convex second-order cost function $\|\mathbf{v}\|_2^2$ and linear constraints $v_l \ge 0$, $\langle\mathbf{v},\mathbf{c}\rangle \le 1$ and $\langle\mathbf{v},\mathbf{q}\rangle \le \rho^2$. Therefore, according to KKT conditions, if there exist $\mu,\lambda \ge 0$ and $\boldsymbol{\nu} \in \mathbb{R}_+^{\mathcal{L}}$ satisfying for some $\mathbf{v}^* \in \mathbb{R}_+^{\mathcal{L}}$ the conditions $2\mathbf{v}^* + \lambda\mathbf{c} - \mu\mathbf{q} - \boldsymbol{\nu} = 0$ and $\lambda(\langle\mathbf{v}^*,\mathbf{c}\rangle - 1) = 0$, $\mu(\langle\mathbf{v}^*,\mathbf{q}\rangle - \rho^2) = 0$ and $\nu_l v_l^* = 0$ for all l, then $\mathbf{v}^*$ is a solution to the minimization problem (34). Under the conditions of the proposition, one easily checks that these KKT conditions are fulfilled with $\lambda = 2/\sum_l c_l(T_\rho q_l - c_l)_+$, $\mu = 2T_\rho/\sum_l c_l(T_\rho q_l - c_l)_+$ and $\nu_l = 2(c_l - T_\rho q_l)_+/\sum_l c_l(T_\rho q_l - c_l)_+$.

A.4. Proof of Theorem 1

To ease notation, we set $\mathcal{N}_{n,\gamma} = \mathcal{N}(T_{n,\gamma})$. We first check that under the assumptions of the theorem all the conditions required in Propositions 1 and 2 are fulfilled. Since $\|\widehat{\mathbf{w}}_n^*\|_0 = |\mathcal{N}_{n,\gamma}|$ and $\|\widehat{\mathbf{w}}_n^*\|_\infty^2 \le \max_{l\in\mathcal{N}_{n,\gamma}} q_l^2/\sum_{l\in\mathcal{N}_{n,\gamma}} \left(q_l - \frac{c_l}{T_{n,\gamma}}\right)^2$,

condition [**C1**] implies the first condition of Proposition 1. Conditions [**C3**] and [**C4**] imply respectively the third and the second conditions of Proposition 1. Finally, condition [**C6**] implies the fourth condition of Proposition 1. Thus, we have checked that under the conditions of the theorem, the claim of Proposition 1 holds true. To check that the claim of Proposition 2 holds true as well, it suffices to check the first assumption of that proposition (the second one being identical to [**C7**]). In fact, it is not difficult to check that the first assumption of Proposition 2 follows from [**C2**], [**C4**] and [**C5**] for the sequence $\zeta_n^2 = \min_{l\in\mathcal{N}_{n,\gamma}} q_l^2/4\sum_{l\in\mathcal{N}_{n,\gamma}} q_l^2$.

Therefore, combining the results of Proposition 1 and 2, we upper-bound $\gamma_n(\mathcal{F}_0, \mathcal{F}_1(r^*_{n,\gamma}), \widehat{\phi}^*_n)$ by

$$\Phi(-z_{1-\gamma/2}) + \Phi\big(z_{1-\gamma/2} - \inf_{f\in\mathcal{F}_1(r^*_{n,\gamma})} h_n[f, \widehat{\mathbf{w}}^*_n]\big) + o(1). \tag{35}$$

In view of Proposition 3, the infimum over f of $h_n[f, \widehat{\mathbf{w}}^*_n]$ can be evaluated as follows:

$$\begin{aligned}
\inf_{f\in\mathcal{F}_1(r^*_{n,\gamma})} h_n[f, \widehat{\mathbf{w}}^*_n] &= \Big(\frac{m(m-1)}{2}\Big)^{1/2} \inf_{\substack{\boldsymbol{\theta}\in\mathbb{R}^{\mathcal{L}}:\sum_l c_l\theta_l^2\le 1\\ \sum_l q_l\theta_l^2\ge (r^*_{n,\gamma})^2}} \sum_l \widehat{w}^*_{l,n}\theta_l^2 \\
&= \Big(\frac{m(m-1)}{2}\Big)^{1/2} \inf_{\substack{\mathbf{v}\in\mathbb{R}^{\mathcal{L}}_+:\langle\mathbf{v},\mathbf{c}\rangle\le 1\\ \langle\mathbf{v},\mathbf{q}\rangle\ge (r^*_{n,\gamma})^2}} \langle\widehat{\mathbf{w}}^*_n, \mathbf{v}\rangle \\
&= \Big(\frac{m(m-1)}{2}\Big)^{1/2} \|\mathbf{v}^*\|_2 \\
&= \Big(\frac{m(m-1)}{2}\Big)^{1/2} \frac{\big(\sum_{l\in\mathcal{N}_{n,\gamma}} (T_{n,\gamma}q_l - c_l)^2\big)^{1/2}}{\sum_{l\in\mathcal{N}_{n,\gamma}} c_l(T_{n,\gamma}q_l - c_l)}.
\end{aligned}$$

Inserting this expression in (35) and using (15), we get that

$$\begin{aligned}
\gamma_n(\mathcal{F}_0, \mathcal{F}_1(r^*_{n,\gamma}), \widehat{\phi}^*_n) &\le \Phi(-z_{1-\gamma/2}) + \Phi\big(z_{1-\gamma/2} - 2z_{1-\gamma/2} + o(1)\big) + o(1) \\
&= 2\Phi(-z_{1-\gamma/2}) + o(1) = \gamma + o(1).
\end{aligned}$$

A.5. Proof of Theorem 2

The proof of the lower bound follows the steps of (Ingster and Sapatinas, 2009). However, we considerably modified the way some of these steps are carried out which allowed us to relax several assumptions and resulted in a shorter proof.

Let us recall that $\boldsymbol{\theta}[f] = (\theta_l[f])_{l\in\mathcal{L}} \in \ell_2(\mathcal{L})$ is the array of Fourier coefficients of a function in $L_2(\Delta)$ w.r.t. the system $(\varphi_l)_{l\in\mathcal{L}}$. We introduce the sets $\Theta_1(\rho) = \big\{\boldsymbol{\theta} \in \ell_2(\mathcal{L}) : \langle\mathbf{c}, \boldsymbol{\theta}^2\rangle \le 1,\ \langle\mathbf{q}, \boldsymbol{\theta}^2\rangle \ge \rho^2\big\}$ and $\Theta_0 = \big\{\boldsymbol{\theta} \in \ell(\mathcal{L}) : \langle\mathbf{c}, \boldsymbol{\theta}^2\rangle \le 1,\ \langle\mathbf{q}, \boldsymbol{\theta}^2\rangle = 0\big\}$, where we used the notation $\boldsymbol{\theta}^2 = \{\theta_l^2\}_{l\in\mathcal{L}}$. Clearly, if f belongs to the functional class $\mathcal{F}_1(\rho)$ (resp. $\mathcal{F}_0$) then $\boldsymbol{\theta}[f] \in \Theta_1(\rho)$ (resp. $\boldsymbol{\theta}[f] \in \Theta_0$).

Let $C < 1$ be a constant. Our goal is to prove that $\gamma_n(\mathcal{F}_0, \mathcal{F}_1(Cr^*_{n,\gamma})) \ge \gamma + o(1)$. To get this lower bound, we define prior measures that are essentially

concentrated on the sets Θ_0 and Θ_1. Let π_n^1 and π_n^2 be measures on the space $\ell_2(\mathcal{L})$ such that $\pi_n^1(\Theta_0) = 1 + o(1)$ and $\pi_n^2(\Theta_1(Cr_{n,\gamma}^*)) = 1 + o(1)$. Those priors lead to the corresponding mixtures:

$$P_{\pi_n^i}(A) = \int P_{\boldsymbol{\theta}}(A)\pi_n^i(d\boldsymbol{\theta}) \quad \forall \text{ measurable set } A \subset (\Delta \times \mathbb{R})^n, \quad i = 1, 2.$$

If $\gamma_n(P_{\pi_n^1}, P_{\pi_n^2}) = \inf_{\psi:(\Delta\times\mathbb{R})^n\to\{1,2\}} \{P_{\pi_n^1}(\psi = 2) + P_{\pi_n^2}(\psi = 1)\}$ is the minimal total error probability for testing the simple null hypothesis $H_0 : P = P_{\pi_n^1}$ against the simple alternative $H_1 : P = P_{\pi_n^2}$, then we have (see Proposition 2.11 in Ingster and Suslina (2003))

$$\gamma_n\big(\mathcal{F}_0, \mathcal{F}_1(Cr_{n,\gamma}^*)\big) \geq \gamma_n(P_{\pi_n^1}, P_{\pi_n^2}) + o(1).$$

As shows the next result, to get the desired lower bound, it suffices to show that the Bayesian log-likelihood $\log(dP_{\pi_n^2}/dP_{\pi_n^1})$ is asymptotically equivalent to a Gaussian log-likelihood.

Lemma 5 (section 4.3.1 in Ingster and Suslina (2003))**.** *If there exists a deterministic sequence u_n and a sequence of random variables η_n such that under $P_{\pi_n^1}$-probability η_n converges in distribution to $\mathcal{N}(0,1)$ and*

$$\log(dP_{\pi_n^2}/dP_{\pi_n^1}) = u_n\eta_n - \frac{u_n^2}{2} + o_P(1), \tag{36}$$

then $\gamma_n(P_{\pi_n^1}, P_{\pi_n^2}) \geq 2\Phi(-u_n/2) + o(1)$.

For our purposes, we choose π_n^1 to be the Dirac measure in $\mathbf{0}$ and denote the corresponding mixture probability $P_{\pi_n^1}$ by P_0. It is clear that with this choice $\pi_n^1(\Theta_0) = 1$. We now explain how π_n^2, that we will call π_n from now on, is built. Let $\mathbf{a}_n \in \mathbb{R}_+^{\mathcal{L}}$ be an array containing a finite number of nonzero elements. Let $\mathcal{L}(\mathbf{a}_n)$ be the support of $\mathbf{a}_n$, *i.e.*, $a_l \neq 0$ if and only if $l \in \mathcal{L}(\mathbf{a}_n)$. We assume that $\mathcal{L}(\mathbf{a}_n) \subset S_F$ and define $\pi_n(d\boldsymbol{\theta})$ as the Gaussian product measure such that under π_n the entries θ_l are independent Gaussian with zero mean and variance a_l.

Proposition 6. *Let $\delta \in (0,1)$ be such that $1 - \delta \geq C$. Assume that $\mathbf{a}_n = (1-\delta)\mathbf{v}_n$ and, as $n \to \infty$, the following assumptions are fulfilled:*

[L1] $\langle \mathbf{c}, \mathbf{v}_n\rangle \leq 1$ *and* $\langle \mathbf{q}, \mathbf{v}_n\rangle \geq (r_{n,\gamma}^*)^2$,
[L2] $\max_{l\in\mathcal{L}(\mathbf{v}_n)}(q_l v_l) = o(\langle \mathbf{q}, \mathbf{v}\rangle)$ *and* $\max_{l\in\mathcal{L}(\mathbf{v}_n)}(c_l v_l) = o(\langle \mathbf{c}, \mathbf{v}\rangle)$,
[L3] $\|\mathbf{v}_n\|_0 \to \infty$ *and* $n\|\mathbf{v}_n\|_\infty^2\|\mathbf{v}_n\|_0^2 \log\|\mathbf{v}_n\|_0 \to 0$,
[L4] $n\|\mathbf{v}_n\|_\infty\|\mathbf{v}_n\|_0^{1/3} \to 0$ *and* $\|\mathbf{v}_n\|_3 = o(\|\mathbf{v}_n\|_2)$.
[L5] *For some $L_5 > 0$, it holds* $\sum_{l\in\mathcal{L}(\mathbf{a}_n)} \varphi_l^2(\mathbf{t}) \leq L_5\|\mathbf{a}_n\|_0$.

Then, as $n \to \infty$,

$$\gamma_n(\mathcal{F}_0, \mathcal{F}_1(Cr_{n,\gamma}^*)) \geq 2\Phi\Big(-\frac{n(1-\delta)}{2\sqrt{2}}\|\mathbf{v}_n\|_2\Big) + o(1). \tag{37}$$

Proof. The proof of this proposition will be carried out with the help of several lemmas. The fact that $\pi_n\big(\Theta_1(Cr_{n,\gamma}^*)\big) = 1 + o(1)$ is proved in the following lemma.

Lemma 6. *Assume that* $\mathbf{a}_n = (1-\delta)\mathbf{v}_n$ *satisfies* **[L1]** *and* **[L2]**. *Then, for every* $\delta \in (0,1)$, *it holds that* $\pi_n\big(\Theta_1(Cr_n^*)\big) = 1 + o(1)$.

Proof. Let us denote $\mathcal{H}_1(\boldsymbol{\theta}) = \sum_{l\in\mathcal{L}} q_l\theta_l^2$ and $\mathcal{H}_2(\boldsymbol{\theta}) = \sum_{l\in\mathcal{L}} c_l\theta_l^2$. In view of **[L1]**, we have

$$\int \mathcal{H}_1(\boldsymbol{\theta})\pi_n(d\boldsymbol{\theta}) = \sum_{l\in\mathcal{L}} q_l a_l \ge (r_{n,\gamma}^*)^2(1-\delta),$$
$$\int \mathcal{H}_2(\boldsymbol{\theta})\pi_n(d\boldsymbol{\theta}) = \sum_{l\in\mathcal{L}} c_l a_l \le 1-\delta.$$

On the other hand, since the variance of the sum of independent random variables equals the sum of the variances of these random variables, we get

$$\int \mathcal{H}_1(\boldsymbol{\theta})^2\pi_n(d\boldsymbol{\theta}) - \Big(\int \mathcal{H}_1(\boldsymbol{\theta})\pi_n(d\boldsymbol{\theta})\Big)^2 = 2\sum_{l\in\mathcal{L}} q_l^2 a_l^2 \le 2\langle \mathbf{q}, \mathbf{a}_n\rangle \max_{l\in\mathcal{L}(\mathbf{a}_n)}(q_l a_l).$$

By Tchebychev's inequality, we arrive at

$$\pi_n\big(\boldsymbol{\theta} : \mathcal{H}_1(\boldsymbol{\theta}) < (Cr_{n,\gamma}^*)^2\big) \le \frac{2\max_{l\in\mathcal{L}(\mathbf{v}_n)}(q_l v_l)}{C^2(1-C)^2\langle \mathbf{q}, \mathbf{v}_n\rangle},$$
$$\pi_n\big(\boldsymbol{\theta} : \mathcal{H}_2(\boldsymbol{\theta}) > 1\big) \le \frac{2\max_{l\in\mathcal{L}(\mathbf{v}_n)}(c_l v_l)}{\delta^2\langle \mathbf{c}, \mathbf{v}_n\rangle}.$$

The claim of the lemma follows now from condition **[L2]**. □

Second, we show that for every $p > 2$ and every $L > 0$, the probability $\pi_n(\boldsymbol{\theta} : \|\sum_l \theta_l\varphi_l\|_p > L)$ tends to zero. Indeed, in view of the Tchebychev inequality and Fubini's theorem,

$$\pi_n\Big(\boldsymbol{\theta} : \Big\|\sum\nolimits_l \theta_l\varphi_l\Big\|_p > L\Big) \le L^{-p}\int_\Delta E_{\pi_n}\Big[\Big|\sum_l \theta_l\varphi_l(\mathbf{t})\Big|^p\Big]\,d\mathbf{t}.$$

Using the fact that for every fixed $\mathbf{t}$, the random variable $\sum_l \theta_l\varphi_l(\mathbf{t})$ is Gaussian with zero mean and variance $\sum_l a_l\varphi_l^2(\mathbf{t})$, we get

$$\pi_n\Big(\boldsymbol{\theta} : \Big\|\sum\nolimits_l \theta_l\varphi_l\Big\|_p > L\Big) \le p!L^{-p}\int_\Delta \Big|\sum_l a_l\varphi_l^2(\mathbf{t})\Big|^{p/2}\,d\mathbf{t}$$
$$\le p!L_5^{p/2}L^{-p}(\|\mathbf{a}_n\|_\infty\|\mathbf{a}_n\|_0)^{p/2}.$$

The last expression tends to zero as $n\to\infty$ in view of condition **[L3]**.

We focus now on the proof of (36). Set $m = |\mathcal{L}(\mathbf{a}_n)|$ and let Φ_n be the $m\times n$ matrix having as generic element $(\Phi_n)_{li} = \varphi_l(\mathbf{t}_i)$. Let $\mathtt{A}_n$ be $m\times m$ diagonal matrix having the nonzero entries of $\mathbf{a}_n$ on its main diagonal. It is clear that under P_{π_n}, conditionally to $\mathcal{T}_n$, $\mathbf{x} = (x_1,\dots,x_n)^\top$ is distributed according to a

multivariate Gaussian distribution with zero mean and $n \times n$ covariance matrix $\mathtt{R}_n = \Phi_n^\top \mathtt{A}_n \Phi_n + \mathtt{I}_n$. Therefore, the logarithm of its density w.r.t. P_0 is given by

$$\log\Big(\frac{dP_{\pi_n}}{dP_0}(\mathbf{x}; \mathbf{t}_1, \dots, \mathbf{t}_n)\Big) = -\frac{1}{2}\big(\log\det \mathtt{R}_n + \mathbf{x}^\top(\mathtt{R}_n^{-1} - \mathtt{I}_n)\mathbf{x}\big).$$

In what follows, we denote by $|\!|\!|\mathtt{M}|\!|\!| = \sup_{\|\mathbf{x}\|_2=1} \|\mathtt{M}\mathbf{x}\|_2$ the spectral norm of a matrix $\mathtt{M}$.

Lemma 7. *Let $\bar{\mathtt{R}}_n = n\mathtt{A}_n + \mathtt{I}_m$ and $m = m_n \to \infty$ as n tends to infinity. If $n^2\|\mathbf{a}_n\|_\infty^2\|\mathbf{a}_n\|_0|\!|\!|\frac{1}{n}\Phi_n\Phi_n^\top - \mathtt{I}_m|\!|\!|^2 + |\mathrm{Tr}[\bar{\mathtt{R}}_n^{-1}\mathtt{B}_n]| + E[|\boldsymbol{\xi}^\top\bar{\mathtt{R}}_n^{-1}\mathtt{B}_n\bar{\mathtt{R}}_n^{-1}\boldsymbol{\xi}|] = o_P(1)$, then under P_0 it holds $\log\big(dP_{\pi_n}/dP_0\big) = -\frac{1}{2}\big(\log\det\bar{\mathtt{R}}_n + \boldsymbol{\xi}^\top(\bar{\mathtt{R}}_n^{-1} - \mathtt{I}_m)\boldsymbol{\xi}\big) + o_P(1)$, where $\boldsymbol{\xi} \sim \mathcal{N}_m(0, \mathtt{I}_m)$.*

Proof. Let us denote $\tilde{\mathtt{R}}_n = \mathtt{A}_n^{1/2}\Phi_n\Phi_n^\top\mathtt{A}_n^{1/2} + \mathtt{I}_m$, $\mathtt{B}_n = \tilde{\mathtt{R}}_n - \bar{\mathtt{R}}_n$ and introduce the function $g(z) = \log\det(\bar{\mathtt{R}}_n + z\mathtt{B}_n)$ for $z \in [0,1]$. One easily checks that $g(1) = \log\det\tilde{\mathtt{R}}_n = \log\det\mathtt{R}_n$, $g(0) = \log\det\bar{\mathtt{R}}_n$ and $g'(z) = \mathrm{Tr}[(\bar{\mathtt{R}}_n + z\mathtt{B}_n)^{-1}\mathtt{B}_n]$. Therefore, the relation $g(1) - g(0) = g'(\bar{z})$ for some $\bar{z} \in [0,1]$ implies

$$\begin{aligned}|\log\det\mathtt{R}_n - \log\det\bar{\mathtt{R}}_n| &= |\,\mathrm{Tr}[(\bar{\mathtt{R}}_n + \bar{z}\mathtt{B}_n)^{-1}\mathtt{B}_n]| \\ &\le |\,\mathrm{Tr}[\bar{\mathtt{R}}_n^{-1}\mathtt{B}_n]| + m|\!|\!|(\bar{\mathtt{R}}_n + \bar{z}\mathtt{B}_n)^{-1} - \bar{\mathtt{R}}_n^{-1}|\!|\!||\!|\!|\mathtt{B}_n|\!|\!|.\end{aligned}$$

Using the identity $(\bar{\mathtt{R}}_n + \bar{z}\mathtt{B}_n)^{-1} - \bar{\mathtt{R}}_n^{-1} = -\bar{z}(\bar{\mathtt{R}}_n + \bar{z}\mathtt{B}_n)^{-1}\mathtt{B}_n\bar{\mathtt{R}}_n^{-1}$, we get

$$\begin{aligned}|\log\det\mathtt{R}_n - \log\det\bar{\mathtt{R}}_n| &\le |\,\mathrm{Tr}[\bar{\mathtt{R}}_n^{-1}\mathtt{B}_n]| + m|\!|\!|(\bar{\mathtt{R}}_n + \bar{z}\mathtt{B}_n)^{-1}|\!|\!||\!|\!|\bar{\mathtt{R}}_n^{-1}|\!|\!||\!|\!|\mathtt{B}_n|\!|\!|^2 \\ &\le |\,\mathrm{Tr}[\bar{\mathtt{R}}_n^{-1}\mathtt{B}_n]| + m|\!|\!|\mathtt{B}_n|\!|\!|^2,\end{aligned}$$

where we used that $\bar{\mathtt{R}}_n$ and $\bar{\mathtt{R}}_n + \bar{z}\mathtt{B}_n = \mathtt{I}_m + \bar{z}\mathtt{A}_n^{1/2}\Phi_n\Phi_n^\top\mathtt{A}_n^{1/2} + (1-\bar{z})n\mathtt{A}_n$ have all their eigenvalues ≥ 1. On the other hand, one can check that $|\!|\!|\mathtt{B}_n|\!|\!| \le n|\!|\!|\mathtt{A}_n|\!|\!||\!|\!|\frac{1}{n}\Phi\Phi^\top - \mathtt{I}_m|\!|\!|$. Combining these inequalities with the facts $|\!|\!|\mathtt{A}_n|\!|\!| = \|\mathbf{a}_n\|_\infty$ and $m = \|\mathbf{a}_n\|_0 \to \infty$ we arrive at $\log\det\mathtt{R}_n = \log\det\bar{\mathtt{R}}_n + o_P(1)$.

The term $\mathbf{x}^\top\mathtt{R}_n^{-1}\mathbf{x}$ is dealt with similarly. First, using the singular values decomposition of the matrix $\mathtt{A}_n^{1/2}\Phi_n$, one can note that for an appropriately chosen vector $\boldsymbol{\xi} \sim \mathcal{N}_m(0, \mathtt{I}_m)$, it holds that $\mathbf{x}^\top(\mathtt{R}_n^{-1} - \mathtt{I}_n)\mathbf{x} = \boldsymbol{\xi}^\top(\tilde{\mathtt{R}}_n^{-1} - \mathtt{I}_m)\boldsymbol{\xi}$. Then, we introduce the function $\bar{g}(z) = \boldsymbol{\xi}^\top[\bar{\mathtt{R}}_n + z\mathtt{B}_n]^{-1}\boldsymbol{\xi}$, the derivative of which is given by $g'(z) = -\boldsymbol{\xi}^\top(\bar{\mathtt{R}}_n + z\mathtt{B}_n)^{-1}\mathtt{B}_n(\bar{\mathtt{R}}_n + z\mathtt{B}_n)^{-1}\boldsymbol{\xi}$. Therefore, for some $\bar{z} \in [0,1]$,

$$\begin{aligned}|\boldsymbol{\xi}^\top\tilde{\mathtt{R}}_n^{-1}\boldsymbol{\xi} - \boldsymbol{\xi}^\top\bar{\mathtt{R}}_n^{-1}\boldsymbol{\xi}| &= |\boldsymbol{\xi}^\top(\bar{\mathtt{R}}_n + \bar{z}\mathtt{B}_n)^{-1}\mathtt{B}_n(\bar{\mathtt{R}}_n + \bar{z}\mathtt{B}_n)^{-1}\boldsymbol{\xi}| \\ &\le |\boldsymbol{\xi}^\top\bar{\mathtt{R}}_n^{-1}\mathtt{B}_n\bar{\mathtt{R}}_n^{-1}\boldsymbol{\xi}| + |\boldsymbol{\xi}^\top[(\bar{\mathtt{R}}_n + \bar{z}\mathtt{B}_n)^{-1} - \bar{\mathtt{R}}_n]^{-1}\mathtt{B}_n\bar{\mathtt{R}}_n^{-1}\boldsymbol{\xi}| \\ &\quad + |\boldsymbol{\xi}^\top[(\bar{\mathtt{R}}_n + \bar{z}\mathtt{B}_n)^{-1} - \bar{\mathtt{R}}_n]^{-1}\mathtt{B}_n(\bar{\mathtt{R}}_n + \bar{z}\mathtt{B}_n)^{-1}\boldsymbol{\xi}| \\ &\le |\boldsymbol{\xi}^\top\bar{\mathtt{R}}_n^{-1}\mathtt{B}_n\bar{\mathtt{R}}_n^{-1}\boldsymbol{\xi}| + 2\|\boldsymbol{\xi}\|_2^2|\!|\!|\mathtt{B}_n|\!|\!|^2.\end{aligned}$$

It is well-known that $\|\boldsymbol{\xi}\|_2^2$ being distributed according to the χ^2_m distribution is $O_P(m)$, as $m \to \infty$. This completes the proof of the lemma. □

According to (Vershynin, 2012, Cor. 5.52), under **[C3]**, we have $\|\frac{1}{n}\Phi_n\Phi_n^\top - \mathtt{I}_m\| \le C(\frac{m\log m}{n})^{1/2}$ with probability at least $1-1/n$. Furthermore, using the facts that the $\bar{\mathtt{R}}_n$ is a diagonal matrix with diagonal entries ≥ 1 and that the variance of the sum of independent random variables equals the sum of variances, one readily checks that $E|\operatorname{Tr}[\bar{\mathtt{R}}_n^{-1}\mathtt{B}_n]|^2 + E[|\boldsymbol{\xi}^\top\bar{\mathtt{R}}_n^{-1}\mathtt{B}_n\bar{\mathtt{R}}_n^{-1}\boldsymbol{\xi}|^2] \le 3C_3^2 n\|\mathbf{v}_n\|_\infty^2\|\mathbf{v}_n\|_0^2$. Hence, condition **[L3]** implies that the two conditions of the last lemma are fulfilled and, therefore, its claim holds true. Using the fact that $\mathtt{A}_n$ is diagonal, we get

$$\begin{aligned}\log\left(dP_{\pi_n}/dP_0\right) &= \frac{1}{2}\sum_l\Big(\frac{na_l\xi_l^2}{na_l+1} - \log(na_l+1)\Big) + o_P(1)\\ &= \frac{1}{2}\sum_l\Big(\frac{na_l}{na_l+1} - \log(na_l+1)\Big) + \sum_l\frac{na_l(\xi_l^2-1)}{2(na_l+1)} + o_P(1).\end{aligned} \tag{38}$$

Lemma 8. *Let us denote*

$$u_n = \frac{n\|\mathbf{a}_n\|_2}{\sqrt{2}}, \quad \eta_n = \frac{1}{u_n}\sum_{l\in\mathcal{L}}\frac{na_l(\xi_l^2-1)}{2(na_l+1)}.$$

If the conditions $mn^3\|\mathbf{a}_n\|_\infty^3 \to 0$, *and* $\|\mathbf{a}_n\|_3 = o(\|\mathbf{a}_n\|_2)$ *are fulfilled, then* η_n *converges in distribution to* $\mathcal{N}(0,1)$ *and*

$$\frac{1}{2}\sum_{l\in\mathcal{L}}\Big(\frac{na_l}{na_l+1} - \log(na_l+1)\Big) + \sum_{l\in\mathcal{L}}\frac{na_l(\xi_l^2-1)}{2(na_l+1)} = u_n\eta_n - \frac{u_n^2}{2} + o(1). \tag{39}$$

Proof. Since $n\|\mathbf{a}\|_\infty \to 0$, we have $\frac{na_l}{na_l+1} = na_l - (na_l)^2 + O((na_l)^3)$ and $\log(na_l+1) = na_l - \frac{(na_l)^2}{2} + O((na_l)^3)$. This implies that $\sum_{l\in\mathcal{L}}\left(\frac{na_l}{na_l+1} - \log(na_l+1)\right) = -\frac{1}{2}u_n^2 + O(mn^3\|\mathbf{a}_n\|_\infty^3)$. On the other hand, using the central limit theorem for triangular arrays, we get the weak convergence of η_n to $\mathcal{N}(0,1)$ provided that $u_n^{-3}\sum_l(na_l)^3/(na_l+1)^3$ tends to zero. Since under the conditions of the lemma this convergence trivially holds, we get the claim of the lemma. □

Combining Lemma 5 with (38) and (39), we get (37) and the proposition follows. □

To complete the proof of Theorem 2, we shall show now that if we choose $T_{n,\gamma}$ as in Theorem 1 and define $\mathbf{v}_n$ by

$$v_l = v_{l,n} = \frac{(T_{n,\gamma}q_l - c_l)_+}{\sum_{l\in\mathcal{L}} c_l(T_{n,\gamma}q_l - c_l)_+},$$

then all the conditions of Proposition 6 are fulfilled. We start by noting that **[L1]** is straightforward. To check the first relation in **[L2]**, we use **[C1]** and $|\mathcal{N}(T_{n,\gamma})| \to \infty$, along with the following evaluations: $\forall l \in \mathcal{N}(T_{n,\gamma})$,

$$\frac{q_l v_l}{\langle\mathbf{q},\mathbf{v}\rangle} = \frac{q_l(T_{n,\gamma}q_l - c_l)}{\sum_l q_l(T_{n,\gamma}q_l - c_l)_+} \le \frac{q_l^2}{\sum_l(q_l - \frac{c_l}{T_{n,\gamma}})_+^2} \le \frac{C_1}{|\mathcal{N}(T_{n,\gamma})|}.$$

For the second relation in **[L2]**, in view of (15), $\forall l \in \mathcal{N}(T_{n,\gamma})$ we have

$$\frac{c_l v_l}{\langle \mathbf{c}, \mathbf{v}\rangle} = \frac{c_l(T_{n,\gamma}q_l - c_l)}{\sum_l c_l(T_{n,\gamma}q_l - c_l)_+} \le \frac{T_{n,\gamma}c_l q_l}{\sum_l c_l(T_{n,\gamma}q_l - c_l)_+}$$
$$\le \frac{T_{n,\gamma}c_l q_l\, O(1)}{nT_{n,\gamma}\big(\sum_l (q_l - \frac{c_l}{T_{n,\gamma}})_+^2\big)^{1/2}} \le \frac{\max_{l\in\mathcal{N}(T_{n,\gamma})} c_l}{n|\mathcal{N}(T_{n,\gamma})|^{1/2}}\, O(1).$$

The last term tends to zero due to **[C9]**. From the definition of $\mathbf{v}_n$, equation (15) and condition **[C1]** one can deduce that

$$\|\mathbf{v}_n\|_\infty = \frac{\max_l (T_{n,\gamma}q_l - c_l)_+}{\sum_l c_l(T_{n,\gamma}q_l - c_l)_+} \le \frac{T_{n,\gamma}\max_l q_l}{n\big(\sum_l (T_{n,\gamma}q_l - c_l)_+^2\big)^{1/2}} O(1)$$
$$\le \frac{\max_l q_l}{n|\mathcal{N}(T_{n,\gamma})|^{1/2}\max_l q_l} O(1) = \frac{O(1)}{n|\mathcal{N}(T_{n,\gamma})|^{1/2}}.$$

This inequality yields $n\|\mathbf{v}_n\|_\infty^2\|\mathbf{v}_n\|_0^2 = O(|\mathcal{N}(T_{n,\gamma})|/n)$. Therefore, **[L3]** follows from **[C8]**. Finally, to check that condition **[L4]** is true, we notice that $n\|\mathbf{v}_n\|_\infty\|\mathbf{v}_n\|_0^{1/3} = O(|\mathcal{N}(T_{n,\gamma})|^{\frac{1}{3}-\frac{1}{2}}) = o(1)$ and

$$\frac{\|\mathbf{v}_n\|_3^3}{\|\mathbf{v}_n\|_2^3} = \frac{\sum_l (T_{n,\gamma}q_l - c_l)_+^3}{\big(\sum_l (T_{n,\gamma}q_l - c_l)_+^2\big)^{3/2}} \le \frac{\max_l q_l}{\big(\sum_l (q_l - \frac{c_l}{T_{n,\gamma}})_+^2\big)^{1/2}} \le \frac{C_1^{1/2}}{|\mathcal{N}(T_{n,\gamma})|^{1/2}}.$$

Thus, all the conditions of Proposition 6 are fulfilled and, therefore,

$$\gamma_n(\mathcal{F}_0, \mathcal{F}_1(Cr_{n,\gamma}^*)) \ge 2\Phi\Big(-\frac{n(1-\delta)}{2\sqrt{2}}\|\mathbf{v}_n\|_2\Big) + o(1).$$

Since this equation is true for every $\delta \in (0, 1-C)$, it is also true for $\delta = 0$, and the claim of Theorem 2 follows from (15).

Appendix B: Proofs of lemmas and propositions of Section 3

B.1. Proof of Lemma 1

Let us write $\Pi f = \Pi_1 f + \Pi_2 f$, where Π_1 and Π_2 are the orthogonal projectors in $L_2(\Delta)$ onto the subspaces $\operatorname{span}\{\varphi_l : l \in \mathcal{N}_1(T)\}$ and $\operatorname{span}\{\varphi_l : l \in \mathcal{N}_2(T)\}$, respectively. We first assume that the inequality $\sum_l c_l^{-1} < \infty$ is fulfilled.

On the one hand, using the Cauchy-Schwarz inequality,

$$\|\Pi_2 f\|_4^4 = \int_\Delta \Big(\sum\nolimits_{l\in\mathcal{N}_2(T)} \theta_l[f]\varphi_l(\mathbf{t})\Big)^4 d\mathbf{t} \le 2^{2d}\Big(\sum\nolimits_{l\in\mathcal{N}_2(T)} |\theta_l[f]|\Big)^4$$
$$\le 2^{2d}\Big(\sum\nolimits_{\mathcal{N}_2(T)} c_l\theta_l[f]^2\Big)^2\Big(\sum\nolimits_{\mathcal{N}_2(T)} c_l^{-1}\Big)^2 \le 2^{2d}\Big(\sum\nolimits_{\mathcal{N}_2(T)} c_l^{-1}\Big)^2.$$

On the other hand,

$$\|\Pi_1 f - \widehat{\Pi f}_n\|_4^4 = \int_\Delta \Big(\sum_{l \in \mathcal{N}_1(T)} \big(\widehat{\theta}_l - \theta_l[f]\big)\varphi_l(\mathbf{t}) \Big)^4 d\mathbf{t}$$
$$= \int_\Delta \Big(\frac{1}{n} \sum_{i=1}^n \sum_{l \in \mathcal{N}_1(T)} \big(x_i \varphi_l(\mathbf{t}_i) - \theta_l[f]\big)\varphi_l(\mathbf{t}) \Big)^4 d\mathbf{t}.$$

Using Fubini's theorem and Rosenthal's inequality, for some constant $C > 0$, we get

$$E_f\|\Pi_1 f - \widehat{\Pi f}_n\|_4^4 \le \frac{C}{n^4} \int_\Delta \sum_{i=1}^n E_f \Big(\sum_{l \in \mathcal{N}_1(T)} \big(x_i \varphi_l(\mathbf{t}_i) - \theta_l[f]\big)\varphi_l(\mathbf{t}) \Big)^4 d\mathbf{t}$$
$$+ \frac{C}{n^4} \int_\Delta \Big\{ \sum_{i=1}^n E_f \Big(\sum_{\mathcal{N}_1(T)} \big(x_i \varphi_l(\mathbf{t}_i) - \theta_l[f]\big)\varphi_l(\mathbf{t}) \Big)^2 \Big\}^2 d\mathbf{t}$$

By Hölder's inequality, we get

$$E_f \Big(\sum_{l \in \mathcal{N}_1(T)} \big(x_i \varphi_l(\mathbf{t}_i) - \theta_l[f]\big)\varphi_l(\mathbf{t}) \Big)^4$$
$$\le |\mathcal{N}_1(T)|^3 \sum_{l \in \mathcal{N}_1(T)} E_f \big(x_i \varphi_l(\mathbf{t}_i) - \theta_l[f]\big)^4 \varphi_l(\mathbf{t})^4$$
$$\le 2^{2d} |\mathcal{N}_1(T)|^3 \sum_{l \in \mathcal{N}_1(T)} E_f \big(f(\mathbf{t}_i)\varphi_l(\mathbf{t}_i) + \xi_i \varphi_l(\mathbf{t}_i) - \theta_l[f]\big)^4$$
$$= O(|\mathcal{N}_1(T)|^4),$$

where we used the fact that $E[\xi^4] < \infty$ and that $E[f(t_i)^4] \le 2^{2d}(\sum_l c_l^{-1})^2 < \infty$ under the conditions of the lemma. Similar arguments lead to

$$\int_\Delta \Big\{ \sum_{i=1}^n E_f \Big(\sum_{l \in \mathcal{N}_1(T)} \big(x_i \varphi_l(\mathbf{t}_i) - \theta_l[f]\big)\varphi_l(\mathbf{t}) \Big)^2 \Big\}^2 d\mathbf{t} = O(n^2 |\mathcal{N}_1(T)|^4),$$

which implies that $E_f\|\Pi_1 f - \widehat{\Pi f}_n\|_4^4 = O(|\mathcal{N}_1(T)|^4/n^2)$. Combining the obtained evaluations, we get

$$E_f\|\Pi f - \widehat{\Pi f}_n\|_4^4 \le \frac{|\mathcal{N}_1(T)|^4}{n^2} + C\Big(\sum\nolimits_{l: c_l > T} c_l^{-1} \Big)^2.$$

The required consistency follows from the assumption $|\mathcal{N}_1(T_n)| = o(n^{1/2})$.

Let us consider the case $\Sigma \subset W_2^{\boldsymbol{\sigma}}(R)$. Without loss of generality, we will assume that $\Sigma = W_2^{\boldsymbol{\sigma}}(R)$ and $c_l = \sum_{i=1}^d (2\pi l_i)^{2\sigma_i}/R^2$. The computations remain the same as in the previous case but the term $\|\Pi_2 f\|_4^4$ is bounded using Sobolev

inequality (Kolyada, 1993). Indeed, choosing $\boldsymbol{\sigma}'$ so that $\sigma_i' = (1-\tau)\sigma_i$ and $\tau < 1 - d/(4\bar{\sigma})$ (this implies that $\bar{\sigma}' > d/4$), we get

$$\begin{aligned}\|\Pi_2 f\|_4^2 &\le C\|\Pi_2 f\|_{W_2^{\boldsymbol{\sigma}'}}^2 = C\Big[\sum_{\boldsymbol{l}\in\mathcal{N}_2(T)}\sum_{i=1}^{d}(2\pi l_i)^{2\sigma_i'}\theta_{\boldsymbol{l}}[f]^2\Big]\\ &\le C\Big[\sum_{\boldsymbol{l}\in\mathcal{N}_2(T)} d^{\tau}\big(c_l R^2\big)^{1-\tau}\theta_{\boldsymbol{l}}[f]^2\Big]\\ &\le C(d/T)^{\tau}R^{2(1-\tau)}\Big[\sum_{\boldsymbol{l}\in\mathcal{N}_2(T)} c_l\theta_{\boldsymbol{l}}[f]^2\Big] \le C(d/T)^{\tau}R^{2(1-\tau)}.\end{aligned}$$

This completes the proof, since the last term tends to zero as $T\to\infty$.

B.2. Proof of Lemma 2

Let us introduce $\Pi_J f = \sum_{\boldsymbol{k}\in[1,2^J]^d}\alpha_{J,\boldsymbol{k}}\varphi_{J,\boldsymbol{k}}$. We first decompose the empirical coefficients as follows:

$$\widehat{\alpha}_{J,\boldsymbol{k}} = \frac{1}{n}\sum_{i=1}^{n}\varphi_{J,\boldsymbol{k}}(\mathbf{t}_i)x_i = \frac{1}{n}\sum_{i=1}^{n}\varphi_{J,\boldsymbol{k}}(\mathbf{t}_i)f(\mathbf{t}_i) + \frac{1}{n}\sum_{i=1}^{n}\varphi_{J,\boldsymbol{k}}(\mathbf{t}_i)\xi_i := \tilde{\alpha}_{J,\boldsymbol{k}} + \epsilon_{j,\boldsymbol{k}}.$$

Then, using standard arguments, we have

$$\begin{aligned}\Big\|\Pi f - \widehat{\Pi f}_n\Big\|_4^4 \le{}& 3^3\Big(\Big\|\sum_{\boldsymbol{k}\in[1,2^J]^d}(\alpha_{J,\boldsymbol{k}}-\tilde{\alpha}_{J,\boldsymbol{k}})\varphi_{J,\boldsymbol{k}}\Big\|_4^4 + \Big\|\sum_{\boldsymbol{k}\in[1,2^J]^d}\epsilon_{J,\boldsymbol{k}}\varphi_{J,\boldsymbol{k}}\Big\|_4^4\Big)\\ &+3^3\big\|\Pi f - \Pi_J f\big\|_4^4\end{aligned}$$

with $\big\|\Pi f - \Pi_J f\big\|_4^4 = O(2^{-4J\sigma})$. Furthermore, by well-known properties of wavelet bases (Cohen, 2003) and the Rosenthal inequality,

$$E_f\Big\|\sum_{\boldsymbol{k}\in[1,2^J]^d}(\alpha_{J,\boldsymbol{k}}-\tilde{\alpha}_{J,\boldsymbol{k}})\varphi_{J,\boldsymbol{k}}\Big\|_4^4 = O(2^{Jd})\sum_{\boldsymbol{k}}E_f(\alpha_{J,\boldsymbol{k}}-\tilde{\alpha}_{J,\boldsymbol{k}})^4 = O\Big(\frac{2^{2Jd}}{n^2}\Big)$$

and

$$\begin{aligned}E\Big\|\sum_{\boldsymbol{k}}\epsilon_{J,\boldsymbol{k}}\varphi_{J,\boldsymbol{k}}\Big\|_4^4 &= O(2^{Jd})\sum_{\boldsymbol{k}}E[\epsilon_{J,\boldsymbol{k}}^4] = O(2^{Jd})\sum_{\boldsymbol{k}}E\Big(\frac{1}{n^2}\sum_{i=1}^{n}\varphi_{J,\boldsymbol{k}}^2(\mathbf{t}_i)\Big)^2\\ &= 2^{2Jd}O\Big(\frac{1}{n^2}+\frac{2^{Jd}}{n^3}\Big).\end{aligned}$$

Finally, we obtain, uniformly in $f\in\Sigma$, $E_f\|\Pi f - \widehat{\Pi f}_n\|^4 = O\big(\frac{2^{2Jd}}{n^2}+\frac{2^{3Jd}}{n^3}+2^{-4J\sigma}\big)$, and the announced result follows.

Appendix C: Proof of Proposition 4

We are going to check that all the assumptions of Theorem 1 and Theorem 2 are satisfied. We can use the Sobolev embedding theorem (Kolyada, 1993) for **[C7]**: if $\bar{\sigma} > d/4$, then **[C7]** is satisfied. For the pilot estimator proposed in subsection 3.2, **[C6]** holds as well. Since the Fourier basis is uniformly bounded, checking **[C3]** is straightforward.

Let now $T_{n,\gamma} = (C_\gamma^* r_n^*)^{-2}(1 + 2\kappa^{-1})$, where r_n^* and C_γ^* are defined in Proposition 4. We will show that

- $T_{n,\gamma}$ satisfies (15),
- $r_{n,\gamma}^*$ defined by (16) satisfies $r_{n,\gamma}^* \sim C_\gamma^* r_n^*$,
- conditions **[C1]**, **[C2]**, **[C5]**, **[C8]** and **[C9]** are fulfilled.

To this end, we need an asymptotic analysis of the terms

$$I_0(T) = \sum_{\boldsymbol{l}\in\mathbb{Z}^d} \Big(q_{\boldsymbol{l}} - \frac{c_{\boldsymbol{l}}}{T}\Big)_+^2, \qquad I_1(T) = \sum_{\boldsymbol{l}\in\mathbb{Z}^d} q_{\boldsymbol{l}}\Big(q_{\boldsymbol{l}} - \frac{c_{\boldsymbol{l}}}{T}\Big)_+$$

and $I_2(T) = I_1(T) - I_0(T)$. For the first one, it holds that

$$I_0(T) = \sum_{\boldsymbol{l}\in\mathbb{Z}^d} \Big(\prod_{j=1}^d (2\pi l_j)^{2\alpha_j} - \sum_{i=1}^d \frac{(2\pi l_i)^{2\sigma_i}}{T}\Big)_+^2.$$

For every $i \in \{1, \dots, d\}$, we set

$$m_i = \frac{T^{\gamma_i}}{2\pi}, \quad \gamma_i = \frac{1}{2\sigma_i(1-\delta)} \quad \text{and} \quad x_{\boldsymbol{l},i} = \frac{2\pi l_i}{T^{\gamma_i}} = \frac{l_i}{m_i}.$$

Note that, as $\delta < 1$, we have $\gamma_i > 0$. With this notation,

$$I_0(T) = T^{\frac{2\delta}{1-\delta}} m_1 \cdot \ldots \cdot m_d \sum_{\boldsymbol{l}\in\mathbb{Z}^d} \Big(\prod_{j=1}^d |x_{\boldsymbol{l},j}|^{2\alpha_j} - \sum_{i=1}^d |x_{\boldsymbol{l},i}|^{2\sigma_i}\Big)_+^2 / (m_1 \cdot \ldots \cdot m_d).$$

As $m_i \to \infty$ for every i, we can replace the sums by integrals

$$I_0(T) \sim \frac{T^{\frac{4\delta\bar{\sigma}+d}{2(1-\delta)\bar{\sigma}}}}{(2\pi)^d} \int_{\sum_{i=1}^d |x_j|^{2\sigma_j} < \prod_{j=1}^d |x_j|^{2\alpha_j}} \Big(\prod_{j=1}^d |x_j|^{2\alpha_j} - \sum_{i=1}^d |x_i|^{2\sigma_i}\Big)^2 d\mathbf{x}.$$

Next, we make the change of variables $y_j = x_j^{2\sigma_j}$, $j = 1, \dots, d$ and set $\mathcal{D} = \big\{\mathbf{y} \in \mathbb{R}_+^d : \sum_{j=1}^d y_j < \prod_{i=1}^d y_i^{\alpha_i/\sigma_i}\big\}$. We get

$$I_0(T) \sim \frac{T^{\frac{4\delta\bar{\sigma}+d}{2(1-\delta)\bar{\sigma}}}}{\pi^d \sigma_1 \dots \sigma_d} \int_{\mathcal{D}} \Big(\prod_{i=1}^d y_i^{\frac{\alpha_i}{\sigma_i}} - \sum_{j=1}^d y_j\Big)^2 y_1^{\frac{1}{2\sigma_1}-1} \dots y_d^{\frac{1}{2\sigma_d}-1} d\mathbf{y}.$$

Now, we make another change of variables: $z_i = y_i\big(\prod_{j=1}^d y_j^{\alpha_j/\sigma_j}\big)^{-1}$. Note that $\prod_{i=1}^d z_i^{\alpha_i/\sigma_i} = \big(\prod_{i=1}^d y_i^{\alpha_i/\sigma_i}\big)^{1-\delta}$. Therefore, using the notation $\Sigma_d = \big\{\mathbf{z} \in \mathbb{R}_+^d : \|\mathbf{z}\|_1 \le 1\big\}$,

$$I_0(T) \sim \frac{T^{\frac{4\delta\bar\sigma+d}{2(1-\delta)\bar\sigma}}}{\pi^d\sigma_1\dots\sigma_d}\int_{\Sigma_d}\Big(\prod_{i=1}^d z_i^{\frac{\alpha_i}{\sigma_i}}\Big)^{\frac{4\bar\sigma+d-2d\bar\sigma}{2\bar\sigma(1-\delta)}}(1-\|\mathbf{z}\|_1)^2 z_1^{\frac{1}{2\sigma_1}-1}\dots z_d^{\frac{1}{2\sigma_d}-1}\Delta(\mathbf{z})\,d\mathbf{z},$$

where $\Delta(\mathbf{z})$ is the Jacobian. For the latter, standard algebra yields $\Delta(\mathbf{z}) = \big(\prod_{i=1}^d z_i^{\alpha_i/\sigma_i}\big)^{d/(1-\delta)}/(1-\delta)$. Next we give an explicit form for this integral $I_0(T) \sim \pi^{-d}T^{\frac{4\delta\bar\sigma+d}{2(1-\delta)\bar\sigma}}I$, where

$$I = \frac{1}{\big(\prod_{i=1}^d\sigma_i\big)(1-\delta)}\int_{\Sigma_d}\Big(\prod_{i=1}^d z_i^{\frac{\alpha_i}{\sigma_i}}\Big)^{\frac{4\bar\sigma+d}{2\bar\sigma(1-\delta)}}(1-\|\mathbf{z}\|_1)^2 z_1^{\frac{1}{2\sigma_1}-1}\dots z_d^{\frac{1}{2\sigma_d}-1}d\mathbf{z}.$$

Now, the Liouville formula (see Ingster and Stepanova (2011)) combined with the well-known identity $\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ yields

$$\begin{aligned} I &= \frac{\prod_{i=1}^d\Gamma\big(\frac{1}{2\sigma_i}+\frac{\alpha_i}{\sigma_i}\frac{4\bar\sigma+d}{2\bar\sigma(1-\delta)}\big)}{\big(\prod_{i=1}^d\sigma_i\big)(1-\delta)\Gamma\big(\frac{d}{2\bar\sigma}+(2+\frac{d}{2\bar\sigma})\frac{\delta}{1-\delta}\big)}\int_0^1(1-u)^2u^{\frac{d}{2\bar\sigma}+\frac{(4\bar\sigma+d)\delta}{2\bar\sigma(1-\delta)}-1}du \\ &= \frac{2\prod_{i=1}^d\Gamma(\kappa_i)}{\big(\prod_{i=1}^d\sigma_i\big)(1-\delta)\Gamma(\kappa+3)} = \frac{2\prod_{i=1}^d\Gamma(\kappa_i)}{\big(\prod_{i=1}^d\sigma_i\big)(1-\delta)(\kappa+2)\Gamma(\kappa+2)} \\ &= \frac{2\pi^dC(d,\boldsymbol{\sigma},\boldsymbol{\alpha})}{\kappa+2}. \end{aligned}$$

Therefore,

$$I_0(T) \sim \frac{2C(d,\boldsymbol{\sigma},\boldsymbol{\alpha})}{\kappa+2}T^{\frac{4\delta\bar\sigma+d}{2(1-\delta)\bar\sigma}}.$$

Very similar computations imply that, as $T\to\infty$, we have

$$I_1(T) \sim \frac{T^{\frac{4\delta\bar\sigma+d}{2(1-\delta)\bar\sigma}}}{\pi^d}\frac{\prod_{i=1}^d\Gamma(\kappa_i)}{\big(\prod_{i=1}^d\sigma_i\big)(1-\delta)\Gamma(\kappa+2)} = C(d,\boldsymbol{\sigma},\boldsymbol{\alpha})T^{\frac{4\delta\bar\sigma+d}{2(1-\delta)\bar\sigma}}.$$

Note now that (15) is equivalent to $n^2T^2I_0(T) \sim 8T^4(I_1(T)-I_0(T))^2z_{1-\gamma/2}^2$. Using the asymptotic equivalents for I_0 and I_1 we have derived above, one directly checks that the value of $T_{n,\gamma}$ proposed in Proposition 4 satisfies (15). Furthermore, since (16) is equivalent to $(r_{n,\gamma}^*)^2 = I_1(T_{n,\gamma})/T_{n,\gamma}I_2(T_{n,\gamma})$, we get $r_{n,\gamma}^* = C_\gamma^* r_n^*(1+o(1))$, as announced in proposition.

It remains to check that for the sequence $T_{n,\gamma} \asymp n^{\frac{4\bar\sigma(1-\delta)}{4\bar\sigma+d}}$ conditions **[C1]**, **[C2]**, **[C5]**, **[C8]** and **[C9]** are fulfilled. Using the same method as the one used above to evaluate I_0, we get

$$|\mathcal{N}(T_{n,\gamma})| \asymp n^{\frac{2d}{4\bar\sigma+d}} \quad\text{and}\quad M(T_{n,\gamma}) = \sum_{\boldsymbol{l}\in\mathcal{N}(T_{n,\gamma})} q_{\boldsymbol{l}}^2 \asymp n^{\frac{2(4\delta\bar\sigma+d)}{4\bar\sigma+d}}. \tag{40}$$

The assumption $\bar{\sigma} > d/4$ implies $|\mathcal{N}(T_{n,\gamma})| \log |\mathcal{N}(T_{n,\gamma})| = o(n)$ and, as a consequence, conditions **[C4]** and **[C8]** are true. Furthermore, the second relation in (40) combined with $\delta < 1$ implies **[C2]**. Condition **[C5]** follows from the fact that all the nonzero entries of $\mathbf{q}$ are lower-bounded by 1.

In order to check **[C1]** and **[C9]**, we need to find an upper bound for $\max_{l \in \mathcal{N}(T_{n,\gamma})} q_{\boldsymbol{l}}$. In the following calculations, the term C is a constant which depends only on d, $\boldsymbol{\alpha}$ and $\boldsymbol{\sigma}$ and can vary from line to line. Let $l \in \mathcal{N}(T)$, then $c_{\boldsymbol{l}} \le T q_{\boldsymbol{l}}$, which implies, for every $i = 1, \dots, d$, $l_i^{2(\sigma_i - \alpha_i)} \le CT \prod_{j \ne i} l_j^{2\alpha_j}$. In particular

$$l_1^{2(\sigma_1-\alpha_1)} \le CT \prod_{j\ne 1} l_j^{2\alpha_j}, \quad l_2^{2(\sigma_2-\alpha_2)} \le CT \prod_{j\ne 2} l_j^{2\alpha_j}, \quad l_3^{2(\sigma_3-\alpha_3)} \le CT \prod_{j\ne 3} l_j^{2\alpha_j}. \tag{41}$$

Injecting the first inequality of (41) in the second one, we obtain

$$\begin{aligned} l_2^{2(\sigma_2-\alpha_2)} &\le CT\Big(\prod_{j\ge 3} l_j^{2\alpha_j}\Big)\Big(T\prod_{j\ge 2} l_j^{2\alpha_j}\Big)^{\frac{\alpha_1}{\sigma_1-\alpha_1}} \\ &\le C\Big(T\prod_{j\ge 3} l_j^{2\alpha_j}\Big)^{\frac{\sigma_1}{\sigma_1-\alpha_1}}\Big(l_2^{2\alpha_2}\Big)^{\frac{\alpha_1}{\sigma_1-\alpha_1}}. \end{aligned}$$

Hence

$$l_2^{2\sigma_2} \le C\Big(T\prod\nolimits_{j\ge 3} l_j^{2\alpha_j}\Big)^{\frac{1}{1-\alpha_1/\sigma_1-\alpha_2/\sigma_2}} \tag{42}$$

and by symmetry,

$$l_1^{2\sigma_1} \le C\Big(T\prod\nolimits_{j\ge 3} l_j^{2\alpha_j}\Big)^{\frac{1}{1-\alpha_1/\sigma_1-\alpha_2/\sigma_2}}. \tag{43}$$

Next, using (42), (43) and the third inequality in (41), we get

$$l_3^{2\sigma_3} \le C\Big(T\prod\nolimits_{j\ge 4} l_j^{2\alpha_j}\Big)^{\frac{1}{1-\alpha_1/\sigma_1-\alpha_2/\sigma_2-\alpha_3/\sigma_3}}.$$

Iterations of the previous process lead to the bound $\max_j l_j^{2\sigma_j} \le CT^{1/(1-\delta)}$. Therefore, $\max_{\boldsymbol{l}\in\mathcal{N}(T)} q_{\boldsymbol{l}} = C\prod_{j=1}^d l_j^{2\alpha_j} \le CT^{\frac{\delta}{1-\delta}}$. Combining this bound with $T_{n,\gamma} \asymp n^{\frac{4\bar{\sigma}(1-\delta)}{4\bar{\sigma}+d}}$ and (40) yields the inequalities of **[C1]** and **[C9]**.

Appendix D: Proof of Proposition 5

As in the previous subsection, we begin with the calculation of I_0. Setting $x_{\boldsymbol{l},i} = \frac{2\pi l_i}{T^{\frac{1}{\sigma-1}}}$ and using the same method to get an integral, we have

$$\begin{aligned} I_0 &= T^{\frac{4}{\sigma-1}} \sum_{\boldsymbol{l}\in\mathbb{Z}^d} \Big[\|\mathbf{x}_{\boldsymbol{l}}\|_2^2 - (\boldsymbol{\beta}^\top \mathbf{x}_{\boldsymbol{l}})^2 - \|\mathbf{x}_{\boldsymbol{l}}\|_{2\sigma}^{2\sigma}\Big]_+^2 \\ &\sim \frac{T^{\frac{d+4}{\sigma-1}}}{(2\pi)^d} \int_{\mathbb{R}^d} \Big[\|\mathbf{x}\|_2^2 - (\boldsymbol{\beta}^\top \mathbf{x})^2 - \|\mathbf{x}\|_{2\sigma}^{2\sigma}\Big]_+^2 d\mathbf{x}. \end{aligned}$$

This implies the asymptotic relation $I_0 \sim C_0 T^{\frac{d+4}{\sigma-1}}$ with the constant $C_0 = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\big[\sum_{i=1}^d (x_i - \frac{1}{d}\sum_{j=1}^d x_j)^2 - \sum_{i=1}^d x_i^{2\sigma}\big]_+^2 d\mathbf{x}$. Similar computations yield $I_1 \sim C_1 T^{(d+4)/(\sigma-1)}$ and $I_2 \sim C_2 T^{(d+4)/(\sigma-1)}$, where C_1 and C_2 have the values given in the paragraph preceding the proposition.

The rest of the proof can be carried out exactly in the same way as the proof of the previous proposition, based on the relation $N(T) \asymp T^{\frac{d}{\sigma-1}}$ and $M(T) \asymp T^{\frac{d+4}{\sigma-1}}$.

Appendix E: Proofs of results stated in Section 4

E.1. Proof of Theorem 3

The arguments are almost the same as in the proof of Theorem 1. We use the array $\mathbf{w}_n$ with entries $w_l = q_l \mathbf{1}_{\{l\in\mathcal{N}(T)\}}/M(T)^{1/2}$ and the kernel $G_n(\mathbf{t}_1,\mathbf{t}_2) = \sum_{l\in\mathcal{L}} w_l\varphi_l(\mathbf{t}_1)\varphi_l(\mathbf{t}_2)$ in order to define the linear U-test statistic:

$$U_n = \binom{n}{2}^{-1/2} \sum_{1\le i<j\le n} x_i x_j G_n(\mathbf{t}_i,\mathbf{t}_j).$$

We write as $U_n = U_{n,0} + U_{n,1} + U_{n,2}$, where

$$U_{n,0} = \binom{n}{2}^{-1/2} \sum_{i<j} \xi_i\xi_j G_n(\mathbf{t}_i,\mathbf{t}_j),$$
$$U_{n,1} = \binom{n}{2}^{-1/2} \sum_{i<j} (\xi_i f(\mathbf{t}_j) + \xi_j f(\mathbf{t}_i)) G_n(\mathbf{t}_i,\mathbf{t}_j)$$

and $U_{n,2} = \binom{n}{2}^{-1/2}\sum_{i<j} f(\mathbf{t}_i)f(\mathbf{t}_j)G_n(\mathbf{t}_i,\mathbf{t}_j)$. The first and the second moments of this U-statistic are described in the next result, in which we use the notation $\mathtt{T}_{\mathbf{w}}[f] = \sum_l w_l\theta_l[f]\varphi_l$.

Lemma 9. *Let $\mathbf{w}_n = (w_{l,n})_{l\in\mathcal{L}}$ be an array containing only a finite number of nonzero entries and such that $\sum_{l\in\mathcal{L}} w_{l,n}^2 = 1$. Let $\mathcal{L}(\mathbf{w}_n)$ be the support of $\mathbf{w}_n$. The expectation of the U-statistic U_n is given by:*

$$E_f[U_n] = E_f[U_{n,2}] = \bar{h}_n[f,\mathbf{w}_n] = \Big(\frac{n(n-1)}{2}\Big)^{1/2}\sum_l w_l\theta_l^2[f].$$

Furthermore, if condition **[D2]** *holds true, then $E[U_{n,0}^2] = 1$, $E[U_{n,1}^2] \le 2D_2\|\mathbf{w}_n\|_\infty^2\|\mathbf{w}_n\|_0\|f\|_2^2$ and*

$$\mathrm{Var}[U_{n,2}] \le D_2\|\mathbf{w}_n\|_\infty^2\|\mathbf{w}_n\|_0\|f\|_4^4 + \frac{2n}{3}\|f\cdot\mathtt{T}_{\mathbf{w}}[f]\|_2^2.$$

Proof. This result can be proved along the lines of the proof of Lemma 3. The only difference is in the evaluation of the term $A_{n,2}$, for which we have

$$A_{n,2} = \frac{4}{n(n-1)}\binom{n}{3}\sum_{l,l'\in\mathcal{L}(\mathbf{w}_n)} w_l w_{l'}\theta_l[f]\theta_{l'}[f]\Big\{\int f(\mathbf{t})^2\varphi_l(\mathbf{t})\varphi_{l'}(\mathbf{t})\,d\mathbf{t}\Big\}$$
$$= \frac{2(n-2)}{3}\Big\{\int f(\mathbf{t})^2\Big(\sum_l w_l\theta_l[f]\varphi_l(\mathbf{t})\Big)^2 d\mathbf{t}\Big\} \le \frac{2n}{3}\|f\cdot \mathtt{T}_{\mathbf{w}}[f]\|_2^2.$$

This yields the desired result. □

Let us now study the type I and type II error probabilities of the test $\phi_n(T) = \mathbf{1}_{\{|U_n(T)|>u\}}$.

Evaluation of type I error Using Tchebychev's inequality, for every $u > |E[U_n(T)]|$, we have

$$\sup_{f\in\mathcal{F}_0} P_f\big(|U_n(T)| > u\big) \le \sup_{f\in\mathcal{F}_0} P_f\Big(\big|U_n(T) - E[U_n(T)]\big| > u - \big|E[U_n(T)]\big|\Big)$$
$$\le \sup_{f\in\mathcal{F}_0}\frac{\mathrm{Var}(U_n(T))}{(u-|E[U_n(T)]|)^2}.$$

Let us denote $\nu_{n,T} = nT^{-1}(2M(T))^{-1/2}$. Using Lemma 9, we get

$$|E[U_n(T)]| \le \frac{n}{\sqrt{2M(T)}}\Big|\sum_{l\in\mathcal{N}(T)} q_l\theta_l^2[f]\Big| = T\nu_{n,T}\Big|\sum_{l\in\mathcal{N}(T)} q_l\theta_l^2[f]\Big|.$$

Since, under H_0, we have $Q[f] = \sum_l q_l\theta_l[f]^2 = 0$ and $\sum_l c_l\theta_l[f]^2 \le 1$, the last sum can be bounded as follows: $\big|\sum_{l\in\mathcal{N}(T)} q_l\theta_l^2[f]\big| = \big|\sum_{l:|q_l|<c_l/T} q_l\theta_l^2[f]\big| \le T^{-1}\sum_l c_l\theta_l[f]^2 \le T^{-1}$. Thus, $|E[U_n(T)]| = |\bar{h}_n[\mathbf{w}_n, f]| \le \nu_{n,T}$. Combining this bound with those of Lemma 9, we arrive at

$$\sup_{f\in\mathcal{F}_0} P_f(\phi_n(T)=1) \le \frac{3(1+2D_1D_2D_3^2 + D_1D_2D_3^4 + 2nD_4/(3M(T))}{(u-\nu_{n,T})^2}$$
$$= \frac{B_1 + B_2 nM(T)^{-1}}{2(u-\nu_{n,T})^2}$$

Consequently, if we choose $u \ge \nu_{n,T} + \big(B_1 + B_2 nM(T)^{-1}\big)^{1/2}\gamma^{-1/2}$, then $\sup_{f\in\mathcal{F}_0} P_f(\phi_n(T)=1) \le \frac{\gamma}{2}$.

Evaluation of type II error Using similar arguments, we get

$$\sup_{f\in\mathcal{F}_1(\rho)} P_f(\phi_n(T)=0) = \sup_{f\in\mathcal{F}_1(\rho)} P_f(|U_n(T)|\le u)$$
$$\le \sup_{f\in\mathcal{F}_1(\rho)} P_f\big(|E[U_n(T)]| - |U_n(T) - E[U_n(T)]| \le u\big)$$
$$\le \sup_{f\in\mathcal{F}_1(\rho)} P_f\big(T\nu_{n,T}|Q[f]| - \nu_{n,T} - \big|U_n(T) - E[U_n(T)]\big| \le u\big)$$
$$\le P\big(T\nu_{n,T}\rho^2 - \big|U_n(T) - E[U_n(T)]\big| \le u + \nu_{n,T}\big).$$

This can also be written as:

$$\sup_{f\in\mathcal{F}_1(\rho)} P_f(\phi_n(T)=0) \le P\Big(\big|U_n(T)-E[U_n(T)]\big| \ge (T\rho^2-1)\nu_{n,T}-u\Big).$$

Using the Tchebychev inequality and the evaluations obtained in Lemma 9, we get

$$\sup_{f\in\mathcal{F}_1(\rho)} P_f(\phi_n(T)=0) \le \frac{B_1+B_2nM(T)^{-1}}{2\big((T\rho^2-1)\nu_{n,T}-u\big)^2}.$$

Clearly, the right hand-side of this inequality is lower than $\gamma/2$ if

$$\rho^2 \ge \left[u+\frac{1}{\gamma^{1/2}}\left(B_1+\frac{B_2n}{M(T)}\right)^{1/2}\right]\frac{1}{T\nu_{n,T}}+\frac{1}{T}.$$

This completes the proof of Theorem 3.

E.2. Proof of Corollary 1

It is enough to remark that (since $M(\cdot)$ is increasing and $T_n \le T_n^0$)

$$\frac{\sqrt{M(T_n)}}{n} \le \frac{\sqrt{M(T_n^0)}}{n} \asymp \frac{1}{T_n^0} \le \frac{1}{T_n}$$

and $\frac{1}{\sqrt{n}} \le T_n^{-1}$. In view of these inequalities, the claim of the corollary immediately follows from Theorem 3.

E.3. Proof of Theorem 4

We start by proving that the minimax rate of separation is lower bounded by $n^{-1/4}$. Let $l_a = \operatorname{argmin}_{l\in\mathcal{L}_a}\{c_l\}$ for $a\in\{+,-\}$. We define two functions f_0 and f_1 as linear combinations of the basis functions φ_{l_-} and φ_{l_+}. More precisely, $f_i = \theta_{i,-}\varphi_{l_-}+\theta_{i,+}\varphi_{l_+}$, for $i=0,1$, with

$$\theta_{0,-}^2 = \frac{|q_{l_+}|}{c_{l_-}|q_{l_+}|+c_{l_+}|q_{l_-}|} \quad \theta_{0,+}^2 = \frac{|q_{l_-}|}{c_{l_-}q_{l_+}+c_{l_+}|q_{l_-}|}$$

and, for some $z>0$,

$$\theta_{1,-}=\theta_{0,-} \quad \theta_{1,+}^2 = \theta_{0,+}^2 - z/\sqrt{n}.$$

One easily checks that $f_0\in\mathcal{F}_0$ and $f_1\in\mathcal{F}_1(r_n)$ with $r_n^2 = zq_{l_+}/\sqrt{n}$. Furthermore, the Kullback-Leibler divergence $K(P_{f_0},P_{f_1}) = \int \log\frac{dP_{f_0}}{dP_{f_1}}dP_{f_0}$ between the probability measures P_{f_0} and P_{f_1} can be bounded as follows:

$$K(P_{f_0},P_{f_1}) = E\left(E_{f_0}\left[\log\frac{dP_{f_0}}{dP_{f_1}}(x_1,\dots,x_n,\mathbf{t}_1,\dots,\mathbf{t}_n)\middle|\mathbf{t}_1,\dots,\mathbf{t}_n\right]\right)$$

$$= E\left(E_{f_0}\left[\sum_{i=1}^{n}(x_i - f_1(\mathbf{t}_i))^2 - (x_i - f_0(\mathbf{t}_i))^2 \middle| \mathbf{t}_1, \dots, \mathbf{t}_n\right]\right)$$
$$= nE\big[\big(f_0(\mathbf{t}_1) - f_1(\mathbf{t}_1)\big)^2\big] = n\sum\nolimits_{a\in\{+,-\}}(\theta_{0,a} - \theta_{1,a})^2$$
$$= n\big(\theta_{0,+} - |\theta_{0,+}^2 - zn^{-1/2}|^{1/2}\big)^2 \le z^2(2\theta_{0,+})^{-2}.$$

To conclude, it suffices to use inequality (2.74) from (Tsybakov, 2009), which implies that $\gamma_n(\mathcal{F}_0, \mathcal{F}_1(r_n)) \ge 0.25e^{-z^2(2\theta_{0,+})^{-2}} = \gamma$ for $z = 2\theta_{0,+}[\ln(4\gamma)^{-1}]^{1/2}$.

It remains to prove the second assertion of the theorem. To ease notation, we write T_n instead of T_n^0 and set

$$Q_a[f] = \sum\nolimits_{l\in\mathcal{L}_a} q_l\theta_l^2[f] \quad \text{and} \quad \mathcal{F}_a = \{f : Q_a[f] = 0\}, \quad \text{for} \quad a \in \{+,-\}.$$

Let us assume that $M_+(T_n) \ge M_-(T_n)$. We use the fact that testing $Q[f] = 0$ against $|Q[f]| \ge r_n^2$, with $f \in \Sigma$ is harder than testing $Q_+[f] = 0$ against $Q_+[f] \ge r_n^2$, with $f \in \mathcal{F}_-$.

The rest of the proof follows the same steps as those of the proof of Theorem 2. As indicated in Remark 5, we use as π_n the simplified prior for which θ_l's are independent Gaussian random variables with zero mean and variance $a_l = \frac{q_l}{2T_nM_+(T_n)}\mathbf{1}_{\{l\in\mathcal{L}_+\cup\mathcal{N}(T_n)\}}$. It is an easy exercice to show that conditions **[L1]**-**[L5]** of Proposition 6 are fulfilled with $\delta = 1/2$. This completes the proof of the theorem.

References

Baraud, Y., Huet, S., and Laurent, B. Adaptive tests of linear hypotheses by model selection. *Ann. Statist.*, 31(1):225–251, 2003. MR1962505

Baraud, Y., Huet, S., and Laurent, B. Testing convex hypotheses on the mean of a Gaussian vector. Application to testing qualitative hypotheses on a regression function. *Ann. Statist.*, 33(1):214–257, 2005. MR2157802

Butucea, C. Goodness-of-fit testing and quadratic functional estimation from indirect observations. *Ann. Statist.*, 35(5):1907–1930, 2007. MR2363957

Butucea, C., Matias, C., and Pouet, C. Adaptive goodness-of-fit testing from indirect observations. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45 (2):352–372, 2009. MR2521406

Cohen, A. *Numerical analysis of wavelet methods*, volume 32 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 2003. MR1990555

Collier, O. Minimax hypothesis testing for curve registration. *Electron. J. Statist.*, 6:1129–1154, 2012.

Comminges, L. and Dalalyan, A. S. Tight conditions for consistent variable selection in high dimensional nonparametric regression. *Journal of Machine Learning Research - Proceedings Track*, 19:187–206, 2011.

Comminges, L. and Dalalyan, A. S. Tight conditions for consistency of variable selection in the context of high dimensionality. *Ann. Statist.*, 40(6), doi:10.1214/12-AOS1046, 2012.

Dalalyan, A. S. and Collier, O. Wilks' phenomenon and penalized likelihood-ratio test for nonparametric curve registration. *J. Mach. Learn. Res. - Proceedings Track*, 22:264–272, 2012.

Dalalyan, A. S., Juditsky, A., and Spokoiny, V. A new algorithm for estimating the effective dimension-reduction subspace. *J. Mach. Learn. Res.*, 9:1648–1678, 2008. MR2438819

Donoho, D. and Nussbaum, M. Minimax quadratic estimation of a quadratic functional. *J. Complexity*, 6(3):290–323, 1990. MR1081043

Efromovich, S. On the limit in the equivalence between heteroscedastic regression and filtering model. *Statist. Probab. Lett.*, 63(3):239–242, 2003. MR1986322

Ermakov, M. S. Minimax detection of a signal in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(4):704–715, 1990. MR1090496

Ermakov, M. S. Minimax detection of a signal in weighted Gaussian white noise. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 320(Veroyatn. i Stat. 8):54–68, 226, 2004. MR2115865

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

Fan, J. On the estimation of quadratic functionals. *Ann. Statist.*, 19(3):1273–1294, 1991. MR1126325

Gaïffas, S. and Lecué, G. Optimal rates and adaptation in the single-index model using aggregation. *Electron. J. Stat.*, 1:538–573, 2007. MR2369025

Gayraud, G. and Pouet, C. Minimax testing composite null hypotheses in the discrete regression scheme. *Math. Methods Statist.*, 10(4):375–394 (2002), 2001. Meeting on Mathematical Statistics (Marseille, 2000). MR1887339

Gayraud, G. and Pouet, C. Adaptive minimax testing in the discrete regression scheme. *Probab. Theory Related Fields*, 133(4):531–558, 2005. MR2197113

Hall, P. Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.*, 14(1):1–16, 1984. MR0734096

Horowitz, J. L. and Spokoiny, V. G. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, 69(3):599–631, 2001. MR1828537

Ingster, Yu. and Stepanova, N. Estimation and detection of functions from anisotropic Sobolev classes. *Electron. J. Stat.*, 5:484–506, 2011. MR2813552

Ingster, Yu. I. Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Math. Methods Statist.*, 2(2):85–114, 1993a. MR1257978

Ingster, Yu. I. Asymptotically minimax hypothesis testing for nonparametric alternatives. II. *Math. Methods Statist.*, 2(3):171–189, 1993b. MR1257983

Ingster, Yu. I. Asymptotically minimax hypothesis testing for nonparametric alternatives. III. *Math. Methods Statist.*, 2(4):249–268, 1993c. MR1259685

Ingster, Yu. I. and Sapatinas, T. Minimax goodness-of-fit testing in multivariate nonparametric regression. *Math. Methods Statist.*, 18(3):241–269, 2009. MR2560455

Ingster, Yu. I., Sapatinas, T., and Suslina, I. A. Minimax Signal Detection in Ill-Posed Inverse Problems. *Ann. Statist.*, 40(2):1524–1549, 2012.

Ingster, Yu. I. and Suslina, I. A. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2003. MR1991446

Kneser, H. Sur un théorème fondamental de la théorie des jeux. *C. R. Acad. Sci. Paris*, 234:2418–2420, 1952. MR0050249

Kolyada, V. I. On the embedding of Sobolev spaces. *Mat. Zametki*, 54(3):48–71, 158, 1993. MR1248284

Laurent, B., Loubes, J. M., and Marteau, C. Testing inverse problems: a direct or an indirect problem? *Journal of Statistical Planning and Inference*, 141 (5):1849–1861, 2011. MR2763215

Laurent, B., Loubes, J. M., and Marteau, C. Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. *Electronic Journal of Statistics*, 6:91–122, 2012. MR2879673

Lepski, O., Nemirovski, A., and Spokoiny, V. On estimation of the L_r norm of a regression function. *Probab. Theory Related Fields*, 113(2):221–253, 1999. MR1670867

Pinsker, M. S. Optimal filtration of square-integrable signals in Gaussian noise. *Problems Inform. Transmission*, (16):52–68, 1980. MR0624591

Pouet, C. An asymptotically optimal test for a parametric set of regression functions against a non-parametric alternative. *J. Statist. Plann. Inference*, 98(1-2):177–189, 2001. MR1860233

Samarov, A., Spokoiny, V., and Vial, C. Component identification and estimation in nonlinear high-dimensional regression models by structural adaptation. *J. Amer. Statist. Assoc.*, 100(470):429–445, 2005. MR2160548

Spokoiny, V. G. Adaptive hypothesis testing using wavelets. *Ann. Statist.*, 24(6):2477–2498, 1996. MR1425962

Tsybakov, A. B. *Introduction to nonparametric estimation*. Springer Verlag, 2009. MR2724359

Vershynin, R. Introduction to the non-asymptotic analysis of random matrices. *Compressed sensing*, 210–268, Cambridge Univ. Press, Cambridge, 2012. MR2963170