REVIEW   **Open Access**

Non-commutative large entries for cognitive radio applications

Antonia Maria Masucci[1*] and Merouane Debbah[2]

Abstract

Cognitive radio has been proposed as a solution for the problem of underutilization of the radio spectrum. Indeed, measurements have shown that large portions of frequency bands are not efficiently assigned since large pieces of bandwidth are unused or only partially used. In the last decade, studies in different areas, such as signal processing, random matrix theory, information theory, game theory, etc., have brought us to the current state of cognitive radio research. These theoretical advancements represents a solid base for practical applications and even further developments. However, still open questions need to be answered. In this study, free probability theory, through the free deconvolution technique, is used to attack the huge problem of retrieving useful information from the network with a finite number of observations. Free deconvolution, based on the moments method, has shown to be a helpful approach to this problem. After giving the general idea of free deconvolution for known models, we show how the moments method works in the case where scalar random variables are considered. Since, in general, we have a situation where more complex systems are involved, the parameters of interest are no longer scalar random variables but random vectors and random matrices. Random matrices are non-commutative operators with respect to the matrix product and they can be considered elements of what is called non-commutative probability space.Therefore, we focus on the case where random matrices are considered. Concepts from combinatorics, such as crossing and non-crossing partitions are useful tools to express the moments of Gaussian and Vandermonde matrices, respectively. Our analysis and simulation results show that free deconvolution framework can be used for studying relevant information in cognitive radio such as power detection, users detection, etc.

1 Introduction

In the last decade, recent studies [1] have shown that future communication systems should be designed to be able to adapt to their environment in order to tackle the problem of the underutilization of a precious resource such as the radio spectrum. Measurements have shown that large portions of frequency bands are not efficiently used, that is, for most of the time, large pieces of bandwidth are unoccupied or partially occupied [2]. A possible solution, introduced by Mitola [3,4], is represented by cognitive networks, that can be thought of as self-learning, adaptive and intelligent networks. In cognitive networks, unlicensed (secondary) systems improve spectral efficiency by sensing the environment and filling opportunistically the discovered holes spectrum (or white spaces) of licensed systems (primary), which have exclusive right to

operate in a certain spectrum band [5]. The current development of microelectronics allows us to suppose that these wireless systems, for which the spectrum utilization will play a key role, will be realized in the near future. These systems provide an efficient utilization of the radio spectrum based on the methodology understanding-by-building to learn from the environment and to adapt their parameters to statistical variations in the input stimuli [6].

The current development in cognitive radio research is the result of a multidisciplinary study that allows us to analyze different aspects of cognitive radio. We identify in signal processing, game theory, information theory, random matrix theory, etc., enabling areas for the development of cognitive radio.

Signal processing plays a major role in designing cognitive wireless networks, especially in spectrum sensing to identify spectrum opportunities and in the design of cognitive spectrum access to exploit the identified spectrum holes. We refer to spectrum sensing as the process where

* Correspondence: antonia-maria.masucci@ensea.fr
[1]ETIS/ENSEA - University of Cergy Pontoise - CNRS, Cergy-Pontoise, France
Full list of author information is available at the end of the article

devices look for a signal in the presence of noise for a given frequency band. Several digital signal processing techniques, such as matched filtering, energy detection, and cyclostationary feature detection are analyzed [7,8] to improve radio sensitivity and detect the presence of primary users. In [9], it is proved that the energy detector is an efficient spectrum sensing technique when the secondary user has limited information on the primary user's waveform, i.e., only the power of the local noise is known. The authors of [10] formulate the spectrum sensing problem as a nonlinear optimization problem, minimizing the interference to the primary user and meeting the requirement of opportunistic spectrum utilization. Cooperation between users follows as a consequence of the following constraints: (1) secondary users should not interfere with the primary transmissions and they should be able to detect the primary signal even if decoding the signal may be impossible [9]; (2) secondary users are in general not aware of the exact transmission scheme used by primary users. Cooperation among all cognitive users operating in the same band reduces the detection time and increases the overall agility with which cognitive users are able to shift bands [11-13]. Cooperation is designed in [14] as joint detection among all the cooperating users and in [15] as fusion center that makes the final decision about the occupancy of the band by fusing the decisions made by all cooperating users. In [16], cooperation is analyzed for the partial CSI (channel state information) scenario at the secondary users.

From a game theoretic point of view spectrum sharing may be considered as a competition. The importance of studying cognitive radio networks in a game theoretic framework is multifold. By modeling dynamic spectrum sharing between users as a game, users behaviors and actions can be analyzed in a formalized structure, where the theoretical results in game theory can be fully applied [17,18]. The optimization of spectrum usage is generally a multi-objective optimization problem, which is very difficult to analyze and to solve. Moreover, game theory provides us game models that predict convergence and stability of networks [19]. In [20], a game-theoretic adaptive channel allocation scheme is proposed for cognitive radio networks. In particular, a game is formulated to analyze the selfish and cooperative behaviors of the players. The players of this game were the wireless nodes and their strategies were defined in terms of channel selection. In [21], the convergence dynamics of the different types of games in cognitive radio systems is studied. Then, a game theory framework is proposed for distributed power control to achieve agility in spectrum usage in a cognitive radio network.

Information theory is used to characterize the achievable rates in a cognitive radio network under different assumptions on how the secondary systems interfere with the primary ones. Fundamental understanding on the capacity of the cognitive systems are provided in [22-26]. Using recent results on random matrix theory, the authors of [27,28] propose a new method for signal detection in cognitive radio, based on the eigenvalues of the covariance matrix of received signal at the secondary users. In [29], a spectrum sensing technique that relies on the use of multiple receivers to infer on the structure of the received signals using random matrix theory is proposed. The authors show that their technique is quite robust and does not require the knowledge of signal or noise statistics. These methods do not require any prior information on the primary signal or on the noise power. In [30,31] two hypothesis tests allowing to detect the presence of an unknown transmitter using several sensors are proposed and random matrix theory is used to provide the error associated with both tests.

We recognize as a crucial point of cognitive radio development understanding how much it is possible to infer from the network with the knowledge of just few observations. In the current study, we use free probability theory, through the concept of free deconvolution, to handle the problem of retrieving useful information from the network with a limited number of observations. Free deconvolution, based on the moments method, has shown to be a interesting tool to attack this problem.

In cognitive random networks, devices are autonomous and should take optimal decisions based on their sensing capabilities. We are particularly interested in measures such as capacity, signal to noise ratio, and estimation of the signal power. Such measures are usually related to the eigenvalues of the channel matrix and not on the specific structure, the eigenvectors. The fact that the spectrum of a stationary process is related to the information measure of the underlying process dates back to Kolmogorov [32]. The entropy rate of a stationary Gaussian stochastic process can be expressed by

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

where S is the spectral density of the considered process. Therefore, a complete characterization of the information contained in the process is given in the case the autocorrelation of the process is known. Moreover, the authors of [33,34] have shown that the entropy rate is also related to the minimum mean squared error (MMSE) of the best estimator of the process given the infinite past. In wireless communication, this means that it is possible retrieve one quantity from the other, especially as many receivers incorporate an MMSE component. In the discrete case,

the entropy rate per dimension (or differential entropy) of a Gaussian stochastic process $\mathbf{x}_i$ of size n is given by

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

where $\mathbf{R} = \mathbb{E}(\mathbf{x}_i\mathbf{x}_i^H)$ is the covariance and λ_i are its eigenvalues. The knowledge these eigenvalues provides us with the information on Gaussian networks. In fact, in order to estimate the rate and in extension the capacity which is the difference between two differential entropies or any other measure which involves performance criteria, one needs to compute the eigenvalues of the covariance. For a number of observations K of the vector $\mathbf{x}_i$, $i = 1,..., K$, the covariance $\mathbf{R}$ is usually estimated by

$$\hat{\mathbf{R}} = \frac{1}{p}\sum_{i=1}^{p}\mathbf{x}_i\mathbf{x}_i^H$$
$$= \mathbf{R}^{1/2}\mathbf{SS}^H\mathbf{R}^{1/2}, \tag{1}$$

where $\mathbf{S} = [\mathbf{s}_1,..., \mathbf{s}_K]$ is an $n \times K$ i.i.d zero mean Gaussian vector with variance $\frac{1}{K}$. In cognitive random networks, the number of samples K is of the same order as n, due to the high mobility of the network and to the fact that the statistics are considered to be the same within a K number of samples. Because of this, the use of classical asymptotic signal processing techniques is not more efficient since they require a number of samples $K >> n$. Therefore,our main problem consists in retrieving information within a window of limited samples. In this sense, free probability theory, through the concept of free deconvolution, is a very appealing framework for the study of cognitive networks. The main advantage of free deconvolution framework is that it provides us with helpful techniques to obtain useful informations from a finite number of observations. The deconvolution framework comes here from the fact that we would like to invert Equation (1) and express $\mathbf{R}$ with respect to $\hat{\mathbf{R}}$, since we can only have access to the sample covariance matrix. This is not possible, however, one can compute the eigenvalues of $\mathbf{R}$ knowing only the eigenvalues of $\hat{\mathbf{R}}$.

In the following, the general idea of free deconvolution is presented. We show how the moments method works in the case where scalar random variables are considered. However, since in practical situations systems are more complex, the parameters of interest are no longer scalar random variables and they need to be represented by random matrices. Therefore, we analyze the case where random matrices are considered. We analyze moments method for matrices which show the *freeness* property and

we show that it can be used to propose algorithmic methods to compute moments of finite Gaussian random matrices. Moreover, we analyze the case of matrices for which *freness* does not hold: Vandermonde, Hankel, Toeplitz. In the end, we present applications showing how the moments method approach can be used for studying cognitive radio: power detection, users detection, etc. In last section, we discuss our results presenting conclusions and open problems.

2 Information plus noise model

The example given in (1) is rarely met in practice in wireless communication since the transmitted signal $\mathbf{s}_i$ is, usually, distorted by a medium, given by $\mathbf{m}_i = f(\mathbf{s}_i)$ with f any function, and the received signal $\mathbf{y}_i$ is altered by some additive noise $\mathbf{n}_i$. We consider a finite number K of observations of the following $n \times 1$ received signal, known as Information plus Noise model

$$\mathbf{y}_i = \mathbf{m}_i + \mathbf{n}_i \quad i = 1, \ldots, K \tag{2}$$

which can be rewritten in a matrix form stacking all observations as

$$\mathbf{Y} = \mathbf{M} + \mathbf{N} \tag{3}$$

with $\mathbf{M}$ and $\mathbf{N}$ $K \times n$ independent random matrices. We are interested in retrieving information about the transmitted signal from the received signal, more explicitly to obtain the eigenvalues of $\mathbf{MM}^H$ from the eigenvalues of $\mathbf{YY}^H$ and $\mathbf{NN}^H$. This is exactly the goal of deconvolution.

In more general terms, the idea of deconvolution is related to the following problem [35]: Given $\mathbf{A}$, $\mathbf{B}$ two $n \times n$ independent square complex Hermitian (or real symmetric) random matrices:

(1) Can one derive the eigenvalue distribution of $\mathbf{A}$ from those of $\mathbf{A} + \mathbf{B}$ and $\mathbf{B}$? If feasible in the large n-limit, this operation is named additive free deconvolution,

(2) Can one derive the eigenvalue distribution of $\mathbf{A}$ from those of $\mathbf{AB}$ and $\mathbf{B}$? If feasible in the large n-limit, this operation is named multiplicative free deconvolution.

The techniques generally used to compute the operation of deconvolution in the large n-limit are the moments method [35] and the Stieltjes transform method [36]. Each of these methods has its advantages and its drawbacks. The moments method only works for measures with moments and characterizes the convolution only by giving its moments but it is easily implementable and, in many applications, one needs only a subset of the moments depending on the number of parameters to be estimated. Instead, the Stieltjes transform method works for any measure and it allows, when computations are possible, to

recover the densities. Unfortunately, this method works only in very few cases, since the operations which are necessary are almost always impossible to implement in practice and combining patterns of matrices naturally leads to more complex equations for the Stieltjes transform and can only be performed in the large n-limit.

We analyze the concept of free deconvolution based on the moments method which uses the empirical moments of the eigenvalue distribution of random matrices to obtain information about the eigenvalues. The moments method has shown to be a fruitful technique in both the asymptotic and the finite setting to compute deconvolution, as well as the simplest patterns, sums and products, and products of many independent matrices.

3 Moments method

3.1 Scalar case

We start by showing how moments method works for the scalar case. We consider X and Y two independent random variables and $Z = X + Y$. We are interested in retrieving the distribution of X knowing the distribution of Z and Y. The idea is to consider the moment generating function

$$M_Z(t) = \mathbb{E}\left[e^{tZ}\right] = M_X(t) M_Y(t)$$

from which

$$M_X(t) = \frac{M_Z(t)}{M_Y(t)}.$$

The knowledge of $M_X(t)$ gives us the distribution of the random variable X. However, it is not always easy to recover the distribution of X from $M_X(t)$. Another approach to solve the problem is to express the independence in terms of moments or cumulants. The cumulants are given by derivatives (at zero) of the function $g_X(t) = \log \mathbb{E}\left[e^{tX}\right]$. We denote by c_n the cumulant of order n:

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

The main advantage of using cumulants is due to the fact that for independent variables X and Y

$$g_{X+Y}(t) = \log\left[\mathbb{E}\left(e^{t(X+Y)}\right)\right] = \log\left[\mathbb{E}\left(e^{tX}\right)\mathbb{E}\left(e^{tY}\right)\right] = \log\left[\mathbb{E}\left(e^{tX}\right)\right] + \log\mathbb{E}\left(e^{tY}\right) = g_X(t) + g_Y(t),$$

this means that

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

Moments, denoted by $m_n(X) = \mathbb{E}[X^n]$, and cumulants of a random variable can be deduced from each other by the moment-cumulant formula

$$m_n(X) = \sum_{i=1}^{n} \sum_{\substack{k_1,\ldots,k_i \geq 1 \\ k_1 + \cdots + k_i = n}} c_{k_1}(X) \ldots c_{k_i}(X). \tag{4}$$

Therefore to obtain the distribution of X from the ones of $X + Y$ and Y one can compute the cumulants of X by the formula $c_n(X) = c_n(X + Y) - c_n(Y)$ and then deduce the moments of X from its cumulants.

In the multiplicative case, we consider X and Y independent random variables and we are interested in retrieving the distribution of X from XY and Y. In this case, the problem can be easily solved since

$$\mathbb{E}[(XY)^n] = \mathbb{E}[X^n]\mathbb{E}[Y^n],$$

then, we obtain $\mathbb{E}[X^n] = \frac{\mathbb{E}[(XY)^n]}{\mathbb{E}[Y^n]}$. Therefore, using the moments approach, we can compute the moments of X.

The moments method for scalar random variables seems to be very straightforward, however in general we have more complex situations. The generalization to multi-user multi-antenna communication systems has dramatically changed the nature of wireless communication problems. Furthermore, multi-dimensional stochastic problems need to be solved since cognitive devices are required to be simultaneously smarter and able to collaborate with one another. The random parameters in these problems are no longer scalar random variables but potentially vectors and matrices. The computation of deconvolution for random matrices is more complex than the scalar case and it is explained in the following.

3.2 Historical perspective

The origin of the moment approach for the derivation of the eigenvalue distribution of random matrices dates back to the work of Wigner [37]. Wigner was interested to the energy levels of nuclei (the positively charged central core of an atom). These energy levels are linked to the Hamiltonian operator by the Schrondinger equation, and the fact that these energy levels can be represented as the eigenvalues of the matrix representation of this operator, led Wigner to replace the exact matrix by a random matrix having the same properties. In most of the cases, it could be considered the following hermitian random matrix

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

where the upper diagonal elements are i.i.d generated with a binomial distribution. His study revealed 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 (see Figure 1). The idea to show this is to compute, as the dimension increases, the moments of the matrix **H**, that is the trace at different exponent. Consider

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

then the moments of the eigenvalue distribution of H are given by:

$$m_1(\mathbf{H}) = \frac{1}{n}\mathrm{Tr}(\mathbf{H}) = \frac{1}{n} \sum_{i=1}^{n} \lambda_i = \int \lambda dF_n(\lambda)$$

$$m_2(\mathbf{H}) = \frac{1}{n}\mathrm{Tr}(\mathbf{H}^2) = \frac{1}{n} \sum_{i=1}^{n} \lambda_i^2 = \int \lambda^2 dF_n(\lambda)$$

$$\cdots = \cdots$$

$$m_k(\mathbf{H}) = \frac{1}{n}\mathrm{Tr}(\mathbf{H}^k) = \frac{1}{n} \sum_{i=1}^{n} \lambda_i^k = \int \lambda^k dF_n(\lambda)$$

The traces above can be computed, as the dimension increases, using combinatorial tools. It turn out that all odd moments converge to zero, whereas all even moments converge to what is known as the *Catalan* numbers. The only distribution which has all its odd moments null and all its even moments equal to the Catalan numbers is the semi-circular law (see Figure 1) provided by

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

with $|x| \leq 2$. In this way, the moments approach is shown to be a useful method for computing the eigenvalues distribution of classical known matrices.

When more than one matrix is considered, the concept of asymptotic freeness [38] leaves us to compute the eigenvalue distribution of sums and products of random matrices.

3.3 Free probability framework
Free probability theory [38] was introduced by Voiculescu in the 1980s in order to attack some problems
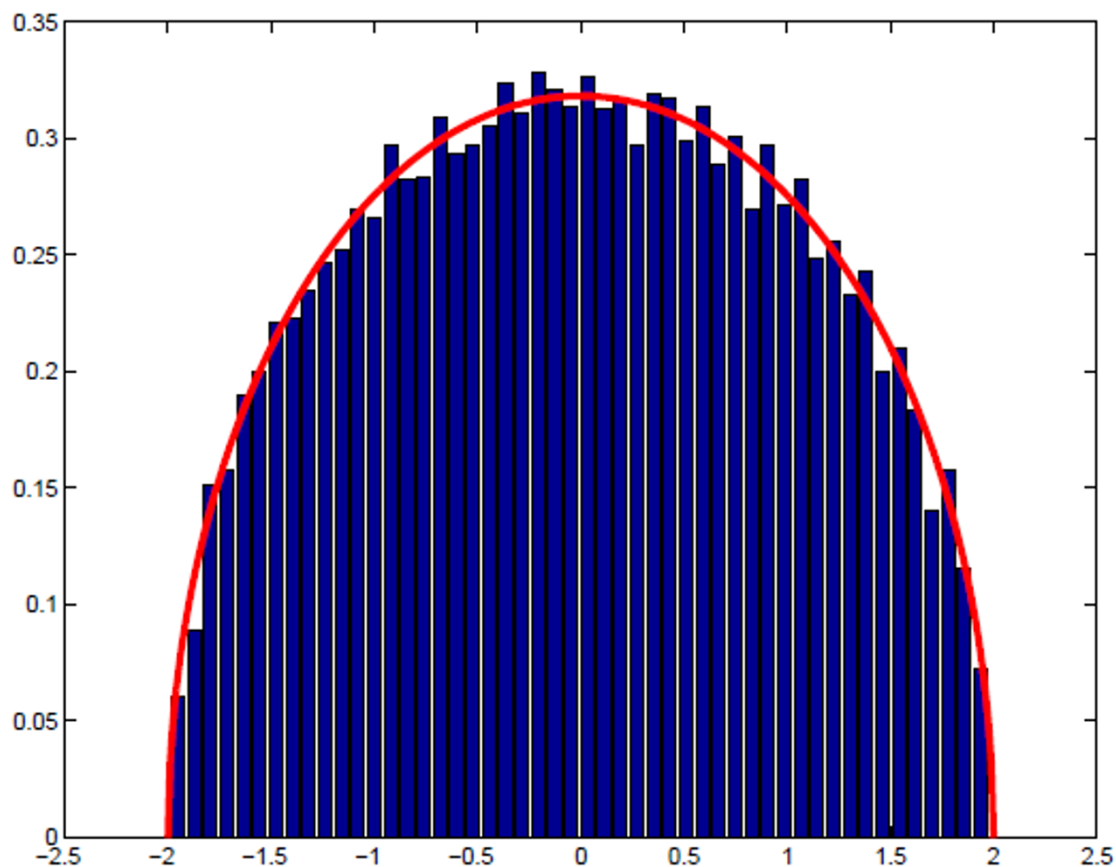

Figure 1 Semicircle law and simulation for a 512 × 512 Wigner matrix.

related to operator algebras and it can be considered as a generalization of classical probability theory to non-commutative algebras. The analogy between the concept of freeness and the independence in classical probability leaves us to work with noncommutative operators like matrices that can be considered elements in what is called a noncommutative probability space. The algebra of Hermitian random matrices is a particular case of such a probability space, for which the random variables, i.e., the random matrices, do not commute with respect to the matrix product.

Definition 3.1 *A non-commutative probability space* $(\mathcal{A}, \varphi)$ *consists of a unital*[a] *non-commutative algebra* $\mathcal{A}$ *over* $\mathbb{C}$ *and a unital linear function*

$$\varphi : \mathcal{A} \to \mathbb{C}, \varphi (1_{\mathcal{A}}) = 1$$

The elements of $\mathcal{A}$ are called non-commutative random variables. In our case, $\mathcal{A}$ will consist of $n \times n$ matrices or random matrices. For matrices, ϕ will be the normalized trace defined for any $\mathbf{A} \in \mathcal{A}$ by

$$\mathrm{tr}(\mathbf{A}) = \frac{1}{n}\mathrm{Tr}(\mathbf{A}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{A}(i, i),$$

while for random matrices, ϕ will be the linear functional τ defined by

$$\tau(\mathbf{A}) = \frac{1}{n}\mathbb{E}(\mathrm{Tr}(\mathbf{A})) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\mathbf{A}(i, i)].$$

Definition 3.2 *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be* $n \times n$ *hermitian random matrices and the functional* $\varphi(\mathbf{A}) := \lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\mathrm{Tr}(\mathbf{A})]$ *we say that* $\mathbf{A}$ *and* $\mathbf{B}$ *are asymptotically free if whenever there exist polynomials* p_i *and* q_i *such that* $\phi[p_i(\mathbf{A})] = 0$ *for all* i *and* $\phi[q_j(\mathbf{B})] = 0$ *for all* j, *then necessarily*

$$\varphi[p_1(\mathbf{A})q_1(\mathbf{B})p_2(\mathbf{A})q_2(\mathbf{B})\ldots] = 0.$$

Given $\mathbf{A}$, $\mathbf{B}$ $n \times n$ hermitian and asymptotically free random matrices such that their eigen-values distributions converge to some probability measure $\mu_{\mathbf{A}}$ and $\mu_{\mathbf{B}}$, respectively, then the eigenvalue distributions of $\mathbf{A} + \mathbf{B}$ and $\mathbf{AB}$ converge to a probability measure which depends on $\mu_{\mathbf{A}}$ and $\mu_{\mathbf{B}}$, called additive and multiplicative free convolution, and denoted by $\mu_{\mathbf{A}} \boxplus \mu_{\mathbf{B}}$ and $\mu_{\mathbf{A}} \boxtimes \mu_{\mathbf{B}}$, respectively.

Additive free deconvolution: The additive free deconvolution of a measure ρ by a measure v is (when it exists) the only measure μ such that $\rho = \mu \boxplus v$. In this case, μ is denoted by $\mu = \rho \boxminus v$.

Multiplicative free deconvolution The multiplicative free deconvolution of a measure ρ by a measure v is (when it exists) the only measure μ such that $\rho = \mu \boxtimes v$. In this case, μ is denoted by $\mu = \rho \boxslash v$.

For a given $n \times n$ random matrix $\mathbf{A}$, the p-th moment of $\mathbf{A}$ is defined, if it exists, as:

$$m_{\mathbf{A}}^{n,p} = \mathbb{E}[\mathrm{tr}(\mathbf{A}^p)] = \int \lambda^p d\rho_n(\lambda) \tag{5}$$

where $d\rho_n(\lambda) = \mathrm{E}(\frac{1}{n}\sum_{i=1}^{n}\delta(\lambda - \lambda_i))$ is the associated empirical mean measure, and λ_i are the eigenvalues of $\mathbf{A}$. The idea of additive and multiplicative free deconvolution stems from the fact that in the asymptotic case

$$m_{\mathbf{A}+\mathbf{B}}^{p} := \lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\mathrm{Tr}((\mathbf{A}+\mathbf{B})^p)] = f(m_{\mathbf{A}}^{(1)}, \ldots, m_{\mathbf{A}}^{(p)}, m_{\mathbf{B}}^{(1)}, \ldots, m_{\mathbf{B}}^{(p)})$$

$$m_{\mathbf{AB}}^{p} := \lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\mathrm{Tr}((\mathbf{AB})^p)] = g(m_{\mathbf{A}}^{(1)}, \ldots, m_{\mathbf{A}}^{(p)}, m_{\mathbf{B}}^{(1)}, \ldots, m_{\mathbf{B}}^{(p)})$$

which means that we can express the moments of $\mathbf{A} + \mathbf{B}$ and the moments of $\mathbf{AB}$ as a function of the moments of A and the moments of $\mathbf{B}$. In other words, the joint distribution of $\mathbf{A} + \mathbf{B}$ and the joint distribution of AB depend only on the marginal distributions of $\mathbf{A}$ and $\mathbf{B}$.

Even if matrices with finite dimensions are not free, the free probability framework, based on the moments, can still be used to propose an algorithmic method to compute these operations for finite size matrices. This means that

$$m_{\mathbf{A}+\mathbf{B}}^{n,p} = \frac{1}{n}\mathbb{E}[\mathrm{Tr}((\mathbf{A}+\mathbf{B})^p)] = f(m_{\mathbf{A}}^{(1)}, \ldots, m_{\mathbf{A}}^{(p)}, m_{\mathbf{B}}^{(1)}, \ldots, m_{\mathbf{B}}^{(p)})$$

$$m_{\mathbf{AB}}^{n,p} = \frac{1}{n}\mathbb{E}[\mathrm{Tr}((\mathbf{AB})^p)] = g(m_{\mathbf{A}}^{(1)}, \ldots, m_{\mathbf{A}}^{(p)}, m_{\mathbf{B}}^{(1)}, \ldots, m_{\mathbf{B}}^{(p)})$$

Hence, when, for $n \to \infty$, the moment $m_{\mathbf{A}}^{n,p}$ converges almost surely to an analytical expression $m_{\mathbf{A}}^p$ that depends only on some specific parameters of $\mathbf{A}$ (such as the distribution of its entries).[b] Therefore, in the finite setting one is still able by recursion to express all the moments of $\mathbf{A}$ with respect only to the moments of $\mathbf{A} + \mathbf{B}$ and $\mathbf{B}$, or $\mathbf{AB}$ and $\mathbf{B}$.

We will give a characterization of free deconvolution in terms of free cumulants, which are polynomials in the moments with a nice behaviour with respect to the freeness. The nomenclature comes from classical probability theory where corresponding objects are well known. There exists a combinatorial description of these classical cumulants, which depends on partitions of sets. In the same way, free cumulants can also be described combinatorially, the only difference to the classical case is the replacement of partitions by the so called non-crossing partitions [39].

Definition 3.3 *A partition* π *of a set* $\{1, 2,..., n\}$ *is a decomposition in subsets* V_i: $\pi = \{V_1,..., V_r\}$ *such that* $\bigcup_{i=1}^{r} V_i = \{1, \ldots, n\}$, *with* $V_i \neq \varnothing$ *and* $V_i \cap V_j = \varnothing$ *for all* $i \neq j$.

The set of all partitions of $\{1, 2,..., n\}$ is denoted by $\mathcal{P}(n)$, and V_i are called blocks of π.

For a given random variable X, the relationship between moments and cumulants given in (4) can be combinatorially expressed by

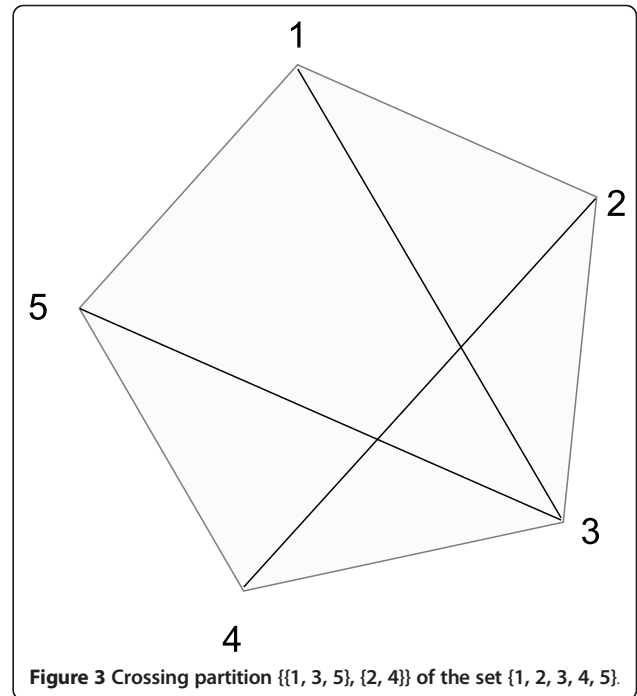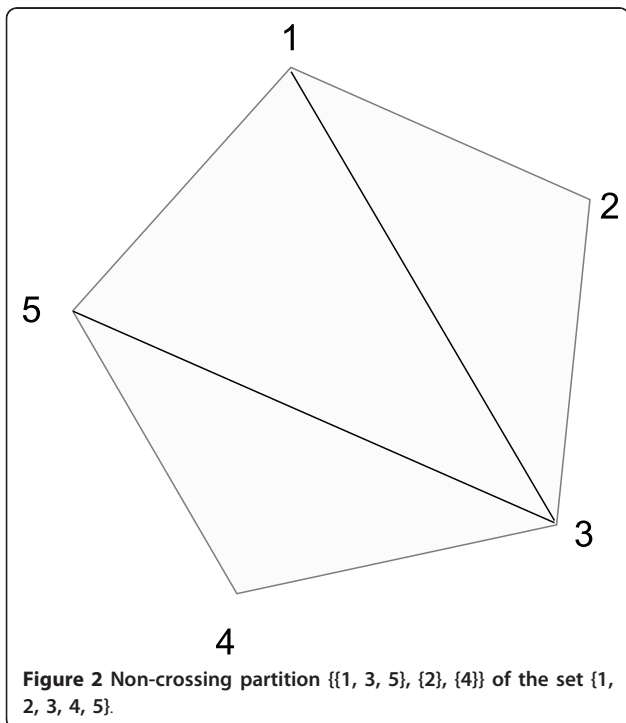$$\mathbb{E}(X^n) = \sum_{\pi \in \mathcal{P}(n)} c_\pi$$

where $c_\pi = \prod_{i=1}^{|\pi|} c_{|\pi_i|}$ when $\pi = \{\pi_1,..., \pi_{|\pi|}\}$.

Definition 3.4 *A partition π of $\{1,..., n\}$ is non-crossing if whenever we have four numbers $1 \leq i < k < j < l \leq n$ such that i and j are in the same block, k and l are in the same block, we also have that i, j, k, l belong to the same block.*

We denote by $NC(n)$ the set of non-crossing partition of $\{1,..., n\}$, and if this situation does not happen, then we call π a crossing partition. Examples of non-crossing and crossing partitions are give in Figures 2 and 3.

The computation of free deconvolution by the moments method approach is based on the moment-cumulant formula, which gives a relation between the moments $m_A^p \equiv m_{\mu_A}^p$ and the free cumulants $\kappa_A^p \equiv \kappa_{\mu_A}^p$ of a matrix A, where μ_A is the associated measure. It turns out that the cumulants are quantities much easier to compute, also thanks to the concept of non-crossing partitions. The moment-cumulant formula says that

$$m_A^p = \sum_{\pi = \{V_1,...,V_k\} \in NC(p)} \prod_{i=1}^{k} \kappa_A^{|V_i|}, \tag{6}$$

Figure 3 Crossing partition {{1, 3, 5}, {2, 4}} of the set {1, 2, 3, 4, 5}.

where $|V_i|$ is the cardinality of the block V_i. From (6) it follows that the first p cumulants can be computed from the first p moments, and viceversa.

The following characterization allows us to compute easily the additive free convolution using free cumulants.

Theorem 3.5 [38] *Given A and B asymptotically free random matrices, $\mu_A \boxplus \mu_B$ is the only law such that for all $p \geq 1$*

$$\kappa_{\mu_A \boxplus \mu_B}^p = \kappa_{\mu_A}^p + \kappa_{\mu_B}^p \tag{7}$$

Hence, the deconvolution of μ_{A+B} by μ_B, denoted by $\mu_{(A+B)} \boxminus \mu_B$, is characterized by the fact that for all $p \geq 1$

$$\kappa_{\mu_{A+B} \boxminus \mu_B}^p = \kappa_{\mu_{A+B}}^p - \kappa_{\mu_B}^p. \tag{8}$$

The implementation of additive free deconvolution is based on the following steps: for the two matrices ($A + B$) and B, we first compute the free cumulants, then, considering the relation between the cumulants and the moments, we can obtain information about the distribution of the eigenvalues of A.

The moments method, in the multiplicative case, is based on the relation between the moments $m_A^p \equiv m_{\mu_A}^p$ and the coefficients $s_A^p \equiv s_{\mu_A}^p$ of the S-transform[c] of the measure associated to A. They can be deduced one from each other from the following relations for all $p \geq 1$

$$m_A^1 s_A^1 = 1, \quad s_A^p = \sum_{k=1}^{p+1} s_A^k + \sum_{\substack{p_1,...,p_k \geq 1 \\ p_1 + \cdots + p_k = p+1}} m_A^{p_1} \ldots m_A^{p_k}.$$

Figure 2 Non-crossing partition {{1, 3, 5}, {2}, {4}} of the set {1, 2, 3, 4, 5}.

Hence, we can compute multiplicative free convolution by the following characterization.

Theorem 3.6 [38]*Given* **A** *and* **B** *asymptotically free random matrices,* $\mu_{\mathbf{A}} \boxtimes \mu_{\mathbf{B}}$ *is the only law such that:*

$$S_{\mu_{\mathbf{A}} \boxtimes \mu_{\mathbf{B}}} = S_{\mu_{\mathbf{A}}} S_{\mu_{\mathbf{B}}}$$

The multiplicative free deconvolution of $\mu_{\mathbf{AB}}$ by $\mu_{\mathbf{B}}$, $\mu_{(\mathbf{AB})} \boxtimes \mu_{\mathbf{B}}$, is characterized by the fact that for all $p \geq 1$

$$s^p_{\mu_{\mathbf{AB}} \boxtimes \mu_{\mathbf{B}}} s^1_{\mu_{\mathbf{B}}} = s^p_{\mu_{\mathbf{AB}}} - \sum_{k=1}^{p-1} s^k_{\mu_{\mathbf{AB}} \boxtimes \mu_{\mathbf{B}}} s^{p+1-k}_{\mu_{\mathbf{B}}}. \qquad (9)$$

Even though freeness usually does not hold for finite matrices, the moments method can still be used to propose algorithmic methods to compute their moments. Focusing on the study of random matrices in the finite case, the authors of [40] were able to derive the explicit series expansion of the eigenvalue distribution of various models, namely the case of non-central Wishart distributions as well as one sided correlated zero mean Wishart distributions. In particular, they proposed a general finite dimensional statistical inference framework based on the moments method in the finite case, which takes a set of moments as input and produces sets of moments as output with the dimensions of the matrices considered finite. They focus on the finite Gaussian case. The formulas of the moments presented in their contributions have been generated by iterations through partitions and permutations and concepts from combinatorics. The first and simplest result concerns the moments of a product of a deterministic matrix and a Wishart matrix. Let n, N be positive integers, **X** be $n \times N$ standard, complex, Gaussian[d] matrix and **D** a (deterministic) $n \times n$ matrix. Denoting the moments $D_p = \text{tr}(\mathbf{D}^p)$ and $M_p = \mathbb{E}[\text{tr}((\mathbf{D}\frac{1}{N}\mathbf{XX}^H)^p)]$ for any positive integer p, Theorem 1 in [40] allows us to express the moments M_p in terms of the moments D_p. In particular, the first three moments can be written as

$$M_1 = D_1$$
$$M_2 = D_2 + cD_1^2$$
$$M_3 = \left(1 + \frac{1}{N^2}\right)D_3 + 3cD_2D_1 + c^2D_1^3$$

where $c = \frac{n}{N}$. By a simple recursion, we can express D_p from M_p. For the first three moments these recursions become

$$D_1 = M_1$$
$$D_2 = M_2 - cM_1^2$$
$$D_3 = (M_3 - 3c(M_2 - cM_1^2)M_1 + c^2M_1^3)\left(1 + \frac{1}{N^2}\right)^{-1}.$$

Considering the sum of a **D** deterministic $n \times N$ matrix and **X** a $n \times N$ standard, complex, Gaussian matrix, in accordance with the [40, Theorem 2], for any positive integer p the moments $M_p = \mathbb{E}[\text{tr}((\frac{1}{N}\mathbf{D} + \mathbf{X})(\mathbf{D} + \mathbf{X})^H)^p)]$ can be expressed in terms of the moments $D_p = \text{tr}((\frac{1}{n}\mathbf{DD}^H)^p)$ as the following formulas:

$$M_1 = D_1 + 1$$
$$M_2 = D_2 + (2 + 2c)D_1 + (1 + c)$$
$$M_3 = D_3 + (3 + 3c)D_2 + 3cD_1^2 + \left(3 + 9c + 3c^2 + \frac{3}{N^2}\right)D_1 + \left(1 + 3c + c^2 + \frac{1}{N^2}\right)$$

In this case also, by a simple recursion, one can express D_p from M_p. It is clear how the operation of deconvolution can be viewed as operating on the moments: explicit expression for the moments of the Gram matrices associated to our models (sum or product of a deterministic matrix and a complex standard Gaussian matrix) are found, and are expressed in terms of the moments of the matrices involved. Hence, deconvolution means to express the moments, in this case of the deterministic matrices, in function of the moments of the Gram matrices.

Similar results are found when the Gaussian matrices are assumed to be square and selfad-joint. The implementation of the results is also able to generate the moments of many types of combinations of independent Gaussian and Wishart random matrices.

The algorithms are based on iterations through partitions and permutations and they give us rather complex expression. However, the author of [41] have generated Matlab codes based on concepts as partitions and permutations in order to implement the above results.

Once known the moments, the Newton-Girard formulas [42] can be used to retrieve the eigenvalues from the moments. These formulas state a relationship between the elementary symmetric polynomials

$$\Pi_1(\lambda_1, \ldots, \lambda_n) = \lambda_1 + \cdots + \lambda_n$$
$$\Pi_2(\lambda_1, \ldots, \lambda_n) = \sum_{1 \leq i < j \leq n} \lambda_i \lambda_j$$
$$\vdots$$
$$\Pi_n(\lambda_1, \ldots, \lambda_n) = \lambda_1 \ldots \lambda_n,$$

and the sums of the powers of their variables

$$S_p(\lambda_1, \ldots, \lambda_n) = \sum_{1 \leq i \leq n} \lambda_i^p = nt^p$$

(with t^p being the p-th moment) through the recurrence relation

$$(-1)^m m \prod_m (\lambda_1, \ldots, \lambda_n) + \sum_{i=1}^m (-1)^{k+m} S_k(\lambda_1, \ldots, \lambda_n) \prod_{m-k} (\lambda_1, \ldots, \lambda_n) = 0. \qquad (10)$$

If the sums of powers $S_p(\lambda_1,..., \lambda_n)$ are known for $1 \leq p \leq n$, the relation (10) can be used to compute the elementary symmetric polynomials $\prod_m(\lambda_1,..., \lambda_n)$ for $1 \leq m \leq n$. Therefore, the characteristic polynomial

$$(\lambda - \lambda_1)...(\lambda - \lambda_n)$$

(the roots of which provide the eigenvalues of the associated matrix) can be fully charac-terized since its $n - k$ coefficient is given by $(-1)^k \prod_k(\lambda_1,..., \lambda_n)$. In this way the entire characteristic polynomial can be computed, and the eigenvalues can also be found.

4 Non free case

In recent works, deconvolution, based on the moments method, has been analyzed when $n \to \infty$ for some particular matrices **A** and **B**. For instance, when **A** is a random Vandermonde matrix and B is a deterministic diagonal matrix [43], or when **A** and **B** are two independent random Vandermonde matrices [44]. The authors in [43] developed analytical methods for finding moments of random Vandermonde matrices with entries on the unit circle and provide explicit expressions for the moments of the Gram matrix associated to the models considered. The explicit expressions found for the moments are useful for performing deconvolution. In these cases the moments technique has been shown to be very appealing and powerful in order to derive the exact asymptotic moments of "non free matrices". This type of matrices occurs in cognitive radio [45].

Definition 4.1 *An $N \times M$ random Vandermonde matrix with entries on the unit circle has the form*

$$\mathbf{V} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & \cdots & 1 \\ e^{-j\omega_1} & \cdots & e^{-j\omega_M} \\ \vdots & \cdots & \vdots \\ e^{-j(N-1)\omega_1} & \cdots & e^{-j(N-1)\omega_M} \end{bmatrix},$$

where the phases $\omega_1,..., \omega_M$ are i.i.d. random variables in $[0, 2\pi]$.

The asymptotic behaviour of random Vandermonde matrices is analyzed when N and M are large, both go to infinity at a given ratio $\frac{M}{N} \to c$, with c constant. The scaling factor $\frac{1}{\sqrt{N}}$ and the assumption that the entries lies on the unit circle guarantee that the analysis will give limiting asymptotic behaviour.

Definition 4.2 *For $\rho = \{W_1, ..., W_r\} \in \mathcal{P}(n)$, we define*

$$K_{\rho,\omega,N} =$$
$$\frac{1}{N^{n+1-|\rho|}} \int\limits_{(0,2\pi)^{|\rho|}} \prod_{k=1}^{n} \frac{1 - e^{jN(\omega_{b(k-1)} - \omega_{b(k)})}}{1 - e^{j(\omega_{b(k-1)} - \omega_{b(k)})}} d\omega_1 \ldots d\omega_{|\rho|}, \quad (11)$$

where $\omega_{W_1}, \ldots, \omega_{W_{|\rho|}}$ are i.i.d. (indexed by the blocks of ρ) with the same distribution of ω and where $b(k)$ is the block of ρ which contains k. If the limit

$$K_{\rho,\omega} = \lim_{N \to \infty} K_{\rho,\omega,N} \quad (12)$$

exists, we call it a Vandermonde mixed moment expansion coefficient.

These quantities do not behave exactly as cumulants, but rather as weights which tell us how a partition in the moment formula we present should be weighted. In this respect, the formulas presented for the moments are different from classical or free moment-cumulant formulas, since these do not perform this weighting. The limits $K_{\rho,\omega}$ may not always exist, and necessary and sufficient conditions for their existence seem to be hard to find. In [43], it has been proved that the limit in (12) exists if the density of ω is continuous. The calculation is based on combinatorial computation using crossing partitions since the matrices are not free.

Theorem 4.3 *Let $\omega_1,..., \omega_M$ be phases independent and identically distributed in $[0, 2\pi]$. we have that*

$$m_n = \lim_{N \to \infty} \mathbb{E}[\mathrm{tr}_M(\mathbf{P}\mathbf{V}^H\mathbf{V})^n] = \sum_{\rho \in \mathcal{P}(n)} K_{\rho,\omega} c^{|\rho|-1} P_\rho \quad (13)$$

exist when $\frac{M}{N} \to c > 0$, $P_n = \lim_{N \to \infty} \mathrm{tr}_M(\mathbf{P}^n)$, $P_\rho = \prod_{i=1}^{k} P_{W_i}$.

Remark 1. The fact that all moments exist is not enough to guarantee that there exists a limit probability measure having these moments. However, it is proved in [46] that the Carleman's condition is satisfied.

Uniform phase distribution plays an important role or Vandermonde matrices.

Theorem 4.4 *For Vandermonde matrices with uniform phase distribution u, the Vander-monde mixed moment expansion coefficient*

$$K_{\rho,u} = \lim_{N \to \infty} K_{\rho,u,N} \quad (14)$$

exists $\forall \rho$. Moreover, $K_{\rho,u}$ satisfies the following properties

- $0 \leq K_{\rho,u} \leq 1$;
- $K_{\rho,u}$ *are rational numbers $\forall \rho$;*
- $K_{\rho,u} = 1 \Leftrightarrow \rho$ *is non-crossing partition;*
- *Let $V_{\omega,n} = \lim_{N \to \infty} \mathbb{E}[\mathrm{tr}_M(\mathbf{V}_\omega^H\mathbf{V}_\omega)^n]$ (with $\mathbf{V}_\omega$ a Vandermonde matrix with phase distribution ω), then $V_{u,n} \leq V_{\omega,n}$.*

The importance of uniform phase distribution is also expressed by the following theorem.

Theorem 4.5 *The Vandermonde mixed moment expansion coefficient $K_{\rho,\omega} = \lim\limits_{N\to\infty} K_{\rho,\omega,N}$ exist whenever the density p_ω of ω is continuous on $[0, 2\pi]$, then*

$$K_{\rho,\omega} = K_{\rho,u}^{|\rho|-1} \left(\int_0^{2\pi} p_\omega(x)^{|\rho|} dx \right). \tag{15}$$

The behaviour of Vandermonde matrices is different when the density of ω has singularities and depends on the density growth rates near the singularities points. Indeed, for the case of generalized Vandermonde matrices, whose columns do not consist of uniformly distributed power, it is possible to define mixed moment expansion coefficients but the formulas are more complex.

When many independent Vandermonde matrices are considered, the following relations hold.

Theorem 4.6 *If $\mathbf{V}_1, \mathbf{V}_2,...$ are independent Vandermonde matrices with the same phase distribution ω with continuous density, then*

$$\lim\limits_{N\to\infty} \mathbb{E}[\mathrm{tr}_M(\mathbf{P}_1(N)\mathbf{V}_{i_1}^H\mathbf{V}_{i_2}\cdots\mathbf{P}_n(N)\mathbf{V}_{i_n}^H\mathbf{V}_{i_1})] = \sum_{\rho\le\sigma\in\mathcal{P}(n)} K_{\rho,\omega}c^{|\rho|-1}P_\rho \tag{16}$$

exist when $\frac{M}{N}$, with $\sigma = \{\sigma_1, \sigma_2\} = \{\{1, 3,... \}, \{2, 4,... \}\}$, and "$\le$"denotes the refinement order, i.e., any block of ρ is contained within a block of σ.

If $\mathbf{V}_1, \mathbf{V}_2,...$ are independent Vandermonde matrices with different phase distribution ω_i with continuous density p_{ω_i}, Equation (16) still holds with $K_{\rho,\omega}$ replaced by

$$K_{\rho,u}^{|\rho|-1} \left(\int_0^{2\pi} \prod_{i=1}^n p_{\omega_i}(x)^{|\rho|} dx \right).$$

In [44], the authors generalize the above results replacing convergence in distribution with almost sure convergence in distribution.

Such matrices are applied to cognitive radio in [45], where authors consider a scenario with a primary and a secondary user wish to communicate with their corresponding receivers simultaneously over frequency selective channels is considered. Under realistic assumptions that the primary user is ignorant of the secondary user's presence and that the secondary transmitter has no side information about the primary's message, the authors propose a Vandermonde precoder that cancels the interference from the secondary user by exploiting the redundancy of a cyclic prefix.

4.1 Toeplitz and Hankel matrices

The same strategy used to compute the moments of Vandermonde matrices can be used for Hankel and Toeplitz matrices.

Definition 4.7 *We define a Toeplitz matrix*

$$\mathbf{T} = \frac{1}{\sqrt{N}} \begin{bmatrix} X_0 & X_1 & X_2 & \cdots & X_{N-2} & X_{N-1} \\ X_1 & X_0 & X_1 & \cdots & & X_{N-2} \\ X_2 & X_1 & X_0 & & & \vdots \\ \vdots & & & & & \vdots \\ X_{N-1} & X_{N-2} & \cdots & & X_1 & X_0 \end{bmatrix}, \tag{17}$$

where X_i are i.i.d., real-valued random variables with unit variance.

Theorem 4.8 [44] *We denote by M_i the asymptotic moments of order $2i$ of a Toeplitz matrix $\mathbf{T}$, then we have*

$$M_1 = 1$$
$$M_2 = \frac{8}{3}$$
$$M_3 = 11.$$

Similar results hold for Hankel matrices.

Definition 4.9 *We define an Hankel matrix*

$$\mathbf{H} = \frac{1}{\sqrt{N}} \begin{bmatrix} X_1 & X_2 & \cdots & X_{N-1} & X_N \\ X_2 & X_3 & & X_N & X_{N+1} \\ \vdots & & & X_{N+1} & X_{N+2} \\ & & & & \ddots \\ X_N & N+1 & \cdots & X_{2N-2} & X_{2N-1} \end{bmatrix}, \tag{18}$$

where X_i are i.i.d., real-valued random variables with unit variance.

Theorem 4.10 [44] *We denote by M_i the asymptotic moments of order $2i$ of a Hankel matrix $\mathbf{H}$, then we have*

$$M_1 = 1$$
$$M_2 = \frac{8}{3}$$
$$M_3 = 14.$$

Toeplitz and Hankel matrices are structured matrices used for compressive wide-band spectrum sensing schemes [47,48] and for direction of arrival estimation [49].

5 Application
5.1 Power estimation

We consider a multi-user MIMO system where the received signal can be expressed by

$$\mathbf{y}_i = \mathbf{W}\mathbf{P}^{\frac{1}{2}}\mathbf{s}_i + \sigma\mathbf{n}_i \tag{19}$$

where $\mathbf{W}$, $\mathbf{P}$, $\mathbf{s}_i$, and $\mathbf{n}_i$ are respectively the $N \times K$ channel gain matrix, the $K \times K$ diagonal power matrix due to the different distances from which the users emit, the $K \times 1$ matrix of signals and the $N \times 1$ matrix

representing the noise with variance σ. In particular, $\mathbf{W}$, $\mathbf{s}_i$, $\mathbf{n}_i$ are independent standard, complex, Gaussian matrices and vectors. We are interested in estimating the power with which the users send information, from M observations (during which the channel gain matrix stays constant) of the vector $\mathbf{y}_i$. We consider the 2×2-matrix

$$\mathbf{P}^{\frac{1}{2}} = \begin{pmatrix} 1.5 & 0 \\ 0 & 0.5 \end{pmatrix} \qquad (20)$$

and we apply additive deconvolution at first, and then multiplicative deconvolution twice (each application takes care of one Gaussian matrix). We can estimate the eigenvalues of $\mathbf{P}$ when we have an increasing number L of observations of the matrix $\mathbf{Y} = [\mathbf{y}_1,...,\mathbf{y}_M]$, representing the signals received (we average across several block fading channels). Hence, we estimate the moments of the matrix $\mathbf{P}$ based only on the moments of the matrix $\mathbf{Y}\mathbf{Y}^H$. Knowing the moments of $\mathbf{P}$, we can estimate the eigenvalues using Newton-Girard formulas. When L increases, we get a prediction of the eigenvalues which is closer to the true eigenvalues of $\mathbf{P}$. Figure 4 illustrates the estimation of eigenvalues up to $L = 1000$ observations. The actual powers are 2.25 and 0.25, the variance σ of the noise is assumed to be equal to 0.1.

5.2 Understanding the network in a finite time

In cognitive MIMO networks, one must learn and control the "black box" (for instance the wireless channel) with multiple inputs and multiple outputs within a fraction of time and with finite energy. The fraction of time constraint is due to the fact that the channel (black box) changes over time. Of particular interest is the estimation of the capacity within the window of observation.

Let $\mathbf{y}$ be the output vector, $\mathbf{x}$ and $\mathbf{n}$ respectively the input signal and the noise vector, so that

$$\mathbf{y} = \mathbf{x} + \sigma\mathbf{n}. \qquad (21)$$

In the Gaussian case, the rate is given by

$$\begin{aligned} C &= H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}) \\ &= \log_2 \det(\pi e \mathbf{R}_Y) - \log_2 \det(\pi e \mathbf{R}_N) \\ &= \log_2 \left(\frac{\det(\mathbf{R}_Y)}{\det(\mathbf{R}_N)} \right) \end{aligned}$$

where $\mathbf{R}_Y$ is the covariance of the output signal and $\mathbf{R}_N$ is the covariance of the noise. Therefore, one can
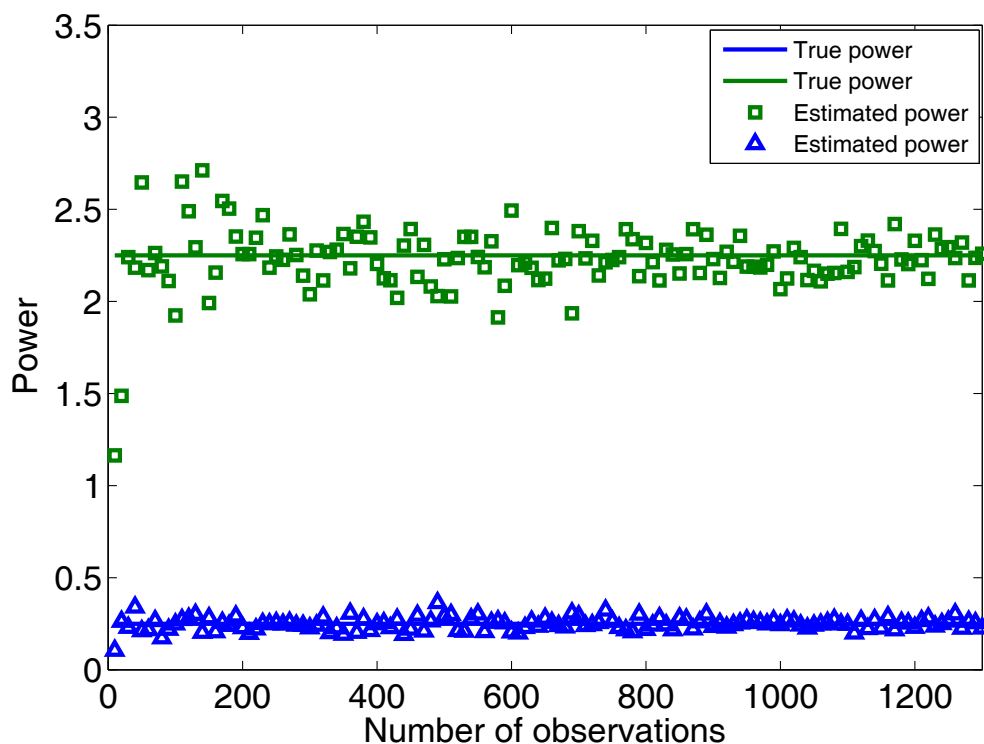

Figure 4 Estimation of the powers for the model (19), where the number L of observations increases, the sizes of the matrices are $K = N = M = 2$ and $\sigma = 0.1$. The actual powers are 2.25 and 0.25.

fully describe the information transfer in the system by knowing only the eigenvalues of $\mathbf{R}_Y$ and $\mathbf{R}_N$. Unfortunately, the receiver has only access to a limited number L of observations of $\mathbf{y}$ and not to the covariance of $\mathbf{R}_Y$. However, in the case where $\mathbf{x}$ and $\mathbf{n}$ are Gaussian vectors, $\mathbf{y}$ can be written as $\mathbf{y} = \mathbf{R}_Y^{\frac{1}{2}} \mathbf{u}$ where $\mathbf{u}$ is an i.i.d standard Gaussian vector. The problems falls therefore in the realm of inference with a correlated Wishart model $\left(\frac{1}{L} \sum_{i=1}^{L} \mathbf{y}_i \mathbf{y}_i^H = \mathbf{R}_i^{\frac{1}{2}} \frac{1}{L} \sum_{i=1}^{L} \mathbf{u}_i \mathbf{u}_i^H \mathbf{R}_Y^{\frac{1}{2}} \right)$.

In the simulation we have taken n as an i.i.d. standard Gaussian vector of dimension 2 and

$$\mathbf{R}_Y = \begin{pmatrix} 1.2^2 & 0 \\ 0 & 0.4^2 \end{pmatrix}. \qquad (22)$$

Considering L observations of (21), we stack the observations as columns in a compound matrix to get an unbiased estimate of the moments of $\mathbf{R}_Y$. In Figure 5, we can observe the convergence of the estimated capacity to the true one.

In Figure 6, we estimate the eigenvalues of the matrix $\mathbf{R}_Y$ versus the number of observations L. Once again, we observe the convergence of the estimated eigenvalues to the true eigenvalues.

5.3 Users detection

We consider M mobile users, each with a single antenna, communicating with a base station equipped with N receiving antennas, arranged as a uniform linear array (ULA). The $N \times 1$ received signal at the base station is given by

$$\mathbf{y} = \mathbf{V}\mathbf{P}^{\frac{1}{2}}\mathbf{x} + \mathbf{n} \qquad (23)$$

where $\mathbf{x}$ is the $M \times 1$ input signal transmitted by the M users, satisfying $\mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{I}_M$, and $\mathbf{n}$ is the additive Gaussian noise such that $\mathbb{E}[\mathbf{n}\mathbf{n}^H] = \sigma^2 \mathbf{I}_M$. We suppose that the components in $\mathbf{x}$ and $\mathbf{n}$ are independent. The matrix $\mathbf{P} = \mathrm{diag}(p_1,..., p_M)$ represents the power with which users send information. In the case of a line of sight between the mobile users and the base station, the $N \times M$ matrix $\mathbf{V}$ has the following form

$$\mathbf{V}(\theta) = \frac{1}{\sqrt{N}}[\mathbf{v}(\theta_1), \ldots, \mathbf{v}(\theta_M)],$$
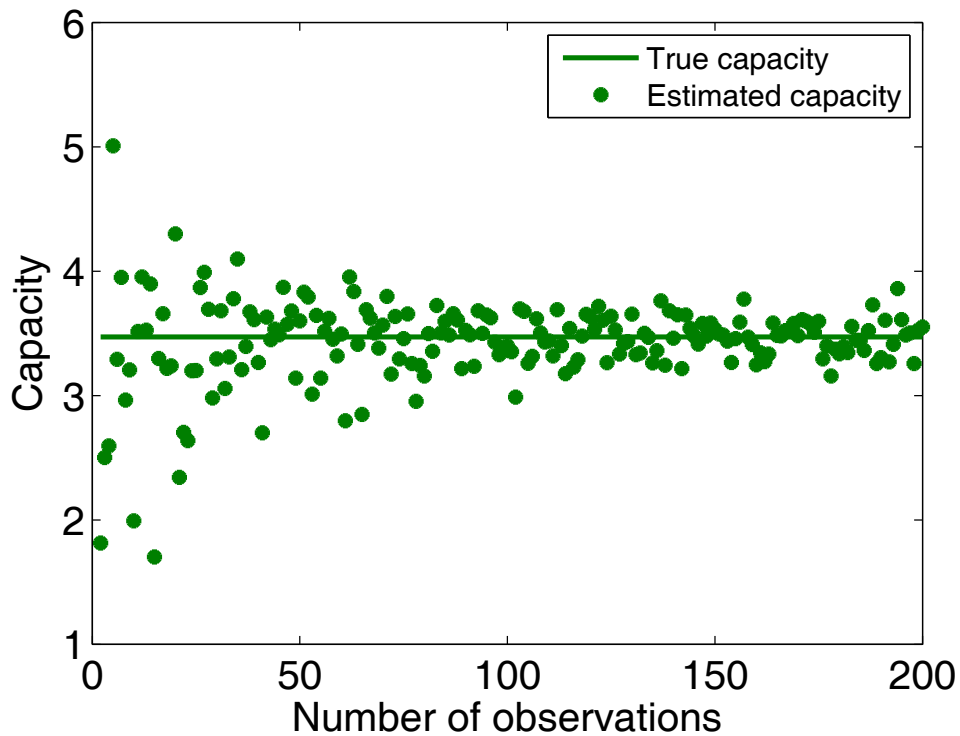

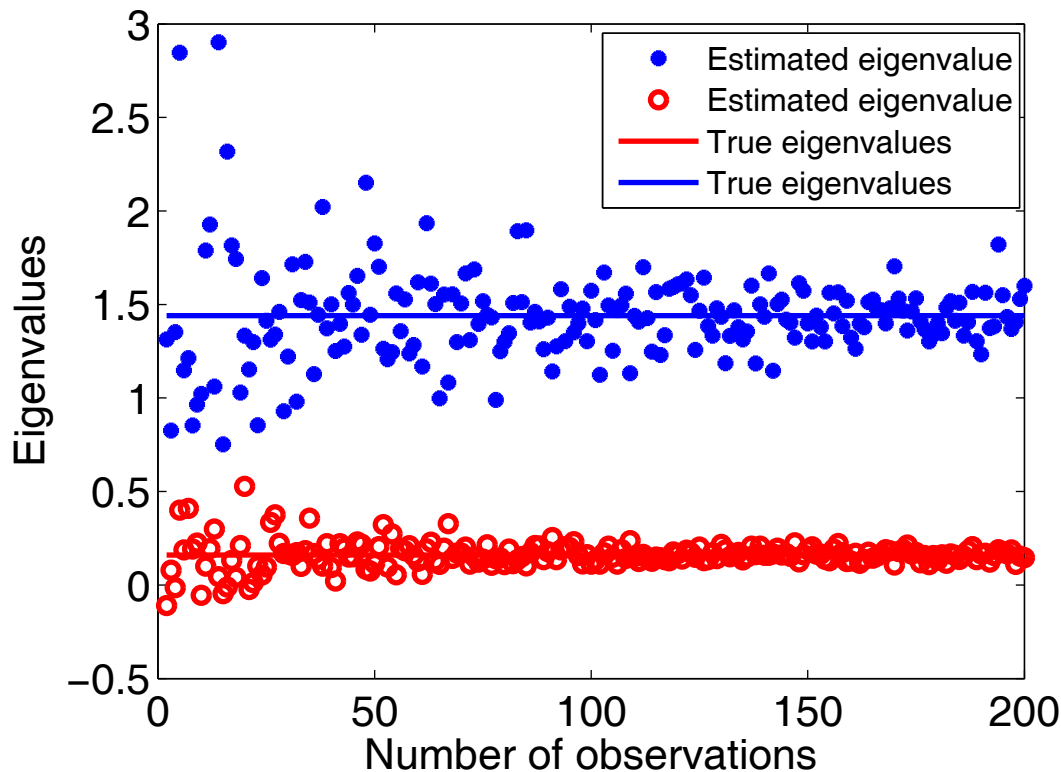**Figure 5 Estimation of the capacity for the model (21) up to $L = 500$ observations, with $\sigma = 0.5$.**

Figure 6 Estimation of the eigenvalues for the 2×2-matrix R_Y of (22) for various number of observations.

where

$$\mathbf{v}(\theta) = \left[1, \ e^{-j2\pi \frac{d}{\lambda} \sin(\theta)}, \ \ldots, \ e^{-j2\pi \frac{d}{\lambda} (N-1) \sin(\theta)} \right]^T,$$

and $\theta_1,\ldots, \theta_M$ are the angles of the users and are supposed to be i.i.d. and uniform on $[-\alpha, \alpha]$, d is the interspacing distance between the antennas of the ULA, and λ is the wave-length of the signal.

We are interested in estimating the number of users from a finite number K of observations of the received signal. Stacking all observations in a matrix

$$\mathbf{Y} = [\mathbf{y}_1, \ \ldots, \ \mathbf{y}_K] = \mathbf{VP}^{\frac{1}{2}} [\mathbf{x}_1, \ \ldots, \ \mathbf{x}_K] + [\mathbf{n}_1, \ \ldots, \ \mathbf{n}_K] \quad (24)$$

we can have access to the sample covariance matrix

$$\mathbf{W} = \frac{1}{K} \mathbf{YY}^H. \quad (25)$$

For estimation of the number of users M, we assume that the power distribution of $\mathbf{P}$ is known. Based on the knowledge of the power distribution, we are able to estimate the number of users in the system. Thanks to the

moments method it is possible to estimate the moments of the sample covariance matrix in (25) from the moments of the power matrix $\mathbf{P}$.

Proposition 5.1 [43]*Given the phase distribution ω and p_ω its density function, we define*

$$I_n = (2\pi)^{n-1} \int_{-\pi}^{\pi} p_\omega(x)^n dx.$$

Denoting the moments of $\mathbf{P}$ and the moments of the sample covariance matrix $\mathbf{W}$, respectively, by

$$P_i = \mathrm{tr}_M(\mathbf{P}^i) \quad (26)$$

$$W_i = \mathrm{tr}_N(\mathbf{W}^i) \quad (27)$$

then

$$W_1 = c_2 P_1 + \sigma^2$$

$$W_2 = c_2 p_2 + (c_2^2 I_2 + c_2 c_3)(P_1)^2 + 2\sigma^2(c_2 + c_3) P_1 + \sigma^4(1 + c_1)$$

$$W_3 = c_2 \left(1 + \frac{1}{K^2}\right) P_3 + \left(3 \left(1 + \frac{1}{K^2}\right) c_2^2 I_2 + 3 c_2 c_3\right) P_1 P_2 +$$
$$\left(c_2^3 I_3 \left(1 + \frac{1}{K^2}\right) + 3 c_2^2 c_3 I_2 + c_2 c_3^2\right) (P_1)^3 +$$
$$3\sigma^2 \left((1 + c_1) + \frac{c_1 c_2}{KM}\right) c_2 P_2 +$$
$$3\sigma^2 \left[\left((1 + c_1) + \frac{c_1 c_2}{KM}\right) c_2^2 I_2 + c_3 (c_3 + 2c_2)\right] (P_1)^2 +$$
$$3\sigma^4 \left(c_1^2 + 3c_1 + 1 + \frac{1}{K^2}\right) c_2 P_1 +$$
$$\sigma^6 \left(c_1^2 + 3c_1 + 1 + \frac{1}{K^2}\right),$$

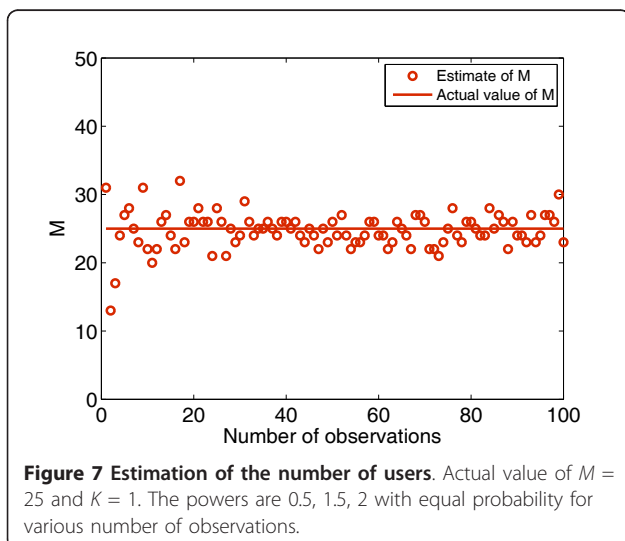*where* $\quad c_1 = \lim_{N \to \infty} \frac{N}{K}, \quad c_2 = \lim_{N \to \infty} \frac{M}{N} \quad$ *and* $c_3 = \lim_{N \to \infty} \frac{M}{K}$.

Knowing the matrix **P**, we can compute the moments P_i. From the moments P_i, using the above expressions is possible to get the moments W_i of the sample covariance matrix. We consider some candidate values of the number M of users. The estimate of M is chosen as the one which minimizes the sum of the square errors between the moments W_i and the moments of the observed sample covariance matrix.

In Figures 7 and 8 we have taken $N = 100$, $M = 30$, $\sigma = \sqrt{0.1}$, the distance $d = 1$ and the wavelength $\lambda = 2d$. We take the values of the matrix **P** equals to 0.5, 1.5, 2 with equal probability. Therefore just the first three moments are considered. We see that in Figure 8 the approximation is better even for a small number of observations.
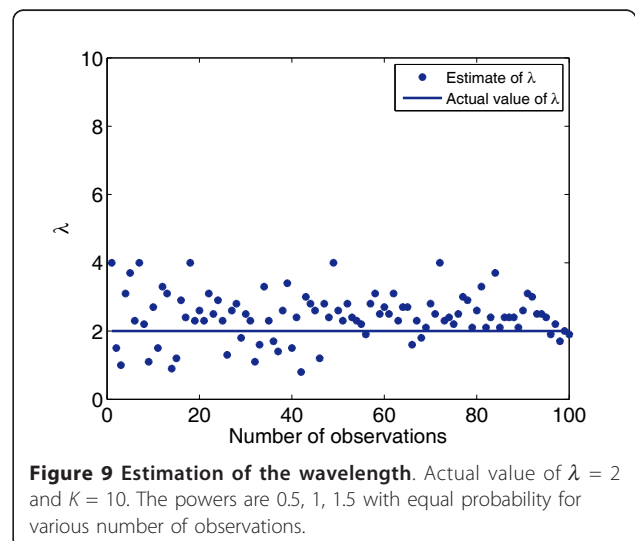
5.4 Wavelength detection
We know that in cognitive radio, mobile users are interested in understanding which band of transmission is occupied. Considering the model (23), we estimate the

Figure 8 Estimation of the number of users. Actual value of $M = 25$ and $K = 10$. The powers are 0.5, 1.5, 2 with equal probability for various number of observations.

wavelength λ using the moments method. As before, we consider some realizations of the sample covariance matrix and we estimate its moments in function of the moments of the power matrix, supposed to be known. In Figure 9, we consider $K = 10$, $L = 30$, $N = 100$, and $\sigma = \sqrt{0.1}$, in addition to $\lambda = 2$, $d = 1, \alpha = \frac{\pi}{4}$. Candidate values of the wavelength are taken in the interval [0.5, 4] with step 0.1 and the estimate value is chosen as the one which minimizes the sum of the squared errors of the first three moments between the moments W_i and the moments of the observed sample covariance matrix.

6 Conclusions and open problems
In the last decade, researchers and practitioners have devised cognitive radio as a possible solution for the problem of underutilization of the radio spectrum. These theoretical ad-vancements in cognitive radio research have

Figure 7 Estimation of the number of users. Actual value of $M = 25$ and $K = 1$. The powers are 0.5, 1.5, 2 with equal probability for various number of observations.

Figure 9 Estimation of the wavelength. Actual value of $\lambda = 2$ and $K = 10$. The powers are 0.5, 1, 1.5 with equal probability for various number of observations.

set up a solid base for practical applications and even further developments. However, still open questions need to be answered. In particular, in the current work, we use free probability theory, through the concept of free deconvolution, to attack the problem of retrieving useful information from the network with a limited number of observations. Free deconvolution, based on the moments method, has shown to be a interesting tool to tackle this problem. First, we show how the moments method works in the case where scalar random variables are considered. Then, since in practical situations systems are so complex that the parameters of interest need to be represented by vectors and matrices and can not be modeled by scalar random variables, we analyze the case where random matrices are considered. We propose algorithm method to compute the moments of various models such as Gaussian and Vandermonde matrices. Matlab codes for cognitive radio is developed to implement this algorithm method. In the applications free deconvolution framework can be used for retrieving relevant information such as power with which users send information, number of users, etc.

We have analyzed how free deconvolution framework works for random matrices and how random matrices behave differently depending on their structure. Different directions of research can be followed in this framework. In Vandermonde matrix model, the deconvolution techniques have been performed taking into account only diagonal matrices. It could be interesting to address the case of general deterministic matrices. In this way, correlation between users can be considered. The knowledge of the correlation could be a relevant element to improve the cooperation among the users in a cognitive system.

The extension of free deconvolution techniques to more general functions of matrices is a hard task. The difficulty is related to the fact that up to now there is not a general hypothesis that guarantees the application of free deconvolution to any random matrix. This extension can take into account more general models that represent more realistic situations.

For future perspective we would like also to take into account a second order analysis. The study of the covariance matrices can improve the accuracy of the estimations related to the free deconvolution framework.

Endnotes
[a]An algebra is unital if it contains a multiplicative identity element, i.e., an element $1_\mathcal{A}$ with the property $1_\mathcal{A} \cdot x = x \cdot 1_\mathcal{A} = x$ for all elements x of the algebra. [b]Note that in the following, when speaking of moments of matrices, we refer to the moments of the associated measure. [c]Let X be a

random variable and $M_X(z) := \sum_{m=0}^{\infty} \varphi(X^n) z^n$, we define the *S*-Transform of X as

$$S_X(z) = \frac{1+z}{z} M_X^{<-1>}(z)$$

where $(\cdot) < -1 >$ denotes the inverse (under composition) (i.e., $M_X^{<-1>}(M_X(z)) = M_X(M_X^{<-1>}(z)) = z$). [d]A standard complex Gaussian matrix **X** has i.i.d. complex Gaussian entries with zero mean and unit variance (in particular, the real and imaginary parts of the entries are independent, each with zero mean and variance 1/2).

Author details
[1]ETIS/ENSEA - University of Cergy Pontoise - CNRS, Cergy-Pontoise, France [2]Supélec, 3, rue Joliot-Curie, 91190 Gif-sur-Yvette, France

References
1. SEW Group, Report of the Spectrum Efficiency Working Group. Technical report, FCC (2002)
2. G Staple, K Werbach, The end of spectrum scarcity [spectrum allocation and utilization]. IEEE Spectrum. **41**(3), 48–52 (2004). doi:10.1109/MSPEC.2004.1270548
3. IJ Mitola, Cognitive radio for flexible mobile multimedia communications, in *IEEE International Workshop on Mobile Multimedia Communications 1999 (MoMuC '99)*, San Diego, California, USA, 3–10 (1999)
4. IJ Mitola, Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio, PhD Thesis, (Royal Institute of Technology (KTH) Stockholm, Sweden, May 2000)
5. I Akyildiza, I Leea, M Vuran, S Mohantya, Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey. Computer Networks. **50**(6), 2127–2159 (2006)
6. S Haykin, Cognitive radio: brain empowered wireless communications. IEEE J Sel Areas Commun. **23**(2), 201–220 (2005)
7. D Cabric, S Mishra, R Brodersen, Implementation issues in spectrum sensing for cognitive radios, in *Proc of 38th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove (CA), 772–776 (2004)
8. D Cabric, R Brodersen, Physical layer design issues unique to cognitive radio systems, in *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 759–763 (2005)
9. A Sahai, N Hoven, R Tandra, Some fundamental limits on cognitive radio, in *Proc Allerton Conference Communication, Control, and Computing* (2004)
10. Z Quan, S Cui, A Sayed, Optimal linear cooperation for spectrum sensing in cognitive radio networks. IEEE Journal of Selected Topics on Signal Processing. **2**(1), 28–40 (2008)
11. G Ganesan, Y Li, Cooperative spectrum sensing in cognitive radio, part I: two user networks. IEEE Transactions on Wireless Communication. **6**(6), 2204–2213 (2007)
12. G Ganesan, Y Li, Cooperative spectrum sensing in cognitive radio, in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DyS-PAN)*, Baltimore, Maryland, **2005**, 137–143 (2005)
13. A Ghasemi, E Sousa, Cooperative spectrum sensing in cognitive radio: The Cooperation-Processing Tradeoff. Wireless Communications and Mobile Computing. **7**(9), 1049–1060 (2007). doi:10.1002/wcm.480
14. S Mishra, A Sahai, RW Broderson, Cooperative sensing among cognitive radios, in *Proc of IEEE International Conference on Communications (ICC)*, Istanbul, Turkey, **4**, 1658–1663 (2006)
15. J Unnikrishnan, V Veeravalli, Cooperative sensing for primary detection in cognitive radio. IEEE Journal of Selected Topics on Signal Processing. **2**(1), 18–27 (2008)
16. I Nevat, G Peters, I Collings, J Yuan, Cooperative spectrum sensing with partial CSI, *IEEE Statistical Signal Processing Workshop*, Nice, France, 373–376 (2011)
17. Z Ji, K Liu, Cognitive radios for dynamic spectrum access-dynamic spectrum sharing: a game theoretical overview. IEEE Communications Magazine. **45**(5), 88–94 (2007)

18. J Neel, R Buehrer, B Reed, R Gilles, Game theoretic analysis of a network of cognitive radios, in *Procedings of 45th Midwest Symposium on Circuits and Systems*, Tulsa, Oklahoma, 409–412 (2002)

19. D Niyato, E Hossain, Competitive pricing for spectrum sharing in cognitive radio networks: dynamic game, inefficiency of Nash equilibrium, and collusion. IEEE Journal on Selected Areas in Communications. **26**(1), 192–202 (2008)

20. N Nie, C Comaniciu, Adaptive channel allocation spectrum etiquette for cognitive radio networks, in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, Baltimore, Maryland, USA, 269–278 (2005)

21. J Neel, J Reed, R Gilles, Convergence of cognitive radio networks, in *Proc IEEE Wireless Communications and Networking Conference (WCNC'04)*, Atlanta (GA), USA, 2250–2255 (2004)

22. N Devroye, P Mitran, V Tarokh, Achievable rates in cognitive radio channels. IEEE Transactions on Information Theory. **52**(5), 1813–1827 (2006)

23. A Jovicic, P Viswanath, Cognitive radio: an information-theoretic perspective. IEEE Transactions on Information Theory. **55**(9), 3945–3958 (2009)

24. S Jafar, S Srinivasa, Capacity limits of cognitive radio with distributed and dynamic spectral activity. IEEE Journal on Selected Areas in Communications. **25**, 529–537 (2007)

25. S Srinivasa, S Jafar, Cognitive radios for dynamic spectrum access-the throughput potential of cognitive radio: a theoretical perspective. IEEE Communications Magazine. **45**(5), 73–79 (2007)

26. A Goldsmith, S Jafar, I Maric, S Srinivasa, Breaking spectrum gridlock with cognitive radios: an information theoretic perspective. *Proceedings of IEEE*. **97**(5), 894–914 (2009)

27. Y Zeng, Y Liang, Maximum-minimum eigen value detection for cognitive radio, in *IEEE Communication Magazine IEEE 18th International Symposiumon Personal, Indoor and Mobile Radio Communications (PIMRC)*, Toronto, Canada, 1–5 (2011)

28. Y Zeng, Y Liang, Eigenvalue-based spectrum sensing algorithms for cognitive radio. IEEE Transaction on Communications. **57**(6), 1784–1793 (2009)

29. L Cardoso, M Debbah, P Bianchi, J Najim, Cooperative spectrum sensing using random matrix theory, in *IEEE 3rd International Wireless Pervasive Computing (ISWPC)*, Santorini, Greece, 334–338 (2008)

30. P Bianchi, J Najim, M Maida, M Debbah, Performance analysis of some eigen-based hypothesis tests for collaborative sensing, in *IEEE/SP 15th Workshop on Statistical Signal Processing (SSP)*, (Cardiff, United Kingdom, 2009), pp. 5–8

31. P Bianchi, M Debbah, M Maida, J Najim, Performance of statistical tests for single source detection using random matrix theory. IEEE Trans Inf Theory. **57**(4), 2400–2419 (2011)

32. T Cover, J Thomas, *Elements of Information Theory* Wiley, New York, (1991)

33. D Guo, S Shamai, S Verdú, Mutual information and minimum mean-square error in Gaussian channels. IEEE Transactions on Information Theory. **51**(4), 1261–1282 (2006)

34. D Palomar, S Verdú, Gradient of mutual information in linear vector Gaussian channels. IEEE Transactions on Information Theory. **52**(1), 141–154 (2005)

35. F Benaych-Georges, M Debbah, Free deconvolution: from theory to practice. submitted to IEEE Transactions on Information Theory (2008)

36. R Dozier, J Silverstein, On the empirical distribution of eigenvalues of large dimensional information plus noise-type matrices. J Multivar Anal. **98**(4), 678–694 (2007). doi:10.1016/j.jmva.2006.09.006

37. E Wigner, On the distribution of roots of certain symmetric matrices. Ann Math. **67**(2), 325–327 (1958). doi:10.2307/1970008

38. F Hiai, D Petz, *The semicircle law, free random variables and entropy - Mathematical Surveys and Monographs No. 77* (American Mathematical Society, Providence, RI, USA, 2006)

39. R Speicher, Free probability theory and non-crossing partitions. Lecture Notes 39e Seminaire Lotharingien de Combinatoire, Thurnau (April 1997)

40. Ø Ryan, A Masucci, S Yang, M Debbah, Finite dimensional statistical inference. IEEE Transactions on Information Theory. **57**(4), 2457–2473 (2011)

41. Ø Ryan, Tools for convolution with finite Gaussian matrices. (2009), http://ifi.uio.no/~oyvindry/finitegaussian/

42. R Seroul, D O'Shea, *Programming for Mathematicians*, (Springer, 2000)

43. Ø Ryan, M Debbah, Asymptotic behaviour of random Vandermonde matrices with entries on the unit circle. IEEE Transactions on Information Theory. **55**(7), 3115–3148 (2009)

44. Ø Ryan, M Debbah, Convolution operations arising from Vandermonde matrices. IEEE Transactions on Information Theory. **57**(7), 4647–4659 (2011)

45. L Cardoso, M Kobayashi, Ø Ryan, M Debbah, Vandermonde frequency division multiplexing for cognitive radio, in *Proceedings of the 9th Workshop on Signal Processing Advances in Wireless Communications, SPAWC*, Recife, Brazil, 421–425 (2008)

46. G Tucci, P Whiting, Eigenvalue results for large scale random Vandermonde Matrices With Unit Complex Entries. IEEE Transactions on Information Theory. **57**(6), 3938–3954 (2011)

47. Y Polo, Y Wang, A Pandharipande, G Leus, Compressive wide-band spectrum sensing, in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Taipei, Taiwan, 1–4 (2009)

48. Y Wang, A Pandharipande, Y Polo, G Leus, Distributed compressive wide-band spectrum sensing, in *IEEE Information Theory and Applications Workshop*, San Diego (CA), 178–183 (2009)

49. Y He, K Hueske, E Coersmeier, J Gotze, Efficient computation of joint direction-of-arrival and frequency estimation, in *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, Sarajevo, Bosnia and Herzegovina, 144–149 (2008)