Finite Dimensional Statistical Inference

Öyvind Ryan, Member, IEEE, Antonia Masucci, Sheng Yang, Member, IEEE, and Mérouane Debbah, Senior Member, IEEE

Abstract—In this paper, we derive the explicit series expansion of the eigenvalue distribution of various models, namely the case of non-central Wishart distributions as well as one sided correlated zero mean Wishart distributions. The tools used are borrowed from the free probability framework which have been quite successful for high dimensional statistical inference (when the size of the matrices tends to infinity), also known as free deconvolution. This contribution focuses on the finite Gaussian case and proposes algorithmic methods to compute the moments. Cases where the asymptotic results fail to apply are discussed.

Index Terms—Gaussian matrices, Random Matrices, convolution, limiting eigenvalue distribution.

I. INTRODUCTION

Random matrix and free probability theory have fruitful applications in many fields of research, such as digital communication [1], mathematical finance [2] and nuclear physics [3]. In particular, the free probability framework [4], [5], [6], [7], [8] can be used for high dimensional statistical inference (or free deconvolution), i.e., to retrieve the eigenvalue distributions of involved functionals of random matrices. The general idea of deconvolution is related to the following problem [9]:

Given \( A, B \) two \( n \times n \) independent square Hermitian (or symmetric) random matrices:
1) Can one derive the eigenvalue distribution of \( A \) from the ones of \( A + B \) and \( B \)? If feasible in the large \( n \)-limit, this operation is named additive free deconvolution,
2) Can one derive the eigenvalue distribution of \( A \) from the ones of \( AB \) and \( B \)? If feasible in the large \( n \)-limit, this operation is named multiplicative free deconvolution.

In the literature, deconvolution for the large \( n \)-limit has been studied, and the methods generally used to compute it are the moments method [4] or the Stieltjes transform method [10]. The expressions turn out to be quite simple if some kind of asymptotic freeness [8] of the matrices involved is assumed. However, freeness usually does not hold for finite matrices. Quite remarkably, the method of moments can still be used to propose an algorithmic method to compute these operations. The goal of this contribution is exactly to propose a general finite dimensional statistical inference framework based on the method of moments, which is implemented in software. As the calculations are quite tedious and for sake of clarity we focus in this contribution on Gaussian matrices\(^1\).

The moments method [9] is based on the relations between the moments of the different matrices involved. It provides a series expansion of the eigenvalue distribution of the involved matrices. For a given \( n \times n \) matrix \( A \), the \( p \)-th moment of \( A \) is defined as:

\[
t^n_p = \mathbb{E} \left[ \text{tr}(A^p) \right] = \int \lambda^p d\rho_n(\lambda)
\]

where \( \text{tr} \) is the normalized trace, and \( d\rho_n \) is the associated empirical mean measure defined by \( d\rho_n(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \delta(\lambda - \lambda_i) \), where \( \lambda_i \) are the eigenvalues of \( A \). Quite remarkably, when \( n \to \infty \), \( t^n_p \) converges in many cases almost surely to an analytical expression \( t^n_p \) that depends only on some specific parameters of \( A \) (such as the distribution of its entries)\(^2\). This enables to reduce the dimensionality of the problem and simplifies the computation of convolution of measures. In recent works deconvolution has been analyzed when \( n \to \infty \) for some particular matrices \( A \) and \( B \), such as when \( A \) and \( B \) are free [13], or \( A \) random Vandermonde and \( B \) diagonal [11], [12].

The inference framework described in this contribution is based on the moments method in the finite case, i.e. it takes a set of moments as input and produces sets of moments as output, with the dimensions of the matrices considered finite. The framework is flexible enough to allow for repeated combinations of the random matrices we consider, and the patterns in such combinations are reflected nicely in the algorithms. The framework also lends itself naturally to combinations with other types of random matrices, for which support has already been implemented in the framework [12]. This flexibility, exploited with the method of moments, is somewhat in contrast to methods such as the Stieltjes transform method [10], where combining patterns of matrices naturally leads to more complex equations for the Stieltjes transforms (when possible) and can only be performed in the large \( n \)-limit. While the simplest patterns we consider are sums and products, we also consider products of many independent matrices. The algorithms are based on iterations through partitions and permutations as in [14], where the case of a Wishart matrix was considered. Our methods build heavily on the simple form which the moments of complex Gaussian random variables have, as exploited in [14]. We remark that it is possible to implement the method of moments in a different way also [15], [16].

\(^1\)Cases such as Vandermonde matrices can also be implemented in the same vein [11], [12]. The general case is, however, more difficult.

\(^2\)Note that in the following, when speaking of moments of matrices, we refer to the moments of the associated measure.
However, we are not aware of any attempts to make an
inference framework as general as the one presented here. The case presented in [16], for instance, handles only the case where one Wishart matrix is involved.

An important contribution of this paper is to present some finite-dimensional random matrix patterns where the methods presented here are needed. For such patterns, we will not be able to combine a set of observations into a larger compound observation matrix on the same form, which would make asymptotic results applicable. For the other cases presented here, such a stacking of observations is possible. In addition, we can then also stack the observations in different ways, and apply the most optimal stacking [17].

The paper is organized as follows. Section II provides background essentials on random matrix theory, while Section III provides background essentials on combinatorics needed to state the main results. Section III is rather technical, but it is not necessary to understand all details therein to understand the statement of the main results. These are summarized in Section IV. First, algorithms for the simplest patterns (sums and products of random matrices) in the finite dimensional statistical inference framework are presented. Then, recursive algorithms for products of many Wishart matrices and a statistical inference framework are presented. Then, recursive algorithms for products of many Wishart matrices and a statistical inference framework as general as the one presented here. Section VI presents details on the software implementation of the finite dimensional random matrix patterns where the methods apply to the most optimal stacking [17].

II. RANDOM MATRIX BACKGROUND ESSENTIALS

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

\[
\text{Tr}(A) = \sum_{i=1}^{n} a_{ii},
\]

where \(a_{ii}\) are the diagonal elements of the \(n \times n\) matrix \(A\). We also let tr be the normalized trace, defined by \(\text{tr}(A) = \frac{1}{n} \text{Tr}(A)\). When \(A\) is non-random, there is of course no need to take the expectation in (1). \(D\) will in general be used to denote such non-random matrices, and if \(D_1, \ldots, D_s\) are such matrices, we will write

\[
D_{i_1, \ldots, i_s} = \text{tr}(D_{i_1} \cdots D_{i_s}),
\]

whenever \(1 \leq i_1, \ldots, i_s \leq r\). (2) are also called mixed moments.

With the empirical eigenvalue distribution of a hermitian random matrix \(A\), we mean the (random) function

\[
F_A(\lambda) = \frac{\#\{i|\lambda_i \leq \lambda\}}{n},
\]

where \(\lambda_i\) are the (random) eigenvalues of \(A\). In many cases, the moments determine the distribution of the eigenvalues [18]. Due to the expectation in (1), the results in this paper thus apply to the mean eigenvalue distribution of certain random matrices.

We will also encounter quantities on the form

\[
\mathbb{E} [\text{tr}(A^{p_1})\text{tr}(X^{p_2}) \cdots \text{tr}(A^{p_k})],
\]

which are much more general than moments. (4) will in the following be given an interpretation in terms of circles: the matrix indices in (4) will be visualized as points on a circle, with one circle associated with each trace. (4) is thus represented as \(k\) disjoint circles, where there are \(p_1, \ldots, p_k\) points on the different circles, respectively. Since we will work much with such circular representations, we need the following definition:

**Definition 1:** With addition/subtraction modulo \(p_1, \ldots, p_k\) we will mean addition/subtraction performed in such a way that the result stays within the same interval from \([1, \ldots, 2p_1], [2p_1 + 1, \ldots, 2p_2], \ldots, [2p_{k-1} + 1, 2p_k]\).

Note that this is nothing other than the addition/subtraction modulo operation confined to take values in the interval which the start value lies in. In the following, the values \(p_1, \ldots, p_k\) will be implicitly assumed, and therefore we will only mention these when this is strictly needed.

The matrix sizes in the following will be denoted \(n \times N\) for rectangular matrices, \(n \times n\) for square matrices. We remark that some formulas in this paper would be simpler if non-normalized traces were used [14]. We have still used the normalized trace, for compatibility with the asymptotic case where \(\lim_{N \to \infty} \frac{n}{N} = c\), i.e., where the number of columns and the number of rows grow at the same rate. The random matrices we consider will be using Gaussian matrices as building blocks, and they may be either selfadjoint or complex. A standard complex Gaussian matrix \(X\) has i.i.d. complex Gaussian entries with zero mean and unit variance (in particular, the real and imaginary parts of the entries are independent, each with variance \(\frac{1}{2}\)). A selfadjoint Gaussian matrix \(X\) has i.i.d. entries only above the main diagonal, with the real and imaginary part independent with variance \(\frac{1}{2}\) [8].

III. COMBINATORIAL CONCEPTS

To prove the results of this paper, random matrix concepts need to be combined with concepts from partition theory. \(\mathcal{P}(n)\) will denote the partitions of \(\{1, \ldots, n\}\). For a partition \(\rho = \{W_1, \ldots, W_r\} \in \mathcal{P}(n)\), \(W_1, \ldots, W_r\) denote its blocks, while \(|\rho| = r\) denotes the number of blocks. We will write \(k \sim_{\rho} l\) when \(k\) and \(l\) belong to the same block of \(\rho\). Partition notation is adapted to mixed moments in the following way:

**Definition 2:** For \(\rho = \{W_1, \ldots, W_k\}\), with \(W_i = \{w_{i,1}, \ldots, w_{i,|W_i|}\}\), we define

\[
D_{W_i} = D_{w_{i,1}, \ldots, w_{i,|W_i|}}
\]

and

\[
D_{\rho} = \prod_{i=1}^{k} D_{W_i}.
\]
We need the following definitions, taken from [14].

Definition 3: Let \( p \) be a positive integer, \( S_p \) the symmetric group of \( \{1, 2, \ldots, p\} \), and \( \pi \in S_p \). Then \( \pi \in S_{2p} \) is defined by

\[
\hat{\pi}(2j - 1) = 2\pi^{-1}(j), \quad j \in \{1, 2, \ldots, p\} \\
\hat{\pi}(2j) = 2\pi(j) - 1, \quad j \in \{1, 2, \ldots, p\}.
\]

The numbers \( 1, \ldots, 2p \) should in this definition be interpreted as the factors in a matrix product on the form \( (XX^H)^p \), with \( \hat{\pi} \) describing which Gaussians are equal when considering all possible contributions in the matrix product. We will visualize each such factor as an edge on a circle, indexed in the same order as the factors, and the vertices connecting the edges as the corresponding matrix indices, so that edges \( 2i - 1 \) and \( 2i \) are connected with vertex \( 2i \). This is illustrated in Figure 1. Note that \( \hat{\pi} \) has period two, i.e., \( \hat{\pi} \) applied twice gives the identity permutation.

Definition 4: We associate to \( \pi \) an equivalence relation \( \rho = \rho(\pi) \) on \( \{1, \ldots, 2p\} \) generated by

\[
j \sim_\rho \hat{\pi}(j) + 1,
\]

(7)

This equivalence relation will be used to describe exactly when the matrix indices in our circular representation are equal. In the following, equivalence relations will interchangeably also be referred to as partitions. If \( X \) is \( n \times N \), each odd vertex \( (1, 3, 5, \ldots) \) is associated with a number between 1 and \( n \) corresponding to a choice of row in \( X \) (i.e., there are \( n \) possibilities for each odd vertex). Each even vertex \( (2, 4, 6, \ldots) \) is associated with a number between 1 and \( N \), corresponding to a choice of column in \( D \) or \( X \) (i.e., there are \( N \) possibilities for each even vertex). Since \( \hat{\pi} \) maps odd to even numbers and vice versa, blocks of \( \rho \) consist either of even numbers only, or of odd numbers only. The following definition therefore makes sense:

Definition 5: We denote the number of blocks of \( \rho \) consisting of even numbers and odd numbers by \( k(\rho) \) and \( l(\rho) \), respectively. We define \( k_1, \ldots, k_{k(\rho)} \) and \( l_1, \ldots, l_{l(\rho)} \) to be the cardinalities of these blocks.

Throughout the paper, we will as in Definition 5 let \( k \) refer to even numbers, \( l \) refer to odd numbers. In particular we have \( \rho \leq \{1, 3, 5, \ldots\}, \{2, 4, 6, \ldots\} \), where \( \leq \) is the refinement order (i.e. \( \rho_1 \leq \rho_2 \) whenever any block of \( \rho_1 \) is contained within a block of \( \rho_2 \)). The restriction of \( \rho \) to the odd numbers thus defines another partition, which we will denote \( \rho(\text{odd}) \). Definitions 3 and 4 will be needed in order to express (4).

When we compute the moments \( \text{tr}((D + X)(D + X)^H)^p) \), with \( D \) deterministic and \( X \) Gaussian, we multiply out to obtain a sum of many terms of length \( 2p \) on the form \( x_1x_2 \cdots x_{2p} \), where \( x_i = X, X^H, D, \) or \( D^H \), with \( - \) and \( -^H \) terms appearing in alternating order. We need the following definitions, which naturally extend Definition 3 to the case where \( \pi \) constitutes a partial permutation of elements:

Definition 6: Let \( p \) be a positive integer. By a partial permutation we mean a one-to-one mapping \( \pi \) between two subsets \( \rho_1, \rho_2 \) of \( \{1, \ldots, p\} \). We denote by \( SP_p \) the set of partial permutations of \( p \) elements. When \( \pi \in SP_p \), \( \hat{\pi} \in S_{2p} \) is defined by

\[
\hat{\pi}(2j - 1) = 2\pi^{-1}(j), \quad j \in \rho_2 \\
\hat{\pi}(2j) = 2\pi(j) - 1, \quad j \in \rho_1.
\]

It is clear that any partial permutation is uniquely determined by a triple \( \{\rho_1, \rho_2, q\} \), where \( 0 \leq |\rho_1| = |\rho_2| \leq p \), \( q \in S_{\rho_2} \), and where \( S_{\rho_2} \) are the permutations of the elements of \( \rho_2 \). It is also clear that \( \hat{\pi} \) is a partial permutation from \( (2\rho_1) \cup (2\rho_2 - 1) \) to itself (also with period two), where \( 2\rho_2 - 1 = \{2k - 1 | k \in \rho_2\} \).

Definition 7: Let \( \pi \) be a partial permutation, and let \( \hat{\pi} \) be determined by \( \rho_1 = \rho_2 \) and \( q \). We associate to \( \pi \) an equivalence relation \( \rho = \rho(\pi) \) on \( \mathbb{X} = \rho_1 \cup (\rho_1 + 1) \) generated by

\[
j \sim_\rho \hat{\pi}(j) + 1, \quad j \in \rho_1.
\]

(8)

As before, we let \( k(\rho) \) and \( l(\rho) \) denote the number of blocks of \( \rho \) consisting of only even or odd numbers, respectively.

For a partial permutation \( \pi \in SP_p \) determined by \( \rho_1, \rho_2, q \), in the term \( x_1x_2 \cdots x_{2p} \), \( p_1 \) corresponds to those \( k \) such that \( x_{2k} = X^H \), \( \rho_2 \) to those \( k \) such that \( x_{2k-1} = X \) (i.e., they correspond to the Gaussian terms in \( x_1x_2 \cdots x_{2p} \)).

These concepts are illustrated in Figure 2 for the term \( XX^HDD^HXX^HD^H \).

![Figure 1](image1.png)

![Figure 2](image2.png)

Fig. 1. Ordering of vertices and edges in a circle corresponding to the term \((XX^H)^4\). The odd edges \( 1, 3, 5, 7 \) correspond to \( X \); the even edges \( 2, 4, 6, 8 \) correspond to \( X^H \); the odd vertices \( 1, 3, 5, 7 \) correspond to a choice among \( 1, \ldots, n \); the even vertices \( 2, 4, 6, 8 \) correspond to a choice among \( 1, \ldots, N \). The bars to differ between edges and vertices are only used in the figures.

Fig. 2. A partial permutation of edges in the term \( x_1x_2 \cdots x_8 = XX^HDD^HXX^HD^H \) from \((D + X)(D + X)^H\) \cite{IEEEhowto}. Bold, odd edges are \( X \)-terms, bold even edges are \( X^H \)-terms, other edges are \( D \)- or \( D^H \)-terms. \( \rho_1 = \{1, 3\} \) represent the (even) \( X^H \)-edges \( (2, 6) \), \( \rho_2 = \{3, 1\} \) represent the (odd) \( X \)-edges \( (5, 1) \), \( \rho \) represents the pairing \( q \) of \( \rho_1 \) and \( \rho_2 \). Dashed edges connect vertices which belong to the same block of \( \rho \).

We only consider terms where \( 0 \leq |\rho_1| = |\rho_2| \leq p \) due to the fact that, as will be shown in Appendix A, only conjugate pairings of Gaussian elements contribute. The permutation \( q \in S_{\rho_2} \) dictates their pairings. \( \mathbb{X} \) corresponds to matrix indices.
which occur for any Gaussian matrix $X$ or $X^H$. Definition 3 corresponds to the case where $X = \{1, \ldots, 2n\}$.

**Definition 8:** Let $\pi$ be a partial permutation, and let $\tilde{\pi}$ be determined by $\rho_1 = \rho_2$ and $q$. Let $\sigma = \sigma(\pi)$ be the equivalence relation on $D = \rho_1^\dagger$ generated by the relations

\begin{align*}
k &\sim_{\sigma} k + 1, \quad \text{if } k, k + 1 \in D \\
k &\sim_{\sigma} l, \quad \text{if } k, l \in D, k + 1 \sim_{\rho} l.
\end{align*}

Let also $kd(\rho)$ be the number of blocks of $\rho$ contained within the even numbers which intersect $D \cup (D + 1)$, and let $ld(\rho)$ be the number of blocks of $\rho$ contained within the odd numbers which intersect $D \cup (D + 1)$.

$D$ should be interpreted as the edges corresponding to the deterministic elements $D, D^H$ in $x_1 x_2 \cdots x_{2p}$. Two such edges belong to the same block of $\sigma$ if, after identifying vertices with $\rho$, they are connected with a path of edges from $D$. A block of $\rho$ intersecting $D \cup (D + 1)$ corresponds to a Gaussian element which occurs next to a deterministic element. These concepts are illustrated in Figure 3 for the same term used for Figure 2.

Fig. 3. The graph resulting from the identification of edges illustrated in Figure 2. The blocks of $\sigma$ consist of the connected components of the $D$- and $D^H$-edges, i.e., $\sigma = \{\{3, 4\}, \{7, 8\}\}$. Counting the vertices in this case gives that $k(\rho) = 7, l(\rho) = 2, kd(\rho) = 6, ld(\rho) = 2$.

We will also state results for the case when the Gaussian matrices are assumed selfadjoint instead of complex. For this we need to slightly change the definitions of the partitions $\rho$ and $\sigma$:

**Definition 9:** Let $\pi \in \text{SP}_p$ be determined by disjoint subsets $\rho_1, \rho_2$ of $\{1, \ldots, p\}$ with $|\rho_1| = |\rho_2|$ (in particular, $2|\rho_1| \leq p$). We define by $\rho_{sa} = \rho_{sa}(\pi)$ the equivalence relation on

$$X = \rho_1 \cup (\rho_1 + 1) \cup \rho_2 \cup (\rho_2 + 1)$$
generated by

$$i \sim_{\rho_{sa}} \pi(i) + 1, \quad \text{for } i \in \rho_1,$$

$$\pi^{-1}(i) + 1 \sim_{\rho_{sa}} i, \quad \text{for } i \in \rho_2.$$

As before, $\rho_1$ corresponds to the choices from $X^H$, $\rho_2$ corresponds to the choices from $X$, when the selfadjoint Gaussian matrix is expressed with the help of a complex Gaussian matrix $X$. For sums, we also need the following modified version of Definition 8:

**Definition 10:** With $\pi, \rho_1, \rho_2$ as in Definition 9, let $\sigma_{sa} = \sigma_{sa}(\pi)$ be the equivalence relation on $D = (\rho_1 \cup \rho_2)^\cap$ generated by the relations

\begin{align*}
k &\sim_{\sigma_{sa}} k + 1, \quad \text{if } k, k + 1 \in D \\
k &\sim_{\sigma_{sa}} l, \quad \text{if } k, l \in D, k + 1 \sim_{\rho_{sa}} l \text{ or } k \sim_{\rho_{sa}} l + 1.
\end{align*}

Define also $d(\rho_{sa})$ as the number of blocks of $\sigma_{sa}$ which intersect $D \cup (D + 1)$.

Note that in all definitions in this section, we can generalize to addition and subtraction modulo $p_1, \ldots, p_k$ in the obvious way. This means that the definitions also apply to the more general moments (4), which we will have use for when we state the most general results in the next section. The concrete values of $p_1, \ldots, p_k$ will be provided in the given setting only when needed. Figure 4 illustrates how the previous figures could look when more than one circle is involved.

Fig. 4. Identification of edges across two circles. Such identifications arise in the computations of (4).

IV. STATEMENT OF MAIN RESULTS

The main results of the paper are split in three sections. First, basic sums and products are considered, basic meaning that one of the involved matrices is deterministic. After this, we expand to the case where both matrices are assumed random, in which case we bring all expectations of any product of traces into the picture. In these two sections, all Gaussian matrices are assumed complex and rectangular, for which the results relate to the moments of the singular law of the matrices. In the last section we state similar results for the case where the Gaussian matrices instead are assumed square and selfadjoint.

A. Basic sums and products

Our first and simplest result concerns the moments of a product of a deterministic matrix and a Wishart matrix:

**Theorem 1:** Let $n, N$ be positive integers, $X$ be $n \times N$ standard, complex, Gaussian and $D$ a deterministic $n \times n$ matrix. For any positive integer $p$, we have

\begin{equation}
\mathbb{E} \left[ \text{tr} \left( \left( \frac{1}{N} DX^H \right)^p \right) \right] = \sum_{\pi \in S_p} N^{k(\rho)-pH(l(\rho)-1)D_p[\text{odd}]}. \tag{13}
\end{equation}

Theorem 1 is proved in Appendix A. The results following Theorem 1 will build on the same geometric interpretation of circles, edges, and vertices as explained in Section III, and also on the same ingredient proved in Appendix A, namely that only conjugate pairings of Gaussian elements contribute. The following results will therefore be given shorter proofs, explaining only the basic difference in how terms are grouped. We remark also that other closed forms of (13) have been found in the literature when $D$ is assumed positive definite [16].

We now turn to the case of the sum of a deterministic matrix $D$ and a Gaussian matrix $X$. When the matrices are rectangular, we turn to the moments of $(D + X)(D + X)^H$. In the
large $n$, $N$-limit, this corresponds to the concept of rectangular free convolution [19], which admits a nice implementation in terms of moments [20]. When $n$ and $N$ are finite, the following result is needed.

**Theorem 2:** Let $X$ be an $n \times N$ standard, complex, Gaussian matrix, $D$ a deterministic $n \times N$ matrix, and set $D_p = \text{tr} \left( \left( \frac{1}{n} D D^H \right)^p \right)$. We have that

$$
E \left[ \text{tr} \left( \left( \frac{1}{N} (D + X)(D + X)^H \right)^p \right) \right] = \sum_{\pi \in S(p)} \frac{1}{n^n |\rho_1|} \sum_{\sigma \in S(p)} \left( c(k(\rho_1)) \right)^{d(\rho_1)} \left( c(k(\rho_2)) \right)^{d(\rho_2)} \cdots \left( c(k(\rho_s)) \right)^{d(\rho_s)} \prod_{i \in \sigma} D_{|\sigma(i)|/2}.
$$

(14)

The proof can be found in Appendix B. Note that in Theorem 2, $n$- and $N$-terms have not been grouped together. This has been done to make clear in the proof the origin of the different terms.

**B. Expectations of products of traces**

Theorem 1 and 2 can be recursively applied, once one replaces $D(1)$ with random matrices. We thus instead compute an expectation of products of traces in each step. The only other difference in the formulas is that additional powers of $\frac{1}{n}$ may be included (each one being a normalizing factor in taking the trace). The recursive version of Theorem 1 looks as follows.

**Theorem 3:** Assume that the $n \times N$ random matrix $R$ is independent from the $n \times N$ standard, complex, Gaussian matrix $X$, and define

$$
R_{p_1, \ldots, p_k} = E \left[ \text{tr} \left( R^{p_1} \right) \text{tr} \left( R^{p_2} \right) \cdots \text{tr} \left( R^{p_k} \right) \right]
$$

$$
M_{p_1, \ldots, p_k} = E \left[ \text{tr} \left( \left( \frac{1}{N} X^H X \right)^{p_1} \right) \right. \times \text{tr} \left( \left( \frac{1}{N} X^H X \right)^{p_2} \right) \cdots \times \left. \text{tr} \left( \left( \frac{1}{N} X^H X \right)^{p_k} \right) \right].
$$

Set $p = p_1 + \cdots + p_k$, and let as before $l_1, \ldots, l_r$ be the cardinalities of the blocks of odd numbers only of $\rho$, with $k(\rho), l(\rho)$ the number of blocks consisting of even and odd numbers only, respectively. We have that

$$
M_{p_1, \ldots, p_k} = \sum_{\pi \in S(p)} \sum_{\sigma \in S(p)} \left( c(k(\rho_1)) \right)^{d(\rho_1)} \cdots \left( c(k(\rho_r)) \right)^{d(\rho_r)} \prod_{i \in \sigma} R_{l_1, \ldots, l_r}.
$$

(15)

**Proof:** There are only two differences from Theorem 1. First $n^{-1}$ is replaced with $n^{-k}$, since we now are taking $k$ traces instead of 1, and we thus modify with additional trace normalization factors. Second, we replace the trace of a deterministic matrix with the expectation of a random matrix. It is clear that the only additional thing needed for the proof to be replicated is that the random matrices $X$ and $R$ are independent.

It is also clear that the theorem can be recursively applied to compute the moments of any product of independent Wishart matrices

$$
D_1^{-1} X_1 X_1^H \frac{1}{N_2} X_2 X_2^H \cdots \frac{1}{N_k} X_k X_k^H,
$$

where $D$ is deterministic and $X_i$ is an $n \times N_i$ standard complex Gaussian matrix. The $R$’s during these recursions will simply be

$$
R_1 = D_1^{-1} X_1 X_1^H \frac{1}{N_2} X_2 X_2^H \cdots \frac{1}{N_{k-1}} X_{k-1} X_{k-1}^H
$$

$$
R_2 = D_1^{-1} X_1 X_1^H \frac{1}{N_2} X_2 X_2^H \cdots \frac{1}{N_{k-2}} X_{k-2} X_{k-2}^H
$$

$$
\vdots
$$

$$
R_k = D.
$$

The recursive version of Theorem 2 looks as follows.

**Theorem 4:** Let $X$ be an $n \times N$ standard, complex, Gaussian matrix and $R$ be $n \times N$ and independent from $X$. Set

$$
R_{p_1, \ldots, p_k} = E \left[ \text{tr} \left( \left( \frac{1}{N} R R^H \right)^{p_1} \right) \times \text{tr} \left( \left( \frac{1}{N} R R^H \right)^{p_2} \right) \cdots \times \text{tr} \left( \left( \frac{1}{N} R R^H \right)^{p_k} \right) \right]
$$

$$
M_{p_1, \ldots, p_k} = E \left[ \text{tr} \left( \left( \frac{1}{N} (R + X)(R + X)^H \right)^{p_1} \right) \times \text{tr} \left( \left( \frac{1}{N} (R + X)(R + X)^H \right)^{p_2} \right) \cdots \times \text{tr} \left( \left( \frac{1}{N} (R + X)(R + X)^H \right)^{p_k} \right) \right].
$$

We have that

$$
M_{p_1, \ldots, p_k} = \sum_{\pi \in S(p)} \sum_{\sigma \in S(p)} \left( c(k(\rho_1)) \right)^{d(\rho_1)} \cdots \left( c(k(\rho_r)) \right)^{d(\rho_r)} \prod_{i \in \sigma} R_{l_1, \ldots, l_r},
$$

(16)

where $l_1, \ldots, l_r$ are the cardinalities of the blocks of $\sigma$, divided by 2.

The proof is omitted, since it follows in the same way Theorem 1 was generalized to Theorem 3 above.

**C. Selfadjoint Gaussian matrices**

A standard, selfadjoint, Gaussian $n \times n$ random matrix $X$ can be written on the form $X = \frac{1}{\sqrt{2}} (Y + Y^H)$, where $Y$ is an $n \times n$ standard complex Gaussian matrix. We can thus compute the moments of $DX$ and $D + X$ (with $D$ deterministic and $X$ selfadjoint Gaussian) by substituting $X = \frac{1}{\sqrt{2}} (Y + Y^H)$ (with $Y$ complex Gaussian) in these expressions, and summing over all possible combinations of $Y$ and $Y^H$. As before, we only need to consider conjugate pairings of the $Y$ and $Y^H$. 


The analogues of Theorem 1 and Theorem 2 when the Gaussian matrices instead are selfadjoint look as follows. Since Theorem 1 and Theorem 2 had straightforward generalizations to the case where $D$ is random, we state the results only when $D$ is assumed random.

**Theorem 5:** Assume that the $n \times n$ random matrix $R$ is independent from the $n \times n$ standard selfadjoint Gaussian matrix $X$, and define

$$ R_{p_1, \ldots, p_k} = E \left[ \text{tr} \left( R^{p_1} \right) \text{tr} \left( R^{p_2} \right) \cdots \text{tr} \left( R^{p_k} \right) \right] $$

$$ M_{p_1, \ldots, p_k} = E \left[ \left( \frac{1}{\sqrt{n}} X \right)^{p_1} \right] \times \left( \frac{1}{\sqrt{n}} X \right)^{p_2} \cdots \times \left( \frac{1}{\sqrt{n}} X \right)^{p_k} . $$

Set $p = p_1 + \cdots + p_k$, and let $l_1, \ldots, l_r$ be the cardinalities of the blocks of $\sigma_{sa}$. We have that

$$ M_{p_1, \ldots, p_k} = \sum_{\pi = \pi_1 \rho_1 p_1 \pi_2 \rho_2 \cdots \pi_r \rho_r \in \mathcal{SP}_{\rho_1 \rho_2 \cdots \rho_r} \text{ disjoint}} 2^{-|\rho_1|} n^{-p/2} R_{l_1, \ldots, l_r} . \quad (17) $$

**Proof:** The proof follows in the same way as the proofs in Appendix A and Appendix B. We therefore only give the following quick description on how the terms in (17) can be identified:

- $2^{-p/2}$ comes from the $p$ normalizing factors $\frac{1}{\sqrt{2}}$ in $\frac{1}{\sqrt{2}}(Y + Y^H)$,
- $n^{-p/2}$ comes from replacing the non-normalized traces with the normalized traces to obtain $R_{l_1, \ldots, l_r}$,
- $n^{-p/2}$ comes from the $p$ normalizing factors $\frac{1}{\sqrt{n}}$ in $\frac{1}{\sqrt{n}} X$,
- $n^{-k}$ comes from the $k$ traces taken in $M_{p_1, \ldots, p_k}$.

The concept of additive free convolution [4] allows us to compute moments of the sum of a selfadjoint, Gaussian matrix, and let $R$ be $n \times n$ and independent from $X$. Set $p = p_1 + \cdots + p_k$, and let $l_1, \ldots, l_r$ be the cardinalities of the blocks of $\sigma_{sa}$ from Definition 10. We have that

$$ M_{p_1, \ldots, p_k} = \sum_{\pi = \pi_1 \rho_1 p_1 \pi_2 \rho_2 \cdots \pi_r \rho_r \in \mathcal{SP}_{\rho_1 \rho_2 \cdots \rho_r} \text{ disjoint}} 2^{-|\rho_1|} n^{-|p|+|\rho(\pi)_{sa}|} $$

$$ \times \ n^{-d(\rho(\pi)_{sa})-k+|\sigma_{sa}|} \times R_{l_1, \ldots, l_r} . \quad (18) $$

**Proof:** The items in (18) are identified as follows:

- $2^{-|\rho_1|}$ comes from the normalizing factors $\frac{1}{\sqrt{2}}$ in the $2|\rho_1|$ choices of $\frac{1}{\sqrt{2}}(Y + Y^H)$,
- $n^{-|\rho_1|}$ comes from the normalizing factors $\frac{1}{\sqrt{n}}$ in the $2|\rho_1|$ choices of $\frac{1}{\sqrt{n}} X$,
- $n^{-|\rho(\pi)_{sa}|}$ comes from counting the vertices which do not come from applications of (12),
- $n^{-k}$ comes from the $k$ traces taken in $M_{p_1, \ldots, p_k}$,
- $n^{-|\sigma_{sa}|}$ comes from replacing the non-normalized traces with the normalized traces to obtain $R_{l_1, \ldots, l_r}$.

Recursive application of theorems 3, 4, 5, and 6, allows us to compute moments of most combinations of independent (selfadjoint or complex) Gaussian random matrices and deterministic matrices. It also enables us to compute the second order moments (i.e., covariances of traces) for many types of matrices. Asymptotic properties of such second order moments have previously been studied [22], [23], [24]. While previous papers allow us to compute such moments and second order moments asymptotically, in many cases the exact result is needed.

### V. SOFTWARE IMPLEMENTATION

Although the formulas presented up to now are rather complex, and depend heavily on correct handling of the normalizing factors in front of the matrices, it is also clear that they are all implementable: all that is required is traversing subsets $(\rho_1, \rho_2)$, permutations $(\pi, q)$, and implement the partitions $\rho, \sigma(\pi), \rho(\pi)_{sa}, \sigma(\pi)_{sa}$ from $\pi$. Code in Matlab for doing so has been implemented for this paper. In [25], an implementation for all these partitions can be found, together with all necessary code to implement the formulas in this paper. In [26], an overview of all public functions in the library this implementation is part of can be found, as well as how the methods here can be combined with other types of matrices. The software can also generate formulas directly in \texttt{M2X}, in addition to performing the convolution or deconvolution numerically in terms of a set of input moments. All formulas in this section have in fact been generated by this implementation. Due to the complexity of the expressions, it is not recommended to compute these by hand. The reason there are no powers of $n$ in the first formulas below is that the implementation has replaced $\frac{1}{n}$ with $c$ everywhere, which makes the results more compatible with the asymptotic case $\lim_{N \to \infty} \frac{n}{N} = c$. 

$$ R_{p_1, \ldots, p_k} = E \left[ \text{tr} \left( \frac{1}{\sqrt{n}} R^{p_1} \right) \right] $$

$$ M_{p_1, \ldots, p_k} = E \left[ \left( \frac{1}{\sqrt{n}} (D + X) \right)^{p_1} \right] \times \left( \frac{1}{\sqrt{n}} (D + X) \right)^{p_2} \cdots \times \left( \frac{1}{\sqrt{n}} (D + X) \right)^{p_k} . $$
A. Automatically generated formulas for Theorem 1

We obtain the following expression for the first four moments in (13) with our implementation, where we have denoted

\[ D_p = \text{tr} (D^p) \]
\[ M_p = \mathbb{E} \left[ \text{tr} \left( \left( \frac{1}{N} \mathbf{X} \mathbf{X}^H \right)^p \right) \right], \]

in accordance with Theorem 1:

\[
M_1 = D_1 \\
M_2 = D_2 + cD_1^2 \\
M_3 = \left( 1 + \frac{1}{N^2} \right) D_3 + 3cD_2D_1 + c^2 D_1^3 \\
M_4 = \left( 1 + \frac{5}{N^2} \right) D_4 + \left( 2c + \frac{c^2}{N^2} \right) D_2^2 \\
\quad + \left( 4c + \frac{4c}{N^2} \right) D_3D_1 + 6c^2 D_2D_1^2 + c^3 D_1^4
\]

By a simple recursion, we can express \( D_p \) from \( M_p \). For the first three moments these recursions become

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

More generally, in order to compute the moments of products of Wishart matrices, it is needed to implement Theorem 3 as a function which takes \( p_1, \ldots, p_k \) as input, iterates through the set of all permutations \( \pi \), computes the numbers \( k(\rho), l(\rho), l_1, \ldots, l_r \), and performs this recursively. \(^3\)

The implementation is also able to generate the moments of the product of a deterministic matrix and any number of independent Wishart matrices [26].

B. Automatically generated formulas for Theorem 2

Defining

\[ D_p = \text{tr} \left( \left( \frac{1}{N} \mathbf{D} \mathbf{D}^H \right)^p \right) \]
\[ M_p = \mathbb{E} \left[ \text{tr} \left( \left( \frac{1}{N} (\mathbf{D} + \mathbf{X})(\mathbf{D} + \mathbf{X})^H \right)^p \right) \right], \]

in accordance with Theorem 2, the implementation generates the following formulas:

\[
M_1 = D_1 + 1 \\
M_2 = D_2 + (2 + 2c) D_1 + (1 + c) \\
M_3 = D_3 + (3 + 3c) D_2 \\
\quad + 3cD_1^2 + \left( 3 + 9c + 3c^2 + \frac{3}{N^2} \right) D_1 \\
\quad + \left( 1 + 3c + c^2 + \frac{1}{N^2} \right) \\
M_4 = D_4 + (4 + 4c) D_3 + 8cD_2D_1 \\
\quad + \left( 6 + 16c + 6c^2 + \frac{16}{N^2} \right) D_2 \\
\quad + \left( 14c + 14c^2 \right) D_1^2 \\
\quad + \left( 4 + 24c + 24c^2 + 4c^3 + \frac{20 + 20c}{N^2} \right) D_1 \\
\quad + \left( 1 + 6c + 6c^2 + c^3 + \frac{5 + 5c}{N^2} \right)
\]

C. Automatically generated formulas for Theorem 5

Defining

\[ D_p = \text{tr} (\mathbf{D}^p) \]
\[ M_p = \mathbb{E} \left[ \text{tr} \left( \left( \frac{1}{\sqrt{n}} \mathbf{D} \mathbf{X} \mathbf{X}^H \right)^p \right) \right], \]

in accordance with Theorem 5, the implementation generates the following formulas:

\[
M_2 = D_2^3 \\
M_4 = n^{-2} D_4 + 2D_2 D_1^2 \\
M_6 = n^{-2} D_6 + 3n^{-2} D_4 D_2 \\
\quad + 6n^{-2} D_5 D_1 + 3D_2^2 D_1^2 + 2D_3 D_1^3 \\
M_8 = 21n^{-4} D_8 + 4n^{-2} D_6 D_2 \\
\quad + 6n^{-2} D_4 D_2^2 + 16n^{-2} D_4 D_1 D_3 \\
\quad + 24n^{-2} D_5 D_2 D_1 + 20n^{-2} D_6 D_1^2 \\
\quad + 4D_2^3 D_1^2 + 8D_3 D_2 D_1^2 + 2D_4 D_1^4.
\]

The implementation is also able to generate the moments of the product of a deterministic matrix and any number of independent, selfadjoint Gaussian matrices [26]. In fact, the implementation also accomplishes this without any assumption on whether the Gaussian component matrices are selfadjoint or complex, and without any assumptions on the order how these appear.

D. Automatically generated formulas for Theorem 6

Defining

\[ D_p = \text{tr} \left( \left( \frac{1}{\sqrt{n}} \mathbf{D} \right)^p \right) \]
\[ M_p = \mathbb{E} \left[ \text{tr} \left( \left( \frac{1}{\sqrt{n}} (\mathbf{D} + \mathbf{X}) \right)^p \right) \right], \]

\(^3\)In [25], a method is explained, also with an example, which can be called in a recursive fashion for this.
in accordance with Theorem 6, the implementation generates the following formulas:

\[
\begin{align*}
    M_1 &= D_1 \\
    M_2 &= D_2 + 1 \\
    M_3 &= D_3 + 3D_1 \\
    M_4 &= D_4 + 4D_2 + 2D_1^2 + (2 + n^{-2}).
\end{align*}
\]

The implementation will become increasingly slower when more matrices are multiplied. In the literature, up to three Gaussian matrices have been considered [27].

VI. APPLICATIONS

In this section, we consider some wireless communications examples where the presented inference framework is used. Note that for some of these simulations, asymptotic results could also have been used when there are many observations present, since the observations can in the presented cases be stacked into one large, compound observation matrix. A case where such a stacking can not be performed, and hence the asymptotic result does not suffice, is presented in Section VI-C. When the asymptotic result can be used, inference on the moments becomes simpler, since one does not need to take into account expectations of products of traces, due to the almost sure convergence of the trace of the matrices [8]. In this case, it turns out that Theorem 1, Theorem 2, and Theorem 6 can all be implemented by direct application of the moment-cumulant formula [21], for which an efficient implementation exists [13], without the need for iterating through all partitions. Theorem 5 can be implemented in terms of the \( S \)-transform [5], which has an implementation in terms of power series [28], also without the need for iterating through all partitions.

A. MIMO rate estimation

In many MIMO (Multiple Input Multiple Output) antenna based sounding and MIMO channel modelling applications, one is interested in obtaining an estimator of the rate in a noisy and mobile environment. In this setting, one has \( M \) noisy observations of the channel \( Y_i = D + \sigma N_i \), where \( D \) is an \( n \times N \) deterministic channel matrix, \( N_i \) is an \( n \times N \) standard, complex, Gaussian matrix representing the noise, and \( \sigma \) is the noise variance. Typically, a limitation on the rank, \( \text{rank}(D) \leq k \), is known. We will see that only the first \( k \) moments of \( D \) are needed in order to estimate the rate in this case, so that the presented moment formulas can be applied. The channel \( D \) is supposed to stay constant during \( M \) symbols. The rate estimator is given by

\[
C = \frac{1}{n} \log_2 \det \left( I_n + \frac{\rho}{N} DD^H \right) = \frac{1}{n} \log_2 \prod_{i=1}^{n} (1 + \rho \lambda_i),
\]

where \( \rho = \frac{1}{\text{SNR}} \) is the SNR, and \( \lambda_i \) are the eigenvalues of \( \frac{1}{N} DD^H \). This problem falls exactly within a finite dimensional statistical inference problem. Note that \( \prod_{i=1}^{n} (1 + \rho \lambda_i) \)

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig5}
\caption{Estimation of the channel capacity using the moment method for the 2 \( \times \) 2-matrix (20) for various number of observations. \( \rho = 5 \).}
\end{figure}

can be written in terms of the elementary symmetric polynomials given by

\[
\begin{align*}
    \Pi_1(\lambda_1, \ldots, \lambda_n) &= \lambda_1 + \cdots + \lambda_n \\
    \Pi_2(\lambda_1, \ldots, \lambda_n) &= \sum \lambda_i \lambda_j \\
    \vdots \\
    \Pi_n(\lambda_1, \ldots, \lambda_n) &= \lambda_1 \cdots \lambda_n,
\end{align*}
\]

and that only the first \( k \) of these can be nonzero, since there can only be \( k \) nonzero eigenvalues due to the rank restriction. The first \( k \) elementary symmetric polynomials can be written in terms of the first \( k \) moments of \( \frac{1}{N} DD^H \) using the Newton-Girard formulas [29]. These moments can again be estimated from the moments of \( \frac{1}{M}(D+\sigma N_i)(D+\sigma N_i)^H \) using our inference framework. We estimate the expected moments of \( \frac{1}{M}(D+\sigma N_i)(D+\sigma N_i)^H \) by taking an average of the moments of \( M \) observations of our model.

We have tested the procedure for two cases. In the first case, a 2 \( \times \) 2-matrix

\[
D = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix}
\] (20)

was used with \( \rho = 5 \) and different number of observations. The corresponding simulation is shown in Figure 5 and shows the convergence to the true rate. The fact that the channel matrix is diagonal is irrelevant for the rate estimation.

In the second case a 4 \( \times \) 4-matrix

\[
D = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}
\] (21)

was used, with \( \rho = 10 \) and different number of observations. The corresponding simulation is shown in Figure 6. As the number of variables to be estimated is higher than in the 2 \( \times \) 2-matrix case (4 eigenvalues instead of 2), one needs more symbols to obtain the same accuracy estimation.
Fig. 6. Estimation of the channel capacity using the moment method for the 4 × 4-matrix (21) for various number of observations. ρ = 10.

Fig. 7. Cognitive MIMO Networks

B. Understanding the network in a finite time

In cognitive MIMO Networks, one must learn and control the “black box” (wireless channel for example) with multiple inputs and multiple outputs (Figure 7) within a fraction of time and with finite energy. The fraction of time constraint is due to the fact that the channel (black box) changes over time. Of particular interest is the estimation of the rate within the window of observation.

Let y be the output vector, x and n respectively the input signal and the noise vector, so that

\[ y = x + σn. \]  

(22)

In this case and in the Gaussian case, the rate is given by

\[ C = H(y) - H(y|x) = \log_2 \frac{\det(πeR_Y)}{\det(πeR_N)} \]

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

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

\[ R_Y = \begin{pmatrix} 1 & 0 \\ 0 & 0.5^2 \end{pmatrix}. \]  

(23)

Considering \( L \) observations of (22), we use the same procedure followed in the Section VI-A in order to apply our inference framework and estimate the eigenvalues of \( R_Y \). The simulation is shown in Figure 8, where we observe the convergence of the estimated rate to the true one.

In Figure 9, we estimate the eigenvalues of the matrix \( R_Y \) versus the number of observations \( L \). Once again, we observe the convergence of the estimated eigenvalues to the true eigenvalues.
C. Power estimation

In many multi-user MIMO applications, one needs to determine the power with which the users send information. We consider the system given by

\[ y_i = WP^{\frac{1}{2}}s_i + \sigma n_i \]  

where \( W, P, s_i, \) and \( n_i \) are respectively the \( N \times K \) channel gain matrix, the \( K \times K \) diagonal power matrix due to the different distances from which the users emit, the \( K \times 1 \) matrix of signals and the \( N \times 1 \) matrix representing the noise with variance \( \sigma \). In particular, \( W, s_i, n_i \) are independent standard, complex, Gaussian matrices and vectors. We suppose that we have \( M \) observations (during which the channel gain matrix stays constant) of the vector \( y_i \). Considering the \( 2 \times 2 \)-matrix

\[ P^{\frac{1}{2}} = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix} \]  

and applying Theorem 4 first, and then Theorem 3 twice (each application takes care of one Gaussian matrix), we can estimate the eigenvalues of \( P \) when we have an increasing number \( L \) of observations of the matrix \( Y = [y_1, \ldots, y_M] \), representing the signals received (we average across several block fading channels). Hence, we estimate the moments of the matrix \( P \) based only on the moments of the matrix \( YY^H \). Knowing the moments of \( P \), we can estimate the eigenvalues as in Section VI-A. When \( L \) increases, we get a prediction of the eigenvalues which is closer to the true eigenvalues of \( P \). Figure 10 illustrates the estimation of eigenvalues up to \( L = 1200 \) observations.

It is possible to compute the variance of the moment estimators for the model (24). We do not write down expressions for these, but remark that these are analyzed in [27]. These expressions turn out to involve combinations of \( K, M, \) and \( N \) in the denominators, so that in order for the variance to be low, large values for \( K, M, N \) are required. In Figures 11 - 12, we note that the variance decreases much faster when we increase \( K, M, N \) jointly, than when we increase the number of observations.

VII. Conclusion

In this paper, we have introduced a framework which enables us to compute the moments of many types of combinations of independent Gaussian and Wishart random matrices, without any assumptions on the matrix dimensions. We also explained an accompanying software implementation, and also some useful applications where the framework has been used for simulations. Future work will focus on applying and extending the framework to other types of matrix models. While the formulas presented here have been generated by traversing sets of partitions, there may exist expressions for the same formulas which are more efficient to compute (in at least one case [16], this is known to be the case). Future work
will also attempt to find such simpler expressions.

APPENDIX A

The proof of Theorem 1

In order to prove Theorem 1, we will expand the moments

$$\mathbb{E} \left[ \text{tr} (DXX^H)^p \right]$$

following in the footsteps of [14], and in the process generalize results therein, since no deterministic part was involved in that paper. We will thus in the following rewrite some of the important parts in the proofs in [14], since these are needed in our generalizations. First, we will need the following proposition.

Proposition 1: Let $X$ be $n \times N$ standard, complex, Gaussian, and $D$ a deterministic $n \times n$ matrix. Let $p$ be a positive integer, then

$$\mathbb{E} \left[ \text{tr} (DXX^H)^p \right] = \sum_{\pi \in S_p} \mathbb{E} \left[ \text{tr} (D_1X_1^H \cdots D_pX_p^H)^p \right] \quad (27)$$

where $X_1, \ldots, X_p$ are independent $n \times N$ standard, complex, Gaussian matrices.

Proof: Let $(X_j)_{j \in \mathbb{N}}$ be a sequence of independent $n \times N$ standard, complex, Gaussian matrices with entries $x(u,v,i)$, $1 \leq u \leq n$, $1 \leq v \leq N$. For any $s \in \mathbb{N}$, the matrix $s^{-1/2}(X_1 + \cdots + X_s)$ is again $n \times N$ standard, complex, Gaussian. Hence, we can write

$$\mathbb{E} \left[ \text{tr} (DXX^H)^p \right] = \mathbb{E} \left\{ D \left( s^{-1/2}(X_1 + \cdots + X_s) \right)^H \right\} \quad (29)$$

$$= s^{-p} \sum_{1 \leq i_1, j_1, \ldots, i_p, j_p \leq s} \mathbb{E} \left[ \text{tr} (D_{i_1}X_{j_1}^H \cdots D_{i_p}X_{j_p}^H) \right] \quad (30)$$

For $\pi \in S_p$, we define

$$M(\pi, s) = \left\{ (i_1, j_1, \ldots, i_p, j_p) \in \{1, 2, \ldots, s\}^{2p} \mid j_1 = i_{\pi(1)}, \ldots, j_p = i_{\pi(p)} \right\} \quad (31)$$

Just as in [14],

$$\mathbb{E} \left[ \text{tr} (D_{i_1}X_{j_1}^H \cdots D_{i_p}X_{j_p}^H) \right] = n^{-1} \times \sum_{1 \leq u_1, v_1, \ldots, u_p, v_p \leq n} d(u_1, v_1) \cdots d(u_p, v_p) \mathbb{E}[x(v_1, u_1, i_1) \times x(u_2, w_1, j_1) \cdots x(v_p, w_p, i_p) x(u_1, w_p, j_p)]$$

$$\times x(u_{h+1}, w_h, j_h) = x(v_{\pi(h)}, w_{\pi(h)}, i_{\pi(h)})$$

for all $h \in \{1, 2, \ldots, p\}$. Hence, we only have to sum over those 2-tuples $(i_1, j_1, \ldots, i_p, j_p)$ that are in $M(\pi, s)$ for some $\pi \in S_p$, i.e.

$$\mathbb{E} \left[ \text{tr} (DXX^H)^p \right] = s^{-p} \times$$

$$\sum_{(i_1, j_1, \ldots, i_p, j_p) \in M(\pi, s)} \mathbb{E} \left[ \text{tr} (D_{i_1}X_{j_1}^H \cdots D_{i_p}X_{j_p}^H) \right] \quad (32)$$

We observe that the sets $M(\pi, s)$ are not disjoint, but if we put

$$\mathcal{D}(s) = \{ (i_1, j_1, \ldots, i_p, j_p) \in \{1, 2, \ldots, s\}^{2p} \mid i_1, i_2, \ldots, i_p \text{ are distinct} \}$$

the sets $M(\pi, s) \cap \mathcal{D}(s)$, $\pi \in S_p$, are disjoint. Thus, we can write

$$\mathbb{E} \left[ \text{tr} (DXX^H)^p \right] = s^{-p} \sum_{\pi \in S_p} \mathbb{E} \left[ \text{tr} (D_{i_1}X_{j_1}^H \cdots D_{i_p}X_{j_p}^H) \right] \quad (33)$$

If $(i_1, j_1, \ldots, i_p, j_p) \in M(\pi, s) \cap \mathcal{D}(s)$, then $X_1, \ldots, X_s$ are $n \times N$ standard, complex, Gaussian, and we can write the first term of (30) as

$$s^{-p} \sum_{\pi \in S_p} \mathbb{E} \left[ \text{tr} (D_{i_1}X_{j_1}^H \cdots D_{i_p}X_{j_p}^H) \right] \quad (34)$$

Since the cardinality of $M(\pi, s) \cap \mathcal{D}(s)$ is equal to $s(s - 1) \cdots (s - p + 1)$, we have

$$\lim_{s \to \infty} s^{-p} \text{card}(M(\pi, s) \cap \mathcal{D}(s)) = 1,$$
Proof: Let $X_1, \ldots, X_p$ be independent $n \times N$ standard complex Gaussian matrices. We have shown that

$$
\mathbb{E} \left[ \text{tr} \left( DX_1X_H^{(1)} \cdots DX_pX_H^{(p)} \right) \right] = n^{-1} \times \sum_{1 \leq u_1, v_1, \ldots, u_p, v_p \leq n} d(u_1, v_1) \cdots d(u_p, v_p) \mathbb{E} \left[ x(v_1, u_1, 1) \times \prod_{1 \leq u_1, v_1, \ldots, u_p, v_p \leq N} x(u_2, w_1, \pi(1)) \cdots x(v_p, u_p, p) x(u_1, w_p, \pi(p)) \right],
$$

where we assume that the indices give rise to groupings of the Gaussian $x$-variables into pairwise conjugate pairs.

For the rest of the proof, note first that (29) is equivalent to

$$
u_{h+1} = v_{\pi(h)}, \quad (32)$$

$$w_h = w_{\pi(h)}, \quad (33)$$

$$j_h = i_{\pi(h)}, \quad (34)$$

for each $h \in \{1, 2, \ldots, p\}$. Also, $\rho$ restricted to the even numbers is generated by the relations

$$2j \sim 2\pi(j), \quad j \in \{1, \ldots, p\}.$$  

Mapping even numbers $\leq 2p$ onto $\{1, \ldots, p\}$, we see that this also is equivalent to $j \sim \pi(j), \quad j \in \{1, \ldots, p\}$, i.e., the blocks consisting of even numbers are in one-to-one correspondence with the cycles of $\pi$. (33) above also states that equality of the $w_i$’s is decided by this cycle structure, so there are as many degrees of freedom in these as there are blocks of even numbers only in $\rho$. This is $k(\rho)$ by Definition 5. Summing over all $1 \leq w_i \leq N$, we thus get $N^{k(\rho)}$ possibilities, which is one of the contributing factors in (13).

In the same fashion, $\rho$ restricted to the odd numbers is generated by the relations

$$2j - 1 \sim 2\pi^{-1}(j) + 1 = 2(\pi^{-1}(j) + 1) - 1, \quad j \in \{1, \ldots, p\}.$$  

Mapping odd numbers $\leq 2p - 1$ onto $\{1, \ldots, p\}$, we see that this also is equivalent to $j \sim \pi^{-1}(j) + 1$, which is equivalent to $\pi(j) \sim j + 1, \quad j \in \{1, \ldots, p\}$. Also, (32) above states that

$$d(u_{\pi(h)}, v_{\pi(h)}) d(u_{h+1}, v_{h+1}) = d(u_{\pi(h)}, u_{h+1}) d(u_{h+1}, v_{h+1}),$$

so that the factors are grouped into a matrix product, and this grouping coincides with the cycle structure of the blocks consisting of odd numbers only. Each cycle is grouped into a non-normalized trace $\text{Tr}(D^l) = n \text{tr}(D^l)$, where $l$ is the number of elements in the corresponding block. Going through all blocks, we obtain

$$d(u_1, v_1) \cdots d(u_p, v_p) = n^{l(\rho)} \prod_{i=1}^{l(\rho)} \text{tr}(D^{l_i}) = n^{l(\rho)} D_{\rho(\text{odd})},$$

where $l_1, \ldots, l_{l(\rho)}$ are the cardinalities of the blocks consisting of odd numbers only.

In summary, the contribution from the blocks consisting of even numbers only was $N^{k(\rho)}$, and the contribution from the blocks consisting of odd numbers only was $n^{l(\rho)} D_{\rho(\text{odd})}$. Multiplying these together with the factor $n^{-1}$ from taking the trace, the result follows.

APPENDIX B

THE PROOF OF THEOREM 2

For the proof of Theorem 2, the argument showing that only conjugate pairings of Gaussian variables contribute is exactly the same as in Appendix A, hence we need only consider partial permutations. The contribution from the partial permutation $\pi = \pi(\rho_1, \rho_2, q)$ can be written

$$
\mathbb{E} \left[ \text{tr} \left( X_1X_H^{(1)} \cdots X_pX_H^{(p)} \right) \right] = n^{-1} \times \sum_{1 \leq u_1, v_1, \ldots, u_p, v_p \leq n} d(v_1, u_1) \prod_{i \in \rho_1^c} d(v_i, u_i) \times \prod_{p \in \rho_1^c} x(v_{\rho_1}(v_{\rho_1}(v_{\rho_1}(v_{\rho_1}))), w_{\rho_1}(w_{\rho_1}(w_{\rho_1}(w_{\rho_1})))) \times \cdots \times \prod_{p \in \rho_1^c} x(v_{\rho_1}(v_{\rho_1}(v_{\rho_1}(v_{\rho_1}))), w_{\rho_1}(w_{\rho_1}(w_{\rho_1}(w_{\rho_1}))))
$$

Note that if $2k - 1, 2k \in \mathcal{D}$ (i.e. the first relation (9) generating $\sigma$), then $k \in \rho_1^c \cap \rho_2^c$, we find $d(v_k, w_k) d(v_{k+1}, w_k)$ as a part in the matrix product above, which is a part of the matrix product $DD^H$. Similarly, if $2k, 2k + 1 \in \mathcal{D}$, we find a part of the matrix product $D^H D$.

On the other hand, if $2k - 1, 2l \in \mathcal{D}$ with $(2k - 1) + 1 = 2k \sim 2l$ (i.e. the second relation (10) generating $\sigma$), we find that $w_k = w_l$ as in Appendix A, so that we find $d(v_k, w_k) d(v_{k+1}, w_k)$ as a part in the matrix product, which again is a part of the matrix product $DD^H$. We can reason similarly when $k$ and $l$ swap roles, to find a part of the matrix product $D^H D$.

In conclusion, the relations (9) and (10) reflect a cyclic matrix product of the deterministic elements, the length of the product equaling the number of elements in the corresponding block of $\sigma$. Moreover, it is clear that the $D$ and $D^H$ appear in alternating order in the matrix product. In particular, all blocks of $\sigma$ have even cardinality. The cyclic matrix product constitutes a non-normalized trace. Thus, if $\sigma_i$ is the $i$’th block in $\sigma$, $|\sigma_i|$ is even, and the matrix product of the deterministic elements is

$$
\prod_i \text{Tr}(DD^H)^{|\sigma_i|/2} = n^{|\sigma_i|} \prod_i \text{tr}(DD^H)^{|\sigma_i|/2}).
$$

(35), which is seen to be the last term in (14), thus contributes in $\text{tr}(DD^H X D^H X)^p$. The other terms in (14) are identified as follows:

- the first $n$ in the first term $\frac{1}{n^{N|\rho_1|}}$ comes from taking the trace, while $N|\rho_1|$ comes from the normalizing factor for the Gaussian terms (the normalizing factors for the deterministic terms were absorbed in their definition).
- $N^{k(\rho) - kd(\rho)}$ corresponds to the number of all the choices of blocks of $\rho$ with even numbers only, which do not intersect $\mathcal{D} \cup (\mathcal{D} + 1)$,
- $n^{l(\rho) - ld(\rho)}$ corresponds to the number of all the choices of blocks of $\rho$ with odd numbers only, which do not intersect $\mathcal{D} \cup (\mathcal{D} + 1)$,

REFERENCES


