Appl. Comput. Harmon. Anal. 46 (2019) 288–311

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Minimax optimal procedures for testing the structure of multidimensional functions

John Aston [a], Florent Autin [b], Gerda Claeskens [c], Jean-Marc Freyermuth [a,*], Christophe Pouet [b]

[a] *Statistical Laboratory, University of Cambridge, Wilberforce Road Cambridge CB3 0WB, England, United Kingdom*
[b] *Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France*
[c] *ORSTAT and Leuven Statistics Research Center, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium*

A R T I C L E I N F O

Article history:
Received 1 June 2016
Received in revised form 22 March 2017
Accepted 1 May 2017
Available online 18 May 2017
Communicated by Christopher Heil

Keywords:
Adaptation
Anisotropy
Atomic dimension
Besov spaces
Gaussian noise model
Hyperbolic wavelets
Hypothesis testing
Minimax rate
Sobol decomposition
Structural modeling

A B S T R A C T

We present a novel method for detecting some structural characteristics of multidimensional functions. We consider the multidimensional Gaussian white noise model with an anisotropic estimand. Using the relation between the Sobol decomposition and the geometry of multidimensional wavelet basis we can build test statistics for any of the Sobol functional components. We assess the asymptotical minimax optimality of these test statistics and show that they are optimal in presence of anisotropy with respect to the newly determined minimax rates of separation. An appropriate combination of these test statistics allows to test some general structural characteristics such as the atomic dimension or the presence of some variables. Numerical experiments show the potential of our method for studying spatio-temporal processes.

© 2017 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Multidimensional data often exhibit a simpler underlying structure, meaning that their effective dimension is smaller than the observed dimension. Detecting the presence of such structure can enhance the understanding of the data and allows for more effective modeling and inferential strategies. There is a flourishing literature dealing with nonparametric methods for structure detection. These contributions are concerned with different types of structures (such as additivity, small atomic dimension and variable selection) and with different modeling approaches according to the nature of the noise, the smoothness

* Corresponding author.
E-mail address: jmfreyermuth1342@gmail.com (J.-M. Freyermuth).

http://dx.doi.org/10.1016/j.acha.2017.05.003
1063-5203/© 2017 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

assumptions for instance. A brief overview of the most directly related contributions are given hereafter. The main characteristic of this paper is to provide a rigorous theoretical study of structure detection in the presence of an anisotropically smooth estimand. We consider a multidimensional Gaussian white noise model and derive test statistics for the atomic dimension of multidimensional objects (i.e., the maximal degree of interaction between the variables) and for variable selection. Then, following the Sieve estimation of Birgé and Massart [5], we build a data-driven procedure. It is first tested in 'idealistic' numerical experiments before being applied to a more sophisticated context of time series data analysis.

The main ingredient of our methodology is to project the data onto an hyperbolic (wavelet) basis and to build test statistics based on the projection coefficients. Its motivation relies on the relation that emerges between the geometry of that basis and the 'functional components' of the Sobol decomposition of the estimand (see Sobol [34]). The use of an appropriate basis allows to benefit from a sparse representation and an optimal adaptation to the anisotropic smoothness of the estimand. In this paper we construct optimal testing procedures for the functional components accordingly to the minimax hypothesis testing framework of Ingster [11–13]. Appropriately combined these functional components can be rephrased in terms of more general structures such as the atomic dimension.

This project is motivated by the recent results of Autin et al. [3,4] who studied the ability of a tensor-product wavelet basis, the so-called hyperbolic wavelet basis to estimate multidimensional functions having anisotropic smoothness. Tensor-product bases (Fourier or wavelets) have already been widely used in signal detection notably to test for additivity (i.e. atomic dimension equals to 1) as in De Canditiis and Sapatinas [8]. Nevertheless, in the latter they do not provide deep theoretical results on the performance of their method. Abramovich et al. [1] describe another procedure for testing additivity. They propose an adaptive procedure based on the thresholding wavelet coefficients and derive interesting theoretical results. Nevertheless, they exploit a standard (isotropic) wavelet basis that cannot optimally deal with the more realistic situation of having anisotropic estimands. Moreover, their method is limited to test for additivity and does not exploit the full directional representation of a multidimensional wavelet basis. The functional framework (where the dimension $d \to \infty$) has also been investigated by Gayraud and Ingster [10] who studied the problem of signal detection in the case of sparse additive functions. Other developments include, for example, multichannel signal detection as in Ingster and Suslina [18], and detection in inverse models as in Ingster et al. [14].

Comminges and Dalalyan [6] made profound theoretical contributions on hypothesis testing procedures based on quadratic functionals. They study various testing problems involving multidimensional anisotropic functions. They exploit a tensor-product Fourier basis to build non adaptive procedures, i.e. procedures that are built on the knowledge of the regularity parameters of the anisotropic function spaces in the alternative. In this paper, we focus on the consequences of dealing with anisotropic estimands. Therefore, we define two classes of testing procedures, referred to as minimax optimal and adaptive minimax optimal methods for the Besov balls, see Section 4. The former methodology is built using the full knowledge of the regularity parameters while the latter uses only a part of it.

These methods differ by the nature of the truncation applied to the hyperbolic wavelet coefficients sequence. An intuitive way to think about this is to consider the better known multidimensional function estimation context. The truncation rule characterizing the minimax optimal method can be viewed as a linear but directionally dependent smoothing. While the truncation rule of the adaptive minimax optimal method finds its roots in the approximation of the hyperbolic cross that is naturally associated to tensor-product spaces (see for instance Schmeisser [31], Schmeisser and Sickel [32] and Sickel and Ullrich [33]). Tensor-product spaces are also often considered in various statistical contexts such as in estimation with for example the tensor-product space ANOVA model as in Lin [24] or in signal detection as in Ingster and Stepanova [15]. They are of great interest since one can hope to reach performances close to the unidimensional case. It is sometimes promoted as a way to 'tackle' the curse of dimensionality, nevertheless it is a restrictive assumption. In the context of functional ANOVA, for example, assuming a tensor-product

model means that the interaction term's complexity has to be inversely proportional to its degree. In this paper we do not restrict to the structure detection of functions belonging to tensor-product of Besov spaces but to larger classes such as the anisotropic Nikolskii class. We show that our methods attain the optimal separation rates between the null and the alternative hypothesis and we discuss how to combine sets of testing procedures in a way that is adapted to the anisotropic case.

Such testing procedures find applications in many fields. We illustrate its usage in time series data analysis. Our aim is to test the structure of the spatio-spectrum associated to a spatio-temporal process that is time stationary. This allows us also to illustrate how to adapt such testing procedure in the context of multidimensional data where the number of observations is very large, making p-values useless.

This paper is structured as follows. First, in Section 2, we give the fundamental knowledge on hyperbolic wavelet bases and their relation to the Sobol decomposition of a multidimensional function. In Section 3 we introduce the multidimensional Gaussian white noise model and the minimax hypothesis testing framework. Then we describe our test statistics for Sobol functional components in Section 4 before introducing tests for global structural characteristics in Section 5. Section 6 describes our data-driven adaptation and validation in a practical setting of the adaptive minimax optimal procedure and a final discussion is provided in Section 7. The proofs of the theoretical results are postponed to the appendix.

2. Hyperbolic wavelet bases

We start from a one-dimensional periodized wavelet basis $\mathcal{B}_1$ of $L_2([0,1))$ which is built from the dilations and translations of a scaling function ϕ and a wavelet ψ with V (for some $V \geq 1$) vanishing moments,

$$\mathcal{B}_1 = \left\{ \phi(.),\ \psi_{j,k}(.) = 2^{j/2}\psi(2^j. - k) :\ j \in \mathbb{N},\ k \in \{0, \dots, 2^j - 1\} \right\}.$$

A d-dimensional hyperbolic wavelet basis results by forming d-variate functions taking appropriate products of the univariate functions ϕ and ψ as follows,

$$\mathcal{B}_d = \left\{ \phi_{\underline{0},\underline{0}}(.),\ \psi^{\underline{i}}_{\underline{j},\underline{k}}(.) : \quad \underline{i} \in \{0,1\}^d \setminus \underline{0},\ \underline{j} \in \mathbb{J}^{\underline{i}},\ \underline{k} \in \mathbb{K}_{\underline{j}} \right\}$$

where $\underline{0} = (0, \dots, 0)$, $\phi_{\underline{0},\underline{0}}(.) = \phi(.) \times \cdots \times \phi(.)$, $\psi^{\underline{i}}_{\underline{j},\underline{k}}(.) = \psi^{i_1}_{j_1,k_1}(.) \times \cdots \times \psi^{i_d}_{j_d,k_d}(.)$, $\underline{i} = (i_1, \dots, i_d)$ with the following notations:

$$\psi^{i_u}_{j_u,k_u}(\cdot) = \begin{cases} 2^{j_u/2}\phi(2^{j_u} \cdot - k_u) & \text{if } \ i_u = 0 \\ 2^{j_u/2}\psi(2^{j_u} \cdot - k_u) & \text{if } \ i_u = 1 \end{cases}$$

and

$$\mathbb{J}^{\underline{i}} = \left\{ \underline{j} = (j_1, \dots, j_d) : \quad \forall u \in \{1, \dots, d\},\ j_u = j'_u i_u,\ j'_u \in \mathbb{N} \right\},$$
$$\mathbb{K}_{\underline{j}} = \left\{ \underline{k} = (k_1, \dots, k_d) : \quad \forall u \in \{1, \dots, d\},\ k_u \in \{0, \dots, 2^{j_u} - 1\} \right\}.$$

The basis $\mathcal{B}_d$ is generated by a set of a scaling function $\phi_{\underline{0},\underline{0}}$ and translated and dilated wavelet functions $\psi^{\underline{i}}_{\underline{j},\underline{k}}$. If all the components of the vector of directional dilations $\underline{j}$ are equal, this results in the standard (isotropic) wavelet basis. Hereafter we consider that the components of $\underline{j}$ can take different values. The wavelet functions are then supported on hyper-rectangles and the resulting wavelet basis, the so-called hyperbolic or tensor-product wavelet basis, is proven to be able to optimally deal with anisotropic functions [3,4]. We say that each of the $2^d - 1$ elements of $\underline{i}$ defines an orientation. In such a basis, any $f \in L_2([0,1)^d)$ can be decomposed as follows:

$$f = \langle f, \phi_{\underline{0},\underline{0}} \rangle_2 \phi_{\underline{0},\underline{0}} + \sum_{\underline{i} \neq \underline{0}} \left(\sum_{\underline{j} \in \mathbb{J}^{\underline{i}}} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \langle f, \psi^{\underline{i}}_{\underline{j},\underline{k}} \rangle_2 \psi^{\underline{i}}_{\underline{j},\underline{k}} \right) \equiv f_{\underline{0}} + \sum_{\underline{i} \neq \underline{0}} f_{\underline{i}} \tag{1}$$

where $\langle ., . \rangle_2$ is the L_2-scalar product. This representation facilitates a characterization of a specific structure, such as additivity. Indeed, for an additive function f_{add}, that is a function with Sobol decomposition $f_{\text{add}}(x_1, \ldots, x_d) = f_{\underline{0}} + \sum_{u=1}^{d} f_u(x_u)$, the coefficients $\theta^{\underline{i}}_{\underline{j},\underline{k}} = \langle f, \psi^{\underline{i}}_{\underline{j},\underline{k}} \rangle_2$ in each orientation $\underline{i}$ with $|\underline{i}| = i_1 + i_2 + \cdots + i_d > 1$ are exactly equal to 0. Information about the component functions $f_{\underline{i}}$ in equation (1) can be retrieved via the wavelet coefficient sequence through the geometry of the hyperbolic wavelet basis. This enables the design of specific testing procedures for structural information for multivariate functions, more general than merely testing for additivity.

Dalalyan et al. [7] introduced the atomic dimension in the physical domain that we denote as $\delta(f)$. It reflects the maximal degree of interaction between the d variables in the Sobol decomposition. For instance, the additive model f_{add} has $\delta(f_{\text{add}}) = 1$, while a function $f(x_1, \ldots, x_d) = \sum_{u=1}^{d} f_u(x_u) + \sum_{u=1}^{d} \sum_{v=1, v \neq u}^{d} x_u x_v$ has $\delta(f) = 2$. In the sequel, we use the definition from Autin et al. [3] of the atomic dimension in the wavelet coefficient domain relating the orientations of the nonzero wavelet coefficients with corresponding Sobol functional components.

Definition 2.1. Let $f \in L_2([0,1)^d)$ be decomposed in the hyperbolic wavelet basis as

$$f = f_{\underline{0}} + \sum_{\underline{i} \neq \underline{0}} f_{\underline{i}} = \theta^{\underline{0}}_{\underline{0},\underline{0}} \phi_{\underline{0},\underline{0}} + \sum_{i \notin \underline{0}} \left(\sum_{j \in \mathbb{J}^{\underline{i}}} \sum_{k \in \mathbb{K}_{\underline{j}}} \theta^{\underline{i}}_{\underline{j},\underline{k}} \psi^{\underline{i}}_{\underline{j},\underline{k}} \right).$$

Define $\mathcal{A}_f = \{\underline{i} \in \{0,1\}^d \setminus \underline{0};\ \theta^{\underline{i}}_{\underline{j},\underline{k}} \neq 0 \text{ for some } (\underline{j}, \underline{k})\}$. The atomic dimension of f with respect to $\mathcal{B}_d$ is the integer $\delta_{\mathcal{B}_d} = \delta_{\mathcal{B}_d}(f) \in \{0, \ldots, d\}$ such that:

$$\delta_{\mathcal{B}_d} = \begin{cases} \max\{|\underline{i}|; \underline{i} \in \mathcal{A}_f\} & \text{if } \mathcal{A}_f \neq \emptyset \\ 0 & \text{if } \mathcal{A}_f = \emptyset. \end{cases}$$

Definition 2.1 gives us more flexibility than its definition in the physical domain. Indeed, the atomic dimension $\delta_{\mathcal{B}_d}$ depends on the number of vanishing moment (even directional ones) of the considered basis $\mathcal{B}_d$. More precisely, the atomic dimension $\delta(f)$ of a d-dimensional function f in the physic domain and the atomic dimension $\delta_{\mathcal{B}_d}(f)$ of the same function f in the coefficient domain are clearly tied by the following inequality: $\delta_{\mathcal{B}_d}(f) \leq \delta(f)$. We mention here two situations to emphasize the practical importance of being able to characterize the atomic dimension in the coefficient domain. The first example concerns multidimensional function estimation. Autin et al. [3] introduced a novel estimation procedure based on the thresholding of the hyperbolic wavelet coefficients. They show that it outperforms the standard hard thresholding whenever some structural conditions on the estimand are satisfied. These conditions also include some constraint on the value of the atomic dimension $\delta_{\mathcal{B}_d}$. Here we provide a test statistic to test the structure of the estimand before applying the aforementioned estimation method. A second example concerns the exploitation of sparsity in statistical modeling that can be useful for building a statistical model using the hyperbolic wavelet coefficients or in generalized linear modeling. For example, Zhou et al. [38] describe such a generalized linear model with multidimensional covariates, their first step consists in dimension reduction by tensor decomposition. One could instead exploit the sparsity of the hyperbolic wavelet sequence. In such a case, one may seek for the hyperbolic wavelet basis that provides the sparsest situation.

3. Model and minimax approach for hypothesis testing

The observed signal under the Gaussian white noise model is a realization of the process

$$dY_\varepsilon(\underline{x}) = f(\underline{x})d\underline{x} + \varepsilon dW(\underline{x}) \tag{2}$$

where $\underline{x} = (x_1, \dots, x_d) \in [0,1)^d$, $f \in L_2([0,1)^d)$, $W(\underline{x})$ is the Brownian sheet and ε is the noise level, considered here to be close to zero. For technical reasons (see the proofs of Proposition A.3 and Proposition A.4), without loss of generality, we assume that it is smaller than $\varepsilon_o = e^{-e}$, where e denotes the Euler's number. We consider the sequential version of the Gaussian white noise model which consists of the observations of the following random variables

$$\hat{\theta}^{\underline{0}}_{\underline{0},\underline{0}} = \theta^{\underline{0}}_{\underline{0},\underline{0}} + \varepsilon\xi = \langle f, \phi_{\underline{0},\underline{0}}\rangle_{L_2} + \varepsilon\xi \quad \text{and} \quad \hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}} = \theta^{\underline{i}}_{\underline{j},\underline{k}} + \varepsilon\xi^{\underline{i}}_{\underline{j},\underline{k}} = \langle f, \psi^{\underline{i}}_{\underline{j},\underline{k}}\rangle_{L_2} + \varepsilon\xi^{\underline{i}}_{\underline{j},\underline{k}}$$

where, $\xi, \xi^{\underline{i}}_{\underline{j},\underline{k}}$ are i.i.d. $\mathcal{N}(0,1)$ and $(\underline{j},\underline{k}) \in \mathbb{N}^d \times \mathbb{K}_{\underline{j}}$.

For any chosen orientation $\underline{i} \neq \underline{0}$, we may test the following null hypothesis $\mathcal{H}_{\underline{i},0}$ versus one of the two stated alternative hypotheses $\mathcal{H}_{\underline{i},a}$ and $\mathcal{H}'_{\underline{i},a}$:

$$\mathcal{H}_{\underline{i},0} : f \in \mathcal{N}_{\underline{i}}(R) = \left\{ f : \|f\|_2 \leq R,\ f_{\underline{i}}(\underline{x}) = \sum_{j\in\mathbb{J}^{\underline{i}}} \sum_{k\in\mathbb{K}_{\underline{j}}} \theta^{\underline{i}}_{\underline{j},\underline{k}} \psi^{\underline{i}}_{\underline{j},\underline{k}}(\underline{x}) = 0,\ \forall \underline{x} \in [0,1)^d \right\}, \tag{3}$$

$$\mathcal{H}_{\underline{i},a} : f \in \mathcal{A}_{\underline{i}}(R, C, \underline{s}, r_{\varepsilon,\underline{i}}) = \left\{ f : f \in \mathcal{F}^{\underline{s}}(R),\ \|f_{\underline{i}}\|_2 = \left(\sum_{j\in\mathbb{J}^{\underline{i}}} \sum_{k\in\mathbb{K}_{\underline{j}}} \left(\theta^{\underline{i}}_{\underline{j},\underline{k}}\right)^2 \right)^{\frac{1}{2}} \geq C r_{\varepsilon,\underline{i}} \right\}, \tag{4}$$

$$\mathcal{H}'_{\underline{i},a} : f \in \cup_{\underline{s}:\ \sum_u i_u s_u^{-1} = \gamma_{\underline{i}}^{-1}} \mathcal{A}_{\underline{i}}(R, C, \underline{s}, r'_{\varepsilon,\underline{i}}). \tag{5}$$

In the above expression, C is a positive constant which does not depend on ε and $r_{\varepsilon,\underline{i}}$, $r'_{\varepsilon,\underline{i}}$ are continuous sequences of positive real numbers tending to 0 as ε goes to 0. $\mathcal{F}^{\underline{s}}(R)$ denotes a ball of radius R in a function space characterized by a vector of directional regularities $\underline{s}$ with harmonic sum in the orientation $\underline{i}$ denoted as $\gamma_{\underline{i}}^{-1}$.

Given a simultaneous control on the probabilities of type I and II errors, we consider the asymptotic minimax setup: we aim at providing, for any orientation $\underline{i}$, the order of the optimal separation rates in the sense of Section 2.6 in Ingster and Suslina [17] between the null hypothesis $\mathcal{H}_{\underline{i},0}$ and each one of the alternative hypotheses $\mathcal{H}_{\underline{i},a}$ and $\mathcal{H}'_{\underline{i},a}$ associated to a particular choice of $\mathcal{F}^{\underline{s}}$ when:

- $\underline{s}$ is known;
- only the harmonic sum $\gamma_{\underline{i}}^{-1} = \sum_u i_u s_u^{-1}$ of $\underline{s}$ with respect to orientation $\underline{i}$ is known.

Related to this, we search for both a sequence of $\mathcal{F}^{\underline{s}}$-minimax optimal testing procedures and a sequence of $\mathcal{F}^{\underline{s}}$-adaptive minimax optimal testing procedures.

Definition 3.1. Let $\alpha \in \left(0, \frac{1}{2}\right)$, $\underline{s} \in (0, +\infty)^d$ and $\underline{i} \in \{0,1\}^d \setminus \underline{0}$. We say that $r_{\varepsilon,\underline{i}}$ corresponds to the $\mathcal{F}^{\underline{s}}$-minimax rate separating the hypotheses $\mathcal{H}_{\underline{i},0}$ and $\mathcal{H}_{\underline{i},a}$, up to a constant, if the two following statements are satisfied:

1. (Upper bound) for any $C > C_{\underline{i},\alpha}$, with $C_{\underline{i},\alpha}$ an absolute positive constant, there exists a sequence of testing procedures $\left(\Delta^*_{\varepsilon,\alpha}\right)_\varepsilon$ such that

$$\sup_{0<\varepsilon<\varepsilon_o} \left(\sup_{f\in\mathcal{N}_{\underline{i}}(R)} \mathbb{P}_f(\Delta^*_{\varepsilon,\alpha} = 1) \quad + \sup_{f\in\mathcal{A}_{\underline{i}}(R,C,\underline{s},r_{\varepsilon,\underline{i}})} \mathbb{P}_f(\Delta^*_{\varepsilon,\alpha} = 0) \right) \leq \alpha,$$

2. (Lower bound) for any $C < c_{\underline{i},\alpha}$, with $c_{\underline{i},\alpha}$ an absolute positive constant, the following inequality holds

$$\inf_{0<\varepsilon<\varepsilon_o} \inf_{\Delta} \left(\sup_{f\in\mathcal{N}_{\underline{i}}(R)} \mathbb{P}_f(\Delta=1) \quad + \sup_{f\in\mathcal{A}_{\underline{i}}(R,C,\underline{s},r_{\varepsilon,\underline{i}})} \mathbb{P}_f(\Delta=0) \right) > \alpha.$$

The sequence of testing procedures $\left(\Delta^*_{\varepsilon,\alpha}\right)_\varepsilon$ is said to be $\mathcal{F}^{\underline{s}}$-minimax optimal.

Definition 3.2. Let $\alpha \in \left(0, \frac{1}{2}\right)$. Consider $\underline{i} \in \{0,1\}^d$ such that $|\underline{i}| > 1$ and let $\gamma_{\underline{i}} > 0$. We say that $r'_{\varepsilon,\underline{i}}$ corresponds to the $\mathcal{F}^{\underline{s}}$-adaptive minimax rate separating the hypotheses $\mathcal{H}_{\underline{i},0}$ and $\mathcal{H}'_{\underline{i},a}$, up to a constant, if the two following statements are satisfied:

1. (Upper bound) for any $C > C'_{\underline{i},\alpha}$, with $C'_{\underline{i},\alpha}$ an absolute positive constant, there exists a sequence of testing procedure $\left(\Delta^\star_{\varepsilon,\alpha}\right)_\varepsilon$ such that

$$\sup_{0<\varepsilon<\epsilon_o} \left(\sup_{f\in\mathcal{N}_{\underline{i}}(R)} \mathbb{P}_f(\Delta^\star_{\varepsilon,\alpha}=1) \quad + \sup_{\underline{s}:\ \sum_u i_u s_u^{-1}=\gamma_{\underline{i}}^{-1}} \sup_{f\in\mathcal{A}_{\underline{i}}\left(R,C,\underline{s},r'_{\varepsilon,\underline{i}}\right)} \mathbb{P}_f(\Delta^\star_{\varepsilon,\alpha}=0) \right) \le \alpha,$$

2. (Lower bound) for any $C < c'_{\underline{i},\alpha}$, with $c'_{\underline{i},\alpha}$ an absolute positive constant, the following inequality holds

$$\inf_{0<\varepsilon<\varepsilon_o} \inf_{\Delta} \left(\sup_{f\in\mathcal{N}_{\underline{i}}(R)} \mathbb{P}_f(\Delta=1) \quad + \sup_{\underline{s}:\ \sum_u i_u s_u^{-1}=\gamma_{\underline{i}}^{-1}} \sup_{f\in\mathcal{A}_{\underline{i}}\left(R,C,\underline{s},r'_{\varepsilon,\underline{i}}\right)} \mathbb{P}_f(\Delta=0) \right) > \alpha.$$

The sequence of testing procedures $\left(\Delta^\star_{\varepsilon,\alpha}\right)_\varepsilon$ is said to be $\mathcal{F}^{\underline{s}}$-adaptive minimax optimal.

In the latter situation, we have only the information about the harmonic sum of the directional smoothness parameters of the function class in the alternative. To determine the minimax and adaptive minimax separation rates, we naturally consider the combination of directional smoothness with a given harmonic sum that is giving the worst type II error rate. The fully adaptive case corresponds to an unknown value $\gamma_{\underline{i}}$. It is a direct extension studied here which could be called partially adaptive, that is to say when $\gamma_{\underline{i}}$ is known. In this work for the sake of simplicity we use adaptive instead of partially adaptive.

Remark 3.1. The definition of minimax and adaptive minimax rates are equivalent to the ones given in Section 1 of Lepski and Tsybakov [22] (Section 1) and slightly more precise than the ones given in Section 2.6 of Ingster and Suslina [17]. We will show that both the minimax separation rate and the adaptive minimax separation rate of testing depend on the smoothness of the underlying functional space and on the noise level ε. At the end of the appendix, we highlight the relationships between the absolute constants $C_{\underline{i},\alpha}, c_{\underline{i},\alpha}, C'_{\underline{i},\alpha}, c'_{\underline{i},\alpha}$ and the parameters appearing in the problem such as R and the smoothness.

4. Nonparametric tests for a given orientation

In this paragraph we present test statistics and testing procedures that are used in the sequel. These test statistics are going to provide sequences of testing procedures that are $\mathcal{F}^{\underline{s}}$-minimax and $\mathcal{F}^{\underline{s}}$-adaptive minimax optimal in the particular case where $\mathcal{F}^{\underline{s}}$ corresponds to a Besov space with $\underline{s}$ as smoothness parameter.

Definition 4.1 *(Test statistic).* Let $\underline{i} \neq \underline{0}$ and $\underline{j} = (j_1, \dots, j_d) \in \mathbb{J}^{\underline{i}}$. The nonnegative random variable

$$T_{\underline{i},\underline{j}} = \sum_{\underline{j}'\in\mathbb{J}^{\underline{i}}:\ j'_u<\max(j_u,1),\forall u} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}'}} (\hat{\theta}^{\underline{i}}_{\underline{j}',\underline{k}})^2$$

is called the $\underline{i}$-oriented test statistic with level parameter $\underline{j}$.

Definition 4.2 *(Testing procedure).* Let $\underline{i} \neq \underline{0}$, $\underline{j} = (j_1, \dots, j_d) \in \mathbb{J}^{\underline{i}}$ and $t > 0$. The random variable

$$\Delta_{\underline{i},\underline{j}}(t) = 1_{\{T_{\underline{i},\underline{j}} > t\}}$$

is called the $(\underline{i}, \underline{j}, t)$-testing procedure.

4.1. $B^{\underline{s}}_{2,\infty}$-minimax optimal sequence of testing procedures

We define appropriate test statistics based on the estimated empirical energy that is the sum of the squared values of the empirical hyperbolic wavelet coefficients within the specified orientation over certain scales. We show that this sequence of tests constructed using hyperbolic wavelet bases gets optimal minimax properties when f belongs to a Besov ball with smoothness parameter $\underline{s}$.

Definition 4.3 *(Besov ball).* Let $R > 0$ and $\underline{s} = (s_1, \dots, s_d) \in (0, +\infty)^d$. We say that $f \in L_2([0,1)^d)$ belongs to the Besov ball $B^{\underline{s}}_{2,\infty}(R)$ if and only if

$$\sup_{\underline{i} \neq \underline{0}} \sup_{\underline{j} \in \mathbb{J}^{\underline{i}}} \max_{1 \leq u \leq d} \{2^{2 j_u s_u}\} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^2 \leq R^2.$$

In this section we assume that we have the full knowledge about the smoothness parameter $\underline{s}$ of the functions in the alternative. In other words it corresponds to the non adaptive set up. We define a vector $\underline{j}^* = (j_1^*, \dots, j_d^*) \in \mathbb{J}^{\underline{i}}$ of truncation scales such that

$$2^{-j_u^*} \leq \varepsilon^{4 i_u \gamma_{\underline{i}}/(1+4\gamma_{\underline{i}}) s_u} < 2^{1-j_u^*} \tag{6}$$

with $\gamma_{\underline{i}} = (\sum_{u=1}^d i_u s_u^{-1})^{-1}$.

Theorem 4.1. *Let $\alpha \in \left(0, \frac{1}{2}\right)$, $R > 0$, $\underline{s}$ in $(0, +\infty)^d$ and $\underline{i} \in \{0,1\}^d \setminus \underline{0}$. Consider, as in (3) and (4), testing the hypotheses $\mathcal{H}_{\underline{i},0}$ versus $\mathcal{H}_{\underline{i},a}$ where $\mathcal{F}^{\underline{s}}(R) = B^{\underline{s}}_{2,\infty}(R)$. Then the sequence of the $(\underline{i}, \underline{j}^*, t_{\varepsilon,\underline{i},\alpha})$-testing procedures $\left(\Delta_{\underline{i},\underline{j}^*}(t_{\varepsilon,\underline{i},\alpha})\right)_\varepsilon$ is $B^{\underline{s}}_{2,\infty}$-minimax optimal with $\underline{j}^*$ as in (6), and with $\varepsilon^{-2} t_{\varepsilon,\underline{i},\alpha} = \chi_{1-\alpha/2}(2^{|\underline{j}^*|})$, the $1-\alpha/2$ quantile of the Chi-Square distribution with $2^{|\underline{j}^*|} = 2^{j_1^* + \dots + j_d^*}$ degrees of freedom. Moreover the $B^{\underline{s}}_{2,\infty}$-minimax rate separating $\mathcal{H}_{\underline{i},0}$ and $\mathcal{H}_{\underline{i},a}$ is of order of*

$$r_{\varepsilon,\underline{i}} = \varepsilon^{\frac{4\gamma_{\underline{i}}}{1+4\gamma_{\underline{i}}}}.$$

The proof of Theorem 4.1 is stated in the appendix. The absolute constants $C_{\underline{i},\alpha}$ and $c_{\underline{i},\alpha}$ from Definition 3.1 will be given there.

This theorem can be compared to the results given in the book of Ingster and Suslina [17] for Besov bodies in the one-dimension case. In this book, Theorem 3.9 states distinguishability conditions and sharp asymptotics are given in Theorem 4.4. The results stated above in Theorem 4.1 is similar but in any dimension; therefore the smoothness parameter is replaced by the average smoothness value obtained from the harmonic sum.

The reader could note that the worst rate is associated with $\gamma_{\underline{1}}$ with $\underline{1} = (1, \dots, 1)$, i.e. to the interaction term of highest degree. This case is similar to the one considered in Ingster and Stepanova [16] (Remark 3.3). We obtain the same minimax rate of testing hypotheses involving the harmonic sum of smoothness as the one they found for a signal function f in a Sobolev space. Nevertheless we can point that their null hypothesis

$H_0 : f = 0$ is different from the one considered here and briefly reformulated as $\mathcal{H}_{\underline{1},0} : f_{\underline{1}} = 0$ which consists of functions with an atomic dimension strictly smaller than the dimension d. The separation rate can be different for every orientation. When it comes to test for the structure of a d-dimensional function, a sequential testing procedure based on these different test statistics seems to be ideal because one can benefit from better separation rates.

Remark 4.1. In the context of multidimensional function estimation, the maxiset approach introduced by Kerkyacharian and Picard [20] has been proved useful to study the performance of wavelet-based estimators [3]. It consists in determining the largest function space such that the risk of an estimator is controlled at a chosen rate. If the choice of the rate is often an issue, it is common to use the minimax or near-minimax rates over some 'standard' functional spaces $\mathcal{F}$. Then the maxiset approach reveals a set of functions that contains $\mathcal{F}$ strictly or not. This is an optimistic approach in the sense that finally, we can estimate at the minimax or near-minimax rate more functions that previously thought. In minimax hypothesis testing problems, we can get inspired from the maxiset approach to remark that under the alternative, we naturally model the 'estimand' as a function in a d-variate smoothness space, here $B^{\underline{s}}_{2,\infty}(R)$. Nevertheless, since we are only concerned by the presence of information along given directions, the behavior along the directions that are not of interest can be let 'uncontrolled'. Hence, the sequence space in the alternative can be enlarged to the set of all Besov spaces $B^{\underline{s}'}_{2,\infty}(R)$ with $s'_u \leq s_u$ for $i_u = 0$ and $s'_u = s_u$ otherwise.

4.2. $B^{\underline{s}}_{2,\infty}$-adaptive minimax procedure

We now suppose that the regularity parameter $\underline{s}$ appearing in the alternative hypothesis is unknown but its harmonic sum $\gamma_{\underline{i}}^{-1} = \sum_{u=1}^{d} i_u s_u^{-1}$ is known for a chosen $\underline{i}$ such that $|\underline{i}| > 1$. In some sense, we consider now the adaptive setup.

To define the test statistic using the knowledge about γ, we denote by $J_{\underline{i}}$ the integer such that

$$2^{-J_{\underline{i}}} \leq \left(\varepsilon^4 \log\log \varepsilon^{-1}\right)^{\frac{1}{1+4\gamma_{\underline{i}}}} < 2^{1-J_{\underline{i}}}. \tag{7}$$

Theorem 4.2. *Let $\alpha \in \left(0, \frac{1}{2}\right)$, $R > 0$, $\underline{i} \in \{0,1\}^d$ satisfying $|\underline{i}| > 1$ and let $\gamma_{\underline{i}} > 0$. Consider, as in (3) and (5), testing the hypotheses $\mathcal{H}_{\underline{i},0}$ versus $\mathcal{H}'_{\underline{i},a}$ where $\mathcal{F}^{\underline{s}}(R) = B^{\underline{s}}_{2,\infty}(R)$, where $\gamma_{\underline{i}}^{-1} = \sum_{u=1}^{d} i_u s_u^{-1}$. Then the sequence of testing procedures $\left(\Delta_{\underline{i},max}(t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha})\right)_{\varepsilon}$ built from the maximum of the $(\underline{i}, \underline{j}, t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha})$-testing procedures $\Delta_{\underline{i},\underline{j}}(t'_{\varepsilon,\underline{i},\gamma,\alpha})$ for which $\underline{j}$ is satisfying $|\underline{j}| = J_{\underline{i}}$ is $B^{\underline{s}}_{2,\infty}$-adaptive minimax optimal provided that, for any $\varepsilon \in (0, \varepsilon_o)$, $\varepsilon^{-2} t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha} = \chi_{1-\alpha/2K_{\gamma_{\underline{i}}}}(2^{J_{\underline{i}}})$, with $K_{\gamma_{\underline{i}}} = \#\{\underline{j} \in \mathbb{J}^{\underline{i}} : |\underline{j}| = J_{\underline{i}}\}$. Moreover the $B^{\underline{s}}_{2,\infty}$-adaptive minimax rate separating $\mathcal{H}_{\underline{i},0}$ and $\mathcal{H}_{\underline{i},a}$ is of order of*

$$r'_{\varepsilon,\underline{i}} = \left(\varepsilon^4 \log\log \varepsilon^{-1}\right)^{\frac{\gamma_{\underline{i}}}{1+4\gamma_{\underline{i}}}}.$$

The proof of Theorem 4.2 is stated in the appendix. The absolute constants $C'_{\underline{i},\alpha}$ and $c'_{\underline{i},\alpha}$ from Definition 3.2 will be given there.

Results given for adaptation in Theorem 4.2 can be compared to usual results obtained for adaptation in hypothesis testing problems. The unavoidable loss due to adaptation is the usual one that can also be found for example in Theorem 7.2 in Ingster and Suslina [17]. The results stated above in Theorem 4.2 can be compared to the loss obtained by Spokoiny [35] in the Gaussian white noise model in one-dimension for the signal detection problem.

Remark 4.2. Note that compared to the $B^{\underline{s}}_{2,\infty}$-minimax case, the optimal separation rate is slower because $\log\log \varepsilon^{-1} \to \infty$ as $\varepsilon \to 0$. This attests of the deterioration of the information about the function space in

the alternative. Nevertheless, in the alternative hypothesis, the union of sequence spaces $B^{\underline{s}}_{2,\infty}(R)$ can be replaced by another larger sequence space, denoted $A_{\gamma_{\underline{i}},2}(R)$ and defined hereafter.

Definition 4.4 *(Truncation ball).* Let $R > 0$, $\underline{i} \in \{0,1\}^d$ satisfying $|\underline{i}| > 1$ and $\gamma_{\underline{i}} > 0$. We say that $f \in L_2([0,1)^d)$ belongs to the truncation ball $A_{\gamma_{\underline{i}},2}(R)$ if and only if

$$\sup_{\underline{j}\in\mathbb{J}^{\underline{i}}} 2^{2|\underline{j}|\gamma_{\underline{i}}} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} (\theta^{\underline{i}}_{\underline{j},\underline{k}})^2 \le R^2.$$

The truncation ball $A_{\gamma_{\underline{i}},2}(R)$ is similar to a Besov ball in the sense that it controls the decay of the energy of the hyperbolic wavelet coefficients over the scales and hence it is used to control the approximation error. Following Lemma 2.2 of Neumann [26], the following inclusion property holds.

Proposition 4.1. *Fix $\underline{i} \in \{0,1\}^d \setminus \underline{0}$. Let $\gamma_{\underline{i}} > 0$. For any $R > 0$, there exists $R' > 0$ such that*

$$\bigcup_{\underline{s}:\ i_1 s_1^{-1}+\ldots i_d s_d^{-1}=\gamma_{\underline{i}}^{-1}} B^{\underline{s}}_{2,\infty}(R) \subset A_{\gamma_{\underline{i}},2}(R').$$

Considering the functional component $f_{\underline{i}}$ of f does not only enable us to explore fine structures but it is also a way to improve the performance of the testing procedures for more general structure, see Section 6.

5. Tests that combine different orientations

Many interesting hypotheses on the structure of a function can be tested by 'aggregating' the previously described optimal testing procedures for the Sobol functional components. In this paragraph we describe such results in the context of the $B^{\underline{s}}_{2,\infty}$-minimax framework, but it can be easily given for the $B^{\underline{s}}_{2,\infty}$-adaptive minimax setting as well. We first state the following proposition:

Proposition 5.1. *[Max-testing] Fix $\alpha \in \left(0, \frac{1}{2}\right)$, $\underline{s} \in (0,+\infty)^d$ and consider $\beta = 1 - \left(1 - \frac{\alpha}{2}\right)^{|\mathcal{I}|}$ such that $|\mathcal{I}|$ denotes the cardinality of a nonempty set of orientations $\mathcal{I}$ that characterizes the alternative hypothesis.*

Consider the following testing hypotheses

$$\begin{aligned} \mathcal{H}_0 : f \in \mathcal{S}(R,\mathcal{I}) &= \{f : \|f\|_2 \le R, f_{\underline{i}}(\underline{x}) = 0,\ \forall \underline{i} \in \mathcal{I},\quad \forall \underline{x} \in [0,1)^d\}, \\ \mathcal{H}_a : f \in \mathcal{T}(R,\mathcal{I}) &= \{f : f \in B^{\underline{s}}_{2,\infty}(R), \exists\ \underline{i} \in \mathcal{I}, \|f_{\underline{i}}\|_2 \ge C r_{\varepsilon,\underline{i}}\}. \end{aligned} \tag{8}$$

Then, for any $\varepsilon \in (0,\varepsilon_o)$, the testing procedure $\Lambda_{\mathcal{I},\varepsilon,\alpha} = \max_{\underline{i}\in\mathcal{I}} \Delta_{\underline{i},\underline{j}^*}(t_{\varepsilon,\underline{i},\alpha})$ is β-level, i.e. its probability of type I error is equal to β.

The proof of Proposition 5.1 is an immediate consequence of Theorem 4.1. Since the testing procedures involved are independent for different orientations, the limiting distributions can be exactly computed and we do not need the probably conservative FWER. For any $\varepsilon \in (0,\varepsilon_o)$,

$$\sup_{f\in\mathcal{S}(R,\mathcal{I})} \mathbb{P}_f(\Lambda_{\mathcal{I},\varepsilon,\alpha} = 1) = \sup_{f\in\mathcal{S}(R,\mathcal{I})} \left(1 - \mathbb{P}_f\left(\max_{\underline{i}\in\mathcal{I}} \Delta_{\underline{i},\underline{j}^*}(t_{\varepsilon,\underline{i},\alpha}) = 0\right)\right) = 1 - \left(1 - \frac{\alpha}{2}\right)^{|\mathcal{I}|} = \beta.$$

Remark 5.1. Under the alternative, we naturally model the estimand as a function in a d-variate smoothness space, here and once again a $B^{\underline{s}}_{2,\infty}(R)$. Nevertheless, inspired from Remark 4.1, we can state another analogy with the maxiset approach. Let us consider the testing problem with the same hypotheses as in (8) but with a rate that is chosen to be the worst optimal one with respect to $\underline{s}$, that is $r_{\varepsilon,\underline{1}} = \varepsilon^{4\gamma/(1+4\gamma)}$. Since

we are only concerned by the presence of information along at least one given direction that belongs to $\mathcal{I}$, no assumption on the smoothness is needed in the other orientations. Hence, the sequence space in the alternative associated with the chosen rate $r_{\varepsilon,\underline{1}}$ can be considered larger than the initial one, $B^{\underline{s}}_{2,\infty}(R)$. More precisely it could be replaced by the sequence space defined by the union of all Besov balls $B^{\underline{s}'}_{2,\infty}(R)$ with $\underline{s}' \in (0,+\infty)^d$ such that $\sum_u i_u s'_u = \gamma^{-1}$ for at least one $\underline{i} \in \mathcal{I}$.

In what follows, we present some examples where the optimal testing procedures that combine different orientations could be useful.

5.1. Testing the structure of the goal function

Take the set $\mathcal{I}(\delta_0) = \{\underline{i} \in \{0,1\}^d : |\underline{i}| > \delta_0\}$ that contains all orientations for which the number of contributing components of the d-vector is strictly larger than a specified value δ_0. The corresponding hypotheses for testing for the atomic dimension are:

$$\begin{aligned} \mathcal{H}_0 : f \in \mathcal{S}(R,\mathcal{I}(\delta_0)) &= \{f : \|f\|_2 \le R, \delta_{\mathcal{B}_d}(f) \le \delta_0\}, \\ \mathcal{H}_a : f \in \mathcal{T}(R,\mathcal{I}(\delta_0)) &= \{f : f \in B^{\underline{s}}_{2,\infty}(R), \exists\ \underline{i} \in \mathcal{I}(\delta_0), \|f_{\underline{i}}\|_2 \ge Cr_{\varepsilon,\underline{i}}\}. \end{aligned} \tag{9}$$

Tests for the atomic dimension are more versatile than merely testing for additivity, the latter which is a special case with $\delta_0 = 1$. Concluding that a structure is simpler than a full d-dimensional object leads to an efficiency gain in estimation by using the appropriate methods escaping (at least partly) the curse of dimensionality when $\delta(f)$ is smaller than d.

Combining tests over multiple orientations $\underline{i} \in \mathcal{I}$ can be done by computing the maximum of testing procedures in the separate orientations.

From our testing procedures $\Delta_{\underline{i},\underline{j}^*}(t_{\varepsilon,\underline{i},\alpha})$, $\underline{i} \neq \underline{0}$, there is a quite natural way to estimate the structure of the goal function f by considering

$$\hat{\mathcal{A}}_{f,t_{\varepsilon,\underline{i},\alpha}} = \left\{\underline{i} \in \{0,1\}^d \setminus \underline{0} : \ \Delta_{\underline{i},\underline{j}^*}(t_{\varepsilon,\underline{i},\alpha}) = 1\right\}$$

as an estimator of $\mathcal{A}_f$. Nevertheless this estimator could lead to many false discoveries that are any orientation $\underline{i} \in \mathcal{A}_{f,t_{\varepsilon,\underline{i},\alpha}}$ such that $f_{\underline{i}}$ is the null function. Obviously the larger α, the bigger the number of expected false discoveries.

There is also a way to naturally get an estimator $\hat{\delta}$ of the atomic dimension $\delta(f)$ of a d-dimensional signal f by putting, for a chosen $\alpha \in \left(0, \frac{1}{2}\right)$,

$$\hat{\delta}_{\mathcal{B}_d} = \begin{cases} \max\{|\underline{i}|; \underline{i} \in \hat{\mathcal{A}}_{f,t_{\varepsilon,\underline{i},\alpha}}\} & \text{if } \hat{\mathcal{A}}_{f,t_{\varepsilon,\underline{i},\alpha}} \neq \emptyset \\ 0 & \text{if } \hat{\mathcal{A}}_{f,t_{\varepsilon,\underline{i},\alpha}} = \emptyset. \end{cases}$$

Rejecting a combined orientations null hypothesis as in (9) when $\hat{\delta}_{\mathcal{B}_d} \ge \delta_0$ is equivalent with rejecting $\mathcal{H}_0$ when $\Lambda_{\mathcal{I}(\delta_0),\varepsilon,\alpha} = 1$.

5.2. Reduction of model: selection of variables

Another application is to test whether some variables among $x_1, \ldots, x_d$ may be omitted from the model or not. Consider the case of leaving out variable x_m for some $m \in \{1, \ldots, d\}$. In this case $\mathcal{I}_m = \{\underline{i} = (i_1, \ldots, i_m, \ldots, i_d) \in \{0,1\}^d : i_m = 1\}$ and the relevant hypotheses are

$$\begin{aligned} \mathcal{H}_0 : f \in \mathcal{U}(R,m) &= \{f : \|f\|_2 \le R, \forall \underline{i} \in \mathcal{I}_m, f_{\underline{i}}(\underline{x}) = 0, \quad \forall \underline{x} \in [0,1)^d\}, \\ \mathcal{H}_a : f \in \mathcal{V}(R,m) &= \{f : f \in B^{\underline{s}}_{2,\infty}(R), \exists\ \underline{i} \in \mathcal{I}_m, \|f_{\underline{i}}\|_2 \ge Cr_{\varepsilon,\underline{i}}\}. \end{aligned} \tag{10}$$

In addition to possibly reduce the model, when considering such a testing problem could find applications in modeling time series (see Section 6 for details).

Following the two previous paragraphs, an interesting point must be underlined here. Estimating the atomic dimension of a d-dimensional object or reducing the number of its relevant variables can be considered as a first step to guarantee that the object is a function of a number of variables smaller than d (usually called 'sparse variable function'). This assumption was often assumed in recent works in estimation problem as in Autin et al. [3] and in testing problem as in Ingster and Suslina [19]. From the data, our methodology also permits to test whether this assumption is reasonable or not.

6. Numerical experiments

In this section we first describe a practical implementation of the adaptive method. We present its performance in terms of empirical power curves w.r.t. signal to noise ratio for testing the atomic dimension of some multidimensional functions corrupted by an additive Gaussian white noise. Then, we propose an application in time series analysis that consists in exploring the structure of the spatio-spectrum of spatio-temporal processes.

6.1. Data-driven structure detection

To link our theory to the practical setting we refer the reader to Reiss [30] for the equivalence in the Le Cam's sense between the multidimensional nonparametric regression problem (11) and the Gaussian white noise model (2). This equivalence is assumed to hold and typically relies on some minimal regularity conditions and an appropriate calibration of the noise level $\varepsilon = \sigma N^{-\frac{d}{2}}$. Let $\zeta_{l_1,\dots,l_d}$ be i.i.d. $\mathcal{N}(0,1)$,

$$Y_{l_1,\dots,l_d} = f\left(\frac{l_1}{N},\dots,\frac{l_d}{N}\right) + \sigma\zeta_{l_1,\dots,l_d},\ 1 \le l_u \le N,\ 1 \le u \le d. \tag{11}$$

Both, the $B^{\underline{s}}_{2,\infty}$-minimax and $B^{\underline{s}}_{2,\infty}$-adaptive minimax optimal procedures studied previously truncate the wavelet coefficient sequence at scales that are calibrated based on the knowledge of the unknown smoothness of the estimand (based either on the knowledge of all the directional regularities or only of their harmonic sum). Following the idea of Sieve estimation described in Birgé and Massart [5], we build a data-driven version of the adaptive testing procedure. In this numerical part, we consider only the $B^{\underline{s}}_{2,\infty}$-adaptive minimax optimal method since its algorithmic complexity is of about $O(\log_2(N))$ compared to $O\left(\log_2(N)^{|\underline{i}|}\right)$ for the $B^{\underline{s}}_{2,\infty}$-minimax optimal method.

Let us consider the sets of all possible adaptive models $\mathcal{F}_{\mathcal{M}^{\mathrm{A}}_N}$ based on ε^{-2} observations,

$$\mathcal{F}_{\mathcal{M}^{\mathrm{A}}_N} = \left\{\hat{f}_{m^{\mathrm{A}}} = \hat{\theta}^{\underline{0}}_{\underline{0},\underline{0}}\phi_{\underline{0},\underline{0}} + \sum_{\underline{i}\neq 0}\sum_{(\underline{j},\underline{k})\in m^{\mathrm{A}}_{\underline{i}}} \hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}}\psi^{\underline{i}}_{\underline{j},\underline{k}};\ m^{\mathrm{A}} := \{m^{\mathrm{A}}_{\underline{i}} \in \mathcal{M}^{\mathrm{A}}_{N,\underline{i}}\}\right\},$$

$$\text{with}\ \ \mathcal{M}^{\mathrm{A}}_{N,\underline{i}} := \left\{m'_{\underline{i},J} := \left\{(\underline{j},\underline{k}) \in \mathbb{J}^i \times \mathbb{K}_{\underline{j}};\ |\underline{j}| \le J;\ j_u \le \log_2(N),\ \forall u\right\}; J \le |\underline{i}|\log_2(N)\right\}.$$

The oracle approach leads to consider the estimator $\hat{f}_{m^{\mathrm{A}}_o}$ of $\mathcal{F}_{\mathcal{M}^{\mathrm{A}}_N}$ such that $\mathbb{E}\|\hat{f}_{m^{\mathrm{A}}_o} - f\|_2^2$ is minimal. This corresponds to find m^{A}, denoted as m^{A}_o, such that the following quantity is minimal:

$$-\sum_{(\underline{i},\underline{j},\underline{k})\in m^{\mathrm{A}}}\left((\theta^{\underline{i}}_{\underline{j},\underline{k}})^2 - \frac{\sigma^2}{N^d}\right).$$

The optimizer is found by solving the following problem for every $\underline{i} \in \{0,1\}^d \backslash \{\underline{0}\}$,

$$m^{\mathrm{A}}_{\underline{i},o} = \arg \min_{m^{\mathrm{A}}_{\underline{i}} \in \mathcal{M}^{\mathrm{A}}_{N,\underline{i}}} \left\{ - \sum_{(\underline{j},\underline{k}) \in m^{\mathrm{A}}_{\underline{i}}} \left((\theta^{\underline{i}}_{\underline{j},\underline{k}})^2 - \frac{\sigma^2}{N^d} \right) \right\}.$$

In practice, we plug in empirical quantities and adjust for the variability in the data proposing the following optimization, with $\hat{\lambda} = \hat{\sigma}\sqrt{2dN^{-d}\log N}$ the universal threshold,

$$\hat{m}^{\mathrm{A}}_{\underline{i},o} = \arg \min_{m^{\mathrm{A}}_{\underline{i}} \in \mathcal{M}^{\mathrm{A}}_{N,\underline{i}}} \left\{ - \sum_{(\underline{j},\underline{k}) \in m^{\mathrm{A}}_{\underline{i}}} \left((\hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}})^2 - \hat{\lambda}^2 \right) \right\}. \tag{12}$$

6.2. Empirical power functions

In this section we consider 3-dimensional functions f and we test for their atomic dimension $H_0 : \delta(f) = 1$ versus $H_1 : \delta(f) > 1$. Therefore we consider the hyperbolic Haar wavelet basis so that $\delta_{\mathcal{B}_d}(f) = \delta(f)$. For sake of simplicity, we generate them using the Sobol decomposition of a d-variate function $f \in L_2([0,1)^d)$ into orthogonal summands of growing dimensions,

$$f(x_1, \ldots, x_d) = \sum_{u=1}^{d} \sum_{i_1 < \cdots < i_u} f_{i_1 \ldots i_u}(x_{i_1}, \ldots, x_{i_u}). \tag{13}$$

The marginal functional components are chosen to be univariate functions that are frequently considered in the wavelet literature [2]. The interaction terms are obtained as weighted product of the marginal functions. In this experiment we choose the following test functions we describe below.

A: $f_1(x_1)$: 'blip', $f_2(x_2)$: 'blip', $f_3(x_3)$: 'blip',
B: $f_1(x_1)$: 'blip', $f_2(x_2)$: 'wave', $f_3(x_3)$: 'bumps'.

We consider the null hypothesis with $\delta(f) = 1$ and two functions in the alternative with atomic dimension $\delta(f) = 2$ or $\delta(f) = 3$.

- under $H_0 : \delta(f) = 1$, $f(x_1, x_2, x_3) = f_1(x_1) + f_2(x_2) + f_3(x_3)$
- under $H_1 : \delta(f) = 2$, $f(x_1, x_2, x_3) = f_1(x_1) + f_2(x_2) + f_3(x_3) + Df_1(x_1) f_2(x_2)$
- under $H_1 : \delta(f) = 3$, $f(x_1, x_2, x_3) = f_1(x_1) + f_2(x_2) + f_3(x_3) + Df_1(x_1) f_2(x_2) + D'f_1(x_1) f_2(x_2) f_3(x_3)$.

Here, Fig. 1 gives an illustration of some of these test functions.

The total energy of these functions $f(x_1, x_2, x_3)$ is normalized. Then the observed data are generated by adding Gaussian white noise to the test functions. The empirical power is computed as a function of the SNR as follows:

$$P(j) = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}_{\left\{ \Lambda(\delta_0)_{m,SNR_j} = 1 \right\}},$$

where the signal to noise ratio (SNR) is defined as the ratio of the standard deviation of the function values to the standard deviation of the noise. $\Lambda(\delta_0)_{m,SNR_j}$ is the result of the procedure $\Lambda_{\mathcal{I}(\delta_0),\varepsilon,\alpha}$ (with $\delta_0 \in \{1,2,3\}$) at the SNR_j and m-th Monte Carlo iteration. We generate these data sets with sample sizes

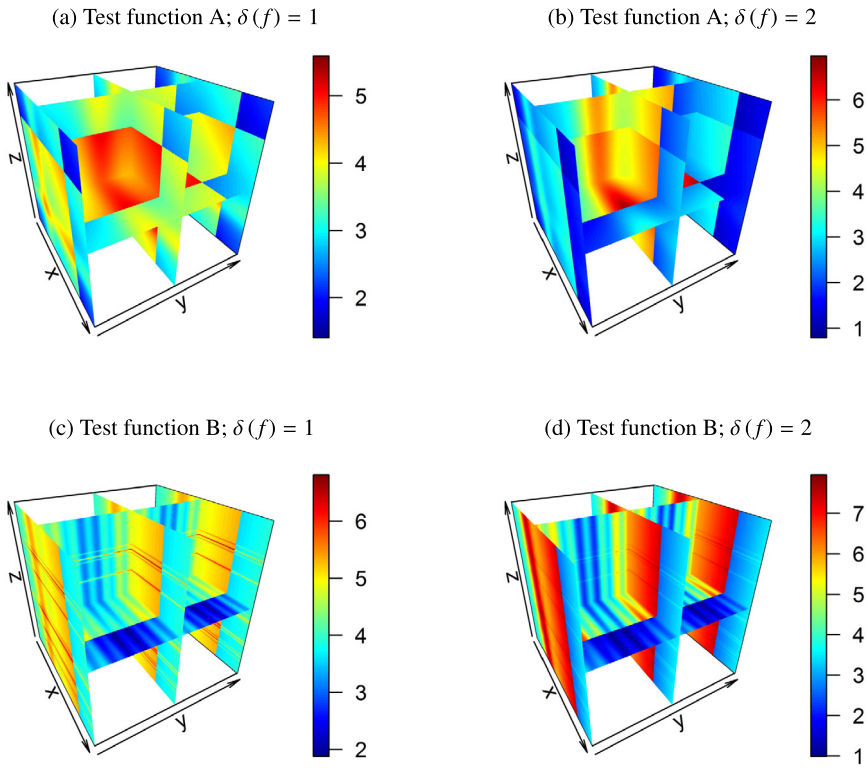

Fig. 1. Three dimensional test functions that are used in the numerical experiments.

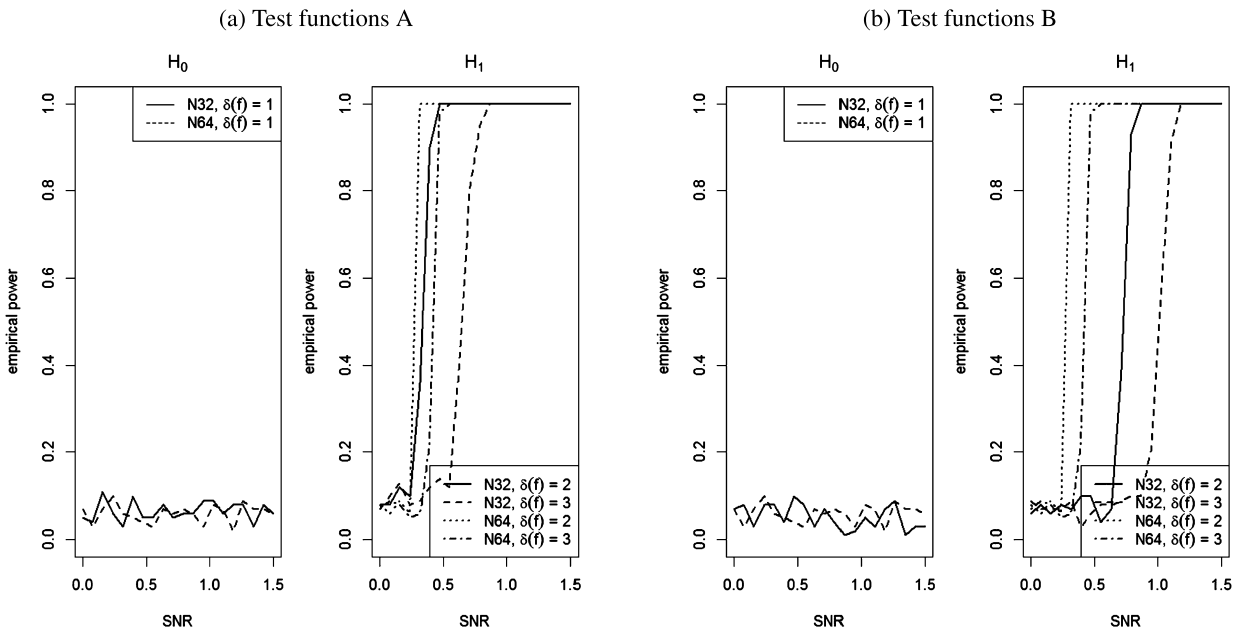

Fig. 2. Simulated rejection probabilities under H_0 and H_1 for both sets of test functions A and B, as a function of the signal to noise ratio.

$N \in \{32, 64\}$, $1 \leq j \leq 50$ and with $M = 100$ Monte Carlo replications at every values of SNR_j. The results are summarized in Fig. 2.

Fig. 2 shows that for both test functions, the nominal level of the test at $\alpha = 5\%$ under the null hypothesis H_0 is maintained over the different values of the SNR. Under the alternative, we observe firstly

that the detection appears at very low SNR values, secondly, an increase in the sample size improves the detection power of the method, and thirdly, under the two different alternatives, when the energy of the function is more spread over orientations ($\delta(f) = 3$), the detection power decreases. These tests confirm the proper calibration of the data-driven adaptive method. We can now apply it to a more realistic data example.

6.3. Application in time series analysis

Neumann and von Sachs [28] consider testing the stationarity of a time series by studying the structure of its time-varying spectrum. They form local contrast test statistics in the time direction also based on hyperbolic wavelet coefficients of an estimator of the time-varying spectral density, namely the preperiodogram. A crucial problem for practical application of this method in testing stationarity is due to the inherent nature of the preperiodogram data which suffers from interferences, very low SNR and non Gaussian noise distribution. This may strongly affect the performances of the data-driven choice of the truncation scales. Adapting our testing methodology to this context is an interesting research problem in itself. In the sequel, we test the structure of the spatio-spectrum of spatio-temporal processes that are time stationary. A specific concern for this application is the high dimensional setting. We face the problem of having small p-values due to a large number of observations [23]. We illustrate how to properly use our results to assess the structure of such data. We first define a class of spatio-temporal processes, its spatio-spectrum, explain how to estimate this and finally we apply our data-driven method to simulated data.

6.3.1. Spatio-temporal model

Following Ombao et al. [29], we define the Cramér representation of a spatio-temporal process $\{X_t(\underline{u});\ \underline{u} = (u_1, u_2) \in [0,1)^2, t \in \mathbb{Z}\}$, where $\underline{u}$ is the spatial index, as follows:

$$X_t(\underline{u}) = \int_{-\pi}^{\pi} A(\underline{u}, \omega) \exp(i\omega t)\, dZ(\omega) \tag{14}$$

where $Z(\omega)$ is a complex-valued stochastic process with zero mean and orthogonal increments, which satisfies

$$\mathbb{E}[dZ(\omega)] = 0, \quad \mathrm{Cov}(dZ(\omega_1), dZ(\omega_2)) = Dirac(\omega_1 - \omega_2)\, d\omega_1$$

where $Dirac(.)$ means the Dirac function at mass point 0 and $A(\underline{u}, \lambda)$ is the spatial complex-valued transfer function. It is Hermitian $A(., -\omega) = A^*(., \omega)$, where A^* is the adjoint operator. The location-dependent spectrum is given by

$$f(\underline{u}, \omega) := |A(\underline{u}, \omega)|^2 .$$

6.3.2. Estimation

In a practical setting we observe the spatio-temporal process over a discrete grid of dimensions $(N_1 \times N_2 \times T) \in \mathbb{N}^3$. Since we assume it to be time stationary, we can compute the bias-corrected log-periodogram at every spatial location

$$\log\left(\frac{1}{2\pi T}\left|\sum_{t=1}^{T} X_t(\underline{u}) \exp(i\omega t)\right|^2\right) + \kappa$$

where $\kappa = 0.57721$ is the Euler–Mascheroni constant, as in Wahba [37].

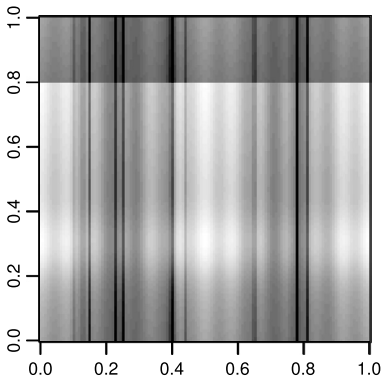

Fig. 3. Spatial AR(1) parameter for the time series application. The value of the AR(1) parameter at each spatial location ranges from 0.2 (black) to 0.99 (white color).

From these spatial log-periodograms we test the structure of the spatio-spectrum $f(\underline{u}, \omega)$. Hereafter we directly apply our methodology without adaptation for the non Gaussian nature of the log-periodogram ordinates. The central limit theorem in the coefficient domain permits to consider that the wavelet coefficients are approximately normally distributed up to comparatively fine scales. The actual procedure could be improved in that specific context, for example by adapting the methodology of Gao [9] in the penalization of the data-driven selection of the truncation scales. Nevertheless, omitting this adaptation helps us to illustrate the practical application of our method considering a slightly suboptimal procedure.

In the previous section we checked that our method is well-calibrated. In this time series example, we are away from this idealistic framework. We are now dealing with periodogram data that are non Gaussian and the spatio-spectrum may have a complicated structure (in terms of energy along the different orientations). In this high-dimensional and more realistic context, the p-values have not longer any practical importance. For (very) large sample sizes, p-values are almost always extremely small, suggesting the rejection of the null hypothesis since, with real life data, the situation of total absence of information in some orientation is rarely exactly true. A solution in this context is to use our test statistic values to describe the proportions to the total energy of the function that are in the different orientations, similarly as in principal components analysis. This is the approach adopted hereafter for time series data examples.

6.3.3. Results

We generate the following time series models using a discretized version of their Cramér representation given by equation (14). We consider two spatial AR(1) processes, i.e., their spatio-spectrum is given by

$$f(\underline{u}, \omega) = \frac{\sigma^2}{2\pi} \left(1 - 2\phi(\underline{u}) \cos\omega + \phi(\underline{u})^2\right)^{-1}.$$

Let 'Model 1' be a constant parameter over space, i.e., $\phi(\underline{u}) = \phi = 0.9,\ \forall \underline{u} \in [0,1]^2$, and 'Model 2' to have a complex spatial AR(1) structure over space as illustrated in Fig. 3.

Table 1 gives the result of the Monte Carlo experiments on generated time series data for $N_1 = N_2 = 128$ and $T = 128$. Looking at the energy repartition for the logarithm of the true spatio-spectrum under Model 1, it is clear that all the information is just contained in the marginal function of frequency. Under the Model 2, one can realize that the difference in the energy in orientations involving the space variable is pretty low. For what concerns the repartition estimation in the different orientation, there is no problem for identifying the first leading orientation, nevertheless, under model 2, we do not get a clear message about the second dominant orientation.

Table 1
True vs Estimated repartition of the total energy (in %).

orientation $\underline{i}$	Model 1		Model 2	
	True repartition	Estimated	True repartition	Estimated
$(1,0,0)$	0	0 (0.013)	0.1	0.1 (0.09)
$(0,1,0)$	0	0 (0.01)	0	0.1 (0.078)
$(1,1,0)$	0	0 (0.01)	0	0.1 (0.069)
$(0,0,1)$	100	99.9 (0.014)	91.8	90.5 (0.150)
$(1,0,1)$	0	0 (0.001)	4.9	3.3 (0.051)
$(0,1,1)$	0	0 (0.001)	3.1	5.9 (0.074)
$(1,1,1)$	0	0 (0.001)	0	0 (0.008)

7. Discussion

In this paper we describe how the Sobol decomposition in relation to the geometry of the multidimensional hyperbolic wavelet basis allows to build simple test statistics with impressive theoretical performance in the presence of anisotropic estimand. Appropriately combined, these test statistics allow to test for general structural properties such as the atomic dimension. We determine the minimax rates for separating the null and alternative hypothesis for detecting the Sobol functional component in cases of full or partial knowledge about the smoothness parameter. We propose two sequences of testing procedures that are based on different knowledges of the smoothness parameter (full or partially known) and we show that they are asymptotically optimal in the minimax sense. Interestingly we observe on the one hand that a loss of information about the smoothness parameter is accompanied with a deterioration of the minimax rate of separation; on the other hand the set of functions in the alternative hypothesis can be larger. Finally, we describe a methodology for a data driven version of the testing procedure and its application to time series data.

In the context of time series analysis, we consider extending our application example to deal with spatial and time non-stationary processes. We are therefore interested in spatial time-varying spectrum of these processes. We can compute at every spatial locations an estimator of the time-varying spectrum such as the preperiodogram (see Neumann and von Sachs [28]). Neumann and von Sachs [27] provide arguments about the asymptotic equivalence of the preperiodogram estimation problem, the Gaussian white noise model and the multidimensional nonparametric regression model so that our theory actually can easily be extended to this context. Nevertheless, in a practical setting, the nature of the preperiodogram data requires to develop an adapted methodology for choosing the truncation scales. In combination with the fully adaptive testing procedure, this method certainly encounters success for practical applications. Already in the context of estimation of the time-varying spectral density, the adaptive smoothing of the preperiodogram is still an actual research topic (see for instance van Delft and Eichler [36]).

Acknowledgments

The authors would like to thank the two reviewers for useful comments and suggestions on a preliminary version of the manuscript. G. Claeskens and J.-M. Freyermuth acknowledge the support of the Fund for Scientific Research Flanders, KU Leuven grant GOA/12/14 and of the IAP Research Network P7/06 of the Belgian Science Policy. Jean-Marc Freyermuth's and John Aston's research was supported by the Engineering and Physical Sciences Research Council [EP/K021672/2].

Appendix A

A.1. Proof of Theorem 4.1

The proof of Theorem 4.1 is a direct consequence of Propositions A.1 and A.2 that respectively deal with the upper bound and the lower bound of the minimax result.

Proposition A.1. *Under the same assumptions of Theorem 4.1 and for any $C > C_{\underline{i},\alpha}$, the following inequalities hold*

$$\sup_{0<\varepsilon<\epsilon_o} \sup_{f\in\mathcal{N}_{\underline{i}}(R)} \mathbb{P}_f\left(\Delta_{\underline{i},\underline{j}^*}\left(t_{\varepsilon,\underline{i},\alpha}\right)=1\right)=\frac{\alpha}{2} \quad \textit{and} \quad \sup_{0<\varepsilon<\epsilon_o} \sup_{f\in\mathcal{A}_{\underline{i}}(R,C,\underline{s},r_{\varepsilon,\underline{i}})} \mathbb{P}_f\left(\Delta_{\underline{i},\underline{j}^*}\left(t_{\varepsilon,\underline{i},\alpha}\right)=0\right)\leq\frac{\alpha}{2}.$$

Remark A.1. The constant $C_{\underline{i},\alpha}$ corresponds to the absolute positive constant $C_{\underline{i},\alpha}$ appearing in Definition 3.1 and will be given at the end of the proof below.

Proof. Consider $\varepsilon \in (0, \varepsilon_o)$. Suppose that $f \in \mathcal{N}_{\underline{i}}(R)$. Then $\varepsilon^{-2}T_{\underline{i},\underline{j}^*}$ has a Chi-Square distribution with $2^{|\underline{j}^*|}$ degrees of freedom. Hence

$$\mathbb{P}_f\left(\Delta_{\underline{i},\underline{j}^*}\left(t_{\varepsilon,\underline{i},\alpha}\right)=1\right)=\mathbb{P}_f\left(T_{\underline{i},\underline{j}^*}>t_{\varepsilon,\underline{i},\alpha}\right)=\mathbb{P}_f\left(\varepsilon^{-2}T_{\underline{i},\underline{j}^*}>\chi_{1-\frac{\alpha}{2}}\left(2^{|\underline{j}^*|}\right)\right)=\frac{\alpha}{2}.$$

Suppose now that $f \in \mathcal{A}_{\underline{i}}(R, C, \underline{s}, r_{\varepsilon,\underline{i}})$. Then $\varepsilon^{-2}T_{\underline{i},\underline{j}^*}$ is a noncentral Chi-Square distribution with $2^{|\underline{j}^*|}$ degrees of freedom. Moreover, if $V^{\underline{i}}_{\underline{j}^*}$ characterizes the following linear span of

$$\left\{\psi^{\underline{i}}_{\underline{j},\underline{k}}:\ \ \underline{j}\in\mathbb{J}^{\underline{i}}:\ j_u<\max(j^*_u,1),\forall u \ \text{ and } \underline{k}\in\mathbb{K}_{\underline{j}}\right\},$$

then,

$$\mathbb{E}_f(T_{\underline{i},\underline{j}^*})=\varepsilon^2 2^{|\underline{j}^*|}+\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}:\ j_u<\max(j^*_u,1),\forall u}\ \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\left(\theta^{\underline{i}}_{\underline{j},\underline{k}}\right)^2=\varepsilon^2 2^{|\underline{j}^*|}+\|Proj_{V^{\underline{i}}_{\underline{j}^*}}(f_{\underline{i}})\|_2^2,$$

$$\mathbb{V}ar_f(T_{\underline{i},\underline{j}^*})=2\varepsilon^4 2^{|\underline{j}^*|}+\ 4\varepsilon^2\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}:\ j_u<\max(j^*_u,1),\forall u}\ \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\left(\theta^{\underline{i}}_{\underline{j},\underline{k}}\right)^2=2\varepsilon^4 2^{|\underline{j}^*|}+4\varepsilon^2\|Proj_{V^{\underline{i}}_{\underline{j}^*}}(f_{\underline{i}})\|_2^2.$$

By considering the Cramér–Chernoff method (see Chapter 2 of Massart [25]), we easily obtain the following concentration inequality for the noncentral Chi-Square distribution with $2^{|\underline{j}^*|}$ degrees of freedom,

$$\mathbb{P}_f\left(\mathbb{E}_f(T_{\underline{i},\underline{j}^*})-T_{\underline{i},\underline{j}^*}\geq u\right)\leq\exp\left(-\frac{u^2}{2\mathbb{V}ar_f(T_{\underline{i},\underline{j}^*})}\right),\qquad \forall u>0.$$

So,

$$\begin{aligned}
\mathbb{P}_f\left(\Delta_{\underline{i},\underline{j}^*}\left(t_{\varepsilon,\underline{i},\alpha}\right)=0\right)&=\mathbb{P}_f\left(T_{\underline{i},\underline{j}^*}\leq t_{\varepsilon,\underline{i},\alpha}\right)\\
&=\mathbb{P}_f\left(\mathbb{E}_f(T_{\underline{i},\underline{j}^*})-T_{\underline{i},\underline{j}^*}\geq\mathbb{E}_f(T_{\underline{i},\underline{j}^*})-t_{\varepsilon,\underline{i},\alpha}\right)\\
&\leq\exp\left(-\frac{\left(\mathbb{E}_f(T_{\underline{i},\underline{j}^*})-t_{\varepsilon,\underline{i},\alpha}\right)^2}{2\mathbb{V}ar_f(T_{\underline{i},\underline{j}^*})}\right)\\
&=\exp\left(-\frac{\left(\varepsilon^2 2^{|\underline{j}^*|}+\|Proj_{V^{\underline{i}}_{\underline{j}^*}}(f_{\underline{i}})\|_2^2-t_{\varepsilon,\underline{i},\alpha}\right)^2}{4\varepsilon^4 2^{|\underline{j}^*|}+8\varepsilon^2\|Proj_{V^{\underline{i}}_{\underline{j}^*}}(f_{\underline{i}})\|_2^2}\right)\\
&=\exp\left(-\frac{\left(\varepsilon^2 2^{|\underline{j}^*|}+\|f_{\underline{i}}\|_2^2-\|f_{\underline{i}}-Proj_{V^{\underline{i}}_{\underline{j}^*}}(f_{\underline{i}})\|_2^2-t_{\varepsilon,\underline{i},\alpha}\right)^2}{4\varepsilon^4 2^{|\underline{j}^*|}+8\varepsilon^2\|f_{\underline{i}}\|_2^2}\right).
\end{aligned}$$

Remark that $f_{\underline{i}}$ belongs to $B^{\underline{s}}_{2,\infty}(R)$ because f does.

Since $\|f_{\underline{i}}\|_2^2 \geq C^2 r^2_{\varepsilon,\underline{i}}$ and $\chi_{1-\frac{\alpha}{2}}\left(2^{|\underline{j}^*|}\right) \leq 2^{|\underline{j}^*|} + 2\left(2^{|\underline{j}^*|}\log(2\alpha^{-1})\right)^{\frac{1}{2}}$ (see the exponential inequality for Chi-Square distribution given in Lemma 1 of Laurent and Massart [21]), one gets,

$$\begin{aligned}
&\left(-\log\left(\mathbb{P}_f\left(\Delta_{\underline{i},\underline{j}^*}\left(t_{\varepsilon,\underline{i},\alpha}\right)=0\right)\right)\right)^{-1} \\
&\leq \frac{4\varepsilon^4 2^{|\underline{j}^*|} + 8\varepsilon^2\|f_{\underline{i}}\|_2^2}{\left(\varepsilon^2 2^{|\underline{j}^*|} + \|f_{\underline{i}}\|_2^2 - \|f_{\underline{i}} - Proj_{V^{\underline{i}}_{\underline{j}^*}}(f_{\underline{i}})\|_2^2 - t_{\varepsilon,\underline{i},\alpha}\right)^2} \\
&\leq \frac{4\varepsilon^4 2^{|\underline{j}^*|} + 8\varepsilon^2\|f_{\underline{i}}\|_2^2}{\left(\varepsilon^2 2^{|\underline{j}^*|} + \|f_{\underline{i}}\|_2^2 - R^2 2^{-2|\underline{j}^*|\gamma_{\underline{i}}} - t_{\varepsilon,\underline{i},\alpha}\right)^2} \\
&\leq \frac{4\varepsilon^4 2^{|\underline{j}^*|}}{\left((C^2-R^2)2^{-2|\underline{j}^*|\gamma_{\underline{i}}} - 2\varepsilon^2\left(2^{|\underline{j}^*|}\log(2\alpha^{-1})\right)^{\frac{1}{2}}\right)^2} + \frac{8\varepsilon^2\|f_{\underline{i}}\|_2^2}{\left(\|f_{\underline{i}}\|_2^2 - R^2 2^{-2|\underline{j}^*|\gamma_{\underline{i}}} - 2\varepsilon^2\left(2^{|\underline{j}^*|}\log(2\alpha^{-1})\right)^{\frac{1}{2}}\right)^2} \\
&= \frac{4\varepsilon^4 2^{(1+4\gamma_{\underline{i}})|\underline{j}^*|}}{\left(C^2 - R^2 - 2^{1+\frac{|\underline{i}|}{2}}\left(\log(2\alpha^{-1})\right)^{\frac{1}{2}}\right)^2} + \frac{8C^4\varepsilon^2\|f_{\underline{i}}\|_2^{-2}}{\left(C^2 - R^2 - 2^{1+\frac{|\underline{i}|}{2}}\left(\log(2\alpha^{-1})\right)^{\frac{1}{2}}\right)^2} \\
&\leq \frac{2^{2+(1+4\gamma_{\underline{i}})|\underline{i}|} + 8C^2 e^{-\frac{2\varepsilon}{1+4\gamma_{\underline{i}}}}}{\left(C^2 - R^2 - 2^{1+\frac{|\underline{i}|}{2}}\left(\log(2\alpha^{-1})\right)^{\frac{1}{2}}\right)^2} \qquad (15) \\
&\leq \left(\log\left(2\alpha^{-1}\right)\right)^{-1}
\end{aligned}$$

provided that $C \geq C_{\underline{i},\alpha}$. The constant $C_{\underline{i},\alpha}$ is the largest value of C such that the last inequality can be replaced by an equality. That leads to $\mathbb{P}_f\left(\Delta_{\underline{i},\underline{j}^*}\left(t_{\varepsilon,\underline{i},\alpha}\right)=0\right) \leq \frac{\alpha}{2}$ for any C large enough and uniformly in f.

Note that the smaller α the larger $C_{\underline{i},\alpha}$. This remark is not a surprise obviously because when α is chosen to be small, the problem of detection becomes harder. Nevertheless the order of the minimax rate does not depend on α. We also note that (15) is obtained because we assumed that $\varepsilon < e^{-e}$. □

Proposition A.2. *Under the same assumptions of Theorem 4.1 and for any $C < c_{\underline{i},\alpha}$, the following inequality holds*

$$\inf_{0<\varepsilon<\epsilon_o}\inf_{\Delta}\left(\sup_{f\in\mathcal{N}_{\underline{i}}(R)}\mathbb{P}_f(\Delta=1) \quad + \sup_{f\in\mathcal{A}_{\underline{i}}(R,C,\underline{s},r_{\varepsilon,\underline{i}})}\mathbb{P}_f(\Delta=0)\right) > \alpha.$$

Remark A.2. The constant $c_{\underline{i},\alpha}$ corresponds to the absolute positive constant $c_{\underline{i},\alpha}$ appearing in Definition 3.1 and will be given at the end of the proof below.

Proof. For any $\varepsilon \in (0,\varepsilon_o)$ let us consider

$$f_\zeta = \sum_{\underline{k}\in\mathbb{K}_{\underline{j}^*}} \theta^{\underline{i}}_{\underline{j}^*,\underline{k}}\psi^{\underline{i}}_{\underline{j}^*,\underline{k}} = \sum_{\underline{k}\in\mathbb{K}_{\underline{j}^*}}\left(L\prod_{u=1}^{d} 2^{-j^*_u(\gamma_{\underline{i}}+\frac{1}{2})}\zeta_{\underline{k}}\right)\psi^{\underline{i}}_{\underline{j}^*,\underline{k}}$$

where $L>0$ and $\zeta_{\underline{k}} \in \{-1,1\}$ for any $\underline{k}$. The values j^*_u are chosen as in the upper bound. We recall that

$$2^{-j^*_u} \leq \varepsilon^{\frac{4i_u\gamma_{\underline{i}}}{(1+4\gamma_{\underline{i}})s_u}} < 2^{1-j^*_u}, \quad \forall\, u=1,\ldots,d.$$

First we check that f_ζ is in the alternative, that is to say it belongs to the Besov ball of radius R with the expected parameter of regularity and it is separated from 0 sufficiently. According to the definition of the Besov ball with radius R, we have to check that

$$\max_{1\le u\le d}\{2^{2j_u^* s_u}\}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}^*}}\left|\theta^{\underline{i}}_{\underline{j}^*,\underline{k}}\right|^2 < R^2.$$

Because of our choice of $\theta^{\underline{i}}_{\underline{j}^*,\underline{k}}$, we have

$$\max_{1\le u\le d}\{2^{2j_u^* s_u}\}2^{\sum_{u=1}^d j_u^*}L^2\prod_{u=1}^d 2^{-2j_u^*(\gamma_{\underline{i}}+\frac{1}{2})} = L^2\max_{1\le u\le d}\{2^{2j_u^* s_u}\}\prod_{u=1}^d 2^{-2j_u^*\gamma_{\underline{i}}} \le 2^{\min_{1\le u\le d} s_u}L^2.$$

A first condition on the constant L is that it has to be chosen smaller than $\frac{R}{\sqrt{2^{\min_u s_u}}}$. It ensures that f_ζ belongs to the Besov ball under consideration.

Next we compute the square of the L^2-distance between f_ζ and 0. We have

$$\sum_{\underline{k}\in\mathbb{K}_{\underline{j}^*}}\left(\theta^{\underline{i}}_{\underline{j}^*,\underline{k}}\right)^2 = L^2\prod_{u=1}^d 2^{-2j_u^*\gamma_{\underline{i}}}$$
$$\ge L^2 2^{-2|\underline{i}|\gamma_{\underline{i}}}\varepsilon^{\frac{8\gamma_{\underline{i}}}{4\gamma_{\underline{i}}+1}}.$$

Therefore L has to be larger than $2^{|\underline{i}|\gamma_{\underline{i}}}C$. This kind of choice guarantees that $\|f_\zeta\|_2 \ge Cr_{\varepsilon,\underline{i}}$.

Following Propositions 2.11 and 2.12 of Ingster and Suslina [17], we have that:

$$\inf_\Delta\left(\sup_{f\in\mathcal{N}_{\underline{i}}(R)}\mathbb{P}_f(\Delta=1) \quad + \sup_{f\in\mathcal{A}_{\underline{i}}(R,C,\underline{s},r_{\varepsilon,\underline{i}})}\mathbb{P}_f(\Delta=0)\right) \ge 1-\frac{1}{2}\sqrt{\mathbb{E}\left(\left(\mathbb{E}_\pi\left(L_{f_\zeta}\right)\right)^2\right)-1} \tag{16}$$

where π is any prior probability measure concentrated on the set of alternatives $\mathcal{A}_{\underline{i}}(R,C,\underline{s},r_{\varepsilon,\underline{i}})$ and

$$L_{f_\zeta} = \exp\left(\varepsilon^{-2}\int f_\zeta(x)\,dY(x) - \frac{1}{2}\varepsilon^{-2}\int f_\zeta(x)^2\,dx\right).$$

Therefore, according to (16), it suffices to prove that for judicious choices of π we can get

$$\mathbb{E}\left(\left(\mathbb{E}_\pi\left(L_{f_\zeta}\right)\right)^2\right) < 4(1-\alpha)^2+1. \tag{17}$$

We consider now the probability measure π such that the $\zeta_{\underline{k}}$'s are independent Rademacher random variables with parameter $\frac{1}{2}$. Note that

$$L_{f_\zeta} = \exp\left(\varepsilon^{-2}\int f_\zeta(x)\,dY(x) - \frac{1}{2}\varepsilon^{-2}\int f_\zeta(x)^2\,dx\right)$$
$$= \exp\left(-\frac{1}{2}\varepsilon^{-2}L^2 2^{|\underline{j}^*|}\prod_{u=1}^d 2^{-2j_u^*(\gamma_{\underline{i}}+\frac{1}{2})}\right)\prod_{\underline{k}\in\mathbb{K}_{\underline{j}^*}}\exp\left(\varepsilon^{-1}L\,\zeta_{\underline{k}}\,\xi^{\underline{i}}_{\underline{j}^*,\underline{k}}\prod_{u=1}^d 2^{-j_u^*(\gamma_{\underline{i}}+\frac{1}{2})}\right).$$

As a first step we take the expectation according to the prior probability measure π.

$$\mathbb{E}_\pi\left(L_{f_\zeta}\right) = \exp\left(-\frac{1}{2}\varepsilon^{-2}L^2 2^{|\underline{j}^*|}\prod_{u=1}^{d} 2^{-2j_u^*(\gamma_{\underline{i}}+\frac{1}{2})}\right)$$
$$\times \prod_{\underline{k}\in\mathbb{K}_{\underline{j}^*}} \frac{1}{2}\left[\exp\left(\varepsilon^{-1}L\,\xi^{\underline{i}}_{\underline{j}^*,\underline{k}}\prod_{u=1}^{d} 2^{-j_u^*(\gamma_{\underline{i}}+\frac{1}{2})}\right) + \exp\left(-\varepsilon^{-1}L\,\xi^{\underline{i}}_{\underline{j}^*,\underline{k}}\prod_{u=1}^{d} 2^{-j_u^*(\gamma_{\underline{i}}+\frac{1}{2})}\right)\right].$$

As a second step we take the expectation according to the independent standard Gaussian random variables $\xi^{\underline{i}}_{\underline{j}^*,\underline{k}}$:

$$\begin{aligned}\mathbb{E}\left(\left(\mathbb{E}_\pi\left(L_{f_\zeta}\right)\right)^2\right) &= \prod_{\underline{k}\in\mathbb{K}_{\underline{j}^*}} \cosh\left(\varepsilon^{-2}L^2\prod_{u=1}^{d} 2^{-j_u^*(2\gamma_{\underline{i}}+1)}\right) \\ &\le \exp\left(\frac{1}{2}2^{|\underline{j}^*|}\varepsilon^{-4}L^4\prod_{u=1}^{d} 2^{-2j_u^*(2\gamma_{\underline{i}}+1)}\right) \\ &\le \exp\left(\frac{L^4}{2}\right) \\ &< 4(1-\alpha)^2+1.\end{aligned} \tag{18}$$

The last inequality is obtained by choosing $L < \min\left(\left(2\log(1+4(1-\alpha)^2)\right)^{\frac{1}{4}}, \frac{R}{\sqrt{2}^{\min_u s_u}}\right)$. So (17) is proved, when considering C small enough, that is $C \le c_{\underline{i},\alpha} = 2^{-|\underline{i}|\gamma_{\underline{i}}}L$.

A.2. Proof of Theorem 4.2

The proof of Theorem 4.2 follows the same path as the proof of Theorem 4.1. It is a direct consequence of Proposition A.3 and Proposition A.4 that respectively deal with the upper bound and the lower bound of the minimax result.

Proposition A.3. *Under the same assumptions of Theorem 4.2 and for any $C > C'_{\underline{i},\alpha}$, the following inequalities hold*

$$\sup_{0<\varepsilon<\epsilon_o}\ \sup_{f\in\mathcal{N}_{\underline{i}}(R)} \mathbb{P}_f\left(\Delta_{\underline{i},max}\left(t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha}\right)=1\right) \le \frac{\alpha}{2} \quad \textit{and}$$
$$\sup_{0<\varepsilon<\epsilon_o}\ \sup_{\underline{s}:\ \sum_u i_u s_u^{-1}=\gamma_{\underline{i}}^{-1}}\ \sup_{f\in\mathcal{A}_{\underline{i}}\left(R,C,\underline{s},r'_{\varepsilon,\underline{i}}\right)} \mathbb{P}_f\left(\Delta_{\underline{i},max}\left(t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha}\right)=0\right) \le \frac{\alpha}{2}.$$

Remark A.3. The constant $C'_{\underline{i},\alpha}$ corresponds to the absolute positive constant $C'_{\underline{i},\alpha}$ appearing in Definition 3.2 and will be given at the end of the proof below.

Proof. Consider $\varepsilon \in (0,\varepsilon_o)$. Suppose that $f \in \mathcal{N}_{\underline{i}}(R)$. Then, for any $\underline{j} \in \mathbb{J}^{\underline{i}}$, $\varepsilon^{-2}T_{\underline{i},\underline{j}}$ has a Chi-Square distribution with $2^{|\underline{j}|}$ degrees of freedom. Hence

$$\begin{aligned}\mathbb{P}_f\left(\Delta_{\underline{i},max}\left(t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha}\right)=1\right) &= \mathbb{P}_f\left(\max_{\underline{j}\in\mathbb{J}^{\underline{i}}:|\underline{j}|=J_{\underline{i}}} \Delta_{\underline{i},\underline{j}}\left(t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha}\right)=1\right) \\ &\le \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}:|\underline{j}|=J_{\underline{i}}} \mathbb{P}_f\left(\Delta_{\underline{i},\underline{j}}\left(t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha}\right)=1\right) \\ &= \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}:|\underline{j}|=J_{\underline{i}}} \mathbb{P}_f\left(T_{\underline{i},\underline{j}} > t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha}\right)\end{aligned}$$

$$= \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}:|\underline{j}|=J_{\underline{i}}} \frac{\alpha}{2K_{\gamma_{\underline{i}}}} = \frac{\alpha}{2}.$$

Suppose now that $f \in \mathcal{A}_{\underline{i}}(R, C, \underline{s}, r'_{\varepsilon,\underline{i}})$ where $\underline{s}: \ \sum_u i_u s_u^{-1} = \gamma_{\underline{i}}^{-1}$. Then, for any fixed $\underline{j}' \in \mathbb{J}^{\underline{i}}$ such that $|\underline{j}'| = J_{\underline{i}}$, $\varepsilon^{-2}T_{\underline{i},\underline{j}'}$ is a noncentral Chi-Square distribution with $2^{J_{\underline{i}}}$ degrees of freedom. Hence, if $V^{\underline{i}}_{\underline{j}'}$ characterizes the following linear span of

$$\left\{\psi^{\underline{i}}_{\underline{j},\underline{k}}: \ \ \underline{j} \in \mathbb{J}^{\underline{i}}: \ j_u < \max(j'_u, 1), \forall u \ \text{ and } \underline{k} \in \mathbb{K}_{\underline{j}}\right\},$$

then,

$$\begin{aligned}
\mathbb{P}_f\left(\Delta_{\underline{i},max}\left(t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha}\right) = 0\right) &\le \mathbb{P}_f\left(T_{\underline{i},\underline{j}'} \le t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha}\right) \\
&= \mathbb{P}_f\left(\mathbb{E}_f(T_{\underline{i},\underline{j}'}) - T_{\underline{i},\underline{j}'} \ge \mathbb{E}_f(T_{\underline{i},\underline{j}'}) - t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha}\right) \\
&\le \exp\left(-\frac{\left(\mathbb{E}_f(T_{\underline{i},\underline{j}'}) - t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha}\right)^2}{2\mathbb{V}ar_f(T_{\underline{i},\underline{j}'})}\right) \\
&= \exp\left(-\frac{\left(\varepsilon^2 2^{J_{\underline{i}}} + \|Proj_{V^{\underline{i}}_{\underline{j}'}}(f_{\underline{i}})\|_2^2 - t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha}\right)^2}{4\varepsilon^4 2^{J_{\underline{i}}} + 8\varepsilon^2\|Proj_{V^{\underline{i}}_{\underline{j}'}}(f_{\underline{i}})\|_2^2}\right).
\end{aligned}$$

Since $\|f_{\underline{i}}\|_2^2 \ge C^2(r'_{\varepsilon,\underline{i}})^2$ and $\chi_{1-\alpha/2K_\gamma}\left(2^{J_{\underline{i}}}\right) \le 2^{J_{\underline{i}}} + 2\left(2^{J_{\underline{i}}}\log(2\alpha^{-1}K_{\gamma_{\underline{i}}})\right)^{\frac{1}{2}}$ (see [21]) with $(K_0 \log\varepsilon^{-1})^{|\underline{i}|-1} \le K_{\gamma_{\underline{i}}} \le (K_1 \log\varepsilon^{-1})^{|\underline{i}|-1}$ for some $K_0, K_1 > 0$, when computing similar calculus as in the non adaptive framework, we get

$$\begin{aligned}
&\left(-\log\left(\mathbb{P}_f\left(\Delta_{\underline{i},\max}\left(t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha}\right) = 0\right)\right)\right)^{-1} \\
&\quad\le \frac{4\varepsilon^4 2^{J_{\underline{i}}} + 8\varepsilon^2\|f_{\underline{i}}\|_2^2}{\left(\varepsilon^2 2^{J_{\underline{i}}} + \|f_{\underline{i}}\|_2^2 - \|f_{\underline{i}} - Proj_{V^{\underline{i}}_{\underline{j}'}}(f_{\underline{i}})\|_2^2 - t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha}\right)^2} \\
&\quad\le \frac{4\varepsilon^4 2^{J_{\underline{i}}} + 8\varepsilon^2\|f_{\underline{i}}\|_2^2}{\left(\varepsilon^2 2^{J_\gamma} + \|f_{\underline{i}}\|_2^2 - R^2 2^{-2J_{\underline{i}}\gamma_{\underline{i}}} - t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha}\right)^2} \\
&\quad\le \frac{4\varepsilon^4 2^{J_{\underline{i}}} + 8\varepsilon^2\|f_{\underline{i}}\|_2^2}{\left(\|f_{\underline{i}}\|_2^2 - R^2 2^{-2J_{\underline{i}}\gamma_{\underline{i}}} - 2\varepsilon^2\left(2^{J_{\underline{i}}}\log(2\alpha^{-1}K_{\gamma_{\underline{i}}}\right)^{\frac{1}{2}}\right)^2} \\
&\quad= \frac{4\varepsilon^4 2^{(1+4\gamma_{\underline{i}})J_{\underline{i}}}}{\left(C^2 - R^2 - 2^{\frac{3}{2}}\left(\log(2\alpha^{-1}) + d(\log K_1 + 1)\right)^{\frac{1}{2}}\right)^2} + \frac{8C^4\varepsilon^2\|f_{\underline{i}}\|_2^{-2}}{\left(C^2 - R^2 - 2^{\frac{3}{2}}\left(\log(2\alpha^{-1}) + d(\log K_1 + 1)\right)^{\frac{1}{2}}\right)^2} \\
&\quad\le \frac{2^{2+(1+4\gamma_{\underline{i}})}(\log\log\varepsilon^{-1})^{-1}}{\left(C^2 - R^2 - 2^{\frac{3}{2}}\left(\log(2\alpha^{-1}) + d(\log K_1 + 1)\right)^{\frac{1}{2}}\right)^2} + \frac{8C^4\varepsilon^2\|f_{\underline{i}}\|_2^{-2}}{\left(C^2 - R^2 - 2^{\frac{3}{2}}\left(\log(2\alpha^{-1}) + d(\log K_1 + 1)\right)^{\frac{1}{2}}\right)^2} \\
&\quad\le \frac{2^{2+(1+4\gamma_{\underline{i}})} + 8C^2 e^{-\frac{2e}{1+4\gamma_{\underline{i}}}}}{\left(C^2 - R^2 - 2^{\frac{3}{2}}\left(\log(2\alpha^{-1}) + d(\log K_1 + 1)\right)^{\frac{1}{2}}\right)^2} \qquad (19) \\
&\quad\le \left(\log\left(2\alpha^{-1}\right)\right)^{-1}
\end{aligned}$$

provided that $C \geq C'_{\underline{i},\alpha}$. The constant $C'_{\underline{i},\alpha}$ is the value of C for which the last inequality can be replaced by an equality. That leads to $\mathbb{P}_f\left(\Delta_{\underline{i},max}\left(t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha}\right)=0\right) \leq \frac{\alpha}{2}$ for any C large enough and uniformly in f. We also note that (19) is obtained because we assumed that $\varepsilon < e^{-e}$.

Note, once again, that the smaller α the larger $C'_{\underline{i},\alpha}$.

Proposition A.4. *Under the same assumptions of Theorem 4.2 and for any $C < c'_{\underline{i},\alpha}$, the following inequality holds*

$$\inf_{0<\varepsilon<\epsilon_o} \inf_{\Delta} \left(\sup_{f\in\mathcal{N}_{\underline{i}}(R)} \mathbb{P}_f(\Delta=1) \quad + \sup_{\underline{s}:\ \sum_u i_u s_u^{-1}=\gamma_{\underline{i}}^{-1}} \ \sup_{f\in\mathcal{A}_{\underline{i}}\left(R,C,\underline{s},r'_{\varepsilon,\underline{i}}\right)} \mathbb{P}_f(\Delta=0) \right) > \alpha.$$

Remark A.4. The constant $c'_{\underline{i},\alpha}$ corresponds to the absolute positive constant $c'_{\underline{i},\alpha}$ appearing in Definition 3.2 and will be given at the end of the proof below.

Proof. For any $\varepsilon \in (0,\varepsilon_o)$ let us consider the family of d-uple $\mathcal{J} = \left\{\underline{j} \in \mathbb{J}^{\underline{i}} : |\underline{j}| = J_{\underline{i}}\right\}$.

Remember that:

- $J_{\underline{i}}$ is such that $2^{-J_{\underline{i}}} \leq \left(\varepsilon^4 \log\log\varepsilon^{-1}\right)^{\frac{1}{1+4\gamma_{\underline{i}}}} < 2^{1-J_{\underline{i}}}$,
- $\#\mathcal{J} = K_{\gamma_{\underline{i}}} \geq (K_0 \log\varepsilon^{-1})^{|\underline{i}|-1}$ for some K_0.

For $\underline{j} \in \mathcal{J}$, define

$$f_{\zeta,\underline{j}} = \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \theta^{\underline{i}}_{\underline{j},\underline{k}} \psi^{\underline{i}}_{\underline{j},\underline{k}} = \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \left(L \prod_{u=1}^{d} 2^{-j_u(\gamma_{\underline{i}}+\frac{1}{2})} \zeta_{\underline{k}} \right) \psi^{\underline{i}}_{\underline{j},\underline{k}}$$

where $L > 0$ and $\zeta_{\underline{k}} \in \{-1,1\}$ for any $\underline{k}$.

As in the proof of Proposition A.2, we can easily check that $f_{\zeta,\underline{j}}$ is in the alternative, provided that $2^{\gamma}C \leq L \leq \frac{R}{\sqrt{2^{\gamma_{\underline{i}}}}}$.

Since

$$\begin{aligned} &\inf_{\Delta} \left(\sup_{f\in\mathcal{N}_{\underline{i}}(R)} \mathbb{P}_f(\Delta=1) + \sup_{\underline{s}:\ \sum_u i_u s_u^{-1}=\gamma_{\underline{i}}^{-1}} \ \sup_{f\in\mathcal{A}_{\underline{i}}\left(R,C,\underline{s},r'_{\varepsilon,\underline{i}}\right)} \mathbb{P}_f(\Delta=0) \right) \\ &\geq 1 - \frac{1}{2}\sqrt{\mathbb{E}\left(\left(\frac{1}{K_\gamma}\sum_{\underline{j}\in\mathcal{J}} \mathbb{E}_{\pi'}\left(L_{f_{\zeta,\underline{j}}}\right)\right)^2\right) - 1} \end{aligned} \tag{20}$$

where π' is a prior probability measure concentrated on the set of alternatives $\mathcal{A}_{\underline{i}}\left(R,C,\underline{s},r'_{\varepsilon,\underline{i}}\right)$ and

$$L_{f_{\zeta,\underline{j}}} = \exp\left(\varepsilon^{-2}\int f_{\zeta,\underline{j}}(x)\ dY(x) - \frac{1}{2}\varepsilon^{-2}\int f_{\zeta,\underline{j}}(x)^2\ dx\right).$$

Therefore, according to (20), it suffices to prove once again that for judicious choices of π' we can get

$$\mathbb{E}\left(\left(\frac{1}{K_\gamma}\sum_{\underline{j}\in\mathcal{J}} \mathbb{E}_{\pi'}\left(L_{f_{\zeta,\underline{j}}}\right)\right)^2\right) = \frac{1}{K_\gamma^2}\mathbb{E}\sum_{\underline{j}\in\mathcal{J}}\left(\mathbb{E}_{\pi'}\left(L_{f_{\zeta,\underline{j}}}\right)^2\right) < 4(1-\alpha)^2 + 1. \tag{21}$$

The left hand equality is due to the orthogonality of the hyperbolic wavelets.

We consider now the probability measure π' such that the $\zeta_{\underline{k}}$'s are independent Rademacher random variables with parameter 1/2. Similarly to (18), we can easily get the following inequality

$$\begin{aligned}\mathbb{E}\left(\left(\mathbb{E}_{\pi'}\left(L_{f_{\zeta,\underline{j}}}\right)\right)^2\right) &= \prod_{\underline{k}\in\mathbb{K}_{\underline{j}}} \cosh\left(\varepsilon^{-2}L^2\prod_{u=1}^{d}2^{-j_u(2\gamma_{\underline{i}}+1)}\right)\\ &\le \exp\left(\frac{1}{2}2^{J_{\underline{i}}}\varepsilon^{-4}L^4\prod_{u=1}^{d}2^{-2j_u(2\gamma_{\underline{i}}+1)}\right)\\ &\le \exp\left(\frac{L^4}{2}\log\log\varepsilon^{-1}\right)\\ &\le \left(\log\varepsilon^{-1}\right)^{\frac{L^4}{2}}.\end{aligned}$$

Since $K_{\gamma_{\underline{i}}} \sim (K\log\varepsilon^{-1})^{|\underline{i}|-1}$, we get the following inequalities:

$$\begin{aligned}\frac{1}{K_{\gamma_{\underline{i}}}^2}\mathbb{E}\sum_{\underline{j}\in\mathcal{J}}\left(\mathbb{E}_{\pi'}\left(L_{f_{\zeta,\underline{j}}}\right)^2\right) &\le \frac{1}{K_{\gamma_{\underline{i}}}^2}\mathbb{E}\sum_{\underline{j}\in\mathcal{J}}\left(\log\varepsilon^{-1}\right)^{\frac{L^4}{2}}\\ &= K_{\gamma_{\underline{i}}}^{-1}\left(\log\varepsilon^{-1}\right)^{\frac{L^4}{2}}\\ &\le K_0^{1-|\underline{i}|}\left(\log\varepsilon^{-1}\right)^{1-|\underline{i}|+\frac{L^4}{2}}\\ &\le K_0^{1-|\underline{i}|}e^{1-|\underline{i}|+\frac{L^4}{2}}\\ &< 4(1-\alpha)^2+1.\end{aligned}$$

The last two inequalities are obtained by choosing

$$L < \min\left(\left(2\log(K_0^{|\underline{i}|-1}(1+4(1-\alpha)^2))+2(|\underline{i}|-1)\right)^{\frac{1}{4}}, \frac{R}{\sqrt{2^{\gamma_{\underline{i}}}}}\right)$$

and by assuming that $\varepsilon < e^{-e}$. So (21) is proved, when considering C small enough, that is $C \le c'_{\underline{i},\alpha} = 2^{-\gamma_{\underline{i}}}L$.

Following the proofs of Propositions A.1 to A.4, it clearly appears that there is a relationship between the constants $C_{\underline{i},\alpha}, c_{\underline{i},\alpha}, C'_{\underline{i},\alpha}, c'_{\underline{i},\alpha}$, R and the level α. Moreover, as the threshold values $t_{\varepsilon,\underline{i},\alpha}$ and $t'_{\varepsilon,\underline{i},\gamma_{\underline{i}},\alpha}$ of our sequence of testing procedures depend on α, it also means that they are related to the constants C and R. From a theoretical point of view, we decided to fix the constant R and the level α and we exhibit the constant C leading to the required level. Nevertheless, from a practical point of view, one can think of fixing two of these constants and get some information about the third one.

References

[1] F. Abramovich, I. Feis, T. Sapatinas, Optimal testing for additivity in multiple nonparametric regression, Ann. Inst. Statist. Math. 61 (3) (2008) 691–714.

[2] A. Antoniadis, J. Bigot, J. Sapatinas, Wavelet estimators in nonparametric regression: a comparative simulation study, J. Stat. Softw. 6 (6) (2001) 1–83.

[3] F. Autin, G. Claeskens, J.-M. Freyermuth, Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach, Appl. Comput. Harmon. Anal. 36 (2) (2014) 239–255.

[4] F. Autin, G. Claeskens, J.-M. Freyermuth, Asymptotic performance of projection estimators in standard and hyperbolic wavelet bases, Electron. J. Stat. 9 (2) (2015) 1852–1883.

[5] L. Birgé, P. Massart, Minimum contrast estimators on sieves: exponential bounds and rates of convergence, Bernoulli 4 (3) (1998) 329–375.

[6] L. Comminges, A.S. Dalalyan, Minimax testing of a composite null hypothesis defined via a quadratic functional in the model of regression, Electron. J. Stat. 7 (2013) 146–190.
[7] A. Dalalyan, Y.I. Ingster, A.B. Tsybakov, Statistical inference in compound functional models, Probab. Theory Related Fields 158 (3) (2014) 512–532.
[8] D. De Canditiis, T. Sapatinas, Testing for additivity and joint effects in multivariate nonparametric regression using Fourier and wavelet methods, Stat. Comput. 14 (2004) 235–249.
[9] H.-Y. Gao, Choice of threshold for wavelet shrinkage estimate of the spectrum, J. Time Series Anal. 18 (1997) 231–251.
[10] G. Gayraud, Y.I. Ingster, Detection of sparse additive functions, Electron. J. Stat. 6 (2012) 1409–1448.
[11] Y.I. Ingster, Asymptotically minimax hypothesis testing for nonparametric alternatives i, Math. Methods Statist. 2 (114) (1993) 85–114.
[12] Y.I. Ingster, Asymptotically minimax hypothesis testing for nonparametric alternatives ii, Math. Methods Statist. 2 (114) (1993) 171–189.
[13] Y.I. Ingster, Asymptotically minimax hypothesis testing for nonparametric alternatives iii, Math. Methods Statist. 2 (114) (1993) 249–268.
[14] Y.I. Ingster, B. Laurent, C. Marteau, Signal detection for inverse problems in a multidimensional framework, Math. Methods Statist. 23 (2014) 279–305.
[15] Y.I. Ingster, N. Stepanova, Estimation and detection of functions from weighted tensor product spaces, Math. Methods Statist. 18 (4) (2009) 310–330.
[16] Y.I. Ingster, N. Stepanova, Estimation and detection of functions from anisotropic Sobolev classes, Electron. J. Stat. 5 (2011) 484–506.
[17] Y.I. Ingster, I.A. Suslina, Nonparametric Goodness-of-Fit Testing Under Gaussian Models, Lectures Notes in Statistics, Springer, New York, 2003.
[18] Y.I. Ingster, I.A. Suslina, Detection of a signal of known shape in a multichannel system, J. Math. Sci. 127 (1) (2005) 1723–1736.
[19] Y.I. Ingster, I.A. Suslina, Detection of a sparse variable function, J. Math. Sci. 206 (2) (2015) 181–196.
[20] G. Kerkyacharian, D. Picard, Thresholding algorithms, maxisets and well concentrated bases, TEST 9 (2) (2000) 283–344.
[21] B. Laurent, P. Massart, Adaptive estimation of a quadratic functional by model selection, Ann. Statist. 28 (5) (2000) 1302–1338.
[22] O. Lepski, A.B. Tsybakov, On asymptotically exact testing of nonparametric hypotheses, Probab. Theory Related Fields 117 (1) (2000) 17–48.
[23] M. Lin, H.C. Lucas, G. Shmueli, Too big to fail: large samples and the p-value problem, Inf. Syst. Res. (2013) 1–12.
[24] Y. Lin, Tensor product space ANOVA models, Ann. Statist. 28 (3) (2000) 734–755.
[25] P. Massart, Concentration Inequalities and Model Selection: Ecole D'Eté De Probabilités De Saint-Flour XXXIII — 2003, Lecture Notes in Mathematics, vol. 1738, Springer, 2003.
[26] M.H. Neumann, Multivariate wavelet thresholding in anisotropic function spaces, Statist. Sinica 10 (2000) 399–431.
[27] M.H. Neumann, R. von Sachs, Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra, Ann. Statist. 25 (1) (1997) 38–76.
[28] M.H. Neumann, R. von Sachs, A wavelet-based test for stationarity, J. Time Series Anal. 21 (5) (2000) 597–613.
[29] H. Ombao, X. Shao, E. Rykhlevskaia, M. Fabiani, G. Gratton, Spatio-spectral analysis of brain signal, Statist. Sinica 18 (2008) 1465–1482.
[30] M. Reiss, Asymptotic equivalence for nonparametric regression with multivariate and random design, Ann. Statist. 36 (4) (2008) 1957–1982.
[31] H.-J. Schmeisser, Recent developments in the theory of function spaces with dominating mixed smoothness, in: Nonlinear Analysis, Function Spaces and Applications, vol. 8, 2006, pp. 145–204.
[32] H.-J. Schmeisser, W. Sickel, Spaces of functions of mixed smoothness and approximation from hyperbolic crosses, J. Approx. Theory 128 (2) (2004) 115–150.
[33] W. Sickel, T. Ullrich, Tensor products of Sobolev–Besov spaces and applications to approximation from the hyperbolic cross, J. Approx. Theory 161 (2009) 748–786.
[34] I. Sobol, Multidimensional Quadrature Formulas and Haar Functions, Nauka, 1969 (in Russian).
[35] V. Spokoiny, Adaptive and spatially adaptive testing of a nonparametric hypothesis, Math. Methods Statist. 7 (3) (1998) 245–273.
[36] A. van Delft, M. Eichler, Data-adaptive estimation of time-varying spectral densities, arXiv:1512.00825, 2015.
[37] G. Wahba, Automatic smoothing of the log-periodogram, J. Amer. Statist. Assoc. 75 (1980) 122–132.
[38] H. Zhou, L. Li, H. Zhu, Tensor regression with applications in neuroimaging data analysis, J. Amer. Statist. Assoc. 108 (502) (2013) 540–552.