Open Issues on the Statistical Spectrum Characterization of Random Vandermonde Matrices

Giusi Alfano, Mérouane Debbah, Oyvind Ryan

***Abstract*— Recently, analytical methods for finding moments of random Vandermonde matrices with entries on the unit circle have been proposed in the literature. Vandermonde matrices play an important role in signal processing and wireless applications, among which the multiple-antenna channel modeling, precoding or sparse sampling theory. Recent investigations allowed to extend the combinatorial approach usually exploited to characterize the spectral behavior of large random matrices with independent and identically distributed (i.i.d.) entries to Vandermonde structured matrices, under fairly broad assumptions on the entries distributions. While in several cases explicit expressions of the moments of the associated Gram matrix, as well as more advanced models involving the Vandermonde matrix could be provided, several issues are still open in the spectral behavior characterization, with applications either in signal processing (deconvolution, compressed sensing) and/or wireless communications (capacity estimation, topology information retrieving, etc).**

I. Problem description

A Vandermonde matrix [6] with entries on the unit circle has the following form

$$\mathbf{V} = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & \cdots & 1 \\ e^{-j\omega_1} & \cdots & e^{-j\omega_L} \\ \vdots & \ddots & \vdots \\ e^{-j(N-1)\omega_1} & \cdots & e^{-j(N-1)\omega_L} \end{pmatrix}; \qquad (1)$$

we will mainly focus on the case where ω_1,..., ω_L are i.i.d. random variables taking values on $[0, 2\pi)$. Throughout the paper, the ω_i will be called *phase distributions*, $\mathbf{V}$ will denote any Vandermonde matrix, of dimensions $N \times L$, with a given phase distribution. Let c denote the aspect ratio of the abovedefined matrix, i.e. $\lim_{L,N\to+\infty} \frac{L}{N} \to c$. Other models of particular interest are the generalized Vandermonde matrices, whose columns do not consist of uniformly distributed powers, namely

$$\mathbf{V} = \frac{1}{\sqrt{N}} \begin{pmatrix} e^{-j\lfloor Nf(0)\rfloor\omega_1} & \cdots & e^{-j\lfloor Nf(0)\rfloor\omega_L} \\ e^{-j\lfloor Nf(\frac{1}{N})\rfloor\omega_1} & \cdots & e^{-j\lfloor Nf(\frac{1}{N})\rfloor\omega_L} \\ \vdots & \ddots & \vdots \\ e^{-j\lfloor Nf(\frac{N-1}{N})\rfloor\omega_1} & \cdots & e^{-j\lfloor Nf(\frac{N-1}{N})\rfloor\omega_L} \end{pmatrix}, \qquad (2)$$

where f is called the power distribution, and is a map from $[0, 1)$ to itself. More general cases can also be considered, for instance by replacing f with a random variable, i.e.

$$\mathbf{V} = \frac{1}{\sqrt{N}} \begin{pmatrix} e^{-jN\lambda_1\omega_1} & \cdots & e^{-jN\lambda_1\omega_L} \\ e^{-jN\lambda_2\omega_1} & \cdots & e^{-jN\lambda_2\omega_L} \\ \vdots & \ddots & \vdots \\ e^{-jN\lambda_N\omega_1} & \cdots & e^{-jN\lambda_N\omega_L} \end{pmatrix}, \qquad (3)$$

with λ_i's, i.i.d. on $[0, 1)$, and also independent from the ω_j's. The basic quantities of interest are defined in the following:

Definition 1: Let us consider an $N \times N$ Hermitian matrix $\mathbf{A}$. The averaged empirical cumulative distribution function of the eigenvalues (also referred to as the averaged empirical spectral distribution (ESD)) of $\mathbf{A}$ is defined as

$$\mathsf{F}^N_{\mathbf{A}}(\lambda) = \frac{1}{N}\sum_{i=1}^{N} 1\mathbb{E}\left[\{\lambda_i(\mathbf{A}) \le \lambda\}\right],$$

where $\lambda_1(\mathbf{A}), \ldots, \lambda_N(\mathbf{A})$ are the eigenvalues of $\mathbf{A}$ and $1\{\cdot\}$ is the indicator function. If $\mathsf{F}^N_{\mathbf{A}}(\cdot)$ converges as $N \to \infty$, then the corresponding limit (asymptotic ESD, AESD) is denoted by $\mathsf{F}_{\mathbf{A}}(\cdot)$. The corresponding asymptotic probability density function is denoted by $f_{\mathbf{A}}(\cdot)$.

Explicit AESD expression is available only for a rather limited number of cases, depending either on the pdf of the entries of the random matrix under exam, or on the amount and structure of correlation between them. In alternative, implicit characterization, often through some integral transforms [14], plays an essential role in the analysis of systems which can be adequately represented through sums and products of random matrices.

In general, other than the AESD, one would like to have the joint eigenvalue distribution of matrices or some marginal distribution (e.g., the extremal eigenvalues distributions).

In particular, if we consider a probability measure ρ on the real line, which has moments of all order, and denoting by $(m_k(\rho) := \int t^k d\rho(t))_{k\ge 0}$ the sequence of its moments, we know that such a given sequence of moments $\{m_k, k \ge 0\}$ does not uniquely determine the associated probability distribution. A trivial sufficient condition, however, is the existence of the moment generating function[1]. In any case, for computing the eigenvalue distribution, one needs to determine the moments of all orders.

SUPELEC, Alcatel-Lucent Chair on Flexible Radio, Plateau de Moulon, 3 rue Joliot-Curie, 91192 GIF SUR YVETTE CEDEX, France and Centre of Mathematics for Applications, University of Oslo, P.O. Box 1053 Blindern, 0316 Oslo, Norway

[1] A more sophisticated one is the Carleman condition, which states that the sequence characterizes a distribution if the following holds:

$$\sum_{i=1}^{\infty} m_{2i}^{-\frac{1}{2i}} = \infty$$

Moments, as well as eigenvalues distribution computation, are of crucial interest in performance analysis of several wireless communications and signal processing systems. In the following, we will describe some application scenarios, stressing the dependence of the performance indices on the moments and, more in general, on some functions of the eigenvalues of (generalized and not) random Vandermonde matrices.

While in some cases computation have already been performed, the general result is not yet available, thus making the problem an actual challenge in application of random matrix theory to information- and/or estimation-theoretic analysis of wireless systems.

The main problem could be thus stated as

Problem 1: Compute, for any integer n, the expression of the so called *mixed moments*

$$M_n = \lim_{N\to\infty} E[tr_L(\mathbf{D}_1(N)\mathbf{V}^H\mathbf{V}\mathbf{D}_2(N)\mathbf{V}^H\mathbf{V} \cdots \times \mathbf{D}_n(N)\mathbf{V}^H\mathbf{V})], \quad (4)$$

for any sequence of matrices $\{\mathbf{D}_r(N)\}_{1\leq r\leq n}$ whose joint limit distribution exists.

This has been already performed in some relevant cases [10] for the phase distributions of $\mathbf{V}$, and for several values of n. In particular, results have been provided for the case of uniformly distributed phases, and it has also been proved that the uniform phase distribution minimizes the value of the mixed moments with respect to all phase distribution. The analysis has been carried out always under the assumption of diagonal $\{\mathbf{D}_r(N)\}_{1\leq r\leq n}$, independent on the Vandermonde matrices. Other open issues of major concern may be the following

Problem 2: Compute, for any integer n,

$$M_n = \lim_{N\to\infty} E[tr_L(\mathbf{D}_1(N)\mathbf{V}^H\mathbf{V}\mathbf{D}_2(N)\mathbf{V}^H\mathbf{V} \cdots \times \mathbf{D}_n(N)\mathbf{V}^H\mathbf{V})], \quad (5)$$

for non-diagonal $\{\mathbf{D}_r(N)\}_{1\leq r\leq n}$ whose joint limit distribution exists, and/or for any sequence of matrices $\mathbf{D}_n(N)$, whose joint limit distribution exists, with independent but not identically distributed columns.

Problem 3: Obtain explicit expressions for the probability density function (p.d.f.) of the AESD of a matrix of the type $\mathbf{V}\mathbf{V}^\dagger$.

Problem 4: Obtain explicit expressions for the p.d.f. of the extremal eigenvalues of $\mathbf{V}\mathbf{V}^\dagger$, say λ_{min} and λ_{max}.

A. Analytical difficulties and discussion

The solution to the abovementioned issues is relevant in several applications. Before to go into application details, we briefly notice some of the difficulties met up to now.

First of all, notice that the proposed evaluation method for the asymptotic mixed moments would fail both for the evaluation of mixed moments with non-diagonal matrices $\mathbf{D}$, as well as for the derivation of an explicit AESD expression. Indeed, as shown in [10], the support of such a function is not compact, hence the availability of the moments would not be sufficient to obtain the sought-for expression. Also, the extremal eigenvalues characterization, too, cannot be performed based on the tools developed in [10]; rather, a good candidate strategy would be to develop for Vandermonde matrices an analytical machinery, analogous to the Stieltjes, R and S transforms, usually exploited in free probability [5]. It is worth to stress that, in our case, the machinery would obviously depend on the explicit distribution of the entries. Indeed, one of the interesting features of free probability (when freeness is proved between matrices) is that the machinery is the same whatever the exact distribution of the entries of the matrices involved. Here, the *new* R and S transforms would be distribution entries dependent.

Finally, recall that we have been only concerned in analyzing the spectral behavior of random Vandermonde objects with entries lying on the unit circle. The generalization to the case of non unit modulus elements is even more challenging than previous issues. Indeed, the factor $\frac{1}{\sqrt{N}}$, as well as the assumption that the Vandermonde entries $e^{-j\omega_i}$ lie on the unit circle, are included in (1) to ensure that the analysis will give limiting asymptotic behaviour. Without this assumption, the problem at hand is more involved, since the rows of the Vandermonde matrix with the highest powers would dominate in the calculations of the moments for large matrices, and also grow faster to infinity than the $\frac{1}{\sqrt{N}}$ factor in (1), making asymptotic analysis difficult.

II. Motivation

The main motivation is to compute free deconvolution with random Vandermonde matrices, which is at the heart of cognitive (cooperative) inference techniques. Free deconvolution is a recent application of free probability to signal processing. It enables to compute the eigenvalues of involved models of sum or product of random matrices using combinatorial techniques. It has some some strong connections with other works on G-estimation [4].

As a straightforward example, suppose that $\mathbf{A}$ and $\mathbf{B}$ are independent large square Hermitian (or symmetric) random matrices, then under some very general conditions, free deconvolution enables to :

- 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 $\mathbf{AB}$ and $\mathbf{B}$.

The concept is even broader as it provides a method to retrieve the eigenvalue distribution of, say, $\mathbf{A}$, from any functional $f(\mathbf{A},\mathbf{B})$ and from $\mathbf{B}$[2].

Often more than in the case of free deconvolution, however, functionals of the eigenvalues of the channel matrix are enough to characterize the measure of information flow or to retrieve information.

Typically, as for the information rate, one is interested in

$$C_n = \frac{1}{n}\sum_{i=1}^{n} f(\lambda_i),$$

[2]The applications of such a deconvolution in wireless communications with free matrices [1, and references therein] have provided a useful framework when the number of observations is of the same order as the dimensions of the system, as is the case also in the G-estimation theory.

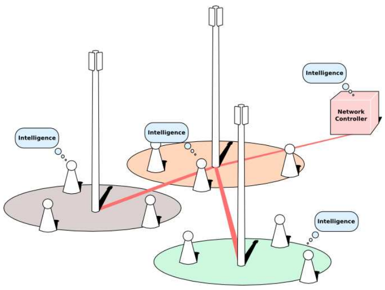

Fig. 1. Mobile Random networks

where f is a continous function ($\log(1+x)$ for example)[3]. Note that C_n is a random variable (due to the fact that the eigenvalues λ_i are random) and can be expressed as:

$$\begin{aligned} C_n &= \int f(\lambda)\frac{1}{n}\delta(\lambda-\lambda_i)d\lambda \quad (6)\\ &= \int f(\lambda)d\rho_n(\lambda) \quad (7)\end{aligned}$$

We call here ρ_n the eigenvalue distribution which turns out to be nothing else than a projection of the vector of eigenvalues $[\lambda_1, ..., \lambda_n]$ into a single quantity. Note that this approach is meaningful when one wants information about the whole set of values taken by the coordinated of the vector and not about each coordinate. It is therefore less explicit than the joint eigenvalue distribution but contains all what is required for our problem. ρ_n is also a random variable but as the dimension of the system grows, the behavior of the eigenvalue distribution becomes deterministic in many cases. We will denote it ρ. Of course, other projections may also be of interest in other applications such as the maximum of the vector $[\lambda_1, ..., \lambda_n]$.

In wireless intelligent random networks, devices are autonomous and should take optimal decisions based on their sensing capabilities (the number of samples they acquire, see also figure 1). 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 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.

However, once again, practical applications show that the limiting eigenvalue distributions in wireless communications depends only a subset of parameters, typically:

$$d\rho(\lambda) = \frac{1}{L}\sum_{i=1}^{L}\delta(\lambda-\lambda_i),$$

where L is small compared to p and is related to the problem of interest (class of users with a given power in multi-user systems, number of scatterers in an environment, rank of the MIMO matrix in multiple antenna systems for example).

In this case, the moments are related to the eigenvalues by the following relations:

[3]In general, the function f should have other constraints (bounded) but for clarity reasons, we do not go into more details

$$m_k(\rho) := \frac{1}{L}\sum_{i=1}^{L}\lambda_i^k. \quad (8)$$

As detailed in [11], [9], [3], one needs only to compute L moments to retrieve the eigenvalues in equation (8). This simplifies drastically the problems and favors a moment approach to the free deconvolution framework rather than deriving the explicit spectrum.

The *Newton-Girard Formulas* [11] can be used to retrieve the eigenvalues from the moments. These formulas state a relationship between the elementary symmetric polynomials

$$\Pi_j(\lambda_1, ..., \lambda_L) = \sum_{i_1<\cdots<i_j\leq L}\lambda_{i_1}\cdots\lambda_{i_j}, \quad (9)$$

and

$$\begin{aligned} S_p(\lambda_1, ..., \lambda_L) &= \sum_{i=1}^{L}\lambda_i^p\\ &= L\times m_p(\rho)\end{aligned}$$

through the recurrence relation

$$\begin{aligned} &(-1)^m m\Pi_m(\lambda_1, ..., \lambda_L)\\ +\ &\textstyle\sum_{k=1}^{m}(-1)^{k+m}S_k(\lambda_1, ..., \lambda_L)\Pi_{m-k}(\lambda_1, ..., \lambda_L) = 0. \end{aligned} \quad (10)$$

Interestingly, the characteristic polynomial

$$(\lambda-\lambda_1)\cdots(\lambda-\lambda_L)$$

(which roots provides the eigenvalues of the associated matrix) can be fully characterized as its $L-k$ coefficient is given by: $(-1)^k\Pi_k(\lambda_1, ..., \lambda_L)$. As $m_p(\rho)$ (we will show later on how these quantities can be computed) are known for $1\leq p\leq L$, (10) can be used repeatedly to compute $\Pi_m(\lambda_1, ..., \lambda_L)$, $1\leq m\leq L$.

For a given L, the previous algorithm works quite fine. However, in practice, L is not a priori known and a neat framework is still under study, considering even general cases where measures have densities (and are not a discrete sum of diracs). Non-optimal minimization methods for finding the appropriate L have been proposed in [9]. The general idea is to take $L = n$, in other words corresponding to the real dimension of the matrix. This of course incurs a burden in terms of complexity as the dimension of the matrix increases. In practice, as previously recalled, the problems ware well structured and therefore, one can use a finite small L which yields a good approximation of the distribution.

III. Examples in Wireless Random Networks

There are several examples in the literature which fall under the previous example. Some have been thoroughly studied in [9].

A. *Topology information*

The most simple example is the case where f is identity. This case is of practical interest when one performs channel sounding measurements. The transmitter sends an impulse on a given band to sound the environment. The channel response (or more precisely its power delay profile through the covariance of the received signal) contains information on the structure of the environment. By appropriate ray tracing techniques [7], [8], localization can be performed with a single receiver. The time-delayed channel impulse response can be written as:

$$x(\tau) = \sum_{k=1}^{L} \sigma_k s_k g(\tau - \tau_k),$$

where s_k are zero mean unit gaussian variables and σ_k are their associated variances (due to the topology), L represent the total number of scatterers and g is the transmit filter. In the frequency domain, the received vector for a given frequency f_i in the presence of noise, can be written:

$$y_i = x_i + n_i,$$

where $x_i = \sum_{k=1}^{L} s_k G(f_i) e^{-j2\pi f_i \tau_k}$; notice that, in general, $\tau_i < \tau_{i+1}$. In matrix form,

$$\mathbf{y} = \mathbf{R}^{\frac{1}{2}} \mathbf{s} + \mathbf{n},$$

where $\mathbf{R}^{\frac{1}{2}} = \mathbf{G\Theta\Sigma}$. Here, $\mathbf{G}$ is a diagonal matrix with entries $G(f_i)$, $\mathbf{\Theta}$ is $n \times L$ matrix with entries $e^{-j2\pi f_i \tau_k}$ and $\mathbf{\Sigma}$ is a diagonal matrix with entries σ_k. $\mathbf{s}$ and $\mathbf{n}$ are respectively $L \times 1$ and $n \times 1$ zero mean unit variance Gaussian vectors. The free deconvolution framework enables to infer on the L non-zero eigenvalues of $\mathbf{R}$ and therefore σ_k as suggested in [10].

B. Capacity and SINR estimation

In the case of cognitive TDD (Time Division Duplex) MIMO systems (the transmitter and the receiver have multi-antenna elements), the receiver would like to infer on the rate based only on the knowledge of the variance of the noise σ^2, but without any training systems and using only p samples. The TDD mode here enables channel reciprocity by providing the same rate on both ends. The received signal can be written as:

$$\mathbf{y}_i = \mathbf{H}\mathbf{s}_i + \mathbf{n}_i,$$

where $\mathbf{H}$ is the $n \times n$ MIMO matrix. The information rate is given by [12]:

$$\begin{aligned} C &= \log\det(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}\mathbf{H}^*) \\ &= \sum_{i=1}^{n} \log(1+\lambda_i). \end{aligned} \quad (11)$$

One can also be interested in the estimation of the SINR (Signal to Interference plus Noise Ratio) at the output of the MMSE receiver (if Bit Error Rate requirements are imposed) which is asymptotically given by [13]:

$$\begin{aligned} \text{SINR} &= \frac{1}{n}\,\text{trace}\left(\mathbf{H}\mathbf{H}^* + \sigma^2\right)^{-1} \\ &= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\lambda_i + \sigma^2}. \end{aligned}$$

In both cases, the number of non-zero eigenvalues is also limited to L in general as the medium (matrix $\mathbf{H}$) provides only a finite number of degrees of freedom. One can compute these eigenvalues by using the free deconvolution framework on $\mathbf{YY}^*$. Indeed, this has been already performed for matrices with i.i.d Gaussian entries. However, in the case of line of sight, $\mathbf{H}$ turns out to be a product of two Vandermonde matrices with independent, but not necessarily identically distributed entries.

C. Power estimation

In TDD heterogenous systems where a terminal is connected to several base stations, determining the power of the signal received from each base station is important as it will induce the adequate rate splitting between the different base stations. Suppose that each base station in the downlink has a given signature vector of size $n \times 1$ $\mathbf{h}_k$[4] (OFDM, CDMA) with random i.i.d components, the received signal can be written as:

$$\mathbf{y} = \sum_{k=1}^{L} \mathbf{h}_k \sqrt{P_k} s_k + n,$$

where P_k is the power received from each base station. L is the number of base stations, s_k is the signal transmitted by base station k and $\mathbf{y}$ and $\mathbf{n}$ are respectively the $n \times 1$ received signal and additive noise. It turns out here once again that one can infer on the powers P_k knowing only $\mathbf{YY}^*$ as shown in [3].

IV. Conclusion

The open problem of characterizing the spectral behavior of products of Vandermonde structured random matrices and matrices independent from them has been stated and recently computed partial solutions have been discussed. The relevance of the computation of the asymptotic moments of such products of random matrices, as well as of some marginal eigenvalues distributions, to signal processing applications has been stressed and illustrated through several examples.

References

[1] F. Benaych-Georges and M. Debbah. Free deconvolution: from theory to practice. *IEEE Transactions on Information Theory, submitted.*

[2] Leonardo S. Cardoso, M. Kobayashi, Mérouane Debbah, and Øyvind Ryan. Vandermonde frequency division multiplexing for cognitive radio. *CoRR*, abs/0803.0875, 2008.

[3] R. Couillet and M. Debbah. Free deconvolution for ofdm multicell snr detection. *PIMRC 2008, Cognitive Radio Workshop, Cannes, France*, Sept, 2008.

[4] V. L. Girko. *Theory of Stochastic Canonical Equations, Volumes I and II.* Kluwer Academic Publishers, 2001.

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

[6] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[7] J. P. Rossi and A. J. Levy. Propagation analysis in cellular environment with the help of models using ray theory. *43rd IEEE Vehicular Technology Conference*, pages 253–2567, 1993.

[8] J. P. Rossi, A. J. Levy, and J. C. Bic. A ray tracing method to study propagation in mobile urban environment. *23th G.A. of U.R.S.I., Prague (Cz)*, page 580, 1990.

[4] Indeed, in the framework of Vandermonde multiplexing, such $\mathbf{h}_k$ may also be a Vandermonde signature, see e.g. [2].

[9] 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.

[10] Ø. Ryan and M. Debbah. Asymptotic behaviour of random vandermonde matrices with entries on the unit circle. *revised, IEEE Trans. on Information Theory*, 2009.

[11] R. Seroul and D. O'Shea. *Programming for Mathematicians*. Springer.

[12] E. Telatar. Capacity of Multi-Antenna Gaussian Channels. *Eur. Trans. Telecomm. ETT*, 10(6):585–596, November 1999.

[13] D. Tse and S. Hanly. Linear multiuser receivers: Effective interference, effective bandwidth and user capacity. *IEEE Transactions on Information Theory*, 45(2):641–657, 1999.

[14] A.M. Tulino and S. Verdú. *Random Matrix Theory and Wireless Communications*. www.nowpublishers.com, 2004.