Asymptotic Behaviour of Random Vandermonde Matrices with Entries on the Unit Circle

Øyvind Ryan, *Member, IEEE*
Centre of Mathematics for Applications,
University of Oslo, P.O. Box 1053 Blindern, 0316 Oslo, Norway
Phone: +47 93 24 83 21
Fax: +47 22 85 43 49
Email: `oyvindry@ifi.uio.no`
Mérouane Debbah, *Senior Member, IEEE*
SUPELEC,
Alcatel-Lucent Chair on Flexible Radio, Plateau de Moulon,
3 rue Joliot-Curie,
91192 GIF SUR YVETTE CEDEX, France
Phone: +33 1 69 85 20 07
Fax: +33 1 69 85 12 59
Email: `merouane.debbah@supelec.fr`

***Abstract*—Analytical methods for finding moments of random Vandermonde matrices with entries on the unit circle are developed. Vandermonde Matrices play an important role in signal processing and wireless applications such as direction of arrival estimation, precoding or sparse sampling theory just to name a few. Within this framework, we extend classical freeness results on random matrices with i.i.d entries and show that Vandermonde structured matrices can be treated in the same vein with different tools. We focus on various types of Vandermonde matrices, namely Vandermonde matrices with or without uniformly distributed phase distributions, as well as generalized Vandermonde matrices (with non-uniform distribution of powers). In each case, we provide explicit expressions of the moments of the associated Gram matrix, as well as more advanced models involving the Vandermonde matrix. Comparisons with classical i.i.d. random matrix theory are provided and free deconvolution results are also discussed. We review some applications of the results to the fields of signal processing and wireless communications.**

***Index Terms*—Vandermonde matrices, Random Matrices, deconvolution, limiting eigenvalue distribution, MIMO.**

I. INTRODUCTION

Vandermonde matrices have had for a long time a central position in signal processing due to their connections with other important matrices in the field such as the FFT [1] or Hadamard [2] transforms to name a few. The matrices have various applications in different fields [3], [4], [5], [6]. The applied research has been somewhat tempered by the fact that very few theoretical results were available.

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

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

We will consider the case where ω_1,...,ω_L are independent and identically random variables taking values on $[0, 2\pi)$. Throughout the paper, the ω_i will be called *phase distributions*. Also, $\mathbf{V}$ will be used only to denote Vandermonde matrices with a given phase distribution, and the dimensions of the Vandermonde matrices will always be $N \times L$. Such matrices occur frequently in many applications, such as finance [3], signal array processing [7], [8], [9], [10], [11], ARMA processes [12], cognitive radio [4], security [6], wireless communications [13], and biology [5], and have been much studied. The main results are related to the distribution of the determinant of (1) [14]. The large majority of known results on the eigenvalues of the associated Gram matrix concern Gaussian matrices [15] or matrices with independent entries. Very few results are available in the literature on matrices whose structure is strongly related to the Vandermonde case [16], [17]. For the Vandermonde case, the results depend heavily on the distribution of the entries, and do not give any hint on the asymptotic behaviour as the matrices become large. In the realm of wireless channel modelling, [18] has provided some insight on the behaviour of the eigenvalues of random Vandermonde matrices for a specific case, without any formal proof. We prove here that the case is in fact more involved than what was claimed.

In many applications, N and L are quite large, and we may be interested in studying the case where both go to ∞ at a given ratio, with $\frac{L}{N} \to c$. Results in the literature say very

This project is partially sponsored by the project BIONET (INRIA).
This work was supported by Alcatel-Lucent within the Alcatel-Lucent Chair on flexible radio at SUPELEC as well as the ANR project SESAME.

little on the asymptotic behaviour of (1) under this growth condition. The results, however, are well known for other models. 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. In general, often the moments, not the moments of the determinants, are the quantities we seek. Results in the literature say also very little on the moments of Vandermonde matrices, and also on the mixed moments of Vandermonde matrices and matrices independent from them. This is in contrast to Gaussian matrices, where exact expressions [19] and their asymptotic behaviour [20] are known using the concept of freeness [20] which is central for describing the mixed moments.

Remarkably, the results in this paper show that, asymptotically, the moments of the Vandermonde matrices depend only on the ratio c and the phase distributions, and have explicit expressions. The derivation of the moments is a useful basis for performing deconvolution. Deconvolution for our purposes will mean retrieving the "moments" $tr_L((\mathbf{D}_1(N))^i), ..., tr_L((\mathbf{D}_n(N))^i)$, from the "mixed moments"

$$\begin{aligned} E[tr_L(& \mathbf{D}_1(N)\mathbf{V}^H\mathbf{V}\mathbf{D}_2(N)\mathbf{V}^H\mathbf{V} \\ & \cdots \times \mathbf{D}_n(N)\mathbf{V}^H\mathbf{V})]. \end{aligned} \tag{2}$$

In Section V we will see that this can be very useful in many applications, since the retrieved moments can give useful information about the system under study. Deconvolution has been handled in cases where the matrix $\mathbf{V}$ in (2) is replaced with a Gaussian matrix [21], [22], [19], [23]. Similarly flavored results will here be proved for Vandermonde matrices. Concerning the moments, it will be the asymptotic moments of random matrices of the form $\mathbf{V}^H\mathbf{V}$ which will be studied, where $(.)^H$ denotes hermitian transpose. We will also consider mixed moments of the form $\mathbf{D}\mathbf{V}^H\mathbf{V}$, where $\mathbf{D}$ is a square diagonal matrix independent from $\mathbf{V}$. As will be seen, the way the phase distribution influences these moments can be split into several cases. Uniform phase distribution plays a central role in that it minimizes the moments. When the phase distribution has a bounded density, a nice connection with the moments for uniform phase distribution can be given. When the density of the phase distribution has singularities, such as point masses, it turns out that the asymptotics of the moments change drastically.

We will also extend our results to generalized Vandermonde matrices, i.e. matrices where the columns do not consist of uniformly distributed powers. They are of the form

$$\mathbf{V} = \frac{1}{\sqrt{N}} \begin{pmatrix} e^{-j\lfloor Nf(0)\rfloor\omega_1} & \cdots & e^{-j\lfloor Nf(0)\rfloor\omega_L} \\ e^{-j\lfloor Nf(\frac{1}{N})\rfloor\omega_1} & \cdots & e^{-j\lfloor Nf(\frac{1}{N})\rfloor\omega_L} \\ \vdots & \ddots & \vdots \\ e^{-j\lfloor Nf(\frac{N-1}{N})\rfloor\omega_1} & \cdots & e^{-j\lfloor Nf(\frac{N-1}{N})\rfloor\omega_L} \end{pmatrix}, \tag{3}$$

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

$$\mathbf{V} = \frac{1}{\sqrt{N}} \begin{pmatrix} e^{-jN\lambda_1\omega_1} & \cdots & e^{-jN\lambda_1\omega_L} \\ e^{-jN\lambda_2\omega_1} & \cdots & e^{-jN\lambda_2\omega_L} \\ \vdots & \ddots & \vdots \\ e^{-jN\lambda_N\omega_1} & \cdots & e^{-jN\lambda_N\omega_L} \end{pmatrix}, \tag{4}$$

with the λ_i mutually independent and distributed as λ, taking values in $[0,1)$, and also independent from the ω_j. Generalized Vandermonde matrices are important for applications to finance [3]. The tools used for standard Vandermonde matrices in this paper will allow us to find the asymptotic behaviour of many generalized Vandermonde matrices.

While we provide the full computation of lower order moments, we also describe how the higher order moments can be computed. Tedious evaluation of many integrals is needed for this, but numerical methods can also be applied. Surprisingly, it turns out that the first three limit moments can be expressed in terms of the Marčhenko Pastur law [20], [24]. For higher order moments this is not the case, although we state an interesting inequality involving the Vandermonde limit moments and the moments of the classical Poisson distribution and the Marčhenko Pastur law, also known as the free Poisson distribution [20]. Note that the framework as well as the presented results are reminiscent of similar results concerning i.i.d. random matrices [25] which have shed light in the design of many important wireless communication problems such as CDMA [26], MIMO [27] or OFDM [28]. This contribution aims to to the same.

The paper is organized as follows: Section III states the main results of the paper. It starts with a general result for the mixed moments of Vandermonde matrices and matrices independent from them. We will differ between the case where the phase distribution ω in (1) are uniformly distributed on $[0,2\pi)$, and the more general cases. Results for the uniform phase distributions are stated next, and it turns out that one has nice expressions, for both the asymptotic moments, as well as for the lower order moments. Next we consider the more general case when ω has a continuous density, and show how the asymptotics can be described in terms of the case when ω is uniform. The case where the density of ω has singularities displays different asymptotic behaviour, and is handled separately. The section ends with results on generalized Vandermonde matrices, and mixed moments of (more than one) independent Vandermonde matrices. Section IV discusses our results and puts them in a general deconvolution perspective, comparing with other deconvolution results, such as those for Gaussian deconvolution. Section V presents some simulations and useful applications showing the implications of the presented results in various applied fields, and discusses the validity of the asymptotic claims in the finite regime. First we apply the presented Vandermonde deconvolution framework for wireless systems, where we estimate the number of paths, the transmissions powers of the users, the number of sources, and the wavelength. Next we apply the results on Vandermonde matrices to the very active field of sparse signal reconstruction. Interestingly, one can provide a general framework where only the sampling distribution matters asymptotically, and the sampling distribution can be estimated with the help of the presented results.

II. Random matrix background essentials

In the following, upper (lower boldface) symbols will be used for matrices (column vectors) whereas lower symbols will represent scalar values, $(.)^T$ will denote transpose operator, $(.)^\star$ conjugation and $(.)^H = \left((.)^T\right)^\star$ hermitian transpose. $\mathbf{I}_n$ will represent the $n \times n$ identity matrix. We let Tr be the (non-normalized) trace for square matrices, defined by,

$$Tr(\mathbf{A}) = \sum_{i=1}^{n} a_{ii},$$

where a_{ii} are the diagonal elements of the $n \times n$ matrix $\mathbf{A}$. We also let tr_n be the normalized trace, defined by $tr_n(\mathbf{A}) = \frac{1}{n}Tr(\mathbf{A})$.

Results in random matrix theory often refer to the empirical eigenvalue distribution of certain random matrices:

Definition 1: With the empirical eigenvalue distribution of an $N \times N$ hermitian random matrix $\mathbf{T}$ we mean the (random) function

$$F_{\mathbf{T}}^N(\lambda) = \frac{\#\{i|\lambda_i \leq \lambda\}}{N}, \tag{5}$$

where λ_i are the (random) eigenvalues of $\mathbf{T}$.

In the following, $\mathbf{D}_r(N), 1 \leq r \leq n$ will denote deterministic diagonal $L \times L$ matrices, where we implicitly assume that $\frac{L}{N} \to c$. We will assume that the $\mathbf{D}_r(N)$ have a joint limit distribution as $N \to \infty$ in the following sense:

Definition 2: We will say that the $\{\mathbf{D}_r(N)\}_{1 \leq r \leq n}$ have a joint limit distribution as $N \to \infty$ if the limit

$$D_{i_1,...,i_s} = \lim_{N\to\infty} tr_L\left(\mathbf{D}_{i_1}(N)\cdots\mathbf{D}_{i_s}(N)\right) \tag{6}$$

exists for all choices of $i_1, ..., i_s \in \{1, .., n\}$. For $\rho = \{W_1, ..., W_k\}$, with $W_i = \{w_{i1}, ..., w_{i|\rho_i|}\}$, we also define

$$\begin{aligned} D_{W_i} &= D_{i_{w_{i1}},...,i_{w_{i|\rho_i|}}} \\ D_\rho &= \prod_{i=1}^{k} D_{W_i}. \end{aligned}$$

Although the matrices $\mathbf{D}_i(N)$ are assumed to be deterministic matrices throughout the paper, all presented formulas extend naturally to the case when $\mathbf{D}_i(N)$ are random matrices independent from the Vandermonde matrices. The difference when the $\mathbf{D}_i(N)$ are random is that covariances of traces come into play. $D_{\{\{1\},\{2,3\}\}}$ would for instance be

$$\lim_{N\to\infty} E\left[tr_L\left(\mathbf{D}(N)\right) tr_L\left((\mathbf{D}(N))^2\right)\right],$$

which is the covariance of two traces when $\mathbf{D}(N)$ is centered ($E[tr_L\mathbf{D}(N)] = 0$) and random.

Most theorems in this paper will present expressions for various mixed moments, defined in the following way:

Definition 3: By a mixed moment we mean the limit

$$\begin{aligned} M_n = \lim_{N\to\infty} E[tr_L(\quad & \mathbf{D}_1(N)\mathbf{V}^H\mathbf{V}\mathbf{D}_2(N)\mathbf{V}^H\mathbf{V} \\ & \cdots \times \mathbf{D}_n(N)\mathbf{V}^H\mathbf{V})], \end{aligned} \tag{7}$$

whenever this exists.

A joint limit distribution of $\{\mathbf{D}_r(N)\}_{1 \leq r \leq n}$ is always assumed in the presented results on mixed moments. A second type of mixed moments will also be considered, where several independent Vandermonde matrices are used instead of the diagonal matrices $\mathbf{D}_r(N)$. Note that when $\mathbf{D}_1(N) = \cdots = \mathbf{D}_n(N) = \mathbf{I}_L$, the M_n compute to the asymptotic moments of the Vandermonde matrices themselves, defined by

$$V_n = \lim_{N\to\infty} E\left[tr_L\left(\left(\mathbf{V}^H\mathbf{V}\right)^n\right)\right].$$

V_n corresponds also to the limit moments of the empirical eigenvalue distribution $F_{\mathbf{V}^H\mathbf{V}}^N$ defined by (5), i.e.

$$V_n = \lim_{N\to\infty} E\left[\int \lambda^n dF_{\mathbf{V}^H\mathbf{V}}^N(\lambda)\right].$$

Similarly, when $\mathbf{D}_1(N) = \cdots = \mathbf{D}_n(N) = \mathbf{D}(N)$, we will also write

$$D_n = \lim_{N\to\infty} tr_L(D(N)^n).$$

Note that this is in conflict with the notation $D_{i_1,...,i_s}$, but the name of the index will resolve such conflicts.

To better understand the presented expressions for mixed moments, the notions of classical and free cumulants will be helpful. These are defined in terms of concepts from partition theory. We denote by $\mathcal{P}(n)$ the set of all partitions of $\{1, ..., n\}$, and use ρ as notation for a partition in $\mathcal{P}(n)$. The set of partitions will be equipped with the refinement order $\leq$, i.e. $\rho_1 \leq \rho_2$ if and only if any block of ρ_1 is contained within a block of ρ_2. Also, we will write $\rho = \{W_1, ..., W_k\}$, where W_j will be used repeatedly to denote the blocks of ρ, and let $|\rho| = k$ denote the number of blocks in ρ. We denote by 0_n the partition with n blocks, and by 1_n the partition with 1 block.

Free cumulants are defined in terms of noncrossing partitions.

Definition 4: A partition ρ is called noncrossing if whenever we have $i < j < k < l$ with $i \sim k$, $j \sim l$ ($\sim$ meaning belonging to the same block), we also have $i \sim j \sim k \sim l$ (i.e. i, j, k, l are all in the same block). The set of noncrossing partitions of $\{1, , , ., n\}$ is denoted $NC(n)$.
The noncrossing partitions have already shown their usefulness in expressing what is called the freeness relation [29] in a particularly nice way.

Definition 5: Assume that $\mathbf{A}_1, ..., \mathbf{A}_n$ are $L \times L$-random matrices. By the free cumulants of $\mathbf{A}_1, ..., \mathbf{A}_n$ we mean the unique set of multilinear functionals κ_r ($r \geq 1$) which satisfy

$$E\left[tr_L\mathbf{A}_1\cdots\mathbf{A}_n\right] = \sum_{\rho\in NC(n)} \kappa_\rho[\mathbf{A}_1, ..., \mathbf{A}_n], \tag{8}$$

where

$$\begin{aligned} \kappa_\rho[\mathbf{A}_1, ..., \mathbf{A}_n] &= \kappa_{W_1}[\mathbf{A}_1, ..., \mathbf{A}_n]\cdots\kappa_{W_k}[\mathbf{A}_1, ..., \mathbf{A}_n] \\ \kappa_{W_i}[\mathbf{A}_1, ..., \mathbf{A}_n] &= \kappa_{|W_i|}[\mathbf{A}_{w_{i1}}, ..., \mathbf{A}_{w_{i|\rho_i|}}], \end{aligned}$$

where $\rho = \{W_1, ..., W_k\}$, with $W_i = \{w_{i1}, ..., w_{i|\rho_i|}\}$.

By the classical cumulants of $\mathbf{A}_1, ..., \mathbf{A}_n$ we mean the unique set of multilinear functionals which satisfy (8) with $NC(n)$ replaced by the noncrossing partitions $\mathcal{P}(n)$.

We have restricted our definition of cumulants to those of random matrices, although their definition as they appear in Lecture 11 of [29] is in terms of more general probability spaces. (8) is also called the (free or classical) moment-cumulant formula. The importance of the free moment-cumulant formula comes from the fact that, had we replaced Vandermonde matrices with Gaussian matrices, it could help us compute the quantities $D_{i_1,...,i_s}$ [23]. For this, the cumulants of the Gaussian matrices are needed, which asymptotically have a very nice form. For Vandermonde matrices, it is not known what a useful definition of cumulants would be. However, from the calculations in Appendix A, it will turn out that the following quantities are useful in describing limit distributions of Vandermonde matrices.

Definition 6: Define

$$\begin{aligned} K_{\rho,\omega,N} = \ & \frac{1}{N^{n+1-|\rho|}} \times \\ & \int_{(0,2\pi)^{|\rho|}} \prod_{k=1}^{n} \frac{1-e^{jN(\omega_{b(k-1)}-\omega_{b(k)})}}{1-e^{j(\omega_{b(k-1)}-\omega_{b(k)})}} \\ & d\omega_1 \cdots d\omega_{|\rho|}, \end{aligned} \tag{9}$$

where $\omega_{W1},...,\omega_{W_{|\rho|}}$ are i.i.d. (indexed by the blocks of ρ), all with the same distribution as ω, and where $b(k)$ is the block of ρ which contains k (where notation is cyclic, i.e. $b(-1) = b(n)$). If the limit

$$K_{\rho,\omega} = \lim_{N\to\infty} K_{\rho,\omega,N}$$

exists, then it is called 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 our formulas for the moments are diferent 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. However, it is easy to prove from their definition that they do not exist if the density of ω has singularities (for instance when the density is a sum of point masses). On the other side, Theorem 3 will show that they exist when the same density is continuous.

In the following sections, we will also encounter the complementation map of Kreweras (p. 147 of [29]), which is an order-reversing isomorphism of $NC(n)$ onto itself. To define this we need the circular representation of a partition: We mark n equidistant points $1,...,n$ (numbered clockwise) on the circle, and form the convex hull of points lying in the same block of the partition. This gives us a number of convex sets H_i, equally many as there are blocks in the partition, which do not intersect if and only if the partition is noncrossing. Put names $\bar{1},...,\bar{n}$ on the midpoints of the $1,...,n$ (so that $\bar{i}$ is the midpoint of the segment from i to $i+1$). The complement of the set $\cup_i H_i$ is again a union of disjoint convex sets $\tilde{H}_i$.

Definition 7: The Kreweras complement of ρ, denoted $K(\rho)$, is defined as the partition on $\{\bar{1},...,\bar{n}\}$ determined by

$$i \sim j \text{ in } K(\rho) \iff \bar{i},\bar{j} \text{ belong to the same convex set } \tilde{H}_k.$$

An important property for the Kreweras complement we will use is that (p. 148 of [29])

$$|\rho| + |K(\rho)| = n+1.$$

III. Statement of main theorems

We first state the main result of the paper, which applies to Vandermonde matrices with any phase distribution. It restricts to the case when the expansion coefficients $K_{\rho,\omega}$ exist. Different versions of it adapted to different Vandermonde matrices will be stated in the succeeding sections.

Theorem 1: Assume that the $\{\mathbf{D}_r(N)\}_{1\le r\le n}$ have a joint limit distribution as $N\to\infty$. Assume also that all Vandermonde mixed moment expansion coefficients $K_{\rho,\omega}$ exist. Then the limit

$$\begin{aligned} M_n = \lim_{N\to\infty} E[tr_L(\ & \mathbf{D}_1(N)\mathbf{V}^H\mathbf{V}\mathbf{D}_2(N)\mathbf{V}^H\mathbf{V} \\ & \cdots \times \mathbf{D}_n(N)\mathbf{V}^H\mathbf{V})] \end{aligned} \tag{10}$$

also exists when $\frac{L}{N}\to c$, and equals

$$\sum_{\rho\in\mathcal{P}(n)} K_{\rho,\omega} c^{|\rho|-1} D_\rho. \tag{11}$$

The proof of Theorem 1 can be found in Appendix A. In the following, we will often make the substitutions

$$\begin{aligned} m_n &= (cM)_n = c \lim_{N\to\infty} E\left[tr_L\left(\left(\mathbf{D}(N)\mathbf{V}^H\mathbf{V}\right)^n\right)\right] && (12)\\ d_n &= (cD)_n = c \lim_{N\to\infty} tr_L\left(\mathbf{D}^n(N)\right), && (13) \end{aligned}$$

Several of the following theorems will be stated in terms of the scaled moments m_n, d_n rather than M_n, D_n. The reason for this is that the dependency on the matrix aspect ratio c often can be absorbed in m_n, d_n, so that the result itself can be expressed independently of c and refer only to m_n, d_n. The same usage of scaled moments has been applied for large Wishart matrices [23].

Typically, the first 5 moments can be expressed as:

$$\begin{aligned} m_1 &= K_1 d_1 && (14)\\ m_2 &= K_2 d_2 + K_{1,1} d_1^2 && (15)\\ m_3 &= K_3 d_3 + K_{2,1} d_2 d_1^2 + K_{1,1,1} d_1^3 && (16)\\ m_4 &= K_4 d_4 + K_{3,1} d_3 d_1 + K_{2,2} d_2^2 + K_{2,1,1} d_2 d_1^2 + \\ & \quad K_{1,1,1,1} d_1^4 && (17)\\ m_5 &= K_5 d_5 + K_{4,1} d_4 d_1 + + K_{3,2} d_3 d_2 + \\ & \quad K_{3,1,1} d_3 d_1^2 + K_{2,2,1} d_2^2 d_1 + K_{2,1,1,1} d_2 d_1^3 + \\ & \quad K_{1,1,1,1,1} d_1^5. && (18) \end{aligned}$$

Theorem 1 explains how convolution with Vandermonde matrices can be performed, and also provides us with an extension of the concept of free convolution to Vandermonde matrices. It also gives us means for performing deconvolution. Indeed, suppose $\mathbf{D}_1(N) = \cdots = \mathbf{D}_n(N) = \mathbf{D}(N)$, and that one knows all the moments M_n. One can then infer on the moments D_n by inspecting (11) for increasing values of n. For instance, the first two equations can also be written

$$\begin{aligned} D_{1_1} &= \frac{M_1}{K_{1_1,\omega}} \\ D_{1_2} &= \frac{M_2 - cK_{0_2,\omega} D_{0_2}}{K_{1_2,\omega}}, \end{aligned}$$

which gives us the first moments D_1 and D_2, since $D_{1_1} = D_1$, $D_{0_2} = D_1^2$, and $D_{1_2} = D_2$.

A. Uniformly distributed ω

Next we derive and analyze the Vandermonde mixed moment expansion coefficients for the case of uniform phase distribution. It turns out that the noncrossing partitions play a central role for such matrices, but that the role is somewhat different than the relation for freeness. We will let u denote the uniform distribution on $[0, 2\pi)$.

Proposition 1: Assume that the $\{\mathbf{D}_r(N)\}_{1\leq r\leq n}$ have a joint limit distribution as $N \to \infty$, Then the Vandermonde mixed moment expansion coefficient

$$K_{\rho,u} = \lim_{N\to\infty} K_{\rho,u,N}$$

exists for all ρ. Moreover, $0 < K_{\rho,u} \leq 1$, the $K_{\rho,u}$ are rational numbers for all ρ, and $K_{\rho,u} = 1$ if and only if ρ is noncrossing.

The proof of Proposition 1 can be found in Appendix B. we remark that the proof is similar to that given in the appendices of [16], where the mixed moment expansion coefficient are given an equivalent description. Due to Proposition 1, Theorem 1 guarantees that the asymptotic mixed moments (10) exist when $\frac{L}{N} \to c$ for uniform phase distribution, and are given by (11). $K_{\rho,u}$ are in general hard to compute for higher order ρ with crossings. It turns out that the following computations suffice to obtain the 7 first moments.

Proposition 2: The following holds:

$$\begin{aligned}
K_{\{\{1,3\},\{2,4\}\},u} &= \frac{2}{3}\\
K_{\{\{1,4\},\{2,5\},\{3,6\}\},u} &= \frac{1}{2}\\
K_{\{\{1,4\},\{2,6\},\{3,5\}\},u} &= \frac{1}{2}\\
K_{\{\{1,3,5\},\{2,4,6\}\},u} &= \frac{11}{20}\\
K_{\{\{1,5\},\{3,7\},\{2,4,6\}\},u} &= \frac{9}{20}\\
K_{\{\{1,6\},\{2,4\},\{3,5,7\}\},u} &= \frac{9}{20}.
\end{aligned}$$

The proof of Proposition 2 is given in Appendix C. Combining Proposition 1 and Proposition 2 into this form, we will prove the following:

Proposition 3: Assume that $\mathbf{D}_1(N) = \cdots = \mathbf{D}_n(N)$. When $\omega = u$, we have that

$$\begin{aligned}
m_1 &= d_1\\
m_2 &= d_2 + d_1^2\\
m_3 &= d_3 + 3d_2d_1 + d_1^3\\
m_4 &= d_4 + 4d_3d_1 + \frac{8}{3}d_2^2 + 6d_2d_1^2 + d_1^4\\
m_5 &= d_5 + 5d_4d_1 + \frac{25}{3}d_3d_2 + 10d_3d_1^2 +\\
&\quad \frac{40}{3}d_2^2d_1 + 10d_2d_1^3 + d_1^5\\
m_6 &= d_6 + 6d_5d_1 + 12d_4d_2 + 15d_4d_1^2 +\\
&\quad \frac{151}{20}d_3^2 + 50d_3d_2d_1 + 20d_3d_1^3 +\\
&\quad 11d_2^3 + 40d_2^2d_1^2 + 15d_2d_1^4 + d_1^6\\
m_7 &= d_7 + 7d_6d_1 + \frac{49}{3}d_5d_2 + 21d_5d_1^2 +\\
&\quad \frac{497}{20}d_4d_3 + 84d_4d_2d_1 + 35d_4d_1^3 +\\
&\quad \frac{1057}{20}d_3^2d_1 + \frac{693}{10}d_3d_2^2 + 175d_3d_2d_1^2 +\\
&\quad 35d_3d_1^4 + 77d_2^3d_1 + \frac{280}{3}d_2^2d_1^3 +\\
&\quad 21d_2d_1^5 + d_1^7.
\end{aligned}$$

Proposition 1 and Proposition 2 reduce the proof of Proposition 3 to a simple count of partitions. Proposition 3 is proved in Appendix D. To compute higher moments, $K_{\rho,u}$ must be computed for partitions of higher order also. The computations performed in Appendix C and D should convince the reader that this can be done, but that it is very tedious.

Following the proof of Proposition 1, we can also obtain formulas for the covariance of mixed moments of Vandermonde matrices. We state two examples of this. First we have that

$$\begin{aligned}
&\lim_{N\to\infty} E\left[tr_L\left(\left(\mathbf{D}(N)\mathbf{V}^H\mathbf{V}\right)^n\right)\left(tr_L\left(\mathbf{D}(N)\mathbf{V}^H\mathbf{V}\right)\right)^m\right]\\
&= \lim_{N\to\infty} E\left[tr_L\left(\left(\mathbf{D}(N)\mathbf{V}^H\mathbf{V}\right)^n\right)\right] D_1^m,
\end{aligned} \tag{19}$$

which follows immediately by noting that $tr_L\left(\mathbf{D}(N)\mathbf{V}^H\mathbf{V}\right) = tr_L\mathbf{D}(N) \to D_1$, since $\mathbf{V}^H\mathbf{V}$ has 1 in all diagonal entries. Secondly, we have that

$$\begin{aligned}
&c\lim_{N\to\infty} E\left[Tr\left(\left(\mathbf{D}(N)\mathbf{V}^H\mathbf{V}\right)^2\right) tr_L\left(\left(\mathbf{D}(N)\mathbf{V}^H\mathbf{V}\right)^2\right)\right]\\
&= \tfrac{4}{3}d_2^2 + 4d_2d_1^2 + 4d_3d_1 + d_4.
\end{aligned} \tag{20}$$

The proof for (20) is a bit more involved, and is therefore omitted. The proof relies on the same type of calculations as those in Appendix C. Following the proof of Proposition 1 again, we can also obtain exact expressions for moments of lower order random Vandermonde matrices with uniform phase distribution, not only the limit. We state these only for the first four moments.

Theorem 2: Assume $\mathbf{D}_1(N) = \mathbf{D}_2(N) = \cdots = \mathbf{D}_n(N)$, set $c = \frac{L}{N}$, and define

$$m_n^{(N,L)} = cE\left[tr_L\left(\left(\mathbf{D}(N)\mathbf{V}^H\mathbf{V}\right)^n\right)\right] \tag{21}$$

$$d_n^{(N,L)} = ctr_L\left(\mathbf{D}^n(N)\right). \tag{22}$$

When $\omega = u$ we have that

$$\begin{aligned}
m_1^{(N,L)} &= d_1^{(N,L)} \\
m_2^{(N,L)} &= \left(1 - N^{-1}\right) d_2^{(N,L)} + (d_1^{(N,L)})^2 \\
m_3^{(N,L)} &= \left(1 - 3N^{-1} + 2N^{-2}\right) d_3^{(N,L)} \\
&\quad +3\left(1 - N^{-1}\right) d_1^{(N,L)} d_2^{(N,L)} + (d_1^{(N,L)})^3 \\
m_4^{(N,L)} &= \left(1 - \frac{20}{3}N^{-1} + 11N^{-2} - \frac{37}{6}N^{-3}\right) d_4^{(N,L)} \\
&\quad + \left(4 - 12N^{-1} + 8N^{-2}\right) d_3^{(N,L)} d_1^{(N,L)} \\
&\quad + \left(\frac{8}{3} - 5N^{-1} + \frac{19}{6}N^{-2}\right) (d_2^{(N,L)})^2 \\
&\quad +6\left(1 - N^{-1}\right) d_2^{(N,L)} (d_1^{(N,L)})^2 + (d_1^{(N,L)})^4.
\end{aligned}$$

Theorem 2 is proved in Appendix E. Exact formulas for the higher order moments also exist, but they become increasingly complex, as entries for higher order terms L^{-k} also enter the picture. These formulas are also harder to prove for higher order moments. In many cases, exact expressions are not what we need: first order approximations (i.e. expressions where only the L^{-1}-terms are included) can suffice for many purposes. In Appendix E, we explain how the simpler case of these first order approximations can be computed. It seems much harder to prove a similar result when the phase distribution is not uniform.

The final result we address for the uniform phase distribution is the following:

Proposition 4: The asymptotic mean eigenvalue distribution of a Vandermonde matrix with uniform phase distribution has unbounded support.

Proposition 4 is proved in Appendix F.

B. ω with continuous density

The following result tells us that the limit $K_{\rho,\omega}$ exists for many ω, and also gives a useful expression for them in terms of the density of ω, and $K_{\rho,u}$.

Theorem 3: The Vandermonde mixed moment expansion coefficients $K_{\rho,\omega} = \lim_{N\to\infty} K_{\rho,\omega,N}$ exist whenever the density p_ω of ω is continuous on $[0, 2\pi)$. If this is fulfilled, then

$$K_{\rho,\omega} = K_{\rho,u}(2\pi)^{|\rho|-1}\left(\int_0^{2\pi} p_\omega(x)^{|\rho|}dx\right). \quad (23)$$

The proof is given in Appendix G. In Section V, several examples are provided where the integrals (23) are computed. An important consequence of Theorem 3 is the following, which gives the uniform phase distribution an important role.

Proposition 5: Let $\mathbf{V}_\omega$ denote a Vandermonde matrix with phase distribution ω, and set

$$V_{\omega,n} = \lim_{N\to\infty} E\left[tr_L\left(\left(\mathbf{V}_\omega^H\mathbf{V}_\omega\right)^n\right)\right].$$

Then we have that

$$V_{u,n} \le V_{\omega,n}.$$

The proof is given in Appendix H. An immediate consequence of this and Proposition 4 is that all phase distributions, not only uniform phase distribution, give Vandermonde matrices with unbounded mean eigenvalue ditrsibutions in the limit. Besides providing us with a deconvolution method for finding the mixed moments of the $\{\mathbf{D}_r(N)\}_{1\le r\le n}$, Theorem 3 also provides us with a way of inspecting the phase distribution ω, by first finding the moments of the density, i.e. $\int_0^{2\pi} p_\omega(x)^k dx$. However, note that we can not expect to find the density of ω itself, only the density of the density of ω. To see this, define

$$Q_\omega(x) = \mu\left(\{x|p_\omega \le x\}\right)$$

for $0 \le x \le \infty$, where μ is uniform measure on the unit circle. Write also $q_\omega(x)$ as the corresponding density, so that $q_\omega(x)$ is the density of the density of ω. Then it is clear that

$$\int_0^{2\pi} p_\omega(x)^{|\rho|}dx = \int_0^\infty x^n q_\omega(x)dx. \quad (24)$$

These quantities correspond to the moments of the measure with density q_ω, which can help us obtain the density q_ω itself. However, the density p_ω can not be obtained, since we see that any reorganization of its values which do not change the density q_ω will provide the same values in (24).

C. ω with density singularities

The asymptotics of Vandermonde matrices are different when the density of ω has singularities, and depends on the density growth rates near the singular points. It will be clear from these results that one can not perform deconvolution for such ω to obtain the higher order moments of the $\{\mathbf{D}_r(N)\}_{1\le r\le n}$, as only their first moment can be obtained. The asymptotics are first described for ω with atomic density singularities, as this is the simplest case to prove. After this, densities with polynomial growth rates near the singularities are addressed.

Theorem 4: Assume that $p_\omega = \sum_{i=1}^r p_i\delta_{\alpha_i}$ is atomic (where $\delta_{\alpha_i}(x)$ is dirac measure (point mass) at α_i), and denote by $p^{(n)} = \sum_{i=1}^r p_i^n$ the. Then

$$\begin{aligned}
\lim_{N\to\infty} \quad E[Tr(\quad & \mathbf{D}_1(N)\frac{1}{N}\mathbf{V}^H\mathbf{V}\mathbf{D}_2(N)\frac{1}{N}\mathbf{V}^H\mathbf{V} \\
& \cdots \times \mathbf{D}_n(N)\frac{1}{N}\mathbf{V}^H\mathbf{V})] \\
= \quad & c^{n-1}p^{(n)} \lim_{N\to\infty}\prod_{i=1}^n tr_L\left(\mathbf{D}_i(N)\right).
\end{aligned}$$

Note here that the non-normalized trace is used.

The proof can be found in Appendix I. In particular, Theorem 4 states that the asymptotic moments of $\frac{1}{N}\mathbf{V}^H\mathbf{V}$ coincide with the moments of p_ω, up to the scaling factor c^{n-1}. The theorem is of great importance for the estimation of the angles α_i and the point masses p_i in our Vandermonde deconvolution framework. In blind seismic and telecommunication applications, one would like to detect the angles α_i through deconvolution. Unfortunately, Theorem 4 tells us that this is impossible, since the $p^{(n)}$ (which are moments which we can find through deconvolution), do not depend on them (this parallels Theorem 3, since also there we could not recover the density p_ω itself). Having found the $p^{(n)}$ through deconvolution, one can, however, find the point masses p_i, by solving for $p_1, p_2, ...$ in the Vandermonde equation

$$\begin{pmatrix} p_1 & p_2 & \cdots & p_r \\ p_1^2 & p_2^2 & \cdots & p_r^2 \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \end{pmatrix} = \begin{pmatrix} p^{(1)} \\ p^{(2)} \\ \vdots \end{pmatrix},$$

even if the number of atoms may be unknown.

The case when the density has non-atomic singularities is more complicated. We provide only the following result, which addresses the case when the density has polynomic growth rate near the singularities.

Theorem 5: Assume that

$$\lim_{x\to\alpha_i} |x-\alpha_i|^s p_\omega(x) = p_i \text{ for some } 0 < s < 1$$

for a set of points $\alpha_1, ..., \alpha_r$, with p_ω continuous for $\omega \neq \alpha_1, ..., \alpha_r$. Then

$$\begin{aligned} \lim_{N\to\infty} \quad E[Tr(\quad & \mathbf{D}_1(N)\frac{1}{N^s}\mathbf{V}^H\mathbf{V}\mathbf{D}_2(N)\frac{1}{N^s}\mathbf{V}^H\mathbf{V} \\ & \cdots \times \mathbf{D}_n(N)\frac{1}{N^s}\mathbf{V}^H\mathbf{V})] \\ = \quad & c^{n-1}q^{(n)} \lim_{N\to\infty} \prod_{i=1}^n tr_L\left(\mathbf{D}_i(N)\right) \end{aligned}$$

where

$$\begin{aligned} q^{(n)} \;=\; & \left(2\Gamma(1-s)\cos\left(\frac{(1-s)\pi}{2}\right)\right)^n p^{(n)} \times \\ & \int_{[0,1]^n} \prod_{k=1}^n \frac{1}{|x_{k-1}-x_k|^{1-s}} dx_1\cdots dx_n, \end{aligned} \quad (25)$$

and $p^{(n)} = \sum_i p_i^n$. Note here that the non-normalized trace is used.

The proof can be found in Appendix J. Also in this case it is only the point masses p_i which can be found through deconvolution, not the singularity locations α_i. Note that the integral in (25) can also be written as an m-fold convolution. Similarly, the definition of $K_{\rho,\omega,N}$ given by (9) can also be viewed as a 2-fold convolution when ρ has two blocks, and as a 3-fold convolution when ρ has three blocks (but not for ρ with more than 3 blocks).

A very useful application of Theorem 5 is the case when $\omega = \sin(x)$, with x uniformly distributed. The density will then be of the form $\frac{d\arcsin(\omega)}{d\omega} = \frac{1}{\sqrt{1-\omega^2}}$, which goes to infinity near $\omega = \pm 1$ (which correspond to $x = \pm\pi/2$) at rate $x^{-1/2}$. Theorem 5 thus applies with $s = 1/2$. For this case, however, the "edges" at $\pm\pi/2$ are never reached in practice. Indeed, in array processing [30], the antenna array is a sector antenna which scans an angle interval which never includes the edges. We can therefore restrict ω in our analysis to clusters of intervals $U_i[\alpha_i, \beta_i]$ not containing ± 1, for which the results of Section III-B suffice. In this way, we also avoid the computation of the cumbersome integral (25).

D. Generalized Vandermonde matrices

We will define mixed moment expansion coefficients for generalized Vandermonde matrices also. The difference is that, while we in Definition 6 simplified using the geometric sum formula, we can not do this now since we do not assume uniform power distribution anymore.

To define expansion coefficients for generalized Vandermonde matrices of the form (3), define first functions f_N from $[0, N-1]$ to $[0, N-1]$ by $f_N(r) = \lfloor f\left(\frac{r}{N}\right)\rfloor$. Let p_{f_N} be the corresponding density for f_N. The procedure is similar for matrices of the form (4). The following definition captures both cases:

Definition 8: For (3) and (4), define

$$\begin{aligned} K_{\rho,\omega,f,N} = \quad & \frac{1}{N^{1-|\rho|}} \times \\ & \int_{(0,2\pi)^{|\rho|}} \prod_{k=1}^n \left(\sum_{r=0}^{N-1} p_{f_N}(r) e^{jr(\omega_{b(k-1)}-\omega_{b(k)})}\right) \\ & d\omega_1\cdots d\omega_{|\rho|} \\ K_{\rho,\omega,\lambda,N} = \quad & \frac{1}{N^{1-|\rho|}} \times \\ & \int_{(0,2\pi)^{|\rho|}} \prod_{k=1}^n \left(\int_0^1 e^{jN\lambda(\omega_{b(k-1)}-\omega_{b(k)})} d\lambda\right) \\ & d\omega_1\cdots d\omega_{|\rho|}, \end{aligned} \quad (26)$$

where $\omega_{W1}, ..., \omega_{W_{|\rho|}}$ are as in definition 6.

If the limits

$$\begin{aligned} K_{\rho,\omega,f} &= \lim_{N\to\infty} K_{\rho,\omega,f,N} \\ K_{\rho,\omega,\lambda} &= \lim_{N\to\infty} K_{\rho,\omega,\lambda,N}, \end{aligned}$$

exist, then they are called *Vandermonde mixed moment expansion coefficients*.

Note that (1) corresponds to (3) with $f(x) = x$. The following result holds:

Theorem 6: Theorem 1 holds also with Vandermonde matrices (1) replaced with generalized Vandermonde matrices on either form (3) or (4), and with $K_{\rho,\omega}$ replaced with either $K_{\rho,\omega,f}$ or $K_{\rho,\omega,\lambda}$.

The proof follows the same lines as those in the proof of Theorem 1 in Appendix A, and is therefore only explained briefly at the end of that appendix.

As for matrices of the form (1), it is the case of uniform phase distribution which is most easily described how to compute for generalized Vandermonde matrices also. Appendix B shows how the computation of $K(\rho, u)$ boils down to computing certain integrals. The same comments are valid for matrices of the form (3) or (4) in order to compute $K_{\rho,\omega,f}$ and $K_{\rho,\omega,\lambda}$. This is commented at the end of that appendix.

We will not consider generalized Vandermonde matrices with density singularities.

E. The joint distribution of independent Vandermonde matrices

In the case when many independent random Vandermonde matrices are involved, the following holds:

Theorem 7: Assume that the $\{\mathbf{D}_r(N)\}_{1\leq r\leq n}$ have a joint limit distribution as $N\to\infty$. Assume also that $\mathbf{V}_1, \mathbf{V}_2, ...$ are independent Vandermonde matrices with the same phase distribution ω, and that the density of ω is continuous. Then the limit

$$\begin{aligned} \lim_{N\to\infty} E[tr_L(\quad & \mathbf{D}_1(N)\mathbf{V}_{i_1}^H\mathbf{V}_{i_2}\mathbf{D}_2(N)\mathbf{V}_{i_2}^H\mathbf{V}_{i_3} \\ & \cdots \times \mathbf{D}_n(N)\mathbf{V}_{i_n}^H\mathbf{V}_{i_1})] \end{aligned}$$

also exists when $\frac{L}{N}\to c$, and equals

$$\sum_{\rho\leq\sigma\in\mathcal{P}(n)} K_{\rho,\omega} c^{|\rho|-1} D_\rho, \quad (27)$$

where σ is the partition where k and j are in the same block if and only if $i_k = i_j$.

The proof of Theorem 7 can be found in Appendix K. That appendix also contains some remarks on the case when the matrices $\mathbf{D}_i(N)$ are placed in different positions relative to the Vandermonde matrices. From Theorem 7, the following corolllary is immediate:

Corollary 1: The first three mixed moments

$$V_n^{(2)} = \lim_{N\to\infty} E\left[tr_L\left(\left(\mathbf{V}_1^H\mathbf{V}_2\mathbf{V}_2^H\mathbf{V}_1\right)^n \right)\right]$$

of independent Vandermonde matrices $\mathbf{V}_1, \mathbf{V}_2$ are given by

$$\begin{aligned} V_1^{(2)} &= I_2 \\ V_2^{(2)} &= \frac{2}{3}I_2 + 2I_3 + I_4 \\ V_3^{(2)} &= \frac{11}{20}I_2 + 4I_3 + 9I_4 + 6I_5 + I_6, \end{aligned}$$

where

$$I_k = (2\pi)^{|\rho|-1}\left(\int_0^{2\pi} p_\omega(x)^{|\rho|}dx\right).$$

In particular, when the phase distributions are uniform, the first three mixed moments are given by

$$\begin{aligned} V_1^{(2)} &= 1 \\ V_2^{(2)} &= \frac{11}{3} \\ V_3^{(2)} &= \frac{411}{20} \end{aligned}$$

The results here can also be extended to the case with independent Vandermonde matrices with different phase distributions:

Theorem 8: Assume that $\{\mathbf{V}_i\}_{1\leq i\leq s}$ are independent Vandermonde matrices, where $\mathbf{V}_i$ has continuous phase distribution ω_i. Denote by p_{ω_i} the density of ω_i. Then equation (27) still holds, with $K_{\rho,\omega}$ replaced by

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

where ρ_i is the partition of σ_i consisting of the blocks of ρ contained in σ_i.

The proof is omitted, as it is a straightforward extension of the proofs of Theorem 3 and Theorem 7.

IV. Discussion

We have already explained that one can perform deconvolution with Vandermonde matrices in a similar way to how one can perform deconvolution for Gaussian matrices. We have, however, also seen that there are many differences.

A. Convergence rates

In [19], almost sure convergence of Gaussian matrices was shown by proving exact formulas for the distribution of lower order Gaussian matrices. These deviated from their limits by terms of the form $1/L^2$. In Theorem 2, we see that terms of the form $1/L$ are involved. This slower rate of convergence may not be enough to make a statement on whether we have almost sure convergence of random Vandermonde matrices. However, [31] shows some almost sure convergence properties for certain Hankel and Toeplitz matrices. These matrices are seen in that paper to have similar combinatorial descriptions for the moments, when compared to Vandermonde matrices in this paper. Therefore, it may be the case that the techniques in [31] can be generalized to address almost sure convergence of Vandermonde matrices also. Figures 1, 2 show the speed of convergence of the moments of Vandermonde matrices (with uniform phase distributions) towards the asymptotic moments as the matrix dimensions grow, and as the number of samples grow. The differences between the asymptotic moments and the exact moments are also shown. To be more precise, the MSE of figures 1 and 2 is computed as follows:

1) K samples $\mathbf{V}_i$ are independently generated using (1).
2) The 4 first sample moments $\hat{v}_{ji} = \frac{1}{L}tr_n\left(\left(\mathbf{V}_i^H\mathbf{V}_i\right)^j\right)$ ($1\leq j\leq 4$) are computed from the samples.
3) The 4 first estimated moments $\hat{V}_j$ are computed as the mean of the sample moments, i.e. $\hat{V}_j = \frac{1}{K}\sum_{i=1}^K \hat{m}_{ji}$.
4) The 4 first exact moments E_j are computed using Theorem 2.
5) The 4 first asymptotic moments A_j are computed using Proposition 3.
6) The mean squared error (MSE) of the first 4 estimated moments from the exact moments is computed as $\sum_{j=1}^4\left(\hat{V}_j - E_j\right)^2$.
7) The MSE of the first 4 exact moments from the asymptotic moments is computed as $\sum_{j=1}^4 (E_j - A_j)^2$.

Figures 1 and 2 are in sharp contrast with Gaussian matrices, as shown in Figure 3. First of all, it is seen that the asymptotic moments can be used just as well instead of the exact moments (for which expressions can be found in [32]), due to the $O(1/N^2)$ convergence of the moments. Secondly, it is seen that only 5 samples were needed to get a reliable estimate for the moments.

B. Inequalities between moments of Vandermonde matrices and moments of known distributions

We will state an inequality involving the moments of Vandermonde matrices, and the moments of known distributions from probability theory. The classical Poisson distribution with rate λ and jump size α is defined as the limit of

$$\left(\left(1-\frac{\lambda}{n}\right)\delta_0 + \frac{\lambda}{n}\delta_\alpha\right)^{*n}$$

as $n\to\infty$ [29], where $*$ denotes classical (additive) convolution, and $*n$ denotes n-fold convolution with itself. For our analysis, we will only need the classical Poisson distribution with rate c and jump size 1. We will denote this quantity by ν_c. The free Poisson distribution with rate λ and jump size α is defined similarly as the limit of

$$\left(\left(1-\frac{\lambda}{n}\right)\delta_0 + \frac{\lambda}{n}\delta_\alpha\right)^{\boxplus n}$$

as $n\to\infty$, where $\boxplus$ is the free probability counterpart of classical additive convolution [29], [20], and where where $\boxplus n$

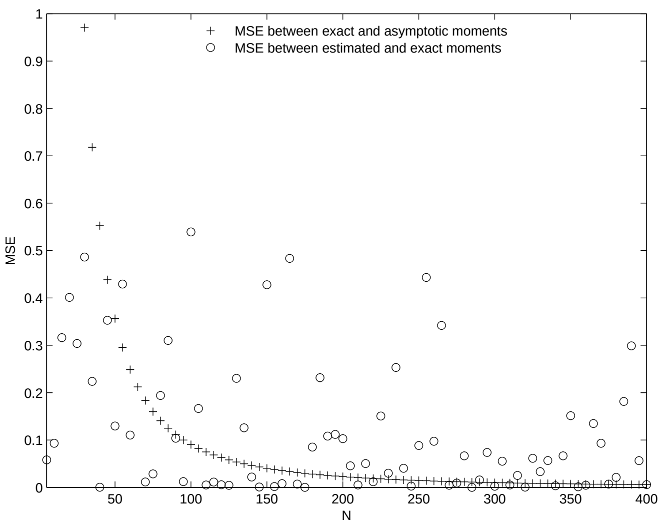


Fig. 1. MSE of the first 4 estimated moments from the exact moments for 80 samples for varying matrix sizes, with $N = L$. Matrices are on the form $\mathbf{V}^H\mathbf{V}$ with $\mathbf{V}$ a Vandermonde matrix with uniform phase distributions. The MSE of the first 4 exact moments from the asymptotic moments is also shown.

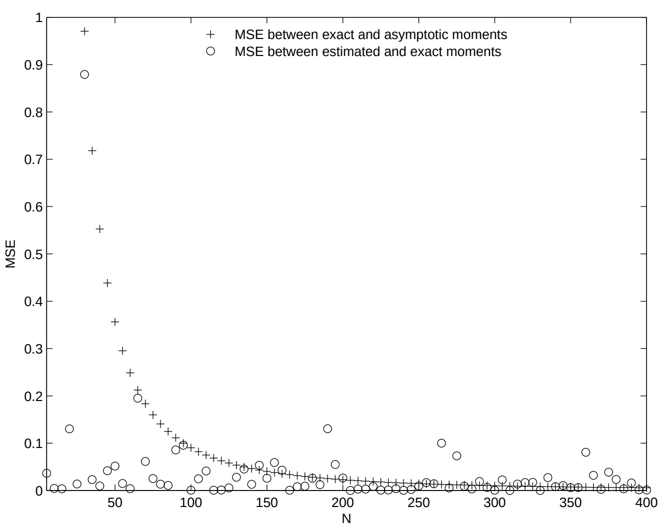


Fig. 2. MSE of the first 4 moments from the actual moments for 320 samples for varying matrix sizes, with $N = L$. Matrices are on the form $\mathbf{V}^H\mathbf{V}$ with $\mathbf{V}$ a Vandermonde matrix with uniform phase distributions. The MSE of the moments and the asymptotic moments is also shown.

denotes n-fold free convolution with itself. For our analysis, we will only need the free Poisson distribution with rate $\frac{1}{c}$ and jump size c. We will denote this quantity by μ_c. μ_c is the same as the better known Marčhenko Pastur law, i.e. it has the density [20]

$$f^{\mu_c}(x) = (1-\frac{1}{c})^+\delta_0(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi c x}, \tag{28}$$

where $(z)^+ = \max(0,z)$, $a = (1-\sqrt{c})^2$, $b = (1+\sqrt{c})^2$. Since the classical (free) cumulants of the classical (free) Poisson distribution are $\lambda\alpha^n$ [29], we see that the (classical) cumulants of ν_c are $c, c, c, c, \ldots$, and that the (free) cumulants of μ_c are $1, c, c^2, c^3, \ldots$. In other words, if a_1 has the distribution μ_c,

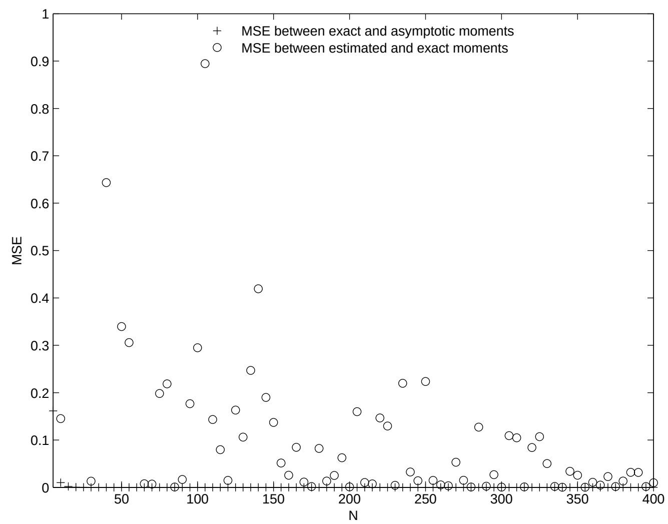


Fig. 3. MSE of the first 4 moments from the actual moments for 5 samples for varying matrix sizes, with $N = L$. Matrices are on the form $\frac{1}{N}\mathbf{X}\mathbf{X}^H$ with $\mathbf{X}$ a complex standard Gaussian matrix. The MSE of the moments and the asymptotic moments is also shown.

then

$$\begin{array}{rcl} \phi(a_1^n) & = & \sum_{\rho\in NC(n)} c^{n-|\rho|} = \sum_{\rho\in NC(n)} c^{|K(\rho)|-1} \\ & = & \sum_{\rho\in NC(n)} c^{|\rho|-1}. \end{array} \tag{29}$$

Here we have used the Kreweras complementation map (here ϕ is the expectation in a non-commutative probability space). Also, if a_2 has the distribution ν_c, then

$$E(a_2^n) = \sum_{\rho\in\mathcal{P}(n)} c^{|\rho|}. \tag{30}$$

We immediately recognize the $c^{|\rho|-1}$-entry of Theorem 1 in (29) and (30) (except for an additional power of c in (30)). Combining Proposition 1 with $\mathbf{D}_1(N) = \cdots = \mathbf{D}_n(N) = \mathbf{I}_N$, (29), and (30), we thus get the following corollary to Proposition 1:

Corollary 2: Assume that $\mathbf{V}$ has uniform phase distributions. Then the limit moment

$$V_n = \lim_{N\to\infty} E\left[tr_L\left(\left(\mathbf{V}^H\mathbf{V}\right)^n\right)\right]$$

satsifies the inequality

$$\phi(a_1^n) \leq V_n \leq \frac{1}{c}E(a_2^n),$$

where a_1 has the distribution μ_c of the Marčhenko Pastur law, and a_2 has the Poisson distribution ν_c. In particular, equality occurs for $m = 1,2,3$ and $c = 1$ (since all partitions are noncrossing for $m = 1,2,3$).

Corollary 2 thus states that the moments of Vandermonde matrices with uniform phase distributions are bounded above and below by the moments of the classical and free Poisson distributions, respectively. The left part of the inequality in Corollary 2 was also observed in [16] Section VI. The different Poisson distributions enter here because their (free and classical) cumulants resemble the $c^{|\rho|-1}$-entry in Theorem 1, where we also can use that $K_{\rho,u} = 1$ if and only if ρ is noncrossing to get a connection with the Marčhenko Pastur law. To see how close the asymptotic Vandermonde moments are to these upper and lower bounds, the following corollary to Proposition 3 contains the first moments:

Corollary 3: When $c = 1$, the limit moments

$$V_n = \lim_{N\to\infty} E\left[tr_L\left(\left(\mathbf{V}^H\mathbf{V}\right)^n\right)\right],$$

the moments fp_n of the Marčhenko Pastur law μ_1, and the moments p_n of the Poisson distribution ν_1 satisfy

$$\begin{array}{lclcl} fp_4 = 14 & \leq & V_4 = \frac{44}{3} \approx 14.67 & \leq & p_4 = 15 \\ fp_5 = 42 & \leq & V_5 = \frac{146}{3} \approx 48.67 & \leq & p_5 = 52 \\ fp_6 = 132 & \leq & V_6 = \frac{3571}{20} \approx 178.55 & \leq & p_6 = 203 \\ fp_7 = 429 & \leq & V_7 = \frac{2141}{3} \approx 713.67 & \leq & p_7 = 877. \end{array}$$

The first three moments coincide for the three distributions, and are $1, 2$, and 5, respectively.

The numbers fp_n and p_n are simply the number of partitions in $NC(n)$ and $\mathcal{P}(n)$, respectively. The number of partitions in $NC(n)$ equals the Catalan number $C_n = \frac{1}{n+1}\binom{2n}{n}$ [29], so they are easily computed. The number of partitions of $\mathcal{P}(n)$ are also known as the Bell numbers B_n [29]. They can easily be computed from the recurrence relation

$$B_{n+1} = \sum_{k=0}^{n} B_k \binom{n}{k}.$$

In Figure 4, the mean eigenvalue distribution of 640 samples of a 1600×1200 (i.e. $c = 0.75$) Vandermonde matrix with uniform phase distribution is shown. While the Poisson distribution ν_1 is purely atomic and has masses at 0, 1, 2, and 3 which are e^{-1}, e^{-1}, $e^{-1}/2$, and $e^{-1}/6$ (the atoms consist of all integer multiples), the Vandermonde histogram shows a more continuous eigenvalue ditribution, with the peaks which the Poisson distribution has at integer multiples clearly visible here as well. The peaks are not as sharp though. We remark that the support of $\mathbf{V}^H\mathbf{V}$ for a fixed N goes all the way up to N, but lies within $[0, N]$. It is unknown whether the peaks at integer multiples in the Vandermonde histogram grow to infinity as we let $N \to \infty$. From the histogram, only the peak at 0 seems to be of atomic nature. The effect of decreasing c amounts to stretching the eigenvalue density vertically, and compressing it horizontally, just as the case for the different Marčhenko Pastur laws. An eigenvalue histogram for Gaussian matrices which in the limit give the corresponding (in the sense of Corollary 2) Marčhenko Pastur law for Figure 4 (i.e. $\mu_{0.75}$) is shown in Figure 6. We have also shown another eigenvalue histogram to demonstrate the case of a non-uniform phase distribution. In Figure 5, the mean eigenvalue distribution of 640 samples of a 1600×1200 Vandermonde matrix with phase distribution with density

$$p_\omega(x) = \frac{1}{2\alpha\sqrt{0.04\pi^2 - x^2}} \tag{31}$$

on $[-\frac{\pi\sin\alpha}{5}, \frac{\pi\sin\alpha}{5}]$ is shown, with $\alpha = \frac{\pi}{4}$. One can see that the effect of high values for this density near the origin is that the Vandermonde matrix has a high concentration of the eigenvalues near the origin, and also a higher proportion of larger eigenvalues, when compared to the uniform phase distribution. The corresponding density is shown in Figure 7.

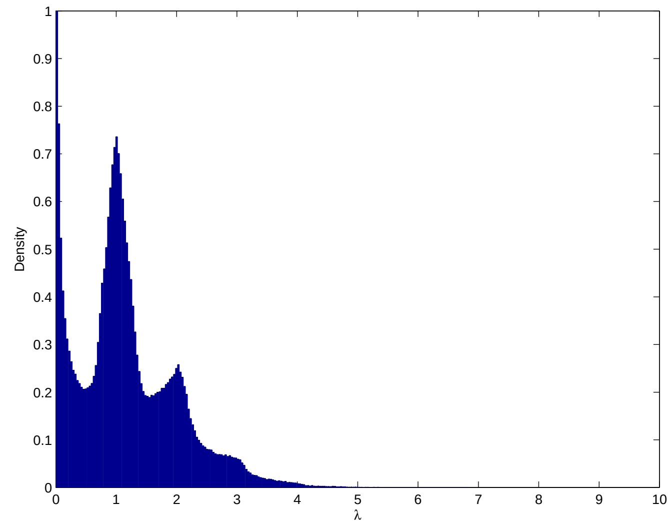


Fig. 4. Histogram of the mean eigenvalue distribution of 640 samples of $\mathbf{V}^H\mathbf{V}$, with $\mathbf{V}$ a 1600×1200 Vandermonde matrix with uniform phase distributions.

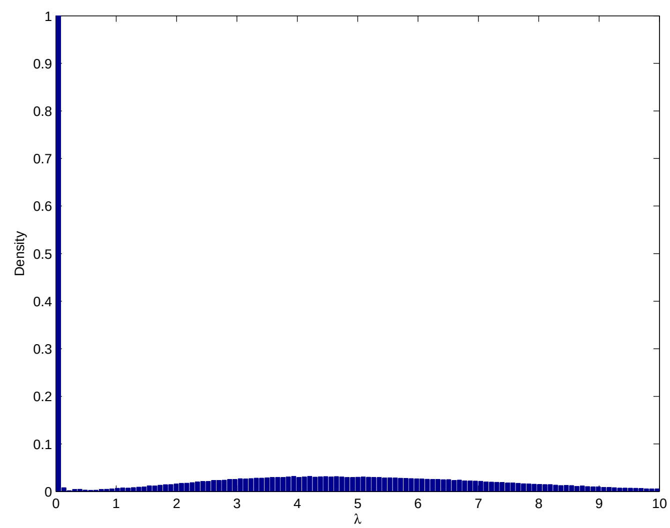


Fig. 5. Histogram of the mean eigenvalue distribution of 640 samples of $\mathbf{V}^H\mathbf{V}$, with $\mathbf{V}$ a 1600×1200 Vandermonde matrix with phase distribution p_ω defined in (31).

It is unknown whether the inequalities for the moments can be extended to inequalities for the associated capacity If $\mathbf{X}$ is an $N\times N$ standard, complex, Gaussian matrix, then an explicit expression for the asymptotic capacity exists [24]:

$$\begin{aligned} \lim_{N\to\infty} \tfrac{1}{N}\log_2\det\left(\mathbf{I}_N + \rho\left(\tfrac{1}{N}\mathbf{X}\mathbf{X}^H\right)\right) = \\ 2\log_2\left(1+\rho-\tfrac{1}{4}\left(\sqrt{4\rho+1}-1\right)^2\right) \\ -\tfrac{\log_2 e}{4\rho}\left(\sqrt{4\rho+1}-1\right)^2. \end{aligned} \tag{32}$$

In Figure 8(a), several realizations of the capacity are computed for Gaussian matrix samples of size 36×36. The asymptotic capacity (32) is also shown. In Figure 8(b), several realizations of the capacity are computed for Vandermonde matrix samples of the same size, for the case of uniform phase distribution. It is seen that the variance of the Vandermonde

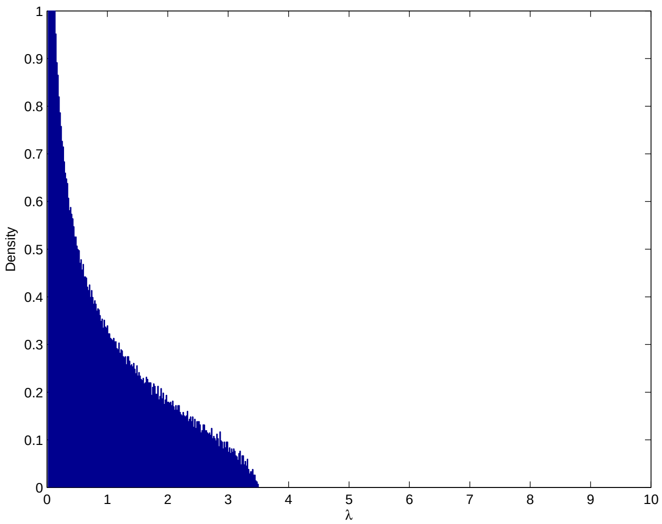


Fig. 6. Histogram of the mean eigenvalue distribution of 20 samples of $\frac{1}{N}\mathbf{X}\mathbf{X}^H$, with $\mathbf{X}$ an $L \times N = 1200 \times 1600$ complex, standard, Gaussian matrix.

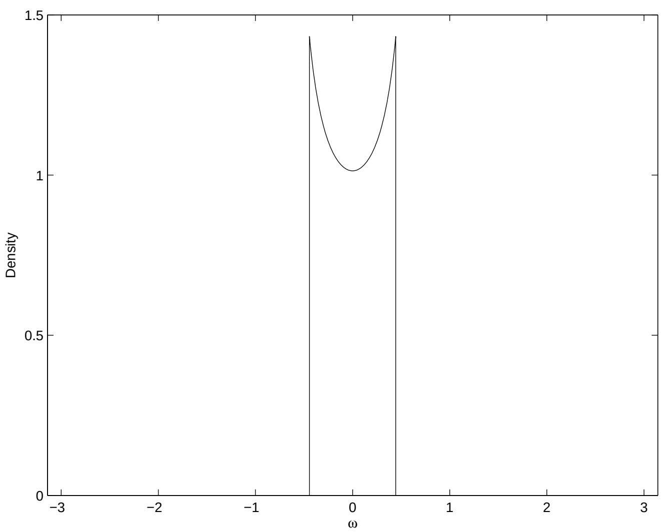
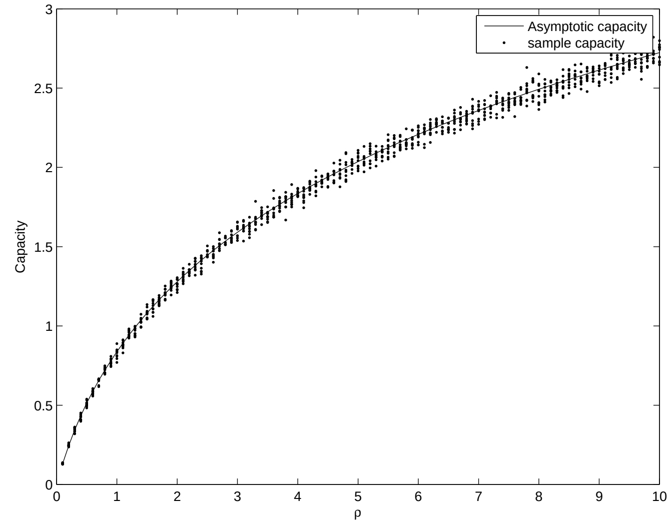
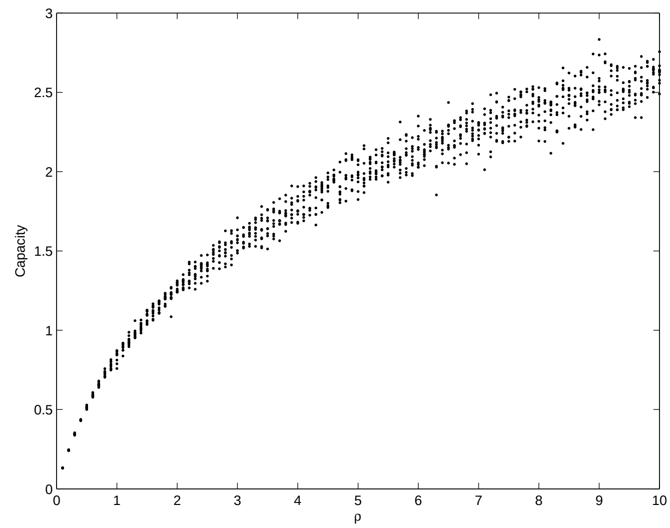


Fig. 7. The density $p_\omega(x)$ used in this paper. $\alpha = \frac{\pi}{4}$ and $\lambda = 10d$.



(a) Realizations of $\frac{1}{N}\log_2 \det\left(\mathbf{I}_N + \rho\frac{1}{N}\mathbf{X}\mathbf{X}^H\right)$ when $\mathbf{X}$ is standard, complex, Gaussian. The asymptotic capacity (32) is also shown.



(b) Realizations of $\frac{1}{N}\log_2 \det\left(\mathbf{I}_N + \rho\mathbf{V}\mathbf{V}^H\right)$ when ω has uniform phase distribution.

Fig. 8. Realizations of the capacity for Gaussian and Vandermonde matrices of size 36×36.

capacities is higher than for the Gaussian counterparts. This should come as no surprise, due to the slower convergence to the asymptotic limits for Vandermonde matrices. Although the capacities of Vandermonde matrices with uniform phase distribution and Gaussian matrices seem to be close, we have no proof that the capacities of Vandermonde matrices are even finite due to the unboundedness of its support.

C. Deconvolution

Deconvolution with Vandermonde matrices (as stated in (11) in Theorem 1) differs from the Gaussian deconvolution counterpart [29] in the sense that there is no multiplicative [29] structure involved, since $K_{\rho,\omega}$ is not multiplicative in ρ. The Gaussian equivalent of Proposition 3 (i.e. $\mathbf{V}^H\mathbf{V}$ replaced with $\frac{1}{N}\mathbf{X}\mathbf{X}^H$, with $\mathbf{X}$ an $L \times N$ complex, standard, Gaussian matrix) is

$$
\begin{aligned}
m_1 &= d_1 && (33)\\
m_2 &= d_2 + d_1^2 && (34)\\
m_3 &= d_3 + 3d_2d_1 + d_1^3 && (35)\\
m_4 &= d_4 + 4d_3d_1 + 2d_2^2 + 6d_2d_1^2 + d_1^4 && (36)\\
m_5 &= d_5 + 5d_4d_1 + 5d_3d_2 + 10d_3d_1^2 + \\
&\quad 10d_2^2d_1 + 10d_2d_1^3 + d_1^5 && (37)\\
m_6 &= d_6 + 6d_5d_1 + 6d_4d_2 + 15d_4d_1^2 + \\
&\quad 3d_3^2 + 30d_3d_2d_1 + 20d_3d_1^3 + \\
&\quad 5d_2^3 + 10d_2^2d_1^2 + 15d_2d_1^4 + d_1^6 && (38)\\
m_7 &= d_7 + 7d_6d_1 + 7d_5d_2 + 21d_5d_1^2 + \\
&\quad 7d_4d_3 + 42d_4d_2d_1 + 35d_4d_1^3 + \\
&\quad 21d_3^2d_1 + 21d_3d_2^2 + 105d_3d_2d_1^2 + \\
&\quad 35d_3d_1^4 + 35d_2^3d_1 + 70d_2^2d_1^3 + \\
&\quad 21d_2d_1^5 + d_1^7, && (39)
\end{aligned}
$$

(where the m_i and the d_i are computed as in (12)-(13) by scaling the respective moments by c). This follows immediately from asymptotic freeness, and from the fact that $\frac{1}{N}\mathbf{X}\mathbf{X}^H$ converges to the Marčhenko Pastur law μ_c. In particular, when all $\mathbf{D}_i(N) = \mathbf{I}_L$ and $c = 1$, we obtain the limit moments: 1,2,5,14,42,132,429, which also were listed in Corollary 3. One can also write down a Gaussian equivalent to the covariance of traces of Vandermonde matrices (19) and (20) (covariance of traces of Gaussian matrices are handled more thoroughly in [33]). These are

$$\begin{aligned} & E\left[\left(tr_n\left(\mathbf{D}(N)\frac{1}{N}\mathbf{X}\mathbf{X}^H\right)\right)^2\right] \\ = & \ (tr_n(\mathbf{D}(N))^2 + \frac{1}{nN}tr_n(\mathbf{D}(N)^2) \qquad (40) \\ & E\left[\left(tr_n\left(\mathbf{D}(N)\frac{1}{N}\mathbf{X}\mathbf{X}^H\right)\right)^n\right] \\ = & \ (tr_n(\mathbf{D}(N))^n + O(N^{-2}) \qquad (41) \\ & E\left[tr_n\left(\mathbf{D}(N)\frac{1}{N}\mathbf{X}\mathbf{X}^H\right)tr_n\left(\left(\mathbf{D}(N)\frac{1}{N}\mathbf{X}\mathbf{X}^H\right)^2\right)\right] \\ = & \ tr_n(\mathbf{D}(N))tr_n(\mathbf{D}(N)^2) + O(N^{-2}). \qquad (42) \end{aligned}$$

These equations can be proved using the same combinatorical methods as in [32], and it is not needed that the matrices $\mathbf{D}(N)$ be diagonal. Only the first equation is here stated as an exact expression. The second and third equations also have exact counterparts, but their computations are more involved. Similarly, one can write down a Gaussian equivalent to Theorem 2 for the exact moments. For the first three moments (the fourth moment is dropped, since this is more involved), these are

$$\begin{aligned} m_1 &= d_1 \\ m_2 &= d_2 + d_1^2 \\ m_3 &= \left(1 + N^{-2}\right) d_3 + 3d_1d_2 + d_1^3. \end{aligned}$$

This follows from a careful count of all possibilities after the matrices have been multiplied together (see also [32], where one can see that the restriction that the matrices $\mathbf{D}_i(N)$ are diagonal can be dropped in the Gaussian case). It is seen, contrary to Theorem 2 for Vandermonde matrices, that the second exact moment equals the second asymptotic moment (34), and also that the convergence is faster (i.e. $O(n^{-2})$) for the third moment (this will also be the case for higher moments).

The two types of (de)convolution also differ in how they can be computed in practice. In [23], an algorithm for free convolution with the Marčhenko Pastur law was sketched. A similar algorithm may not exist for Vandermonde convolution. However, Vandermonde convolution can be subject to numerical approximation: To see this, note first that Theorem 3 splits the numerics into two parts: The approximation of the integrals $\int p_\omega(x)^{|\rho|}dx$, and the approximation of the $K_{\rho,u}$. A strategy for obtaining the latter quantities could be to randomly generate many numbers between 0 and 1 and estimate the volume as the ratio of the solutions which satisfy (76) in Appendix B. Implementations of the various Vandermonde convolution variants given in this paper can be found in [34].

In practice, one often has a random matrix model where independent Gaussian and Vandermonde matrices are both present. In such cases, it should be possible to combine the individual results for both of them. In Section V, examples on how this can be done are presented.

V. Simulations

The simulations presented here all use the deconvolution framework for Vandermonde matrices. Since additive, white, Gaussian noise also is taken into account, Vandermonde deconvolution is also combined with Gaussian deconvolution. In the eigenvalue histograms for Vandermonde matrices shown in figures 4 and 5, as large matrices as needed were used in order for the eigenvalue distribution to stabilize on something close to the asymptotic eigenvalue distribution. In practical scenarios, N and L are much smaller than what was used in these figures, which partially explains the uncertainty in some of the simulations we will present next (all simulations operate on other values for N and L). In particular, the uncertainty for the p_ω in (31) is high, since exact expressions for the lower order moments are not known, contrary to the case of uniform phase distribution. In all the following, d is the distance between the antennas whereas λ is the wavelength. The ratio $\frac{d}{\lambda}$ is a figure of the resolution with which the system will be able to separate (and therefore estimate the position) of users in space.

A. Detection of the number of sources

Let us consider a basestation equipped with N receiving antennas, and with L mobiles (each with a single antenna) in the cell. The received signal at the base station is given by

$$\mathbf{r}_i = \mathbf{V}\mathbf{P}^{\frac{1}{2}}\mathbf{s}_i + \mathbf{n}_i. \qquad (43)$$

Here $\mathbf{r}_i$ is the the $N \times 1$ received vector $\mathbf{s}_i$ is the $L \times 1$ transmit vector by the L users which is assumed to satisfy $\mathbb{E}\left[\mathbf{s}_i\mathbf{s}_i^H\right] = \mathbf{I}_L$, $\mathbf{n}_i$ is $N \times 1$ additive, white, Gaussian noise of variance $\frac{\sigma}{\sqrt{N}}$ (alle components in $\mathbf{s}_i$ and $\mathbf{n}_i$ are assumed independent). In the case of a line of sight between the users and the base station, and considering a Uniform Linear Array (ULA), the matrix $\mathbf{V}$ has the following form:

$$\mathbf{V} = \frac{1}{\sqrt{N}}\begin{pmatrix} 1 & \cdots & 1 \\ e^{-j2\pi\frac{d}{\lambda}\sin(\theta_1)} & \cdots & e^{-j2\pi\frac{d}{\lambda}\sin(\theta_L)} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi(N-1)\frac{d}{\lambda}\sin(\theta_1)} & \cdots & e^{-j2\pi\frac{d}{\lambda}\sin(\theta_L)} \end{pmatrix} \qquad (44)$$

Here, θ_i is the angle of the user in the cell and is supposed to be uniformly distributed over $[-\alpha,\alpha]$. $\mathbf{P}^{\frac{1}{2}}$ is an $L \times 1$ power matrix due to the different distances from which the users emit. In other words, we assume that the phase distribution has the form $2\pi\frac{d}{\lambda}\sin(\theta)$ with θ uniformly distributed on $[-\alpha,\alpha]$. The fact that the phase has the form $2\pi\frac{d}{\lambda}\sin(\theta)$ is a well known result in array processing [30]. The users' distribution can be known (in the case of these simulations, the uniform distribution has been accounted for without loss of generality) through measurements in wireless systems up to some parameters (here, α typically). This is usually done to have a better understanding of the user's behavior. It is easily seen, by taking inverse functions, that the density is, when $\frac{2d\sin\alpha}{\lambda} < 1$,

$$p_\omega(x) = \frac{1}{2\alpha\sqrt{\frac{4\pi^2d^2}{\lambda^2} - x^2}} \tag{45}$$

on $[-\frac{2\pi d\sin\alpha}{\lambda}, \frac{2\pi d\sin\alpha}{\lambda}]$, and 0 elsewhere. A special case of this was considered in Figure 7, where we set $\lambda = 10d, \alpha = \frac{\pi}{4}$.

Throughout the paper we will assume, as in Figure 5, that $\alpha = \frac{\pi}{4}$, $d = 1$, and $5\lambda = d$ when model (44) is used. With this assumption, $\frac{2d\sin\alpha}{\lambda} < 1$ is always fulfilled.

The goal is to detect the number of sources L and their respective power based on the sample covariance matrix supposing that we have K observations, of the same order as N. When the number of observation is quite higher than N (and the noise variance is known), classical subspace methods [35] provide tools to detect the number of sources. Indeed, let $\mathbf{R}$ be the true covariance matrix given by

$$\mathbf{V}\mathbf{P}\mathbf{V}^H + \sigma^2\mathbf{I}_N,$$

where σ^2 is the noise variance. This matrix has $N - L$ eigenvalues equal to σ^2 and L eigenvalues strictly superior to σ^2. One can therefore determine the number of source by counting the number of eigenvalues different from σ^2. However, in practice, one has only access to the sample covariance matrix given by

$$\mathbf{W} = \frac{1}{K}\mathbf{Y}\mathbf{Y}^H,$$

with

$$\mathbf{Y} = [\mathbf{r}_1, ...\mathbf{r}_K] = \mathbf{V}\mathbf{P}^{\frac{1}{2}}[\mathbf{s}_1, ..., \mathbf{s}_K] + [\mathbf{n}_1, ..., \mathbf{n}_K] \tag{46}$$

If one simply has the sample covariance matrix $\mathbf{W}$, (43) has three independent parts which must be dealt with in order to get an estimate of $\mathbf{P}$: the Gaussian matrices $\mathbf{S} = [\mathbf{s}_1, ..., \mathbf{s}_K]$ and $\mathbf{N} = [\mathbf{n}_1, ..., \mathbf{n}_K]$, and the Vandermonde matrix $\mathbf{V}$. It should thus be possible to combine Gaussian deconvolution [32] and Vandermonde deconvolution by performing the following steps:

1) Estimate the moments of $\frac{1}{K}\mathbf{V}\mathbf{P}^{\frac{1}{2}}\mathbf{S}\mathbf{S}^H\mathbf{P}^{\frac{1}{2}}\mathbf{V}^H$ using multiplicative free convolution as described in [23]. This is the denoising part.
2) Estimate the moments of $\mathbf{P}\mathbf{V}^H\mathbf{V}$, again using multiplicative free deconvolution.
3) Estimate the moments of $\mathbf{P}$ using Vandermonde deconvolution as described in this paper.

Putting these steps together, we will prove the following:

Proposition 6: Define

$$I_n = (2\pi)^{n-1}\int_0^{2\pi} p_\omega(x)^n dx, \tag{47}$$

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

$$\begin{aligned} P_i &= tr_L(\mathbf{P}^i) \\ W_i &= tr_N(\mathbf{W}^i). \end{aligned}$$

Then the equations

$$\begin{aligned} W_1 &= c_2P_1 + \sigma^2 \\ W_2 &= c_2P_2 + (c_2^2I_2 + c_2c_3)(P_1)^2 \\ &\quad +2\sigma^2(c_2 + c_3)P_1 + \sigma^4(1 + c_1) \\ W_3 &= c_2P_3 + (3c_2^2I_2 + 3c_2c_3)P_1P_2 \\ &\quad + \left(c_2^3I_3 + 3c_2^2c_3I_2 + c_2c_3^2\right)(P_1)^3 \\ &\quad +3\sigma^2(1 + c_1)c_2P_2 \\ &\quad +3\sigma^2((1 + c_1)c_2^2I_2 + c_3(c_3 + 2c_2))(P_1)^2 \\ &\quad +3\sigma^4(c_1^2 + 3c_1 + 1)c_2P_1 \\ &\quad +\sigma^6(c_1^2 + 3c_1 + 1) \end{aligned}$$

provide an asymptotically unbiased estimator for the moments P_i from the moments of W_i (or vice versa) when $\lim_{N\to\infty}\frac{N}{K} = c_1$, $\lim_{N\to\infty}\frac{L}{N} = c_2$, and where $\lim_{N\to\infty}\frac{L}{K} = c_3$.

The proof of this can be found in Appendix L. Note that $c_3 = c_1c_2$, so that the definition of c_3 is really not necessary. We still include it however, since c_1, c_2 and c_3 are matrix aspect ratios which are used in different deconvolution stages, so that they all are used when these stages are implemented and combined serially. Note also that the statement applies to any ω with continuous density due to Theorem 3, not only the densities we restrict to here. In the simulations, Proposition 6 is put to the test when $\mathbf{P}$ has three sets of powers, 0.5, 1, and 1.5 (with equal probability), with phase distribution given by (44). Both the number of sources and the powers are estimated. For the phase distribution (44), the integrals I_2 and I_3 can be computed exactly (for general phase distributions they are computed numerically), and are [36]

$$\begin{aligned} I_2 &= \frac{\lambda}{4d\alpha^2}\ln\left(\frac{1+\sin\alpha}{1-\sin\alpha}\right) \\ I_3 &= \frac{\lambda^2\tan\alpha}{4d^2\alpha^3}. \end{aligned}$$

Under the assumptions $\alpha = \frac{\pi}{4}$ and $\lambda = 10d$ used throughout this paper, the integrals above take the values

$$\begin{aligned} I_2 &= \frac{40}{\pi^2}\ln\left(\frac{2+\sqrt{2}}{2-\sqrt{2}}\right) \\ I_3 &= \frac{1600}{\pi^3}. \end{aligned}$$

For estimation of the powers, knowing that we have only three sets of powers with equal probability, it suffices to estimate the three lowest moments in order to get an estimate of the powers (which are the three distinct eigenvalues of $\mathbf{P}$). Therefore, in the following simulations, Proposition 6 is first used to get an estimate of the moments of $\mathbf{P}$. Then these are used to obtain an estimate of the three distinct eigenvalues of $\mathbf{P}$ using the Newton-Girard formulas [37]. These should then lie close to the three powers of $\mathbf{P}$.

For the model (44), it turns out that power estimation does not work particularly well. The result is shown in the first plot of Figure 11. In the plot, $K = L = N = 576$, and $\sigma = \sqrt{0.1}$. Even though the matrices are quite large, the estimated powers are quite far from the actual powers.

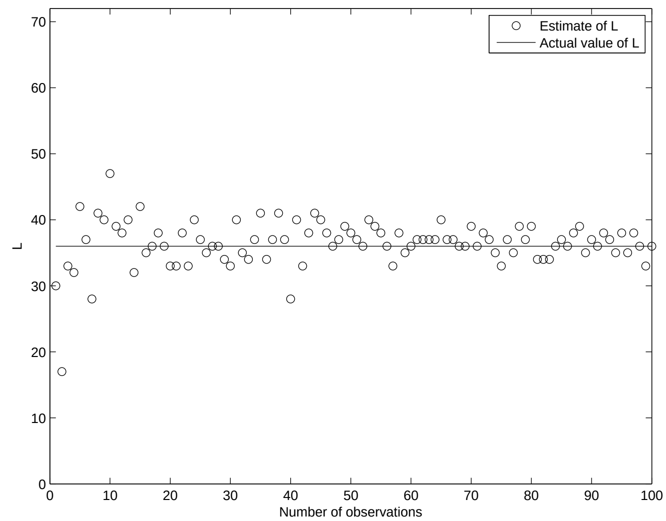
Actually, the estimation process is so far off that it computes eigenvalues which are complex conjugate pairs instead of the true, real ones $(0.5, 1, 1.5)$ (this is an explanation for that the two lowest eigenvalues in the plot seem to coincide, since it is only the absolute values of the eigenvalues which are plotted). Increasing the matrix sizes further results in estimates which are closer to the true powers, but one would need matrices of size larger than 2000×2000 to get much closer to the true powers. As will be seen, power estimation works much better for the phase distribution model in the next section. A tentative explanation for this is the difference between the corresponding eigenvalue histograms of those two Vandermonde matrices, which are shown in Figure 5 for model (44), and in Figure 4 for the model of the next section.

For estimation of the number of users L, we assume that the power distribution of $\mathbf{P}$ is known, but not L itself. Since L is unknown, in the simulations we enter different candidate values of it into the following procedure:

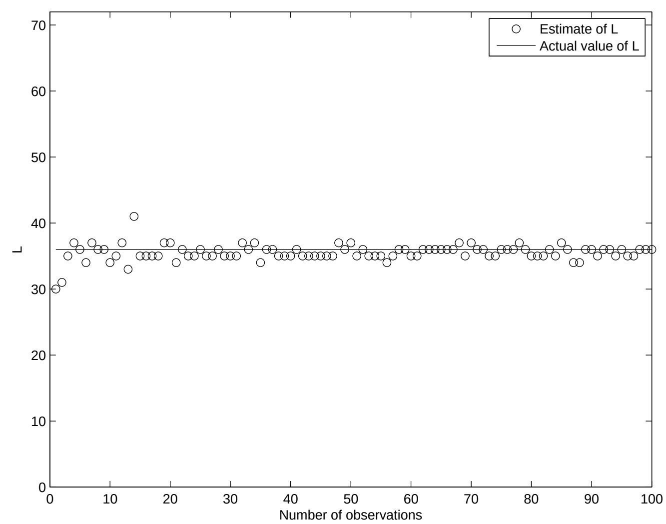
1) Computing the moments $P_i = tr_L(\mathbf{P}^i)$ of $\mathbf{P}$.
2) The moments $tr_L(\mathbf{P}^i)$ are fed into the formulas of Proposition 6, and we thus obtain candidate moments W_i of the sample covariance matrix $\mathbf{W}$.
3) Compute the sum of the square errors between these candidate moments, and the moments $\hat{W}_i$ of the observed sample covariance matrix $\hat{\mathbf{W}}$, i.e. compute $\sum_{i=1}^{3} |W_i - \hat{W}_i|^2$.

The estimate L for the number of users is chosen as the one which gives the minimum value for the sum of square errors after these steps.

In Figure 9, we have set $\sigma = \sqrt{0.1}$, $N = 100$, and $L = 36$. $\mathbf{P}$ has three sets of powers, 0.5, 1, and 1.5 (with equal probability). We tried the procedure described above for 1 all the way up to 100 observations. It is seen that only a small number of observations are needed in order to get an accurarate estimate of L. When $K = 1$, it is seen that more observations are needed to get an accurate estimate of L, when compared to $K = 10$.



(a) $K = 1$



(b) $K = 10$

Fig. 9. Estimate for the number of users. Actual value of L is 36. Also, $\sigma = \sqrt{0.1}$, $N = 100$.

B. Estimation of the number of paths

In many channel modeling applications, one needs to determine the number of paths in the channel [38]. For this purpose, consider a multi-path channel of the form:

$$h(\tau) = \sum_{i=1}^{L} s_i \delta(\tau - \tau_i)$$

Here, s_i are i.d. Gaussian random variables with power P_i and τ_i are uniformly distributed delays over $[0, T]$. The s_i represent the attenuation factors due to the different reflections. L is the total number of paths. In the frequency domain, the channel is given by:

$$H(f) = \sum_{i=1}^{L} s_i G(f) e^{-j2\pi f \tau_i}$$

We consider which samples the channel in frequency. Sampling the continuous frequency signal at $f_i = i\frac{W}{N}$ where W is the bandwidth, the model becomes (for a given channel realization):

$$\mathbf{H} = \mathbf{V}\mathbf{P}^{\frac{1}{2}}\mathbf{s}$$

where:

$$\mathbf{V} = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & \cdots & 1 \\ e^{-j2\pi \frac{W\tau_1}{N}} & \cdots & e^{-j2\pi \frac{W\tau_L}{N}} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi(N-1)\frac{W\tau_1}{N}} & \cdots & e^{-j2\pi(N-1)\frac{W\tau_L}{N}} \end{pmatrix}, \tag{48}$$

We will here set $W = T = 1$, which means that the ω_i of (1) are uniformly distributed over $[0, 2\pi]$. When additive noise (n_i) again is taken into consideration, our model again becomes that of (43), the only difference being that the phase distribution of the Vandermonde matrix now is uniform. L now is the number of paths, N the number of frequency samples, and $\mathbf{P}$ is the unknown $L \times L$ diagonal power matrix. Taking K observations we arrive at the same form as in (46). In this case with uniform phase distribution, we can do even better than Proposition 6, in that one can write down estimators for the moments which are unbiased for any number of observations and frequency samples:

Proposition 7: Assume that $\mathbf{V}$ has uniform phase distribution, and let P_i be the moments of $\mathbf{P}$, and $W_i = tr_N(\mathbf{W}^i)$ the moments of the sample covariance matrix. Define also $c_1 = \frac{N}{K}$, $c_2 = \frac{L}{N}$, and $c_3 = \frac{L}{K}$. Then

$$
\begin{aligned}
E[W_1] &= c_2 P_1 + \sigma^2 \\
E[W_2] &= c_2\left(1 - \frac{1}{N}\right) P_2 + c_2(c_2+c_3)(P_1)^2 \\
&\quad +2\sigma^2(c_2+c_3)P_1 + \sigma^4(1+c_1) \\
E[W_3] &= c_2\left(1+\frac{1}{K^2}\right)\left(1-\frac{3}{N}+\frac{2}{N^2}\right)P_3 \\
&\quad +\left(1-\frac{1}{N}\right)\left(3c_2^2\left(1+\frac{1}{K^2}\right)+3c_2c_3\right)P_1P_2 \\
&\quad +\left(c_2^3\left(1+\frac{1}{K^2}\right)+3c_2^2c_3+c_2c_3^2\right)(P_1)^3 \\
&\quad +3\sigma^2\left((1+c_1)c_2+\frac{c_1c_2^2}{KL}\right)\left(1-\frac{1}{N}\right)P_2 \\
&\quad +3\sigma^2\left(\frac{c_1c_2^3}{KL}+c_2^2+c_3^2+3c_2c_3\right)(P_1)^2 \\
&\quad +3\sigma^4\left(c_1^2+3c_1+1+\frac{1}{K^2}\right)c_2P_1 \\
&\quad +\sigma^6\left(c_1^2+3c_1+1+\frac{1}{K^2}\right)
\end{aligned}
$$

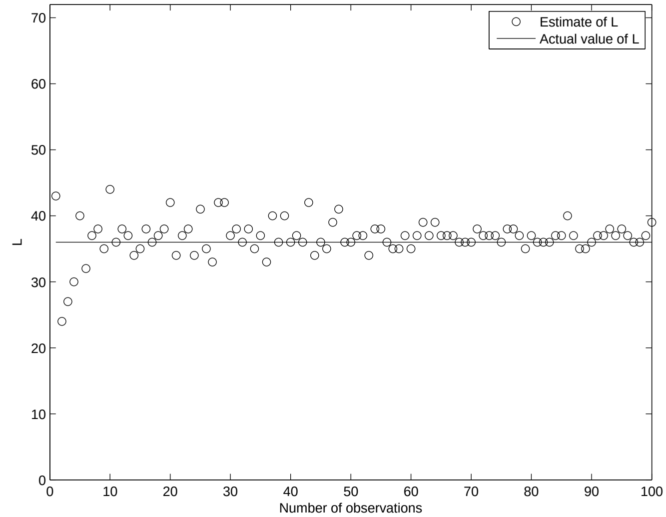
Just as Proposition 6, this is proved in Appendix L. In the following, this result is used in order to determine the number of paths as well as the power of each path. The different convergence rates of the approximations are clearly seen in the plots.

In Figure 10, the number of paths is estimated based on the procedure sketched above. We have set $\sigma = \sqrt{0.1}$, $N = 100$, and $L = 36$. The procedure is tried for 1 all the way up to 100 observations. The plot is very similar to Figure 9, in that only a small number of observations are needed in order to get an accurate estimate of L. When $K = 1$, it is seen that more observations are needed to get an accurate estimate of L, when compared to $K = 10$.

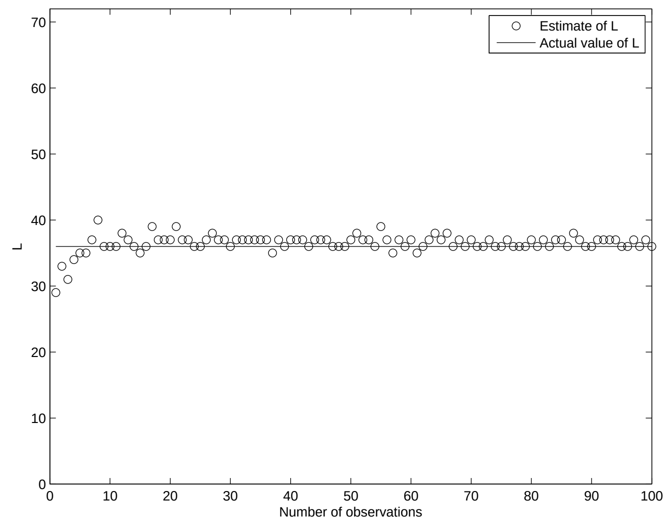
For the estimation of powers simulation, we have set $K = N = L = 144$, and $\sigma = \sqrt{0.1}$, following the procedure also described above, up to 1000 observations. The second plot in Figure 11 shows the results which confirms the usefulness of the approach. We see that even for smaller matrix sizes than the model of the previous section, the estimates are much closer to the true powers.

C. Estimation of wavelength

In the field of MIMO cognitive sensing [39], [40], terminals must decide on the band on which to transmit and in particular sense which band is occupied. One way of doing is to find the wavelength λ in (44), based on some realizations of the sample covariance matrix. In our simulation, we have set $d = 5$ and $\lambda = 10$, $K = 10$, $L = 36$, $N = 100$, and $\sigma = \sqrt{0.1}$. We have tried the values $1, 2, ..., 100$ as candidate wavelengths, and chosen the one which gives the smallest deviation (in the same sense as above, i.e. the sum of the squared errors of the first three moments are taken) from a different number of realizations of sample covariance matrices. The resulting plot is shown in Figure 12, and shows that the Vandermonde deconvolution method can also be used for wavelength estimation. It is seen that the estimation gets better when the number of observations is increased.



(a) $K = 1$

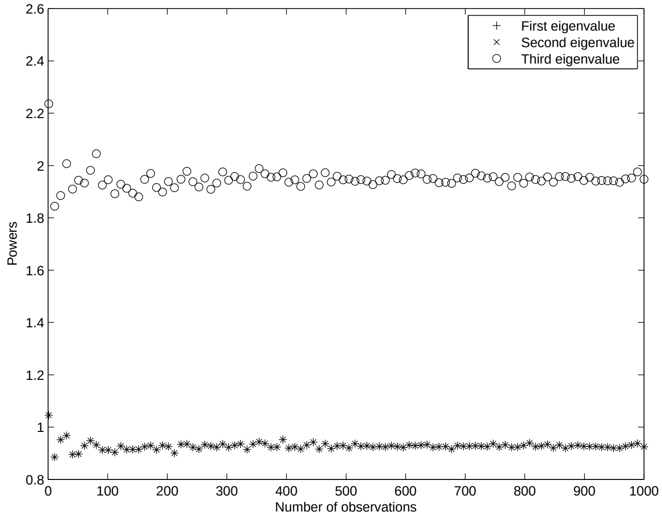


(b) $K = 10$

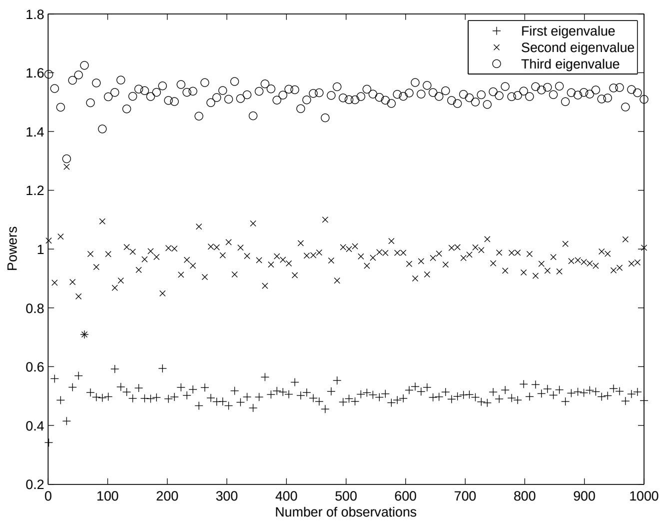
Fig. 10. Estimate for the number of paths. Actual value of L is 36. Also, $\sigma = \sqrt{0.1}$, $N = 100$.

D. Signal reconstruction and estimation of the sampling distribution

Several works have investigated how irregular sampling affects the performance of signal reconstruction in the presence of noise in different fields namely sensor networks [41], [42], image processing [43], [44], geophysics [45], compressive



(a) Estimation of powers for various number of observations for the model (44) of Section V-B. $K = N = L = 576$, and $\sigma = \sqrt{0.1}$.



(b) Estimation of powers for various number of observations for the model (48) of Section V-A. $K = N = L = 144$, and $\sigma = \sqrt{0.1}$.

Fig. 11. Estimation of powers for the two models (44) and (48) of this section, for various number of observations.

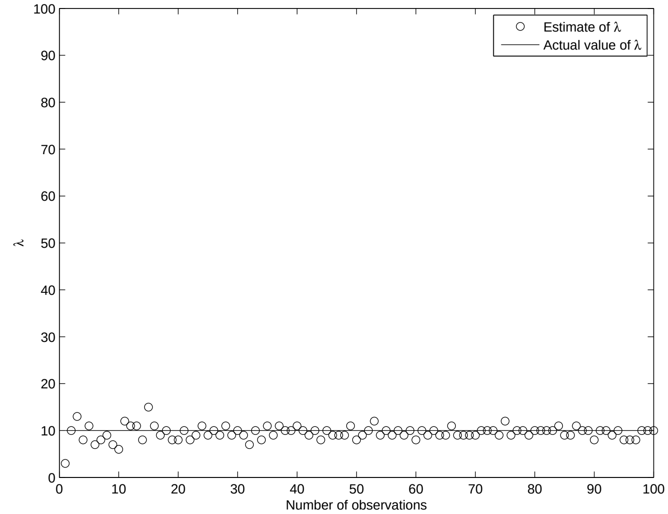


Fig. 12. Estimation of wavelength. Deconvolution was performed for varying number of observations, assuming different wavelengths, In the true model (44), $d = 1$, $\lambda = 10$, $K = 10$, $L = 36$, $N = 100$, and $\sigma = \sqrt{0.1}$.

sampling [46]. The usual Nyquist Theorem states that for a signal with maximum frequency f_{max}, one needs to sample the signal at a rate which is at least twice this number. However, in many cases, this can not be performed or one has an observation of a signal at only a subset of the frequencies. Moreover, one feels that if the signal has a sparse spectrum, one can take fewer samples and still have the same information on the original signal. One of the central motivations of sparse sampling is exactly to understand under which condition one can still have less samples and recover the original signal up to an error of ϵ [47]. Let us consider the signal of interest as a superposition of its frequency components (this is also the case for a unidimensional bandlimited physical signal) i.e

$$r(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} s_k e^{\frac{-j2\pi kt}{N}}$$

and suppose that the signal is sampled at various instants $[t_1, ..., t_L]$ with $t_i \in [0, 1]$. This can be identically written as

$$r(\omega) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} s_k e^{-jk\omega},$$

or $\mathbf{r} = \mathbf{V}^T \mathbf{a}$. In the presence of noise, one can write

$$\mathbf{r} = \mathbf{V}^T \mathbf{s} + \mathbf{n}, \tag{49}$$

where $\mathbf{r} = [r(\omega_1), ...r(\omega_L)]^T$, $\mathbf{s}$ and $\mathbf{n}$ are as in (43), and with $\mathbf{V}$ on the form (1). Contrary to (43), (49) does not include the diagonal matrix $\mathbf{P}$. A similar analysis for such cases can be found in [16], [48].

In the following, we suppose that one has K observations of the received sampled vector $\mathbf{r}$:

$$\mathbf{Y} = [\mathbf{r}_1, ...\mathbf{r}_K] = \mathbf{V}^T [\mathbf{s}_1, ..., \mathbf{s}_K] + [\mathbf{n}_1, ..., \mathbf{n}_K] \tag{50}$$

The vector $\mathbf{r}$ is the discrete output of the sampled continuous signal $r(w)$ for which the distribution is unknown (however, c is known). This case happens when one has an observation without the knowledge of the sampling rate for example. The difference in (50) from the model (46) lies in that the adjoint of a Vandermonde matrix is used, and in that there is no additional diagonal matrix $\mathbf{P}$ included. The following result can now be stated and proved similarly to Proposition 6 and 7:

Proposition 8:

$$\begin{aligned} E\left[tr_n\left(\mathbf{W}\right)\right] &= 1 + \sigma^2 & (51)\\ E\left[tr_n\left(\mathbf{W}^2\right)\right] &= c_2 I_2 + (1 + c_3)(1 + \sigma^2)^2 & (52)\\ E\left[tr_n\left(\mathbf{W}^3\right)\right] &= 1 + 3c_2(1 + c_3) I_2 \\ & \quad 3c_3 + c_3^2 + c_2^2 I_3 \\ & \quad 3\sigma^2(1 + 3c_3 + c_3^2 + c_2(1 + c_3) I_2) \\ & \quad 3\sigma^4 c_2(c_3^2 + 3c_3 + 1) \\ & \quad \sigma^6(c_3^2 + 3c_3 + 1), & (53) \end{aligned}$$

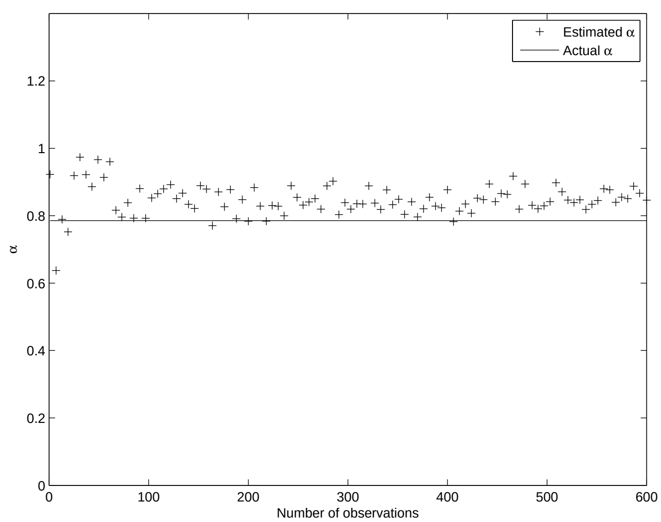


Fig. 13. Estimated values of α using (51)-(53), for various number of observations, and for $K = 10, L = 36, N = 100, \sigma = \sqrt{0.1}$. The actual value of α was $\frac{\pi}{4}$.

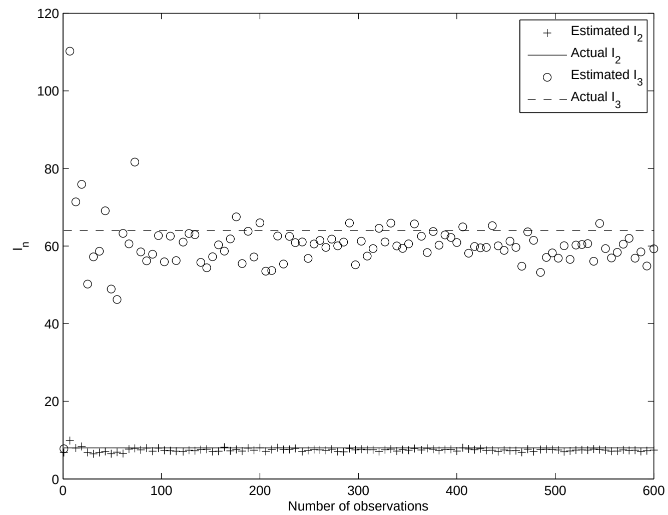


Fig. 14. Estimated values of I_2 and I_3 using (51)-(53), for various number of observations, and for $K = 10, L = 36, N = 100, \sigma = \sqrt{0.1}$. The actual value of α was $\frac{\pi}{4}$.

where $\lim_{N\to\infty} \frac{N}{K} = c_1, \lim_{N\to\infty} \frac{L}{N} = c_2, \lim_{N\to\infty} \frac{L}{K} = c_3$, I_n is defined as in Proposition 6, and $\mathbf{W} = \frac{1}{K}\mathbf{Y}\mathbf{Y}^H$.

The proof of Proposition 8 is commented in Appendix L. We have tested (51)-(53) as follows: we have taken a phase distribution ω which is uniform on $[0, \alpha]$, and 0 elsewhere. The density is thus $\frac{2\pi}{\alpha}$ on $[0, \alpha]$, and 0 elsewehere. In this case we can compute that

$$\begin{aligned} I_2 &= \frac{2\pi}{\alpha} \\ I_3 &= \left(\frac{2\pi}{\alpha}\right)^2. \end{aligned}$$

The first of these equations, combined with (51)-(53), enables us to estimate α. This is tested in Figure 13 for various number of observations. In Figure 14 we have also tested estimation of I_2, I_3 from the observations using the same equations. When one has a distribution which is not uniform, the integrals $I_3, I_4, ...$ would also be needed in finding the characteristics of the underlying phase distribution. Figure 14 shows that the estimation of I_2 requires far fewer observation than the estimation of I_3. In both figures, the values $K = 10, L = 36, N = 100$, and $\sigma = \sqrt{0.1}$ were used and α was $\frac{\pi}{4}$. It is seen that the estimation of I_3 is a bit off even for higher number of observations. This is to be expected, since an asymptotoic result is applied.

VI. Related work

In the recent work [16], the Vandermonde model (1) is encountered in the context of reconstruction of multidimensional signals in wireless sensor networks. The authors also recognize a similar expression for the Vandermonde mixed moment expansion coefficient as in Definition 6. They also state that, for the case of uniform phase distribution, closed form expressions for the moments can be found, building on an analysis of partitions and calculation of volumes of convex polytopes described by certain constraints. This is very similar to what is done in this paper. However [16] does not perform concrete calculations up to the first seven moments. Also, the connection between the uniform case and the general case, as in Theorem 3, was not made. Mixed moments with independent matrices were also not computed, as the case of several independent Vandermonde matrices.

Interestingly, [17] shows that, in cases where the matrices have entries of the form $A_{i,j} = F(\omega_i - \omega_j)$, analytical expressions for the moments can be found. This may be interesting for the Vandermonde matrices we consider, since

$$\left(\frac{1}{N}\mathbf{V}^H\mathbf{V}\right)_{i,j} = \frac{\sin\left(\frac{N}{2}(\omega_i - \omega_j)\right)}{N\sin\left(\frac{1}{2}(\omega_i - \omega_j)\right)}.$$

Unfortunately, the function $F_N(x) = \frac{\sin\left(\frac{N}{2}x\right)}{N\sin\left(\frac{1}{2}x\right)}$ depends on the matrix dimension N, so that we can not find a function F which fits the result from [17].

VII. Conclusion and further directions

We have shown how asymptotic moments of random Vandermonde matrices can be computed analytically, and treated many different cases. Vandermonde matrices with uniform phase distributions proved to be the easiest case and was given separate treatment, and it was shown how the case with more general phases could be expressed in terms of the case of uniform phase distributions. The case where the phase distribution has singularities was also given separate treatment, as this case displayed different asymptotic behaviour. Also mixed moments of independent Vandermonde matrices were computed, as well as the moments of generalized Vandermonde matrices. In addition to the general asymptotic expressions stated, exact expressions for the first moments of Vandermonde matrices with uniform phase distributions were also stated. We have also provided some useful applications of random Vandermonde matrices. The applications concentrated on deconvolution and signal sampling analysis. As shown, many useful system models use independent Vandermonde matrices and Gaussian matrices combined in some way. The presented examples show how random Vandermonde matrices applied for such systems can be handled in practice to obtain estimates on quantities such as the number of paths in channel modeling, the transmission powers of the users in wireless transmission or the sampling distribution for signal recovery. The paper has only touched upon a limited number of applications but the results already provide benchmark figures in the non-asymptotic regime.

From a theoretical perspective, it would also be interesting to find methods for obtaining $K_{\rho,\omega,\lambda}$ from $K_{\rho,u,u}$, similar to what has been done with $K_{\rho,\omega}$ from $K_{\rho,u}$. This could also shed some light on whether uniform phase and power distribution also minimizes moments of generalized Vandermonde matrices, similary to how we showed that uniform phase distribution minimizes moments Vandermonde matrices of the form (1).

Throughout the paper, we assumed that only diagonal matrices were involved in mixed moments of Vandermonde matrices. The case of non-diagonal matrices is harder to address, and should be addressed in future research. The analysis of the maximum and minimum eigenvalue is also of importance. The methods presented in this paper can not be used directly to obtain explicit expressions for the asymptotic mean eigenvalue distribution, so this is also a case for future research. A way of attacking this problem could be to develop for Vandermonde matrices analytic counterparts to what one has in free probability (such as the R- and S-transform and their connection with the Stieltjes transform).

Finally, another case for future research is the asymptotic behaviour of Vandermonde matrices when the matrix entries lie outside the unit circle. The asymptotics are very different in this case. The choice of Vandermonde matrix entries on the unit circle was applied for this paper since the asymptotic behaviour is more easily addressed in this case.

Appendix A
The proof of Theorem 1

We can write

$$E\left[tr_L\left(\mathbf{D}_1(N)\mathbf{V}^H\mathbf{V}\mathbf{D}_2(N)\mathbf{V}^H\mathbf{V}\cdots\mathbf{D}_n(N)\mathbf{V}^H\mathbf{V}\right)\right] \quad (54)$$

as

$$\begin{aligned} L^{-1}\sum_{\substack{i_1,\ldots,i_n\\ j_1,\ldots,j_n}} E(\quad & \mathbf{D}_1(N)(j_1,j_1)\mathbf{V}^H(j_1,i_2)\mathbf{V}(i_2,j_2)\\ & \mathbf{D}_2(N)(j_2,j_2)\mathbf{V}^H(j_2,i_3)\mathbf{V}(i_3,j_3)\\ & \vdots\\ & \mathbf{D}_n(N)(j_n,j_n)\mathbf{V}^H(j_n,i_1)\mathbf{V}(i_1,j_1)) \end{aligned} \quad (55)$$

The $(j_1,\ldots,j_n)$ give rise to a partition ρ of $\{1,\ldots,n\}$, where each block W_j consists of equal values, i.e.

$$W_j = \{k | j_k = j\}.$$

Write

$$W_j = \{w_{j1}, w_{j2}, \ldots, w_{j|W_j|}\}.$$

When $(j_1,\ldots,j_n)$ give rise to ρ, we see that since

$$j_{w_{j1}} = j_{w_{j2}} = \cdots = j_{w_{j|W_j|}},$$

we also have that

$$\omega_{j_{w_{j1}}} = \omega_{j_{w_{j2}}} = \cdots = \omega_{j_{w_{j|W_j|}}},$$

and we will denote their common value by ω_{W_j} as in Definition 6. With this in mind, it is straightforward to verify that (55) can be written as

$$\begin{aligned} &\sum_{\rho\in\mathcal{P}(n)} \sum_{(i_1,\ldots,i_n)} \sum_{\substack{(j_1,\ldots,j_n)\\ \text{giving rise to } \rho}} N^{-n}L^{-1}\\ &\times\prod_{k=1}^{|\rho|} E\left(e^{j\left(\sum_{k\in W_j} i_{k-1}-\sum_{k\in W_j} i_k\right)\omega_{W_k}}\right)\\ &\times\mathbf{D}_1(N)(j_1,j_1)\times\cdots\times\mathbf{D}_n(N)(j_n,j_n)., \end{aligned} \quad (56)$$

where $i_1,\ldots,i_n$ takes values between 0 and $N-1$. We will in the following switch between the form (56) and the form

$$\begin{aligned} &\sum_{\rho\in\mathcal{P}(n)} \sum_{\substack{(j_1,\ldots,j_n)\\ \text{giving rise to } \rho}} \sum_{(i_1,\ldots,i_n)} N^{|\rho|-n-1}c^{|\rho|-1}L^{-|\rho|}\\ &\times E\left(\prod_{k=1}^{n}\left(e^{j(\omega_{b(k-1)}-\omega_{b(k)})i_k}\right)\right)\\ &\times\mathbf{D}_1(N)(j_1,j_1)\times\cdots\times\mathbf{D}_n(N)(j_n,j_n), \end{aligned} \quad (57)$$

where we also have reorganized the powers of N and L in (56), and changed the order of summation (i.e. summed over the different $i_1,\ldots,i_n$ first). Noting that

$$\begin{aligned} &\sum_{(i_1,\ldots,i_n)} N^{|\rho|-n-1}E\left(\prod_{k=1}^{n} e^{j(\omega_{b(k-1)}-\omega_{b(k)})i_k}\right) && (58)\\ =\;& N^{|\rho|-n-1}E\left(\sum_{(i_1,\ldots,i_n)}\prod_{k=1}^{n} e^{j(\omega_{b(k-1)}-\omega_{b(k)})i_k}\right) && (59)\\ =\;& N^{|\rho|-n-1}E\left(\prod_{k=1}^{n}\left(\sum_{i_k=0}^{N-1} e^{j(\omega_{b(k-1)}-\omega_{b(k)})i_k}\right)\right) && (60)\\ =\;& N^{|\rho|-n-1}E\left(\prod_{k=1}^{n}\frac{1-e^{jN(\omega_{b(k-1)}-\omega_{b(k)})}}{1-e^{j(\omega_{b(k-1)}-\omega_{b(k)})}}\right) && (61)\\ =\;& N^{|\rho|-n-1}\times\\ & \int_{(0,2\pi)^{|\rho|}}\prod_{k=1}^{n}\frac{1-e^{jN(\omega_{b(k-1)}-\omega_{b(k)})}}{1-e^{j(\omega_{b(k-1)}-\omega_{b(k)})}}\\ & d\omega_1\cdots d\omega_{|\rho|} && (62)\\ =\;& K_{\rho,\omega,N}, && (63) \end{aligned}$$

Definition 6 of the Vandermonde mixed moment expansion coefficients come into play, so that (57) can also be written

$$\begin{aligned} &\sum_{\rho\in\mathcal{P}(n)} \sum_{\substack{(j_1,\ldots,j_n)\\ \text{giving rise to } \rho}} c^{|\rho|-1}L^{-|\rho|}K_{\rho,\omega,N}\\ &\times\mathbf{D}_1(N)(j_1,j_1)\cdots\times\times\mathbf{D}_n(N)(j_n,j_n). \end{aligned} \quad (64)$$

The notation for a joint limit distribution simplifies (57). Indeed, add to (57) for each ρ the terms

$$\sum_{\rho' \in \mathcal{P}(n), \rho' > \rho} \sum_{(j_1,...,j_n) \text{ giving rise to } \rho'} c^{|\rho|-1} L^{-|\rho|} K_{\rho,\omega,N} \times \mathbf{D}_1(N)(j_1,j_1) \cdots \times \mathbf{D}_n(N)(j_n,j_n). \quad (65)$$

These go to 0 as $N \to \infty$, since they are bounded by

$$c^{|\rho|-1} L^{-|\rho|} K_{\rho,\omega,N} L^{|\rho'|} = K_{\rho,\omega,N} c^{|\rho|-1} L^{|\rho'|-|\rho|} = O(L^{-1}).$$

After this addition, the limit of (64) can be written

$$\sum_{\rho \in \mathcal{P}(n)} c^{|\rho|-1} K_{\rho,\omega} D_\rho, \quad (66)$$

which is what we had to show. ∎

We also need to comment on the statement of Theorem 6, where generalized Vandermonde matrices are considered. In this case, the derivations after (57) are different since the power distribution is not uniform. For the case of (3), we can in (60) replace $\sum_{i_k=1}^{n} e^{j(\omega_{b(k-1)}-\omega_{b(k)})i_k}$ with $\sum_{r=0}^{N-1} N p_{f_N}(r) e^{jr(\omega_{b(k-1)}-\omega_{b(k)})}$, since the number of occurences of the power $e^{jr(\omega_{b(k-1)}-\omega_{b(k)})}$ is $N p_{f_N}(r)$. The rest of the proof of Theorem 6 follows by cancelling n powers of N after this replacement. The details are similar for the case (4), where the law of large numbers is applied to arrive at the second formula in (26).

(11) will also be useful on the scaled form

$$cM_n = \sum_{\rho \in \mathcal{P}(n)} K_{\rho,\omega} (cD)_\rho. \quad (67)$$

When $\mathbf{D}_1(N) = \mathbf{D}_2(N) = \cdots = \mathbf{D}_n(N)$, we denote their common value $\mathbf{D}(N)$, and define the sequence $D = (D_1, D_2, ...)$ with $D_n = \lim_{N\to\infty} tr_L\left((\mathbf{D}(N))^n\right)$. In this case D_ρ does only depend on the block cardinalities $|W_j|$, so that we can group together the $K_{\rho,\omega}$ for ρ with equal block cardinalities. If we group the blocks of ρ so that their cardinalities are in descending order, and set

$$\mathcal{P}(n)_{r_1,r_2,...,r_k} = \{\rho = \{W_1, ..., W_k\} \in \mathcal{P}(n) | |W_i| = r_i \forall i\},$$

where $r_1 \geq r_2 \geq \cdots \geq r_k$, and also write

$$K_{r_1,r_2,...,r_k} = \sum_{\rho \in \mathcal{P}(n)_{r_1,r_2,...,r_k}} K_{\rho,\omega}, \quad (68)$$

After the substitutions (12-(13), (67) can be written

$$m_n = \sum_{\substack{r_1,...,r_k \\ r_1+\cdots+r_k=n}} K_{r_1,r_2,...,r_k} \prod_{j=1}^{k} d_{r_j}. \quad (69)$$

For the first 5 moments this becomes

$$\begin{aligned}
m_1 &= K_1 d_1 && (70)\\
m_2 &= K_2 d_2 + K_{1,1} d_1^2 && (71)\\
m_3 &= K_3 d_3 + K_{2,1} d_2 d_1^2 + K_{1,1,1} d_1^3 && (72)\\
m_4 &= K_4 d_4 + K_{3,1} d_3 d_1 + K_{2,2} d_2^2 + K_{2,1,1} d_2 d_1^2 + \\
&\quad K_{1,1,1,1} d_1^4 && (73)\\
m_5 &= K_5 d_5 + K_{4,1} d_4 d_1 + + K_{3,2} d_3 d_2 + \\
&\quad K_{3,1,1} d_3 d_1^2 + K_{2,2,1} d_2^2 d_1 + K_{2,1,1,1} d_2 d_1^3 + \\
&\quad K_{1,1,1,1,1} d_1^5. && (74)
\end{aligned}$$

This reorganization of the terms will be used in the following.

Appendix B
The proof of Proposition 1

Note that for each block W_j,

$$E\left(e^{j\left(\sum_{k\in W_j} i_{k-1} - \sum_{k\in W_j} i_k\right)\omega_{W_j}}\right) = 0$$

when

$$\sum_{k\in W_j} i_{k-1} \neq \sum_{k\in W_j} i_k,$$

and 1 if

$$\sum_{k\in W_j} i_{k-1} = \sum_{k\in W_j} i_k. \quad (75)$$

We thus define $S_{\rho,N}$ to be the set of all n-tuples $(i_1, ..., i_n)$ where $1 \leq i_k \leq N$ ($1 \leq k \leq n$), and where

$$\sum_{k\in W_j} i_{k-1} = \sum_{k\in W_j} i_k$$

for all $j \in \{1, ..., |\rho|\}$. We also define $|S_{\rho,N}|$ to be the cardinality of $S_{\rho,N}$. With this definition in place, it is obvious that

$$K_{\rho,u} = \lim_{N\to\infty} K_{\rho,u,N} = \lim_{N\to\infty} \frac{1}{N^{n+1-|\rho|}} |S(\rho, N)|.$$

Finding the limit distribution thus boils down to finding $|S_{\rho,N}|$, which is equivalent to finding the number of solutions to equations of the form (75), where the variables are integers constrained to lie between 0 and $N-1$. For Proposition 2 we will compute $|S_{\rho,N}|$ for certain ρ of lower order. To prove Proposition 1, we need not compute specific $|S_{\rho,N}|$.

First we explain why $K_{\rho,u} \leq 1$. It is clear from (75) that $|S_{\rho,N}|$ is the number of integer solutions $\mathbf{i} = [i_1 \cdots i_n]^T$ (with i_j between 0 and $N-1$) to a system of the form $\mathbf{A}\mathbf{i} = \mathbf{0}$, where

1) $\mathbf{A}$ is $|\rho| \times n$,
2) all entries in $\mathbf{A}$ are either -1, 0, or 1,
3) each column of A contains either exactly one -1 and one 1, or just zeroes.

Such a matrix $\mathbf{A}$ has rank $|\rho|-1$, as can be found through elementary row reduction. Hence, there are $|\rho|-1$ pivot columns in $\mathbf{A}$, and $n+1-|\rho|$ free variables among $(i_1, ..., i_n)$ in the solution set of $\mathbf{A}\mathbf{i} = \mathbf{0}$. Therefore, $|S_{\rho,N}| \leq N^{n+1-|\rho|}$, which proves that $K_{\rho,u} \leq 1$.

By dividing the equations (75) by N, and letting N go to infinity, $K_{\rho,u}$ can alternatively be expressed as the volume in $\mathbb{R}^{n+1-|\rho|}$ of the solution set of

$$\sum_{k\in W_j} x_{k-1} = \sum_{k\in W_j} x_k, \tag{76}$$

with $0 \leq x_k \leq 1$ (the volume is computed after expressing the remaining $|\rho|-1$ variables in the $n+1-|\rho|$ free variables). It is clear that the integral for this volume computes to a rational number greater than 0 but less than 1, due to the integral bounds given by (76), and since the volume is contained within a (higher-dimensional) unit cube. It is also clear that the integral computes to 1 if and only if the reduced row echelon form of $\mathbf{A}$ only contains rows with zeros, and rows with 2 nonzero entries (these entries will then be 1 and -1, respectively). This corresponds to a solution set where each pivot variable equals one of the free variables. For the rest of the proof, it therefore suffices to show that such a solution set occurs if and only if the partition ρ is noncrossing.

If ρ is noncrossing, there exists a block W_1 (after renumbering the blocks if necessary) which consists of a single interval of numbers, say $\{r, r+1, ..., r+|W_1|\}$. This block's equation in (75) is easily seen to imply that

$$i_{r-1} = i_{r+|W_1|},$$

and that $i_r, ..., i_{r+|W_1|-1}$ can be chosen arbitrarily. Therefore, this block gives rise to $|W_1|-1$ free variables.

Let W_2 be the block which contains $r+|W_1|+1$ (after renumbering the blocks if necessary). We add together the two equations represented by W_1 and W_2 in (76), and replace these two equations with this sum. The new set of equations gives rise to a new matrix $\mathbf{A}$, where columns $r, ..., r+|W_1|$ are easily seen to contain only 0's. These columns represent $|W_1|-1$ free variables, and can be chosen independently from the rest of the variables. We therefore remove these columns from $\mathbf{A}$. The new equation system corresponds to the equation system for a noncrossing partition of $\{1, ..., n-|W_1|\}$ with $|\rho|-1$ blocks, created by merging the blocks W_1 and W_2. The step where we find a block which is an interval can now be repeated to merge two more blocks, and this process can be repeated until we remain with 1 block with $|W_{|\rho|}|$ elements after $|\rho|-1$ block merges. It is clear that this last block gives rise to $|W_{|\rho|}|$ free variables. If we sum up the total number of free variables we get

$$|W_{|\rho|}| + \sum_{i=1}^{|\rho|-1} (|W_i|-1) = n - (|\rho|-1) = n+1-|\rho|.$$

All in all we see that the solution set is as described above (i.e. each constrained variable is equal to one of the free variables), so that $N^{n+1-|\rho|}$ choices of $i_1, ..., i_n$ satisfy (75), which shows that $K_{\rho,u} = 1$ when ρ is noncrossing. It is easy to see that, when ρ has crossings, the procedure followed above will fail, so that at least one of the constrained variables is not equal to a free variable. But then $K_{\rho,u} < 1$ for such ρ, which proves the theorem.

We remark that it is the form (76) which will be used in the other appendices to compute $K_{\rho,u}$ for certain lower order ρ. From the proof, we see that when ρ is noncrossing, there exists a partition of $\{1, ..., n\}$ into $n+1-|\rho|$ blocks, where two elements are defined to be in the same block if and only if their corresponding variables are equal. It is obvious from the construction above that this partition is the Kreweras complement $K(\rho)$ of ρ. This fact is used elsewhere in this paper.

We will also briefly explain why the computations in this appendix are also useful for generalized Vandermonde matrices with uniform phase distribution. For (3), the number of solutions $i_1, ..., i_k$ to (75) needs to be multiplied by

$$Np_{f_N}(i_1) \cdots Np_{f_N}(i_k),$$

since each i_j now may occur $Np_{f_N}(i_j)$ times. This means that $K_{\rho,\omega,f}$ can be computed as the ingegrals in this appendix, but that we also need to multiply with the density p_f for each variable. The computations of these new integrals become rather involved when f is not uniform, and are therefore dropped.

Appendix C
The proof for Proposition 2

We will in the following compute the volume of the solution set of (76), as a volume in $[0,1]^{n+1-|\rho|} \subset \mathbb{R}^{n+1-|\rho|}$, as explained in the proof of Proposition 1. These integrals are very tedious to compute, and many of the details are skipped. The formula

$$\frac{r!s!}{(r+s+1)!} = \int_0^1 x^r (1-x)^s dx$$

can be used to simplify some of the calculations for higher values of n.

A. Computation of $K_{\{\{1,3\},\{2,4\}\},u}$

This is equivalent to finding the volume of the solution set of

$$x_1 + x_3 = x_2 + x_4$$

in $\mathbb{R}^3$. Since this means that

$$x_4 = x_1 + x_3 - x_2 \text{ lies between } 0 \text{ and } 1,$$

we can set up the following integral bounds: When $x_1 + x_3 \leq 1$, we must have that $0 \leq x_2 \leq x_1 + x_3$, so that we get the contribution

$$\int_0^1 \int_0^{1-x_1} \int_0^{x_1+x_3} dx_2 dx_3 dx_1,$$

which computes to $\frac{1}{3}$. When $1 \leq x_1 + x_3$, we must have that $x_1 + x_3 - 1 \leq x_2 \leq 1$, so that we get the contribution

$$\int_0^1 \int_{1-x_1}^1 \int_{x_1+x_3-1}^1 dx_2 dx_3 dx_1,$$

which also computes to $\frac{1}{3}$. Adding the contributions together we get $\frac{2}{3}$, which is the stated value for $K_{\{\{1,3\},\{2,4\}\},u}$.

It turns out that when the blocks of ρ are cyclic shifts of each other, the computation of $K_{\rho,u}$ can be simplified. Examples of such ρ are $\{\{1,3\},\{2,4\}\}$ (for which we just computed

$K_{\rho,u}$), $\{\{1,3,5\},\{2,4,6\}\}$, and $\{\{1,4\},\{2,5\},\{3,6\}\}$. We will in the following describe this simplified computation. Let $a_l^{(m)}(x)$ be the polynomial which gives the volume in $\mathbb{R}^{m-1}$ of the solutions set to $x_1+\cdots+x_m=x$ (constrained to $0\leq x_i\leq 1$) for $l\leq x\leq l+1$. It is clear that these satisfy the integral equations

$$a_l^{(m+1)}(x)=\int_{x-1}^{l} a_{l-1}^{(m)}(t)dt+\int_l^x a_l^{(m)}(t)dt, \qquad (77)$$

which can be used to compute the $a_l^m(x)$ recursively. Note first that $a_0^{(1)}(x)=1$. For $m=2$ we have

$$\begin{aligned} a_0^{(2)}(x) &= \int_0^x a_0^{(1)}(t)dt = x \\ a_1^{(2)}(x) &= \int_{x-1}^1 a_0^{(1)}(t)dt = 2-x. \end{aligned}$$

For $m=3$ we have

$$\begin{aligned} a_0^{(3)}(x) &= \int_0^x a_0^{(2)}(t)dt = \frac{1}{2}x^2 \\ a_1^{(3)}(x) &= \int_{x-1}^1 a_0^{(2)}(t)dt+\int_1^x a_1^{(2)}(t)dt \\ &= 1-\frac{1}{2}(x-1)^2-\frac{1}{2}(2-x)^2 \\ a_2^{(3)}(x) &= \int_{x-1}^2 a_1^{(2)}(t)dt = \frac{1}{2}(3-x)^2. \end{aligned}$$

By integrating the $a_0^{(2)}(x)$, we can double-check our computation of $K_{\{\{1,3\},\{2,4\}\},u}$ above:

$$\int_0^1 (a_0^{(2)})^2(t)dt+\int_1^2 (a_1^{(2)})^2(t)dt=\frac{2}{3}.$$

B. Computation of $K_{\{\{1,3,5\},\{2,4,6\}\},u}$

For $m=3$, integration gives

$$\int_0^1 (a_0^{(3)})^2(t)dt+\int_1^2 (a_1^{(3)})^2(t)dt+\int_2^3 (a_2^{(3)})^2(t)dt,$$

which computes to $\frac{11}{20}$. This is the stated expression for $K_{\{\{1,3,5\},\{2,4,6\}\},u}$.

C. Computation of $K_{\{\{1,4\},\{2,5\},\{3,6\}\},u}$

This is equivalent to finding the volume of the solution set of

$$x_1+x_4=x_2+x_5=x_3+x_6$$

in $\mathbb{R}^4$, which is computed as

$$\int_0^1 (a_0^{(2)})^3(t)dt+\int_1^2 (a_1^{(2)})^3(t)dt,$$

which computes to $\frac{1}{2}$. This is the stated expression for $K_{\{\{1,4\},\{2,5\},\{3,6\}\},u}$.

D. Computation of $K_{\{\{1,4\},\{2,6\},\{3,5\}\},u}$

This is equivalent to finding the volume of the solution set of

$$\begin{aligned} x_1+x_4 &= x_2+x_5 \\ x_2+x_6 &= x_3+x_1 \end{aligned}$$

in $\mathbb{R}^4$. Since this means that

$$\begin{aligned} x_5 &= x_1-x_2+x_4 \text{ lies between } 0 \text{ and } 1, \\ x_6 &= x_1-x_2+x_3 \text{ lies between } 0 \text{ and } 1, \end{aligned}$$

we can set up the following integral bounds:

For $x_2\geq x_1$ we must have $x_2-x_1\leq x_3,x_4\leq 1$, so that we get the contribution

$$\int_0^1\int_{x_1}^1\int_{x_2-x_1}^1\int_{x_2-x_1}^1 dx_4dx_3dx_2dx_1,$$

which computes to $\frac{1}{4}$. It is clear that for $x_1\geq x_2$ we get the same result by symmetry, so that the total contribution is $\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$, which proves the claim.

E. Computation of $K_{\{\{1,5\},\{3,7\},\{2,4,6\}\},u}$

This is equivalent to finding the volume of the solution set of

$$\begin{aligned} x_1+x_5 &= x_2+x_6 \\ x_3+x_7 &= x_4+x_1 \end{aligned}$$

in $\mathbb{R}^5$, or

$$\begin{aligned} x_6 &= x_5+x_1-x_2 \text{ lies between } 0 \text{ and } 1, \\ x_7 &= x_4+x_1-x_3 \text{ lies between } 0 \text{ and } 1\ . \end{aligned} \qquad (78)$$

This can be split into the following volumes:

1) $x_1\leq x_2\leq x_3$,
2) $x_1\leq x_3\leq x_2$,
3) $x_3\leq x_2\leq x_1$,
4) $x_2\leq x_3\leq x_1$,
5) $x_2\leq x_1\leq x_3$,
6) $x_3\leq x_1\leq x_2$.

Each of these volumes can be computed by setting up an integral with corresponding bounds. Computing these integras, we get the values $\frac{1}{15}$, $\frac{1}{15}$, $\frac{1}{15}$, $\frac{1}{15}$, $\frac{11}{120}$, $\frac{11}{120}$, respectively. Adding these contributions together, we get

$$\frac{4}{15}+\frac{11}{60}=\frac{27}{60}=\frac{9}{20},$$

which proves the claim.

F. The computation of $K_{\{\{1,6\},\{2,4\},\{3,5,7\}\},u}$

This is equivalent to finding the volume of the solution set of

$$\begin{aligned} x_1+x_6 &= x_2+x_7 \\ x_2+x_4 &= x_3+x_5 \end{aligned}$$

in $\mathbb{R}^5$, or

$$\begin{aligned} x_6 &= x_7+x_2-x_1 \text{ lies between } 0 \text{ and } 1, \\ x_5 &= x_4+x_2-x_3 \text{ lies between } 0 \text{ and } 1,\ . \end{aligned}$$

This can be obtained from (78) by a permutation of the variables, so the contribution from $K_{\{\{1,6\},\{2,4\},\{3,5,7\}\},u}$ must also be $\frac{9}{20}$, which proves the claim.

APPENDIX D
THE PROOF FOR PROPOSITION 3

We will have use for the following result, taken from [29]:

Lemma 1: The number of noncrossing partitions in $NC(n)$ with r_1 blocks of length 1, r_2 blocks of length 2 and so on (so that $r_1 + 2r_2 + 3r_3 + \cdots nr_n = n$) is

$$\frac{n!}{r_1!r_2!\cdots r_n!(n+1-r_1-r_2\cdots r_n)!}.$$

Using this and a similar formula for the number of partitions with prescribed block sizes, we obtain cardinalities for noncrossing partitions and the set of all partitions with a given block structure. These numbers are the used in the following calculations. For the proof of Proposition 3, we need to compute (68) for all possible block cardinalities $(r_1, ..., r_k)$, and insert these in (70)-(74). The formulas for the three first moments are obvious, since all partitions of length ≤ 3 are noncrossing. For the remaining computations, the following two observations save a lot of work:

- If $\rho_1 \in \mathcal{P}(n_1)$, $\rho_2 \in \mathcal{P}(n_2)$ with $n_1 < n_2$, and ρ_1 can be otained from ρ_2 by omitting elements k in $\{1, ..., n_2\}$ such that k and $k+1$ are in the same block, then we must have that $K_{\rho_1,u} = K_{\rho_2,u}$. This is straightforward to prove since it follows from the proof of Proposition 1 that i_{k+1} can be chosen arbitrarily between 0 and $N-1$ in such a case.
- $K_{\rho_1,u} = K_{\rho_2,u}$ if the set of equations (76) for ρ_1 can be obtained by a permutation of the variables in the set of equations for ρ_2. Since the rank of the matrix for (76) equals the number of equations -1, we actually need only have that $|\rho_1| - 1$ of the $|\rho_1|$ equations can be obtained from permutation of $|\rho_2| - |$ equations of the $|\rho_2|$ equations in the equation system for ρ_2

A. The moment of fourth order

The result is here obvious except for the case for the three partitions with block cardinalities $(2, 2)$ (for all other block cardinalities, all partitions are noncrossing, so that $K_{r_1,r_2,...,r_k}$ is simply the number of noncrossing partitions with block cardinalities $(r_1, ..., r_k)$. this number can be computed from Lemma 1). Two of the partitions with blocks of cardinality $(2, 2)$ are noncrossing, the third one is not. We see from Proposition 2 that the total contribution is

$$\begin{aligned} K_{2,2} &= 2 + K_{\{\{1,3\},\{2,4\}\},u} \\ &= 2 + \tfrac{2}{3} = \tfrac{8}{3}. \end{aligned}$$

The formula for the fourth moment follows.

B. The moment of fifth order

Here two cases require extra attention:

1) $\rho = \{W_1, W_2\}$ *with* $|W_1| = 3$, $|W_2| = 2$*:* There are 10 such partitions, and 5 of them have crossings and constribute with $K_{\{\{1,3\},\{2,4\}\},u}$. The total contribution is therefore

$$\begin{aligned} & 5 + 5 \times K_{\{\{1,3\},\{2,4\}\},u} \\ = \quad & 5 + 5 \times \tfrac{2}{3} = \tfrac{25}{3}. \end{aligned}$$

2) $\rho = \{W_1, W_2, W_3\}$ *with* $|W_1| = |W_2| = 2$, $|W_3| = 1$*:* There are 15 such partitions, of which 5 have crossings. The total contribution is therefore

$$\begin{aligned} & 10 + 5 \times K_{\{\{1,3\},\{2,4\}\},u} \\ = \quad & 10 + 5 \times \tfrac{2}{3} = \tfrac{40}{3}. \end{aligned}$$

The computations for the sixth and seventh order moments are similar, but the details are skipped. These are more tedious in the sense that one has to count the number of partitions with a given block structure, and identify each partition with one of the coefficients listed in Proposition 2.

APPENDIX E
THE PROOF OF THEOREM 2

In order to get the exact expressions in Theorem 2, we now need to keep track of the $K_{\rho,u,N}$ defined by (9), not only the limits $K_{\rho,u}$ (if we had not assumed $\omega = u$, the calculations for $K_{\rho,\omega,N}$ would be much more cumbersome). When ρ is a partition of $\{1, ..., n\}$ and $n \leq 4$, we have that $K_{\rho,u,N} = K_{\rho,u} = 1$ when $\rho \neq \{\{1, 3\}, \{2, 4\}\}$. We also have that

$$K_{\{\{1,3\},\{2,4\}\},u,N} = \frac{2}{3} + \frac{1}{N} + \frac{1}{6N^2}, \tag{79}$$

where we have used that $\sum_{i=1}^{N} i^2 = \frac{N}{3}(N + 1)(N + \frac{1}{2})$ [36]. We also need the exact expression for the quantity

$$T_\rho = \sum_{\substack{(j_1,...,j_n) \\ \text{giving rise to } \rho}} L^{-|\rho|} \mathbf{D}_1(N)(j_1, j_1) \times \cdots \times \mathbf{D}_n(N)(j_n, j_n)$$

from (64) (i.e. we can not add (65) to obtain the approximation (66) here). Setting $D_n^{(N,L)} = tr_L(\mathbf{D}^n(N))$, and $D_\rho^{(N,L)} = \prod_{i=1}^k D_{W_i}^{(N,L)}$, we see that

$$T_\rho = D_\rho^{(N,L)} - \sum_{\rho' > \rho} L^{|\rho'| - |\rho|} T_{\rho'}, \tag{80}$$

which can be used recursively to express the T_ρ in terms of the $D_\rho^{(N,L)}$. We obtain the following formulas for $n = 4$:

$$\begin{aligned}
T_{\{\{1,2,3,4\}\}} &= D_4^{(N,L)} && (81)\\
T_{\{\{1,2,3\},\{4\}\}} &= D_3^{(N,L)} D_1^{(N,L)} - L^{-1} D_4^{(N,L)} && (82)\\
T_{\{\{1,2\},\{3,4\}\}} &= (D_2^{(N,L)})^2 - L^{-1} D_4^{(N,L)} && (83)\\
T_{\{\{1,2\},\{3\},\{4\}\}} &= D_2^{(N,L)} (D_1^{(N,L)})^2 \\
&\quad -2L^{-1}(D_3^{(N,L)} D_1^{(N,L)} \\
&\quad -L^{-1} D_4^{(N,L)}) \\
&\quad -L^{-1}\left((D_2^{(N,L)})^2 - L^{-1} D_4^{(N,L)}\right) \\
&\quad -L^{-2} D_4^{(N,L)} \\
&= D_2^{(N,L)} (D_1^{(N,L)})^2 \\
&\quad -L^{-1}(D_2^{(N,L)})^2 \\
&\quad -2L^{-1} D_3^{(N,L)} D_1^{(N,L)} \\
&\quad +2L^{-2} D_4^{(N,L)} && (84)\\
T_{\{\{1\},\{2\},\{3\},\{4\}\}} &= (D_1^{(N,L)})^4 \\
&\quad -6L^{-1}(D_2^{(N,L)}(D_1^{(N,L)})^2 \\
&\quad -L^{-1}(D_2^{(N,L)})^2 \\
&\quad -2L^{-1} D_3^{(N,L)} D_1^{(N,L)} \\
&\quad +2L^{-2} D_4^{(N,L)}) \\
&\quad -3L^{-2}(D_2^{(N,L)})^2 \\
&\quad +3L^{-3} D_4^{(N,L)} \\
&\quad -4L^{-2} D_3^{(N,L)} D_1^{(N,L)} \\
&\quad +4L^{-3} D_4^{(N,L)} \\
&\quad -L^{-3} D_4^{(N,L)} \\
&= -6L^{-3} D_4^{(N,L)} \\
&\quad +L^{-2}(8 D_3^{(N,L)} D_1^{(N,L)} \\
&\quad +3(D_2^{(N,L)})^2) \\
&\quad -6L^{-1} D_2^{(N,L)} (D_1^{(N,L)})^2 + \\
&\quad (D_1^{(N,L)})^4. && (85)
\end{aligned}$$

For $n = 3$ and $n = 2$ the formulas are

$$\begin{aligned}
T_{\{\{1,2,3\}\}} &= D_3^{(N,L)} && (86)\\
T_{\{\{1,2\},\{3\}\}} &= D_1^{(N,L)} D_2^{(N,L)} - L^{-1} D_3^{(N,L)} && (87)\\
T_{\{\{1\},\{2\},\{3\}\}} &= (D_1^{(N,L)})^3 - 3L^{-1} D_1^{(N,L)} D_2^{(N,L)} \\
&\quad +2L^{-2} D_3^{(N,L)} && (88)\\
T_{\{\{1,2\}\}} &= D_2^{(N,L)} && (89)\\
T_{\{\{1\},\{2\}\}} &= (D_1^{(N,L)})^2 - L^{-1} D_2^{(N,L)}. && (90)
\end{aligned}$$

It is clear that (81)-(85) and (86)-(90) cover all possibilities when it comes to partition block sizes. Using (12)-(13), and putting (79), (81)-(85), and (86)-(90) into (64) we get the expressions in Theorem 2 after some calculations.

A. First order approximations to Theorem 2

If we are only interested in first order approximations rather than exact expressions, (80) gives us

$$T_\rho \approx D_\rho - \sum_{\substack{\rho' > \rho \\ |\rho| - |\rho'| = 1}} L^{-1} D_{\rho'},$$

which is easier to compute. Also, we need only first order approximations to $K_{\rho,u,N}$, which is much easier to compute than the exact expression. For (79), this is

$$K_{\{\{1,3\},\{2,4\}\},u,N} \approx \frac{2}{3} + \frac{1}{N},$$

Inserting these two approximations in (64) gives a first order approximation of the moments.

Appendix F
The proof of Proposition 4

We only state the proof for the case $c = 1$. In [31] it is stated that the asymptotic moment $2n$ (m_{2n}) of certain Hankel and Toeplitz matrices can be expressed in terms of the quantities

$$m_{2n} = \sum_{\substack{\rho \in \mathcal{P}(2n) \\ \rho \text{ has two elements in each block}}} K_{\rho,u} \tag{91}$$

In the language of [31], the formula is not stated exactly like this, but rather in terms of volumes of solution sets of equations of the form (76). This translates to (91), since we in Appendix B interpreted $K_{\rho,u}$ as such volumes. In Proposition A.1 in [31], unbounded support was proved by showing that $(m_{2n})^{1/n} \to \infty$. Again denoting the asymptotic moments of Vandermonde matrices with uniform phase distribution by V_n, we have that $m_{2n} \leq V_{2n}$, since we sum over a greater class of partitions than in (91) when computing the Vandermonde moments. This means that $(V_{2n})^{1/n} \to \infty$ also, so that the asymptotic mean eigenvalue distribution of the Vandermonde matrices have unbounded support also.

Appendix G
The proof of Theorem 3

We will use the fact that

$$\begin{aligned}
K_{\rho,u,N} = \; & \tfrac{1}{(2\pi)^{|\rho|} N^{n+1-|\rho|}} \times \\
& \int_{(0,2\pi)^{|\rho|}} \prod_{k=1}^n \tfrac{1 - e^{jN(x_{b(k-1)} - x_{b(k)})}}{1 - e^{j(x_{b(k-1)} - x_{b(k)})}} \\
& dx_1 \cdots dx_{|\rho|},
\end{aligned} \tag{92}$$

where integration is w.r.t. Lebesgue measure.

For $\rho = 1_n$ Theorem 3 is trivial. We will thus assume that $\rho \neq 1_n$ in the following. We first prove that $\lim_{N\to\infty} K_{\rho,\omega,N}$ exists whenever p_ω is continuous. To simplify notation, define

$$\begin{aligned}
F(\omega) &= \prod_{k=1}^n \frac{1 - e^{jN(\omega_{b(k-1)} - \omega_{b(k)})}}{1 - e^{j(\omega_{b(k-1)} - \omega_{b(k)})}} \\
&= \prod_{k=1}^n \frac{\sin\left(N(\omega_{b(k-1)} - \omega_{b(k)})/2\right)}{\sin\left((\omega_{b(k-1)} - \omega_{b(k)})/2\right)},
\end{aligned}$$

and set $\omega = (\omega_1, ..., \omega_{|\rho|})$ and $d\omega = d\omega_1 \cdots d\omega_{|\rho|}$. Since ω is continuous, there exists a p_{max} such that $p_\omega(\omega_i) \leq p_{max}$ for all ω_i. Then we have that

$$\begin{aligned} |K_{\rho,\omega,N}| &\leq \frac{p_{max}^{|\rho|}}{N^{m+1-|\rho|}} \\ &\times \int_{[0,2\pi)^{|\rho|}} \prod_{k=1}^n \left| \frac{\sin\left(N(x_{b(k-1)} - x_{b(k)})/2\right)}{\sin\left((x_{b(k)} - x_{b(k+1)})/2\right)} \right| dx, \end{aligned}$$

where we have converted to Lebesgue measure, and where we have also written $dx = dx_1 \cdots dx_{|\rho|}$. Consider first the set

$$U = \{\omega | |x_{b(k-1)} - x_{b(k)}| \leq \pi \forall k\}.$$

When $\frac{2\pi}{N} \leq |\omega_{b(k-1)} - \omega_{b(k)}| \leq \pi$, it is clear that

$$\left| \frac{\sin\left(N(x_{b(k-1)} - x_{b(k)})/2\right)}{\sin\left((x_{b(k-1)} - x_{b(k)})/2\right)} \right| \leq \left| \frac{4}{x_{b(k-1)} - x_{b(k)}} \right|, \quad (93)$$

since $\left|\sin\left(N(x_{b(k-1)} - x_{b(k)})/2\right)\right| \leq 1$, and since $|\sin(x)| \geq |\frac{x}{2}|$ when $|x| \leq \frac{\pi}{2}$. When $|x_{b(k-1)} - x_{b(k)}| \leq \frac{2\pi}{N}$ we have that

$$\left| \frac{\sin\left(N(x_{b(k-1)} - x_{b(k)})/2\right)}{\sin\left((x_{b(k-1)} - x_{b(k)})/2\right)} \right| \leq N. \quad (94)$$

Let $k_1, ..., k_{|\rho|} \in \mathbb{Z}$, and assume that $k_{|\rho|} = 0$. By using the triangle inequality, it is clear that on the set

$$D_{k_1,...,k_{|\rho|-1}} = \{\omega | \left| x_i - \frac{2k_i\pi}{N} \right| \leq \frac{\pi}{N} \forall 1 \leq i \leq |\rho|\},$$

when $|k_r - k_s| \geq 2$ for all r, s, the i'th factor in $F(x)$ is bounded by $\frac{4N}{\left(|k_{b(r-1)} - k_{b(r)}| - 1\right)\pi}$ due to (93). Also, when $|k_r - k_s| < 2$ for some r, s, the corresponding factors in $F(x)$ are bounded by N on $D_{k_1,...,k_{|\rho|}}$ due to (94). Note also that the volume of $D_{k_1,...,k_{|\rho|-1}}$ is $(2\pi)^{|\rho|-1} N^{1-|\rho|}$. By adding some more terms (to compensate for the different behaviour for $|k_r - k_s| \geq 2$ and $|k_r - k_s| < 2$), we have that we can find a constant D that

$$\begin{aligned} &\frac{1}{N^{n+1-|\rho|}} \int_U |F(x)| dx \\ &\leq \frac{1}{N^{n+1-|\rho|}} N^n \\ &\times \sum_{\substack{0 \leq k_1, ..., k_{|\rho|-1} < N \\ \text{all } k_i \text{ different}}} \left(\prod_{r=1}^n \frac{D}{|k_{b(r-1)} - k_{b(r)}|} \right) 2\pi (2\pi)^{|\rho|-1} N^{1-|\rho|} \\ &= (2\pi)^{|\rho|} D^n \sum_{\substack{0 \leq k_1, ..., k_{|\rho|-1} < N \\ \text{all } k_i \text{ different}}} \prod_{r=1}^n \frac{1}{|k_{b(r-1)} - k_{b(r)}|}, \end{aligned} \quad (95)$$

where we have integrated w.r.t. $x_{|\rho|}$ also (i.e. $k_{|\rho|}$ is kept constant in (95)). A similar analysis as for U applies for the complement set

$$V = \{\omega | \pi \leq |x_{b(k-1)} - x_{b(k)}| \leq 2\pi \text{ for some } k\},$$

so that we can find a constant C such that

$$\begin{aligned} &\frac{1}{N^{n+1-|\rho|}} \int_{[0,2\pi)^{|\rho|}} |F(x)| dx \\ &\leq C \sum_{\substack{0 \leq k_1, ..., k_{|\rho|-1} < N \\ \text{all } k_i \text{ different}}} \prod_{r=1}^n \frac{1}{|k_{b(r-1)} - k_{b(r)}|}, \end{aligned} \quad (96)$$

It is clear this sum converges: First of all, this is only needed to prove for $\rho = 0_n$, since the summands for $\rho \neq 0_n$ is only a subset of the summands for $\rho = 0_n$.

Secondly, for $\rho = 0_n$, (96) can be bounded by considering convolutions of the following function with itself:

$$f(x) = \begin{cases} \frac{1}{|x|} & \text{for } |x| > 1 \\ 0 & \text{for } |x| \leq 1 \end{cases} \quad (97)$$

The assumption that $f(x) = 0$ in a neighbourhood of zero is due to the fact that the k_i are all different. Note that $|f(x)| \leq \frac{1}{|x|^{1-\epsilon}}$ for any $0 < \epsilon < 1$. Also, the $n-2$-fold convolution (we wait with the $n-1$'th convolution till the end) of $\frac{1}{|x|^{1-\epsilon}}$ with itself exist outside 0 whenever $0 < (n-2)\epsilon < 1$, and is on the form $r\frac{1}{|x|^{1-(n-2)\epsilon}}$ for some constant r [36]. Therefore, (96) is bounded by

$$\begin{aligned} \int_{|x|>1} r \frac{1}{|x|^{1-(n-2)\epsilon}} \frac{1}{|x|} dx &= \int_{|x|>1} r \frac{1}{|x|^{2-(n-2)\epsilon}} dx \\ &= \frac{2r}{(n-2)\epsilon - 1}. \end{aligned}$$

This proves that the entire sum (96) is bounded, and thus also the statement on the existence of the limit $K(\rho, \omega)$ in Theorem 3 when the density is continuous.

For the rest of the proof of Theorem 3 , we first record the following result:

Lemma 2: For any $\epsilon > 0$,

$$\lim_{N \to \infty} \frac{1}{N^{n+1-|\rho|}} \int_{B_{\epsilon,r}} F(\omega) d\omega = 0, \quad (98)$$

where

$$B_{\epsilon,r} = \{(\omega_1, ..., \omega_{|\rho|}) | |\omega_{b(r-1)} - \omega_{b(r)}| > \epsilon\}.$$

Proof: The set $B_{\epsilon,r}$ corresponds to those $k_1, ..., k_{|\rho|}$ in (96) for which $|k_{b(r-1)} - k_{b(r)}| > \frac{N}{2\pi}\epsilon$. Thus, for large N, we sum over $k_1, ..., k_{|\rho|}$ in (96) for which $|k_{b(r-1)} - k_{b(r)}|$ is arbitrarily large. By the convergence of the Fourier integral of $\frac{1}{|x|}$, it is clear that this converges to zero. ■

Define

$$B_\epsilon = \{(\omega_1, ..., \omega_{|\rho|}) | |\omega_i - \omega_j| > \epsilon \text{ for some } i, j\}.$$

If $\omega \in B_\epsilon$, there must exist an r so that $|\omega_{b(r-1)} - \omega_{b(r)}| > \frac{2\epsilon}{n}$, so that $\omega \in B_{r,2\epsilon/n}$. This means that

$$B_\epsilon \subset \cup_r B_{r,2\epsilon/n},$$

so that by Lemma 2 also

$$\lim_{N \to \infty} \frac{1}{N^{n+1-|\rho|}} \int_{B_\epsilon} F(\omega) d\omega = 0.$$

This means that in the integral for $K_{\rho,\omega,N}$, we need only integrate over the ω which are arbitrarily close to the diagonal, (where $\omega_1 = \cdots = \omega_{|\rho|}$). We thus have

$$\begin{aligned} K_{\rho,\omega} &= \lim_{N \to \infty} \frac{1}{N^{n+1-|\rho|}} \int_{[0,2\pi)^{|\rho|}} F(x) \prod_{r=1}^{|\rho|} p_\omega(x_r) dx \\ &= \lim_{N \to \infty} \frac{1}{N^{n+1-|\rho|}} \int_{[0,2\pi)^{|\rho|}} F(x) p_\omega(x_{|\rho|})^{|\rho|} dx \\ &= \lim_{N \to \infty} \frac{1}{N^{n+1-|\rho|}} \int_0^{2\pi} p_\omega(x_{|\rho|})^{|\rho|} \\ &\quad \left(\int_{[0,2\pi)^{|\rho|-1}} F(x) dx_1 \cdots dx_{|\rho|-1} \right) \\ &\quad dx_{|\rho|}. \end{aligned}$$

We used here the fact that the density is continuous. Using that

$$\begin{aligned} &\lim_{N \to \infty} \frac{1}{N^{n+1-|\rho|}} \int_{[0,2\pi)^{|\rho|-1}} F(x) dx_1 \cdots dx_{|\rho|-1} \\ &= (2\pi)^{|\rho|-1} K_{\rho,u} \end{aligned} \quad (99)$$

when $x_{|\rho|}$ is kept fixed at an arbitrary value (this is straightforward by using the methods from the proof of Proposition 1 and

(92)), and again using the fact that the density is continuous, we get that the above equals

$$K_{\rho,u}(2\pi)^{|\rho|-1}\int_0^{2\pi} p_\omega(x_{|\rho|})^{|\rho|}dx_{|\rho|},$$

which is what we had to show.

Appendix H
The proof of Proposition 5

Proposition 5 will follow directly if we can prove the following result:

Lemma 3: Let ω_k ($1\leq k\leq n$) be the uniform distribution on $[\frac{2\pi(k-1)}{n},\frac{2\pi k}{n}]$ and define $\omega_{\lambda_1,\ldots,\lambda_n}$ ($0\leq\lambda_i\leq 1, \lambda_1+\cdots+\lambda_n=1$) as the phase distribution with density $p_{\omega_{\lambda_1,\ldots,\lambda_n}} = \lambda_1 p_{\omega_1}+\cdots+\lambda_n p_{\omega_n}$. Then

$$K_{\rho,\omega_{\frac{1}{n},\ldots,\frac{1}{n}}}\leq K_{\rho,\omega_{\lambda_1,\ldots,\lambda_n}}.$$

Proof: This follows immediately by noting that

$$\begin{aligned}
& K_{\rho,\omega_{\lambda_1,\ldots,\lambda_n}}\\
= \;& K_{\rho,u}(2\pi)^{|\rho|-1}\left(\int_0^{2\pi} p_{\omega_{\lambda_1,\ldots,\lambda_n}}(x)^{|\rho|}dx\right)\\
= \;& K_{\rho,u}(2\pi)^{|\rho|-1}\\
& \times\int_0^{2\pi}(\lambda_1 p_{\omega_1}(x)+\cdots+\lambda_n p_{\omega_n}(x))^{|\rho|}dx\\
= \;& K_{\rho,u}(2\pi)^{|\rho|-1}\times\\
& ((\lambda_1)^{|\rho|}\int_0^{2\pi} p_{\omega_1}(x)^{|\rho|}dx+\cdots\\
& +(\lambda_n)^{|\rho|}\int_0^{2\pi} p_{\omega_n}(x)^{|\rho|}dx)\\
= \;& K_{\rho,u}(2\pi)^{|\rho|-1}\times\\
& ((\lambda_1)^{|\rho|}\int_0^{2\pi} p_{\omega_1}(x)^{|\rho|}dx+\cdots\\
& +(\lambda_n)^{|\rho|}\int_0^{2\pi} p_{\omega_1}(x)^{|\rho|}dx)\\
= \;& K_{\rho,u}(2\pi)^{|\rho|-1}\left((\lambda_1)^{|\rho|}+\cdots+(\lambda_n)^{|\rho|}\right)\\
& \times\int_0^{2\pi} p_{\omega_1}(x)dx\\
\geq \;& K_{\rho,u}(2\pi)^{|\rho|-1}\left(\left(\frac{1}{n}\right)^{|\rho|}+\cdots+\left(\frac{1}{n}\right)^{|\rho|}\right)\\
& \times\int_0^{2\pi} p_{\omega_1}(x)^{|\rho|}dx\\
= \;& K_{\rho,\omega_{\frac{1}{n},\ldots,\frac{1}{n}}},
\end{aligned}$$

where we have used that $x_1^{|\rho|}+\cdots x_n^{|\rho|}$ constrained to $x_1+\cdots+x_n=1$ achieves its minimum for $x_1=\cdots=x_n=\frac{1}{n}$. ■

Appendix I
The proof of Theorem 4

The contribution in the integral $K_{\rho,\omega,N}$ comes only from when the ω_i coincide with the atoms of p. Actually, we evaluate $\frac{1-e^{jN\omega}}{1-e^{j\omega}}$ in points on the form $\omega=\alpha_i-\alpha_j$. This evaluates to $N^n p_i^n$ when all ω_i are chosen equal to the same atom α_j. Since $\lim_{N\to\infty}\frac{1-e^{jN\omega}}{N(1-e^{j\omega})}=0$ for any fixed $\omega\neq 0$, $\lim_{N\to\infty}K_{\rho,\omega,N}N^{-n}=0$ when ω is chosen from nonequal atoms. (57) (with additional $1/N$-factors) thus becomes

$$\begin{array}{l}
\sum_{\rho\in\mathcal{P}(n)}\\
\sum_{\substack{(j_1,\ldots,j_n)\\ \text{giving rise to }\rho}}\\
\sum_{(i_1,\ldots,i_n)}\\
\quad N^{|\rho|-2n-1}c^{|\rho|-1}L^{-|\rho|}\\
\quad (\sum_i N^n p_i^n + a_{\rho,N}N^n))\\
\quad \mathbf{D}_1(N)(j_1,j_1)\mathbf{D}_2(N)(j_2,j_2)\\
\quad \cdots\times\mathbf{D}_n(N)(j_n,j_n),
\end{array}\tag{100}$$

where $\lim_{N\to\infty}a_{\rho,N}=0$. Multiplying both sides with N and letting N go to infinity gives

$$\lim_{N\to\infty}\sum_{\rho\in\mathcal{P}(n)}N^{|\rho|-n}c^{|\rho|-1}\left(\sum_i p_i^n+a_{\rho,N}\right)D_\rho.$$

It is clear that this converges to 0 when $\rho\neq 0_n$ (since $|\rho|<n$ in this case), so that the limit is

$$c^{n-1}\left(\sum_i p_i^n\right)\alpha_{0_n}=c^{n-1}p^{(n)}\lim_{N\to\infty}\prod_{i=1}^n tr_L\left(\mathbf{D}_i(N)\right),$$

which proves the claim

Appendix J
The proof of Theorem 5

We need the following identity [36]:

$$\int_0^\infty x^{-s}e^{jnx}dx=\frac{\Gamma(1-s)}{|n|^{1-s}}e^{\frac{jsgn(n)(1-s)\pi}{2}},$$

where $sgn(x)=1$ if $x>0$, $sgn(x)=-1$ if $x<0$, and 0 otherwise. From this it follows that

$$\begin{array}{l}\int_{-\infty}^{\infty}p_i|x-\alpha_i|^{-s}e^{jnx}dx=\\ 2p_ie^{jn\alpha_i}\frac{\Gamma(1-s)}{|n|^{1-s}}\cos\left(\frac{(1-s)\pi}{2}\right).\end{array}\tag{101}$$

Note that the measure with density p, has the same asymptotics near α_i as the measure with density $p_i|x-\alpha_i|^{-s}$ on

$$\left(-\left(\frac{1-s}{2p_i}\right)^{\frac{1}{1-s}},\left(\frac{1-s}{2p_i}\right)^{\frac{1}{1-s}}\right).$$

As in the proof in Appendix I, the integral for the expansion coefficients is dominated by the behaviour near the points $(\alpha_i,...,\alpha_i)$. To see this, note that the behaviour near the singular points on the diagonal is $O\left(s(|\rho|-n)-1\right)$ when polynomic growth of order s of the density near the singular points is assumed. This is very much related to (96) in Appendix G, since $K_{\rho,\omega}$ here in a similar way can be bounded by (taking into account new powers of N)

$$\begin{array}{l}C\frac{1}{N^{n+ns+1-|\rho|}}N^nN^{-|\rho|}N^{|\rho|s}\\ \times\sum_{\substack{0\leq k_1,\ldots,k_{|\rho|}<N\\ \text{all }k_i\text{ different}}}\prod_{r=1}^n\frac{1}{|k_{b(r-1)}-k_{b(r)}|}\prod_{t=1}^{|\rho|}k_t^{-s}.\end{array}\tag{102}$$

In (102), the N^n-factor appears in exactly the same way as in the proof of Theorem 3 in Appendix G, $N^{-|\rho|}$ appears as a volume in $\mathbb{R}^{|\rho|}$, and $N^{|\rho|s}$ comes from evaluation of the density in the points $x_i = \frac{2k_i\pi}{N}$, $1 \leq i \leq |\rho|$). Since $\frac{1}{|x|^s}$ has a bounded integral around 0, and since the sum still converges (it is dominated by (96)), (102) is

$$O\left(s(|\rho|-n)-1\right).$$

This has it's highest order when $|\rho| = n$, so that we can restrict to looking at 0_n. Note also that we may just as well assume that $p_\omega(x)$ is identical to $p_i|x-\omega_i|^{-s}$ at an interval around ω_i, since $\lim_{x\to\alpha_i}|x-\alpha_i|^s p_\omega(x) = p_i$ implies that

$$p_\omega(x) = p_i|x-\omega_i|^{-s} + k(x)|x-\omega_i|^{-s} \tag{103}$$

where $\lim_{x\to\omega_i} k(x) = 0$. It is straightforward to see that the contribution of the second part in (103) to (102) vanishes as $N \to \infty$, so that we may just as well assume that $p_\omega(x)$ is identical to $p_i|x-\omega_i|^{-s}$ at an interval around ω_i, as claimed. Also, since

$$\lim_{n\to\infty}\int_{|x|>\epsilon} x^{-s}e^{jnx}dx = 0$$

for all $\epsilon > 0$, and since the contributions from large n dominate in (104) below (since $\sum_n |n|^{-s}$ diverges), it is clear that we can restrict to an interval around ω_i when computing the limit also (since p_ω is continuous outside the singularity points, this follows from Theorem 3, and due to the additional $\frac{1}{N^s}$-factor added to (1)). After restricting to 0_n, multiplying both sides with N, summing over all singularity points, and using (101), we obtain the approximation

$$\begin{aligned}\sum_{(i_1,\ldots,i_n)}\sum_a & \\ & N^{-ns}c^{n-1} \\ & \times\left(2p_a\Gamma(1-s)\cos\left(\frac{(1-s)\pi}{2}\right)\right)^n \\ & \times\prod_{k=1}^n \frac{e^{j(i_{k-1}-i_k)\alpha_a}}{|i_{k-1}-i_k|^{1-s}} \\ & \times tr_L(\mathbf{D}_1(N))\times\cdots\times tr_L(\mathbf{D}_n(N))\end{aligned} \tag{104}$$

to (57). Since $\prod_{k=1}^n e^{j(i_{k-1}-i_k)\alpha_a} = 1$, we recognize

$$\begin{aligned} q^{(n,N)} = & \left(2\Gamma(1-s)\cos\left(\frac{(1-s)\pi}{2}\right)\right)^n \left(\sum_a p_a^n\right)\times \\ & \sum_{(i_1,\ldots,i_n)} N^{-ns}\prod_{k=1}^n \frac{1}{|i_{k-1}-i_k|^{1-s}},\end{aligned}$$

as a factor in (104) such that the limit of (104) as $N \to \infty$ can be written

$$c^{n-1}\lim_{N\to\infty} q^{(n,N)} \lim_{N\to\infty}\prod_{i=1}^n tr_L\left(\mathbf{D}_i(N)\right).$$

It therefore suffices to prove that $\lim_{N\to\infty} q^{(n,N)} = q^{(n)}$. To see this, write

$$\begin{aligned}\frac{N^{-s}}{|i_{k-1}-i_k|^{1-s}} &= \frac{1}{N}\frac{1}{\left(\frac{1}{N}\right)^{1-s}|i_{k-1}-i_k|^{1-s}} \\ &= \frac{1}{N}\frac{1}{\left|\frac{i_{k-1}}{N}-\frac{i_k}{N}\right|^{1-s}}.\end{aligned}$$

Summing over all $1 \leq i_1,\ldots,i_n \leq N$, it is clear from this that $q^{(n,N)}$ can be viewed as a Riemann sum which converges to $q^{(n)}$ as $N \to \infty$.

Appendix K
The proof of Theorem 7 and Corollary 1

Proof of Theorem 7: we define S_j to be the blocks of σ, i.e.

$$S_j = \{k | i_k = j\}.$$

Note that Theorem 3 guarantees that the limit $K_{\rho,\omega} = \lim_{N\to\infty} K_{\rho,\omega,N}$ exists. The partition ρ simply is a grouping of random variables into independent groups. It is therefore impossible for a block in ρ to contain elements from both S_1 and S_2, so that any block is contained in either S_1 or S_2. As a consequence, $\rho \leq \sigma$. ∎

Until now, we have not treated mixed moments of the form

$$\mathbf{D}_1(N)\mathbf{V}_{i_2}\mathbf{V}_{i_2}^H\mathbf{D}_2(N)\mathbf{V}_{i_3}\mathbf{V}_{i_3}^H\cdots\times\mathbf{D}_n(N)\mathbf{V}_{i_1}\mathbf{V}_{i_1}^H,$$

which are the same as the mixed moments of Theorem 7 except for the position of the $\mathbf{D}_i(N)$. We will not go into depths on this, but only remark that this case can be treated in the same vein as generalized Vandermonde matrices by replacing the density p_f (or p_λ in case of continuous generalized Vandermonde matrices) with functions $p_{D_i}(x)$ defined by $p_{D_i}(x) = \mathbf{D}_i(N)(\lfloor Lx\rfloor, \lfloor Lx\rfloor)$ for $0 \leq x \leq 1$. This also covers the case of mixed moments of independent, generalized Vandermonde matrices (and, in fact, there are no restrictions on the horizontal and vertical phase densities p_{ω_i} and p_{λ_j} for each matrix. They may all be different). The proof for this is straightforward.

Proof of Corollary 1: this follows in the same way as Proposition 3 is proved from Proposition 2, by only considering ρ which are less than σ, and also by using Theorem 3. σ are for the listed moments $\{\{1\},\{2\}\}$, $\{\{1,3\},\{2,4\}\}$, and $\{\{1,3,5\},\{2,4,6\}\}$, respectively. ∎

Appendix L
The proofs of Proposition 6 and 7

The moments $E\left[tr_n\left(\mathbf{W}^i\right)\right]$ will be related to the moments P_i through three convolution stages:

1) relating the moments of $\mathbf{W}$ with the moments of

$$\mathbf{\Gamma} = \mathbf{V}\mathbf{P}^{\frac{1}{2}}\left(\frac{1}{K}\mathbf{S}\mathbf{S}^H\right)\mathbf{P}^{\frac{1}{2}}\mathbf{V}^H, \tag{105}$$

from which we easily get the moments of

$$\tilde{\mathbf{S}} = \left(\frac{1}{K}\mathbf{S}\mathbf{S}^H\right)\mathbf{P}^{\frac{1}{2}}\mathbf{V}^H\mathbf{V}\mathbf{P}^{\frac{1}{2}}, \tag{106}$$

2) relating the moments of $\mathbf{S}$ with the moments of

$$\mathbf{T} = \mathbf{P}\mathbf{V}^H\mathbf{V}, \tag{107}$$

3) relating the moments of $\mathbf{T}$ with the moments of $\mathbf{P}$.

For the first stage, the moments of $\hat{\mathbf{W}}$ and $\mathbf{\Gamma}$ relate through the formulas

$$\begin{aligned} E\left[tr_n\left(\mathbf{W}\right)\right] &= E\left[tr_N\left(\mathbf{\Gamma}\right)\right]+\sigma^2 && (108)\\ E\left[tr_n\left(\mathbf{W}^2\right)\right] &= E\left[tr_N\left(\mathbf{\Gamma}^2\right)\right]\\ &\quad +2\sigma^2(1+c_1)E\left[tr_N\left(\mathbf{\Gamma}\right)\right]\\ &\quad +\sigma^4(1+c_1) && (109)\\ E\left[tr_n\left(\mathbf{W}^3\right)\right] &= E\left[tr_N\left(\mathbf{\Gamma}^3\right)\right]\\ &\quad +3\sigma^2(1+c_1)E\left[tr_N\left(\mathbf{\Gamma}^2\right)\right]\\ &\quad +3\sigma^2 c_1 E\left[\left(tr_N\left(\mathbf{\Gamma}\right)\right)^2\right]\\ &\quad +3\sigma^4\left(c_1^2+3c_1+1+\frac{1}{K^2}\right)E\left[tr_N\left(\mathbf{\Gamma}\right)\right]\\ &\quad +\sigma^6\left(c_1^2+3c_1+1+\frac{1}{K^2}\right),, && (110)\end{aligned}$$

which are obtained by replacing $\mathbf{R}$ in [32] by $\mathbf{V}\mathbf{P}^{\frac{1}{2}}\mathbf{S}$, with $c=c_1=\frac{N}{K}$. For the second part of the first stage, note that

$$\begin{aligned} E\left[tr_N\left(\mathbf{\Gamma}^k\right)\right] &= c_2 E\left[tr_L\left(\tilde{\mathbf{S}}^k\right)\right] && (111)\\ E\left[\left(tr_N\left(\mathbf{\Gamma}\right)\right)^k\right] &= c_2^k E\left[\left(tr_L\left(\tilde{\mathbf{S}}\right)\right)^k\right], && (112)\end{aligned}$$

where $c_2=\frac{L}{N}$. We can now apply Theorem 2 to obtain

$$\begin{aligned} c_3 E\left[tr_L\left(\tilde{\mathbf{S}}\right)\right] &= c_3 E\left[tr_L\left(\mathbf{T}\right)\right] && (113)\\ c_3 E\left[tr_L\left(\tilde{\mathbf{S}}^2\right)\right] &= c_3 E\left[tr_L\left(\mathbf{T}^2\right)\right]\\ &\quad +c_3^2 E\left[\left(tr_L\left(\mathbf{T}\right)\right)^2\right] && (114)\\ c_3 E\left[tr_L\left(\tilde{\mathbf{S}}^3\right)\right] &= \left(1+K^{-2}\right)c_3 E\left[tr_L\left(\mathbf{T}^3\right)\right]\\ &\quad +3c_3^2 E\left[\left(tr_L\mathbf{T}\right)tr_L\left(\mathbf{T}^2\right)\right]\\ &\quad +c_3^3 E\left[\left(tr_L\left(\mathbf{T}\right)\right)^3\right] && (115)\\ E\left[\left(tr_L\left(\tilde{\mathbf{S}}\right)\right)^2\right] &= E\left[\left(tr_L\left(\mathbf{T}\right)\right)^2\right]\\ &\quad +\frac{1}{KL}E\left[tr_L\left(\mathbf{T}^2\right)\right], && (116)\end{aligned}$$

where $c_3=\frac{L}{K}$, and $\mathbf{T}=\mathbf{P}\mathbf{V}^H\mathbf{V}$. (108)-(110), (111)-(112), and (113)-(116) can be combined to

$$\begin{aligned} E\left[tr_n\left(\mathbf{W}\right)\right] &= c_2 E\left[tr_L\left(\mathbf{T}\right)\right]+\sigma^2 && (117)\\ E\left[tr_n\left(\mathbf{W}^2\right)\right] &= c_2 E\left[tr_L\left(\mathbf{T}^2\right)\right]+c_2c_3E\left[\left(tr_L\left(\mathbf{T}\right)\right)^2\right]\\ &\quad +2\sigma^2(c_2+c_3)E\left[tr_L\left(\mathbf{T}\right)\right]+\sigma^4(1+c_1) && (118)\\ E\left[tr_n\left(\mathbf{W}^3\right)\right] &= c_2\left(1+\frac{1}{K^2}\right)E\left[tr_L\left(\mathbf{T}^3\right)\right]\\ &\quad +3c_2c_3E\left[\left(tr_L\left(\mathbf{T}\right)\right)\left(tr_L\left(\mathbf{T}^2\right)\right)\right]\\ &\quad +c_2c_3^2E\left[\left(tr_L\left(\mathbf{T}\right)\right)^3\right]\\ &\quad +3\sigma^2\left((1+c_1)c_2+\frac{c_1c_2^2}{KL}\right)E\left[tr_L\left(\mathbf{T}^2\right)\right]\\ &\quad +3\sigma^2c_3(c_3+2c_2)E\left[\left(tr_L\left(\mathbf{T}\right)\right)^2\right]\\ &\quad +3\sigma^4\left(c_1^2+3c_1+1+\frac{1}{K^2}\right)c_2E\left[tr_L\left(\mathbf{T}\right)\right]\\ &\quad +\sigma^6\left(c_1^2+3c_1+1+\frac{1}{K^2}\right). && (119)\end{aligned}$$

Up to now, all formulas have provided exact expressions for the expectations. For the next step, exact expressions for the expectations are only known when the phase distributions are uniform, in which case the formulas are given by Theorem 2:

$$\begin{aligned} c_2E\left[tr_L\left(\mathbf{T}\right)\right] &= c_2tr_L(\mathbf{P}) && (120)\\ c_2E\left[tr_L\left(\mathbf{T}^2\right)\right] &= \left(1-N^{-1}\right)c_2tr_L(\mathbf{P}^2)\\ &\quad +c_2^2(tr_L(\mathbf{P}))^2 && (121)\\ c_2E\left[tr_L\left(\mathbf{T}^3\right)\right] &= \left(1-3N^{-1}+2N^{-2}\right)c_2tr_L(\mathbf{P}^3)\\ &\quad +3\left(1-N^{-1}\right)c_2^2tr_L(\mathbf{P})tr_L(\mathbf{P}^2)\\ &\quad +c_2^3(tr_L(\mathbf{P}))^3 && (122)\\ E\left[\left(tr_L\left(\mathbf{T}\right)\right)^2\right] &= tr_L(\mathbf{P})^2 && (123)\\ E\left[\left(tr_L\left(\mathbf{T}\right)\right)^3\right] &= tr_L(\mathbf{P})^3 && (124)\end{aligned}$$

$$\begin{aligned} &E\left[\left(tr_L\left(\mathbf{T}\right)\right)\left(tr_L\left(\mathbf{T}^2\right)\right)\right]\\ &=\left(1-N^{-1}\right)tr_L(P)tr_L(\mathbf{P}^2)+c_2(tr_L(\mathbf{P}))^3. \end{aligned} \quad (125)$$

If the phase distribution ω is not uniform, Theorem 1 and Theorem 3 gives the following approximation:

$$\begin{aligned} c_2E\left[tr_L\left(\mathbf{T}\right)\right] &= c_2tr_L(\mathbf{P}) && (126)\\ c_2E\left[tr_L\left(\mathbf{T}^2\right)\right] &\approx c_2tr_L(\mathbf{P}^2)+c_2^2I_2(tr_L(\mathbf{P}))^2 && (127)\\ c_2E\left[tr_L\left(\mathbf{T}^3\right)\right] &\approx c_2tr_L(\mathbf{P}^3)+3c_2^2I_2tr_L(\mathbf{P})tr_L(\mathbf{P}^2)\\ &\quad +c_2^3I_3(tr_L(\mathbf{P}))^3 && (128)\\ E\left[\left(tr_L\left(\mathbf{T}\right)\right)^2\right] &= (tr_L\mathbf{P})^2 && (129)\\ E\left[\left(tr_L\left(\mathbf{T}\right)\right)^3\right] &= (tr_L\mathbf{P})^3 && (130)\end{aligned}$$

$$\begin{aligned} &E\left[\left(tr_L\left(\mathbf{T}\right)\right)\left(tr_L\left(\mathbf{T}^2\right)\right)\right]\\ &\approx tr_L(\mathbf{P})tr_L(\mathbf{P}^2)+c_2I_2(tr_L(\mathbf{P}))^3, \end{aligned} \quad (131)$$

where the approximation is $O(N^{-1})$, and where I_k is defined by (47).

Proposition 7 is proved by combining (117)-(119) with (120)-(125), while Proposition 6 is proved by combining (117)-(119) with (126)-(131). Proposition 8 is proved by first observing that the roles of L and N are interchanged, since the Vandermonde matrix is replaced by its transpose. This means that we obtain the formulas (117)-(119), with c_1 and c_3 interchanged, and c_2 replaced with $\frac{1}{c_2}$. The matrix $\mathbf{T}$ is now instead $\mathbf{V}\mathbf{V}^H$, and these can be scaled to obtain the moments of $\mathbf{V}^H\mathbf{V}$. Finally the integrals I_n or the angle α can be estimated from these moments, using (126)-(131) with the moments of $\mathbf{P}$ replaced with 1 (since no additional power matrix is included in the model).

Matlab code for implementing the different steps here (like (108)-(110), (113)-(116), and (120)-(125)) can be found in [34].

REFERENCES

[1] C. S. Burrus and T. W. Parks, *DFT/FFT and Convolution Algorithms*. New York: John Wiley, 1985.

[2] G. H. Golub and C. F. V. Loan, *Matrix Computations*. John Hopkins University Press, 1983.

[3] R. Norberg, "On the Vandermonde matrix and its application in mathematical finance," *working paper no. 162, Laboratory of Actuarial Mathematics, Univ. of Copenhagen*, 1999.

[4] L. Sampaio, M. Kobayashi, Ø. Ryan, and M. Debbah, "Vandermonde matrices for security applications," *work in progress*, 2008.

[5] J. J. Waterfall, J. Joshua, F. P. Casey, R. N. Gutenkunst, K. S. Brown, C. R. Myers, P. W. Brouwer, V. Elser, and J. P. Sethna, "Sloppy-model universality class and the Vandermonde matrix," *Physical Review Letters*, vol. 97, no. 15, 2006.

[6] L. Sampaio, M. Kobayashi, Ø. Ryan, and M. Debbah, "Vandermonde frequency division multiplexing," *9th IEEE Workshop on Signal Processing Advances for wireless applications, Recife, Brazil*, 2008.

[7] R. Schmidt, "Multiple emitter localization and signal parameter estimation," in *Proceedings of the RADC, Spectal Estimation Workshop, Rome*, 1979, pp. 243–258.

[8] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, pp. 387–392, 1985.

[9] D. H. Johnson and D. E. Dudgeon, *Array Signal processing: Concepts and Techniques*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[10] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, pp. 984–995, July 1989.

[11] B. Porat and B. Friedlander, "Analysis of the asymptotic relative efficiency of the MUSIC algorithm," *IEEE Transactions Acoustics Speech and Signal Processing*, vol. 36, pp. 532–544, apr. 1988.

[12] A. Klein and P. Spreij, "On Stein's equation, Vandermonde matrices and Fisher's information matrix of time series processes. part I: The autoregressive moving average process." *Universiteit van Amsterdam, AE-Report 7/99*, 1999.

[13] Z. Wang, A. Scaglione, G. Giannakis, and S. Barbarossa, "Vandermonde-Lagrange mutually orthogonal flexible transceivers for blind CDMA in unknown multipath," in *Proc. of IEEE-SP Workshop on Signal Proc. Advances in Wireless Comm.*, May 1999, pp. 42–45.

[14] V. Girko, *Theory of Random Determinants*. Kluwer Academic Publishers, 1990.

[15] M. L. Mehta, *Random Matrices*, 2nd ed. New York: Academic Press, 1991.

[16] A. Nordio, C.-F. Chiasserini, and E. Viterbo, "Reconstruction of multidimensional signals from irregular noisy samples," *to appear in IEEE Trans. Signal Process.*, 2008, arxiv.org/abs/0804.3255.

[17] C. Bordenave, "Eigenvalues of Euclidean random matrices," 2008, arxiv.org/abs/math.PR/0606624.

[18] R. R. Muller, "A random matrix model of communication via antenna arrays," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2495–2506, 2002.

[19] S. Thorbjørnsen, "Mixed moments of Voiculescu's Gaussian random matrices," *J. Funct. Anal.*, vol. 176, no. 2, pp. 213–246, 2000.

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

[21] T. Anderson, "Asymptotic theory for principal component analysis," *Annals of Mathematical Statistics*, vol. 34, pp. 122–148, mar. 1963.

[22] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems," *IEEE Trans. on Information Theory*, vol. 43, pp. 499–511, mar. 1977.

[23] Ø. Ryan and M. Debbah, "Free deconvolution for signal processing applications," *Submitted to IEEE Trans. on Information Theory*, 2007, http://arxiv.org/abs/cs.IT/0701025.

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

[25] V. Marchenko and L. Pastur, "Distribution of Eigenvalues for Some Sets of Random Matrices," *Math USSR Sb.*, vol. 1, pp. 457–483, 1967.

[26] S. Shamai and S. Verdu, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. Information Theory*, pp. 1302–1326, May 2001.

[27] E. Telatar, "Capacity of multi-antenna gaussian channels," *Eur. Trans. Telecomm. ETT*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[28] M. Debbah, W. Hachem, P. Loubaton, and M. de Courville, "MMSE Analysis of Certain Large Isometric Random Precoded Systems," *IEEE Transactions on Information Theory, Volume: 49 Issue: 5, Page(s): 1293 -1311*, May 2003.

[29] A. Nica and R. Speicher, *Lectures on the Combinatorics of Free Probability*. Cambridge University Press, 2006.

[30] M. V. H. Krim, "Two decades of array signal processing research," *IEEE Signal Processing Magazine*, pp. 67 – 94, July 1996.

[31] W. Bryc, A. Dembo, and T. Jiang, "Spectral measure of large random Hankel, Marko and Toeplitz matrices," *The Annals of Probability*, vol. 34, no. 1, pp. 1–38, 2006.

[32] Ø. Ryan and M. Debbah, "Channel capacity estimation using free probability theory," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5654–5667, November 2008.

[33] J. A. Mingo and R. Speicher, "Second order freeness and fluctuations of random matrices: I. Gaussian and Wishart matrices and cyclic Fock spaces," *J. Funct. Anal.*, vol. 235, no. 1, pp. 226–270, 2006.

[34] Ø. Ryan, *Tools for convolution with Vandermonde matrices*, 2008, http://ifi.uio.no/˜oyvindry/vandermonde/.

[35] P. Comon, "Independent Component Analysis, a new concept?" *Signal Processing Elsevier, special issue on Higher-Order Statistics*, vol. 36, pp. 287–314, 1994.

[36] K. Rottmann, *Mathematische Formelsammlung*. B.I. Wissenschaftsverlag, 1991.

[37] R. Seroul and D. O'Shea, *Programming for Mathematicians*. Springer, 2000.

[38] T. Rappaport, *Wireless Communications, Principles and Practice*. New Jersey: Prentice-Hall, 1996.

[39] L. S. Cardoso, M. Debbah, P. Bianchi, and J. Najim, "Cooperative Spectrum Sensing Using Random Matrix Theory," *International Symposium on Wireless Pervasive Computing*, 2008.

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

[41] D. Ganesan, S. Ratnasamy, H. Wang, and D. Estrin, "Coping with Irregular Spatio-temporal Sampling in Sensor Networks," *ACM SIGCOMM*, vol. 34, pp. 125–130, 2004.

[42] A. Nordio, C.-F. Chiasserini, and E. Viterbo, "Performance of Field Reconstruction Techniques with Noise and Uncertain Sensor Locations," *IEEE Transactions on Signal Processing, submitted*, 2006.

[43] T. Strohmer, T. Binder, and M. Süssner, "How to Recover Smooth Object Boundaries in Noisy Medical Images," *IEEE ICIP'96, Lausanne*, pp. 331–334, 1996.

[44] D. S. Early and D. G. Long, "Image Reconstruction and Enhanced Resolution Imaging from Irregular Samples," *IEEE Transactions on Geoscience and Remotee Sensing*, vol. 39, pp. 291–302, 2001.

[45] M. Rauth and T. Strohmer, "A frequency Domain Approach to the Reconvery of Geophysical Potenials," *Proc. Conf. SampTA'97, Aveiro/Portugal*, pp. 109–114, 1997.

[46] E. J. Candès and J. Romberg, "Sparsity and Incoherence in Compressive Sampling," *Inverse Problems 23*, pp. 969–985, 2007.

[47] H. G. Feichtinger, K. Gröchenig, and T. Strohmer, "Efficient Numerical Methods in Non-Uniform Sampling Theory," *Numerische Mathematik*, vol. 69, pp. 423–440, 1995.

[48] A. Nordio, C.-F. Chiasserini, and E. Viterbo, "Signal reconstruction in multidimensional sensor fields," *2008 International Zurich Seminar on Communications (IZS), Zurich*, 2008.