Free Deconvolution: from Theory to Practice

Florent Benaych-Georges, Mérouane Debbah, *Senior Member, IEEE*

***Abstract*— In this paper, we provide an algorithmic method to compute the singular values of sum of rectangular matrices based on the free cumulants approach and illustrate its application to wireless communications. We first recall the algorithms working for sum/products of square random matrices, which have already been presented in some previous papers and we then introduce the main contribution of this paper which provides a general method working for rectangular random matrices, based on the recent theoretical work of Benaych-Georges. In its full generality, the computation of the eigenvalues requires some sophisticated tools related to free probability and the explicit spectrum (eigenvalue distribution) of the matrices can hardly be obtained (except for some trivial cases). From an implementation perspective, this has led the community to the misconception that free probability has no practical application. This contribution takes the opposite view and shows how the free cumulants approach in free probability provides the right shift from theory to practice.**

***Index Terms*— Free Probability Theory, Random Matrices, Rectangular Free Convolution, Deconvolution, Free Cumulants, Wireless Communications.**

I. Introduction

A. General introduction

A question that naturally arises in cognitive random networks [1] is the following: "From a set of p noisy measurements, what can an intelligent device with n dimensions (time, frequency or space) infer on the rate in the network?". It turns that these questions have recently found answers in the realm of free deconvolution [2], [3]. Cognitive Random Networks have been recently advocated as the next big evolution of wireless networks. The general framework is to design self-organizing secure networks where terminals and base stations interact through cognitive sensing capabilities. The mobility in these systems require some sophisticated tools based on free probability to process the signals on windows of observations of the same order as the dimensions (number of antennas, frequency band, number of chips) of the system. Free probability theory [4] is not a new tool but has grown into an entire field of research since the pioneering work of Voiculescu in the 1980's ([5], [6], [7], [8]). However, the basic definitions of free probability are quite abstract and this has hinged a burden on its actual practical use. The original goal was to introduce an analogy to independence in classical probability that can be used for non-commutative random variables like matrices. These more general random variables are elements of what is called a *noncommutative probability space*, which we do not introduce as our aim is to provide a more practical approach to these methods. Based on the moment/cumulant approach, the free probability framework has been quite successfully applied recently in the works [2], [3] to infer on the eigenvalues of very simple models i.e the case where one of the considered matrices is unitarily invariant. This invariance has a special meaning in wireless networks and supposes that there is some kind of symmetry in the problem to be analyzed. In the present contribution, although focused on wireless communications, we show that the cumulant/moment approach can be extended to more general models and provide explicit algorithms to compute spectrums of matrices. In particular, we give an explicit relation between the spectrums of random matrices $(\mathbf{M}+\mathbf{N})(\mathbf{M}+\mathbf{N})^*$, $\mathbf{M}\mathbf{M}^*$ and $\mathbf{N}\mathbf{N}^*$, where $\mathbf{M}, \mathbf{N}$ are large rectangular independent random matrices, at least one of them having a distribution which is invariant under multiplication, on any side, by any othogonal matrix. This had already been done ([9], [10], [11]), but only in the case where $\mathbf{M}$ or $\mathbf{N}$ is Gaussian.

Florent Benaych-Georges: LPMA, UPMC Univ Paris 6, Case courier 188, 4, Place Jussieu, 75252 Paris Cedex 05, France and CMAP, École Polytechnique, Route de Saclay, Palaiseau Cedex, 91128, France. florent.benaych@gmail.com

Mérouane Debbah is with the Alcatel-Lucent Chair on Flexible Radio, Supélec, 3 rue Joliot-Curie 91192 GIF SUR YVETTE CEDEX France, merouane.debbah@supelec.fr

B. Organization of the paper, definitions and notations

In the following, upper (lower) boldface symbols will be used for matrices (column vectors) whereas roman symbols will represent scalar values, $(.)^*$ will denote hermitian transpose. $\mathbf{I}$ will represent the identity matrix. Tr denotes the *trace*.

The paper is organized as follows.

1) Section II: In section II, we introduce the moments approach for computing the eigenvalues of classical known matrices.

2) Sections III and IV: In these sections, we shall review some classical results of free probability and show how (as long as moments of the distributions are considered) one can, for $\mathbf{A}, \mathbf{B}$ independent large square Hermitian (or symmetric) random matrices (under some general hypothesis that will be specified):

- derive the eigenvalue distribution of $\mathbf{A}+\mathbf{B}$ from the ones of $\mathbf{A}$ and $\mathbf{B}$.
- derive the eigenvalue distribution of $\mathbf{AB}$ or of $\mathbf{A}^{\frac{1}{2}}\mathbf{B}\mathbf{A}^{\frac{1}{2}}$ from those of $\mathbf{A}$ and $\mathbf{B}$.

The framework of computing the eigenvalue of the sum/product of matrices is known in the literature as free convolution ([12]), and denoted respectively by $\boxplus, \boxtimes$.

We will also see how one can:

- Deduce the eigenvalue distribution of $\mathbf{A}$ from those of $\mathbf{A}+\mathbf{B}$ and $\mathbf{B}$.
- Deduce the eigenvalue distribution of $\mathbf{A}$ from those of $\mathbf{AB}$ or of $\mathbf{A}^{\frac{1}{2}}\mathbf{B}\mathbf{A}^{\frac{1}{2}}$ and $\mathbf{B}$.

These last operations are called free deconvolutions ([9]) and denoted respectively by $\boxminus, \boxslash$.

3) Section V: This section will be devoted to rectangular random matrices. We shall present how the theoretical results that the first named author proved in his thesis can be made practical in order to solve some of the network problems presented in Section **??**. The method presented here also uses the classical results of free probability mentioned above.

We consider the general case of two independent real rectangular random matrices $\mathbf{M}, \mathbf{N}$, both of size $n \times p$. We shall suppose that n, p tend to infinity in such a way that n/p tends to $\lambda \in [0, 1]$. We also suppose that at least one of these matrices has a joint distribution of the entries which is invariant by multiplication on any side by any orthogonal matrix. At last, we suppose that the *eigenvalue distributions* of $\mathbf{MM}^*$ and $\mathbf{NN}^*$ (i.e. the uniform distributions on their eigenvalues with multiplicity) both converge to non random probability measures. From a historical and purely mathematical perspective, people have focused on these types of random matrices because the invariance under actions of the orthogonal group is the – quite natural – notion of *isotropy*. The Gram[1] approach was mainly due to the fact that the eigenvalues of $\mathbf{MM}^*$ (which are real and positive) are easier to characterize than those of $\mathbf{M}$. From an engineering perspective, for a random network modeled by a matrix $\mathbf{M}$, the eigenvalues of $\mathbf{MM}^*$ contain in many cases the information needed to characterize the performance limits of the system. In fact, the eigenvalues relate mainly to the energy of the system. We shall explain how one can deduce, in a computational way, the limit eigenvalue distribution of $(\mathbf{M}+\mathbf{N})(\mathbf{M}+\mathbf{N})^*$ from the limit eigenvalue distributions of $\mathbf{MM}^*$ and $\mathbf{NN}^*$. The underlying operation on probability measures is called the *rectangular free convolution with ratio* λ, denoted by $\boxplus_\lambda$ in the literature ([13], [14], [15]). Our machinery will also allow the inverse operation, called *rectangular deconvolution with ratio* λ: the derivation of the eigenvalue distribution of $\mathbf{MM}^*$ from the ones of $(\mathbf{M}+\mathbf{N})(\mathbf{M}+\mathbf{N})^*$ and $\mathbf{NN}^*$.

4) Sections VII and VI: In section VII, we present some applications of the results of section V to the analysis of random networks and we compare them with other results, due to other approaches, in section VI.

II. Moments for single random matrices

A. Historical Perspective

The moment approach for the derivation of the eigenvalue distribution of random matrices dates back to the work of Wigner [16], [17]. Wigner was interested in deriving the energy levels of nuclei. It turns out that energy levels are linked to the Hamiltonian operator by the following Schröndinger equation:

$$\mathbf{H}\mathbf{\Phi}_i = \mathbf{E_i}\mathbf{\Phi}_i,$$

where $\mathbf{\Phi}_i$ is the wave function vector, $\mathbf{E_i}$ is the energy level. A system in quantum mechanics can be characterized by a self-adjoint linear operator in Hilbert space: its hamiltonian operator. We can think of this as a Hermitian matrix of a number of infinitely many dimensions, having somehow introduced a coordinate system in a Hilbert space. Hence, the energy levels of the operator $\mathbf{H}$ are nothing else but the eigenvalues of the matrix representation of that operator. For a specific nucleus, finding the exact eigenvalues is a very complex problem as the number of interacting particles increases. The genuine idea of Wigner was to replace the exact matrix by a random matrix having the same properties. Hence, in some cases, the matrix can be replaced by the following Hermitian random matrix where the upper diagonal elements are i.i.d. generated with a binomial distribution.

$$\mathbf{H} = \frac{1}{\sqrt{n}} \begin{bmatrix} 0 & +1 & +1 & +1 & -1 & -1 \\ +1 & 0 & -1 & +1 & +1 & +1 \\ +1 & -1 & 0 & +1 & +1 & +1 \\ +1 & +1 & +1 & 0 & +1 & +1 \\ -1 & +1 & +1 & +1 & 0 & -1 \\ -1 & +1 & +1 & +1 & -1 & 0 \end{bmatrix}$$

It turns out that, as the dimension of the matrix increases, the eigenvalues of the matrix become more and more predictable irrespective of the exact realization of the matrix. This striking result enabled to determine the energy levels of many nuclei without considering the very specific nature of the interactions. In the following, we will provide the different steps of the proof which are of interest for understanding the free moments approach.

B. The semi-circular law

The main idea is to compute, as the dimension increases, the trace of the matrix $\mathbf{H}$ at different exponents. Typically, let

$$dF_n(\lambda) = \frac{1}{n}\sum_{i=1}^{n} \delta\left(\lambda - \lambda_i\right).$$

Then the moments of the distribution are given by:

$$\begin{aligned} m_1^n &= \frac{1}{n}\operatorname{trace}\left(\mathbf{H}\right) = \frac{1}{n}\sum_{i=1}^{n}\lambda_i = \int x dF_n(x) \\ m_2^n &= \frac{1}{n}\operatorname{trace}\left(\mathbf{H}^2\right) = \int x^2 dF_n(x) \\ \vdots &= \vdots \\ m_k^n &= \frac{1}{n}\operatorname{trace}\left(\mathbf{H}^k\right) = \int x^k dF_n(x) \end{aligned}$$

Quite remarkably, as the dimension increases, the traces can be computed using combinatorial and non-crossing partitions techniques. All odd moments converge to zero, whereas all even moments converge to the Catalan numbers [18]:

$$\begin{aligned} \lim_{n\to\infty}\frac{1}{n}\operatorname{trace}(\mathbf{H}^{2k}) &= \int_{-2}^{2} x^{2k} f(x) dx \\ &= \frac{1}{k+1}C_k^{2k}. \end{aligned}$$

More importantly, the only distribution which has all its odd moments null and all its even moments equal to the Catalan numbers is known to be the semi-circular law provided by:

$$f(x) = \frac{1}{2\pi}\sqrt{4-x^2},$$

[1]For a matrix $\mathbf{M}$, $\mathbf{MM}^*$ is called the Gram matrix associated to $\mathbf{M}$.

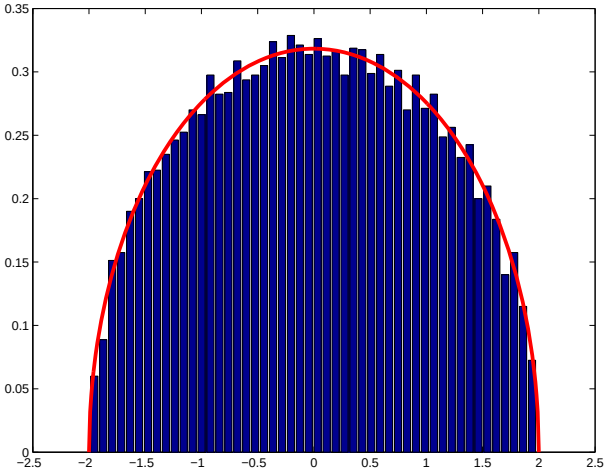

Fig. 1. Semicircle law and simulation for a 512×512 Wigner matrix.

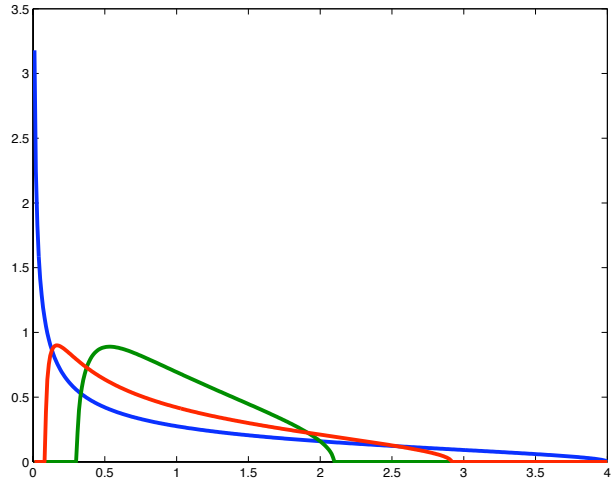

Fig. 2. Density function of (1) for $\lambda = 1, 0.5, 0.2$.

with $|x| \leq 2$. One can verify it directly by calculus based on recursion:

$$\begin{aligned}
\alpha_{2k} &= \frac{1}{\pi}\int_{-2}^{2} x^{2k}\sqrt{4-x^2}dx \\
&= -\frac{1}{2\pi}\int_{-2}^{2} \frac{-x}{\sqrt{4-x^2}} x^{2k-1}(4-x^2)dx \\
&= \frac{1}{2\pi}\int_{-2}^{2} \sqrt{4-x^2}(x^{2k-1}(4-x^2))'dx \\
&= 4(2k-1)\alpha_{2k-2} - (2k+1)\alpha_{2k}.
\end{aligned}$$

In this way, the recursion is obtained:

$$\alpha_{2k} = \frac{2(2k-1)}{k+1}\alpha_{2k-2}.$$

C. The Marchenko-Pastur law

Let us give another example to understand the moments approach for a single random matrix. Suppose that one is interested in the empirical eigenvalue distribution of $\mathbf{S}\mathbf{S}^H$ where $\mathbf{S}$ is an $n \times p$ random matrix which entries are independent centered gaussian random variables with entries of variance $\frac{1}{n}$ with $\frac{n}{p} \to \lambda$. In this case, in the same manner, the moments of this distribution are given by:

$$\begin{aligned}
m_1^n &= \frac{1}{n}\operatorname{trace}\left(\mathbf{S}\mathbf{S}^H\right) = \frac{1}{n}\sum_{i=1}^{n}\lambda_i \to \lambda \\
m_2^n &= \frac{1}{n}\operatorname{trace}\left(\mathbf{S}\mathbf{S}^H\right)^2 = \frac{1}{n}\sum_{i=1}^{N}\lambda_i^2 \to 1+\lambda \\
m_3^n &= \frac{1}{n}\operatorname{trace}\left(\mathbf{S}\mathbf{S}^H\right)^3 = \frac{1}{n}\sum_{i=1}^{n}\lambda_i^3 \to 1+3\lambda+\lambda^2
\end{aligned}$$

It turns out that the only distribution which has the same moments is known to be (a dilation of) the *Marchenko-Pastur law*.

Definition 2.1: (**Marchenko-Pastur Law** (see, e.g. [19])) The eigenvalue distribution of $\mathbf{S}\mathbf{S}^*$ tends to the law μ_λ with density

$$\frac{\sqrt{4\lambda-(x-1-\lambda)^2}}{2\pi\lambda x} \quad \text{on} \quad [(1-\sqrt{\lambda})^2, (1+\sqrt{\lambda})^2]. \quad (1)$$

This law it is the law of λ times a random variable distributed according to the so-called *Marchenko-Pastur law* with parameter $1/\lambda$.

Remark: In many cases, one would obviously think that the eigenvalues of

$$n \underbrace{\left\{\begin{bmatrix} \mathbf{S} \end{bmatrix}\right.}_{p} \begin{bmatrix} \mathbf{S}^H \end{bmatrix}$$

when $n \to \infty$, $n/p \to \lambda$ are equal to one. Indeed, asymptotically, all the diagonal elements tend to to one and the extra-diagonal elements tend to zero. However, the matrix is not the identity matrix. Indeed, there are $n^2 - n$ extra-diagonal terms which tend to zero at a rate of $O(\frac{1}{n^2})$. Therefore, the distance of the matrix to the identity matrix (in the Fröbenius norm sense) is not zero.

D. Conclusion

The moments technique is very appealing and powerful in order to derive the exact asymptotic moments. It requires combinatorial skills and can be used for a large class of random matrices. Recent works on random Vandermonde matrices [20], [21] and Euclidean matrices [22] have shown again its potential. The main drawback of the technique (compared to other tools such as the Stieljes transform method [11]) is that it can rarely provide the exact eigenvalue distribution. However, in many wireless applications, one needs only a subset of the moments depending on the number of parameters to be estimated.

III. Sum of two random matrices

A. Scalar case: $X + Y$

Let us consider two independent random variables X, Y and suppose that we know the distribution of $X + Y$ and Y and would like to infer on the distribution of X. The distribution of $X + Y$ is the convolution of the distribution of X with the distribution of Y. However, the expression is not straightforward to obtain. Another way of computing the spectrum is to form the moment generating functions

$$M_X(s) = \mathbb{E}(e^{sX}),\ M_{X+Y}(s) = \mathbb{E}(e^{s(X+Y)}).$$

It is then immediate to see that

$$M_X(s) = M_{X+Y}(s)/M_Y(s).$$

The distribution of X can be recovered from $M_X(s)$. This task is however not always easy to perform as the inversion formula does not provide an explicit expression. It is rather advantageous to express the independence in terms of moments of the distributions or even cumulants. We denote by C_k the cumulant of order k:

$$C_k(X) = \frac{\partial^n}{\partial t^n}_{|t=0} \log\left(\mathbb{E}\left(e^{tX}\right)\right).$$

They behave additively with respect to the convolution, i.e, for all $k \geq 0$,

$$C_k(X + Y) = C_k(X) + C_k(Y).$$

Moments and cumulants of a random variable can easily be deduce from each other by the formula

$$\forall n \geq 1, m_n(X) = \sum_{p=1}^{n} \sum_{\substack{k_1 \geq 1, \ldots, k_p \geq 1 \\ k_1 + \cdots + k_p = n}} C_{k_1}(X) \cdots C_{k_p}(X)$$

(recall that the moments of a random variable X are the numbers $m_n(X) = \mathbb{E}(X^n)$, $n \geq 1$).

Thus the derivation of the law of X from the laws of $X+Y$ and Y can be done by computing the cumulants of X by the formula $C_k(X) = C_k(X + Y) - C_k(X)$ and then deducing the moments of X from its cumulants.

B. Matrix case: additive free convolution $\boxplus$

1) Definition: It is has been proved by Voiculescu [12] that for $\mathbf{A}_n, \mathbf{B}_n$ free[2] large n by n Hermitian (or symmetric) random matrices (both of them having i.i.d entries, or one of them having a distribution which is invariant under conjugation by any orthogonal matrix), if the eigenvalue distributions of $\mathbf{A}_n, \mathbf{B}_n$ converge as n tends to infinity to some probability measures μ, ν, then the eigenvalue distribution of $\mathbf{A}_n + \mathbf{B}_n$ converges to a probability measure which depends only on μ, ν, which is called the *additive free convolution* of μ and ν, and which will be denoted by $\mu \boxtimes \nu$.

[2]The concept of freeness is different from independence.

2) Computation of $\mu \boxplus \nu$ by the moment/cumulants approach: Let us consider a probability measure ρ on the real line, which has moments of all orders. We shall denote its moments by $m_n(\rho) := \int t^n d\rho(t)$, $n \geq 1$. (Note that in the case where ρ is the eigenvalue distribution of a $d \times d$ matrix $\mathbf{A}$, these moments can easily be computed by the formula: $m_n(\rho) = \frac{1}{d}\operatorname{Tr}(\mathbf{A}^n)$). We shall associate to ρ another sequence of real numbers, $(\mathfrak{K}_n(\rho))_{n\geq 1}$, called its *free cumulants*. The sequences $(m_n(\rho))$ and $(\mathfrak{K}_n(\rho))$ can be deduced one from each other by the fact that the formal power series

$$K_\rho(z) = \sum_{n\geq 1} \mathfrak{K}_n(\rho) z^n \text{ and } M_\rho(z) = \sum_{n\geq 1} m_n(\rho) z^n \qquad (2)$$

are linked by the relation

$$K_\rho(z(M_\rho(z) + 1)) = M_\rho(z). \qquad (3)$$

Equivalently, for all $n \geq 1$, the sequences $(m_0(\rho), \ldots, m_n(\rho))$ and $(\mathfrak{K}_1(\rho), \ldots, \mathfrak{K}_n(\rho))$ can be deduced one from each other via the relations

$$\begin{aligned} m_0(\rho) &= 1 \\ m_n(\rho) &= \mathfrak{K}_n(\rho) + \sum_{k=1}^{n-1} \left[\mathfrak{K}_k(\rho) \sum_{\substack{l_1, \ldots, l_k \geq 0 \\ l_1 + \cdots + l_k = n-k}} m_{l_1}(\rho) \cdots m_{l_k}(\rho) \right] \end{aligned}$$

for all $n \geq 1$.

Example 3.1: As an example, it is known (e.g. [19]) that the law μ_λ of definition 2.1 has free cumulants $\mathfrak{K}_n(\mu_\lambda) = \lambda^{n-1}$ for all $n \geq 1$.

The additive free convolution can be computed easily with the free cumulants via the following characterization [23]:

Theorem 3.2: For μ, ν compactly supported, $\mu \boxplus \nu$ is the only law ρ such that for all $n \geq 1$,

$$\mathfrak{K}_n(\rho) = \mathfrak{K}_n(\mu) + \mathfrak{K}_n(\nu).$$

3) Simulations for finite size matrices: In Figure 3, we plot, for $k = 1, \ldots, 25$, the quantity

$$\left| \frac{m_k(\rho[\mathbf{A} + \mathbf{M}\mathbf{M}^*])}{m_k(\nu \boxplus \mu_\lambda)} - 1 \right| \qquad (4)$$

for $\mathbf{A}$ a diagonal random matrix with independent diagonal entries distributed according to $\nu = (\delta_0 + \delta_1)/2$ and $\mathbf{M}$ a gaussian $n \times p$ matrix as in definition 2.1. $\rho[\mathbf{X}]$ denotes the eigenvalue distribution of the matrix $\mathbf{X}$. The dimensions of $\mathbf{A}$ are 1500×1500, those of $\mathbf{M}$ are 1500×2000. Interestingly, the values of (4) provide a good match, which shows that the moments/free cumulants approach is a good way to compute the spectrum of sums of free random matrices.

C. Matrix case: additive free deconvolution $\boxminus$

1) Computation of $\boxminus$: The moments/cumulants method can also be useful to implement the free additive deconvolution. The *free additive deconvolution* of a measure ρ by a measure ν is (when it exists) the only measure μ such that $\rho = \mu \boxplus \nu$. In this case, μ is denoted by $\rho \boxminus \nu$. By Theorem 3.2, when it exists, $\rho \boxminus \nu$ is characterized by the fact that for all $n \geq 1$, $\mathfrak{K}_n(m \boxminus \nu) = \mathfrak{K}_n(\rho) - \mathfrak{K}_n(\nu)$. This operation is very useful in denoising applications [2], [3]

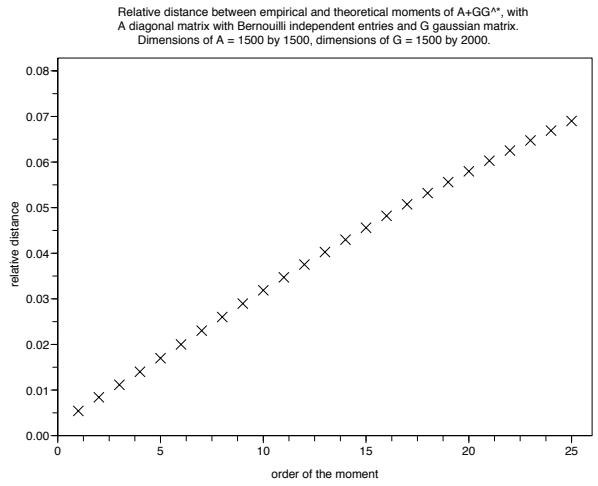

Fig. 3. Relative distance between empirical and theoretical moments of $\mathbf{A}+\mathbf{M}\mathbf{M}^*$, with $\mathbf{A}$ diagonal 1500 by 1500 matrix with Bernouilli independent diagonal entries and $\mathbf{M}$ gaussian 150 by 2000 matrix.

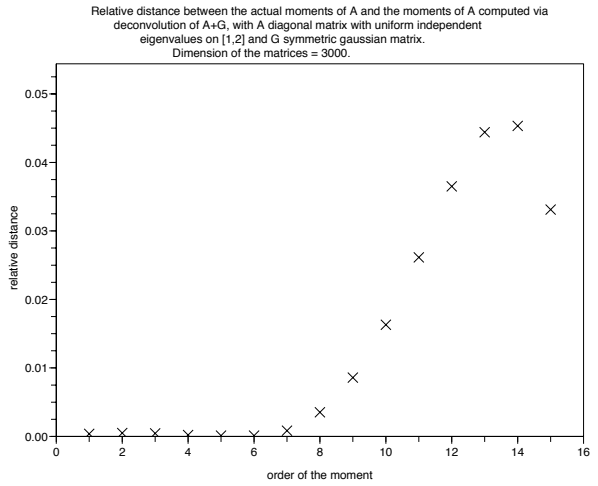

Fig. 4. Relative distance between the actual moments of $\mathbf{A}$ and the moments of $\mathbf{A}$ computed via deconvolution of $\mathbf{A}+\mathbf{B}$, with $\mathbf{A}$ diagonal 3000×3000 matrix with uniform independent diagonal entries and $\mathbf{B}$ gaussian symmetric 3000×3000 matrix.

2) Simulations for finite size matrices: In Figure 4, we plot for $n = 1, \ldots, 15$, the following values:

$$\left| \frac{m_n(\rho[(\mathbf{A}+\mathbf{B}) \boxminus \mathbf{B}])}{m_n(\rho[\mathbf{A}])} - 1 \right| \tag{5}$$

for $\mathbf{A}$ a diagonal random matrix which spectrum is chosen at random (the diagonal entries of $\mathbf{B}$ are independent and uniformly distributed on $[0, 1]$) and $\mathbf{B}$ a Gaussian symmetric matrix. The dimension of the matrices is 3000. Again, the fact that the values of (5) match even in the non-asymptotic case shows that the computational method, for deconvolution, is effecient.

IV. Product of two random matrices

A. Scalar case: XY

Suppose now that we are given two classical random variables X, Y, assumed to be independent. How do we find the distribution of X when only the distributions of XY and Y are given? The solution is quite straightforward since $\mathbb{E}((XY)^k) = \mathbb{E}(X^k)\mathbb{E}(Y^k)$, so that $\mathbb{E}(X^k) = \mathbb{E}((XY)^k)/\mathbb{E}(Y^k)$. Hence, using the moments approach, one has a simple algorithm to compute all the moments of the distribution. The case of matrices is rather involved and is explained in the following.

B. Matrix case: multiplicative free convolution $\boxtimes$

1) Definition: It is has been proved by Voiculescu [12] that for $\mathbf{A}_n, \mathbf{B}_n$ free large $n \times n$ positive Hermitian (or symmetric) random matrices (both of them having i.i.d entries, or one of them having a distribution which is invariant under multiplication by any orthogonal matrix), if the eigenvalue distributions of $\mathbf{A}_n, \mathbf{B}_n$ converge, as n tends to infinity, to some probability measures μ, ν, then the eigenvalue distribution of $\mathbf{A}_n\mathbf{B}_n$, which is equal to the eigenvalue distribution of $\mathbf{A}^{\frac{1}{2}}\mathbf{B}\mathbf{A}^{\frac{1}{2}}$ converges to a probability measure which depends only on μ, ν. The measure is called the *multiplicative free convolution* of μ and ν and will be denoted by $\mu \boxtimes \nu$.

2) Computation of $\mu \boxtimes \nu$ by the moment/cumulants approach: Let us consider a probability measure ρ on $[0, +\infty[$, which is not the Dirac mass at zero and which has moments of all order. We shall denote by $\left\{m_n(\rho) := \int t^n d\rho(t)\right\}_{n\geq 0}$ the sequence of its moments. We shall associate to ρ another sequence of real numbers, $\{\mathfrak{s}_n(\rho)\}_{n\geq 0}$, which are the coefficients of what is called its *S-transform*. The sequences $\{m_n(\rho)\}$ and $\{\mathfrak{s}_n(\rho)\}$ can be deduced one from each other by the fact that the formal power series

$$S_\rho(z) = \sum_{n\geq 1} \mathfrak{s}_n(\rho) z^{n-1} \text{ and } M_\rho(z) = \sum_{n\geq 1} m_n(\rho) z^n \tag{6}$$

are linked by the relation

$$M_\rho(z) S_\rho(M_\rho(z)) = z(1 + M_\rho(z)). \tag{7}$$

Equivalently, for all $n \geq 1$, the sequences $\{m_1(\rho), \ldots, m_n(\rho)\}$ and $\{\mathfrak{s}_1(\rho), \ldots, \mathfrak{s}_n(\rho)\}$ can be deduced one from each other via the relations

$$\begin{cases} m_1(\rho)\mathfrak{s}_1(\rho) = 1, \\ m_n(\rho) = \sum_{k=1}^{n+1} \mathfrak{s}_k(\rho) \sum\limits_{\substack{l_1,\ldots,l_k \geq 1 \\ l_1+\cdots+l_k = n+1}} m_{l_1}(\rho) \cdots m_{l_k}(\rho). \end{cases}$$

Remark Note that these equations allow computations which run faster than the ones already implemented (e.g. [10]), because those ones are based on the computation of the coefficients $\mathfrak{s}_n$ via non crossing partitions and the Kreweras complement, which use more machine time.

Example 4.1: As an example, it can be shown that for the law μ_λ of definition 2.1, that for all $n \geq 1$, $\mathfrak{s}_n(\mu_\lambda) = (-\lambda)^{n-1}$.

The multiplicative free convolution can be computed easily with the free cumulants via the following characterization [23].

Theorem 4.2: For μ, ν compactly supported probability measures on $[0, \infty[$, non of them being the Dirac mass at zero,

$\mu \boxtimes \nu$ is the only law ρ such that $S_\rho = S_\mu S_\nu$, i.e. such that for all $n \geq 1$,

$$\mathfrak{s}_n(\rho) = \sum_{\substack{k,l \geq 1 \\ k+l=n+1}} \mathfrak{s}_k(\mu)\mathfrak{s}_l(\nu).$$

The algorithm for the computation of the spectrum of the product of two random matrices following from this theorem is presented in paragraph IV-C.

C. Matrix case: The multiplicative free deconvolution

The moments/cumulants method can also be useful to implement the multiplicative free deconvolution. The *multiplicative free deconvolution* of a measure m by a measure ν is (when it exists) the only measure μ such that $\rho = \mu \boxtimes \nu$. In this case, μ is denoted by $m \boxslash \nu$. By theorem 4.2, when it exists, $m \boxslash \nu$ is characterized by the fact that for all $n \geq 1$,

$$\mathfrak{s}_n(\rho \boxslash \nu)\mathfrak{s}_1(\nu) = \mathfrak{s}_n(\rho) - \sum_{k=1}^{n-1} \mathfrak{s}_k(\rho \boxslash \nu)\mathfrak{s}_{n+1-k}(\nu).$$

1) Simulations for finite size matrices: In Figure 5, we illustrated the performance of the combinatorial methods

- to predict the spectrum of $\mathbf{AMM}^*$ from those of $\mathbf{A}$ and of $\mathbf{MM}^*$ (free multiplicative convolution)
- to recover the spectrum of $\mathbf{A}$ from those of $\mathbf{MM}^*$ and of $\mathbf{AMM}^*$ (free multiplicative deconvolution).

We simulated a random $n \times n$ diagonal matrix $\mathbf{A}$ which eigenvalue distribution is $\mu = (\delta_1 + \delta_{4/3} + \delta_{5/3} + \delta_2)/4$ and a random $n \times p$ Gaussian matrix $\mathbf{M}$ such that the eigenvalue distribution of $\mathbf{GG}^*$ is approximately the measure μ_λ of definition 2.1. Then on one hand, we compared the moments of $\mathbf{AGG}^*$ with their theoretical limit values obtained by free multiplicative convolution of μ and μ_λ: on the left graph, we plotted, for $n = 1, \ldots, 10$,

$$\left| \frac{m_n(\rho[\mathbf{AMM}^*])}{m_n[\mu \boxtimes \mu_\lambda]} - 1 \right|. \tag{8}$$

On the other hand, we compared the actual moments of $\mathbf{A}$ with their approximations obtained by free multiplicative deconvolution of the eigenvalue distribution of $\mathbf{AMM}^*$ and the limit measure of $\mathbf{MM}^*$ (which is the measure μ_λ of definition 2.1): on the right graphic, we plot, for $n = 1, \ldots, 10$,

$$\left| \frac{m_n(\rho[\mathbf{A}])}{m_n(\rho[\mathbf{AMM}^*]) \boxslash \mu_\lambda]} - 1 \right|. \tag{9}$$

Again, the fact that the values of (8) and (9) are very small for non-asymptotic values show that the computational methods, for convolution as well as for deconvolution, are efficient.

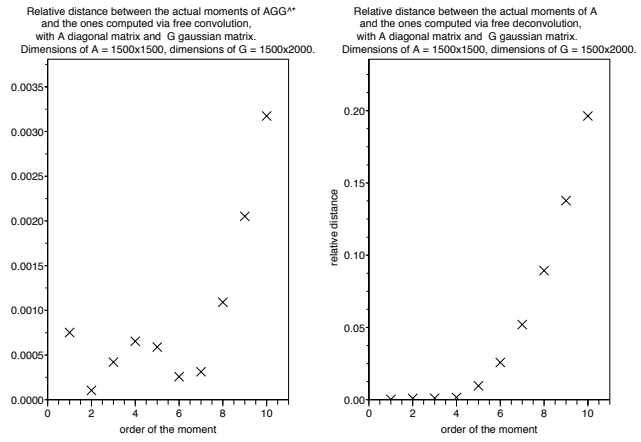

Fig. 5. *On the left:* relative distance between the actual moments of the eigenvalue distribution of $\mathbf{AMM}^*$ and the ones computed via free convolution. *On the right:* relative distance between the actual moments of the eigenvalue distribution of A and the ones computed via free deconvolution of $\mathbf{AMM}^*$ by the theoretical limit of $\mathbf{MM}^*$. *Dimensions of* $\mathbf{A}$*:* 1500×1500*, dimensions of* $\mathbf{M}$*:* 1500×2000.

V. Singular values of sums of rectangular matrices

A. Main result

In this section, we still consider two independent rectangular random matrices $\mathbf{M}, \mathbf{N}$, both having size $n \times p$. We shall suppose that n, p tend to infinity in such a way that n/p tends to $\lambda \in [0, 1]$. We also suppose that at least one of these matrices has a distribution which is invariant by multiplication on both sides by any orthogonal (or unitary, in the case where the matrices are not real but complex) matrix. At last, we suppose that the *eigenvalue distributions* of $\mathbf{MM}^*$ and $\mathbf{NN}^*$ both converge to non random probability measures. Here, we shall denote σ, τ the limit eigenvalue distributions of $\mathbf{MM}^*$ and $\mathbf{NN}^*$ respectively.

Note that in the previously presented results, the case of the limit eigenvalue distribution of $(\mathbf{M} + \mathbf{N})(\mathbf{M} + \mathbf{N})^*$ has not been treated. The reason is that these results rely on the work of Voiculescu, who "only" found a general way to compute the limit normalized trace of product of independent *square* random matrices with large dimension, which is sufficient to compute the moments of the eigenvalue distribution of either $\mathbf{MM}^* + \mathbf{NN}^*$ or $\mathbf{MM}^*\mathbf{NN}^*$, but which is not enough to compute the moments of the eigenvalue distribution of $(\mathbf{M} + \mathbf{N})(\mathbf{M} + \mathbf{N})^*$. In a recent work [13], the authors generalized Voiculescu's work to rectangular random matrices, which allowed to prove that, under the hypothesis made here, the eigenvalue distribution of $(\mathbf{M} + \mathbf{N})(\mathbf{M} + \mathbf{N})^*$ converges to a probability measure which only depends on σ, τ and λ, and is denoted by $\sigma \boxplus_\lambda^+ \tau$.

Remark: The symmetric square root[3] of the distribution $\sigma \boxplus_\lambda^+ \tau$ is called the *rectangular free convolution with ratio* λ of the symmetric square roots $\sqrt{\sigma}, \sqrt{\tau}$ of σ and τ, and denoted by $\sqrt{\sigma} \boxplus_\lambda \sqrt{\tau}$. The operation $\boxplus_\lambda$ is, rather than $\boxplus_\lambda^+$, the one introduced in [13]. It is essentially equivalent to $\boxplus_\lambda$.

[3]For any probability measure ρ on $[0, \infty[$, the *symmetric square root* of ρ, denoted $\sqrt{\rho}$, is the only symmetric probability measure on the real line which push-forward by the $t \mapsto t^2$ function is ρ. Note that ρ is completely determined by $\sqrt{\rho}$, and *vice versa*. In [13], the use of symmetric measures turned out to be more appropriate. However, in the present paper, as the focus is on practical aspects, we shall not symmetrize distributions.

B. Computing $\boxplus_\lambda^+$

We fix $\lambda \in [0,1]$. Let us consider a probability measure ρ on $[0,+\infty[$, which has moments of all orders. Denote $\left\{m_n(\rho) = \int t^n d\rho(t)\right\}_{n\geq 0}$ the sequence of its moments. We associate to ρ another sequence of real numbers, $\{c_n(\rho)\}_{n\geq 1}$, depending on λ, called its *rectangular free cumulants*[4] *with ratio* λ, defined by the fact that the sequences $\{m_n(\rho)\}$ and $\{c_n(\rho)\}$ can be deduced from one another by the relation

$$C_\rho[z(\lambda M_{\rho^2}(z)+1)(M_{\rho^2}(z)+1)] = M_{\rho^2}(z) \qquad (10)$$

with the power series

$$C_\rho(z) = \sum_{n\geq 1} c_n(\rho) z^n \text{ and } M_{\rho^2}(z) = \sum_{n\geq 1} m_n(\rho) z^n. \qquad (11)$$

Equivalently, for all $n \geq 1$, the sequences $\{m_0(\rho),\ldots,m_n(\rho)\}$ and $\{c_1(\rho),\ldots,c_n(\rho)\}$ can be deduced from one another via the relations (involving an auxiliary sequence $\{m'_0(\rho),\ldots,m'_n(\rho)\}$)

$$m_0(\rho) = m'_0(\rho) = 1,$$

$$\forall n \geq 1, \quad m'_n(\rho) = \lambda m_n(\rho),$$

$$\forall n \geq 1, \quad m_n(\rho) =$$

$$c_n(\rho) + \sum_{k=1}^{n-1} c_k(\rho) \sum_{\substack{l_1,l'_1,\ldots,l_k,l'_k\geq 0 \\ l_1+l'_1+\cdots+l_k+l'_k=n-k}} m_{l_1}(\rho) m'_{l'_1}(\rho)\cdots m_{l_k}(\rho) m'_{l'_k}(\rho).$$

Example 5.1: As an example, it is proved in [15] that the law μ_λ of definition 2.1 has rectangular free cumulants with ratio λ given by $c_n(\mu_\lambda) = \delta_{n,1}$ for all $n \geq 1$.

The additive free convolution can be computed from the free cumulants via the following characterization [13].

Theorem 5.2: For σ,τ compactly supported, $\sigma \boxplus_\lambda^+ \tau$ is the only distribution m such that for all $n \geq 1$, $c_n(m) = c_n(\sigma) + c_n(\tau)$.

C. The rectangular free deconvolution

The moments/cumulants method can also be used to implement the rectangular free deconvolution. The *rectangular free deconvolution with ratio* λ of a probability measure m on $[0,+\infty[$ by a measure τ is (when it exists) the only measure σ such that $m = \sigma \boxplus_\lambda^+ \tau$. In this case, σ is denoted by $m \boxminus_\lambda \tau$. By Theorem 5.2, when it exists, $m \boxminus_\lambda \tau$ is characterized by the fact that for all $n \geq 1$,

$$c_n(m \boxminus_\lambda \tau) = c_n(m) - c_n(\tau).$$

[4]In [13], these numbers were not called rectangular free cumulants of ρ, but rectangular free cumulants of its symmetrized square root.

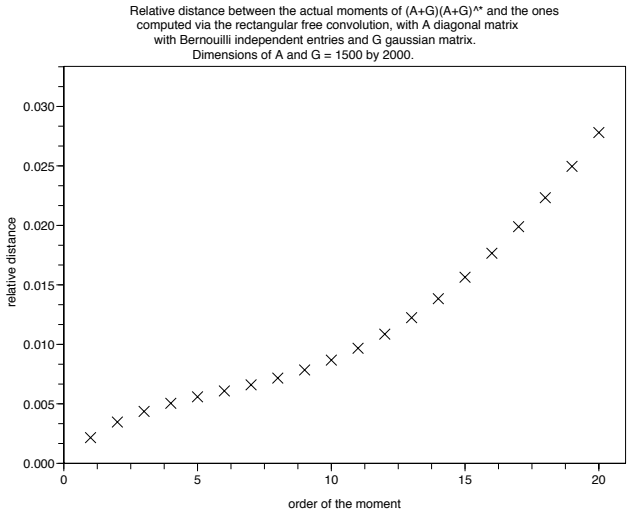

Fig. 6. Relative distance between the actual moments of the eigenvalue istribution of $(\mathbf{M}+\mathbf{N})(\mathbf{M}+\mathbf{N})^*$ and the ones computated via rectangular free convolution. Dimensions of $\mathbf{M}$ and $\mathbf{N}$: 1500×2000.

D. Simulations for finite size matrices

In Figure 6, one can read the value, for $n = 1,\ldots,20$, of

$$\left| \frac{m_n(\rho[(\mathbf{M}+\mathbf{N})(\mathbf{M}+\mathbf{N})^*])}{m_n(\nu \boxplus_\lambda^+ \mu_\lambda)|} - 1 \right| \qquad (12)$$

for $\lambda = 0.75$, $\mathbf{A}$ diagonal $n \times p$ random matrix with independent diagonal entries distributed according to $\nu = (\delta_0+\delta_1)/2$ and $\mathbf{G}$ Gaussian $n \times p$ $(= n/\lambda)$ as in definition 2.1. The dimensions of $\mathbf{M}$ and $\mathbf{N}$ are 1500×2000. It can be seen on this figure that the values of (12) are very small, which means that the moments/free cumulants approach is a good way to compute the spectrum of sums of independent random matrices. In Figure 7, we illustrated the efficiency of the combinatorial method to recover, for $\mathbf{M}$ large rectangular matrix, the spectrum of $\mathbf{M}\mathbf{M}^*$ from those of $\mathbf{N}\mathbf{N}^*$ and of $(\mathbf{M}+\mathbf{N})(\mathbf{M}+\mathbf{N})^*$. We simulated a random $n \times p$ diagonal matrix $\mathbf{M}$ such that $\mathbf{M}\mathbf{M}^*$ has eigenvalue distribution $\mu = (\delta_1+\delta_0)/4$ and a random $n \times p$ Gaussian matrix $\mathbf{N}$ such that the eigenvalue distribution of $\mathbf{N}\mathbf{N}^*$ is the measure μ_λ of definition 2.1. Then we compared the actual moments of $\mathbf{M}\mathbf{M}^*$ with their approximations obtained by free rectangular deconvolution of the (symmetric square root of) the eigenvalue distribution of $(\mathbf{M}+\mathbf{N})(\mathbf{M}+\mathbf{N})^*$ by the limit (symmetric square root of) the eigenvalue distribution of $\mathbf{N}\mathbf{N}^*$. We plotted, for $n = 1,\ldots,11$,

$$\left| \frac{m_n(\rho[(\mathbf{M}+\mathbf{N})(\mathbf{M}+\mathbf{N})^*] \boxminus_\lambda \mu_\lambda)}{m_n(\rho[\mathbf{M}\mathbf{M}^*])} - 1 \right|. \qquad (13)$$

E. Special cases

1) $\lambda = 0$: It is proved in [13] that if $\lambda = 0$, then $\boxplus_\lambda^+ = \boxplus$. This means that if $\mathbf{M},\mathbf{N}$ are independent $n \times p$ random matrices which dimensions n,p both tend to infinity such that $n/p \to 0$, then (under the hypothesis that $\mathbf{M}$ or $\mathbf{N}$ is invariant, in distribution, under multiplication by unitary matrices)

$$\text{eigenvalue distribution}((\mathbf{M}+\mathbf{N})(\mathbf{M}+\mathbf{N})^*)$$

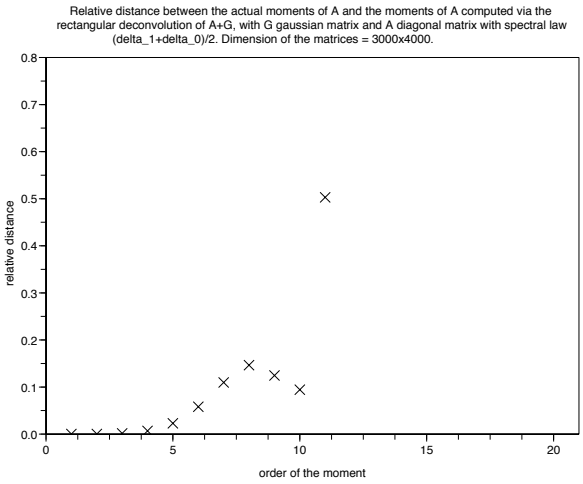

Fig. 7. Relative distance between the actual moments of the eigenvalue istribution of $\mathbf{M}\mathbf{M}^*$ and the ones computated via rectangular free deconvolution of $(\mathbf{M}+\mathbf{N})(\mathbf{M}+\mathbf{N})^*$, with $\mathbf{N}$ gaussian. Dimensions of $\mathbf{M}$ and $\mathbf{N}$: 3000×40000.

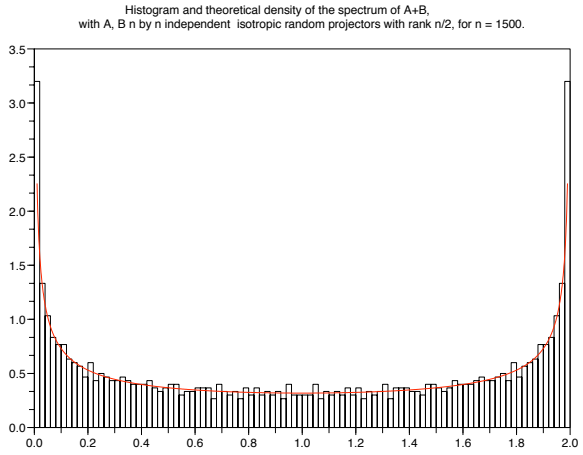

Fig. 8. Histogram and theoretical density of the spectrum of $\mathbf{A}+\mathbf{B}$, with $\mathbf{A}, \mathbf{B}$ independent n by n isotropic random projectors with rank $n/2$, for $n = 1500$.

$$\simeq \text{eigenvalue distribution}(\mathbf{M}\mathbf{M}^* + \mathbf{N}\mathbf{N}^*).$$

2) $\lambda = 1$: It is proved in [13] that if $\lambda = 1$, then for all σ, τ probability measures on $[0, +\infty[$, $\sigma \boxplus_\lambda^+ \tau$ is the push forward by the function $t \mapsto t^2$ of the free convolution $\sqrt{\sigma} \boxplus \sqrt{\tau}$ of the symmetrized square roots $\sqrt{\sigma}, \sqrt{\tau}$ of σ and τ.

VI. Discussion

A. Methods based on analytic functions

1) The R-transform method: The previous method of computation of the additive free convolution is very appealing and can be used in practice. From a more general theoretical point of view, it has two drawbacks. The first drawback is the fact that the method only works for measures with moments and the second one is that for μ, ν measures, it characterizes the measures $\mu \boxplus \nu$ only by giving its moments. We shall now expose a method, developed by Voiculescu and Bercovici [24], but also, in an indirect way, by Pastur, Bai, and Silverstein which works for any measure and which allows, when computations are not involved, to recover the densities. Unfortunately, this method is interesting only in very few cases, because the operations which are necessary here (the inversion of certain functions, extension of analytic functions) are almost always impossible to realize practically. It has therefore very little practical use. However, we provide in the following an example where this method works and give a simulation to sustain the theoretical results. In particular, note that in [25], Rao and Edelman provided similar computations of this method.

For ρ probability a measure on the real line, the *R-transform* K_ρ of ρ is the analytic function on a neighborhood of zero in $\{z \in \mathbb{C}\,;\, \Im z > 0\}$ defined by the equation

$$G_\rho\left(\frac{K_\rho(z)+1}{z}\right) = z, \tag{14}$$

The *Cauchy transform* of ρ is $G_\rho(z) = \int_{t\in\mathbb{R}} \frac{1}{z-t} d\rho(t)$.

Note that in the case where ρ is compactly supported, the notations of (14) and of (2) are in accordance. Hence we have, for all pair μ, ν of probability measures on the real line:

$$K_{\mu\boxplus\nu}(z) = K_\mu(z) + K_\nu(z). \tag{15}$$

Note that since, as proved in [19], for any probability measure ρ on the real line, ρ is the weak limit, as $y \to 0^+$, of the measure with density $x \mapsto -\frac{1}{\pi}\Im(G_\rho(x+iy))$, which allows us to recover $G_{\mu\boxplus\nu}$ *via* (14). Therefore, (15) theoretically determines $\mu \boxplus \nu$.

Example 6.1: As an example, for $\mu = \frac{\delta_0+\delta_1}{2}$, $\mu \boxplus \mu$ is the measure with density $\frac{1}{\pi\sqrt{x(2-x)}}$ on $[0, 2]$.

Figure 8 illustrates the previous example.

2) The S-transform method: As for the additive free convolution (see sectionVI-A.1), there exists an analytic method for the theoretical computation of $\boxtimes$, called the S-transform [24]: for $\rho \neq \delta_0$ a probability measure on $[0, +\infty[$, the *S-transform* S_ρ of ρ is the analytic function on a neighborhood of zero in $\mathbb{C}\backslash[0, +\infty)$ defined by the equation

$$M_\rho\left(\frac{z}{1+z}S_\rho(z)\right) = z, \tag{16}$$

where $M_\rho(z) = \int_{t\in\mathbb{R}} \frac{zt}{1-zt} d\rho(t)$. But the method based on this transform works only in some limited cases, because the formulas of S-transforms are almost never explicit.

3) The rectangular R-transform method: $\lambda \in [0, 1]$ is still fixed. For τ probability measure on $[0, +\infty[$, we shall define an analytic function $C_\tau(z)$ in a neighborhood of zero (to be more precise, in a neighborhood of zero in $\mathbb{C}\backslash\mathbb{R}^+$) which, in the case where τ is compactly supported, has the following series expansion:

$$C_\tau(z) = \sum_{n\geq 1} c_n(\tau) z^n. \tag{17}$$

It implies, by Theorem 5.2, that for all compactly supported probability measures σ, τ,

$$C_{\sigma\boxplus_\lambda^+\tau}(z) = C_\sigma(z) + C_\tau(z). \tag{18}$$

Hence the analytic transform C_ρ somehow "linearizes" the binary operation $\boxplus_\lambda^+$ on the set of probability measures on $[0,+\infty[$. The analytic function C_ρ is called the *rectangular R-transform*[5] *with ratio* λ of ρ.

It happens, as shown below, that for any probability measure ρ on $[0,+\infty[$, C_ρ can be computed in a direct way, without using the definition of (17).

Let us define $M_\rho(z) = \int_{t\in\mathbb{R}^+} \frac{zt}{1-zt} d\rho(t)$. Then the analytic function C_ρ is defined in a neighborhood of zero (in $\mathbb{C}\backslash\mathbb{R}^+$) to be the solution of

$$C_\rho[z(\lambda M_\rho(z)+1)(M_\rho(z)+1)] = M_\rho(z). \qquad (19)$$

which tends to zero for $z \to 0$.

To give a more explicit definition of C_ρ, let us define

$$H_\rho(z) = z(\lambda M_\rho(z)+1)(M_\rho(z)+1).$$

Then

$$C_\rho(z) = U\left(\frac{z}{H_\rho^{-1}(z)} - 1\right)$$

with

$$U(z) = \begin{cases} \frac{-\lambda-1+((\lambda+1)^2+4\lambda z)^{\frac{1}{2}}}{2\lambda} & \text{if } \lambda \neq 0, \\ z & \text{if } \lambda = 0, \end{cases}$$

where $z \mapsto z^{\frac{1}{2}}$ is the analytic version of the square root defined on $\mathbb{C}\backslash\mathbb{R}^-$ such that $1^{\frac{1}{2}} = 1$ and H_ρ^{-1} is the inverse (in the sense of the composition) of H_ρ.

To recover ρ from C_ρ, one has to go the inverse way:

$$H_\rho^{-1}(z) = \frac{z}{(\lambda C_\rho(z)+1)(C_\rho(z)+1)}$$

and

$$M_\rho(z) = U\left(\frac{H_\rho(z)}{z} - 1\right),$$

from what one can recover ρ, via its Cauchy transform.

Note that all this method is working for non compactly supported probability measures, and that (18) is valid for any pair of symmetric probability measures.

As for $\boxplus$ and $\boxtimes$, analytic functions give us a new way to compute the multiplicative free convolution of two symmetric probability measures τ, σ. However, the operations which are necessary in this method (the inversion of certain functions, the extension of certain analytic functions) are almost always impossible to realize in pratice. However, in the following example, computations are less involved.

Example 6.2: Suppose $\lambda > 0$. Then $\delta_1 \boxplus_\lambda^+ \delta_1$ has support $[(2-\kappa),(2+\kappa)]$ with $\kappa = 2(\lambda(2-\lambda))^{1/2} \in [0,2]$, and it has a density

$$\delta_1 \boxplus_\lambda^+ \delta_1 = \frac{\left[\kappa^2-(x-2)^2\right]^{1/2}}{\pi\lambda x(4-x)} \qquad (20)$$

on its support.

It means that if $\mathbf{A}$ is an $n\times p$ matrix with ones on the diagonal and zeros everywhere else, and $\mathbf{U},\mathbf{V}$ are random $n\times n$, $p\times p$ orthogonal matrices with Haar distribution, then as n,p tend to infinity such that $n/p \to \lambda$,

$$\text{eigenvalue distribution of } (\mathbf{A}+\mathbf{UAV})(\mathbf{A}+\mathbf{UAV})^*$$

has density

$$\simeq \frac{\left[\kappa^2-(x-2)^2\right]^{1/2}}{\pi\lambda x(4-x)} \text{ on } [2-\kappa, 2+\kappa].$$

Indeed, $\frac{\left[\kappa^2-(x-2)^2\right]^{1/2}}{\pi\lambda x(4-x)}$ is the density of the square of a random variable with density (20).

This point is illustrated in Figure 9.

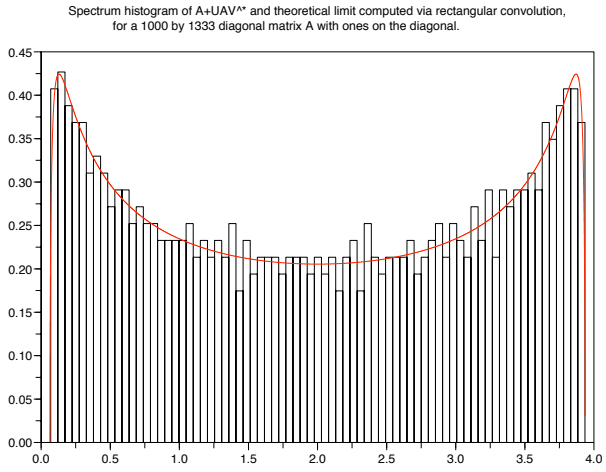

Fig. 9. Spectrum histogram of $\mathbf{A}+\mathbf{UAV}^*$ and theoretical limit computed via rectangular convolution, for a 1000×1333 diagonal matrix $\mathbf{A}$ with ones on the diagonal.

B. Other methods

Deconvolution techniques based on statistical eigen-inference methods using large Wishart matrices [26], random Matrix theory [27] or other deterministic equivalents *à la* Girko [28], [29], [30], [31] exist and are possible alternatives. Each one has its advantages and drawbacks. We focus here mainly on the random matrix approach. These are semi-parametric, grid-based techniques for inferring the empirical distribution function of the population from the sample eigen-spectrum. Unfortunately, they can only apply when all the matrices ($\mathbf{M}$ or $\mathbf{N}$) under consideration are Gaussian. For ease of understanding, let us focus on the particular case where m_i $(i=1,\ldots,p)$ are zero mean Gaussian vectors of size $n\times 1$ with covariance $\mathbf{\Sigma}$ and n_i $(i=1,\ldots,p)$ are i.i.d. Gaussian vectors with covariance $\sigma^2\mathbf{I}$. In this case, $\mathbf{H} = \mathbf{M}+\mathbf{N}$ is a Gaussian vector with covariance $\mathbf{\Sigma_c} = \mathbf{\Sigma}+\sigma^2\mathbf{I}$ and can be rewritten $\mathbf{H} = \mathbf{W}\mathbf{\Sigma_c}^{\frac{1}{2}}$ where $\mathbf{Y}$ is an $n\times p$ matrix whose entries are i.i.d. The deconvolution framework allows to recover the eigenvalues of $\mathbf{\Sigma_c}$ based only on the knowledge of $\mathbf{H}$. Before going further, let us introduce some definitions. In this case, one can use the following theorem due to Silverstein [32]:

Theorem 6.3: Let the entries of the $n\times p$ matrix $\mathbf{W}$ be i.i.d. with zero mean and variance $\frac{1}{n}$. Let $\mathbf{\Sigma}_c$ be a $p\times p$ a Hermitian

[5]Again, to make the notations of this paragraph coherent with the ones of the paper [13], where the rectangular machinery was build, one needs to use the duality between measures on $[0,+\infty)$ and their symmetrized square roots, which are symmetric measures on the real line.

matrix with an empirical distribution converging almost surely in distribution to a probability distribution $P_{\mathbf{\Sigma_c}}(x)$ as $p \to \infty$. Then almost surely, the empirical eigenvalue distribution of the random matrix:

$$\mathbf{H}\mathbf{H}^* = \mathbf{W}\mathbf{\Sigma_c}\mathbf{W}^*$$

converges weakly, as $p, n \to \infty$ with $\frac{p}{n}$ fixed, to the unique nonrandom distribution function whose Stieltjes transform satisfies:

$$-\frac{1}{G_{\mathbf{H}\mathbf{H}^*}(z)} = z - \frac{p}{n}\int \frac{ydP_{\mathbf{\Sigma_c}}(y)}{1+yG_{\mathbf{H}\mathbf{H}^*}(z)}.$$

Let us now explain the deconvolution framework in this specific case.

- In the first step, one can compute $G_{\mathbf{H}\mathbf{H}^*}(z) = \frac{1}{n}\operatorname{trace}(\mathbf{H}\mathbf{H}^* - z\mathbf{I})^{-1}$ for any z. The algorithm starts by evaluating $G_{\mathbf{H}\mathbf{H}^*}$ on a grid of a set of values $(z_j)_{j=1}^{J_n}$. The more values are considered (big value of J_n), the more accurate the result will be.
- In the second step, $dP_{\mathbf{\Sigma_c}}$ is approximated by a weighted sum of point masses:

  $$dP_{\mathbf{\Sigma_c}}(y) \sim \sum_{k=1}^{K} w_k \delta(t_k - y)$$

  where $t_k, k = 1, ..K$ is a chosen grid of points and $w_k \geq 0$ are such that $\sum_{k=1}^{K} w_k = 1$. The approximation turns the problem of optimization over measures into a search for a vector of weights $\{w_1, ...w_K\}$ in $\mathbb{R}_+^K$.
- In this case, we have that:

  $$\int \frac{ydP_{\mathbf{\Sigma_c}}(y)}{1+yG_{\mathbf{H}\mathbf{H}^*}(z)} \sim \sum_{k=1}^{K} w_k \frac{t_k}{1+t_k G_{\mathbf{H}\mathbf{H}^*}(z)}$$

  Hence the whole problem becomes an optimization problem: minimize

  $$\max_{j=1,..J_n} \max\left(|\ \Re(e_j)\ |, |\ \Im(e_j)\ |\right)$$

  subject to $\sum_{k=1}^{K} w_k = 1$ and $w_k \geq 0$ for all k, where

  $$\begin{aligned} e_j &= \frac{1}{\frac{1}{n}\operatorname{trace}(\mathbf{H}\mathbf{H}^* - z_j\mathbf{I})^{-1}} + z_j \\ &- \frac{p}{n}\sum_{k=1} K w_k \frac{t_k}{1+t_k\frac{1}{n}\operatorname{trace}(\mathbf{H}\mathbf{H}^* - z_j\mathbf{I})^{-1}}. \end{aligned}$$

The optimization procedure is a Linear program which can be solved numerically. Note that although there are several sources of errors due to the approximations of the integrals and the replacement of asymptotic quantities by non-asymptotic ones, [27] shows remarkably that the solution is consistent (in terms of size of available data and grid points on which to evaluate the functions).

The assumptions for the cumulants approach are a bit more general, because we do not suppose the random matrices we consider to be Gaussian. But the main difference lies in the way we approximate the distribution $P_{\mathbf{\Sigma_c}}$: we give moments, whereas here, $P_{\mathbf{\Sigma_c}}$ is approximated by a linear combination of Dirac masses.

VII. Entropy rate and spectrum

In wireless intelligent random networks, devices are autonomous and should take decisions based on their sensing capabilities. Of particularly interest are information measures such as capacity, signal to noise ratio, estimation of powers or even topology identification. Information measures are usually related to the spectrum (eigenvalues) of the matrices modelling the underlying network and not on the specific structure (eigenvectors). This entails many simplifications that make free deconvolution a very appealing framework for the design of these networks. In the following, we provide only one example of the applications of free deconvolution to wireless communications. Others can be found in [9].

The entropy rate [33]of a stationnary Gaussian stochastic process, upon which many information theoretic performance measures are based, can be expressed as:

$$H = \log(\pi e) + \frac{1}{2\pi}\int_{-\pi}^{\pi} \log(S(f))df,$$

where S is the spectral density of the process. Hence, if one knows the autocorrelation of the process, one has therefore a full characterization of the information contained in the process. Moreover, as side result, one can also show that the entropy rate is also related to the minimum mean squared error of the best estimator of the process given the infinite past [34], [35]. This remarkable result is of main interest for wireless communications as one can deduce one quantity from the other, especially as many receivers incorporate an MMSE (Minimum Mean Square Error) component. These results show the central importance of the autocorrelation function for Gaussian processes. In the discrete case when considering a random Gaussian vector $\mathbf{x}$ of size n, the entropy rate per dimension (or differential entropy) is given by:

$$\begin{aligned} H &= \log(\pi e) + \frac{1}{n}\log\det(\mathbf{R}) \\ &= \log(\pi e) + \frac{1}{n}\sum_{i=1}^{n}\log(\lambda_i), \end{aligned} \tag{21}$$

where $\mathbf{R} = \mathbb{E}(\mathbf{x}_i\mathbf{x}_i^*)$ and λ_i the associated eigenvalues. The covariance matrix (and more precisely its eigenvalues) carries therefore all the information of Gaussian networks. The Gaussianity of these networks is due to the fact that the noise, the channel and the signaling are very often Gaussian. Hence, in order to get a reliable estimate of the rate (and in extension of the capacity which is the difference between two differential entropies), one needs to compute the eigenvalues of the covariance matrix. For a number of observations p of the vector $\mathbf{x}_i, i = 1, ..., p$, the covariance matrix $\mathbf{R}$ is usually estimated by:

$$\hat{\mathbf{R}} = \frac{1}{p}\sum_{i=1}^{p}\mathbf{x}_i\mathbf{x}_i^* \tag{22}$$

$$= \mathbf{R}^{\frac{1}{2}}\mathbf{S}\mathbf{S}^*\mathbf{R}^{\frac{1}{2}} \tag{23}$$

Here, $\mathbf{S} = [\mathbf{s}_1, .., \mathbf{s}_p]$ is an $n \times p$ i.i.d zero mean Gaussian vector of variance $\frac{1}{p}$. In many cases, the number of samples p is of the same order as n. This is mainly due to the fact that

the network is highly mobile and the statistics are considered to be the same within a number p of samples, which restricts the use of classical asymptotic signal processing techniques. Therefore, information retrieval must be performed within a window of limited samples. Example (23) is unfortunately rarely encountered in practice in wireless communications. The signal of interest $\mathbf{s}_i$ is usually distorted by a medium, given by $\mathbf{m}_i = f(\mathbf{s}_i)$ where f is any function. Moreover, the received signal $\mathbf{y}_i$ is altered by some additive noise $\mathbf{n}_i$ (not necessarily Gaussian) but in many respect unitarily invariant (due to the fact that all the dimensions have the same importance). In this case, the model is known as the Information plus Noise model:

$$\mathbf{y}_i = \mathbf{m}_i + \mathbf{n}_i,$$

which can be rewritten in the following matrix form by stacking all the observations:

$$\mathbf{Y} = \mathbf{M} + \mathbf{N}. \tag{24}$$

We have therefore a signal $\mathbf{Y} = \mathbf{M} + \mathbf{N}$, with $\mathbf{M}, \mathbf{N}$ independent $n \times p$ ($n, p >> 1$, $n/p \simeq \lambda$) random matrices, $\mathbf{M} = \mathbf{R}^{\frac{1}{2}}\mathbf{S}$, with $\mathbf{S}$ having i.i.d. $N(0,1)$ entries, and $\mathbf{N}$ (the noise) has i.i.d. $N(0,\sigma)$ entries. σ is known and supposed to be measured by the device. From (21), the *entropy rate* H is given by:

$$H = \log(\pi e) + \frac{1}{n}\sum_{i=1}^{n}\log(\lambda_i),$$

where the λ_i's are the eigenvalues of $\mathbf{R}$. Hence H can be computed with the previous method, since $H = \log(\pi e) + \int \log(x)\mathrm{d}\mu_{\mathbf{R}}(x)$, where $\mu_{\mathbf{R}}$ denotes the eigenvalue distribution of $\mathbf{R}$. Indeed, $\mu_{\mathbf{R}}$ is given by the formula:

$$\mu_{\mathbf{R}} = (\mu_{\frac{1}{p}\mathbf{Y}\mathbf{Y}^*} \boxminus_\lambda \mu_{\frac{1}{p}\mathbf{N}\mathbf{N}^*}) \boxtimes \mu_{\frac{1}{p}\mathbf{S}\mathbf{S}^*},$$

where the eigenvalues distributions $\mu_{\frac{1}{p}\mathbf{N}\mathbf{N}^*}, \mu_{\frac{1}{p}\mathbf{S}\mathbf{S}^*}$ are known, by definition 2.1, to be the Marchenko-Pastur distribution (rescaled by the factor σ in the case of $\mu_{\frac{1}{p}\mathbf{N}\mathbf{N}^*}$). Using a polynomial approximation of the logarithm (because $\mu_{\mathbf{R}}$ is computed via its moments), we have implemented the algorithms based on the previous developments, and plotted the results in Figure 10. The moments are related to the eigenvalues through the Newton-Girard formula as suggested in [2].

Interestingly, the cumulants approach provides a good match between the theoretical and finite size approach.

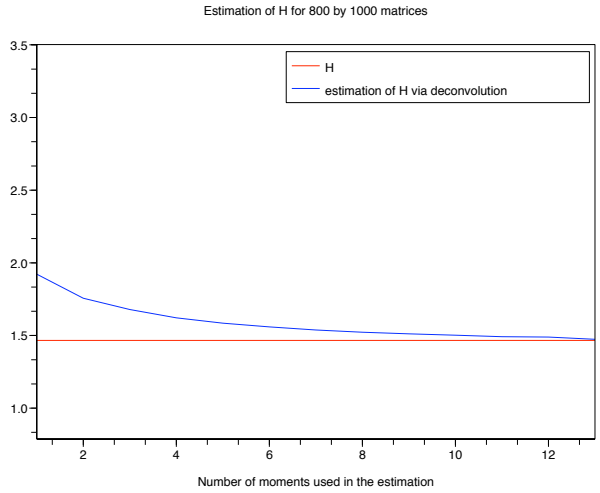

Fig. 10. *In red (horizontal line):* the real value of H. *In blue (other line):* the estimation using deconvolution and approximation of the logarithm by a polynomial which degree is represented on the X axis. *Dimension of the matrices:* $n = 800, p = 1000$.

VIII. Conclusion

In this paper, we reviewed classical and new results on free convolution (and deconvolution) and provided a useful framework to compute free deconvolution based on the cumulants approach when one of the matrices is unitarily invariant. The recursive cumulant algorithm extends the work on product and sum of square random matrices. For the information plus noise model (which is of main interest in wireless communications to compute information measures on the network), we have shown through simulations that the results are still valid for finite dimensions. It turns out that the free deconvolution approach based on moments/cumulants can be extended to much broader classes of matrices as recently advocated in [20], [21]. The research in this field is still in its infancy and the authors are convinced that many other models (than the invariant measure case) can be treated in the same vein.

IX. Acknowledgements

Florent Benaych-Georges would like to thank R. Rao to have encouraged him to work in Applied Mathematics.

References

[1] S. Haykin, "Cognitive Radio: Brain Empowered Wireless Communications," *Journal on Selected Areas in Communications*, vol. 23, pp. 201–220, 2005.

[2] Ø. Ryan and M. Debbah, "Free deconvolution for signal processing applications," *second round review, IEEE Trans. on Information Theory*, 2008, http://arxiv.org/abs/cs.IT/0701025.

[3] ——, "Channel capacity estimation using free probability theory," *Volume 56, Issue 11, Nov. 2008 Page(s):5654 - 5667.*

[4] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*. American Mathematical Society, 2000.

[5] D. Voiculescu, "Addition of certain non-commuting random variables," *J. Funct. Anal.*, vol. 66, pp. 323–335, 1986.

[6] D. V. Voiculescu, "Multiplication of certain noncommuting random variables," *J. Operator Theory*, vol. 18, no. 2, pp. 223–235, 1987.

[7] D. Voiculescu, "Circular and semicircular systems and free product factors," vol. 92, 1990.

[8] ——, "Limit laws for random matrices and free products," *Inv. Math.*, vol. 104, pp. 201–220, 1991.

[9] Ø. Ryan and M. Debbah, "Multiplicative free convolution and information-plus-noise type matrices," *submitted to Journal Of Functional Analysis*, 2007, http://www.ifi.uio.no/~oyvindry/multfreeconv.pdf.

[10] O. Ryan, "Implementation of free deconvolution," *Planned for submission to IEEE Transactions on Signal Processing*, 2006, http://www.ifi.uio.no/~oyvindry/freedeconvsigprocessing.pdf.

[11] B. Dozier and J. Silverstein, "On the empirical distribution of eigenvalues of large dimensional information-plus-noise type matrices," *Submitted.*, 2004, http://www4.ncsu.edu/~jack/infnoise.pdf.

[12] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free random variables*, ser. CRM Monograph Series. Providence, RI: American Mathematical Society, 1992, vol. 1, a noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups.

[13] F. Benaych-Georges, "Rectangular random matrices, related convolution," *To appear in Probability Theory and Related Fields*.

[14] S. Belinschi, F. Benaych-Georges, and A. Guionnet, "Regularization by free additive convolution, square and rectangular cases," *To appear in Complex Analysis and Operator Theory*.

[15] F. Benaych-Georges, "Infinitely divisible distributions for rectangular free convolution: classification and matricial interpretation," *Probab. Theory Related Fields*, vol. 139, no. 1-2, pp. 143–189, 2007.

[16] E. Wigner, "On the distribution of roots of certain symmetric matrices," *The Annals of Mathematics*, vol. 67, pp. 325–327, 1958.

[17] ——, "Characteristic vectors of bordered matrices with infinite dimensions," *The Annals of Mathematics*, vol. 62, pp. 458–564, 1955.

[18] Conway and Guy, *The Book of Numbers*. New York: Copernicus, 1996.

[19] F. Hiai and D. Petz, *The semicircle law, free random variables and entropy*, ser. Mathematical Surveys and Monographs. Providence, RI: American Mathematical Society, 2000, vol. 77.

[20] Ø. Ryan and M. Debbah, "Random Vandermonde matrices-part I: Fundamental results," *Submitted to IEEE Trans. on Information Theory*, 2008.

[21] ——, "Random Vandermonde matrices-part II: Applications," *Submitted to IEEE Trans. on Information Theory*, 2008.

[22] C. Bordenave, "Eigenvalues of euclidean random matrices," *To appear in Random Structures and Algorithms, ArXiv: math.PR/0606624*, 2008.

[23] A. Nica and R. Speicher, *Lectures on the combinatorics of free probability*, ser. London Mathematical Society Lecture Note Series. Cambridge: Cambridge University Press, 2006, vol. 335.

[24] H. Bercovici and D. Voiculescu, "Free convolution of measures with unbounded support," *Indiana Univ. Math. J.*, vol. 42, no. 3, pp. 733–773, 1993.

[25] A. Edelman and N. R. Rao, "The polynomial method for random matrices," *To appear in Foundations of Computational Mathematics*.

[26] N. Rao, J. Mingo, R. Speicher, and A. Edelman, "Statistical eigen-inference from large wishart matrices," *http://front.math.ucdavis.edu/0701.5314*, Jan, 2007.

[27] N. E. Karoui, "Spectrum estimation for large dimensional covariance matrices using random matrix theory," *http://front.math.ucdavis.edu/0609.5418*, Sept, 2006.

[28] V. L. Girko, *An Introduction to Statistical Analysis of Random Arrays*. The Netherlands, VSP, 1998.

[29] ——, *Theory of Stochastic Canonical Equations, Volumes I and II*. Kluwer Academic Publishers, 2001.

[30] W. Hachem, P. Loubaton, and J. Najim, "Deterministic equivalents for certain functionals of large random matrices," *The Annals of Applied Probability*, 2007.

[31] X. Mestre, "Improved estimation of eigenvalues of covariance matrices and their associated subspaces using their sample estimates," *submitted to IEEE Transactions on Information Theory*, 2008.

[32] J. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *J. of Multivariate Analysis*, vol. 55, pp. 331–339, 1995.

[33] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley, 1991.

[34] D. Guo, S. Shamai, and S. Verdu, "Mutual Information and Minimum Mean-Square Error in Gaussian Channels," *IEEE Trans. Information Theory*, vol. 51, pp. 1261–1283, 2005.

[35] D. P. Palomar and S. Verdu, "Gradient of Mutual Information in Linear Vector Gaussian Channels," *IEEE Trans. Information Theory*, vol. 52, pp. 141–154, 2006.