



HAL
open science

Testing for the mean of random curves : a penalization approach

Andre Mas

► **To cite this version:**

Andre Mas. Testing for the mean of random curves : a penalization approach. *Statistical Inference for Stochastic Processes*, 2007, 10 (2), pp.147 - 163. <10.1007/s11203-005-0754-3>. <hal-01817402>

HAL Id: hal-01817402

<https://hal.science/hal-01817402v1>

Submitted on 17 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Testing for the mean of random curves : a penalization approach.

André MAS*

Université Montpellier II

Abstract

Let X_1, \dots, X_n be an i.i.d. sample of random curves, viewed as Hilbert space valued random elements, with mean curve m . An asymptotic test of $m = m_0$ vs $m \neq m_0$ is proposed, when m_0 is a fixed known function. The test statistics converges under very mild assumptions and relies on the pseudo-inversion of the covariance operator (leading to a non standard inverse problem). The power against local alternatives is investigated.

Key words : Random curves, functional data, weak convergence, hypothesis testing, local alternatives.

AMS Mathematics Subject Classification : 62E20, 62G10, 62G20.

1 Introduction

1.1 The functional statistics setting.

Inference on random curves is undoubtedly a soaring area in nonparametric statistics. Modern computational techniques make it possible to deal with "high dimensional random vectors". Data that are obtained from an underlying continuous-time process, for instance, are extremely common in finance, climatology, medicine, etc. Random curves collected from (independent or not) experiments are also likely to be studied by non parametric techniques. Some references are Franck and Friedman (1993), Cavallini, Montanari, Loggini, Lessi and Cacciari (1994), Besse, Cardot and Stephenson (2000) for applications in respectively chemometrics, industry and meteorology.

This trend revealed that, conversely to probabilists, statisticians sometimes lack theoretical results for studying random functions. Many authors anyway paid attention to these topics, developping methods to link this rather formal framework with the statistician's "everyday's life". Amongst these are Kneip and Gasser (1992), Ramsay and Silverman (1997). Dauxois, Pousse and Romain (1982), in an earlier paper, investigated the principal component analysis for Hilbert-valued random variables. Recently several authors generalized to the functional framework standard models in finite dimension : the linear regression model in Cardot, Ferraty and Sarda (1999), the autoregressive model for time series in Bosq (2000), the infinite moving average model in Mas (2002b).

*Mailing address : André Mas, Laboratoire de Probabilités-Statistiques, CC051, Université Montpellier 2, Place Eugène Bataillon, 34095 Montpellier Cedex 5.

We propose an asymptotic procedure to test for the mean of a random function. Let X_1, \dots, X_n be an i.i.d. sample of random curves with mean m (note that m is also a curve) defined on some abstract probability space (Ω, \mathcal{A}, P) . We should write $X_i(\omega, t)$ where for fixed $\omega \in \Omega$, $X_i(\omega, \cdot)$ is the path of the curve and where for fixed t $X_i(\cdot, t)$ is a real random variable. But for the sake of simplicity both indices will be dropped.

In this work we will suppose that the data are truly infinite-dimensional or functional. In practice, the random curves are observed only at discrete times. A version of the underlying process is then rebuilt by interpolation techniques such as splines, wavelets, etc. These issues (as well as the problem of estimating the curve m) will not be considered in the sequel since they give rise to a considerable literature. We refer for instance to Rice and Silverman (1991).

The test may be written :

$$\begin{cases} H_0 : m = m_0 \\ H_a : m \neq m_0 \end{cases}$$

where $m_0 = m_0(\cdot)$ is a fixed known function. The principle of such a test is not new. Several authors developed procedures in the framework of diffusion processes : X_1 is supposed to satisfy the stochastic differential equation

$$dX(t) = m(t) dt + \varepsilon dW(t)$$

where W is the standard Wiener process and ε is the noise level. We refer to Spokoiny (1996), Lepski and Tsybakov (2000) for instance. The function m is supposed to belong to a class of functions (Sobolev, Hölder, or more generally Besov) indexed by some smoothness parameter. This gaussian and typical framework allows these authors to derive very interesting results : adaptive (w.r.t. the function class) tests but also optimal rates of testing- optimal in a sense that is defined by introducing the notion of minimax rate of testing- when the noise level decays to zero. Fan (1996) and Fan and Lin (1998) also proposed, in connection with high-dimensional ANOVA, an adaptive Neyman test. Here adaptivity means that the procedure will automatically reduce the dimension of the model. All these authors apply wavelet thresholding techniques.

A new feature of the paper relies on the **removal of the assumption of Gaussian observations**, which was almost systematic in the above-mentioned articles. This framework makes the test truly new because we need to carry out alternative techniques based on lower moment assumptions, enlarging the class of process which may be investigated. The drawback is : we lose optimality properties. The procedure remains asymptotic and does not rely on a finite dimensional approximation of m . The limiting distribution is however extremely simple. Some may view it as goodness of fit test. It also turns out that the main technical problems arising may be expressed in terms of an ill-posed inverse problem.

The setup of the paper is closer to a recent work by Cardot, Ferraty, Mas and Sarda (2002). The authors propose a test for the regression operator in a linear model with functional inputs. The second serious point in this work is the *mild condition on the rate of decrease of the eigenvalues of the covariance operator*.

It should also be stressed that our problem does not depend on parameters intrinsic to the model or to spaces of functions. All the calculations just rely on the Hilbertian structure of the space, no matter which space is chosen. Consequently minimaxity or adaptivity make no sense.

From now on, the studied random curves will always be seen as random variables defined on the abstract probability space (Ω, \mathcal{A}, P) and with values in an infinite dimensional, real and separable Hilbert space H endowed with inner product $\langle \cdot, \cdot \rangle$ and norm

$\|\cdot\|$. The Hilbert space setting enables to consider different sorts of basis and also to derive similar results for the sequence space l_2 . It also makes computations easier (especially as far as the central limit theorem is concerned). In the special case when $H = L^2([0, 1])$ and for all random functions u and v in H (u and v are consequently defined on $\Omega \times [0, 1]$) $\langle u, v \rangle$ is a random number :

$$\langle u, v \rangle (\omega) = \int_0^1 u(\omega, t)v(\omega, t)dt. \quad (1)$$

Spaces of smooth functions, such as Sobolev spaces, may (and are often) preferred to L^2 spaces for stability reasons. We refer to Silverman (1996) for developments of this approach. Formally, this does not change much the inner product defined in (1) : Lebesgue's measure dt is replaced with another non-finite measure $d\mu(t)$ (about this well-known fact see Rudin (1973) Ch 8.8 : Sobolev Spaces).

1.2 Preliminary facts about operators

Let X_1, \dots, X_n be an i.i.d. sample of Hilbert-valued random variables with mean m . We are first going to assume that the covariance operator of these random variables exists. Assumptions about the second order structure of X_1 will appear in the next paragraph.

We denote Γ (resp. Γ_n) the covariance operator of X_1 (resp. the empirical covariance operator of the sample). These operators are bounded linear selfadjoint and positive mappings from H to H . They are defined this way : for all x in H ,

$$\begin{aligned} \Gamma(x) &= E [\langle X_1 - m, x \rangle (X_1 - m)], \\ \Gamma_n(x) &= \frac{1}{n} \sum_{k=1}^n [\langle X_k - m, x \rangle (X_k - m)]. \end{aligned}$$

In the following, for all u, v in H , $u \otimes v$ just stands for the rank-one operator defined for all x in H by $u \otimes v(x) = \langle u, x \rangle v$. With these notations, we get :

$$\begin{aligned} \Gamma &= E [(X_1 - m) \otimes (X_1 - m)], \\ \Gamma_n &= \frac{1}{n} \sum_{k=1}^n [(X_k - m) \otimes (X_k - m)]. \end{aligned}$$

We will now assume that the random variable X_1 has a finite strong moment of order two i.e. $E \|X_1\|^2 < +\infty$. This means that Γ is a trace class operator on H . If $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ denotes the ordered sequence of its real and positive eigenvalues (associated to the eigenvectors e_1, e_2, \dots) it also implies that the vector $(\lambda_1, \lambda_2, \dots) \in l_1$. For further information about these facts we refer to Vakhania, Tarieladze and Chobanyan (1987).

We will always suppose that Γ is one to one : the sequence of its strictly positive eigenvalues is infinite and that the set $\ker \Gamma = \{h \in H : \Gamma h = 0\}$ is reduced to $\{0\}$. Under the moment assumptions on X mentioned just above, it is well-known that Γ is a positive selfadjoint compact operator. For any linear operator T defined on and with values in H , $\|T\|_\infty$ stands for the usual norm of continuous linears operators i.e. $\|T\|_\infty = \sup \|Tx\|$ for x in the unit ball of H . If we go further in the moments assumption, we even know that $\sqrt{n} \|\Gamma_n - \Gamma\|$ is bounded in probability whenever $E \|X_1\|^4$ is finite.

The operator $\Gamma^{1/2}$ plays a key-role in the sequel. It is defined by

$$\Gamma^{1/2} = \sum_{k=1}^{+\infty} \sqrt{\lambda_k} (e_k \otimes e_k)$$

and is a compact operator. When Γ is one to one, its inverse $\Gamma^{-1/2}$ is defined on a domain \mathcal{D} in H (\mathcal{D} is a dense vector subspace of H). The linear mapping $\Gamma^{-1/2}$ is unbounded (i.e. $\Gamma^{-1/2}$ is continuous *at no point* of \mathcal{D}) which also means $\|\Gamma^{-1/2}\|_{\infty} = +\infty$. As a consequence $\Gamma^{-1/2}X_1$ cannot be considered as a (bounded) random variable. More information about these topics may be found in Dunford and Schwartz (1988).

2 Formulating the problem.

2.1 An original linear inverse problem

Now, in order to understand the particularities of the situation, let us consider first what happens *in the finite dimensional setting*. Let X_1, \dots, X_n be a sample of i.i.d. vectors of R^p with mean $m \in R^p$. The test is the following

$$\begin{cases} H_0 : m = m_0 \\ H_a : m \neq m_0. \end{cases}$$

The operator Γ_n is replaced with a square matrix of size p , say $M_n(p)$, the empirical covariance matrix of the sample, which is often supposed to be invertible for n large enough. Then the test statistics (usually chi-square) is derived from :

$$M_n(p)^{-1/2} \left(\frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - m_0) \right)$$

which converges under the null hypothesis to a gaussian random variable whose covariance matrix is the identity.

In our functional framework we still have : $S_n = n^{-1/2} \sum_{k=1}^n (X_k - m_0)$ converges weakly to a gaussian r.v. G by the central limit theorem for Hilbert space valued random elements but the distribution of G depends on the unknown λ_i 's (see Fact 2 of Appendix 1). These eigenvalues may be viewed as nuisance parameters. Besides Γ_n , conversely to M_n is never invertible in H even for large n since its range (the linear span of X_1, \dots, X_n) is of finite dimension. Considering " $\Gamma_n^{-1/2}S_n$ " makes no sense, as well as " $\Gamma^{-1/2}G$ " as was mentioned above. Consequently our goal is double. First we should approximate the operator $\Gamma^{-1/2}$ by a pseudo inverse, say L_n , built from Γ_n and hence random. Then we will have to study weak convergence for $L_n S_n$. This approach will be made more precise in the next section. Anyway we should expect the norm of L_n to be a non decreasing sequence tending to $+\infty$. We could also say that the sequence of operators L_n is *unstable or ill-conditioned*. Finally, copying the finite dimensional approach to our test procedure leads us to a *non standard ill-posed inverse problem* : G will be approximated by S_n and $\Gamma^{-1/2}$ by L_n but S_n and L_n are connected via the sample X_1, \dots, X_n . Fortunately we will see that it is possible to propose a convergent test procedure.

Approximating $\Gamma^{-1/2}$ by a sequence of bounded operators is the purpose of the forthcoming paragraph.

2.2 Pseudo-inverse estimators.

The problem of approximating inverses of selfadjoint compact operators is absolutely not new. It is addressed in Nashed and Wahba (1974), Arsenin and Tikhonov (1977), Groetsch (1993) amongst many others. The main point is always to regularize a matrix M (resp. an operator S) which is invertible but "not by much" (resp. unbounded). This property implies that for any vector x , Mx (resp. Sx) may have large variations even when x does not vary much. Numerous procedures were proposed. We will keep one which is suited to our problem and may be easily implemented.

The penalized regularization procedure is based on $(\Gamma + \alpha I)^{-1/2}$. It is a ridge-type method and depends on the index α which is a strictly positive real number. The penalization term is αI where I is the identity operator on H . The operator $(\Gamma + \alpha I)^{-1/2}$ is continuous for all strictly positive α with norm $1/\sqrt{\alpha}$.

Another way to regularize $\Gamma^{-1/2}$ consists in deleting all the terms beyond a threshold p . In the basis e_i we get

$$\Gamma_{(p)}^{-1/2} = \text{diag} \left(\lambda_1^{-1/2}, \lambda_2^{-1/2}, \dots, \lambda_p^{-1/2}, 0, 0, \dots \right).$$

$\Gamma_{(p)}^{-1/2}$ is a bounded operator with norm $\lambda_p^{-1/2}$ and is known as the spectral truncation (S.T.) of order p of $\Gamma^{-1/2}$. But $\Gamma_{(p)}^{-1/2}$ which is often used has several drawbacks :

- The S.T. estimator is based on the functional Principal Component Analysis of Γ_n . It is necessary to estimate the eigenvectors and eigenvalues of this random operator before projecting the sample $X_1 - m, \dots, X_n - m$ on these eigenvectors (to obtain principal components as a by-product). The estimation procedure is consequently not that simple and usually entails serious stability problems. Estimating the penalized estimate just requires a **nonrandom basis**, (e.g. spline, Fourier, wavelet). The data are then projected on a known vector space.
- As it may be seen in Cardot, Ferraty and Sarda (1999), and Bosq (1991), convergence rates of estimates stemming from the S.T. procedure always depend on the speed of decay of the (unknown) eigenvalues and eigenvectors of Γ . Several assumptions are usually made, restricting the generality of the results.
- Speed of convergence of estimates involving the S.T. estimator are usually not good, due to the very slow rate of uniform convergence of the empirical eigenvectors of Γ_n to the eigenvectors of Γ for large values of p .

Also note that, conversely to the S.T. estimator, the norm of the penalized one, $1/\sqrt{\alpha_n}$, is **nonrandom** and **does not depend on the rate of decay of the eigenvalues of Γ or on the observed data**. All the previous reasons explain the final choice of a penalization procedure for regularizing the inverse of the square root of the covariance operator.

However even if the regularized operator does not depend explicitly on the eigenvalues, it should be stressed that the test statistics depends on them. A good knowledge of the rate of decay of the eigenvalues is even crucial to compute smoothing parameters intrinsic to our estimates (the above-mentioned α_n as well as k_n , mentioned below-see next section.)

3 Main results

Weak convergence (resp. convergence in probability) will be denoted \xrightarrow{w} (resp. \xrightarrow{P}) in the sequel and will be considered with respect to the norm of the space H . As mentioned above, we denote \mathcal{D} the domain of the operator $\Gamma^{-1/2}$ i.e. the set of points x in H for which $\Gamma^{-1/2}x$ has a finite norm. Clearly if x is expressed in the basis e_i , $x = \sum x_k e_k$. We have :

$$\mathcal{D} = \left\{ x \in H : \sum_k \frac{x_k^2}{\lambda_k} < +\infty \right\}.$$

Before giving the main results, we recapitulate the two assumptions needed :

$$\mathbf{H1} : E \|X_1\|^4 < +\infty.$$

$$\mathbf{H2} : \Gamma \text{ is one to one.}$$

We set once and for all :

$$S_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - m_0).$$

and we know that $S_n \xrightarrow{w} G$ where G is the H valued gaussian random element with mean 0 and covariance operator Γ . For further purpose we finally recall that the following well-known decomposition for G holds almost surely on an appropriate probability space

$$G = \sum_{p=1}^{+\infty} \sqrt{\lambda_p} \eta_p e_p$$

where the η_p 's are gaussian centered random variables with unit variance.

3.1 Weak convergence

Let the sequences k_n (of integers) and α_n (of nonnegative numbers) respectively increase to infinity and decrease to zero We set

$$\hat{c}_n = \sum_{p=1}^{k_n} \frac{\hat{\lambda}_p}{\hat{\lambda}_p + \alpha_n} \quad \text{and} \quad \hat{d}_n = \sqrt{\sum_{p=1}^{k_n} \left(\frac{\hat{\lambda}_p}{\hat{\lambda}_p + \alpha_n} \right)^2}$$

where the $\hat{\lambda}_p$ are the estimated eigenvalues of Γ_n . The random term \hat{c}_n (resp. \hat{d}_n) is a bias term (resp. a normalizing term).

Remark 3.1 *The $\hat{\lambda}_p$'s are directly estimated from the (smooth) PCA of the X_i 's hence from the spectrum of Γ_n . Asymptotic results (almost sure convergence, weak convergence) are given in Dauxois, Pousse and Romain (1982).*

The test statistic is

$$\hat{T}_n = \frac{1}{\hat{d}_n} \left(\left\| (\Gamma_n + \alpha_n)^{-1/2} S_n \right\|^2 - \hat{c}_n \right)$$

Theorem 3.1 : Under **H1** and **H2** there exists two conjugated sequences k_n and α_n such that, when H_0 holds ($m = m_0$) :

$$\widehat{T}_n \xrightarrow[n \rightarrow +\infty]{w} \mathcal{N}(0, 3).$$

Remark 3.2 : Note that Theorem 3.1 which just asserts **the existence of the test procedure is valid without any assumptions on the distribution of the data or on the spectrum of Γ** . The crucial step in the proofs relies on a result by Mas (2002a) on the estimate of closeness of distributions for images of measures on Banach spaces by general families of mappings. The distance between distributions is expressed in terms of the Prokhorov metric. We refer to the Appendix for the main properties of this metric as well as references.

The next theorem determines explicitly both sequences k_n and α_n such that \widehat{T}_n converges. The "price to pay" is quite low and relies on the forthcoming uniform bound on the eigenvalues :

$$\mathbf{H3} : \text{for some } M, \varepsilon > 0, \lambda_p \leq M / (p(\log p)^{1+\varepsilon})$$

Theorem 3.2 Let $0 < \nu < 1/2$. Under **H1**, **H2** and **H3** when $k_n = n^{1/2-\nu}$ and $\alpha_n = (\ln n)^{-\varepsilon/(1+\nu\varepsilon)}$, and when H_0 holds :

$$\widehat{T}_n \xrightarrow[n \rightarrow +\infty]{w} \mathcal{N}(0, 3).$$

Remark 3.3 Assumption **H3** may be viewed as prior information on the eigenvalues. The estimates $\widehat{\lambda}_p$'s obviously provide a good knowledge of the unknown λ_p 's. If however more information is available and if the bound provided in **H3** can be made more precise, one may get deeply different rates of increase (resp. of decrease) for k_n (resp. α_n). An inspection of the proof would lead to the following table :

<i>Rate of decay for the eigenvalues</i>	<i>exponential</i>	<i>geometric</i>	<i>assumption H3</i>
k_n	log	geometric	geometric
α_n	geometric	geometric	log

The point is : the respective "speed" of k_n and α_n are inverted. When the problem is "highly ill-conditionned" (i.e an exponential rate of decrease for the eigenvalues) k_n should increase slowly. Conversely when the eigenvalues are assumed to converge "slowly" to zero (assumption H3) k_n may be large and the penalization parameter may also decrease slowly. The reduction of dimension induced by k_n is implicitly governed by α_n which depends on the "illness" of the inverse problem.

3.2 Study of the power for local alternatives

For the sake of completeness, the test procedure is given below :

- Fix m_0 and compute \widehat{T}_n where $S_n = n^{-1/2} \sum_{k=1}^n (X_k - m_0)$ and

$$\Gamma_n = \frac{1}{n} \sum_{k=1}^n (X_k - m_0) \otimes (X_k - m_0).$$

- Fix a level of significance κ and compute u_κ such that $P(|N(0, 1)| \leq u_\kappa) = 1 - \kappa$
- If $|\widehat{T}_n| \leq \sqrt{3}u_\kappa$ H_0 is accepted otherwise, it is rejected.

A very classical way to investigate the power of the test consists in studying the power function against sequences of alternatives tending to the null. Let us take

$$H_a(\delta_n) : m = m_0 + \delta_n t.$$

where $t \in H$ and $\delta_n > 0$ with $\delta_n \downarrow 0$. We are interested in computing the "optimal" rate of decay for δ_n i.e. the fastest rate of decay for which the test distinguishes $H_a(\delta_n)$ from H_0 at a prescribed level.

At this point we need to recall a notation. Remind that the set

$$\mathcal{D} = \left\{ x \in H : \sum_k \frac{x_k^2}{\lambda_k} < +\infty \right\}$$

is the domain of $\Gamma^{-1/2}$. Consequently $\|\Gamma^{-1/2}t\|$ is finite if and only if $t \in \mathcal{D}$.

Proposition 3.1 *The optimal rate of convergence of the sequence of local alternatives to the null is $\delta_n = \sqrt{d_n/n}$. Let us pick a t in H and distinguish two cases.*

First case : *If $t \notin \mathcal{D}$ the test statistic \widehat{T}_n diverges for the sequence of alternatives $H_a(\sqrt{d_n/n})$*

Second case : *If $t \in \mathcal{D}$, \widehat{T}_n converges weakly to a $\mathcal{N}(\|\Gamma^{-1/2}t\|^2, 3)$ for the sequence of alternatives $H_a(\sqrt{d_n/n})$.*

Remark 3.4 *Since $d_n \leq C\alpha_n^{-1}$, considerations made above prove that the power of the test is directly connected to the "ill-posedness" situation. The eventual bad rate is not only due to the infinite dimensional setting but gets worse because of the inverse problem.*

4 Proofs

Along the proofs C and M will stand for fixed constants whose values do not matter. Remind that the notation $Y_n = O_P(t_n)$ means that the sequence of random variables $Y_n t_n^{-1}$ is bounded in probability. Let us denote

$$\widetilde{T}_n = \frac{1}{d_n} \left(\left\| (\Gamma_n + \alpha_n)^{-1/2} S_n \right\|^2 - c_n \right)$$

where

$$c_n = \sum_{p=1}^{k_n} \frac{\lambda_p}{\lambda_p + \alpha_n}, d_n = \sqrt{\sum_{p=1}^{k_n} \left(\frac{\lambda_p}{\lambda_p + \alpha_n} \right)^2}$$

Proposition 4.1 *If $k_n/(\alpha_n\sqrt{n}) \leq 1$,*

$$\widetilde{T}_n - \widehat{T}_n \xrightarrow{P} 0.$$

Proof of the Proposition :

Straightforward considerations lead in fact to proving

$$\frac{c_n - \widehat{c}_n}{d_n} \xrightarrow{P} 0$$

and

$$d_n - \widehat{d}_n = O_P(1)$$

since obviously \widehat{d}_n is not bounded in probability (d_n increases to infinity when k_n does and $\alpha_n \rightarrow 0$). Elementary calculations give

$$\begin{aligned} |c_n - \widehat{c}_n| &\leq \alpha_n \sum_{p=1}^{k_n} \frac{|\lambda_p - \widehat{\lambda}_p|}{(\lambda_p + \alpha_n)(\widehat{\lambda}_p + \alpha_n)} \\ &\leq \frac{k_n}{\alpha_n} \sup_{1 \leq p \leq k_n} |\lambda_p - \widehat{\lambda}_p| \end{aligned}$$

We invoke Dauxois, Pousse and Romain (1982) to claim that $\sqrt{n} \sup_{1 \leq p \leq k_n} |\lambda_p - \widehat{\lambda}_p|$ is bounded in probability. Hence

$$|c_n - \widehat{c}_n| = O_P\left(\frac{k_n}{\alpha_n \sqrt{n}}\right) \quad (2)$$

The same steps lead to an analogous result to get $|d_n - \widehat{d}_n| = O_P\left(\frac{k_n}{\alpha_n \sqrt{n}}\right)$ again. ■

Proposition 4.2 : *Setting $m_n = \sum_{p=k_n}^{+\infty} \frac{\lambda_p}{\lambda_p + \alpha_n}$ We have :*

$$\frac{1}{d_n} \left| \left\| (\Gamma_n + \alpha_n I)^{-1/2} S_n \right\|^2 - \left\| (\Gamma + \alpha_n I)^{-1/2} S_n \right\|^2 \right| = O_P\left(\frac{\sqrt{c_n + m_n}}{\alpha_n^2 d_n \sqrt{n}}\right). \quad (3)$$

Proof of the Proposition :

$$\begin{aligned} &\frac{1}{d_n} \left| \left\| (\Gamma_n + \alpha_n I)^{-1/2} S_n \right\|^2 - \left\| (\Gamma + \alpha_n I)^{-1/2} S_n \right\|^2 \right| \\ &= \frac{1}{d_n} \left| \langle (\Gamma_n + \alpha_n I)^{-1} S_n, S_n \rangle - \langle (\Gamma + \alpha_n I)^{-1} S_n, S_n \rangle \right| \\ &= \frac{1}{d_n} \left| \langle ((\Gamma_n + \alpha_n I)^{-1} - (\Gamma + \alpha_n I)^{-1}) S_n, S_n \rangle \right| \\ &\leq \frac{1}{d_n} \left\| ((\Gamma_n + \alpha_n I)^{-1} - (\Gamma + \alpha_n I)^{-1}) S_n \right\| \|S_n\| \\ &= \frac{1}{d_n} \left\| ((\Gamma_n + \alpha_n I)^{-1} (\Gamma - \Gamma_n) (\Gamma + \alpha_n I)^{-1}) S_n \right\| \|S_n\| \end{aligned}$$

This last term may be bounded by

$$\begin{aligned} &\frac{1}{d_n} \left\| (\Gamma_n + \alpha_n I)^{-1} (\Gamma - \Gamma_n) (\Gamma + \alpha_n I)^{-1/2} \right\|_{\infty} \left\| (\Gamma + \alpha_n I)^{-1/2} S_n \right\| \|S_n\| \\ &\leq \frac{1}{\alpha_n^{3/2} d_n} \|\Gamma - \Gamma_n\|_{\infty} \left\| (\Gamma + \alpha_n I)^{-1/2} S_n \right\| \|S_n\| \end{aligned}$$

When **H1** holds $\|\Gamma - \Gamma_n\|_\infty = O_P(1/\sqrt{n})$. We turn to $\left\|(\Gamma + \alpha_n I)^{-1/2} S_n\right\|$

$$\begin{aligned} P\left(\left\|(\Gamma + \alpha_n I)^{-1/2} S_n\right\| > M\right) &\leq \frac{1}{M^2} E \left\|(\Gamma + \alpha_n I)^{-1/2} S_n\right\|^2 \\ &= \frac{1}{M^2} E \left\|(\Gamma + \alpha_n I)^{-1/2} X_1\right\|^2 \\ &= \frac{1}{M^2} \sum_{p=1}^{+\infty} \frac{\lambda_p}{\lambda_p + \alpha_n} \leq \frac{c_n + m_n}{M^2} \end{aligned}$$

and $\left\|(\Gamma + \alpha_n I)^{-1/2} S_n\right\| = O_P(\sqrt{m_n + c_n})$ and we can conclude as announced

$$\frac{1}{d_n} \left| \left\|(\Gamma_n + \alpha_n I)^{-1/2} S_n\right\|^2 - \left\|(\Gamma + \alpha_n I)^{-1/2} S_n\right\|^2 \right| = O_P\left(\frac{\sqrt{m_n + c_n}}{\alpha_n^{3/2} d_n \sqrt{n}}\right) \quad (4)$$

■

For the next Proposition we need to introduce Π^{k_n} the projector on the k_n first eigenvectors of Γ

Proposition 4.3 : *It is possible to find coupled sequences α_n and k_n such that the limiting distribution of*

$$\tilde{T}_n = \frac{1}{d_n} \left(\left\|(\Gamma + \alpha_n I)^{-1/2} S_n\right\|^2 - c_n \right)$$

is the same (if it exists) as the limiting distribution of

$$T_{G,n} = \frac{1}{d_n} \left(\left\|\Pi^{k_n} (\Gamma + \alpha_n I)^{-1/2} G\right\|^2 - c_n \right).$$

Proof of the Proposition :

The proof of this lemma is based on a recent result by Mas (2002a). It should be first stressed that calculations show that \tilde{T}_n cannot be written as a sum of n random variables. Indeed

$$\tilde{T}_n = \frac{1}{d_n} \left(\left\|(\Gamma + \alpha_n I)^{-1/2} S_n\right\|^2 - c_n \right).$$

If the square norm is developed, n^2 terms will appear. Consequently, it seems that it is not possible to prove the weak convergence of \tilde{T}_n by standard methods. The famous theorems of preservation of weak convergence by mappings (e.g. Billingsley, 1968 p.30 and 31) cannot be applied either. We introduce π_E the Prokhorov metric for measures on a Banach space E . Reformulating the Proposition and identifying a random variable on H with the measure it induces we will get the desired result if

$$\pi_R(\tilde{T}_n, T_{G,n}) \rightarrow 0.$$

Or equivalently

$$\frac{1}{d_n} \pi_R\left(\left\|(\Gamma + \alpha_n I)^{-1/2} S_n\right\|^2, \left\|\Pi^{k_n} (\Gamma + \alpha_n I)^{-1/2} G\right\|^2\right) \rightarrow 0.$$

First we bound

$$\begin{aligned} & \pi_H \left((\Gamma + \alpha_n I)^{-1/2} S_n, \Pi^{k_n} (\Gamma + \alpha_n I)^{-1/2} G \right) \\ & \leq \pi_H \left(\Pi^{k_n} (\Gamma + \alpha_n I)^{-1/2} S_n, \Pi^{k_n} (\Gamma + \alpha_n I)^{-1/2} G \right) \\ & + \pi_H \left((\Gamma + \alpha_n I)^{-1/2} S_n, \Pi^{k_n} (\Gamma + \alpha_n I)^{-1/2} S_n \right). \end{aligned}$$

The last term may be bounded by $m_n^{1/3}$. We refer to the Appendix for further explanations about this bound. The first term may be bounded, following Theorem 1 p.236 in Vakhania (1977) by :

$$C \frac{k_n^{1/4} |\ln n|^{1/2}}{n^{1/8} \alpha_n^{3/8}}.$$

For the sake of clarity we denote $\pi_H = \pi_H \left((\Gamma + \alpha_n I)^{-1/2} S_n, \Pi^{k_n} (\Gamma + \alpha_n I)^{-1/2} G \right)$.

$$\pi_H \leq m_n^{1/3} + C \frac{k_n^{1/4} |\ln n|^{1/2}}{n^{1/8} \alpha_n^{3/8}} = \tau_n.$$

We will get the final result whenever we can express the Prokhorov distance between $\|U\|^2$ and $\|V\|^2$ only by means of the Prokhorov distance between U and V . This was precisely done in Case 1 of paragraph 4.1 in Mas (2002a). We have :

$$\pi_R \left(\tilde{T}_n, T_{G,n} \right) \leq \frac{C}{d_n} \pi_H \sqrt{-\ln \pi_H}$$

for a sufficiently large n . Since obviously the function $s \rightarrow s\sqrt{-\ln s}$ is non decreasing in a positive neighborhood of zero, the final bound is

$$\pi_R \left(\tilde{T}_n, T_{G,n} \right) \leq C \frac{\tau_n \sqrt{-\ln \tau_n}}{d_n}. \quad (5)$$

We are going to cope with each term separatedly.

On a one hand it is always possible to fix two conjugate sequences k_n and α_n (depending on each other) such that

$$m_n \rightarrow 0. \quad (6)$$

On the other hand, once either k_n or α_n is determined, it is possible to choose the other such that

$$\frac{k_n^{1/4} |\ln n|^{1/2}}{n^{1/8} \alpha_n^{3/8}} \rightarrow 0 \quad (7)$$

hence the Proposition. ■

Proposition 4.4 :

$$\frac{1}{d_n} \left\| \Pi^{k_n} (\Gamma + \alpha_n I)^{-1/2} G \right\|^2 - c_n \xrightarrow{w} \mathcal{N}(0, 3).$$

Proof of the Proposition :

We have successively :

$$\begin{aligned} & \frac{1}{d_n} \left\| \Pi^{k_n} (\Gamma + \alpha_n I)^{-1/2} G \right\|^2 - c_n \\ &= \frac{\sum_{p=1}^{k_n} \frac{\lambda_p}{\lambda_p + \alpha_n} (\eta_p^2 - 1)}{d_n}. \end{aligned}$$

The central limit theorem for triangular arrays of independent random variables yields the final result (see Billingsley (1968) p.42). ■

Proof of Theorem 3.1 :

Collecting Propositions 4.1, 4.2, 4.3 we have to check that the three rates of convergence are not contradictory, namely those obtained in (2), (4), (5). The conclusion will then follow from Proposition (4.4).

We first go back to (6). This conditions still implies an underlying link between k_n and α_n . It does not seem possible to compare condition (7) and condition (4) but both will be fulfilled whenever

$$\frac{\sqrt{c_n} (\ln n)^2}{\alpha_n^{3/2} d_n \sqrt{n}} \rightarrow 0 \quad (8)$$

and (2) is obviously respected in this case. Note also that $\sqrt{c_n} \leq \sqrt{k_n}$. Consequently we are again in the situation of Proposition 4.3 but (7) is replaced with (8) and Theorem 3.1 follows. ■

Proof of Theorem 3.2 :

Assumption **H3** implies that

$$m_n \leq C \frac{1}{\alpha_n (\ln k_n)^\varepsilon}$$

Setting $\ln k_n = \alpha_n^{-\nu-1/\varepsilon}$ where $\nu > 0$ implies that the term above tends to zero. The expression in (8) is now

$$\frac{1}{d_n} \frac{(k_n)^{1/2} (\ln k_n)^{3\varepsilon/2(1+\nu\varepsilon)} (\ln n)^2}{n^{1/2}}$$

Taking at last, for instance $k_n = n^{1/2-\nu}$ gives the final result. ■

Proof of Proposition 3.1 :

The method of the proof for Theorem 3.1 and especially the use of Prokhorov metric directly yields the following : under $H_a(\delta_n)$ the distribution of \widehat{T}_n is asymptotically the same as the distribution of

$$\frac{1}{d_n} \left\| \Pi^{k_n} (\Gamma + \alpha_n I)^{-1/2} (G - \sqrt{n} \delta_n t) \right\|^2 - c_n$$

We prove just below that taking $\delta_n = d_n/\sqrt{n}$ is sufficient for the sequences of alternatives to be detected by the test procedure.

In fact we have to consider two cases. The first one is quite simple : if $t \notin \mathcal{D}$, $\Pi^{k_n} (\Gamma + \alpha_n I)^{-1/2} t$ does not converge at all and the test statistic diverge if $\delta_n = \sqrt{d_n/n}$.

The second case is not that much complicated. If $t \in \mathcal{D}$, $\Pi^{k_n} (\Gamma + \alpha_n I)^{-1/2} t$ converges to $\Gamma^{-1/2} t$ and we have to determine the limiting distribution of the above displayed

expression. But

$$\begin{aligned}
& \frac{1}{d_n} \left\| \Pi^{k_n} (\Gamma + \alpha_n I)^{-1/2} (G - \sqrt{n} \delta_n t) \right\|^2 - c_n \\
&= \frac{1}{d_n} \left\| \Pi^{k_n} (\Gamma + \alpha_n I)^{-1/2} G \right\|^2 - c_n \\
&+ \left\| \Pi^{k_n} (\Gamma + \alpha_n I)^{-1/2} t \right\|^2 \\
&- \frac{2}{\sqrt{d_n}} \left\langle \Pi^{k_n} (\Gamma + \alpha_n I)^{-1/2} G, \Pi^{k_n} (\Gamma + \alpha_n I)^{-1/2} t \right\rangle
\end{aligned}$$

The first term obviously converges weakly to a $\mathcal{N}(0, 3)$, the second is deterministic and converges to $\left\| \Gamma^{-1/2} t \right\|^2$. The last term has the same distribution as a centered gaussian real random variable whose variance is

$$s_n^2 = \frac{4}{d_n^2} \sum_{p=1}^{k_n} \frac{\lambda_p}{(\lambda_p + \alpha_n)^2} t_p^2$$

The series converges to $\sum_{p=1}^{+\infty} \frac{t_p^2}{\lambda_p}$ (apply Lebesgue's dominated convergence theorem) which ensures that s_n tends to zero as $d_n \uparrow +\infty$ and concludes the proof. ■

Appendix

Definition and some properties of the Prokhorov metric

Along the proofs we used a metric defined on spaces of measures. This (Prokhorov) distance metrizes the topology of weak convergence for probability measures on a separable Banach space E . It will be denoted π_E . We refer to Billingsley (1968), Dudley (1968) and Araujo and Giné (1980) for an introduction to the main features of π_E . The definition is the following (it is given for two measures μ_1 and μ_2 on E but is immediately adaptable to random elements) :

$$\pi_E(\mu_1, \mu_2) = \inf \left\{ \varepsilon > 0 : \begin{array}{l} \mu_1(A^\varepsilon) \leq \mu_2(A) + \varepsilon \\ \mu_2(A^\varepsilon) \leq \mu_1(A) + \varepsilon \end{array}, A \in \mathcal{F} \right\}$$

where \mathcal{F} is the family of closed sets in E and A^ε stands for the ε -neighborhood of A (with respect to the norm on E). In fact π_E may be defined for other family of sets than \mathcal{F} .

Within the proofs we need the following property of π_E .

First, Theorem 1 in Dudley (1968) implies that

$$\pi_H(X, Y) \leq \inf \{ \varepsilon > 0 : P(\|X - Y\| > \varepsilon) \leq \varepsilon \}$$

Hence if Π^k denotes the projector on the k first eigenvectors of the covariance operator of the H valued random variable X

$$P(\|X - \Pi^k X\| > \varepsilon) \leq \frac{\sum_{p=k+1}^{+\infty} \lambda_p}{\varepsilon^2}$$

by Tchebychev inequality and

$$\pi_H(X, \Pi^k X) \leq \left(\sum_{p=k+1}^{+\infty} \lambda_p \right)^{1/3} \tag{9}$$

which was the result needed to get m_n in the proof of Proposition 4.3.

References

- [1] Araujo, A. and Giné, E.: *The central limit theorem for real and Banach valued random variables*, Wiley and Sons, 1980.
- [2] Arsenin, V., and Tikhonov, A.: *Solutions of ill-posed problems*, Winston and Sons, Washington D.C., 1977.
- [3] Besse, P., Cardot, H. and Stephenson, D.B.: Autoregressive forecasting of some functional climatic variations, *Scand. Journal Stat.* **27** (2000), 673–687.
- [4] Billingsley, P.: *Convergence of probability measures*, Wiley Series in Probability and Mathematical Statistics, 1968.
- [5] Bosq, D.: Modelization, non-parametric estimation and prediction for continuous time processes, in Roussas (ed), *Nonparametric Functional Estimation and Related Topics*, NATO, ASI Series, 1991, pp. 509–529.

- [6] Bosq, D.: *Linear processes in function spaces*, Lecture Notes in Statistics, Springer-Verlag, 2000.
- [7] Cardot, H., Ferraty, F. and Sarda, P.: Functional linear model, *Statist. & Prob. Letters* **45** (1999), 11–22.
- [8] Cardot, H., Ferraty, F., Mas, A. and Sarda, P.: Testing hypotheses in the functional linear model. *Scand. Journal Stat*, **99** (2002), 1-15.
- [9] Cavallini, A., Montanari, G.C., Loggini, M., Lessi, O. and Cacciari, M.: Nonparametric prediction of harmonic levels in electrical networks, in *Proc. 6th IEEE ICHPS*, Bologna, 1994.
- [10] Dauxois, J., Pousse, A. and Romain, Y.: Asymptotic theory for the principal component analysis of a random vector function: some applications to statistical inference, *Journal Mult. Anal.* **12** (1982), 136–154.
- [11] Dudley, R.M.: Distances of Probability Measures and Random Variables, *Ann. Math. Statist.* **39** (1968), 1563–1572.
- [12] Dunford, N. and Schwartz, J.T.: *Linear operators*, Vol. 2. Wiley, New-York, 1962.
- [13] Fan, J.: Test of significance based on wavelet thresholding and Neyman’s truncation, *JASA* **91** (1996), 674-688.
- [14] Fan, J. and Lin, S.K.: Test of significance when data are curves. *JASA* **93** (1998), 1007-1022.
- [15] Franck, I.E., Friedman, J.H.: A statistical view of some chemometrics regression tools, *Technometrics* **35** (1993), 109–148.
- [16] Groetsch, C.: *Inverse Problems in the Mathematical Sciences*, Vieweg, Wiesbaden, 1993.
- [17] Kneip, A. and Gasser, T.: Statistical tools to analyze data representing a sample of curves, *Ann. Statist.* **20** (1992), 1266–1305.
- [18] Lepski, O.V. and Tsybakov, A.B.: Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point, *Probab. Theory. Relat. Fields* **117** (2000), 17-48.
- [19] Mas, A.: Rates of weak convergence for images of measures by families of mappings, *Stat. Proba. Letters* **56** (2002a), 7–12.
- [20] Mas, A.: Weak convergence for the covariance operators of a Hilbertian linear process, *Stoch. Processes and Appl.* **99** (2002b), 117-135.
- [21] Nashed, M.Z. and Wahba, G.: Generalized inverses in reproducing kernel spaces: An approach to regularization of linear operator equations, *SIAM Journal on Math. Analysis* **6** (1974), 974–987.
- [22] Ramsay, J.O. and Silverman, B.W.: *Functional Data Analysis*, Springer-Verlag, Berlin, 1997.

- [23] Rice, J.A. and Silverman, B.W.: Estimating the mean and covariance structure nonparametrically when the data are curves, *J. Roy. Stat. Soc. Ser. B*, **53** (1991), 233–243.
- [24] Rudin, W.: *Functional Analysis*, Mc Graw-Hill, New-York, 1973.
- [25] Silverman, B.W.: Smoothed principal component Analysis by choice of norm, *Ann. Statist.* **24** (1996), 1–24.
- [26] Spokoiny, V.G.: Adaptive hypothesis testing using wavelets, *Ann. Statist.* **24** (1996), 2477–2498.
- [27] Vakhania, N.N., Tarieladze, V.I. and Chobanyan, S.M.: Probability Distributions on Banach Spaces, Mathematics and its Applications, D. Reidel Publishing, 1987.
- [28] Whitt, W.: Preservation of rates of convergence under mappings, *Z. Wahr. verw. Gebiete* **29** (1974), 39–44.
- [29] Yurinskii, V.V.: On the error of the Gaussian approximation for convolutions, *Theory Prob. Applications* **2** (1977), 236–247.