The Annals of Statistics
2013, Vol. 41, No. 2, 838–869
DOI: 10.1214/13-AOS1093

MINIMAX ADAPTIVE TESTS FOR THE FUNCTIONAL LINEAR MODEL

BY NADINE HILGERT, ANDRÉ MAS AND NICOLAS VERZELEN[1]

INRA, Université Montpellier 2 and INRA

We introduce two novel procedures to test the nullity of the slope function in the functional linear model with real output. The test statistics combine multiple testing ideas and random projections of the input data through functional principal component analysis. Interestingly, the procedures are completely data-driven and do not require any prior knowledge on the smoothness of the slope nor on the smoothness of the covariate functions. The levels and powers against local alternatives are assessed in a nonasymptotic setting. This allows us to prove that these procedures are minimax adaptive (up to an unavoidable $\log\log n$ multiplicative term) to the unknown regularity of the slope. As a side result, the minimax separation distances of the slope are derived for a large range of regularity classes. A numerical study illustrates these theoretical results.

1. Introduction. Consider the following functional linear regression model where the scalar response Y is related to a square integrable random function $X(\cdot)$ through

$$Y = \omega + \int_{\mathcal{T}} X(t)\theta(t)\,dt + \varepsilon. \tag{1}$$

Here, ω is a constant, denoting the intercept of the model, $\mathcal{T}$ is the domain of $X(\cdot)$, $\theta(\cdot)$ is an unknown function representing the slope function and ε is a centered random noise variable. In functional linear regression, much interest focuses on the nonparametric estimation of $\theta(\cdot)$ in (1), given an i.i.d. sample $(X_i, Y_i,)_{1\le i\le n}$ of (X, Y). Testing whether θ belongs to a given finite dimensional linear subspace $\mathcal{V}$ is a question that arises in different problems such as dimension reduction, goodness-of-fit analysis, or lack-of-effect tests of a functional variable. If the properties of estimators of θ are widely discussed in the literature, there is still a great need to have generic test procedures supported by strong theoretical properties. This is the problem addressed in the present paper.

Let us reformulate the functional model (1) as a generic linear regression model in an infinite dimensional space. The random function X is assumed to belong

Received June 2012; revised January 2013.

[1]Supported in part by the French Agence Nationale de la Recherche (ANR 2011 BS01 010 01 projet Calibration).

MSC2010 subject classifications. Primary 62J05; secondary 62G10.

Key words and phrases. Functional linear regression, eigenfunction, principal component analysis, adaptive testing, minimax hypothesis testing, minimax separation rate, multiple testing, ellipsoid, goodness-of-fit.

to some separable Hilbert space, henceforth denoted $\mathcal{H}$, endowed with the inner product $\langle \cdot, \cdot \rangle$. Examples of $\mathcal{H}$ include $\mathcal{L}^2([0,1])$ or Sobolev space $\mathcal{W}_2^m([0,1])$. For the sake of clarity, we consider that $\omega = 0$ and that X and Y are centered. Thus, assuming that θ also belongs to $\mathcal{H}$, the statistical model (1) is rephrased as

$$Y = \langle X, \theta \rangle + \varepsilon, \tag{2}$$

where ε is a centered random variable independent from X with unknown variance σ^2. In the sequel, we note $\mathbf{X}$ and $\mathbf{Y}$ the size n vectors of i.i.d. observations X_i and Y_i $(1 \le i \le n)$, while $\boldsymbol{\varepsilon}$ stands for the size n vector of the noise.

In essence, testing a linear hypothesis of the form "$\theta \in \mathcal{V}$" is as difficult as testing "$\theta = 0$" when a parametric estimator of θ in $\mathcal{V}$ is computed. Therefore we consider the following problem of testing:

$$H_0 : \text{“}\theta = 0\text{”} \quad \text{against} \quad H_1 : \text{“}\theta \neq 0\text{”}$$

given an i.i.d. sample $(\mathbf{X}, \mathbf{Y})$ from model (2). The extension to general subspaces $\mathcal{V}$ is developed in the discussion section.

Most testing procedures are based on ideas that have been originally developed for the estimation of θ. We briefly review the main approaches and the corresponding results in estimation.

A first class of procedures is based on the minimization of a least-square type criterion penalized by a roughness term that assesses the "plausibility" of θ. Such approaches include smoothing spline estimators [6, 13], thresholding projection estimators [8], or reproducing kernel Hilbert space methods [36]. A second class of procedures is based on the functional principal components analysis (PCA) of $\mathbf{X}$ [9, 21]. It consists of estimating θ in a finite dimensional space spanned by the k first eigenfunctions of the empirical covariance operator of $\mathbf{X}$. The main difference with the previous class of estimators lies in the fact that the finite dimensional space is estimated from the observations of the process X. See the survey [10] and references therein for an overview of these two approaches.

The theoretical properties of these classes of estimators have been investigated from different viewpoints: prediction [6, 9, 13, 36] (estimation of $\langle X_{n+1}, \theta \rangle$ where X_{n+1} follows the same distribution as X), pointwise prediction [4] (estimation of $\langle x, \theta \rangle$ for a fixed $x \in \mathcal{H}$) or the inverse problem [11, 21] (estimation of θ). For these three objectives, optimal rates of convergence have been derived, and some of the aforementioned procedures have been shown to asymptotically achieve this rate [4, 13, 21, 36]. Recently, some nonasymptotic results have emerged [11, 12] for estimation procedures that rely on a prescribed basis of functions (e.g., splines). Most of these estimation procedures rely on tuning parameters whose optimal value depend on quantities such as the noise variance, or the smoothness of θ. In fact, there is a longstanding gap in the literature between theory, where the variance σ^2, the smoothness of θ and the smoothness of the covariance operator of X are generally assumed to be known, and practice where they are unknown.

The literature on tests in the functional linear model is scarce. In [5], Cardot et al. introduced a test statistic based on the k first components of the functional PCA of $\mathbf{X}$. Its limiting distribution is derived under H_0 and the power of the corresponding test is proved to converge to one under H_1. The main drawback of the procedure is that the number k of components involved in the statistic has to be set. As for estimation, setting k is arguably a difficult problem. To bypass this calibration issue, one may apply a permutation approach [7] or use bootstrap methodologies [14, 20]. While the levels of the corresponding tests are asymptotically controlled, there is again no theoretical guarantee on the power.

In this paper, our objective is to introduce automatic testing procedures whose powers are optimal from a nonasymptotic viewpoint.

As a first step, we introduce in Section 3 Fisher-type nonadaptive tests, $T_{\alpha,k}$, corresponding to projections of $\mathbf{Y}$ on the k first principal components of $\mathbf{X}$. We study their levels and powers in Sections 3 and 4. Under moment assumptions on ε and mild assumptions on the covariance of X, the level is smaller than α up to a $\log^{-1}(n)$ additional term, and a sharp control of the power is provided. Such results are comparable to state of the art results in nonparametric regression [3, 33]. In our setting, the main difficulty in the proof is to control the randomness of the principal components of $\mathbf{X}$. The arguments rely on the perturbation theory of operators. While other estimation or testing procedures based on the Karhunen–Loève expansion have only been analyzed in an asymptotic setting [4, 5, 21], our nonasymptotic results rely on less restrictive assumptions on X than those commonly used in the literature. In Section 4, we assess the optimality of the parametric test $T_{\alpha,k}$ in the minimax sense. The notion of minimaxity of a level-α test T_α is related to the *separation distance* of T_α over some class of functions Θ (e.g., a Sobolev ball). Intuitively, the power of a reasonable test T_α should be large when the norm of θ is large while the power of T_α is close to α when θ is close to 0. For the problem of testing H_0 : "$\theta = 0$" against $H_{1,\Theta}$: "$\theta \in \Theta \setminus \{0\}$", the separation distance corresponds to the smallest distance ρ such that T_α rejects H_0 with probability larger than $1 - \beta$ for all $\theta \in \Theta$ whose norm is larger than ρ. The smaller the separation distance, the more powerful the test T_α is. The minimax separation distance over Θ is the smallest separation distance that is achieved by a level-α test. A test achieving this minimax separation distance is said to be minimax over Θ. In the nonparametric regression setting, minimax separation distances have been derived in an asymptotic [26–28] and a nonasymptotic [2] setting. In this paper, the separation distances of our testing procedures are nonasymptotically controlled. We derive minimax separation distance in the functional model (2) for a wide class of ellipsoids. We show that the parametric test $T_{\alpha,k}$ achieves the optimal rate of detection when the dimension k is suitably chosen.

In practice, the regularity of θ is unknown. However, the choice of k in $T_{\alpha,k}$ depends on unknown quantities such as the regularity of X or the regularity of θ. Thus, assuming a priori that the function θ belongs to a particular smoothness class Θ and building an optimal test over Θ may lead to poor performances, for instance,

if $\theta \notin \Theta$. For this reason, a more ambitious issue is to build a minimax adaptive testing procedure, that is, a procedure which is simultaneously minimax for a wide range of regularity classes Θ. Minimax adaptive testing procedures have already been studied in the nonparametric regression setting, from an asymptotic [33] and a nonasymptotic [3] viewpoint. As a second step, we combine the parametric tests $T_{\alpha,k}$ with multiple testing techniques in the spirit of [3]. Two such multiple testing procedures are introduced in Section 5. They are completely data-driven: no tuning parameters are required, whose optimal values depend on θ, the distribution of X or on σ. Their levels and powers are analyzed from a nonasymptotic viewpoint in Sections 5 and 6. We prove that our multiple testing procedures are simultaneously minimax over the class of ellipsoids aforementioned (up to an unavoidable $\log\log n$ factor). As in the estimation setting [21], the minimax separation distances involve the common regularity of θ and X.

The two multiple testing procedures are illustrated and compared by simulations in Section 7. Extensions of the approach are discussed in Section 8. Section 9 contains the main proofs while the lemmas involving perturbation theory are given in Section 10. All the technical and side results are postponed to appendices.

2. Preliminaries.

2.1. *Notation*. We remind that $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively, refer to the inner product and the corresponding norm in the Hilbert $\mathcal{H}$. In contrast, $\langle \cdot, \cdot \rangle_n$ and $\| \cdot \|_n$ stand for the inner product and the Euclidean norm in $\mathbb{R}^n$. Furthermore, $\otimes$ refers to the tensor product. We assume henceforth that X is centered and has a second moment that is $\mathbb{E}[\|\mathbf{X}\|^2] < \infty$. The covariance operator of X is defined as the linear operator Γ defined on $\mathcal{H}$ as follows:

$$\Gamma h = \mathbb{E}[X \otimes Xh] = \mathbb{E}\big[\langle h, X\rangle X\big], \qquad h \in \mathcal{H}.$$

It is well known that Γ is a symmetric, positive trace-class hence Hilbert–Schmidt operator, which implies that Γ is diagonalizable in an orthonormal basis. We denote $(\lambda_j)_{j\geq 1}$ the nonincreasing sequence of eigenvalues of Γ, while the sequence $(V_j)_{j\geq 1}$ stands for a corresponding sequence of eigenfunctions. It follows that Γ decomposes as $\Gamma = \sum_{j=1}^{\infty} \lambda_j V_j \otimes V_j$. For any integer $k \geq 1$, we note $\Gamma_k = \sum_{j=1}^{k} \lambda_j V_j \otimes V_j$ the operator such that $\Gamma_k h = \Gamma h$ for $h \in \mathrm{Vect}(V_1, \ldots, V_k)$ and $\Gamma_k h = 0$ if $h \in (V_1, \ldots, V_k)^{\perp}$.

In the sequel, $C, C_1, \ldots$ denote positive universal constants that may vary from line to line. The notation $C(\cdot)$ specifies the dependency on some quantities.

2.2. *Karhunen–Loève expansion and functional PCA*. We recall here a classical tool of functional data analysis: the *Karhunen–Loève expansion*, denoted KL expansion in the sequel.

DEFINITION 2.1. There exists an expansion of X in the basis $(V_j)_{j\geq 1}$: $X = \sum\langle X, V_j\rangle V_j$. The real random variables $\langle X, V_j\rangle$ are centered (when X is centered), uncorrelated and with variance λ_j. As a consequence, there exists a collection $(\eta^{(j)})_{j\geq 1}$ of random variables that are centered, uncorrelated and with unit variance such that

$$X = \sum_{j=1}^{+\infty} \sqrt{\lambda_j}\eta^{(j)} V_j. \tag{3}$$

The decomposition is called the KL-expansion of X.

The eigenfunction V_j is the jth principal direction whose amount of variance coincides with λ_j. When X is a Gaussian process, the $(\eta^{(j)})_{j\in\mathbb{N}}$ form an i.i.d. sequence with $\eta^{(1)} \sim \mathcal{N}(0, 1)$. If the eigenfunctions (V_j) and the eigenvalues (λ_j) are unknown in practice, they can be estimated from the data using functional principal component analysis. In the sequel, we note $\widehat{\Gamma}_n$ the empirical covariance operator defined by

$$\widehat{\Gamma}_n h = \frac{1}{n}\sum_{i=1}^{n} X_i \otimes X_i h = \frac{1}{n}\sum_{i=1}^{n}\langle X_i, h\rangle X_i, \qquad h \in \mathcal{H}.$$

Functional PCA allows us to estimate (λ_j, V_j), $j \geq 1$, by diagonalizing the empirical covariance operator $\widehat{\Gamma}_n$. These empirical counterparts of (λ_j, V_j) are denoted $(\widehat{\lambda}_j, \widehat{V}_j)$ in the sequel.

Functional PCA is usually applied as a dimension reduction technique. One of its appealing features relies on its ability to capture most of the variance of X by a k-dimensional projection on the space $\mathrm{Vect}(\widehat{V}_1, \ldots, \widehat{V}_k)$. For this reason, PCA is at the core of many procedures for functional data. After the seminal paper by Dauxois et al. [15], the convergence of the random eigenelements $(\widehat{\lambda}_j, \widehat{V}_j)$ has been assessed from an asymptotic point of view [22–24, 29]. One issue with such a dimension reduction method is the choice of the tuning parameter k, whose optimal value usually depends on unknown quantities. Besides plugging the $(\widehat{\lambda}_j, \widehat{V}_j)$ into linear estimates creates nonlinearity and usually introduces stochastic dependence.

3. Parametric test.

3.1. *Definition.* In the sequel, k denotes a positive integer smaller than $n/2$. As a first step, we consider the parametric testing problem of the hypotheses

$$H_0 : \text{“}\theta = 0\text{”} \quad \text{against} \quad H_{1,k} : \text{“}\theta \in \mathrm{Vect}\big[(V_j)_{j=1,\ldots,k}\big] \setminus \{0\}\text{”}. \tag{4}$$

Given a dimension k of the Karhunen–Loève expansion, we note $\hat{k}^{\mathrm{KL}}$ as $k \wedge \mathrm{Rank}(\widehat{\Gamma}_n)$. In order to introduce the parametric statistic, let us restate the functional linear model into a finite dimensional linear model. We consider the response vector $\mathbf{Y}$ of size n, the $n \times \hat{k}^{\mathrm{KL}}$ design matrix $\mathbf{W}$ defined by $\mathbf{W}_{i,j} = \langle X_i, \widehat{V}_j\rangle$

for $i = 1, \dots, n$, $j = 1, \dots, \hat{k}^{\mathrm{KL}}$, the parameter vector ϑ defined by $\vartheta_j = \langle \theta, \widehat{V}_j \rangle$, $j = 1, \dots, \hat{k}^{\mathrm{KL}}$ and the size n noise vector $\tilde{\boldsymbol{\varepsilon}}$ defined by $\tilde{\varepsilon}_i = \varepsilon_i + \langle X_i, \theta \rangle - [\mathbf{W}\vartheta]_i$. The functional linear model is equivalently written as

$$\mathbf{Y} = \mathbf{W}\vartheta + \tilde{\boldsymbol{\varepsilon}}.$$

Intuitively, testing "$\vartheta = 0$" is a reasonable proxy for testing H_0 against $H_{1,k}$. For this reason, we propose a Fisher-type statistic.

DEFINITION 3.1. In the sequel, $\widehat{\mathbf{\Pi}}_k$ stands for the orthogonal projection in $\mathbb{R}^n$ onto the space generated by the $\hat{k}^{\mathrm{KL}}$ columns of $\mathbf{W}$. For any $k \le n/2$, we consider the statistic $\phi_k(\mathbf{Y}, \mathbf{X})$ defined by

$$\phi_k(\mathbf{Y}, \mathbf{X}) := \frac{\|\widehat{\mathbf{\Pi}}_k \mathbf{Y}\|_n^2}{\|\mathbf{Y} - \widehat{\mathbf{\Pi}}_k \mathbf{Y}\|_n^2 / (n - \hat{k}^{\mathrm{KL}})}. \tag{5}$$

The main difference with a classical Fisher statistic comes from the fact that the projection $\widehat{\mathbf{\Pi}}_k$ is *random*. This projector is built using the $\hat{k}^{\mathrm{KL}}$ first directions $(\widehat{V}_1, \widehat{V}_2, \dots, \widehat{V}_{\hat{k}^{\mathrm{KL}}})$ of the empirical Karhunen–Loève expansion of $\mathbf{X}$. Let us call $\widetilde{\mathbf{\Pi}}_k$ the orthogonal projector in $\mathbb{R}^n$ onto the space spanned by $(\langle X_i, V_j \rangle)_{i=1,\dots,n}$, $j = 1, \dots, k$. If we knew the basis (V_j), $j \ge 1$, in advance, we would use this orthogonal projector instead of $\widehat{\mathbf{\Pi}}_k$. We shall prove that, under H_0, $\phi_k(\mathbf{Y}, \mathbf{X})/\hat{k}^{\mathrm{KL}}$ behaves like a Fisher distribution with $(\hat{k}^{\mathrm{KL}}, n - \hat{k}^{\mathrm{KL}})$ degrees of freedom.

DEFINITION 3.2 (Parametric tests). Fix $\alpha \in (0, 1)$. We reject H_0 against $H_{1,k}$ when the statistic

$$T_{\alpha,k} := \phi_k(\mathbf{Y}, \mathbf{X}) - \hat{k}^{\mathrm{KL}} \bar{\mathcal{F}}^{-1}_{\hat{k}^{\mathrm{KL}}, n - \hat{k}^{\mathrm{KL}}}(\alpha) \tag{6}$$

is positive.

REMARK 3.1 [Other interpretations of $\phi_k(\mathbf{Y}, \mathbf{X})$]. Consider $\widehat{\theta}_k$ the least-squares estimator of θ in the space generated by $\widehat{V}_j$, $j = 1, \dots, \hat{k}^{\mathrm{KL}}$. It is proved in Section 9.2 that $\|\widehat{\mathbf{\Pi}}_k \mathbf{Y}\|_n^2 = \|\widehat{\Gamma}_n^{1/2} \widehat{\theta}_k\|^2$. Thus, the numerator of (5) corresponds to some norm of $\widehat{\theta}_k$. Intuitively, the larger $\widehat{\theta}_k$, the larger the statistic $\phi_k(\mathbf{Y}, \mathbf{X})$ is. Furthermore, $\|\widehat{\mathbf{\Pi}}_k \mathbf{Y}\|_n^2$ also expresses as the numerator of the statistic D_n considered in Cardot et al. [5]; see Section 9.2 for details.

REMARK 3.2. From the considerations above, we see that the transformed parameter $\Gamma^{1/2}\theta$ naturally occurs in the definition of $\phi_k(\mathbf{Y}, \mathbf{X})$. In fact, hypotheses H_0 and $H_{1,k}$ remain unchanged, if we replace θ by $\Gamma^{1/2}\theta$ in (4) as soon as Γ is injective. The crucial role of this synthetic parameter is underlined in [30] where the functional linear regression model is proved to be asymptotically equivalent to a white noise model with signal $\Gamma^{1/2}\theta$.

3.2. *Size.* We study the type I error of the parametric tests $T_{\alpha,k}$. On one hand, we control exactly the size of the tests when the noise ε is normally distributed. On the other hand, we bound the size of the tests when the noise is only constrained to admit a fourth moment.

3.2.1. *Gaussian noise.*

(A.1) ε follows a Gaussian distribution $\mathcal{N}(0, \sigma^2)$.

PROPOSITION 3.3 (Size of $T_{\alpha,k}$ under Gaussian errors). *Under assumption* (A.1) *and if* $k \le n/2$, *we have for any* $n \ge 2$, $\mathbb{P}_0(T_{\alpha,k} > 0) = \alpha$.

Observe that this control does not require any assumption on the process X.

3.2.2. *Non-Gaussian noise.* In this part, the noise ε is only assumed to admit a fourth order moment, but we perform additional assumptions on X and k.

$$\text{(B.1)} \qquad \sup_{j\ge 1} \mathbb{E}\big[\big(\eta^{(j)}\big)^4\big] \le C_1 \quad \text{and} \quad \frac{\mathbb{E}[\varepsilon^4]}{\sigma^4} \le C_2,$$

where C_1 and C_2 are two positive constants:

(B.2) For some $\gamma > 0$ $\big(j\lambda_j\big((\log^{1+\gamma} j) \vee 1\big)\big)_{j\ge 1}$ is decreasing and $\operatorname{Ker}\Gamma = \{0\}$.

$$\text{(B.3)} \qquad k \le n^{1/4}/\log^4(n).$$

Assumption (B.1) is classical, since we need to control second order moments for the empirical covariance operator $\widehat{\Gamma}_n$. This comes down to inspecting the behavior of the fourth order moments of the $\eta^{(j)}$'s. The second part of (B.2) ensures that the framework is truly functional. The first part of (B.2) is mild and holds for an X that may have very irregular paths (it holds for the Brownian motion for which $\lambda_j \propto j^{-2}$) and for classical examples of eigenvalue sequences: with polynomial decay, exponential decay, or even Laurent sequences such as $\lambda_j = j^{-\delta} \cdot \log^{-\nu}(j)$ for $\delta > 1$ and $\nu \ge 0$. In fact, (B.2) is less restrictive than assumptions commonly used in the literature [4, 5, 21] since it does not require any spacing control between the eigenvalues.

The restriction (B.3) on the dimension of the projection is classical for the analysis of statistical procedures based on the Karhunen–Loève expansion. If we knew the eigenfunctions V_k of Γ in advance, we could consider larger dimensions k. The estimation of the eigenfunctions V_k becomes more difficult when k increases. By considering dimensions k that satisfy assumption (B.3), we prove in the next theorem that the *random* projector $\widehat{\mathbf{\Pi}}_k$ concentrates well around its mean. It may

be noticed that this assumption links k and n independently from the eigenvalues hence from any prior knowledge on the data.

THEOREM 3.4 (Size of $T_{\alpha,k}$). *Under assumptions* (B.1)–(B.3), *there exist positive constants* $C(\alpha,\gamma)$ *and* C_2 *such that the following holds. For any* $n \geq C_2$, *we have*

$$\mathbb{P}_0[T_{\alpha,k} > 0] \leq \alpha + \frac{C(\alpha,\gamma)}{\log(n)}.$$

REMARK 3.3. In the proof of Theorem 3.4, we show that, under H_0, the distribution of $\phi_k(\mathbf{Y},\mathbf{X})$ is close to a χ^2 distribution with k degrees of freedom. The arguments rely on perturbation theory for random operators (see Section 10).

4. Power and minimaxity of $T_{\alpha,k}$. Intuitively, the larger the signal-to-noise ratio $\mathbb{E}[\langle X,\theta\rangle^2]/\sigma^2 = \|\Gamma^{1/2}\theta\|^2/\sigma^2$ is, the easier we can reject H_0. For this reason, we study how large $\|\Gamma^{1/2}\theta\|^2$ has to be, so that the test $T_{\alpha,k}$ rejects H_0 with probability larger than $1-\beta$ for a prescribed positive number β. We provide such type II errors under moment assumption of ε. Additional controls of the power when ε follows a Normal distribution are stated in Appendix A of the supplementary material [25].

4.1. *Power of* $T_{\alpha,k}$.

$$\sup_{j\geq 1}\mathbb{E}\big[\big(\eta^{(j)}\big)^8\big] \leq C. \tag{B.4}$$

THEOREM 4.1 (Power under non-Gaussian errors). *Let* α *and* β *be fixed. Under* (B.1)–(B.4), *there exist positive constants* $C(\gamma)$, C_1, C_2 *and* C_3 *such that the following holds. Assume that* $\alpha \geq e^{-\sqrt{n}}$, $\beta \geq C(\gamma)/\log(n)$ *and that* $n \geq C_3$. *Then*, $\mathbb{P}_\theta(T_{\alpha,k} > 0) \geq 1-\beta$ *for any* θ *satisfying*

$$\|\Gamma^{1/2}\theta\|^2 \geq C_1\|(\Gamma^{1/2}-\Gamma_k^{1/2})\theta\|^2 + C_2\frac{\sigma^2}{n}\left(\sqrt{k\log\left(\frac{1}{\alpha\beta}\right)} + \log\left(\frac{1}{\alpha\beta}\right)\right). \tag{7}$$

REMARK 4.1. If we knew that θ belongs to the space spanned by the k first eigenvectors $(V_1,\ldots,V_k)$ and if we knew these k eigenvectors in advance, then we could consider the statistic defined by

$$\widetilde{\phi}_k(\mathbf{X},\mathbf{Y}) := \frac{\|\widetilde{\mathbf{\Pi}}_k\mathbf{Y}\|_n^2}{\|\mathbf{Y}-\widetilde{\mathbf{\Pi}}_k\mathbf{Y}\|_n^2} - \bar{\mathcal{F}}_{k,n-k}^{-1}(\alpha),$$

where $\widetilde{\mathbf{\Pi}}_k$ is the projection in $\mathbb{R}^n$ onto the space spanned by $(\langle X_i, V_j\rangle)_{i=1,\dots,n}$, $j = 1, \dots, k$. The corresponding test is optimal in the minimax sense and rejects H_0 with probability larger than $1 - \beta$ when

$$\|\Gamma^{1/2}\theta\|^2 \geq C(\alpha, \beta)\sqrt{k}\sigma^2/n. \tag{8}$$

See [35] for a proof when X is a Gaussian process and ε follows a Gaussian distribution, the extension to non-Gaussian processes being straightforward. In (7), we recover an additional term $\|(\Gamma^{1/2} - \Gamma_k^{1/2})\theta\|^2$ because we do not assume that θ belongs to the space spanned by $(V_1, \dots, V_k)$. The statistic $\phi_k(\mathbf{Y}, \mathbf{X})$ only captures the projection of θ onto $\operatorname{span}(V_1, \dots, V_k)$. In fact, the test $T_{\alpha,k}$ rejects with large probability when

$$\|\Gamma_k^{1/2}\theta\|^2 = \|\Gamma^{1/2}\theta\|^2 - \|(\Gamma^{1/2} - \Gamma_k^{1/2})\theta\|^2$$

is large.

REMARK 4.2 (Joint regularity of Γ and θ). Looking more precisely at the bias term, we obtain

$$\|(\Gamma^{1/2} - \Gamma_k^{1/2})\theta\|^2 = \sum_{j=k+1}^{\infty} \lambda_j \langle \theta, V_j\rangle^2.$$

Consequently, the bias term does not only depend on the rate of convergence of the eigenvalues of Γ, it also depends on the behavior of the sequence $\lambda_j\langle\theta, V_j\rangle^2$. In other words, the joint regularity of the covariance operator Γ and of θ [in the expansion of (V_j), $j \geq 1$] plays a role in the bias term. For a fixed θ, the power of $T_{\alpha,k}$ is large for a tuning parameter k that achieves a trade-off between the bias term $\|(\Gamma^{1/2} - \Gamma_k^{1/2})\theta\|^2$ and a variance term $\sqrt{k}\sigma^2/n$.

4.2. *Minimax separation distance over an ellipsoid.* In this section, we assess the optimality of the procedure $T_{\alpha,k}$. To this end, we study the optimal power of a level-α test, when θ is assumed to have a known regularity.

DEFINITION 4.2 (Ellipsoids). Given a nonincreasing sequence $(a_i)_{i\geq 1}$ and a positive number $R > 0$, we define the ellipsoid $\mathcal{E}_a(R)$ by

$$\mathcal{E}_a(R) := \left\{\theta \in \mathcal{H} : \sum_{k=1}^{+\infty} \frac{\langle \theta, V_k\rangle^2}{a_k^2} \leq R^2\sigma^2\right\}.$$

The ellipsoid $\mathcal{E}_a(R)$ contains all the elements $\theta \in \mathcal{H}$ that have a given regularity in the basis (V_k), $k \geq 1$. In other words, it prescribes the rate of convergence of $\langle\theta, V_k\rangle$ toward 0. The faster a_k goes to zero, the more regular θ is assumed to be.

We take some positive numbers α and β such that $\alpha + \beta < 1$. Let us consider a test T taking its values in $\{0, 1\}$. For any subset $\mathcal{C} \subset \mathcal{H} \times \mathbb{R}^+$, $\boldsymbol{\beta}[T; \mathcal{C}]$ denotes the supremum of type II errors of an α-level test T for all parameters $(\theta, \sigma) \in \mathcal{C}$,

$$\boldsymbol{\beta}[T; \mathcal{C}] := \sup_{(\theta,\sigma)\in\mathcal{C}} \mathbb{P}_\theta[T = 0].$$

The (α, β)-*separation distance* of the test T over the ellipsoid $\mathcal{E}_a(R)$, noted $\rho[T; \mathcal{E}_a(R)]$ is the minimal number $\rho > 0$ such that T rejects H_0 with probability larger than $1 - \beta$ for all $\theta \in \mathcal{E}_a(R)$ and $\sigma > 0$ such that $\|\Gamma^{1/2}\theta\|^2/\sigma^2 \geq \rho^2$. Hence, $\rho[T; \mathcal{E}_a(R)]$ corresponds to the minimal distance such that the hypotheses $\{\theta = 0, \sigma > 0\}$ and $\{\theta \in \mathcal{E}_a(R), \sigma > 0, \|\Gamma^{1/2}\theta\|^2/\sigma^2 \geq \rho^2\}$ are well separated by T:

$$\rho\big[T; \mathcal{E}_a(R)\big] := \inf\left\{\rho > 0, \boldsymbol{\beta}\left[T; \left\{\theta \in \mathcal{E}_a(R), \sigma > 0, \frac{\|\Gamma^{1/2}\theta\|^2}{\sigma^2} \geq \rho^2\right\}\right] \leq \beta\right\}.$$

By definition, T has a power larger than $1 - \beta$ for all $\theta \in \mathcal{E}_a(R)$ and $\sigma > 0$ such that $\|\Gamma^{1/2}\theta\|^2/\sigma^2 \geq \rho^2[T, \mathcal{E}_a(R)]$.

DEFINITION 4.3 (Minimax separation distance). We consider

$$\rho^*\big[\alpha; \mathcal{E}_a(R)\big] := \inf_{T_\alpha} \rho\big[T_\alpha; \mathcal{E}_a(R)\big], \tag{9}$$

where the infimum run over all level-α tests. This quantity is called the (α, β)-*minimax separation distance* over the ellipsoid $\mathcal{E}_a(R)$.

REMARK 4.3. The notion of (α, β)-minimax separation distance is a nonasymptotic counterpart of the detection boundaries studied in the Gaussian sequence model [16]. Furthermore, as the variance σ^2 is unknown, this definition of the minimax separation distance considers the power of the testing procedures for all possible values of σ^2.

PROPOSITION 4.4 (Minimax lower bound over an ellipsoid). *There exists a constant $C(\alpha, \beta)$ such that the following holds. Let us assume that X is a Gaussian process and that ε follows a Gaussian distribution. For any ellipsoid $\mathcal{E}_a(R)$, we have*

$$\rho^*\big[\alpha; \mathcal{E}_a(R)\big] \geq \rho^2_{a,R,n} := \sup_{k\geq 1}\left[C(\alpha, \beta)\left(\frac{\sqrt{k}}{n}\right) \wedge \big(R^2 a_k^2 \lambda_k\big)\right]. \tag{10}$$

In other words, for any test T_α of level α, we have

$$\boldsymbol{\beta}\left[T_\alpha; \left\{\theta \in \mathcal{E}_a(R), \sigma > 0, \frac{\|\Gamma^{1/2}\theta\|^2}{\sigma^2} \geq \rho^2_{a,R,n}\right\}\right] \geq \beta.$$

Consequently, the (α, β) minimax-separation distance over $\mathcal{E}_a(R)$ is lower bounded by $\rho^2_{a,R,n}$. The next proposition states the corresponding upper bound.

COROLLARY 4.5 (Minimax upper bound). *Under* (B.1), (B.2) *and* (B.4), *there exists positive constants* $C(\gamma)$, C_2, $C_3(\alpha,\gamma)$ *and* $C_4(\alpha,\beta)$ *such that the following holds. Given an ellipsoid* $\mathcal{E}_a(R)$, *we define*

$$k_n^* := \inf\left\{k \geq 1, a_k^2 \lambda_k R^2 \leq \frac{\sqrt{k}}{n}\right\}. \tag{11}$$

Assume that $\alpha \geq e^{-\sqrt{n}}$, $\beta \geq C(\gamma)/\log(n)$, $n \geq C_2$ *and* $k_n^* \leq n^{1/4}/\log^4(n)$. *Then the test* T_{α,k_n^*} *has a size smaller than* $\alpha + C_3(\alpha,\gamma)/\log(n)$ *and is minimax over* $\mathcal{E}_a(R)$,

$$\boldsymbol{\beta}\left[T_{\alpha,k_n^*}; \left\{\theta \in \mathcal{E}_a(R), \sigma > 0, \frac{\|\Gamma^{1/2}\theta\|^2}{\sigma^2} \geq C_4(\alpha,\beta)\rho_{a,R,n}^2\right\}\right] \leq \beta. \tag{12}$$

This corollary is a straightforward consequence of Theorem 4.1. Hence, the test T_{α,k_n^*} is minimax over $\mathcal{E}_a(R)$; that is, its (α,β)-separation distance equals (up to a multiplicative constant) the (α,β) minimax separation distance. Interestingly, upper bound (12) does not require the error ε to be normally distributed.

REMARK 4.4. As a consequence, the (α,β)-minimax separation distance over $\mathcal{E}_a(R)$ is of order

$$\rho_{a,R,n}^2 := \sup_{k\geq 1}\left[C(\alpha,\beta)\left(\frac{\sqrt{k}}{n}\right) \wedge \left(R^2 a_k^2 \lambda_k\right)\right].$$

It depends on the behavior of the nonincreasing sequence $(\lambda_k a_k^2)$, where the sequence of eigenvalues (λ_k) prescribes the "regularity" of the process X and the sequence (a_k) prescribes the regularity of θ. In order to grasp the quantity $\rho_{a,R,n}^2$, let us specify some examples of sequences $\lambda_k a_k^2$:

COROLLARY 4.6. Polynomial decay. *If* $\lambda_k a_k^2 = k^{-s}$ *with* $s > 7/2$, *then the* (α,β)*-minimax separation is of order* $R^{2/(1+2s)} n^{-2s/(1+2s)}$. *This rate is achieved by the test* $T_{\alpha,k}$ *with* $k \asymp (R^2 n)^{2/(1+2s)}$.

Exponential decay. *If* $\lambda_k a_k^2 = e^{-sk}$ *with* $s > 0$, *then the* (α,β)*-separation distance of* $T_\alpha^{(1)}$ *over* $\mathcal{E}_a(R)$ *is of order* $\frac{\sqrt{\log(n)}}{\sqrt{sn}}$. *This rate is achieved by the test* $T_{\alpha,k}$ *with* $k \asymp \log(n)/s$.

REMARK 4.5. The condition $s > 7/2$ in the polynomial regime arises because of assumption (B.3) [$k \leq n^{1/4}/\log^4(n)$]. This restriction is related to the difficulty to reliably estimate the eigenvalues λ_k and eigenfunctions V_k when k is large; see Lemma 9.7 and its proof. If the process X and the function θ are less regular ($s < 7/2$), our theory only allows us to take $k = n^{1/4}/\log^4(n)$ in $T_{\alpha,k}$ which leads to a rate of testing of order (up to log terms) $n^{-s/4}R^2$ while the minimax lower bound is of order $R^{2/(1+2s)} n^{-2s/(1+2s)}$. Note that similar restrictions also occur in

"state of the art" results for estimation. For instance, condition (3.3) in Hall and Horowitz [21] amounts to $s > 3$.

In conclusion, $T_{\alpha,k}$ achieves the optimal rate of detection when k is suitably chosen. However, the choice of k depends on unknown quantities such as the regularity of X or the regularity of θ. Taking k too small does not allow one to detect nonzero θ such that the bias $\|(\Gamma^{1/2} - \Gamma_k^{1/2})\theta\|^2$ in (7) is too large. In contrast, taking k too large leads to a large variance term $\sqrt{k}/n$ in (7). The best k corresponds to the trade-off between the bias term and the variance term in (7). In the following, we introduce a procedure that nearly achieves this trade-off without requiring any prior knowledge of the regularity of Γ or θ.

5. A multiple testing procedure.

5.1. *Definition.* In the sequel, $\mathcal{K}_n$ stands for a "dyadic" collection of dimensions defined by

$$\mathcal{K}_n = \{2^0, 2^1, 2^2, 2^3, \ldots, \bar{k}_n\}, \tag{13}$$

where $\bar{k}_n$ is a power of 2 that will be fixed later. As k cannot be a priori chosen, we evaluate the statistic $\phi_k(\mathbf{Y}, \mathbf{X})$ for all k belonging to a collection $\mathcal{K}_n$. This choice of the collection $\mathcal{K}_n$ is discussed in the next section.

DEFINITION 5.1 (KL-test). We reject H_0: "$\theta = 0$" when the statistic

$$T_\alpha := \sup_{k \in \mathcal{K}_n, k \le \mathrm{Rank}(\widehat{\Gamma}_n)} \big[\phi_k(\mathbf{Y}, \mathbf{X}) - \hat{k}^{\mathrm{KL}} \bar{\mathcal{F}}^{-1}_{\hat{k}^{\mathrm{KL}}, n-\hat{k}^{\mathrm{KL}}}\{\alpha_{\mathcal{K}_n}(\mathbf{X})\}\big] \tag{14}$$

is positive, where the weight $\alpha_{\mathcal{K}_n}(\mathbf{X})$ is chosen according to one of the procedures P_1 and P_2 explained below.

P_1: (Bonferroni) $\alpha_{\mathcal{K}_n}(\mathbf{X})$ is equal to $\alpha/|\mathcal{K}_n|$.

P_2: Let $\mathbf{Z}$ be a standard Gaussian vector of size n. We take $\alpha_{\mathcal{K}_n}(\mathbf{X}) = q_{\mathbf{X},\alpha}$, the α-quantile of the distribution of the random variable

$$\inf_{k \in \mathcal{K}_n} \bar{\mathcal{F}}_{\hat{k}^{\mathrm{KL}}, n-\hat{k}^{\mathrm{KL}}}\big[\phi_k(\mathbf{Z}, \mathbf{X})/\hat{k}^{\mathrm{KL}}\big] \tag{15}$$

conditionally to $\mathbf{X}$.

In the sequel, $T_\alpha^{(1)}$ (resp., $T_\alpha^{(2)}$) refers to the statistic T_α, defined with procedure P_1 (resp., P_2). $T_\alpha^{(1)}$ corresponds to a Bonferroni multiple testing procedure. In contrast $T_\alpha^{(2)}$ handles better the dependence between the statistics ϕ_k, by using an ad-hoc quantile $q_{\mathbf{X},\alpha}$. We compare these two tests in Section 5.3. This multiple testing approach has already been considered in the nonparametric fixed design regression setting [3].

REMARK 5.1 (Computation of $q_{\mathbf{X},\alpha}$). Let Z be a standard Gaussian random vector of size n independent of $\mathbf{X}$. As $\boldsymbol{\varepsilon}$ is independent of $\mathbf{X}$, the distribution of (15) conditionally to $\mathbf{X}$ is the same as the distribution of

$$\inf_{k\in\mathcal{K}_n} \bar{\mathcal{F}}_{\hat{k}^{\mathrm{KL}},n-\hat{k}^{\mathrm{KL}}}\left(\frac{\|\widehat{\mathbf{\Pi}}_k Z\|_n^2/\hat{k}^{\mathrm{KL}}}{\|Z-\widehat{\mathbf{\Pi}}_k Z\|_n^2/(n-\hat{k}^{\mathrm{KL}})}\right)$$

conditionally to $\mathbf{X}$. As a consequence, one can simulate a random variable that follows the same distribution as (15) conditionally to $\mathbf{X}$. Hence, the quantile $q_{\mathbf{X},\alpha}$ is easily worked out applying a Monte Carlo approach.

REMARK 5.2 (Choice of $\bar{k}_n$). In practice, our advice is to take $\bar{k}_n = 2^{\lfloor \log_2 n\rfloor -1}$ which lies between $n/4$ and $n/2$. This choice is supported by practical experiences and results obtained in Section 5.2 and in Appendix A of [25]. Nevertheless, some of the theoretical results will require taking a slightly smaller value for $\bar{k}_n$.

5.2. *Size of the tests.*

PROPOSITION 5.2 (Size of $T_\alpha^{(1)}$ and $T_\alpha^{(2)}$ under Gaussian errors). *Under assumption* (A.1) *and if* $\bar{k}_n \le n/2$, *we have for any* $n \ge 2$,

$$\mathbb{P}_0\big(T_\alpha^{(1)} > 0\big) \le \alpha, \qquad \mathbb{P}_0\big(T_\alpha^{(2)} > 0\big) = \alpha.$$

If the noise ε follows a Gaussian distribution, the size of $T_\alpha^{(2)}$ is exactly α, while the size of $T_\alpha^{(1)}$ is smaller than α because of the Bonferroni correction. Let us now control the size of $T_\alpha^{(1)}$ assuming that ε admit a finite fourth moment:

(B′.3) $$\bar{k}_n \le n^{1/4}/\log^4(n).$$

The assumption (B′.3) is the counterpart of (B.3) for a multiple testing procedure. Next, we state the counterpart of Theorem 3.4 for $T_\alpha^{(1)}$.

THEOREM 5.3 (Size of $T_\alpha^{(1)}$). *Under assumptions* (B.1), (B.2) *and* (B′.3), *there exist positive constants* $C(\alpha,\gamma)$ *and* C_2 *such that the following holds. For any* $n \ge C_2$, *we have*

$$\mathbb{P}_0\big[T_\alpha^{(1)} > 0\big] \le \alpha + \frac{C(\alpha,\gamma)}{\log(n)}.$$

5.3. *Comparison of* $T_\alpha^{(1)}$ *and* $T_\alpha^{(2)}$. The test $T_\alpha^{(2)}$ is always more powerful than $T_\alpha^{(1)}$ as shown in the next proposition.

PROPOSITION 5.4. *For any parameter $\theta \neq 0$, the tests $T_\alpha^{(1)}$ and $T_\alpha^{(2)}$ satisfy*

$$\mathbb{P}_\theta\big(T_\alpha^{(2)} > 0 | \mathbf{X}\big) \geq \mathbb{P}_\theta\big(T_\alpha^{(1)} > 0 | \mathbf{X}\big)\mathbf{X} \qquad a.s. \tag{16}$$

On one hand, the choice of procedure P_1 is valid even for a non-Gaussian noise and avoids the computation of the quantile $q_{\mathbf{X},\alpha}$. On the other hand, the test $T_\alpha^{(2)}$ has a size of exactly α when the error is Gaussian and is more powerful than the corresponding test with procedure P_1. This comparison is numerically illustrated in Section 7.

6. Power and adaptation of $T_\alpha^{(1)}$. Since $T_\alpha^{(2)}$ is always more powerful than $T_\alpha^{(1)}$, we only consider the power and the minimax optimality of $T_\alpha^{(1)}$.

THEOREM 6.1 (Power under non-Gaussian errors). *Let α and β be fixed. Under* (B.1), (B.2), (B′.3), (B.4), *there exist positive constants $C(\gamma)$, C_1, C_2 and C_3 such that the following holds. Assume that $\alpha \geq e^{-\sqrt{n}}$, $\beta \geq C(\gamma)/\log(n)$ and that $n \geq C_3$. Then, $\mathbb{P}_\theta(T_\alpha^{(1)} > 0) \geq 1 - \beta$ for any θ satisfying*

$$\begin{aligned}\big\|\Gamma^{1/2}\theta\big\|^2 \geq \inf_{k\in\mathcal{K}_n} C_1\big\|(\Gamma^{1/2} - \Gamma_k^{1/2})\theta\big\|^2 \\ + C_2\frac{\sigma^2}{n}\left(\sqrt{k\log\left(\frac{\log n}{\alpha\beta}\right)} + \log\left(\frac{\log n}{\alpha\beta}\right)\right).\end{aligned} \tag{17}$$

REMARK 6.1. Comparing Theorems 4.1 and 6.1, we observe that the rejection region $T_\alpha^{(1)}$ almost contains all the rejection regions of the tests $T_{\alpha,k}$ for all $k \in \mathcal{K}_n$. The price to pay for this feature is an additional $\sqrt{\log\log n}$ in the variance term of (17),

$$\frac{\sigma^2}{n}\left(\sqrt{k\log\left(\frac{\log n}{\alpha\beta}\right)} + \log\left(\frac{\log n}{\alpha\beta}\right)\right).$$

This $\log\log(n)$ term corresponds to the quantity $\log(|\mathcal{K}_n|)$. If we had used a collection of the form $\{1, \ldots, \bar{k}_n\}$ instead of $\mathcal{K}_n$, the $\log\log(n)$ would have been replaced by a $\log(n)$. We prove below that this $\log\log n$ term is in fact unavoidable for an adaptive procedure.

As for $T_{\alpha,k}$, additional controls of the power when ε follows a Normal distribution are stated in Appendix A of [25]. Let us now consider the power of $T_\alpha^{(1)}$ over ellipsoids $\mathcal{E}_a(R)$. In the sequel, $\lfloor\cdot\rfloor$ stands for the integer part, while $\log_2(\cdot)$ corresponds to the binary logarithm.

COROLLARY 6.2 (Power of $T_\alpha^{(1)}$ over ellipsoids). *Under* (B.1), (B.2) *and* (B.4), *there exist positive constants $C(\gamma)$, C_1, C_2, $C_3(\alpha,\beta)$ and $C_4(\alpha,\beta)$ such that*

the following holds. Assume that $\alpha \geq e^{-\sqrt{n}}$, *that* $\beta \geq C(\gamma)/\log(n)$ *and* $n \geq C_2$. *Consider the test* $T_\alpha^{(1)}$ *with* $\bar{k}_n = 2^{\lfloor \log_2[n^{1/4}/\log^4(n)] \rfloor}$. *Fix any ellipsoid* $\mathcal{E}_a(R)$.

(1) *We have* $\mathbb{P}_\theta(T_\alpha^{(1)} > 0) \geq 1 - \beta$ *for any* $\theta \in \mathcal{E}_a(R)$ *satisfying*

$$\frac{\|\Gamma^{1/2}\theta\|^2}{\sigma^2} \geq C_3(\alpha, \beta) \inf_{k=1,2,4,\ldots,\bar{k}_n} \left[\lambda_{k+1} a_{k+1}^2 R^2 + \frac{\sigma^2}{n} (\sqrt{k \log\log n} + \log\log n) \right].$$

(2) *Consider* k_n^* *as in* (11). *If* $\log\log(n) \leq k_n^* \leq \bar{k}_n$, *then* $\mathbb{P}_\theta(T_\alpha^{(1)} > 0) \geq 1 - \beta$ *for any* $\theta \in \mathcal{E}_a(R)$ *satisfying*

$$\frac{\|\Gamma^{1/2}\theta\|^2}{\sigma^2} \geq C_4(\alpha, \beta) \sqrt{\log\log n}\, \rho_{a,R,n}^2,$$

where $\rho_{a,R,n}$ *is defined in* (10).

This is a direct consequence of Theorem 4.1.

REMARK 6.2. If we compare Corollary 6.2 with the minimax lower bound of Proposition 4.4, we observe that the separation distance only matches up to a factor of order $\sqrt{\log\log n}$. As a consequence, $T_{\alpha,k}$ is almost minimax over all ellipsoids $\mathcal{E}_a(R)$ satisfying $\log\log(n) \leq k_n^* \leq \bar{k}_n$. Next, we prove that this $\sqrt{\log\log(n)}$ term loss is unavoidable when the ellipsoid $\mathcal{E}_a(R)$ is unknown.

PROPOSITION 6.3 (Minimax lower bounds over a collection of nested ellipsoids). *There exists a positive constant* $C(\alpha, \beta)$ *such that the following holds. Let us assume that* X *is a Gaussian process, that the noise* ε *follows a Gaussian distribution, and that the rank of* Γ *is infinite. For any ellipsoid* $\mathcal{E}_a(R)$, *we set*

$$\tilde{\rho}_{a,R,n}^2 := \sup_{k \geq 1} \left[C(\alpha, \beta) \left(\frac{\sqrt{\log\log(k \vee 3)}\sqrt{k}}{n} \right) \wedge \left(R^2 a_k^2 \lambda_k \right) \right].$$

For any nonincreasing sequence $(a_k)_{k \geq 1}$ *and any test* T *of level* α, *we have*

$$\boldsymbol{\beta}\left[T; \bigcup_{R>0} \left\{ \theta \in \mathcal{E}_a(R), \sigma > 0, \frac{\|\Gamma^{1/2}\theta\|^2}{\sigma^2} \geq \tilde{\rho}_{a,R,n}^2 \right\} \right] \geq \beta.$$

As a consequence, there is a $\sqrt{\log\log n}$ price to pay if we simultaneously consider a nested collection of ellipsoids. Such impossibility for perfect adaptation has already been observed for the testing problem in the classical nonparametric regression framework [33].

REMARK 6.3. In order to compare the lower and upper bounds of Proposition 6.3 and Corollary 6.2, let us specify the sequence $\lambda_k a_k^2$:

- Polynomial decay. If $\lambda_k a_k^2 = k^{-s}$, then the (α, β)-separation distance of $T_\alpha^{(1)}$ over $\mathcal{E}_a(R)$ is of order

$$R^{2/(1+2s)}\left(\frac{\sqrt{\log\log(n)}}{n}\right)^{2s/(1+2s)}$$

for $s > 7/2$. By Proposition 6.3, this rate is optimal for adaptation.

- Exponential decay. If $\lambda_k a_k^2 = e^{-sk}$, then the (α, β)-separation distance of $T_\alpha^{(1)}$ over $\mathcal{E}_a(R)$ is of order

$$\frac{\sqrt{\log(n)\log\log(n)}}{\sqrt{s}n}$$

for any $s > 0$. By Proposition 6.3, this rate is almost optimal for adaptation [up to a $\sqrt{\log\log(n)/\log\log\log n}$ term].

In conclusion, the procedure $T_\alpha^{(1)}$ is adaptive to the unknown regularity of θ, to the unknown regularity of the eigenvalues $(\lambda_k)_{k\geq 1}$ and to the unknown noise variance σ^2. Interestingly, the minimax rate of testing depends on the decay of the nonincreasing sequence $(\lambda_k a_k^2)_{k\geq 1}$.

7. Simulations.

7.1. *Experiments*.

Setting. The performances of the procedures $T_\alpha^{(1)}$ and $T_\alpha^{(2)}$ are illustrated for various choices of the function θ. In all experiments, the noise ε follows a standard Gaussian distribution with unit variance, while the process X is a Brownian motion defined on $[0, 1]$. The eigenfunctions and eigenvalues of the covariance operator of the Brownian motion have been computed in Ash and Gardner [1],

$$\lambda_j = \frac{1}{(j-0.5)^2\pi^2} \quad \text{and}$$

$$V_j(t) = \sqrt{2}\sin\bigl((j-0.5)\pi t\bigr), \qquad t \in [0,1],\ j = 1, 2, \ldots.$$

In practice $X(t)$ has been simulated using a truncated version of the Karhunen–Loève expansion $\sum_{j=1}^{100}\sqrt{\lambda_j}\eta^{(j)}V_j(t)$, where the $(\eta^{(j)})_{j\in\mathbb{N}}$ form an i.i.d. sequence of standard normal variables. The function $X(t)$ is observed on 1000 evenly spaced points in $[0, 1]$.

Testing procedure. For each experiment, we perform the tests $T_\alpha^{(1)}$ (procedure P_1) and $T_\alpha^{(2)}$ (procedure P_2) with $\bar{k}_n = 2^{\lfloor \log_2 n - 1\rfloor}$. The quantile $q_{\mathbf{X},\alpha}$ involved in P_2 is computed by Monte Carlo simulations. For each experiment, we use 1000 random simulations to estimate this quantile.

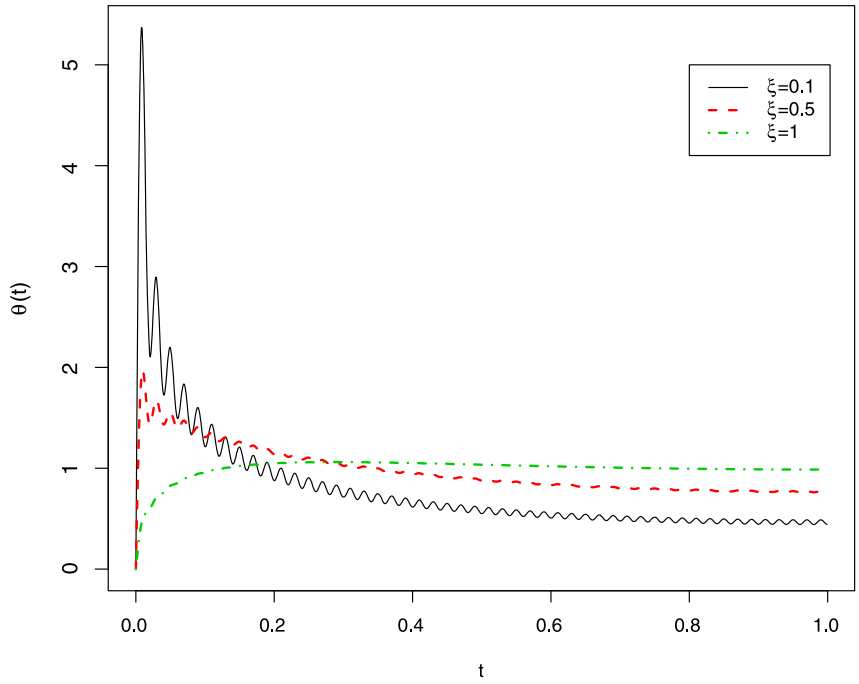

FIG. 1. *Three functions θ in Θ_{KL} when $B = 1$.*

Choice of θ.

(1) In the first experiment, we fix $\theta = 0$ as a way to evaluate the sizes of the testing procedures.

(2) In the second experiment, we build directly the function θ in the KL basis of X. The set Θ_{KL} is made of all the functions $\theta_{B,\xi}$ with $B > 0$, $\xi > 0$ and

$$\theta_{B,\xi}(t) := \frac{B}{\sqrt{\sum_{k=1}^{+\infty} k^{-2\xi-1}}} \sum_{j=1}^{100} j^{-\xi-0.5} V_j(t), \tag{18}$$

where ξ is a smoothness parameter. Observe that B stands for the l_2 norm of the function $\theta_{B,\xi}$. As we show in Figure 1, the smoothness of $\theta_{B,\xi} \in \Theta_{\mathrm{KL}}$ increases with ξ. For this experiment, we have an explicit expression of the joint regularity of θ and Γ,

$$\|(\Gamma^{1/2} - \Gamma_k^{1/2})\theta\|^2 = \frac{B^2}{\pi^2(\sum_{l=1}^{+\infty} l^{-2\xi-1})} \sum_{j=k+1}^{100} (j-0.5)^{-2} j^{-2\xi-1}.$$

In practice, we fix $\xi = 0.1, 0.5, 1$ and $B = 0.1, 0.5, 1$.

(3) In the third experiment, we consider the set Θ_G of functions

$$\theta_{B,\tau}(t) = B \exp\left[-\frac{(t-0.5)^2}{2\tau^2}\right]\left[\int_0^1 \exp\left[-\frac{(x-0.5)^2}{\tau^2}\right] dx\right]^{-1/2},$$

TABLE 1
First simulation study: *Null hypothesis is true*. *Percentages of rejection of* H_0 *and* 95% *confidence intervals*

	$n = 100$	$n = 500$
$T_\alpha^{(1)}$	3.47 ($\pm$ 0.36)	2.61 ($\pm$ 0.31)
$T_\alpha^{(2)}$	4.97 ($\pm$ 0.43)	5.26 ($\pm$ 0.44)

with $B > 0$ and $\tau > 0$. Here, B stands for the l_2 norm of $\theta_{B,\tau}$, and τ is a smoothness parameter. In fact, $\theta_{B,\tau}(t)$ corresponds (up to a constant) to the density of a normal variable with mean 0.5 and variance τ^2. As τ decreases to 0, $\theta_{B,\tau}$ converges to a Dirac function centered on 0.5. In practice, we fix $\tau = 0.01, 0.02, 0.05$ and $B = 0.5, 1, 2$.

Number of experiments. We have set $n = 100$ and $n = 500$. For each set of parameters (n, B, ξ) or (n, B, τ), 10,000 trials were run to estimate the percentages of rejection of H_0 (i.e., the percentages of positive values of $T_\alpha^{(1)}$ and $T_\alpha^{(2)}$ with $\alpha = 5\%$), along with their 95% confidence intervals.

7.2. *Results*. The two procedures P_1 and P_2 have been implemented in R [31] on a 3 GHz Intel Xeon processor, with a 4000 KB cache size and 8 GB total physical memory.

First setting. The percentages of rejection of $T_\alpha^{(1)}$ and $T_\alpha^{(2)}$ under H_0 with $n = 100$ and $n = 500$ are provided in Table 1. As expected, the size of $T_\alpha^{(1)}$ decreases when n increases because we pay a price for the Bonferroni correction. The size of $T_\alpha^{(2)}$ remains close to the nominal level $\alpha = 5\%$.

Second setting. Tables 2 and 3 depict the results for $\theta \in \Theta_{\mathrm{KL}}$ with $n = 100$ and $n = 500$, respectively. As expected, the power of the procedures is increasing with

TABLE 2
Second simulation study: $\theta \in \Theta_{\mathrm{KL}}$, $n = 100$. *Percentages of rejection of* H_0 *and* 95% *confidence intervals*

		$B = 0.1$	$B = 0.5$	$B = 1$
$\xi = 0.1$	$T_\alpha^{(1)}$	3.88 ($\pm$ 0.38)	21.41 ($\pm$ 0.8)	77.24 ($\pm$ 0.82)
	$T_\alpha^{(2)}$	5.8 ($\pm$ 0.46)	26.38 ($\pm$ 0.86)	81.78 ($\pm$ 0.76)
$\xi = 0.5$	$T_\alpha^{(1)}$	4.74 ($\pm$ 0.42)	46.47 ($\pm$ 0.98)	98.68 ($\pm$ 0.22)
	$T_\alpha^{(2)}$	6.65 ($\pm$ 0.49)	52.79 ($\pm$ 0.98)	99.06 ($\pm$ 0.19)
$\xi = 1$	$T_\alpha^{(1)}$	4.8 ($\pm$ 0.42)	62.67 ($\pm$ 0.95)	99.75 ($\pm$ 0.1)
	$T_\alpha^{(2)}$	7.07 ($\pm$ 0.5)	68.3 ($\pm$ 0.91)	99.84 ($\pm$ 0.08)

TABLE 3
Second simulation study: $\theta \in \Theta_{\mathrm{KL}}$, $n = 500$. *Percentages of rejection of* H_0 *and* 95% *confidence intervals*

		$B = 0.1$	$B = 0.5$	$B = 1$
$\xi = 0.1$	$T_\alpha^{(1)}$	5.17 ($\pm$ 0.43)	86.98 ($\pm$ 0.66)	100 ($\pm$ 0)
	$T_\alpha^{(2)}$	8.48 ($\pm$ 0.55)	90.89 ($\pm$ 0.56)	100 ($\pm$ 0)
$\xi = 0.5$	$T_\alpha^{(1)}$	8.81 ($\pm$ 0.56)	99.85 ($\pm$ 0.08)	100 ($\pm$ 0)
	$T_\alpha^{(2)}$	13.07 ($\pm$ 0.66)	99.88 ($\pm$ 0.07)	100 ($\pm$ 0)
$\xi = 1$	$T_\alpha^{(1)}$	11.38 ($\pm$ 0.62)	99.99 ($\pm$ 0.02)	100 ($\pm$ 0)
	$T_\alpha^{(2)}$	16.13 ($\pm$ 0.72)	100 ($\pm$ 0)	100 ($\pm$ 0)

B as $\|\theta\|$ becomes larger. Furthermore, the power also increases with ξ. This corroborates the rates stated in Section 6, since the function $\theta_{B,\xi}$ becomes smoother when ξ increases. In every setting the test $T_\alpha^{(2)}$ with the second procedure performs better than $T_\alpha^{(1)}$.

Third setting. The results of the last experiment are provided in Table 4 for $n = 100$ and Table 5 for $n = 500$. Again, the power is increasing with B, n and τ. Here, τ does not directly correspond to the rate of convergence of the sequence $(\int_0^1 \theta_{B,\tau} V_j(t)\,dt)$, $j \geq 1$, as ξ does in the last example. Nevertheless, it is difficult to detect a function $\theta_{B,\tau}$ when τ decreases, that is, when $\theta_{B,\tau}$ becomes close to a Dirac function.

In each setting, the test under P_2 is more powerful than the test under P_1. Nevertheless, the procedure (P_2) is slightly slower to compute as it requires the evaluations of the quantile $q_{\mathbf{X},\alpha}$ by a Monte Carlo method. Under P_1, the mean computation time is 9 seconds for $n = 100$ and 12 seconds for $n = 500$. In contrast, it, respectively, equals 11 and 18 seconds under P_2.

TABLE 4
Third simulation study: $\theta \in \Theta_G$, $n = 100$. *Percentage of rejection of* H_0 *and* 95% *confidence interval*

		$B = 0.5$	$B = 1$	$B = 2$
$\tau = 0.01$	$T_\alpha^{(1)}$	4.94 ($\pm$ 0.42)	11.85 ($\pm$ 0.63)	46.69 ($\pm$ 0.98)
	$T_\alpha^{(2)}$	7.25 ($\pm$ 0.51)	15.49 ($\pm$ 0.71)	53.56 ($\pm$ 0.98)
$\tau = 0.02$	$T_\alpha^{(1)}$	7.33 ($\pm$ 0.51)	23.09 ($\pm$ 0.83)	80.26 ($\pm$ 0.78)
	$T_\alpha^{(2)}$	10 ($\pm$ 0.59)	28.54 ($\pm$ 0.89)	84.04 ($\pm$ 0.72)
$\tau = 0.05$	$T_\alpha^{(1)}$	13.85 ($\pm$ 0.68)	56.51 ($\pm$ 0.97)	99.48 ($\pm$ 0.14)
	$T_\alpha^{(2)}$	18.13 ($\pm$ 0.76)	63.09 ($\pm$ 0.95)	99.65 ($\pm$ 0.12)

TABLE 5
Third simulation study: $\theta \in \Theta_G$, $n = 500$. *Percentage of rejection of* H_0 *and* 95% *confidence interval*

		$B = 0.5$	$B = 1$	$B = 2$
$\tau = 0.01$	$T_\alpha^{(1)}$	12.41 ($\pm$ 0.65)	54.6 ($\pm$ 0.98)	99.75 ($\pm$ 0.1)
	$T_\alpha^{(2)}$	17.99 ($\pm$ 0.75)	63.16 ($\pm$ 0.95)	99.98 ($\pm$ 0.07)
$\tau = 0.02$	$T_\alpha^{(1)}$	26.11 ($\pm$ 0.86)	88.91 ($\pm$ 0.62)	100 ($\pm$ 0)
	$T_\alpha^{(2)}$	33.95 ($\pm$ 0.93)	92.62 ($\pm$ 0.51)	100 ($\pm$ 0)
$\tau = 0.05$	$T_\alpha^{(1)}$	65.38 ($\pm$ 0.93)	99.95 ($\pm$ 0.04)	100 ($\pm$ 0)
	$T_\alpha^{(2)}$	72.74 ($\pm$ 0.87)	99.99 ($\pm$ 0.02)	100 ($\pm$ 0)

8. Discussion. Two multiple testing procedures of the nullity of the slope function θ have been proposed in this paper. They are completely data-driven and benefit from optimal properties assessed in a nonasymptotic setting. We address here some extensions of our results.

Although we focused on the null-hypothesis "$H_0 : \theta = 0$," our approach easily extends to linear hypotheses $H_{0,\mathcal{V}}$: "$\theta \in \mathcal{V}$", where $\mathcal{V}$ is a given finite dimensional subspace of $\mathcal{H}$ of dimension $p < n/2$. As previously, the procedure relies on parametric statistics for testing $H_{0,\mathcal{V}}$ against $H_{1,k,\mathcal{V}}$: "$\theta \in (\mathrm{Vect}(V_1, \ldots, V_k) + \mathcal{V}) \setminus V$", where k is a positive integer. We consider the $n \times \hat{k}^{\mathrm{KL}}$ design matrix $\mathbf{W}$ defined by $\mathbf{W}_{i,j} = \langle \mathbf{X}_i, \widehat{V}_j \rangle$ for $i = 1, \ldots, n$, $j = 1, \ldots, \hat{k}^{\mathrm{KL}}$. The space generated by the $\hat{k}^{\mathrm{KL}}$ columns of the matrix W is denoted $\underline{\mathcal{W}}_{\hat{k}^{\mathrm{KL}}}$. Considering a basis $(\xi_1, \ldots, \xi_p)$ of $\mathcal{V}$, we define $\underline{\mathcal{V}}_p$ as the space generated by the p columns of the matrix whose (ij)th element is $\langle \mathbf{X}_i, \xi_j \rangle$. In the sequel, $\widehat{\mathbf{\Pi}}_{k,\mathcal{V}}$ stands for the orthogonal projection in $\mathbb{R}^n$ onto $\underline{\mathcal{V}}_p^{\perp} \cap \underline{\mathcal{W}}_{\hat{k}^{\mathrm{KL}}}$ of dimension less or equal to $\hat{k}^{\mathrm{KL}}$, while $\widehat{\mathbf{\Pi}}_{\mathcal{V}}$ stands for the orthogonal projection onto $\underline{\mathcal{V}}_p$. Then, we consider the following parametric statistic:

$$\phi_{k,\mathcal{V}}(\mathbf{Y}, \mathbf{X}) := \frac{\|\widehat{\mathbf{\Pi}}_{k,\mathcal{V}} \mathbf{Y}\|_n^2}{\|\mathbf{Y} - \widehat{\mathbf{\Pi}}_{k,\mathcal{V}} \mathbf{Y} - \widehat{\mathbf{\Pi}}_{\mathcal{V}} \mathbf{Y}\|_n^2 / [n - \dim(\underline{\mathcal{V}}_p + \underline{\mathcal{W}}_{\hat{k}^{\mathrm{KL}}})]}. \tag{19}$$

Under $H_{0,\mathcal{V}}$, $\phi_{k,\mathcal{V}}(\mathbf{Y}, \mathbf{X})/\hat{k}^{\mathrm{KL}}$ behaves like a Fisher distribution with $(\dim(\underline{\mathcal{V}}_p^{\perp} \cap \underline{\mathcal{W}}_{\hat{k}^{\mathrm{KL}}}), n - \dim(\underline{\mathcal{V}}_p + \underline{\mathcal{W}}_{\hat{k}^{\mathrm{KL}}}))$ degrees of freedom. The proof is the same as that for $\phi_k(\mathbf{Y}, \mathbf{X})$. In typical situations, we have $\dim(\underline{\mathcal{V}}_p^{\perp} \cap \underline{\mathcal{W}}_{\hat{k}^{\mathrm{KL}}}) = k$ and $\dim(\underline{\mathcal{V}}_p + \underline{\mathcal{W}}_{\hat{k}^{\mathrm{KL}}}) = k + p$. We reject $H_{0,\mathcal{V}}$ when the statistic

$$T_{\alpha,\mathcal{V}} := \sup_{k \in \mathcal{K}_n, k \le \mathrm{Rank}(\widehat{\Gamma}_n)} \big[\phi_{k,\mathcal{V}}(\mathbf{Y}, \mathbf{X}) - \hat{k}^{\mathrm{KL}} \bar{\mathcal{F}}^{-1}_{\dim(\underline{\mathcal{V}}_p^{\perp} \cap \underline{\mathcal{W}}_{\hat{k}^{\mathrm{KL}}}), n - \dim(\underline{\mathcal{V}}_p + \underline{\mathcal{W}}_{\hat{k}^{\mathrm{KL}}})} \{\alpha_{\mathcal{K}_n}(\mathbf{X})\}\big]$$

is positive, where the weight $\alpha_{\mathcal{K}_n}(\mathbf{X})$ is chosen according to procedure P_1 (Bonferroni) or a slight variation of P_2 (Monte Carlo). All the results stated for $T_\alpha^{(1)}$ and $T_\alpha^{(2)}$ are still valid with $T_{\alpha,\mathcal{V}}$. The extension to affine subspaces $\mathcal{V}$ is also possible.

The power of $T_\alpha^{(1)}$ has been analyzed over the collection of ellipsoids $\mathcal{E}_a(R)$. The considered ellipsoids describing the nonparametric alternatives are determined by the principal directions $(V_j)_{j\geq 1}$, which are generally unknown. In fact, for some functions θ that are well represented by a prescribed basis (as wavelet, spline or Fourier basis) and whose expansion in the eigenfunction basis decreases slowly, projecting the data onto the Karhunen–Loève expansion is not necessarily best suited. Alternatively, one can adopt a similar approach in the context of a prescribed basis (as wavelet, spline or Fourier basis) instead of the eigenfunctions basis discussed above. The size and the power of the corresponding procedures are in fact easier to derive than for a Karhunen–Loève approach as we do not have to control the randomness of the basis. We refer, for instance, to [3] for such results in a fixed design regression problem. As θ is unknown, the best choice of basis (prescribed or estimated by PCA) is also unknown. A solution is to combine testing procedures based on different basis.

9. Main proofs. In this section, we emphasize the core of the proofs. Arguments based on perturbation theory are introduced in the next section. All the technical and side results are gathered in Appendices C–F of [25].

9.1. *Additional notation.* Given any integer $k < \mathrm{Rank}(\Gamma)$, we recall that $\Gamma_k = \sum_{j=1}^k \lambda_i V_j \otimes V_j$, where $\otimes$ stands for the tensor product. Similarly, $\widehat{\Gamma}_{n,k} := \sum_{j=1}^k \widehat{\lambda}_i \widehat{V}_j \otimes \widehat{V}_j$ denotes its empirical counterpart. For any $k < \mathrm{Rank}(\Gamma)$, we note Π_k the orthogonal projection in $\mathcal{H}$ onto the space spanned by V_j, $j = 1, \ldots, k$, while $\widehat{\Pi}_k$ stands for the orthogonal projection onto the space spanned by $\widehat{V}_j$, $j = i, \ldots, k \wedge \mathrm{Rank}(\widehat{\Gamma}_n)$.

In order to translate the definition of the testing procedure into functional data analysis framework, we shall use $\Delta = \mathbb{E}(\langle X, \cdot\rangle Y)$. We note $\Delta_n = \sum_{i=1}^n \langle X_i, \cdot\rangle \mathbf{Y}_i/n$ its empirical counterpart. For any $k \leq \mathrm{Rank}(\Gamma)$, we note $A_k = \sum_{j=1}^k \lambda_j^{-1/2} \langle V_j, \cdot\rangle V_j$ and $\widehat{A}_k = \sum_{j=1}^{k\wedge \mathrm{Rank}(\widehat{\Gamma}_n)} \widehat{\lambda}_j^{-1/2} \langle \widehat{V}_j, \cdot\rangle \widehat{V}_j$ its empirical counterpart.

Let S be a bounded linear operator on the Hilbert space $\mathcal{H}$. The corresponding operator norm will be denoted $\|\cdot\|_\infty$ where $\|S\|_\infty = \sup_{x\in \mathbf{B}(0,1)} \|S(x)\|$ and $\mathbf{B}(0, 1)$ stands for the unit ball of $\mathcal{H}$. Let T be a Hilbert–Schmidt operator. $\|\cdot\|_{\mathrm{HS}}$ denotes the Hilbert–Schmidt norm, and tr stands for the classical trace (defined for trace-class operators). We recall that $\|T\|_{\mathrm{HS}}^2 = \mathrm{tr}(T^*T)$.

In the sequel, we note $\bar{\chi}_k(u)$ the probability that a χ^2 variable with k degrees of freedom is larger than u, while $\bar{\chi}_k^{-1}(u)$ denotes the $1-u$ quantile of a χ^2 random variable.

9.2. *Connection between $\phi_k(\mathbf{Y}, \mathbf{X})$ and the procedure of Cardot et al.* [5]. In fact, the numerator of the statistic ϕ_k is exactly the same as the test statistic $\|\sqrt{n}\widehat{A}_k\Delta_n\|^2$ introduced by Cardot et al. [5], that is,

$$\phi_k(\mathbf{Y}, \mathbf{X}) = \frac{\|\widehat{\mathbf{\Pi}}_k\mathbf{Y}\|_n^2}{\|\mathbf{Y} - \widehat{\mathbf{\Pi}}_k\mathbf{Y}\|_n^2/(n - \hat{k}^{\mathrm{KL}})} = \frac{\|\sqrt{n}\widehat{A}_k\Delta_n\|^2}{\|\mathbf{Y} - \widehat{\mathbf{\Pi}}_k\mathbf{Y}\|_n^2/(n - \hat{k}^{\mathrm{KL}})}. \tag{20}$$

PROOF OF EQUATION (20). Consider the least-squares $\widehat{\theta}_k$ estimator of θ in the space generated by $\widehat{V}_j$, $j = 1, \ldots, \hat{k}^{\mathrm{KL}}$. It follows that $\|\widehat{\mathbf{\Pi}}_k\mathbf{Y}\|_n^2 = n\langle\widehat{\theta}_k, \widehat{\Gamma}_n\widehat{\theta}_k\rangle$. Since $\widehat{\theta}_k = \widehat{\Gamma}_{n,k}^{-}\Delta_n$ where $\widehat{\Gamma}_{n,k}^{-}$ is the Moore–Penrose pseudo-inverse of $\widehat{\Gamma}_{n,k}$, we obtain

$$\|\widehat{\mathbf{\Pi}}_k\mathbf{Y}\|_n^2 = n\langle\widehat{\Gamma}_{n,k}^{-}\Delta_n, \widehat{\Gamma}_n\widehat{\Gamma}_{n,k}^{-}\Delta_n\rangle = n\langle\widehat{A}_k\Delta_n, \widehat{A}_k\widehat{\Gamma}_n\widehat{\Gamma}_{n,k}^{-}\Delta_n\rangle = n\|\widehat{A}_k\Delta_n\|^2. \quad \square$$

9.3. *Proof of the type I error bounds.* We first prove Propositions 3.3 and 5.2. Afterwards, we derive Theorem 5.3, the proof of Theorem 3.4 being postponed to Appendix D in [25].

PROOF OF PROPOSITIONS 3.3 AND 5.2. Let us assume that ε follows a Gaussian distribution and that $\theta = 0$. Conditionally on $\mathbf{X}$, the statistic $\phi_k(\mathbf{Y}, \mathbf{X})/\hat{k}$ defined in (5.1) follows a Fisher distribution with $(\hat{k}, n - \hat{k})$ degrees of freedom. Hence, conditionally on $\mathbf{X}$, the test $T_{\alpha,k}$ has a size exactly α. Conditionally on $\mathbf{X}$, $T_\alpha^{(1)}$ is a Bonferroni procedure of Fisher statistics, and its size is smaller than α. Reintegrating with respect to $\mathbf{X}$, we derive that the size of $T_\alpha^{(1)}$ is smaller than α. Let us turn to the second result. The quantity $q_{\mathbf{X},\alpha}$ satisfies

$$\mathbb{P}_0\Bigg(\sup_{k\in\mathcal{K}_n}\left\{\frac{(n-\hat{k})\|\widehat{\mathbf{\Pi}}_k\boldsymbol{\varepsilon}\|_n^2}{\hat{k}\|\boldsymbol{\varepsilon} - \widehat{\mathbf{\Pi}}_k\boldsymbol{\varepsilon}\|_n^2} - \bar{\mathcal{F}}^{-1}_{\hat{k},n-\hat{k}}(q_{\mathbf{X},\alpha})\right\} > 0\Big|\mathbf{X}\Bigg) = \alpha,$$

which implies that $\mathbb{P}_0(T_\alpha^{(2)}|\mathbf{X}) = \alpha$ $\mathbf{X}$ a.s. $\square$

PROOF OF THEOREM 5.3. First, we state that $\hat{k}^{\mathrm{KL}} = k$ with large probability.

LEMMA 9.1. *Consider the event $\mathcal{A}_n$ defined by*

$$\mathcal{A}_n = \left\{\sup_{1\le j\le\bar{k}_n}\frac{|\widehat{\lambda}_j - \lambda_j|}{\min\{\lambda_j - \lambda_{j+1}, \lambda_{j-1} - \lambda_j\}} \ge 1/2\right\}. \tag{21}$$

Under assumptions (B.2) *and* (B.3), *we have*

$$\mathbb{P}(\mathcal{A}_n) \le C(\gamma)\left[\frac{\bar{k}_n^3\log^2(\bar{k}_n \vee e)}{n}\right] \le C(\gamma)\frac{1}{\log^2(n)}, \tag{22}$$

where γ is a positive constant involved in assumption (B.2).

This result, proved in Appendix E of [25], relies on the perturbation theory of random operators. Observe that under the event $\overline{\mathcal{A}}_n$, we have $\hat{k}^{\mathrm{KL}} = k$ for all $k \le \bar{k}_n$. Consequently, we can replace $\hat{k}^{\mathrm{KL}}$ by k in the definition of the test statistic up to an event of probability less than $C(\gamma)/\log(n)$. In the sequel, we use the alternative expression (20) of ϕ_k, and we replace $\hat{k}^{\mathrm{KL}}$ by k. The proof is split into three main lemmas, Lemmas 9.2–9.4. The first lemma states that $\|\sqrt{n}A_k\Delta_n\|^2/\sigma^2$ behaves like a χ^2 distribution. Its proof (Appendix C in [25]) relies on a multivariate Berry–Esseen theorem. The second lemma, which tells us that $\|\sqrt{n}A_k\Delta_n\|^2/\sigma^2$ is close to $\|\sqrt{n}\widehat{A}_k\Delta_n\|^2/\sigma^2$ is proved below. The third lemma, proved in Appendix F of [25], states that $\|\mathbf{Y} - \widehat{\mathbf{\Pi}}_k\mathbf{Y}\|_n^2/n$ concentrates well around σ^2.

LEMMA 9.2. *Assume that* (B.1) *and* (B′.3) *hold. For any* $k \ge 1$ *and any* $x > 0$, *we have*

$$\begin{aligned}\left|\mathbb{P}\big(\|\sqrt{n}A_k\Delta_n\|^2 \ge x\big) - \bar{\chi}_k\big(x/\sigma^2\big)\right| &\le C\frac{k^{3/2}}{\sqrt{n}}\frac{\mathbb{E}[\varepsilon^4]^{3/4}}{\sigma^3}\sup_{1\le j\le k}\mathbb{E}\big[\big(\eta^{(j)}\big)^4\big]^{3/4}\\ &\le \frac{C}{\log^2(n)},\end{aligned}$$

uniformly over all $k \le \bar{k}_n$.

LEMMA 9.3. *Assume that* (B.1)–(B′.3) *hold. Writing* $x_{n,k} = 1/(k\log^2(n))$, *we have for all* $k \le \bar{k}_n$, *and all* $n \ge 5$,

$$\begin{aligned}&\mathbb{P}\big[\|\sqrt{n}\widehat{A}_k\Delta_n\|^2 \ge (1-x_{n,k})^{-1}\|\sqrt{n}A_k\Delta_n\|^2\big]\\ &\quad\le \mathbb{P}[\mathcal{A}_n] + \frac{C(\gamma)}{\log^2(n)} + \left(\frac{\sqrt{e}}{\log(n)}\right)^k.\end{aligned} \tag{23}$$

LEMMA 9.4. *Uniformly over all* $k \le \bar{k}_n$, *we have*

$$\mathbb{P}\left[\left|\frac{\|\mathbf{Y} - \widehat{\mathbf{\Pi}}_k\mathbf{Y}\|_n^2}{n\sigma^2} - 1\right| \ge \frac{k\log^2(n)}{n} + 8\sqrt{\frac{\log\log n}{n}}\right] \le \frac{C}{\log^2(n)} + \frac{C'}{\sqrt{n}}.$$

Let us upper bound the rejection probability due to the statistic ϕ_k

$$\mathbb{P}\big[\phi_k(\mathbf{Y},\mathbf{X}) \ge k\bar{\mathcal{F}}_{k,n-k}^{-1}\big(\alpha/|\mathcal{K}_n|\big)\big] \le \mathbb{P}\left[\frac{\|\sqrt{n}\widehat{A}_k\Delta_n\|^2}{\|\mathbf{Y} - \widehat{\mathbf{\Pi}}_k\mathbf{Y}\|_n^2/n} \ge k\bar{\mathcal{F}}_{k,n-k}^{-1}\big(\alpha/|\mathcal{K}_n|\big)\right]$$

by the three following probabilities:

$$\begin{aligned}&\mathbb{P}\left[\frac{\|\sqrt{n}A_k\Delta_n\|^2}{(1-x_{n,k})\sigma^2} \ge k\left(1 - 8\sqrt{\frac{\log\log(n)}{n}} - \frac{k\log^2(n)}{n}\right)\bar{\mathcal{F}}_{k,n-k}^{-1}\big(\alpha/|\mathcal{K}_n|\big)\right]\\ &\quad+ \mathbb{P}\big[\|\sqrt{n}\widehat{A}_k\Delta_n\|^2 \ge (1-x_{n,k})^{-1}\|\sqrt{n}A_k\Delta_n\|^2\big]\\ &\quad+ \mathbb{P}\left[\left|\frac{\|\mathbf{Y} - \widehat{\mathbf{\Pi}}_k\mathbf{Y}\|_n^2}{n\sigma^2} - 1\right| \ge \frac{k\log^2(n)}{n} + 8\sqrt{\frac{\log\log(n)}{n}}\right].\end{aligned}$$

Gathering the above results, we obtain that this probability is upper bounded by

$$\frac{C(\gamma)}{\log^2(n)} + \biggl(\frac{\sqrt{e}}{\log(n)}\biggr)^k + \bar{\chi}_k\biggl[k\biggl(1 - 8\sqrt{\frac{\log\log n}{n}} - \frac{k\log^2(n)}{n} - x_{n,k}\biggr)\bar{\mathcal{F}}_{k,n-k}^{-1}\bigl(\alpha/|\mathcal{K}_n|\bigr)\biggr], \tag{24}$$

uniformly over all $k \le \bar{k}_n$.

LEMMA 9.5. *Writing* $t = 8\sqrt{\frac{\log\log n}{n}} + \frac{k\log^2(n)}{n} + \frac{1}{k\log^2(n)}$, *we have for* n *larger than some numerical constant*

$$\bar{\chi}_k\bigl[k(1-t)\bar{\mathcal{F}}_{k,n-k}^{-1}\bigl(\alpha/|\mathcal{K}_n|\bigr)\bigr] \le \frac{\alpha}{|\mathcal{K}_n|}\biggl(1 + \frac{C(\alpha)}{\log(n)}\biggr).$$

The proof of this technical lemma is postponed to Appendix F of [25]. We conclude by combining (24) with Lemma 9.5 and taking an union bound over all $k \in \mathcal{K}_n$ [recall that $|\mathcal{K}_n| \le \log(n)$]. $\square$

PROOF OF LEMMA 9.3. From $\|b\|^2 - \|a\|^2 = 2\langle a, b-a\rangle + \|b-a\|^2$, we get

$$\frac{\|b\|^2 - \|a\|^2}{\|a\|^2} \le \frac{\|b-a\|}{\|a\|}\biggl(2 + \frac{\|b-a\|}{\|a\|}\biggr).$$

Since $x_{n,k} < 1$ for $n \ge 3$, it follows that

$$\begin{aligned}\mathbb{P}\biggl[\|\sqrt{n}\widehat{A}_k\Delta_n\|^2 \ge \frac{\|\sqrt{n}A_k\Delta_n\|^2}{1-x_{n,k}}\biggr] &\le \mathbb{P}\biggl[\frac{\|\sqrt{n}(\widehat{A}_k - A_k)\Delta_n\|}{\|\sqrt{n}A_k\Delta_n\|} \ge \frac{x_{n,k}}{4}\biggr] \\ &\le \mathbb{P}\biggl[\|\sqrt{n}(\widehat{A}_k - A_k)\Delta_n\| \ge \frac{\sigma\sqrt{k}x_{n,k}}{4\log(n)}\biggr] \\ &\quad + \mathbb{P}\biggl[\|\sqrt{n}A_k\Delta_n\| \le \frac{\sqrt{k}\sigma}{\log(n)}\biggr].\end{aligned}$$

By Lemma 11.1 in [34], we know that for any $0 < x < 1$ and any integer $d \ge 1$, $\mathbb{P}[\chi^2(d) \le de^{-1}x^{2/d}] \le x$. We get from Lemma 9.2 and the last deviation inequality that

$$\mathbb{P}\biggl[\|\sqrt{n}A_k\Delta_n\| \le \frac{\sigma\sqrt{k}}{\log(n)}\biggr] \le \frac{C}{\log^2(n)} + \biggl(\frac{\sqrt{e}}{\log(n)}\biggr)^k,$$

uniformly over all $k \le \bar{k}_n$. Let us turn to the other term. By Markov's inequality and by definition of $x_{n,k}$, the first probability $\mathbb{P}[\|\sqrt{n}(\widehat{A}_k - A_k)\Delta_n\| \ge \frac{\sigma\sqrt{k}x_{n,k}}{4\log(n)}]$ is smaller than

$$\frac{16k\log^6(n)}{\sigma^2}\mathbb{E}\bigl[\|\sqrt{n}(\widehat{A}_k - A_k)\Delta_n\|^2 1_{\overline{\mathcal{A}}_n}\bigr] + \mathbb{P}[\mathcal{A}_n].$$

In order to conclude, we only need to bound $\mathbb{E}[\|\sqrt{n}(\widehat{A}_k - A_k)\Delta_n\|^2 1_{\overline{\mathcal{A}}_n}]$. Tedious computations whose details are gathered in Appendix F of [25] lead us to the following bound:

LEMMA 9.6. *The expectation* $\mathbb{E}[\|\sqrt{n}(\widehat{A}_k - A_k)\Delta_n\|^2 1_{\overline{\mathcal{A}}_n}]$ *is smaller than*

$$2\sigma^2\mathbb{E}\big[\mathrm{tr}\big\{\Gamma_k^{-1/2}\big(\Gamma_k^{1/2} - \widehat{\Gamma}_{n,k}^{1/2}\big)\big\}1_{\overline{\mathcal{A}}_n}\big] + \sigma^2 C(\gamma)\frac{\bar{k}_n^{5/2}\log(\bar{k}_n \vee e)}{n}. \tag{25}$$

The main argument in the proof of the theorem is to bound the expectation in the right-hand side of (25). Its proof relies on perturbation theory and is described in the next section.

LEMMA 9.7. *Under assumptions* (B.1) *and* (B.2), *we have for all* $n \geq 1$,

$$\mathbb{E}\big[\mathrm{tr}\big\{\Gamma_k^{-1/2}\big(\Gamma_k^{1/2} - \widehat{\Gamma}_{n,k}^{1/2}\big)\big\}1_{\overline{\mathcal{A}}_n}\big] \leq C(\gamma)\frac{k^3[\log^2(k) \vee 1]}{n} + C'\frac{k}{\sqrt{n}} \tag{26}$$

uniformly over all $k \leq \bar{k}_n$.

Gathering Lemma 9.7 with (25) allows us to conclude

$$\mathbb{P}\left[\|\sqrt{n}(\widehat{A}_k - A_k)\Delta_n\| \geq \frac{\sigma\sqrt{k}x_{n,k}}{4\log(n)}\right] \leq C(\gamma)\left[\frac{\bar{k}_n^4\log^8(n)}{n} + \frac{\bar{k}_n^2\log^6(n)}{\sqrt{n}}\right] + \mathbb{P}[\mathcal{A}_n],$$

where the left-hand side is smaller than $C(\gamma)/\log^2(n)$ by assumption (B′.3). □

9.4. *Proofs of the type II error bounds.*

PROOF OF PROPOSITION 5.4. This proof follows the same steps as the proof of Proposition 3.2 in [35]. □

We derive Theorem 6.1, the proof of Theorem 4.1 being postponed to Appendix D in [25].

PROOF OF THEOREM 6.1. Arguing as in the beginning of the proof of Theorem 5.3, we can replace $\hat{k}^{\mathrm{KL}}$ by k in the definition of the statistic (14). Consider some $k \in \mathcal{K}_n$ and $n > 8$, the numerator of $\phi_k(\mathbf{Y}, \mathbf{X})$ (20) is lower bounded as follows:

$$\begin{aligned}\|\sqrt{n}\widehat{A}_k\Delta_n\|^2 \geq \|\sqrt{n}A_k\Delta_n\|^2\big[1 - \big(\sqrt{k}\log(n)\big)^{-1}\big] \\ - \sqrt{k}\log(n)\|\sqrt{n}(A_k - \widehat{A}_k)\Delta_n\|^2,\end{aligned}$$

since $2ab \leq a^2 + b^2$. Observe that $\Delta_n = \widehat{\Gamma}_n\theta + \Delta_{n,1}$, where $\Delta_{n,1} = \sum_{i=1}^n \langle X_i, \cdot\rangle\varepsilon_i/n$. The proof is based on the two main following lemmas.

LEMMA 9.8. *For any $\beta \in (0,1)$, we have*

$$\|\sqrt{n}A_k\Delta_n\|^2 \geq k\sigma^2 + \frac{n}{5}\|\Gamma_k^{1/2}\theta\|^2 - 2\sigma^2\sqrt{k\log\Big(\frac{2}{\beta}\Big)} - 10\sigma^2\log\Big(\frac{2}{\beta}\Big),$$

with probability larger than $1-\beta/2-C/\log(n)$ uniformly over all $k \leq \bar{k}_n$.

LEMMA 9.9. *Assume that* (B.1)–(B′.3) *hold. For any $n \geq 1$ we have*

$$\mathbb{P}\bigg[\|\sqrt{n}(\widehat{A}_k - A_k)\widehat{\Gamma}_n\theta\| \geq \frac{\sqrt{n}\|\Gamma^{1/2}\theta\|}{k^{1/4}\log(n)}\bigg] \leq \frac{C(\gamma)}{\log(n)},$$

$$\mathbb{P}\bigg[\|\sqrt{n}(\widehat{A}_k - A_k)\Delta_{n,1}\| \geq \frac{\sigma}{\log(n)}\bigg] \leq \frac{C(\gamma)}{\log(n)},$$

uniformly over all $k \leq \bar{k}_n$.

Lemma 9.8 is based on a multivariate Berry–Esseen inequality and is proved in Appendix C in [25]. The second lemma proceeds from the same kind of arguments as Lemma 9.3. Thus, its proof is postponed to Appendix F in [25]. We get by gathering Lemmas 9.8 and 9.9 and since $\sqrt{k}\log(n) \geq 2$ for $n \geq 8$,

$$\|\sqrt{n}\widehat{A}_k\Delta_n\|^2 \geq k\sigma^2 - 3\sigma^2\frac{\sqrt{k}}{\log(n)} - 2n\frac{\|\Gamma^{1/2}\theta\|^2}{\log(n)} + C_1 n\|\Gamma_k^{1/2}\theta\|^2 - C_2\sigma^2\bigg(\sqrt{k\log\Big(\frac{2}{\beta}\Big)} + \log\Big(\frac{2}{\beta}\Big)\bigg),$$

with probability larger than $1-\beta/2-C(\gamma)/\log(n)$. Next, we use a rough control of the denominator, proved in Appendix F.

LEMMA 9.10 (Control of the denominator). *We have*

$$\frac{\|\mathbf{Y} - \widehat{\mathbf{\Pi}}_k\mathbf{Y}\|_n^2}{n-k} \leq \sigma^2\bigg[1 + C\Big(\frac{k}{n} + \sqrt{\frac{\log(n)}{n}}\Big)\bigg] + C'\|\Gamma^{1/2}\theta\|^2/\beta,$$

with probability larger than $1-2/\log(n)-\beta/4$.

Since $\sqrt{\log(2/\beta)} \geq 1/\log(n)$ and $C_1 \geq 2/\log(n)$ for n large enough, we derive from the previous results that with probability larger than $1-3\beta/4-C(\gamma)/\log(n)$, the statistic $\phi_k(\mathbf{Y},\mathbf{X})$ is lower bounded by

$$\bigg(k\sigma^2 + n\|\Gamma_k^{1/2}\theta\|^2 nC_1' - C_2'\sigma^2\big(\sqrt{k\log(1/\beta)} + \log(1/\beta)\big) - 2\frac{n}{\log(n)}\|(\Gamma^{1/2} - \Gamma_k^{1/2})\theta\|^2\bigg) \Big/ \bigg(\sigma^2\bigg[1 + C\Big(\frac{k}{n} + \sqrt{\frac{\log(n)}{n}}\Big)\bigg] + C'\|\Gamma^{1/2}\theta\|^2/\beta\bigg). \tag{27}$$

By Lemma 1 in [3], we can upper bound the quantile of Fisher distribution

$$k\bar{\mathcal{F}}_{k,n-k}^{-1}(\alpha/|\mathcal{K}_n|) \le k + C\left[\sqrt{k\log\left(\frac{|\mathcal{K}_n|}{\alpha}\right)} + \log\left(\frac{|\mathcal{K}_n|}{\alpha}\right)\right], \tag{28}$$

since we assume that $\log(|\mathcal{K}_n|/\alpha) \le \log(n) + \log(1/\alpha) \le 2\sqrt{n}$. Comparing the lower bound (27) with (28) allows us to conclude. We refer to Appendix F in [25] for the details. □

10. Arguments based on perturbation theory.

10.1. *Preliminary facts.* Roughly speaking, several results mentioned below are based on an extension of the classical residue formula on the complex plane (see Rudin [32]) to analytic functions still defined on the complex plane but with values in the space of operators. We refer to Dunford and Schwartz [17], Chapter VII.3 or to Gohberg et al. [18, 19] for an introduction to functional calculus for operators related with Riesz integrals. Let us denote $\mathcal{B}_j$ the oriented circle of the complex plane with center λ_j and radius $\delta_j/2$ where δ_j is defined by

$$\delta_j = \min\{\lambda_j - \lambda_{j+1}, \lambda_{j-1} - \lambda_j\}. \tag{29}$$

The open domain whose boundary is $\mathcal{C}_k := \bigcup_{j=1}^k \mathcal{B}_j$ is not connected, but we can apply the functional calculus for bounded operators; see Dunford and Schwartz [17], Section VII.3, Definitions 8 and 9. Using this formalism it is easy to prove the following formulas:

$$\Pi_k = \frac{1}{2\pi\iota}\int_{\mathcal{C}_k} (zI - \Gamma)^{-1}\,dz \quad \text{and} \quad \Gamma_k^{1/2} = \frac{1}{2\pi\iota}\int_{\mathcal{C}_k} z^{1/2}(zI - \Gamma)^{-1}\,dz.$$

The same is true with the random operator $\widehat{\Gamma}_n$, but the contour $\mathcal{C}_k$ must be replaced by its random counterpart $\widehat{\mathcal{C}}_k = \bigcup_{j=1}^{k\wedge \mathrm{Rank}(\widehat{\Gamma}_n)} \widehat{\mathcal{B}}_j$ where each $\widehat{\mathcal{B}}_j$ is a random ball of the complex plane with center $\widehat{\lambda}_j$ and a radius $\widehat{\delta}_j/2 = \min\{\widehat{\lambda}_j - \widehat{\lambda}_{j+1}, \widehat{\lambda}_{j-1} - \widehat{\lambda}_j\}$. We start with some lemmas.

LEMMA 10.1. *Assume that for some $\gamma > 0$, the sequence $(j\lambda_j \log^{1+\gamma}(j \vee 2))_{j\in\mathbb{N}^*}$ decreases. Then we have*

$$\sum_{j\ge 1, j\ne k} \frac{\lambda_j}{|\lambda_k - \lambda_j|} \le C(\gamma)k[\log k \vee 1].$$

For any positive integer j, let us define the event

$$\mathcal{E}_{j,n} := \Big\{\sup_{z\in\mathcal{B}_j} \big\|(zI - \Gamma)^{-1/2}(\widehat{\Gamma}_n - \Gamma)(zI - \Gamma)^{-1/2}\big\|_\infty \ge 1/2\Big\}.$$

LEMMA 10.2. *Suppose that assumption* (B.1)–(B.2) *holds. For any* $j \geq 1$, *we have the two following bounds*:

$$\mathbb{E} \sup_{z \in \mathcal{B}_j} \big\| (zI - \Gamma)^{-1/2} (\widehat{\Gamma}_n - \Gamma)(zI - \Gamma)^{-1/2} \big\|_{\mathrm{HS}}^2 \leq \frac{C(\gamma)}{n} \big[j(\log j \vee 1) \big]^2,$$

$$\mathbb{P}(\mathcal{E}_{j,n}) \leq \frac{C(\gamma)}{n} \big[j(\log j \vee 1) \big]^2.$$

The proof of Lemma 10.1 (resp., Lemma 10.2) is postponed to Appendix F (resp., Appendix E) in [25].

10.2. *Proof of Lemma* 9.7. In order to upper bound this expectation, we set $\widehat{\lambda}_j = 0$ for any $j > \mathrm{Rank}(\widehat{\Gamma}_n)$. We have

$$\begin{aligned}
\mathrm{tr}\big[\Gamma_k^{-1/2}(\Gamma_k^{1/2} - \widehat{\Gamma}_{n,k}^{1/2}) 1_{\overline{\mathcal{A}}_n}\big] &= k 1_{\overline{\mathcal{A}}_n} - \sum_{j=1}^{k} \sum_{l=1}^{k} \sqrt{\frac{\widehat{\lambda}_j}{\lambda_l}} \langle V_l, \widehat{V}_j \rangle^2 1_{\overline{\mathcal{A}}_n} \\
&\leq k 1_{\overline{\mathcal{A}}_n} - \sum_{j=1}^{k} \sqrt{\frac{\widehat{\lambda}_j}{\lambda_j}} \langle V_j, \widehat{V}_j \rangle^2 1_{\overline{\mathcal{A}}_n} \\
&\leq \sum_{j=1}^{k} (1 - \langle V_j, \widehat{V}_j \rangle^2) 1_{\overline{\mathcal{A}}_n} + \sum_{j=1}^{k} \left| \sqrt{\frac{\widehat{\lambda}_j}{\lambda_j}} - 1 \right| 1_{\overline{\mathcal{A}}_n} \\
&\leq \sum_{j=1}^{k} (1 - \langle V_j, \widehat{V}_j \rangle^2) 1_{\overline{\mathcal{A}}_n} + \sum_{j=1}^{k} \frac{|\widehat{\lambda}_j - \lambda_j|}{\lambda_j} 1_{\overline{\mathcal{A}}_n},
\end{aligned}$$

where the last equation follows from the upper bound $|\sqrt{1+x} - 1| \leq |x|$ for any $x \geq -1$. Observe that under the event $\overline{\mathcal{A}}_n$, $(\widehat{\lambda}_j - \lambda_j)/\lambda_j \leq 1/2$. Applying Lemma 10.2, we obtain the following bound:

$$\begin{aligned}
&\mathbb{E}\big[\mathrm{tr}\big[\Gamma_k^{-1/2}(\Gamma_k^{1/2} - \widehat{\Gamma}_{n,k}^{1/2}) 1_{\overline{\mathcal{A}}_n}\big]\big] \\
&\quad \leq \sum_{j=1}^{k} \mathbb{E}\big[(1 - \langle V_j, \widehat{V}_j \rangle^2) 1_{\overline{\mathcal{A}}_n \cap \overline{\mathcal{E}}_{j,n}}\big] + \sum_{j=1}^{k} \mathbb{E}\left[\frac{|\widehat{\lambda}_j - \lambda_j|}{\lambda_j} 1_{\overline{\mathcal{A}}_n \cap \overline{\mathcal{E}}_{j,n}} \right] \\
&\qquad + \sum_{j=1}^{k} \frac{3}{2} \mathbb{P}[\mathcal{E}_{j,n}] \qquad (30) \\
&\quad \leq \sum_{j=1}^{k} \mathbb{E}\big[(1 - \langle V_j, \widehat{V}_j \rangle^2) 1_{\overline{\mathcal{A}}_n \cap \overline{\mathcal{E}}_{j,n}}\big] + \sum_{j=1}^{k} \mathbb{E}\left[\frac{|\widehat{\lambda}_j - \lambda_j|}{\lambda_j} 1_{\overline{\mathcal{A}}_n \cap \overline{\mathcal{E}}_{j,n}} \right] \\
&\qquad + C(\gamma) \frac{k^3 (\log^2(k) \vee 1)}{n}.
\end{aligned}$$

In the sequel, π_j stands for the orthogonal projector associated to the single jth eigenvector V_j while $\widehat{\pi}_j$ refers to its empirical counterpart. Applying functional calculus tools for linear operators, we get for any $1 \le j \le k$

$$1 - \langle V_j, \widehat{V}_j\rangle^2 = \langle V_j, V_j\rangle^2 - \langle V_j, \widehat{V}_j\rangle^2 = \langle (\pi_j - \widehat{\pi}_j) V_j, V_j\rangle$$
$$= \frac{1}{2\pi\iota}\Big[\int_{\mathcal{B}_j} \langle (zI - \Gamma)^{-1} V_j, V_j\rangle dz - \int_{\hat{\mathcal{B}}_j} \langle (zI - \widehat{\Gamma}_n)^{-1} V_j, V_j\rangle dz\Big],$$

which looks like the definition of Π_k given in the first paragraph of Section 10.1 (note that only the contour changed). Under the event $\overline{\mathcal{A}}_n$, $\widehat{\lambda}_j$ lies inside the circle $\mathcal{B}_j$. In fact, $(zI - \widehat{\Gamma}_n)^{-1}$ has only one pole inside the circle $\mathcal{B}_j$ at $z = \widehat{\lambda}_j$. As a consequence, we have almost surely

$$\int_{\hat{\mathcal{B}}_j} \langle (zI - \widehat{\Gamma}_n)^{-1} V_j, V_j\rangle dz 1_{\overline{\mathcal{A}}_n} = \int_{\mathcal{B}_j} \langle (zI - \widehat{\Gamma}_n)^{-1} V_j, V_j\rangle dz 1_{\overline{\mathcal{A}}_n},$$

so that

$$\begin{aligned}(1 - \langle V_j, \widehat{V}_j\rangle^2) 1_{\overline{\mathcal{E}}_{j,n} \cap \overline{\mathcal{A}}_n} \\ = \frac{1}{2\pi\iota} \int_{\mathcal{B}_j} \langle \{(zI - \Gamma)^{-1} - (zI - \widehat{\Gamma}_n)^{-1}\} V_j, V_j\rangle dz 1_{\overline{\mathcal{E}}_{j,n} \cap \overline{\mathcal{A}}_n}.\end{aligned} \tag{31}$$

Working out this integral, we get

$$\int_{\mathcal{B}_j} \langle \{(zI - \Gamma)^{-1} - (zI - \widehat{\Gamma}_n)^{-1}\} V_j, V_j\rangle dz$$
$$= -\int_{\mathcal{B}_j} \langle (zI - \Gamma)^{-1} (\widehat{\Gamma}_n - \Gamma)(zI - \Gamma)^{-1} V_j, V_j\rangle$$
$$- \int_{\mathcal{B}_j} \langle (zI - \widehat{\Gamma}_n)^{-1} (\widehat{\Gamma}_n - \Gamma)(zI - \Gamma)^{-1} (\widehat{\Gamma}_n - \Gamma)(zI - \Gamma)^{-1} V_j, V_j\rangle dz.$$

The first term is $\int_{\mathcal{B}_j} (z - \lambda_j)^{-2} \langle (\widehat{\Gamma}_n - \Gamma) V_j, V_j\rangle\, dz$. Thus, it is null almost surely by the Cauchy integration theorem. Define $S_n(z) = (zI - \Gamma)^{1/2} (zI - \widehat{\Gamma}_n)^{-1} (zI - \Gamma)^{1/2}$ and $T_n(z) = (zI - \Gamma)^{-1/2} (\widehat{\Gamma}_n - \Gamma)(zI - \Gamma)^{-1/2}$. For any fixed z, we have $S_n(z) = [I - T_n(z)]^{-1}$. Thus, it comes from (31) that

$$\begin{aligned}&\mathbb{E}\big[(1 - \langle V_j, \widehat{V}_j\rangle^2) 1_{\overline{\mathcal{E}}_{j,n} \cap \overline{\mathcal{A}}_n}\big] \\ &\le \mathbb{E}\Big[\Big|\frac{1}{2\pi\iota} \int_{\mathcal{B}_i} \langle (zI - \Gamma)^{-1/2} S_n(z) T_n^2(z) (zI - \Gamma)^{-1/2} V_j, V_j\rangle 1_{\overline{\mathcal{E}}_{j,n} \cap \mathcal{A}_n}\, dz\Big|\Big] \\ &\le C\delta_j \mathbb{E}\Big[\sup_{z \in \mathcal{B}_j} \{\|S_n(z)\|_\infty 1_{\overline{\mathcal{E}}_{j,n}} \|T_n(z)\|_\infty^2 \|(zI - \Gamma)^{-1}\|_\infty\}\Big] \\ &\le C\delta_j \sup_{z \in \mathcal{B}_j} \|(zI - \Gamma)^{-1}\|_\infty \mathbb{E}\Big[\sup_{z \in \mathcal{B}_j} \|T_n(z)\|_\infty^2\Big] \le C(\gamma) \frac{j^2 (\log^2(j) \vee 1)}{n},\end{aligned} \tag{32}$$

since $\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1}\|_\infty \le 2\delta_i^{-1}$, $\sup_{z\in\mathcal{B}_j}\|S_n(z)\|_\infty 1_{\overline{\mathcal{E}}_{j,n}} \le 2$ and $\mathbb{E}[\sup_{z\in\mathcal{B}_i}\|T_n(z)\|_\infty^2] \le \frac{C(\gamma)}{n} j^2(\log^2(j)\vee 1)$ by Lemma 10.2. Hence, we obtain an upper bound for the first term in (30),

$$\mathbb{E}\left[\left(k-\sum_{j=1}^{k}\langle V_j,\widehat{V}_j\rangle^2\right)1_{\overline{\mathcal{A}}_n}\right] \le C(\gamma)\frac{k^3(\log^2(k)\vee 1)}{n}. \tag{33}$$

Turning to the second term in (30), we only provide a sketch of the proof since the approach is the same as for the first term. We have

$$\widehat{\lambda}_j-\lambda_j = \operatorname{tr}(\widehat{\Gamma}_n\widehat{\pi}_j-\Gamma\pi_j) = \operatorname{tr}\big(\widehat{\Gamma}_n(\widehat{\pi}_j-\pi_j)\big)+\operatorname{tr}\big((\widehat{\Gamma}_n-\Gamma)\pi_j\big),$$

so that

$$\begin{aligned}&\frac{|\widehat{\lambda}_j-\lambda_j|}{\lambda_j}1_{\overline{\mathcal{A}}_n\cap\overline{\mathcal{E}}_{j,n}}\\ &\quad\le \frac{|\operatorname{tr}(\widehat{\Gamma}_n(\widehat{\pi}_j-\pi_j))|}{\lambda_j}1_{\overline{\mathcal{A}}_n\cap\overline{\mathcal{E}}_{j,n}}+\frac{|\operatorname{tr}((\widehat{\Gamma}_n-\Gamma)\pi_j)|}{\lambda_j}1_{\overline{\mathcal{A}}_n\cap\overline{\mathcal{E}}_{j,n}}.\end{aligned} \tag{34}$$

The second term in this decomposition is bounded as follows:

$$\mathbb{E}\left[\frac{|\operatorname{tr}((\widehat{\Gamma}_n-\Gamma)\pi_j)|}{\lambda_j}1_{\overline{\mathcal{A}}_n\cap\overline{\mathcal{E}}_{j,n}}\right] \le \mathbb{E}\left[\frac{|\langle(\widehat{\Gamma}_n-\Gamma)V_j,V_j\rangle|}{\lambda_j}\right] \le \frac{1}{\sqrt{n}}. \tag{35}$$

We now consider $\mathbb{E}[|\operatorname{tr}(\widehat{\Gamma}_n(\widehat{\pi}_j-\pi_j))|1_{\overline{\mathcal{A}}_n\cap\overline{\mathcal{E}}_{j,n}}]$, and we use the same method as above for bounding $(1-\langle V_j,\widehat{V}_j\rangle^2)$. This leads us to the upper bound (the details are postponed to Appendix E in [25]).

LEMMA 10.3. *We have*

$$\frac{1}{\lambda_j}\mathbb{E}\big[|\operatorname{tr}\big(\widehat{\Gamma}_n(\widehat{\pi}_j-\pi_j)\big)|1_{\overline{\mathcal{A}}_n\cap\overline{\mathcal{E}}_{j,n}}\big] \le C(\gamma)\frac{j^2(\log^2 j\vee 1)}{n}.$$

Gathering (34) and (35) with this last bound, we get

$$\sum_{j=1}^{k}\mathbb{E}\left[\left|\frac{\widehat{\lambda}_j-\lambda_j}{\lambda_j}\right|1_{\overline{\mathcal{A}}_n\cap\overline{\mathcal{E}}_{j,n}}\right] \le C(\gamma)\frac{k^3(\log^2(k)\vee 1)}{n}+C'\frac{k}{\sqrt{n}}.$$

Combining this last bound with (30) and (33) allows us to conclude.

Acknowledgments. We would like to thank two anonymous referees for their insightful remarks that lead us to significantly improve the presentation of the paper.

SUPPLEMENTARY MATERIAL

Technical Appendix (DOI: 10.1214/13-AOS1093SUPP; .pdf). We provide additional control of the power and we describe the remaining proofs.

REFERENCES

[1] ASH, R. B. and GARDNER, M. F. (1975). *Topics in Stochastic Processes*. *Probability and Mathematical Statistics* **27**. Academic Press, New York. MR0448463

[2] BARAUD, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606. MR1935648

[3] BARAUD, Y., HUET, S. and LAURENT, B. (2003). Adaptive tests of linear hypotheses by model selection. *Ann. Statist.* **31** 225–251. MR1962505

[4] CAI, T. T. and HALL, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34** 2159–2179. MR2291496

[5] CARDOT, H., FERRATY, F., MAS, A. and SARDA, P. (2003). Testing hypotheses in the functional linear model. *Scand. J. Stat.* **30** 241–255. MR1965105

[6] CARDOT, H., FERRATY, F. and SARDA, P. (2003). Spline estimators for the functional linear model. *Statist. Sinica* **13** 571–591. MR1997162

[7] CARDOT, H., GOIA, A. and SARDA, P. (2004). Testing for no effect in functional linear regression models, some computational approaches. *Comm. Statist. Simulation Comput.* **33** 179–199. MR2044864

[8] CARDOT, H. and JOHANNES, J. (2010). Thresholding projection estimators in functional linear models. *J. Multivariate Anal.* **101** 395–408. MR2564349

[9] CARDOT, H., MAS, A. and SARDA, P. (2007). CLT in functional linear regression models. *Probab. Theory Related Fields* **138** 325–361. MR2299711

[10] CARDOT, H. and SARDA, P. (2011). Functional linear regression. In *The Oxford Handbook of Functional Data Analysis* 21–46. Oxford Univ. Press, Oxford. MR2908018

[11] COMTE, F. and JOHANNES, J. (2010). Adaptive estimation in circular functional linear models. *Math. Methods Statist.* **19** 42–63. MR2682854

[12] COMTE, F. and JOHANNES, J. (2012). Adaptive functional linear regression. *Ann. Statist.* **40** 2765–2797.

[13] CRAMBES, C., KNEIP, A. and SARDA, P. (2009). Smoothing splines estimators for functional linear regression. *Ann. Statist.* **37** 35–72. MR2488344

[14] CUEVAS, A. and FRAIMAN, R. (2004). On the bootstrap methodology for functional data. In *COMPSTAT* 2004—*Proceedings in Computational Statistics* 127–135. Physica, Heidelberg. MR2173014

[15] DAUXOIS, J., POUSSE, A. and ROMAIN, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivariate Anal.* **12** 136–154. MR0650934

[16] DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195

[17] DUNFORD, N. and SCHWARTZ, J. T. (1988). *Linear Operators*, *Part I*: *General Theory*. Wiley, New York. MR1009162

[18] GOHBERG, I., GOLDBERG, S. and KAASHOEK, M. A. (1990). *Classes of Linear Operators*. *Vol. I. Operator Theory*: *Advances and Applications* **49**. Birkhäuser, Basel. MR1130394

[19] GOHBERG, I., GOLDBERG, S. and KAASHOEK, M. A. (1993). *Classes of Linear Operators*. *Vol. II. Operator Theory*: *Advances and Applications* **63**. Birkhäuser, Basel. MR1246332

[20] GONZÁLEZ-MANTEIGA, W. and MARTÍNEZ-CALVO, A. (2011). Bootstrap in functional linear regression. *J. Statist. Plann. Inference* **141** 453–461. MR2719509

[21] HALL, P. and HOROWITZ, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35** 70–91. MR2332269

[22] HALL, P. and HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 109–126. MR2212577

[23] HALL, P. and HOSSEINI-NASAB, M. (2009). Theory for high-order bounds in functional principal components analysis. *Math. Proc. Cambridge Philos. Soc.* **146** 225–256. MR2461880

[24] HALL, P. and VIAL, C. (2006). Assessing extrema of empirical principal component functions. *Ann. Statist.* **34** 1518–1544. MR2278366

[25] HILGERT, N., MAS, A. and VERZELEN, N. (2013). Supplement to "Minimax adaptive tests for the functional linear model." DOI:10.1214/13-AOS1093SUPP.

[26] INGSTER, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Math. Methods Statist.* **2** 85–114. MR1257978

[27] INGSTER, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. II. *Math. Methods Statist.* **2** 171–189. MR1257983

[28] INGSTER, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. III. *Math. Methods Statist.* **2** 249–268. MR1259685

[29] MAS, A. and MENNETEAU, L. (2003). Perturbation approach applied to the asymptotic study of random operators. In *High Dimensional Probability, III* (*Sandjberg*, 2002). *Progress in Probability* **55** 127–134. Birkhäuser, Basel. MR2033885

[30] MEISTER, A. (2011). Asymptotic equivalence of functional linear regression and a white noise inverse problem. *Ann. Statist.* **39** 1471–1495. MR2850209

[31] R DEVELOPMENT CORE TEAM (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available at http://www.R-project.org.

[32] RUDIN, W. (1987). *Real and Complex Analysis*, 3rd ed. McGraw-Hill, New York. MR0924157

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

[34] VERZELEN, N. (2012). Minimax risks for sparse regressions: Ultra-high dimensional phenomenons. *Electron. J. Stat.* **6** 38–90. MR2879672

[35] VERZELEN, N. and VILLERS, F. (2010). Goodness-of-fit tests for high-dimensional Gaussian linear models. *Ann. Statist.* **38** 704–752. MR2604699

[36] YUAN, M. and CAI, T. T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *Ann. Statist.* **38** 3412–3444. MR2766857

N. HILGERT
N. VERZELEN
INRA, UMR 729 MISTEA
F-34060 MONTPELLIER
FRANCE
AND
SUPAGRO, UMR MISTEA
F-34060 MONTPELLIER
FRANCE
E-MAIL: nadine.hilgert@supagro.inra.fr
nicolas.verzelen@supagro.inra.fr

A. MAS
I3M
UNIVERSITÉ MONTPELLIER 2
F-34095 MONTPELLIER
FRANCE
E-MAIL: mas@math.univ-montp2.fr