ON THE LIMIT DISTRIBUTION OF THE CANONICAL CORRELATION COEFFICIENTS BETWEEN THE PAST AND THE FUTURE OF A HIGH-DIMENSIONAL WHITE NOISE

D. Tieplova[1,2], *P. Loubaton*[1], *L. Pastur* [2]

[1] Laboratoire IGM, Univ. Paris-Est/Marne la Vallée, UMR CNRS 8049, France

[2] Dept. of Theoretical Physics, Institute for Low Temperature Physics and Engineering, Ukraine

ABSTRACT

It is shown that the distribution of the estimated canonical correlation coefficients between the past and the future of a high-dimensional multivariate white noise sequence converges almost surely towards a limit distribution whose density is given in closed form. A sketch of proof, based on free probability technics, is provided. Finally, it is briefly explained how this result can be used to produce consistent uncorrelatedness tests in the high-dimensional context.

***Index Terms*—** High-dimensional time series, canonical correlation analysis, large random matrix

1. INTRODUCTION

The canonical correlation coefficients between 2 linear subspaces $\mathcal{Y}_1$ and $\mathcal{Y}_2$ contained in some ambient Hilbert space are defined as the singular values of the projection operator from $\mathcal{Y}_1$ onto $\mathcal{Y}_2$. If $(\omega_{i,1})_{i\in I}$ and $(\omega_{j,2})_{j\in J}$ represent any orthonormal bases of $\mathcal{Y}_1$ and $\mathcal{Y}_2$ respectively, the canonical correlation coefficients coincide with the singular values of the matrix with entries $(<\omega_{i,1}, \omega_{j,2}>)_{i\in I, j\in J}$ where $<,>$ represents the scalar product of the ambient space. This concept was introduced in multivariate analysis by Hotelling (see e.g. [8]) when $\mathcal{Y}_1$ and $\mathcal{Y}_2$ represent the spaces generated by the components of 2 Gaussian zero mean random vectors $\mathbf{y}_1$ and $\mathbf{y}_2$. In this context, the canonical correlation coefficients allow in some sense to quantify the information that can be obtained on the linear combinations of the components of $\mathbf{y}_i$ by observing $\mathbf{y}_j$ for $i \neq j$. This led to the introduction of the very popular canonical correlation analysis between 2 sets of random variables. The canonical correlation coefficients can also be defined in time series analysis in order to evaluate the relationships between the past and the future of a given Gaussian zero mean multivariate time series $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ (see e.g. [9]). In this context, the 2 subspaces, denoted here $\mathcal{Y}_p$ (the past) and $\mathcal{Y}_f$ (the future), are defined respectively as the spaces generated by the components of $\mathbf{y}_n$ for $n \leq 0$ and the components of $\mathbf{y}_n$ for $n > 0$. The canonical correlation coefficients between the past and the future of $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ are of fundamental interest if $\mathbf{y}$ has a rational spectrum because the number r of non zero canonical correlation coefficients is finite, and coincides with the minimal dimension of the state-space representations of $\mathbf{y}$. We refer the reader to [10] for an exhaustive presentation of the related results and their important implications on questions such as the identification of state space models or reduction model technics. See also the concise monography [15]. In a number of practical procedures, $\mathcal{Y}_p$ and $\mathcal{Y}_f$ are replaced by the finite dimensional spaces $\mathcal{Y}_{p,L}$ and $\mathcal{Y}_{f,L}$ generated respectively by the components of $\mathbf{y}_n, n = -(L-1), \ldots, 0$ and $\mathbf{y}_n, n = 1, \ldots, L$ for a certain integer $L \geq r$, a condition that implies that the number of non zero coefficients between $\mathcal{Y}_{p,L}$ and $\mathcal{Y}_{f,L}$ is still equal to r. We refer again to [10] for more details on the effects of the truncation. As the second order statistics of $\mathbf{y}$ are very often unknown, the correlation coefficients between $\mathcal{Y}_{p,L}$ and $\mathcal{Y}_{f,L}$ have to be estimated from N avalaible samples $\mathbf{y}_1, \ldots, \mathbf{y}_N$. If $\mathbf{Y}_{p,L}$ and $\mathbf{Y}_{f,L}$ are the two block Hankel $ML \times N$ matrices defined by

$$\mathbf{Y}_{p,L} = \begin{pmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \ldots & \mathbf{y}_{N-1} & \mathbf{y}_N \\ \mathbf{y}_2 & \mathbf{y}_3 & \ldots & \mathbf{y}_N & \mathbf{y}_{N+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{y}_L & \mathbf{y}_{L+1} & \ldots & \mathbf{y}_{N+L-2} & \mathbf{y}_{N+L-1} \end{pmatrix} \tag{1}$$

and

$$\mathbf{Y}_{f,L} = \begin{pmatrix} \mathbf{y}_{L+1} & \mathbf{y}_{L+2} & \ldots & \mathbf{y}_{N-1+L} & \mathbf{y}_{N+L} \\ \mathbf{y}_{L+2} & \mathbf{y}_{L+3} & \ldots & \mathbf{y}_{N+L} & \mathbf{y}_{N+L+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{y}_{2L} & \mathbf{y}_{2L+1} & \ldots & \mathbf{y}_{N+2L-2} & \mathbf{y}_{N+2L-1} \end{pmatrix} \tag{2}$$

the correlation coefficients between $\mathcal{Y}_{p,L}$ and $\mathcal{Y}_{f,L}$ are usually estimated by the canonical correlation coefficients between the row spaces of $\mathbf{Y}_{p,L}$ and $\mathbf{Y}_{f,L}$ (see e.g. [15]). We remark that matrices $\mathbf{Y}_{p,L}$ and $\mathbf{Y}_{f,L}$ depend on the non available samples $\mathbf{y}_n, N+1 \leq n \leq N+2L-1$. As these end effects have no implication in the following, we prefer to use the definitions (1, 2) in order to simplify the notations. The above estimation procedure produces reasonably accurate results when the ratio $c_N = ML/N$ is small enough. However, if $\mathbf{y}$ is high-dimensional, i.e. if M is large, the condition $c_N << 1$ will not be verified as soon as the number of observations is not unlimited. It is therefore important to evaluate the behaviour of the above estimators when c_N is not negligible. In this paper, we address this problem by studying the behaviour of the above estimators in the high-dimensional regime where L is a fixed integer and where M and N both converge towards infinity in such a way that the ratio $c_N = ML/N$ converges towards a non zero constant $c < 1$. As this problem appears difficult in general contexts, we specifically consider the simple case where $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ is an uncorrelated complex Gaussian time series, i.e. $\mathbb{E}(\mathbf{y}_n\mathbf{y}_m^*) = \mathbf{R}\delta_{n-m}$ for some positive definite matrix $\mathbf{R}$. In this context, the true canonical correlation coefficients between the past and the future of $\mathbf{y}$ are of course all equal to 0, i.e. $r = 0$. Using large random matrix methods, more specifically free probability technics, we show that the

Authors supported by Bézout Labex, funded by ANR, reference ANR-10-LABX-58, and by the ANR Project HIDITSA, reference ANR-17-CE40-0003.

estimated canonical correlation coefficients have a limit deterministic distribution that is given in closed form. In practice, this means that for each realization of sequence $\mathbf{y}_1, \ldots, \mathbf{y}_N$, the histogram of the estimated coefficients is close from the graph of the probability density of the above mentioned limit distribution. If $c \leq 1/2$, the limit distribution is absolutely continuous and its support is the interval $[0, 2\sqrt{c(1-c)}]$, and if $c > 1/2$, a Dirac mass at point 1 appears. While this new result is valid in the specific case of a white noise sequence $\mathbf{y}$, we believe it is useful for the following reasons: First, when the observation is the sum of a white noise with a useful signal with a low rank rational spectral density, the use of perturbation technics (see e.g. [3] in the context of simpler random matrix models) should allow to derive the conditions under which the largest estimated canonical correlation coefficients escape from the interval $[0, 2\sqrt{c(1-c)}]$, thus revealing the presence of the useful signal. Second, our results provide various tests, consistent in the high-dimensional context, to verify that the samples $(\mathbf{y}_n)_{n=1,\ldots,N}$ come from an uncorrelated sequence or not.

We finally mention that a number of previous works addressed the behaviour of canonical correlation coefficients in the high-dimensional case. However, the underlying random matrix models are simpler than in the present paper. More specifically, the random matrices $\mathbf{Y}_{p,L}$ and $\mathbf{Y}_{f,L}$ defined by (1, 2) are replaced by independent matrices $\mathbf{Y}_1$ and $\mathbf{Y}_2$ with i.i.d. elements, a property that is not verified by $\mathbf{Y}_{p,L}$ and $\mathbf{Y}_{f,L}$. In 1980, [17] addressed the case of Gaussian i.i.d. entries and derived the limit distribution of the canonical correlation coefficients between the row spaces of $\mathbf{Y}_1$ and $\mathbf{Y}_2$. More recently, [18] extended this result to the case where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are independent matrices with non Gaussian i.i.d. entries. We also note that [19] took benefit of this result to propose independence tests between 2 sets of i.i.d. high-dimensional samples, a question which is not the same than the derivation of high-dimensional whiteness tests. We finally mention that [2] extended the result of [17] to the case where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ have Gaussian i.i.d. entries, but $\mathbb{E}\frac{\mathbf{Y}_1\mathbf{Y}_2^*}{N}$ is a non zero low rank matrix.

2. THE MAIN RESULT

For each n, $\mathbf{y}_n$ can be written as $\mathbf{y}_n = \mathbf{R}^{1/2}\mathbf{y}_{n,iid}$ where $(\mathbf{y}_{n,iid})_{n\in\mathbb{Z}}$ is an i.i.d. sequence of $\mathcal{N}_c(0,\mathbf{I})$ distributed random vectors. It is clear that the row spaces of $\mathbf{Y}_{p,L}$ and $\mathbf{Y}_{f,L}$ coincide with the row spaces of the block Hankel matrices $\mathbf{Y}_{p,L,iid}$ and $\mathbf{Y}_{f,L,iid}$ defined from vectors $(\mathbf{y}_{n,iid})_{n=1,\ldots,N+2L-1}$. Therefore, the correlations coefficients between the 2 pairs of subspaces coincide, and there is no restriction to assume that $\mathbf{R} = \mathbf{I}$ in the following. From now on, we thus assume that $(\mathbf{y}_n)_{n\geq 1}$ is an independent sequence of $\mathcal{N}_c(0,\mathbf{I})$ distributed random vectors. In order to simplify the notations, we denote by $\mathbf{W}, \mathbf{W}_p, \mathbf{W}_f$ the matrices defined by $\mathbf{W} = \frac{1}{\sqrt{N}}(\mathbf{y}_1, \ldots, \mathbf{y}_N)$, $\mathbf{W}_p = \frac{1}{\sqrt{N}}\mathbf{Y}_{p,L}$ and $\mathbf{W}_f = \frac{1}{\sqrt{N}}\mathbf{Y}_{f,L}$. The estimated canonical correlation coefficients therefore coincide with the singular values of matrix $\mathbf{\Sigma} = (\mathbf{W}_f\mathbf{W}_f^*)^{-1/2}\mathbf{W}_f\mathbf{W}_p^*(\mathbf{W}_p\mathbf{W}_p^*)^{-1/2}$ because the rows of $(\mathbf{W}_f\mathbf{W}_f^*)^{-1/2}\mathbf{W}_f$ and $(\mathbf{W}_p\mathbf{W}_p^*)^{-1/2}\mathbf{W}_p$ represent orthonormal bases. In the following, we rather study the singular values to the square, or equivalently the eigenvalues of the $ML \times ML$ matrix $\mathbf{\Sigma\Sigma}^*$, in the asymptotic regime where $c_N = ML/N$ converges towards $c < 1$, L being supposed to remain fixed. This regime will be referred to as $N \to +\infty$ in order to simplify the notations. In the following, we denote by $(\lambda_k)_{k=1,\ldots,ML}$ the eigenvalues of $\mathbf{\Sigma\Sigma}^*$. The main result of this paper is the following Theorem.

Theorem 1 *The empirical eigenvalue distribution $d\hat{\mu}(\lambda)$ of $\mathbf{\Sigma\Sigma}^*$ defined as $d\hat{\mu}(\lambda) = \frac{1}{ML}\sum_{k=1}^{ML}\delta_{\lambda-\lambda_k}$ converges weakly almost surely towards the probability distribution $d\mu(\lambda)$ given by*

$$d\mu(\lambda) = \frac{1}{c}\frac{\sqrt{\lambda\,(4c(1-c)-\lambda)}}{2\pi\lambda(1-\lambda)}\mathbf{1}_{[0,4c(1-c)]}\,d\lambda + \max(2-\frac{1}{c}, 0)\delta_{\lambda-1} \tag{3}$$

Interestingly, measure μ coincides with the limit eigenvalue distribution derived in [17] when matrices $\mathbf{Y}_{p,L}, \mathbf{Y}_{f,L}$ are replaced by 2 independent $ML \times N$ random matrices with $\mathcal{N}_c(0,1)$ i.i.d. entries. We note that a similar phenomenon holds for the limit eigenvalue distribution of $\frac{\mathbf{Y}_i\mathbf{Y}_i^*}{N}$, which, for $i = p, f$, converges towards the Marcenko-Pastur distribution, i.e. the limit eigenvalue distribution of the above matrices if $\mathbf{Y}_{p,L}, \mathbf{Y}_{f,L}$ were replaced by a $ML \times N$ random matrix with $\mathcal{N}_c(0,1)$ i.i.d. entries (see e.g. [11]). We also remark that the support of μ is equal $[0, 4c(1-c)] \cup \{1\}\mathbf{1}_{c>1/2}$. Therefore, while the true canonical correlation coefficients between the finite dimensional past and future $\mathcal{Y}_{p,L}$ and $\mathcal{Y}_{p,L}$ are all zero, the high-dimensionality of the observation produces a spreading of the distribution of the estimated coefficients. We however remark that the density of μ converges towards $+\infty$ when $\lambda \to 0$. In practice, this means that a number of eigenvalues of the matrix $\mathbf{\Sigma\Sigma}^*$ are concentrated around 0. We also notice that if $c > 1/2$, a mass at $\lambda = 1$ appears. If $c > 1/2$, $c_N = ML/N$ is also strictly larger than $1/2$ for M and N large enough, and the intersection of the row spaces of $\mathbf{Y}_{p,L}$ and $\mathbf{Y}_{f,L}$ is a non zero subspace whose dimension is at least equal to $2ML - N = N(2c_N - 1)$. Therefore, 1 is eigenvalue of matrix $\Sigma\Sigma^*$ with multiplicity at least equal to $N(2c_N - 1)$. This in accordance with Theorem 1.

Fig. 1 illustrates Theorem 1. In the corresponding numerical experiment, $N = 1200, M = 75, L = 4$ so that $c_N = \frac{1}{4}$. The histogram of the eigenvalues of a realization of matrix $\mathbf{\Sigma\Sigma}^*$ is represented as well the graph of the corresponding limit probability density. As expected, the 2 plots are close one from each other, and a number of eigenvalues are close from 0.

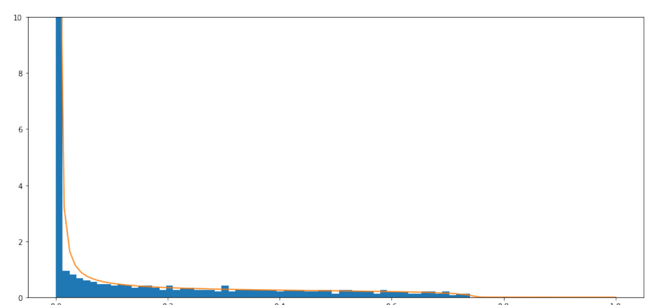

Fig. 1. Histogram of the eigenvalues, $N = 1200, M = 75, L = 4, c = \frac{1}{4}$

Due to the lack of space, we of course cannot provide the proof of Theorem 1. However, we present in the next section a sketch of the main arguments.

3. SKETCH OF PROOF

3.1. Background on free probability theory.

The proof uses free probability technics. We thus provide some background on the corresponding theory in order to make this paper reasonnably self-contained for non expert readers. Due to the lack of space, some concepts are presented in a rather unformal way. We refer the reader to the short summary provided in [4] (see section V) in which the necessary concepts are introduced rigourously. Section 2.4 in [14] is also recommended. For a deeper view of the theory, we refer to [6] and [12]. A non commutative probability space is a couple $(\mathcal{A}, \phi)$ where $\mathcal{A}$ is a non commutative algebra having a unit denoted $\mathbf{1}$ and ϕ is a linear functional such that $\phi(\mathbf{1}) = 1$. We also assume that ϕ verifies $\phi(ab) = \phi(ba)$. An element $a \in \mathcal{A}$ is called a non commutative random variable, and the distribution of a is defined as the linear functional ρ_a defined on the algebra of complex polynomials in 1 variable $\mathbb{C}(X)$ by $\rho_a(P) = \phi(P(a))$. For each integer $k \geq 1$, $\phi(a^k)$ is called the order k moment of ρ_a. In a number of useful contexts, ρ_a is associated to a probability measure μ_a defined on $\mathbb{R}$ by $\phi(a^k) = \int_{\mathbb{R}} \lambda^k d\mu_a(\lambda)$ for each $k \geq 1$. If $a_1, \ldots, a_p$ are p elements of $\mathcal{A}$, the joint distribution of $a_1, \ldots, a_p$ is this time the linear functional defined on the algebra of complex polynomials in p variables by $\rho_a(X_{i_1} \ldots X_{i_q}) = \phi(a_{i_1} \ldots a_{i_q})$ where the indices $i_1, i_2, \ldots, i_q$ belong to $\{1, 2, \ldots, p\}$.

A typical example of non commutative probability space is $(\mathcal{H}_P, \phi_{tr})$ where $\mathcal{H}_P$ represents the set of all $P \times P$ Hermitian matrices and where $\phi_{tr}(\mathbf{H}) = \frac{1}{P}\mathrm{Tr}\mathbf{H}$ for each Hermitian matrix $\mathbf{H}$. In this context, the distribution of $\mathbf{H}$ is associated to the probability distribution $\mu_{\mathbf{H}} = \frac{1}{P}\sum_{k=1}^{P} \delta_{\lambda - \lambda_k(\mathbf{H})}$ where $(\lambda_k(\mathbf{H}))_{k=1,\ldots,P}$ represent the eigenvalues of $\mathbf{H}$. In other words, $\mu_{\mathbf{H}}$ coincides with the empirical eigenvalue distribution of $\mathbf{H}$.

A central notion of the theory is the concept of freeness, which, in some sense, plays the role of the independence in classical probability theory. As the formal definition is may be not very informative, we omit to introduce it, and rather mention that if 2 elements a_1 and a_2 are free, then, the joint distribution of (a_1, a_2) can be retrieved from the individual distributions of a_1 and a_2. In this case, if ρ_{a_1} and ρ_{a_2} are associated to 2 compactly supported probability distributions μ_{a_1} and μ_{a_2}, then the distribution of $a_1 + a_2$ is associated to a certain probability measure $\mu_{a_1} \boxplus \mu_{a_2}$ called the free additive convolution product of μ_{a_1} and μ_{a_2}. If moreover μ_{a_1} and μ_{a_2} are carried by $\mathbb{R}^+$, the probability distribution associated to $a_1 a_2$ is a certain probability measure $\mu_{a_1} \boxtimes \mu_{a_2}$ called the free multiplicative convolution product of μ_{a_1} and μ_{a_2}. These convolution products can in practive be evaluated using some relevant analytic tools. It is also possible to define the freeness of 2 sets of non commutative random variables as well as the mutual freeness of $a_1, \ldots, a_p$.

The connection between free probability theory and large random matrices is based on the observation that certain mutually independent large random Hermitian $P \times P$ matrices, considered as elements of the non commutative probability space $(\mathcal{H}_P, \phi_{tr})$, behave almost surely as free non commutative random variables when their dimension converges towards $+\infty$. In this case, the corresponding random matrices are said to be asymptotically free almost surely. We refer the reader to [6], p. 147, for a formal definition. In particular, if $(\mathbf{H}_{1,P})_{P\geq 1}$ and $(\mathbf{H}_{2,P})_{P\geq 1}$ are two sequences of independent Hermitian unitarily invariant [1] $P \times P$ random matrices whose empirical eigenvalue distributions converge almost surely towards 2 compactly supported probability measures μ_1 and μ_2, then $(\mathbf{H}_{1,P})_{P\geq 1}$ and $(\mathbf{H}_{2,P})_{P\geq 1}$ are asymptotically free almost everywhere. Therefore, $\mathbf{H}_1 + \mathbf{H}_2$ has a limit eigenvalue distribution equal to $\mu_1 \boxplus \mu_2$. If moreover $\mathbf{H}_1$ and $\mathbf{H}_2$ are positive matrices, μ_1 and μ_2 are carried by $\mathbb{R}^+$, and the limit eigenvalue distribution of $\mathbf{H}_1\mathbf{H}_2$ is $\mu_1 \boxtimes \mu_2$ (see also [13] and [16] for direct approaches that do not use free probability theory). Another useful result states that if $(\mathbf{H}_P)_{P\geq 1}$ is a sequence of Hermitian unitarily invariant $P \times P$ random matrices whose empirical eigenvalue distribution converges almost surely, and if $(\mathbf{D}_{i,P})_{i=1,\ldots,I}$ are I $P \times P$ deterministic matrices whose joint distribution defined in $(\mathcal{H}_P, \phi_{tr})$ is convergent (i.e. the joint moments of $(\mathbf{D}_{i,P})_{i=1,\ldots,I}$ converge when $P \to +\infty$), then, $(\mathbf{H}_P)_{P\geq 1}$ and $((\mathbf{D}_{i,P})_{i=1,\ldots,I})_{P\geq 1}$ are almost surely asymptotically free. Both results are consequences of Corollary 4.3.6, p. 156 in [6].

[1] in the sense that for each unitary matrix $\mathbf{U}$, the probability distributions of $\mathbf{H}_i$ and $\mathbf{U}^*\mathbf{H}_i\mathbf{U}$ coincide

3.2. The main steps of the proof of the Theorem.

The first step consists in remarking that if $\mathbf{W}_p$ and $\mathbf{W}_f$ are replaced by finite rank perturbations, matrix $\mathbf{\Sigma\Sigma}^*$ will be affected by a finite rank perturbation which has no influence of its limit eigenvalue distribution. We therefore modify $\mathbf{W}_p$ and $\mathbf{W}_f$ by replacing samples $\mathbf{y}_{N+1}, \mathbf{y}_{N+2}, \ldots, \mathbf{y}_{N+2L-1}$ by the samples $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{2L-1}$. As L remains finite, this modification induces a finite rank perturbation of both matrices. In order to simplify the notations, the new matrices will be still denoted $\mathbf{W}_p$, $\mathbf{W}_f$, and $\mathbf{\Sigma}$. We denote by $\mathbf{\Pi}$ the $N \times N$ permutation matrix defined by $\mathbf{\Pi e}_n = \mathbf{e}_{n+1}$ for $n = 1, \ldots, N-1$ and $\mathbf{\Pi e}_N = \mathbf{e}_1$ where $\mathbf{e}_1, \ldots, \mathbf{e}_N$ represents the canonical basis of $\mathbb{C}^N$. Then, it is easily seen that the new matrices $\mathbf{W}_p$ and $\mathbf{W}_f$ are given by $\mathbf{W}_p = (\mathbf{W}^T, (\mathbf{W\Pi})^T, \ldots, (\mathbf{W\Pi}^{L-1})^T)^T$ and $\mathbf{W}_f = ((\mathbf{W\Pi}^L)^T, (\mathbf{W\Pi}^{L+1})^T, \ldots, (\mathbf{W\Pi}^{2L-1})^T)^T$, where we recall that $\mathbf{W} = \frac{1}{\sqrt{N}}(\mathbf{y}_1, \ldots, \mathbf{y}_N)$. We also introduce the $N \times ML$ orthogonal matrices $\mathbf{\Theta}_p$ and $\mathbf{\Theta}_f$ given by $\mathbf{\Theta}_i = \mathbf{W}_i^*(\mathbf{W}_i\mathbf{W}_i^*)^{-1/2}$ for $i = p, f$. Matrix $\mathbf{\Sigma\Sigma}^*$ coincides with $\mathbf{\Sigma\Sigma}^* = \mathbf{\Theta}_f^*\mathbf{\Theta}_p\mathbf{\Theta}_p^*\mathbf{\Theta}_f$

As for the second step, we notice that up to the eigenvalue 0, the $ML \times ML$ matrix $\mathbf{\Theta}_f^*\mathbf{\Theta}_p\mathbf{\Theta}_p^*\mathbf{\Theta}_f$ has the same eigenvalues than the $N \times N$ matrix $\mathbf{\Theta}_f\mathbf{\Theta}_f^*\mathbf{\Theta}_p\mathbf{\Theta}_p^*$. Therefore, it is sufficient to evaluate the limit eigenvalue distribution of $\mathbf{\Theta}_f\mathbf{\Theta}_f^*\mathbf{\Theta}_p\mathbf{\Theta}_p^*$. Matrices $\mathbf{\Theta}_f\mathbf{\Theta}_f^*$ and $\mathbf{\Theta}_p\mathbf{\Theta}_p^*$ are the orthogonal projection matrices on the row spaces of $\mathbf{W}_f$ and $\mathbf{W}_p$ respectively. Therefore, their empirical eigenvalue distribution both coincide with $c_N\delta_{\lambda-1} + (1-c_N)\delta_\lambda$, and converge towards the same limit $d\nu(\lambda) = c\delta_{\lambda-1} + (1-c)\delta_\lambda$. If $\mathbf{\Theta}_f\mathbf{\Theta}_f^*$ and $\mathbf{\Theta}_p\mathbf{\Theta}_p^*$ were almost surely asymptotically free, the limit eigenvalue distribution of $\mathbf{\Theta}_f\mathbf{\Theta}_f^*\mathbf{\Theta}_p\mathbf{\Theta}_p^*$ would be equal to $\nu \boxtimes \nu$, and easy calculations would imply that the limit distribution of $\mathbf{\Theta}_f^*\mathbf{\Theta}_p\mathbf{\Theta}_p^*\mathbf{\Theta}_f$ is the measure μ defined by (3).

Theorem 1 will thus be proved if we establish that $\mathbf{\Theta}_f\mathbf{\Theta}_f^*$ and $\mathbf{\Theta}_p\mathbf{\Theta}_p^*$ are almost surely asymptotically free. This is the third step of the proof. In order to establish this fundamental property, we state the following adaptation of Lemma 6 in [5].

Lemma 1 *We consider a sequence of $N \times N$ hermitian random matrices $(\mathbf{X}^N)_{N\geq 1}$ and $N \times N$ deterministic matrices $(\mathbf{U}_l^N, \mathbf{V}_l^N)_{l=1,\ldots m}$ such that $\mathbf{X}^N$ and $\{(\mathbf{U}_l^N, \mathbf{V}_l^N)_{l=1,\ldots m}\}$ are almost surely asymptotically free. Then, if $(\mathbf{U}_l^N, \mathbf{V}_l^N)_{l=1,\ldots m}$ satisfy $\mathbf{U}_l^N\mathbf{V}_l^N = \mathbf{V}_l^N\mathbf{U}_l^N = \mathbf{I}_N$ for each $l = 1, \ldots, m$ as well as $\frac{1}{N}\mathrm{Tr}(\mathbf{U}_k^N\mathbf{V}_l^N) = \delta_{k-l}$ for all $k, l = 1 \ldots m$, then the random matrices $\mathbf{U}_1^N\mathbf{X}^N\mathbf{V}_1^N, \ldots, \mathbf{U}_m^N\mathbf{X}^N\mathbf{V}_m^N$ are almost surely asymptotically mutually free.*

We first verify that it is possible to apply Lemma 1 when $m = 2L$, $\mathbf{X} = \mathbf{W}^*\mathbf{W}$, and $\mathbf{U}_l = \mathbf{\Pi}^{*(l-1)}, \mathbf{V}_l = \mathbf{U}_l^{-1} = \mathbf{\Pi}^{l-1}$. It is clear that $\{(\mathbf{\Pi}^{*(l-1)}, \mathbf{\Pi}^{l-1})_{l=1,\ldots,2L}\}$ verify the conditions of the Lemma. Moreover, matrix $\mathbf{W}^*\mathbf{W}$ is unitarily invariant. Hence, as recalled in paragraph 3.1, $\mathbf{W}^*\mathbf{W}$ and $\{(\mathbf{\Pi}^{*(l-1)}, \mathbf{\Pi}^{l-1})_{l=1,\ldots,2L}\}$ are almost surely asymptotically free. Lemma 1 thus leads to the conclusion that $\mathbf{W}^*\mathbf{W}, \mathbf{\Pi}^*\mathbf{W}^*\mathbf{W\Pi}, \ldots, \mathbf{\Pi}^{*(2L-1)}\mathbf{W}^*\mathbf{W\Pi}^{2L-1}$

are almost surely asymptotically mutually free. This immediately implies that $\mathbf{W}_p^*\mathbf{W}_p = \sum_{l=0}^{L-1} \mathbf{\Pi}^{*l}\mathbf{W}^*\mathbf{W}\mathbf{\Pi}^l$ and $\mathbf{W}_f^*\mathbf{W}_f = \sum_{l=L}^{2L-1} \mathbf{\Pi}^{*l}\mathbf{W}^*\mathbf{W}\mathbf{\Pi}^l$ are also almost surely asymptotically free. In order to complete the proof, we mention that it is possible to show that the empirical eigenvalue distributions of $\mathbf{W}_p^*\mathbf{W}_p$ and $\mathbf{W}_f^*\mathbf{W}_f$ converge a.s. towards $c\mu_{MP} + (1-c)\delta_\lambda$ where μ_{MP} is the Marcenko-Pastur distribution with parameter c. Moreover, for each $\epsilon > 0$, almost surely, for each N large enough, all the non zero eigenvalues of $\mathbf{W}_p^*\mathbf{W}_p$ and $\mathbf{W}_f^*\mathbf{W}_f$ are located into $[(1-\sqrt{c})^2 - \epsilon, (1+\sqrt{c})^2 + \epsilon]$. As $c < 1$, we choose ϵ in such a way that $(1-\sqrt{c})^2 - 2\epsilon > 0$ and consider a smooth function $f(\lambda)$ which is equal to 1 on $[(1-\sqrt{c})^2 - \epsilon, (1+\sqrt{c})^2 + \epsilon]$ and which vanishes outside $[(1-\sqrt{c})^2 - 2\epsilon, (1+\sqrt{c})^2 + 2\epsilon]$. Then, we claim that almost surely, $\mathbf{\Theta}_i\mathbf{\Theta}_i^* = f(\mathbf{W}_i^*\mathbf{W}_i)$ for $i = p, f$ for each N large enough. To check this, we express $\mathbf{W}_i^*\mathbf{W}_i$, as $\mathbf{W}_i^*\mathbf{W}_i = \sum_{k=1}^{ML} \gamma_{i,k}\boldsymbol{\theta}_{i,k}\boldsymbol{\theta}_{i,k}^*$ where the $(\gamma_{i,k}, \theta_{i,k})_{k=1,\ldots,ML}$ are the non zero eigenvalues and eigenvectors of $\mathbf{W}_i^*\mathbf{W}_i$. Almost surely, for N large enough, $\gamma_{i,k}$ belongs to $[(1-\sqrt{c})^2 - \epsilon, (1+\sqrt{c})^2 + \epsilon]$ and $f(\gamma_{i,k}) = 1$ for each $k = 1, \ldots, ML$. Therefore, $f(\mathbf{W}_i^*\mathbf{W}_i)$ is equal to $\sum_{k=1}^{ML} \boldsymbol{\theta}_{i,k}\boldsymbol{\theta}_{i,k}^*$, which, of course, coincides with $\mathbf{\Theta}_i\mathbf{\Theta}_i^*$. Approximating uniformly function f on $[0, (1+\sqrt{c})^2 + 2\epsilon]$ by a sequence of polynomials, and using the a.s. asymptotic freeness of $\mathbf{W}_p^*\mathbf{W}_p$ and $\mathbf{W}_f^*\mathbf{W}_f$ eventually lead to the conclusion that $\mathbf{\Theta}_f\mathbf{\Theta}_f^*$ and $\mathbf{\Theta}_p\mathbf{\Theta}_p^*$ are almost surely asymptotically free.

4. APPLICATION TO UNCORRELATEDNESS TESTING

In order to test that the high-dimensional time series $\mathbf{y}$ verifies $\mathbb{E}(\mathbf{y}_{n+l}\mathbf{y}_n^*) = 0$ for $l = 1, \ldots, 2L-1$, to be referred to as the hypothesis H_0, it is possible to compare the empirical eigenvalue distribution $\hat{\mu}$ of matrix $\mathbf{\Sigma}\mathbf{\Sigma}^*$ with its limit μ under H_0. If the 2 measures are close enough, the decision is that H_0 holds, and vice and versa. There are a number of ways to compare $\hat{\mu}$ and μ. It is possible to consider the Kolmogorov-Smirnov statistics $\eta = \sup_\lambda |\hat{F}(\lambda) - F(\lambda)|$ where $\hat{F}$ and F represent the cumulative distribution functions of $\hat{\mu}$ and μ, and to compare η to a threshold. Other kind of distance can also be considered, such as the Wasserstein distance. If f is a test function, another approach is to compare $\int f(\lambda)d\hat{\mu}(\lambda)$ to its limit $\int f(\lambda)d\mu(\lambda)$ (see e.g. [19] when $\mathbf{Y}_{p,L}$ and $\mathbf{Y}_{f,L}$ are replaced by independent i.i.d. non Gaussian matrices). In particular, it is easily seen that the first moment of μ, i.e. $\int \lambda d\mu(\lambda)$ coincides with c. Therefore, it is relevant to compare the test statistics $\gamma = \left|\frac{1}{ML}\mathrm{Tr}\mathbf{\Sigma}\mathbf{\Sigma}^* - c\right|$ to 0.

We illustrate the performance of the Kolmogorov-Smirnov (KS) test (Fig. 2) and of the test associated to the statistics γ (Fig. 3). We plot 3 ROC curves obtained by Monte-Carlo simulations. The signal generated under hypothesis H_1 is a M-dimensional signal $(\mathbf{z}_n)_{n\in\mathbb{Z}}$ whose M components $((\mathbf{z}_{m,n})_{n\in\mathbb{Z}})_{m=1,\ldots,M}$ are mutually independent autoregressive sequences of order 1 with coefficients $(a_m)_{m=1,\ldots,M}$ uniformly distributed between 0.35 and 0.45. The number of observations N is equal to $N = 1200$, the ratio $c_N = ML/N$ is equal to $\frac{1}{4}$, and the integer L take the values $L = 2$, $L = 4$, and $L = 8$. Therefore, M is equal to $M = 150$, $M = 75$ and $M = 37$ respectively. It is observed that the performance of the tests depends on L, or equivalently of M: the larger M, the better the performance. This is because under H_0, the $ML \times ML$ random matrices $\mathbf{W}_f$ and $\mathbf{W}_p$ depend on MN independent scalar random variables. Intuitively, the convergence towards 0 of η and γ depends on MN, so that the observed loss of performance when L increases was expected. We also notice the test statistics γ provides better results than the KS test. While we have not yet evaluated the asymptotic distribution of η and γ under H_0, the better performance of γ is probably due to the fact that γ converges faster towards 0 than η: in the context of simpler models, statistics such as γ and η are $\mathcal{O}_P(\frac{1}{N})$ and $\mathcal{O}_P(\frac{1}{N^{2/5}})$ terms respectively (see e.g. [1] and the references therein). We finally remark that the ROC curves of course also depend on the $(a_m)_{m=1,\ldots,M}$ which control the speed of convergence towards 0 of the autocovariance sequences of the components of $(\mathbf{z}_n)_{n\in\mathbb{Z}}$.

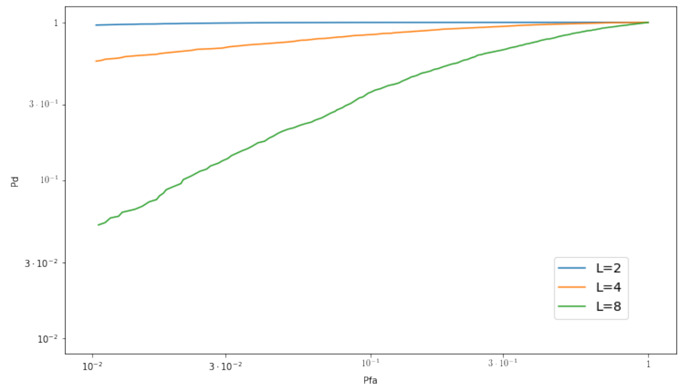

Fig. 2. ROC curves for the KS test

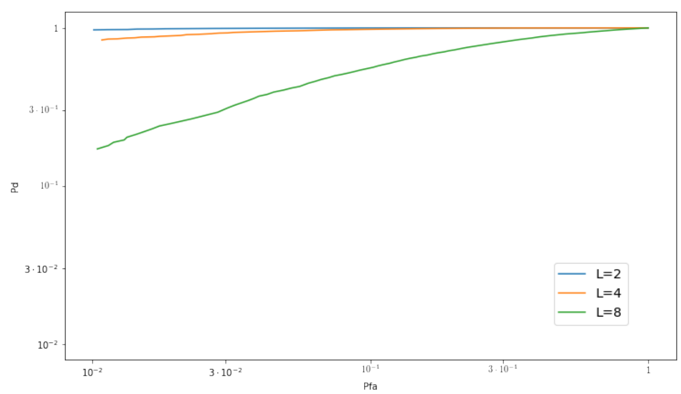

Fig. 3. ROC curves for the test statistics γ

5. CONCLUDING REMARKS.

We conclude by indicating some directions for future research. An interesting problem would be to study the largest canonical correlation coefficients when the observation contains a "useful signal" having low dimensional state-space representations, and to evaluate the conditions under which some coefficients escape from the interval $[0, 2\sqrt{c(1-c)}]$. This kind of result could be used to detect such a signal. Finally, the applications to uncorrelatedness testing, briefly mentioned in the present paper, pose a number of open questions: among others, consider the case where $L \to +\infty$ in order to be able to test that the observation is an i.i.d. sequence, establish some CLT on the above mentioned test statistics, make the appropriate connections with traditional tests used in standard asymptotic regimes such as the multivariate Portmanteau tests (see e.g. [7]).

6. REFERENCES

[1] Z.D. Bai, J.W. Silverstein, "Spectral Analysis of Large Dimensional Random Matrices", Second Edition, Springer Series in Statistics, 2010.

[2] Z. Bao, J. Hu, G. Pan, W. Zhou, "Canonical correlation coefficients of high-dimensional Gaussian vectors: Finite rank case", Ann. Statis., vol. 47, no. 1, pp. 612-640, 2019.

[3] F. Benaych-Georges, R.R. Nadakuditi,"The singular values and vectors of low rank perturbations of large rectangular random matrices", J. Multivariate Anal., vol. 111 (2012), pp. 120-135.

[4] M. Debbah, W. Hachem, Ph. Loubaton, M. de Courville, "MMSE analysis of certain large isometric random precoded systems", IEEE Trans. on Information Theory, vol. 49, no. 5, pp. 1293-1311, May 2003.

[5] J. Evans, D. N. C. Tse, Large System Performance of Linear Multiuser Receivers in Multipath Fading Channels, IEEE Transactions on Information Theory, Vol. 46, No. 6, 2000.

[6] F. Hiai, D. Petz, "The Semi-Circle Law, Free Random Variables and Entropy", Math. Surveys and Monographs, vol. 77, American Mathematical Society, 2000.

[7] J.R.M. Hoskins, "The multivariate Portmanteau test", J. of the American Statis. Asso., vol. 75, no. 371, Sept. 1980, pp. 602-608.

[8] H. Hotelling, "Relations between two sets of variables", Biometrika 28, 3/4 (1936), pp. 321-377.

[9] N.P. Jewell, P. Bloomfield, "Canonical correlations of past and future for time series: definitions and theory", Annals of Stat., vol. 11, no. 3, pp.837-847, 1983.

[10] A. Lindquist, G. Picci, “Linear Stochastic Systems”, Series in Contemporary Mathematics, Vol. 1, Springer, 2015

[11] P. Loubaton, "On the almost sure location of the singular values of certain Gaussian block-Hankel large random matrices", J. of Theoretical Probability, vol. 29, no. 4, pp. 1339-1443, December 2016

[12] J. Mingo, R. Speicher, "Free Probability and Random Matrices", Field Institute Monograph 35, American Mathematical Society, 2017.

[13] L. Pastur, V. Vasilchuk, "On the law of addition of random matrices", Comm. Math. Phys. 214 (2000), 249286.

[14] A. Tulino, S. Verdu, "Random Matrix Theory and Wireless Communications", Foundations and Trends in Communications and Information Theory: Vol. 1: No. 1, pp 1-182.

[15] P. Van Overschee, B. de Moor, "Subspace Identification for Linear Systems: Theory, Implementation, Applications", Kluwer Academic Publishers, 1996.

[16] V. Vasilchuk, "On the law of multiplication of random matrices", Mathematical Physics, Analysis and Geometry, March 2001, vol. 4, no. 1, pp 136.

[17] K. W. Wachter, "The limiting empirical measure of multiple discriminant ratios", Ann. Statist., vol. 8, no. 5, pp. 937-957, 1980.

[18] Y. R. Yang, G. M. Pan, "The convergence of the empirical distribution of canonical correlation coefficients", Electron. J. Probab., 17, 64, 2012.

[19] Y. Yang, G.M.Pan, "Independence test for high dimensional data based on regularized canonical correlation coefficients", Ann. Statis., 43(2), pp. 467500, 2015.