MULTIPLICATIVE FREE CONVOLUTION AND INFORMATION-PLUS-NOISE TYPE MATRICES

BY ØYVIND RYAN and MÉROUANE DEBBAH

University of Oslo and Institut Eurecom

Free probability and random matrix theory has shown to be a fruitful combination in many fields of research, such as digital communications, nuclear physics and mathematical finance. The link between free probability and eigenvalue distributions of random matrices will be strengthened further in this paper. It will be shown how the concept of multiplicative free convolution can be used to express known results for eigenvalue distributions of a type of random matrices called Information-Plus-Noise matrices. The result is proved in a free probability framework, and some new results, useful for problems related to free probability, are presented in this context. The connection between free probability and estimators for covariance matrices is also made through the notion of free deconvolution.

1. Introduction. Applications of free probability have been growing rapidly over the last years. Random matrices and their limit eigenvalue distributions is an area where free probability has proved to be useful [7]. Random matrices are a useful tool for modelling systems, for instance in digital communications [19, 20], nuclear physics [6, 8] and mathematical finance [2]. This paper is a contribution to the random matrix facet of free probability, in that the connection between certain random matrices and free probability is clarified further. We will focus on what we call Information-Plus-Noise Type Matrices, i.e. random matrices on the form

\begin{equation}
W_n = \frac{1}{N}(R_n + \sigma X_n)(R_n + \sigma X_n)^*,
\end{equation}

where $R_n$ and $X_n$ are independent random matrices of dimension $n \times N$. These can be thought of as sample covariance matrices of random vectors $r_n + \sigma x_n$, where $r_n$ can be interpreted as a vector carrying the information in a system, and $x_n$ additive noise, with $\sigma$ the strength of the noise. We impose no assumption on independence between samples. We will use some
common restrictions on the noise: \( X_n \) will contain i.i.d. complex entries of unit variance. \( n \) and \( N \) will be increased so that

\[
\lim_{n \to \infty} \frac{n}{N} = c.
\]

In \cite{DozierSilverstein}, Dozier and Silverstein explain how the limit eigenvalue distribution \( \mu_W \) of the matrix \( W_n \) can be found, based on knowledge of the limit eigenvalue distribution \( \mu_\Gamma \) of the matrix \( \Gamma_n = \frac{1}{N} R_n R_n^* \). The result is expressed in terms of a solution to a function equation (equation (4.1)). We will show that there is an equivalent way of expressing this solution, using the concept of \textit{multiplicative free deconvolution}, denoted by \( \boxminus \) (multiplicative free convolution, as well as freeness and asymptotic freeness are defined in section \( \S \)).

The following is the main result of the paper:

**Theorem 1.1.** Assume that the entries \( X_{ij}^n \) of \( X_n \) are Gaussian, independent and identically distributed with expectation 0 and variance 1. Assume also that the empirical eigenvalue distribution of \( \Gamma_n = \frac{1}{N} R_n R_n^* \) converges in distribution almost surely to a compactly supported probability measure \( \mu_\Gamma \). Then we have that the empirical eigenvalue distribution of \( W_n \) also converges in distribution almost surely to a compactly supported probability measure \( \mu_W \) uniquely identified by

\[
\mu_W \boxminus \mu_c = (\mu_\Gamma \boxminus \mu_c) \boxplus \mu_\sigma^2 I.
\]

Some remarks are needed to explain theorem 1.1. By the \textit{empirical eigenvalue distribution} of an \( n \times n \) random matrix \( X \) we mean the (random) atomic measure

\[
\frac{1}{n} \left( \delta(\lambda_1(X)) + \cdots + \delta(\lambda_n(X)) \right),
\]

where \( \lambda_1(X), \ldots, \lambda_n(X) \) are the (random) eigenvalues of \( X \). That \( \mu_n \) converges in distribution to \( \mu \) means that the moments of \( \mu_n \) converge to the moments of \( \mu \). Theorem 1.1 requires compactly supported measures, and these have moments of all orders.

The conditions in theorem 1.1 are somewhat stronger than those in \cite{DozierSilverstein} due to the restriction to measures with compact support. Contrary to \cite{DozierSilverstein}, we also restrict to noise matrices with Gaussian entries.

Theorem 1.1 yields a short expression for \( \mu_W \), removing the need for solving the equation in \cite{DozierSilverstein} directly. It essentially says that the connection between \( \mu_W \) and \( \mu_\Gamma \) can be expressed compactly in the \textit{deconvolved domain}, where the connection can be viewed as a shift of the spectrum with the noise variance \( \sigma^2 \). The proof of theorem 1.1 is based on methods from free...
probability, with some new results established along the way. Some of these
deserve extra attention, in particular theorem 3.4. This can be thought of as
a version of theorem 1.1 where $\mu_W$ and $\mu_\Gamma$ are interpreted as distributions
of free random variables.

Theorems 3.2 and 3.3 also deserve some extra attention. These address
asymptotic freeness almost everywhere for two random matrices where

1. both converge in distribution almost everywhere to compactly sup-
ported limits, and

2. one of the random matrices are standard unitary (theorem 3.2) or
Gaussian (theorem 3.3).

These results expand known results from [7] for asymptotic freeness. The
proofs of theorems 3.2 and 3.3 use random matrix approximations with
deterministic matrices. Asymptotic freeness of Gaussian/standard unitary
random matrices and uniformly norm-bounded deterministic matrices are
well-known (lemma 4.3.2 in [7]). Unfortunately, norm-bounded deterministic
matrices are not able to approximate the random matrices under consider-
ation. We solve the problem by generalizing to matrices satisfying uniform
$\| \cdot \|_p$-norm bounds instead, where $\| \cdot \|_p$ denotes the Schatten $p$-norm (with
respect to $tr_n$), defined for $p \geq 1$ by $\|A\|_p = tr_n(|A|^p)^{\frac{1}{p}} (A \in M_n(C))$: We
prove that matrices satisfying such bounds can be used to approximate our
random matrices, and that they also give asymptotic freeness as in lemma
4.3.2 in [7] (theorem 3.1).

Theorem 1.1 is actually proved by combining theorems 3.3 and 3.4 through
another approximation argument (see theorem 3.5). While [3] restricts to the
distribution of $\frac{1}{N}(R_n + \sigma X_n)(R_n + \sigma X_n)^*$, we show more in that any mixed
moments of $\frac{1}{N}R_nR_n^*$ and $\frac{1}{N}X_nX_n^*$ are obtained through our asymptotic freeness
results.

Recent works [17, 18] show that multiplicative free convolution also ad-
mits an efficient implementation in terms of the moments of the operand
measures. The basic results on free probability we need for this are proved
in this paper (theorems 2.1 and 2.2). A consequence is that existing compu-
tational frameworks can be used in obtaining $\mu_\Gamma$ and $\mu_W$. In [18], $\mu_\Gamma$
and $\mu_W$ are illustrated in terms of signal processing applications, and simulations
are run using a computational framework building on theorems 2.1
and 2.2. A useful consequence of the link with free probability is that that
the ”inverse problem” (i.e. that of finding $\mu_\Gamma$ from $\mu_W$) can be solved within
the same framework, since the framework embraces convolution as well as
deconvolution.

The eigenvalue distribution of $\Gamma_n$ provides us with possibilities for estimat-
ing the covariance matrices of the system through the so-called $G$-estimators \cite{5}. These will be reviewed, and it will be shown how multiplicative free convolution can be used to rewrite such estimators to a very simple form. It will be apparent from this that the $G^2$-estimator actually can be viewed as a step in expressing $\mu_W$ from $\mu_G$.

While the results mentioned here are hard to prove, some of them should come as no surprise. For instance, \cite{14} has already made the connection between Information-Plus-Noise type matrices and multiplicative free convolution. This paper also indicates that some of the mentioned results are already known, by saying that random matrices with Haar-distributed eigenvectors are asymptotically free from any random matrices independent from them. However, the generality in which this should hold is not indicated. Also, \cite{14} considers only Gaussian matrices, and the connection with already existing estimators of covariance matrices was not made.

This paper is organized as follows. Section 2 contains notation and preliminaries for various free probability tools, like free transforms and combinatorial aspects. The mentioned implementation of free convolution builds on the combinatorial expression of freeness, and the results needed on this are explained in section 2.1. The proof for theorem 1.1 is presented in section 3. A sketch of the proof is first given, followed by the proofs for theorems 3.1, 3.2, 3.3 and 3.4. Section 4 first states the results we need from \cite{3}, and sketch the proof for the equivalence of these and theorem 1.1. This sketch is then followed by the rest of the details. The various transforms used in free probability (section 2) are used in this direction. In section 5 we state the principles of $G$-analysis and the expression for the $G^2$-estimator. We also prove the theorem which expresses the $G^2$-estimator in terms of free probability.

2. Notation and preliminaries. In the following, uppercase symbols will be used for matrices, and $(\cdot)^*$ will denote hermitian transpose. $I_n$ will represent the identity matrix of order $n$. We will focus here on certain non-commutative probability spaces. A noncommutative probability space is a pair $(A, \phi)$ where $A$ is a unital $\ast$-algebra and $\phi$ is a normalized (i.e. $\phi(I) = 1$) linear functional on $A$. The elements of $A$ are called random variables. The probability spaces we will encounter are mostly $(M_n(C), tr_n)$, i.e. $n \times n$-matrices equipped with the normalized trace. Any matrix can be associated with a probability measure through it’s eigenvalue distribution. We will mostly be concerned with probability measures with compact support.

Definition 2.1. A family of unital $\ast$-subalgebras $(A_i)_{i \in I}$ will be called
a free family if

\begin{equation}
\begin{cases}
    a_j \in A_{i_j} \\
    i_1 \neq i_2, i_2 \neq i_3, \ldots, i_{n-1} \neq i_n \\
    \phi(a_1) = \phi(a_2) = \cdots = \phi(a_n) = 0
\end{cases}
\end{equation}

(2.1) enables us to calculate the mixed moments of \(a_1\) and \(a_2\) when they are free. In particular, the moments of \(a_1 + a_2\) and \(a_1 a_2\) can be calculated. This gives us two new probability measures, which depend on the probability measures of \(a_1, a_2\) only (i.e. not on their realizations). Therefore we can define two operations on the set of probability measures: Additive free convolution

\begin{equation}
\mu_1 \boxplus \mu_2
\end{equation}

for the sum of free random variables, and multiplicative free convolution

\begin{equation}
\mu_1 \boxtimes \mu_2
\end{equation}

for the product of free random variables.

Let \(F^{\mu_A}\) denote the empirical distribution function (e.d.f.) of the eigenvalues of \(A\) (so that \(F^{\mu_A}(x)\) is the proportion of eigenvalues of \(A\) which are \(\leq x\)). When we have a series of e.d.f.’s \(F^{\mu_{A_n}}\), we will use the notation

\[ F^{\mu_{A_n}} \xrightarrow{D} F^{\mu} \]

for weak convergence, where \(F^{\mu}\) is the cumulative distribution function of the measure \(\mu\). We will also write a.s. as shorthand notation for almost sure convergence.

Some random matrices and limit distributions occur naturally in many contexts. If the entries of the \(n \times N\) (with \(\lim_{n \to \infty} \frac{n}{N} = c\)) random matrices \(W_n\) have zero mean and unit variance, the empirical eigenvalue distribution of \(\frac{1}{N} W_n W_n^*\) converges almost surely to the so-called Marčenko Pastur law \(\mu_c\) (page 9). These are also called the free Poisson distributions, and are characterized by the density

\begin{equation}
\rho_{\mu_c}(x) = \left(1 - \frac{1}{c}\right)^+ \delta(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi cx},
\end{equation}

where \((z)^+ = \max(0, z)\), \(a = (1 - \sqrt{c})^+\) and \(a = (1 + \sqrt{c})^+\). Similar notation to the Marčenko Pastur law is used for the distribution \(\mu_a\) of a random variable \(a\). We avoid confusion by never using \(c\) to denote random variables. \(\mu_1\) will always mean the Marčenko Pastur law with parameter one. Marčenko
Pastur laws are some of the most basic random matrix building blocks, as they appear as limits for large random matrices in many contexts. This paper will demonstrate that this is indeed the case for the type of systems we consider also.

We will not use the characterization of the Marchenko Pastur law as in (2.4) directly. Rather we will work with equivalent expressions of it through the transforms defined in this section. The transforms we define will only be applied for probability measures with support contained on the positive real line.

The Stieltjes transform (21 page 38) of a probability measure \( \mu \) is the analytic function on \( \mathbb{C}^+ = \{ z \in \mathbb{C} : \Im z > 0 \} \) defined by

\[
(2.5) \quad m_\mu(z) = \int_{-\infty}^{\infty} \frac{1}{\lambda - z} dF^\mu(\lambda).
\]

A convenient inversion formula for the Stieltjes transform also exists, so that \( m_\mu \) uniquely identifies \( \mu \). If \( \mu \) is assumed to have nonnegative support, \( m_\mu \) can be analytically continued to the negative part of the real line. If \( \mu = \mu_X \) for a non-negative random variable \( X \), \( m_\mu \) is strictly monotone on the negative real line, taking values in the interval \([0, E(\frac{1}{X})]\). We will use the fact that if we know \( m_\mu(z) \) in an interval \((-z, 0)\) for \( z < 0 \), we also know \( m_\mu \) for all other values of \( z \), and hence we also know \( \mu \) (use the Stieltjes inversion formula).

The \( \eta \)-transform (21 page 40) is defined for measures \( \mu \) with support on the positive real line, and for nonnegative real numbers by

\[
(2.6) \quad \eta_\mu(z) = \int_{-\infty}^{\infty} \frac{1}{1 + z\lambda} dF^\mu(\lambda).
\]

\( \eta(z) \) is a strictly monotonically decreasing function. As such it simplifies many derivations and statements of results. The inverse is tightly connected to the S-transform (see below). It’s connection with the Stieltjes transform is

\[
(2.7) \quad \eta_\mu(z) = \frac{1}{z} m_\mu(-\frac{1}{z}), \quad m_\mu(z) = -\frac{1}{z} \eta_\mu(-\frac{1}{z}).
\]

Therefore \( \eta_\mu(z) \) uniquely identifies \( m_\mu(z) \), since \( m_\mu(z) \) for real, negative \( z \) can be continued analytically to \( \mathbb{C}^+ \). We will use the fact that if we know \( \eta_\mu(z) \) in an interval \((0, z)\) for \( z > 0 \), we also know \( \mu \).

The \( R \)-transform (21 page 48) has domain of definition \( \mathbb{C}^+ \) and can be defined in terms of the Stieltjes transform as

\[
(2.8) \quad R_\mu(z) = m_\mu^{-1}(-z) - \frac{1}{z}.
\]
The importance of the $R$-transform comes from its additive property for the distribution of the sum of free random variables $A_1$ and $A_2$,

\begin{equation}
R_{\mu_{A_1+A_2}}(z) = R_{\mu_{A_1}}(z) + R_{\mu_{A_2}}(z).
\end{equation}

Slightly different versions of the $R$-transform are encountered in the literature. The one above is from [21]. In connection with free combinatorics, another definition is used, namely $R_\mu(z) = zR_\mu(z)$. Of course, $R_\mu(z)$ also satisfies (2.9).

The $S$-transform ([21] page 50) is defined on $(-1, 0)$. It can be defined in terms of the $\eta$-transform by

\begin{equation}
S_\mu(z) = -\frac{z+1}{z} \eta_\mu^{-1}(z+1).
\end{equation}

The Marchenko Pastur law (2.4) can be shown to have S-transform $S_{\mu_c}(z) = \frac{1}{1+cz}$ ([21] page 51). The importance of the $S$-transform comes from its multiplicative property for the distribution of the product of free random variables $A_1$ and $A_2$:

\begin{equation}
S_{\mu_{A_1A_2}}(z) = S_{\mu_{A_1}}(z)S_{\mu_{A_2}}(z).
\end{equation}

If the values of $\eta_\mu(z)$ or $S_\mu(z)$ are known in an interval, one also knows $\mu$.

Freeness, additive and multiplicative free convolution have a combinatorial description involving these transforms which we will use for in some of our proofs. These combinatorial descriptions build on the concept of noncrossing partitions:

**Definition 2.2.** A partition $\pi$ 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 $\{1, \ldots, n\}$ is denoted $NC(n)$.

$NC(n)$ becomes a lattice under the refinement order of partitions. An ingredient we need in making the connections between freeness and the noncrossing partitions is the complementation map of Kreweras, which is a lattice anti-isomorphism of $NC(n)$. To define this we need the circular representation of a partition: We mark $n$ equidistant points $1, \ldots, n$ (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 $\bar{1}, \ldots, \bar{n}$ on the midpoints of the $1, \ldots, n$ (so that $i$ is the midpoint of the segment from $i$ to $i+1$). The
Fig 1. The circular representation of a partition of \{1, ..., 16\}.

complement of the set \( \cup_i H_i \) is again a union of disjoint convex sets \( \tilde{H}_i \). All this is demonstrated in figure 1, where the \( \tilde{H}_i \) are the scrambled areas with dashed borders, the \( H_i \) are the scrambled areas with non-dashed borders. We will refer to this figure heavily during the proof of theorem 3.4. We can now define the Kreweras complementation map:

**Definition 2.3.** The Kreweras complement of \( \pi \), denoted \( K(\pi) \), is the partition on \( \{\bar{1}, ..., \bar{n}\} \) determined by

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

The connection between the \( R \)-transform and noncrossing partitions comes through the moment-cumulant formula, which relates the moments and the \( R \)-transform coefficients (also called cumulants) for the distribution of a random variable.

**Lemma 2.1.** Write the \( R \)-transform as a power series, \( R_{\mu_0}(z) = \sum_n \alpha_n z^n \).
Then

\[ \phi(a^n) = \sum_{\pi = \{B_1, \ldots, B_k\} \in NC(n)} \prod_{i=1}^{k} \alpha_{|B_i|}, \]

This can be used as an alternative definition of the R-transform. We also need to define the multidimensional R-transform for the joint distribution of a sequence of random variables. Denote by \(C(z_1, \ldots, z_m)\) the space of complex power series in \(m\) noncommuting variables \(z_i\) with vanishing constant term. These can be written in the form

\[ \sum_{k \geq 1} \sum_{i_1, \ldots, i_k} a_{i_1, \ldots, i_k} z_{i_1} \cdots z_{i_k}. \]

In referring to the coefficients of a power series \(f\) on this form we will write

\[ [\text{coef}(i_1, \ldots, i_k)](f) = a_{i_1, \ldots, i_k}, \]

and if \(\pi = \{B_1, \ldots, B_m\}\),

\[ [\text{coef}(i_1, \ldots, i_k)|B_i](f) = a_{(i_j)_{j \in B_i}}, \]
\[ [\text{coef}(i_1, \ldots, i_m); \pi](f) = \prod_{i}[\text{coef}(i_1, \ldots, i_k)|B_i](f). \]

For power series in one variable, the coefficients will also be written in the form \([\text{coef}_k](f)\). For \(n\) random variables \(a_1, \ldots, a_n\) we define their joint moment series as the power series \(M_{\mu a_1, \ldots, a_n} \in C(z_1, \ldots, z_n)\) such that

\[ M_{\mu a_1, \ldots, a_n}(z_1, \ldots, z_k) = \sum_{m \geq 1} \sum_{i_1, \ldots, i_m} \phi(a_{i_1} \cdots a_{i_m}) z_{i_1} \cdots z_{i_m}, \]

and we define their joint R-series as the unique power series \(R_{\mu a_1, \ldots, a_n} \in C(z_1, \ldots, z_n)\) such that

\[ \phi(a_{i_1} \cdots a_{i_m}) = \sum_{\pi \in NC(m)} [\text{coef}(i_1, \ldots, i_m); \pi](R_{\mu a_1, \ldots, a_n}). \]

The result we will use connecting the joint R-series and freeness is the following:

**Lemma 2.2.** \((\{a_1, \ldots, a_n\}, \{b_1, \ldots, b_m\})\) is a free family if and only if

\[ R_{\mu a_1, \ldots, a_n, b_1, \ldots, b_m}(z_1, \ldots, z_{n+m}) = R_{\mu a_1, \ldots, a_n}(z_1, \ldots, z_n) + R_{\mu b_1, \ldots, b_m}(z_{n+1}, \ldots, z_{n+m}). \]
This lemma is often summarized by saying that the joint \( R \)-series of free random variables has no mixed terms. A special form of (2.13) and lemma 2.2 we will use is the following: If \((a_1, a_2)\) is a free family, and a mixed term \(a_{i_1} \cdots a_{i_m}\) is given, form the partition with two blocks \(\sigma = \{\sigma_1, \sigma_2\}\), where \(\sigma_k = \{j | a_{i_j} = k\}\). Then

\[
\phi(a_{i_1} \cdots a_{i_m}) = \sum_{\pi \leq \sigma \in NC(m)} [\text{coef}(i_1, \ldots, i_m); \pi](R_{\mu_{a_1, a_2}}).
\]

Our combinatorial connection with multiplicative free convolution can be made complete with the help of the following definition [10], [11]:

**Definition 2.4.** Given two power series \(f\) and \(g\), their boxed convolution \(\bigtriangledown\) is defined by

\[
[\text{coef}(i_1, \ldots, i_m)](f \bigtriangledown g) = \sum_{\pi \in NC(m)} [\text{coef}(i_1, \ldots, i_m); \pi](f)[\text{coef}(i_1, \ldots, i_m); K(\pi)](g).
\]

Boxed convolution is commutative only on power series in one variable [13]. It satisfies the associative law, but not the distributive law. It does not satisfy linearity properties w.r.t. scalar multiplication. However, the following holds and will be useful to us:

\[
[\text{coef}f_n]((cf)\bigtriangledown cg) = [\text{coef}f_n](c^{n+1}(f \bigtriangledown g))
\]

and

\[
[\text{coef}f_n](\bigtriangledown (cId)) = [\text{coef}f_n](c^n f).
\]

Here we used the shorthand notation \(c^n f\) for the power series defined by \([\text{coef}f_n](c^n f) = c^n[\text{coef}f_n](f)\). The first statement is easily proved using the fact that \(|\pi| + |K(\pi)| = n + 1\) for any \(\pi \in NC(n)\) [13]. The second statement is trivial. The following result holds for multiplicative free convolution [13]:

**Lemma 2.3.** If \(\{a_1, \ldots, a_n\}, \{b_1, \ldots, b_n\}\) is a free family, then

\[
R_{\mu_{a_1 b_1, \ldots, a_n b_n}} = R_{\mu_{a_1, \ldots, a_n}} \bigtriangledown R_{\mu_{b_1, \ldots, b_n}}
\]

One can also define additive and multiplicative free deconvolution in most cases, i.e. finding \(\mu_2\) in (2.3) when \(\mu_1 \boxtimes \mu_2\) are known.

**Definition 2.5.** Given probability measures \(\mu\) and \(\mu_2\). When there is a unique probability measure \(\mu_1\) such that \(\mu = \mu_1 \boxtimes \mu_2\), we will denote \(\mu_1 = \mu \bowtie \mu_2\). We say that \(\mu_1\) is the multiplicative free deconvolution of \(\mu\) with \(\mu_2\).
We can define additive free deconvolution similarly. Note that free deconvolution is defined only for a subset of all probability measures, since measures exist which can’t be expressed on the forms $\mu_1 \boxplus \mu_2$ or $\mu_1 \boxtimes \mu_2$. Deconvolution can, however, also be viewed as a formal operation on a sequence of moments. Viewed as such, multiplicative free deconvolution is well-defined when we have non-vanishing first moments. This can be seen from the combinatorial description of multiplicative free convolution (2.15). In light of (2.16), it is obvious from (2.15) that the cumulants in $R_{\mu_2}$ can be calculated recursively from those of $R_{\mu_1}$ and $R_{\mu_1 \boxtimes \mu_2}$, when the first coefficient of $R_{\mu_1}$ (which equals the first moment) is known. Since the main theorem relates to the moments of the involved measures (it is a statement on convergence in distribution), we will in the following view deconvolution in terms of the moments only.

A form of (2.16) which will be useful to us is for the case $n = 1$. If we write $\mu_a = (\mu_a \boxtimes \mu_c) \boxtimes \mu_c$, we get

$$R_{\mu_a \boxtimes \mu_c} = R_{\mu_a} R_{\mu_c}^{-1}.$$  \hfill (2.17)

The facts we will use concerning boxed convolution are the following, relating moment series, $R$-series of general random variables (in particular projections and free Poisson random variables), the Zeta series, which is defined as

$$\text{Zeta}(z_1, \ldots, z_n) = \sum_{k} \sum_{i_1, \ldots, i_k} z_{i_1} \cdots z_{i_k},$$

the Moeb series (which is the inverse of Zeta under composition with $\boxtimes$) and $Id$ (which is the unit under composition with $\boxtimes$):

$$M_\mu = R_\mu \boxtimes \text{Zeta} \quad \text{and} \quad R_\mu = M_\mu \boxtimes \text{Moeb}$$

$$M_{\mu_p} = c \text{Zeta} \quad \text{and} \quad R_{\mu_c} = c^{-1} \text{Zeta}.$$  \hfill (2.18)

Here $p$ is a projection with $\phi(p) = c$. Our definition of $\mu_c$ differs from that of $\boxtimes$, for purposes of compatibility with $\boxtimes$ [14]. Consequently, the expressions for the $R$-transforms are different. In the terminology of $\boxtimes$, the $R$-series would be $c \text{Zeta}$. The following definition [12] will also be in use:

**Definition 2.6.** A pair $(a, b)$ of noncommutative random variables is called an $R$-diagonal pair if it’s $R$-series is of the form

$$R_{\mu_{a,b}}(z_1, z_2) = \sum_{n=1}^{\infty} \alpha_n ((z_1 z_2)^n + (z_2 z_1)^n).$$  \hfill (2.19)

An element $a$ will be said to be an $R$-diagonal element if $(a, a^*)$ is an $R$-diagonal pair. The one-variable series $\sum_{n=1}^{\infty} \alpha_n z^n$ will be called the determining series of the $R$-diagonal pair $(a, b)$.
We will use the fact that if \( a \) is an \( R \)-diagonal element, it’s determining series can be written as \( R_{\mu_{a,a}} \square \text{Moeb} [7] \). Two important \( R \)-diagonal elements are

1. the \textit{Haar unitary}, which can be defined as a unitary \( u \) satisfying \( \phi(u^n) = 0 \) for all \( n \in \mathbb{Z} \neq 0 \), and
2. the \textit{circular element}, which can be defined as an element \( s \) whose \( * \)-distribution \( \mu_{s,s} \) satisfies \( R_{\mu_{s,s}}(z_1, z_2) = z_1z_2 + z_2z_1 \).

The concept of \( R \)-diagonally was in fact invented in search of a common approach for Haar unitaries and circular elements \[12\]. Haar unitaries are very important in asymptotic random matrix results. In \( W^* \)-probability spaces, when the isometric part of an \( R \)-diagonal element has kernel equal to zero, the isometric part is actually a Haar unitary.

2.1. \textit{Implementation of free convolution}. While free convolution has an abstract definition, the combinatorial description given in this section can actually be used to obtain an efficient implementation. In many practical cases, free convolution with \( \mu_c \) is what we are interested in. Such free convolution is simplified through the following result.

\begin{theorem}
(2.20) \[ (cM_{\mu}) \square \text{Zeta} = c \left( M_{\mu \boxtimes \mu_c} \right) \].
\end{theorem}

\textbf{Proof.} To see this, start by combining (2.16) with (2.18) to get
\[ R_{\mu \boxtimes \mu_c} = R_\mu \square R_{\mu_c} = R_\mu \square (c^m-1 \text{Zeta}) \].
After convolving both sides with \( \text{Zeta} \), we get
\begin{equation}
M_{\mu \boxtimes \mu_c} = M_\mu \square (c^m-1 \text{Zeta}).
\end{equation}

To prove (2.20), rewrite the left hand side as
\[ \sum_{\pi \in \text{NC}(m)} c^{||\pi||} [\text{coef}_{m;\pi}] M_{\mu} \].
Since \( ||\pi|| + |K(\pi)| = m + 1 \), this equals
\begin{align*}
& c \sum_{\pi \in \text{NC}(m)} [\text{coef}_{m;\pi}] (M_{\mu}) c^{m-|K(\pi)|} \\
& = c \sum_{\pi \in \text{NC}(m)} [\text{coef}_{m;\pi}] (M_{\mu}) c^{m-|K(\pi)|} [\text{coef}_{m;K(\pi)}] (\text{Zeta}) \\
& = c \sum_{\pi \in \text{NC}(m)} [\text{coef}_{m;\pi}] (M_{\mu}) c^{m} [\text{coef}_{m;K(\pi)}] (c^{-1} \text{Zeta}) \\
& = c \sum_{\pi \in \text{NC}(m)} [\text{coef}_{m;\pi}] (M_{\mu}) [\text{coef}_{m;K(\pi)}] (c^{m-1} \text{Zeta}) \\
& = c (M_\mu \square (c^{m-1} \text{Zeta})) ,
\end{align*}
substituting (2.21) proves the claim. \( \square \)
In summary, if we need to compute the moments of $\mu \boxtimes \mu_c$, one can first compute the moment series $cM_{\mu}$, then use this to compute the left hand side of (2.20). According to (2.21) and (2.20), the moment series of $\mu \boxtimes \mu_c$ can then be computed from this with an additional scaling with $\frac{1}{c}$.

In other words, convolving with $\mu_c$ is equivalent to convolving with $\mu_1$ (with additional scalings of power series taken into account), since $R_{\mu_1} = Zeta$. It turns out that boxed convolution with $Zeta$ is easy to compute, as the following result shows. The result is stated in terms of the moment-cumulant formula, since the relation between cumulants and moments are given by boxed convolution with $Zeta$.

**Theorem 2.2.**

(2.22) \[ [\text{coef}_m](M_{\mu}) = \sum_{k=1}^{m} [\text{coef}_k](R_{\mu})[\text{coef}_{m-k}](1 + M_{\mu})^k. \]

**Proof.** For each $\pi \in NC(m)$, fix the block $B_1 = \{b_{11}, ..., b_{1k}\}$ in $\pi$ containing 1, and let $NC(m, B_1)$ be the set of all noncrossing partitions which contain $B_1$ as a block. Rewrite the definition of boxed convolution (2.15) to (2.23)

\[ [\text{coef}_m](M_{\mu}) = \sum_{B_1} \sum_{\pi \in NC(m, B_1)} [\text{coef}_m; \pi](R_{\mu})[\text{coef}_m; K(\pi)](Zeta) \]

\[ = \sum_{B_1} \sum_{\pi \in NC(m, B_1)} [\text{coef}_m; \pi](R_{\mu}). \]

Blocks in $\pi \in NC(m, B_1)$ other than $B_1$ must be entirely contained in one of $\{b_{11} + 1, ..., b_{12} - 1\}, ..., \{b_{1k} + 1, ..., b_{11} - 1\}$. This means that the inner summand in (2.23) can be rewritten to

(2.24) \[ [\text{coef}_k](R_{\mu}) \prod_{i=1}^{k} \left( \sum_{\pi \in NC(b_{1(i+1)} - b_{1i} - 1)} [\text{coef}_m; \pi](R_{\mu}) \right). \]

From the moment-cumulant formula it is seen that each sum here is simply a moment, so we can rewrite to

\[ [\text{coef}_k](R_{\mu}) \prod_{i=1}^{k} [\text{coef}_{b_{1(i+1)} - b_{1i} - 1}](1 + M_{\mu}), \]

where the summand 1 in $1 + M_{\mu}$ accounts for elements $i$ in (2.24) with $b_{1(i+1)} = b_{1i} + 1$ (i.e. consecutive elements in a block). All in all, (2.23) can be rewritten to

(2.25) \[ \sum_k \sum_{|B_1|=k} [\text{coef}_k](R_{\mu}) \prod_{i=1}^{k} \left( [\text{coef}_{b_{1(i+1)} - b_{1i} - 1}](1 + M_{\mu}) \right). \]
Write \( a_i = b_{1(i+1)} - b_{1i} - 1 \), and note that
\[
\sum_{i=1}^{k} a_i = \sum_{i=1}^{k} \left( b_{1(i+1)} - b_{1i} - 1 \right) = m - k.
\]
The \( a_i \) are in one-to-one correspondence with all candidates for \( B_1 \), so that we can rewrite (2.25) to
\[
\sum_{k} [\text{coef}_k](R_{\mu}) \sum_{a_1, \ldots, a_k} \prod_{i=1}^{k} ([\text{coef}_{a_i}](1 + M_{\mu})) \quad \sum_{a_i=m-k}
\]
The inner sum here is easily recognized as coefficient \( m - k \) in the power series \((1 + M_{\mu})^k\) (one factor for each \( a_i \)). Putting things together we get (2.22). □

In (2.22) we see that there is no reference to noncrossing partitions. (2.22) can be used easily in calculating moments recursively from cumulants. The coefficients in the power series \((1 + M_{\mu})^k\) can be computed in terms of \( k \)-fold (classical) convolution. This is done in [18], where many multiplicative free convolutions are computed based on (2.22). The actual implementation of (2.22) used in [18] is contained in [16].

Free convolution as introduced here is just defined for compactly supported probability measures.

3. Proof of theorem 1.1
In what follows we first sketch the proof of theorem 1.1. After this follows proofs for theorems needed in the proof.

First we prove the following variant of lemma 4.3.2 in [17], which can be used together with the Borel-Cantelli lemma to prove almost sure convergence. It is slightly more general in the sense that boundedness in the operator norm \( \| \cdot \| \) is not assumed, only boundedness in \( \| \cdot \|_p \), for \( p \geq 1 \). This weaker boundedness assumption is needed since uniformly norm-bounded matrices are not sufficient to approximate all compactly supported probability measures almost surely. Recall that an \( n \times n \) unitary random matrix is called standard unitary if it’s distribution equals the Haar probability measure on \( U(n) \).

**Theorem 3.1.** Let \( U(s, n)_{s \in S} \) be an independent family of \( n \times n \) standard unitary random matrices. Let \( s_1, \ldots, s_l \in S \), \( m_1, \ldots, m_l \in \mathbb{Z} \setminus \{ 0 \} \), and let \( R_p \geq 0, p \geq 1 \) be constants. Then
\[
E \left( |tr_n (U(s_1, n)^{m_1}D_1(n)U(s_2, n)^{m_2}D_2(n)\cdots U(s_l, n)^{m_l}D_l(n))|^2 \right)
\]
MULTIPLICATIVE FREE CONVOLUTION

is \( O(n^{-2}) \) as \( n \to \infty \) uniformly for the choice of any \( D_r(n) \in M_n(C) \) (1 \( \leq r \leq l \)) such that for 1 \( \leq r \leq l \) either

\[
\text{tr}_n(D_r(n)) = 0 \quad \text{and} \quad \|D_r(n)\|_p \leq R_p \quad (n \in \mathbb{N})
\]

or

\[
D_r(n) = I_n \quad (n \in \mathbb{N}) \quad \text{and} \quad s_r \neq s_{r+1} \quad (\text{with } s_{l+1} = s_1).
\]

Also, for a given \( l \), there exists a \( p_l \) such that the same statement holds as long as the \( \| \cdot \|_p \)-norm bounds are satisfied for \( p \leq p_l \) only.

The proof is in section 3.1. It somewhat simplifies the proof of lemma 4.3.2 in [7], and can also be used to simplify the proof of theorem 4.3.5 in [7]. As in [7], theorem 3.1 is sufficient to prove asymptotic freeness almost everywhere for the family

\[
\left( \{U(s,n), U(s,n)^*\}_{s \in S}, \{D(t,n), D(t,n)^* : t \in T\} \right)
\]

when the \( D_r(n) \) is known to have a limit distribution. It will also be useful to us that theorem 3.1 gives us bounds also in cases where the \( D_r(n) \) do not converge to a limit. The \( D_r(n) \) model in our case concerns \( \frac{1}{\sqrt{n}}R_n \) random matrices, for which it is not known whether an almost sure limit exists (only that \( \Gamma_n = \frac{1}{n}R_nR_n^* \) has an almost sure limit). Also, theorem 3.1 gives us grounds for proving that only the lower mixed moments converge to zero. It can be applied to cases where only the lower \( \| \cdot \|_p \)-norms are known to be bounded, in which only lower mixed moments can be bounded.

What we really want is to use random matrices \( R_n \) independent from the \( U_n \) instead of the deterministic matrices \( D_r(n) \). This is addressed by the following theorem. We restrict to the case of one standard unitary random matrix.

**Theorem 3.2.** Let \( U_n \) be \( n \times n \) standard unitary random matrices, and let \( R_n \) be random matrices independent from \( U_n \), such that \( R_nR_n^* \) converges in distribution almost surely to a compactly supported probability measure \( \rho \). Then

\[
|\text{tr}_n \left( U_n^{m_1}P_1(R_n)U_n^{m_2}P_2(R_n) \cdots U_n^{m_l}P_l(R_n) \right) | \to 0 \quad \text{a.s.}
\]

uniformly for any choice of polynomials \( P_1, \ldots, P_l \) such that \( \text{tr}_n(P_3(R_n)) = 0 \) for all \( 1 \leq i \leq l \).

The proof is in section 3.2. As for theorem 3.1, theorem 3.2 is sufficient to prove asymptotic freeness almost everywhere for the family \( (U_n, R_n) \) when the \( R_n \) are additionally known to have a limit distribution.
The proof is split in two: First (3.2) is shown for random matrices satisfying bounds of the form \( \|R_n\|_p \leq R_p \) \((p \geq 1)\). The proof in this case uses theorem 3.1 and is quite short. The more general case of compactly supported probability measures is proved with an approximation argument.

The next step is to pass from standard unitary random matrices \( U_n \) to standard Gaussian random matrices \( X_n \). Note that it could be possible to skip starting with standard unitary random matrices altogether, by building directly on results for almost sure convergence of Gaussian random matrices like those in [15]. We have chosen the approach with standard unitary random matrices for compatibility with [7]. We will prove the following:

**Theorem 3.3.** Let \( X_n \) be \( n \times n \) standard Gaussian random matrices, and let \( R_n \) be random matrices independent from \( X_n \), such that \( R_n R_n^* \) converges in distribution almost surely to a compactly supported probability measure \( \rho \). Then

\[
|\text{tr}_n (Q_1(X_n)P_1(R_n)Q_2(X_n)P_2(R_n) \cdots Q_l(X_n)P_l(R_n))| \to 0 \text{ a.s.}
\]

uniformly for any choice of polynomials \( Q_1, \ldots, Q_l, P_1, \ldots, P_l \) such that

\[
\text{tr}_n (Q_i(R_n)) = 0 \text{ and } \text{tr}_n (P_i(R_n)) = 0
\]

for all \( 1 \leq i \leq l \).

The proof is quite short, and also presented in section 3.2. Note that the approximation argument used in the proof of theorem 4.3.5 in [7] does not work in this case. As for theorem 3.2, theorem 3.3 is enough to prove asymptotic freeness almost everywhere when the \( R_n \) are additionally known to have a limit distribution. Just as theorem 3.1 gives bounds for mixed moments also in cases where the deterministic matrices do not converge in distribution, theorem 3.2 and it’s counterpart for Gaussian random matrices can be used to bound mixed moments in cases where it is only known that the \( R_n \) matrices satisfy \( \| \cdot \|_p \)-norm bounds.

To finish the proof we will model our situation through the following theorem, which is stated independently of a random matrix setting.

**Theorem 3.4.** Suppose that \( a \) and \( \{p, b\} \) are \( \ast \)-free, with a \( R \)-diagonal and \( p \) a projection with \( \phi(p) = c \). In the reduced probability space \( (p, Ap, \phi(p)^{-1} \phi) \), \( \mu_{p(a+b)(a+b)^*} \) is uniquely identified by \( \mu_{paa^*} \) and \( \mu_{pbb^*} \) through the equation

\[
(3.3) \quad \mu_{p(a+b)(a+b)^*} \otimes \mu_c = (\mu_{paa^*} \otimes \mu_c) \boxplus (\mu_{pbb^*} \otimes \mu_c)
\]

In particular, \( \mu_{p(a+b)(a+b)^*} \) has no dependence on mixed moments of \( a \) and \( b \).
This will be proved in section 3.3. Note that there is no assumption on freeness between \( p \) and \( b \). The case \( c = 1 \) is particularly interesting, and corresponds to \( p = I \). In this case, \( \mu_{(a+b)(a+b)^*} \) is uniquely identified by \( \mu_{aa^*} \) and \( \mu_{bb^*} \), and (3.3) is simply

\[
\mu_{(a+b)(a+b)^*} \boxtimes \mu_1 = (\mu_{aa^*} \boxtimes \mu_1) \boxplus (\mu_{bb^*} \boxtimes \mu_1).
\]

This equation has an interpretation in terms of square random matrices.

Due to (3.3), \( R \)-diagonality relieves us from dependencies of many mixed moments, so that some cancellation phenomenon must occur. This also happens in other cases. If \( a \) and \( b \) are free, (13) expresses the distribution of the free commutator, i.e.

\[
R_{\mu_{(a-b)(a-b)^*}}(z) = 2 \left( R_{\mu_a}^{\text{even}} \boxtimes R_{\mu_b}^{\text{even}} \text{Zeta} \right) (z^2),
\]

where \( R_{\mu}^{\text{even}}(z) = \sum_{n=1}^{\infty} \alpha_n z^n \) whenever \( R(z) = \sum_{n=1}^{\infty} \alpha_n z^n \). (3.5) holds also when \( a \) and \( b \) are not \( R \)-diagonal. (3.5) also expresses a connection with multiplicative free convolution with \( \mu_1 \), since boxed convolution with the \( \text{Zeta} \)-series is involved.

Theorem 3.4 has a more general flavour than theorem 4.1 since the limits \( \frac{1}{n} X_n X_n^\ast \) from theorem 4.1 do not include all \( R \)-diagonal pairs. The following limiting version of theorem 3.4 will be useful in finishing the proof of theorem 4.1.

**Theorem 3.5.** Let the random variables \( \{a_n, b_n, p_n\} \in (\mathcal{A}_n, \phi_n) \), \( \{a, p\} \in (\mathcal{A}, \phi) \) be given, where \( a \) is \( R \)-diagonal and \( p, p_n \) are projections with \( \phi(p) = \phi_n(p_n) = c \). Form the random variable \( p a a^* p \) in \( (pAp, \phi(p)^{-1} \phi) \), and the random variables \( p_n b_n b_n^* p_n \) and \( p_n (a_n + b_n)(a_n + b_n)^* p_n \) in \( (p_n A_n p_n, \phi(p_n)^{-1} \phi_n) \). If

\[
\mu_{a_n, a_n^*} \to \mu_{a, a^*}, \quad \mu_{p_n} \to \mu_p,
\]

and

\[
\mu_{p_n b_n b_n^* p_n} \to \mu_c,
\]

in distribution, moments are uniformly bounded in \( n \), and mixed moments of \( \{a_n, \{p_n, b_n\}\} \) go to 0, then \( \mu_{p_n (a_n + b_n)(a_n + b_n)^* p_n} \) converges in distribution and the limit is uniquely identified by the equation

\[
\lim_{n \to \infty} \mu_{p_n (a_n + b_n)(a_n + b_n)^* p_n} \boxtimes \mu_c = \mu_{pa a^* p} \boxtimes \mu_c \boxplus \mu_c = \mu_p \boxtimes \mu_c.
\]

**Proof.** The limiting moments of \( \mu_{p_n (a_n + b_n)(a_n + b_n)^* p_n} \boxtimes \mu_c \) do not change if we "zero out" the mentioned mixed moments (i.e. that we assume freeness
of \((a_n, \{p_n, b_n\})\), due to the assumption on their vanishing and of uniform boundedness on moments. It is also easily seen that the limiting moments do not change if we change the distribution of \(a_n\) to \(\mu_{a_n} = \mu_a\) for all \(n\). But then
\[
\lim_{n \to \infty} \mu_{p_n(a_n+b_n)(a_n+b_n)^* p_n \boxplus \mu_c} = \lim_{n \to \infty} \mu_{p_n(a+b_n)(a+b_n)^* p_n \boxplus \mu_c}
\]
\[
= \lim_{n \to \infty} (\mu_{p_{na} a^* p_n \boxplus \mu_c}) \boxplus (\mu_{p_{nb} b_n^* p_n \boxplus \mu_c}) = (\mu_{p_{na} a^* p \boxplus \mu_c}) \boxplus (\mu_{p \boxplus \mu_c})
\]
where we used theorem 3.4, so that (3.6) holds. \(\square\)

The rest of the proof of theorem 4.1 now goes as follows: The rectangular random matrices \(R_n\) can be viewed as the \(N \times N\) random matrices \(p_n S_n\), where the projection \(p_n\) is a diagonal constant matrix, with the fraction of 1’s on the diagonal equal to \(c\), and \(S_n\) is an extension of the \(n \times N\) matrix \(R_n\) to an \(N \times N\)-matrix, obtained by adding zeros. Similarly, the random matrices \(X_n\) can be viewed as the \(N \times N\)-matrices \(p_n Y_n\), where \(Y_n\) is an extension of the \(n \times N\) matrix \(X_n\) to an \(N \times N\)-matrix, obtained by adding more independent standard Gaussian entries.

Since \(\frac{1}{\sqrt{N}} S_n^* S_n\) almost surely converges to a compactly supported probability measure, \((\frac{1}{\sqrt{N}} Y_n, \frac{1}{\sqrt{N}} S_n)\) satisfies the requirements of theorem 3.3. Thus, mixed moments of \(\frac{1}{\sqrt{N}} X_n\) and \(\frac{1}{\sqrt{N}} S_n\) go to zero almost surely. It is also seen that \(\frac{1}{\sqrt{N}} S_n\) has it’s moments bounded as \(n \to \infty\) almost surely. It is well known [7] that \(\frac{1}{\sqrt{N}} Y_n\) converges in distribution almost surely to the circular law, which is \(R\)-diagonal.

Thus, all assumptions of theorem 3.5 are satisfied for \(a_n = \frac{1}{\sqrt{N}} Y_n\), \(b_n = \frac{1}{\sqrt{N}} S_n\) and \(p_n\), almost surely. Thus, almost surely,
\[
\lim_{n \to \infty} \mu_{p_n \frac{1}{\sqrt{N}} (S_n + \sigma Y_n)(S_n + \sigma Y_n)^* p_n \boxplus \mu_c} = (\mu_{p_{a^2} \frac{1}{\sqrt{N}} YY^* \boxplus \mu_c}) \boxplus (\mu \boxplus \mu_c)
\]
\[
= (\mu_{a^2 I} \boxplus \mu_c \boxplus \mu_c) \boxplus (\mu \boxplus \mu_c) = \mu_{a^2 I} \boxplus (\mu \boxplus \mu_c),
\]
or
\[
\lim_{n \to \infty} \mu_{\frac{1}{\sqrt{N}}(R_n + \sigma X_n)(R_n + \sigma X_n)^* \boxplus \mu_c} = \mu_{a^2 I} \boxplus (\mu \boxplus \mu_c),
\]
which is the statement of theorem 4.1.

The proof as sketched here assumes \(c \leq 1\). An explanation for how the proof goes for \(c > 1\) is given succeeding the proof of theorem 3.4.

imsart-aap ver. 2007/01/24 file: multfreeconv.tex date: January 26, 2022
3.1. The proof of theorem 3.1. The proof will use the (generalized) Hölder inequality:

**Lemma 3.1.** For matrices $A_1, \ldots, A_k$, the following holds:

$$\| A_1 \cdots A_k \|_p \leq \| A_1 \|_{p_1} \cdots \| A_k \|_{p_k} \text{ when } \sum_{i=1}^k \frac{1}{p_i} = \frac{1}{p}. $$

In the proof of lemma 4.3.2 in [7], (3.1) is written as

$$(\frac{1}{n})^2 \sum_{i_1, \ldots, i_{2k}=1}^n \sum_{h=1}^k \left( \prod_{r=1}^l d_{j(h(r), j(h(r)+1)}(t_r, n) \right)$$

(3.7) $$\times \left( \prod_{r=l+1}^{2l} \bar{d}_{i(h(r), j(h(r)+1)}(t_r, n) \right) \mathbb{E} \left( \prod_{h=1}^{2k} u_{i(h), j(h)}(s(h), \varepsilon(h), n) \right),$$

where for $1 \leq r \leq l$,

$$k(r) = |m_1| + \cdots + |m_r|,$$

$$k = k(l), k(l + r) = k + k(r), t_{l+r} = t_r.$$

Moreover, for $h$ such that $k(r - 1) + 1 \leq h \leq k(r)$,

$$s(h) = s_r, \varepsilon(h) = \begin{cases} 1 & \text{if } m_r > 0 \\ -1 & \text{if } m_r < 0 \end{cases}$$

Here $s(h + k) = s(h)$ and $\varepsilon(k + h) = -\varepsilon(h)$ for $1 \leq h \leq k$, and

$$u_{ij}(s, \varepsilon, n) = \begin{cases} U_{ij}(s, n) & \text{if } \varepsilon = 1 \\ \bar{U}_{ji}(s, n) & \text{if } \varepsilon = -1 \end{cases}$$

Since (3.7) is a matrix product written out, the following must hold:

(3.8) $$\begin{cases} j_h = i_{h+1} & \text{for } h \in \{1, \ldots, k\} \setminus \{k(1), \ldots, k(l)\} \\ i_h = j_{h+1} & \text{for } h \in \{k+1, \ldots, 2k\} \setminus \{k(l+1), \ldots, k(2l)\} \end{cases}$$

Also, due to the vanishing of many mixed moments of entries in standard unitary random matrices (lemma 4.2.2 in [7]), two pair partitions $\mathcal{U}$ and $\mathcal{V}$ can be chosen so that if $\{h, h'\} \in \mathcal{U}$ then

(3.9) $$s(h) = s(h'), \varepsilon(h) = 1, \varepsilon(h') = -1, i_h = j_{h'}(= i_{h'} + 1),$$
and if \( \{h, h'\} \in \mathcal{V} \) then

\[
\begin{align*}
\text{(3.10)} & \quad s(h) = s(h'), \; \varepsilon(h) = -1, \; \varepsilon(h') = 1, \; i_h = j_{h'} = i_{h'} + 1
\end{align*}
\]

These two pair partitions and (3.8) cause many equalities among the \( i_1, \ldots, i_{2k} \), and define the equivalence relation \( \mathcal{R}(U, \mathcal{V}) \) on \( \{1, \ldots, 2k\} \) so that \( i_h = i_{h'} \) whenever \( h \) and \( h' \) are in the same equivalence class of \( \mathcal{R}(U, \mathcal{V}) \). We let \( k_0 \) denote the number of equivalence classes of \( \mathcal{R}(U, \mathcal{V}) \), and let \( h(1), \ldots, h(k_0) \) be representatives from the equivalence classes.

Recall the expressions

\[
\begin{align*}
\text{(3.11)} & \quad C_n(t_1, \ldots, t_{k_0}) = \left( \prod_{r=1}^{l} d_j, j_{k(r)} + 1(t_r, n) \right) \left( \prod_{r=l+1}^{2l} \bar{d}_j, j_{k(r)} + 1(t_r, n) \right), \\
\text{(3.12)} & \quad Q_n(t_1, \ldots, t_{k_0}) = E \left( \prod_{h=1}^{2l} u_{i_h, j_h} \left( s(h), \varepsilon(h), n \right) \right)
\end{align*}
\]

from [7], where \((t_1, \ldots, t_{k_0})\) in (3.11) are defined as \((i_{h(1)}, \ldots, i_{h(k_0)})\) (i.e. representatives of the equivalence classes), and \( j_1, \ldots, j_{2k} \) are determined subject to \( \mathcal{U}, \mathcal{V} \). (4.3.6) of [7] says that it is enough to prove that for any partition \( \mathcal{W} \) of \( \{1, \ldots, k_0\} \) we have that

\[
\sum_{(t_1, \ldots, t_{k_0}) : \mathcal{W}} C_n(t_1, \ldots, t_{k_0}) Q_n(t_1, \ldots, t_{k_0}) = O(1) \text{ as } n \to \infty,
\]

where the summation is over \((t_1, \ldots, t_{k_0})\) such that \( t_p = t_q \) if and only if \( p \) and \( q \) are in the same block of \( \mathcal{W} \). For a given \( \mathcal{W} \), it is known that \(|Q_n(t_1, \ldots, t_{k_0})| = O(n^{-k})\) uniformly for \( i_k, j_k \) as \( n \to \infty \), and that it has the same value for all \((t_1, \ldots, t_{k_0})\) taking part in the sum (3.13). So, from (3.14) we deduce that it is enough to show that for any choice of \( \mathcal{U}, \mathcal{V}, \mathcal{W} \) (there is a finite number of such choices),

\[
\sum_{(t_1, \ldots, t_{k_0}) : \mathcal{W}} C_n(t_1, \ldots, t_{k_0})
\]

is \( O(n^k) \). In [7] this is proved using the fact that (3.11) are bounded uniformly. This is not true in our case since only uniform boundedness in \( \| \cdot \|_p \) is assumed. Instead, we will group sums of terms into matrix multiplication units, and use the Hölder inequality together with the \( \| \cdot \|_p \)-norm bounds. Instead of the terms in (3.14), where the sum is over

\((t_1, \ldots, t_{k_0}) : t_p = t_q \) if and only if \( p \) and \( q \) are in the same block of \( \mathcal{W} \),
it will be better for us to sum over

\((t_1, ..., t_{k_0}): t_p = t_q \text{ if } p \text{ and } q \text{ are in the same block of } \mathcal{W}\).

The latter set is more compatible with indices in multiplications of many matrices. This second set is larger than the first, and can be written as

\[
\sum_{\mathcal{W}' \geq \mathcal{W}(t_1, ..., t_{k_0})} \sum C_n(t_1, ..., t_{k_0})
\]

It is obvious that (3.14) can be written

\[
\sum_{\mathcal{W}' \geq \mathcal{W}} a_{\mathcal{W}'} \sum_{\mathcal{W}'' \geq \mathcal{W}'} \sum C_n(t_1, ..., t_{k_0}),
\]

where \(a_{\mathcal{W}'}\) are integer constants which can easily be calculated (proving (3.16) boils down to splitting all values of \(t_1, t_2\) into those where \(t_1 = t_2\), and those where \(t_1 \neq t_2\). This is done recursively and for all \(t_i\) to yield (3.16). Since there is a finite number of elements in the two outer sums in (3.16), to prove that (3.14) is \(O(n^k)\) it is enough to show that (3.15) is \(O(n^k)\) for any choice of \(\mathcal{W}\).

Let \(l_0\) denote the number of equivalence classes in \(\mathcal{R}(\mathcal{U}, \mathcal{V})\) with only one entry, and let \(h(1), ..., h(l_0)\) be the corresponding representatives. According to [7], equivalence classes with only one entry give rise to factors of the form \(\tilde{d}_i\) in (3.11), where \(\tilde{d}\) is either \(d_{i,i}, (t_r, n)\) or \(d_{i,i}(t_r, n)\) for some \(r\). Equivalence classes with only one entry thus leads (through summation over one \(t_i\) appearing in just one factor) to factors in (3.15) which are (non-normalized) traces of the \(D(t_r, n)\). These are zero, so we can assume that \(l_0 = 0\) when we attempt to bound (3.15). Had we used the sum (3.14) instead of (3.15), we would not obtain zero.

So we assume that there are no singleton equivalence classes, i.e. \(k_0 \leq k\). Let \(K_0\) be the number of equivalence classes actually appearing in (3.11) (this is a function of \(\mathcal{U}\) and \(\mathcal{V}\)). we have that \(K_0 \leq k_0\), but equality does not necessarily hold. We will use matrix units \(E_{ij}\) (i.e. \(E_{ij}(i, j) = \delta_{ij}\)). By placing matrix units \(F_i\) with indices from \(t_1, ..., t_{k_0}\) in between the terms in (3.11), (3.15) can be written as

\[
\sum_{\mathcal{W}' \geq \mathcal{W}(t_1, ..., t_{k_0})} \sum ntr_n \left( \prod_{i=1}^{2d} (F_i D_i) \right),
\]

where \(D_i\) are matrices from \(D(t_r, n)\) or one of their transposes/conjugates. Since \(\|E_{ij}\|_p = n^{-\frac{1}{p}}\) and the number of possible choices of matrix units is
\( n^{K_0} \), lemma 3.1 implies that (3.17) is bounded by

\[
 n^{K_0+1} \sum_{i=1}^{2l} \prod_{i=1}^{2l} \| D_i \|_4 = n^{K_0+\frac{1}{2}} \prod_{i=1}^{2l} \| D_i \|_4.
\]

Since \( K_0 \leq k \), this is \( O(n^k) \) except possibly in the case when \( K_0 = k \), i.e. when all equivalence classes have exactly two elements.

So, for the rest of the proof, we assume that all equivalence classes have exactly two elements. Note that the number of times an equivalence class appears as an \( i \) is equal to the number of the times the same class appears as a \( j \) in (3.11). This is obvious from the way the equivalence relation is defined (3.8), (3.9), (3.10) in order to avoid a zero value in (3.12). This means that we can take the first of the \( K_0 \) equivalence classes appearing in (3.11), and rearrange the terms in (3.11) so that the equivalence class appear in alternating order as an \( i \) and as a \( j \). (3.17) can thus be rewritten to

\[
\sum_{W' \geq W_{(i_1, \ldots, i_{k_0})}; W'} ntr_\alpha (F_1 D_1 G_1 D_2 F_2 D_3 G_2 D_4 F_3),
\]

where \( D_1, D_2, D_3, D_4 \) are the matrices where the first equivalence class appear as an \( i \) or a \( j \), and in alternating order. Also, \( F_1 = F_2 = F_3 = E_r \) are matrix units, \( r \) is a given number, and the \( G_i \) are products man of the matrices \( D_i \) in (3.11). (3.19) can also be written

\[
\sum_{W' \geq W_{(i_1, \ldots, i_{k_0})}; W'} ntr_\alpha (diag(D_1 G_1 D_2) diag(D_3 G_2 D_4)),
\]

where \( diag(A) \) stands for the diagonal of the matrix \( A \). Similarly to the calculation of the bound (3.18), (3.20) is seen to be bounded by

\[
n^{K_0} \| diag(D_1 G_1 D_2) \|_2 \| diag(D_3 G_2 D_4) \|_2.
\]

Note that \( \| diag(A) \|_2 \leq \| A \|_2 \), since

\[
\| diag(A) \|_2^2 = \frac{1}{n} \sum_i |a_{ii}|^2 \leq \frac{1}{n} \sum_{i,j} |a_{ij}|^2 = tr_\alpha(A^* A) = \| A \|_2^2,
\]

where \( A = (a_{ij})_{i,j} \). This means that (3.21) is bounded by

\[
n^{K_0} \| D_1 G_1 D_2 \|_2 \| D_3 G_2 D_4 \|_2,
\]

which is \( O(n^{K_0}) \) and hence \( O(n^k) \) since all \( D_i \) are bounded in \( p \)-norm, and the only other factors are matrix units, which have \( p \)-norm \( n^{-\frac{1}{p}} \).

That there exists a \( p_l \) for a given \( l \) as in the last statement of the theorem is obvious from the proof and the way the Hölder inequality was used. This completes the proof.
3.2. The proofs of theorem 3.2 and 3.3. First assume that $R_n$ satisfies $\|R_n\|_p \leq R_p$ ($p \geq 1$) almost surely for some constants $R_p$. $P_1(R_n)$ satisfies similar $\| \cdot \|_p$-norm bounds due to lemma 3.1. Call the underlying probability space $\Omega$. Denote by $\Omega^R$ claimed. Since $\bigcup_n \mathbb{R}$ is compactly supported probability measure, observe enough [7]. For each $R$ for such random matrices $\lim_{n \to \infty} \|P_1(R_n)\|_p \to R$.

Now for theorem 3.3. Write

$$X_n = U_n \Lambda_n U_n^*$$

(3.22)

$$E \left( |\text{tr}_n(U_m^1 P_1(R_n) U_m^2 P_2(R_n) \cdots U_m^1 P_1(R_n))|^2 \right)$$

$$= \int_{\Omega} |\text{tr}_n(U_m^1 P_1(R_n) \cdots U_m^1 P_1(R_n))|^2 ds$$

$$= \int_{M_n(C)} \int_{M_n(C)} |\text{tr}_n(U_m^1 P_1(R) \cdots U_m^1 P_1(R))|^2 f_{U_n, R_n}(U, R) dU dR$$

$$\leq \int_{M_n(C)} Cn^{-2} f_{R_n}(R) dR = Cn^{-2}$$

where we have used the bounds for deterministic matrices from theorem 3.1. Therefore

$$|\text{tr}_n(U_m^1 P_1(R_n) U_m^2 P_2(R_n) \cdots U_m^1 P_1(R_n))| \to 0 \text{ a.s.}$$

for such random matrices $R_n$. If $R_n R_n^*$ is just known to converge in distribution almost surely to a compactly supported probability measure, observe that almost surely there exists a value $R$ so that $\|R_n\|_p \leq R$ for $n$ large enough [8]. For each $l$, choose $p_l$ as in the statement of theorem 3.1. Denote by $\Omega_{p_l, N}$, $p \geq 1$ $N \in \mathbb{N}$ the subset of $\Omega$ determined by values $s$ such that

(3.23) $\|R_n(s)\|_{p_l} \leq N$

for all $n$. Define $R_{n,p_l, N} = \chi_{\Omega_{p_l, N}} R_n$ with $\chi$ denoting the characteristic function. The $R_{n, p_l, N}$ satisfy the estimates (3.22) for mixed moments of length $\leq l$, so that these mixed moments go to zero almost surely in $\Omega_{p_l, N}$. $\cup_{p_l, N} \Omega_{p_l, N}$ has probability 1: Almost surely, the $\| \cdot \|_{p_l}$-norm of $R_n$ stays bounded by some finite value for large enough $n$. Thus, for every $s$ in a set with probability one, we can find a value $N_s$ such that $\|R_n(s)\|_{p_l} \leq N_s$ for ALL $n$. But then $s \in \Omega_{p_l, N_s}$, so that $\cup_{p_l, N} \Omega_{p_l, N}$ has probability 1 as claimed. Since $\cup_{p_l, N} \Omega_{p_l, N}$ has probability 1, theorem 3.2 follows from the fact that $R_n = R_{n, p_l, N}$ on $\Omega_{p_l, N}$. By increasing $l$ we get almost sure convergence of higher mixed moments to zero also.

Now for theorem 3.3 Write
for a unitary random matrix $U_n$, and diagonal random matrix $\Lambda_n$. We may assume that $U_n$ is a standard unitary random matrix, as in the proof of theorem 4.3.5 of [2], since Gaussian random matrices are unitarily invariant. We can also assume that $\Lambda_n$ is independent from $U_n$, so that $(U_n, \{\Lambda_n, R_n\})$ is an independent family. $R_n$ converges to a limit which is compactly supported, and $\Lambda_n$ does the same. Since (3.22) can be easily generalized to the case where the $R_n$ are replaced with many different $R_n$ (with the $\{R_n\}$ all independent from $U_n$) we conclude also for theorem (3.3) that we get almost sure convergence to zero of mixed moments as in definition (2.1).

### 3.3. The proof of theorem (3.4)

First write $\phi((p(a+b)(a+b)^*p)^m)$ as a sum of mixed moments of length $3m$ by multiplying out $(p(a+b)(a+b)^*p)^m$:

\[
\phi((p(a+b)(a+b)^*p)^m) = \sum_{\sigma_1 \leq \sigma} \phi(x_1 x_2^* \cdots x_{2m-1} x_{2m}^*) ,
\]

where $\sigma = \{1, 2, 4, 5, \cdots 3m - 2, 3m - 1\}$ $(\{1, 2, 4, 5, \cdots 3m - 2, 3m - 1\}$ correspond to the indices of the locations of the $x_i, x_i^*$ in the moments $x_1 x_2^* \cdots x_{2m-1} x_{2m}^*$), $\sigma_1$ runs over all subsets of $\sigma$, and $x_i = a$ if $i \in \sigma_1$, $x_i = b$ if $i \in \sigma \setminus \sigma_1$. Denote by $|\sigma_1|$ the cardinality of $\sigma_1$. We denote by $\alpha$ the cumulants of $\mu_{a,a^*}$ and $\beta$ the cumulants of $\mu_{b,b^*p}$, so that the moment-cumulant formula for $(a,a^*)$ is

\[
\phi(x_1 \cdots x_n) = \sum_{\pi = \{B_1, \cdots, B_k\} \in NC(n)} \prod_{i=1}^k \\text{coef}(i_1, \ldots, i_n)|B_i|(R_{\mu_{a,a^*}})
\]

with $x_1 = a$, $x_2 = a^*$ and $i = 1$ or $2$, and the moment-cumulant formula for $(b,b^*,p)$ is

\[
\phi(x_1 \cdots x_n) = \sum_{\pi = \{B_1, \cdots, B_k\} \in NC(n)} \prod_{i=1}^k \\text{coef}(i_1, \ldots, i_n)|B_i|(R_{\mu_{b,b^*p}})
\]

with $x_1 = b$, $x_2 = b^*$, $x_3 = p$ and $i = 1, 2$ or $3$. We will use the shorthand notation

\[
\alpha_{B_i} = \text{coef}(i_1, \ldots, i_n)|B_i|(R_{\mu_{a,a^*}})
\]

\[
\beta_{B_i} = \text{coef}(i_1, \ldots, i_n)|B_i|(R_{\mu_{b,b^*p}})
\]

Due to the freeness of $a$ and $\{p,b\}$, the moment-cumulant formula applied to all moments in (3.24) and (2.14) yields

\[
\sum_{\sigma_1 \leq \sigma} \sum_{\sigma_2 \in NC(\sigma_1)} \sum_{\pi_1 \leq \sigma_1} \sum_{\pi_2 \in NC(\sigma_2)} \alpha_{\pi_1} \beta_{\pi_2},
\]

$p, b, b^*, p$
where $\sigma_1^c = \{1, ..., 3m\} \setminus \sigma_1$ and $|\sigma_1^c|$ is the cardinality of $\sigma_1^c$. $\pi_1$ divides
$\{1, ..., |\sigma_1^c|\}$ into $|K(\pi_1)|$ sets (see definition 2.3 and figure 1) when $\pi_1$ is
viewed as an element in $NC(|\sigma_1|)$. $\pi_1$ also divides $\{1, ..., 3m\}$ into the same
number of sets, according to the circular representation of $\{1, ..., 3m\}$. Let
us denote these blocks by $B_1, \cdots, B_k$, so that $\sigma_1^c = \{B_1, \cdots, B_k\}$ as a sub-
partition of $1_{|\sigma_1|}$. Since $\pi_1$ and $\pi_2$ have no crossings if and only if
$\pi_2 \leq \{B_1, \cdots, B_k\}$, (3.27) can be written as

\[
(3.28) \quad \sum_{\sigma_1 \leq \sigma_1^c} \sum_{\pi_1 \in NC(|\sigma_1|)} \sum_{\pi_2 \leq \{B_1, \cdots, B_k\}} \alpha_{\pi_1} \beta_{\pi_2}
\]

When $\pi_2 \leq \{B_1, \cdots, B_k\}$ we can write $B_i = \pi_2i_1 \cup \pi_2i_2, \cdots$, where the $\pi_2i_j$
are the reindexed blocks of $\pi_2$ which are contained in $B_i$, and where $\pi_2i = \{\pi_2i_1, \pi_2i_2, \cdots\}$. This is in $NC(|B_i|)$ since $\pi_2$ is noncrossing. First rewrite
(3.28) to

\[
(3.29) \quad \sum_{\sigma_1 \leq \sigma_1^c} \sum_{\pi_1 \in NC(|\sigma_1|)} \sum_{\pi_2i \in NC(|B_i|)} \alpha_{\pi_1} \left( \sum_{\pi_2i_1 \in NC(|B_i|)} \prod_{i=1}^k \beta_{\pi_2i_1} \right).
\]

Then note that the $\pi_2i \in NC(|B_i|)$ can be summed independently of one
another, so that we can rewrite to

\[
(3.30) \quad \sum_{\sigma_1 \leq \sigma_1^c} \sum_{\pi_1 \in NC(|\sigma_1|)} \alpha_{\pi_1} \prod_{i=1}^k \left( \sum_{\pi_2i \in NC(|B_i|)} \beta_{\pi_2i} \right).
\]

Note also that only $\pi_1$ with blocks

$$C_k = \{c_{k1}, ..., c_{kr}\}$$

where $x_{i,c_{ki}}$, $\cdots$, $x_{i,c_{kr}}$ are alternating values of $a$ and $a^*$, give contribution
in (3.30), due to $R$-diagonality of $a$. Hold such a $\pi_1$ fixed in (3.30), and
take a look at the inner sum in (3.30) for a given $i$. This is simply the
moment-cumulant formula (3.26) for a moment of length $|B_i|$, where the
mixed moment is on the form

$$p_{bb}^* p_{bb}^* \cdots bb^* p,$$

or on the form

$$b^* p_{bb}^* p_{bb}^* \cdots bb^* pb$$
due to the alternating structure in \[^{[3.21]}\]. In both cases the moment-cumulant formula yields \( \phi \left( (pbb^* p)^{\left| \frac{B_1}{x} \right|} \right) \) for the inner sum in \(^{[3.30]}\). Therefore, we get that \(^{[3.30]}\) equals

\[
(3.31) \quad \sum_{\sigma_1 \leq \sigma} \sum_{\pi_1 \in \mathcal{NC}(|\sigma_1|)} \alpha_{\pi_1} \prod_{i=1}^{k} \phi \left( (pbb^* p)^{\left| \frac{B_i}{x} \right|} \right).
\]

Since \( a \) is \( R \)-diagonal, the \( \alpha_{\pi_1} \) which give contribution in \(^{[3.31]}\) are uniquely identified by the moments \( \phi((aa^*)^m) \). Therefore, the moments

\[
\phi \left( (p(a + b)(a + b)^* p)^m \right)
\]

are entirely identified by the moments \( \phi((aa^*)^m) \) and \( \phi((pbb^* p)^m) \), so that \( \mu_{p(a+b)(a+b)^* p} \) only depends on \( \mu_{aa^*} \) and \( \mu_{pbb^* p} \). All distributions are here in the noncommutative probability space \( (A, \phi) \), not yet in the reduced space \( (pAp, \phi(p)^{-1}\phi) \). If we can prove the theorem when \( p \) and \( b \) are free, it will also hold when \( p \) and \( b \) are not free since \( \mu_{p(a+b)(a+b)^* p} \) only depends on \( \mu_{aa^*} \) and \( \mu_{pbb^* p} \).

We can replace \( \mu_{bb^*} \) with the (unique) \( R \)-diagonal pair \( b_0 \) so that \( \mu_{bb^*} = \mu_{b_0 b_0^*} \). So, we assume that both \( a \) and \( b \) give rise to free \( R \)-diagonal pairs. Their determining series are \( R_{\mu_{aa^*}} \mathbb{M} \text{Meob} \) and \( R_{\mu_{bb^*}} \mathbb{M} \text{Meob} \), respectively. \( (a+b, (a+b)^*) \) is also an \( R \)-diagonal pair, with determining series \( R_{\mu_{aa^*}} \mathbb{M} \text{Meob} + R_{\mu_{bb^*}} \mathbb{M} \text{Meob} \). This means that

\[
(3.32) \quad R_{\mu_{(a+b)(a+b)^*}} \mathbb{M} \text{Meob} = R_{\mu_{aa^*}} \mathbb{M} \text{Meob} + R_{\mu_{bb^*}} \mathbb{M} \text{Meob}.
\]

If \( x \) is free from \( p \), we next calculate \( R_{\mu_{pxx^* p}} \) in the reduced space \( (pAp, \phi(p)^{-1}\phi) \). We call this \( R_{\mu_{pxx^* p}}^{p\text{Ap}} \) in the rest of the proof, with similar notation for the moment series (In the rest of the paper, this notation is dropped since \( pxx^* p \) is assumed to be in \( (\mu_{pAp}, (\phi(p)^{-1}\phi)) \)). Note that \( M_{\mu_{pxx^* p}}^{p\text{Ap}} = \frac{1}{c} \mathcal{M}_{\mu_{pxx^* p}} \). We have

\[
R_{\mu_{pxx^* p}}^{p\text{Ap}} = M_{\mu_{pxx^* p}}^{p\text{Ap}} \mathbb{M} \text{Meob} = \left( \frac{1}{c} \mathcal{M}_{\mu_{pxx^* p}} \right) \mathbb{M} \text{Meob} = \left( \frac{1}{c} \left( R_{\mu_{xx^*}} \mathbb{C} \text{Zeta} \right) \right) \mathbb{M} \text{Meob} = \left( \frac{1}{c} \left( c \mathbb{C} \text{Meob} \right) \right) = \left( \frac{1}{c} \left( c \mathbb{C} \text{Meob} \right) \right) = c^{-n-1} \mathbb{C} \mu_{xx^*} R_{\mu_{xx^*}}.
\]

For a general Marchenko Pastur law \( \mu_d \), \( R_{\mu_d} = d^n \mathbb{C} \text{Zeta} \), and it is easily verified that \( R_{\mu_d}^{-1} = d^n \mathbb{C} \text{Meob} \). We have that

\[
R_{\mu_{pxx^* p}}^{p\text{Ap}} \mathbb{M} \text{Meob} \mu_d^{-1} = c^n d^n \mathbb{C} \mu_{pxx^*}^{-1} \mathbb{C} \mu_{pxx^*} \mathbb{C} d^{-1} \mathbb{C} \text{Meob} = c^n d^n \left( \left( c^{-1} R_{\mu_{xx^*}} \right) \mathbb{C} \left( d^{-1} \mathbb{C} \text{Meob} \right) \right)
\]
if \( c = d \), this can be simplified to

\[
(3.33) \quad c^n c^{-n-1} (R_{\mu_{xx}} \boxplus \text{Moeb}) = c^{n-1} (R_{\mu_{xx}} \boxplus \text{Moeb}) .
\]

Using this for \( x = a \), \( x = b \) and \( x = a + b \), and also using (3.32), we get

\[
R_{\mu_{p(a+b)(a+b)^*}}^{\text{Ap}} \boxplus \text{Moeb}^{-1} = c^{n-1} \left( R_{\mu_{p(a+b)(a+b)^*}} \boxplus \text{Moeb} \right)
= c^{n-1} \left( R_{\mu_{pa} \boxplus \text{Moeb}} + R_{\mu_{bb}} \boxplus \text{Moeb} \right)
= c^{n-1} \left( R_{\mu_{pa} \boxplus \text{Moeb}} + c^{n-1} (R_{\mu_{ab}} \boxplus \text{Moeb}) \right)
= R_{\mu_{paa} \boxplus \text{Moeb}} \boxplus \text{Moeb}^{-1} + R_{\mu_{pbb} \boxplus \text{Moeb}} \boxplus \text{Moeb}^{-1}
\]

Using (2.17), this can be written

\[
R_{\mu_{p(a+b)(a+b)^*}} \boxplus \text{Moeb}^{-1} = R_{\mu_{paa} \boxplus \text{Moeb}} + R_{\mu_{pbb} \boxplus \text{Moeb}}
\]

which can equivalently be stated as

\[
\left( \mu_{p(a+b)(a+b)^*} \boxplus \text{Moeb} \right) = (\mu_{paa} \boxplus \text{Moeb}) \boxplus (\mu_{pbb} \boxplus \text{Moeb})
\]

which is what we had to prove.

Note that if \( c \neq d \), \( R_{\mu_{xx}} \boxplus \text{Moeb} \) does not appear as a factor in (3.33). Therefore, there is no reason why the result should hold for other Marchenko Pastur laws than \( \mu_c \), since (3.32) cannot be used in such cases.

Although the proof of theorem 1.1 is described for \( c \leq 1 \) only, the methods used in the proof of theorem 3.4 here can help us prove the case for \( c > 1 \) also. If \( R_n \) and \( X_n \) are the random matrices from theorem 1.1 and \( c > 1 \), note that

\[
M_{\mu_{p(a+b)(a+b)^*}} \boxplus \text{Moeb} = \frac{1}{c} M_{\mu_{paa} \boxplus R_n R_n} = \frac{1}{c} M_{\mu_{pbb} \boxplus R_n R_n} = c^{-m-1} M_{\mu_{paa} \boxplus R_n R_n},
\]

where \( c^m f \) denoted the power series defined by \( [\text{coef}_k](c^m f) = c^k[\text{coef}_k](f) \). From this one can show that

\[
(3.34) \quad R_{\mu_{p(a+b)(a+b)^*}} \boxplus \text{Moeb}^{-1} = M_{\mu_{paa} \boxplus R_n R_n} \boxplus \text{Moeb}(c^m \text{Moeb})
= \left( c^{-m-1} M_{\mu_{paa} \boxplus R_n R_n} \boxplus \text{Moeb}(c^m \text{Moeb}) \right)
= \left( c^{-m-1} \left( (c^{-m}) M_{\mu_{paa} \boxplus R_n R_n} \boxplus (c^m \text{Moeb}) \right) \right) \boxplus \text{Moeb}
= \left( c^{-m-1} \left( M_{\mu_{paa} \boxplus R_n R_n} \boxplus \text{Moeb}(c^{-m-1} \text{Moeb}) \right) \right)
= c^{-m-1} \left( M_{\mu_{paa} \boxplus R_n R_n} \boxplus \text{Moeb}(c^{-m-1} \text{Moeb}) \right)
= c^{-m-1} \left( R_{\mu_{paa} \boxplus R_n R_n} \boxplus R_{\mu_{paa} \boxplus R_n R_n}^{-1} \right).
\]
Since theorem 1.1 has been proved for \( c \leq 1 \), \( R_{\mu_{\frac{1}{n}X_n}}(R_n + X_n)^* (R_n + X_n) R_{\mu_{\frac{1}{n}}}^{-1} \) will converge to
\[
R_{\mu_{\frac{1}{n}R_n X_n}} R_{\mu_{\frac{1}{n}}}^{-1} + R_{\mu_{\frac{1}{n}X_n X_n}} R_{\mu_{\frac{1}{n}}}^{-1}
\]
as \( n \to \infty \) (\( c > 1 \)), and the result for \( c > 1 \) follows from (3.34).

4. Equivalence with known expressions for limit distributions of Information-Plus-Noise Type Matrices. [14] studies systems where the sample covariance matrix is formed by taking independent samples of a system of the form
\[
y_n = A_n x_n + \sigma w_n
\]
(the noise factor \( \sigma \) does not appear in [14]) where \( x_n \) and \( w_n \) are independent standard (zero mean, unit variance) Gaussian random vectors, and \( A_n \) is an \( n \times L \) matrix. The covariance matrix of the system is \( \Theta_n = A_n A_n^* + \sigma^2 I \). In particular, when there is no noise (i.e. \( \sigma = 0 \)), the covariance is \( \Theta_n = A_n A_n^* \).

Denote by \( \mu_\Theta \) the limiting eigenvalue distribution of \( A_n A_n^* \). [14] states that the limiting eigenvalue distribution of the sample covariance matrix of the system is \( (\mu_\Theta \boxplus \mu_{\sigma^2 I}) \boxtimes \mu_c \). When there is no noise, the limit is \( \mu_\Theta \boxtimes \mu_c \).

This way of passing from \( \mu_\Theta \boxtimes \mu_c \) to \( (\mu_\Theta \boxplus \mu_{\sigma^2 I}) \boxtimes \mu_c \) is of course compatible with theorem 1.1. We will also show that it is equivalent with the results in [3]. The following restrictions taken from [3] will be used:

1. For \( n = 1, 2, \cdots \), \( X_n = (X_{ij}^n, n \times N \), i.d. for all \( i, j, n \), independent across \( i, j \) for each \( n \), and \( E|X_{11}^n|^2 = 1 \)
2. \( R_n \) is \( n \times N \) and independent of \( X_n \), with \( F^{\mu_{WN}} \overset{D}{\to} F^{\mu_W} \)

Theorem 1.1 in [3] expresses a relationship for finding the limiting eigenvalue distribution \( \mu_W \) of \( W_n \) from that of \( \Gamma_n = \frac{1}{N} R_n R_n^* \) (denoted \( \mu_\Gamma \)). More precisely, under the conditions 1) and 2), we have that (in a slightly rewritten form) \( F^{\mu_{WN}} \overset{D}{\to} F^{\mu_W} \) almost surely, where \( F^{\mu_W} \) is a nonrandom p.d.f. characterized by
\[
(4.1) \quad m_{\mu_W}(z) = \int \frac{dF^{\mu_Y}(t)}{1 + \sigma^2 c m_{\mu_W}(z)} - (1 + \sigma^2 c m_{\mu_W}(z)) z + \sigma^2 (1 - c)
\]
for any \( z \in C^+ \). The connection with multiplicative free convolution is hard to see from this formula. To obtain this connection, the following lemma is needed, which will be proved in section 4.1.
Lemma 4.1. \([4.1]\) is equivalent to
\[
(4.2) \quad m_{\mu_W}^{-1} \left( \frac{z}{1 - \sigma^2 cz} \right) = (1 - \sigma^2 cz)^2 m_{\mu_\Gamma}^{-1}(z) + \sigma^2(1 - c)(1 - \sigma^2 cz),
\]
for \(z\) in some interval \((0, z_1)\).

In \([4.2]\) and all other places where the inverse of the Stieltjes transform is taken in this paper, we will mean the unique inverse on the negative real line. The inverse will only be calculated for positive values close to 0. It will turn out that \([4.2]\) can be more conveniently expressed in terms of distributions obtained from multiplicative free deconvolution with the Marchenko Pastur law using the following lemma, which will be proved in section 4.2:

Lemma 4.2. If
\[
(4.3) \quad \mu_\Gamma = \mu_\Theta \boxtimes \mu_c,
\]
then, for \(z\) in some interval \((0, z_1)\),
\[
(4.4) \quad \eta_{\mu_\Gamma}^{-1}(z) = \eta_{\mu_\Theta}^{-1}(z) \frac{1}{1 - c + cz}
\]
and also
\[
(4.5) \quad m_{\mu_\Gamma}^{-1} \left( \frac{z}{1 - c - czm_{\mu_\Theta}^{-1}(z)} \right) = m_{\mu_\Theta}^{-1}(z)(1 - c - czm_{\mu_\Theta}^{-1}(z)).
\]

Using \((4.5)\), the following relationship with multiplicative free convolution will be shown:

Theorem 4.1. Under the conditions 1) and 2), assume that
\[
(4.6) \quad F_{\mu_\Gamma n} \overset{D}{\to} F_{\mu_\Theta \boxtimes \mu_c} \ a.s.
\]
Then
\[
(4.7) \quad F_{\mu_W n} \overset{D}{\to} F_{(\mu_\Theta \boxtimes \mu_{\sigma^2 I}) \boxtimes \mu_c} \ a.s.
\]
Equivalently, assume that
\[
(4.8) \quad F_{\mu_\Gamma n} \overset{D}{\to} F_{\mu_\Gamma} \ a.s.
\]
Then
\[
(4.9) \quad F_{\mu_W n} \overset{D}{\to} F_{\mu_W} \ a.s.
\]
where \(\mu_W\) is uniquely identified by the equation
\[
(4.10) \quad \mu_W \boxplus \mu_c = (\mu_\Gamma \boxplus \mu_c) \boxplus \mu_{\sigma^2 I}.
\]
Theorem 4.1 will be proved in section 4.3.
4.1. The proof of lemma 4.1. Rewritten in terms of the Stieltjes transform, (4.1) says that (with terms somewhat regrouped)

$$\frac{m_{\mu W}}{1+\sigma^2 cm_{\mu W}} = m_{\mu t} \left( (1+\sigma^2 cm_{\mu W})^2 \frac{z}{1+\sigma^2 cz} - \sigma^2 (1-c)(1+\sigma^2 cz) \right),$$

where $m_{\mu W}, m_{\mu t}$ are evaluated in $z$ when the parameter is omitted. We restrict ourselves to $z$ on the negative real line. The relation holds also for such $z$, since we can analytically continue to the negative real line. Evaluating in $m_{\mu t}^{-1}(z)$ we get

$$\frac{z}{1+\sigma^2 cz} = m_{\mu t} \left( (1+\sigma^2 cz)^2 m_{\mu W}^{-1}(z) - \sigma^2 (1-c)(1+\sigma^2 cz) \right)$$

for $z$ in some interval $(0, z_1)$. We will find it convenient to work with the inverse of the Stieltjes transform, so we rewrite the expression to

$$m_{\mu t}^{-1} \left( \frac{z}{1+\sigma^2 cz} \right) = (1+\sigma^2 cz)^2 m_{\mu W}^{-1}(z) - \sigma^2 (1-c)(1+\sigma^2 cz).$$

Substituting $u = \frac{z}{1+\sigma^2 cz}$ (or equivalently $z = \frac{u}{1-\sigma^2 cu}$) (this is an isomorphism of the positive real axis which sends 0 to 0), we get

$$m_{\mu t}^{-1}(z) = m_{\mu W}^{-1} \left( \frac{z}{1-\sigma^2 cz} \right) - \frac{\sigma^2 (1-c)}{1-\sigma^2 cz}$$

for $z$ in some interval $(0, z_1)$, so that

$$m_{\mu W}^{-1} \left( \frac{z}{1-\sigma^2 cz} \right) = (1-\sigma^2 cz)^2 m_{\mu t}^{-1}(z) + \sigma^2 (1-c)(1-\sigma^2 cz),$$

which is (4.2).

4.2. The proof of lemma 4.2. By the multiplicative property of the S-transform we have

$$S_{\mu t}(z) = \frac{S_{\mu \Theta}(z)}{1+cz}.$$ 

Expressed in terms of the $\eta$-transform this can be written

$$\eta_{\mu t}^{-1}(z) = \eta_{\mu \Theta}^{-1}(z) \frac{1}{1-c+cz},$$

which is (4.4). Evaluating in $\eta_{\mu \Theta}(z)$ and applying $\eta_{\mu t}$ on both sides gives

$$\eta_{\mu t} \left( \frac{z}{1-c+cz} \eta_{\mu \Theta}(z) \right) = \eta_{\mu \Theta}(z)$$
for \( z \geq 0 \). This can also be expressed in terms of Stieltjes transforms as

\[
\frac{m_{\mu W} \left( \frac{1-c+cz_{\mu_\Theta}}{z} \right)}{z} = m_{\mu_\Theta} \left( \frac{1}{z} \right)
\]

Regrouping terms and substituting \(-\frac{1}{z}\) for \( z \) we get

\[
m_{\mu W} \left(z(1 - c - cz_{\mu_\Theta}(z))\right) = \frac{m_{\mu_\Theta}(z)}{1 - c - cz_{\mu_\Theta}(z)}
\]

for \( z \leq 0 \). Substituting \( m_{\mu_\Theta}^{-1}(z) \) for \( z \) and taking the inverse Stieltjes transform \( m_{\mu W}^{-1} \) we get

\[
m_{\mu W}^{-1} \left( \frac{z}{1 - c - cz_{\mu_\Theta}(z)} \right) = m_{\mu_\Theta}^{-1}(z)(1 - c - cz_{\mu_\Theta}(z))
\]

for \( z \) in some interval \((0, z_1)\), which is (4.5).

4.3. The proof of theorem 4.1. Note that if

\[
z = \frac{z_1}{1 - c - cz_{\mu_\Theta}(z_1)}
\]

for \( z_1 \) positive and close to 0, then we have

\[
(4.11) \quad z \geq 1 - c^2 cz = \frac{z_1}{1 - c - cz_{\mu_\Theta}(z_1) + \sigma^2 (1 - c^2 c z)}.
\]

Note also that \( z \geq 1 - \frac{z}{\sigma^2 c z} \geq 0 \) as long as \( z < \frac{1}{\sigma^2 c} \). Substituting (4.11) and (4.5) in (4.2) we get \( m_{\mu W}^{-1} \left( \frac{z}{1 - \sigma^2 cz} \right) = \)

\[
(4.12) \quad \left(1 - \frac{\sigma^2 c z_1}{1 - c - cz_{\mu_\Theta}(z_1)}\right)^2 m_{\mu_\Theta}^{-1}(z_1)(1 - c - cz_{\mu_\Theta}(z_1))
\]

\[ + \sigma^2 (1 - c) \left(1 - \frac{\sigma^2 c z_1}{1 - c - cz_{\mu_\Theta}(z_1)}\right)^2 m_{\mu_\Theta}^{-1}(z_1)(1 - c - cz_{\mu_\Theta}(z_1))
\]

\[ = \left(1 - c - cz_{\mu_\Theta}(z_1)\right)^2 m_{\mu_\Theta}^{-1}(z_1)(1 - c - cz_{\mu_\Theta}(z_1)) + \sigma^2 (1 - c) \left(1 - c - cz_{\mu_\Theta}(z_1)\right)^2 m_{\mu_\Theta}^{-1}(z_1)(1 - c - cz_{\mu_\Theta}(z_1))
\]

\[ = \left(1 - c - cz_{\mu_\Theta}(z_1)\right)^2 m_{\mu_\Theta}^{-1}(z_1)(1 - c - cz_{\mu_\Theta}(z_1)) + \sigma^2 (1 - c) \left(1 - c - cz_{\mu_\Theta}(z_1)\right)^2 m_{\mu_\Theta}^{-1}(z_1)(1 - c - cz_{\mu_\Theta}(z_1))
\]

\[ = \left(m_{\mu_\Theta}^{-1}(z_1) + \sigma^2 \right) \left(1 - c - cz_{\mu_\Theta}(z_1)\right)^2 m_{\mu_\Theta}^{-1}(z_1)(1 - c - cz_{\mu_\Theta}(z_1))
\]

\[ = m_{\mu_\Theta}^{-1}(z_1) + \sigma^2 \left(1 - c - cz_{\mu_\Theta}(z_1)\right)^2 m_{\mu_\Theta}^{-1}(z_1)(1 - c - cz_{\mu_\Theta}(z_1))
\]
Here we have used that $m_{\mu_1 \boxplus \mu_2^{-1}}(z_1) = m_{\mu_1}^{-1}(z_1) + \sigma^2$, which follows from the additivity property of the $R$-transform and the fact that the inverse of the Stieltjes transform is used to define the $R$-transform (perform additive free convolution with $\mu_\sigma 1$). (4.12) is thus nothing else than (4.5) (with $m_{\mu_\sigma}$ replaced by $m_{\mu_1 \boxplus \mu_2^{-1}}$).

Since (4.5) is just an equivalent expression for multiplicative free convolution, we therefore have

$$\mu_W = (\mu_\Theta \boxplus \mu_\sigma 1) \boxtimes \mu_c,$$

or equivalently

$$\mu_W \ominus \mu_c = \mu_\Theta \boxplus \mu_\sigma 1 = (\mu_\Gamma \ominus \mu_c) \boxplus \mu_\sigma 1.$$

This completes the proof.

5. Using $G$-analysis to estimate the spectral function of covariance matrices. It turns out that multiplicative free deconvolution can also be used to estimate covariance matrices. The general statistical analysis of observations, also called $G$-analysis [5] is a mathematical theory for complex systems where the number of parameters of the underlying mathematical model increase together with the growth of the number of observations of the system. The mathematical models which approach the system in some sense are called $G$-estimators. The main difficulty in $G$-analysis is to find good $G$-estimators. $G$-estimators have already shown their usefulness in many applications [9]. We denote by $N$ the number of observations of the system, and by $n$ the number of parameters of the mathematical model. The condition used in $G$-analysis expressing the growth of the number of observations vs. the number of parameters in the mathematical model, is called the $G$-condition. The $G$-condition used throughout this paper is (1.2).

Girko restricts to systems where a number of independent random vector observations are taken, and where the random vectors have identical distributions. If a random vector $r_n$ has length $n$, we will let $\Theta_n$ denote it’s covariance, while $\Gamma_n$ will still denote sample covariance matrices. The $\Gamma_n$ we analyze in this section are more restrictive than in previous sections, since independence across samples is assumed. Girko calls estimators for the Stieltjes transform of covariance matrices $G^2$-estimators. In chapter 2.1 of [4] he introduces the following expression as candidate for a $G^2$-estimator:

$$G^2_n(z) = \frac{\hat{\theta}(z)}{z} m_{\mu_\Gamma_n}(\hat{\theta}(z)),$$

where the function $\hat{\theta}(z)$ is the solution to the equation

$$\hat{\theta}(z) cm_{\mu_\Gamma_n}(\hat{\theta}(z)) - (1 - c) + \frac{\hat{\theta}(z)}{z} = 0.$$
Girko claims that a function \( G_n^2(z) \) satisfying (5.2) and (5.1) is a good approximation for the Stieltjes transform of the covariance matrices \( m_\Theta_n(z) = \text{tr}_n \{ \Theta_n - zI_n \}^{-1} \). More precisely, he shows that when (5.1), (5.2) and the \( G \)-condition (1.2) are fulfilled, under certain conditions there exists a \( c > 0 \) such that
\[
\lim_{n \to \infty} \sup_{0 < c \leq G(z) \leq S} |G_n^2(z) - m_\Theta_n(z)| = 0,
\]
with probability one for every \( S > 0 \) and \( T > 0 \). According to Girko, analytical continuation of \( G_n^2(z) \) can be performed to obtain limits for other \( z \) than the ones in (5.3).

As it turns out, the \( G^2 \)-estimator can equivalently be expressed in terms of multiplicative free convolution:

**Theorem 5.1.** For the \( G^2 \)-estimator given by (5.1), (5.2), the following holds for real \( z < 0 \):
\[
G_n^2(z) = m_{\mu^\Gamma_n \boxtimes \mu_c}.
\]

**Proof.** (5.2) can be rewritten to
\[
-cn_{\mu^R_n} \left( \frac{1}{\hat{\theta}(z)} \right) - (1 - c) + \frac{\hat{\theta}(z)}{z} = 0
\]
\[
\eta_{\mu^R_n}^{-1} \left( \frac{1}{c} \left( \frac{\hat{\theta}(z)}{z} - (1 - c) \right) \right) = -\frac{1}{\hat{\theta}(z)},
\]
which we will write
\[
\frac{1}{\eta_{\mu^R_n}^{-1} \left( \frac{1}{c} \left( \frac{\hat{\theta}(z)}{z} - (1 - c) \right) \right)} = \hat{\theta}(z).
\]
Denote by \( \mu \) the measure with Stieltjes transform \( G_n^2(z) \). (5.1) can be rewritten using the \( \eta \)-transform as
\[
\eta_\mu \left( \frac{-1}{z} \right) = \eta_{\mu^R_n} \left( \frac{-1}{\hat{\theta}(z)} \right).
\]
Since \( \eta_{\mu^R_n} \) and \( \eta_\mu \) are monotone, it is easily seen from this that \( \hat{\theta} \) is monotone since it is a combination of monotone functions. Forming the inverse functions on both sides, and also applying \( \hat{\theta} \), yields
\[
\hat{\theta} \left( \frac{-1}{\eta_{\mu^R_n}^{-1}(z)} \right) = -\frac{1}{\eta_{\mu^R_n}^{-1}(z)}.
\]
Showing $\mu_{R_n} = \mu \otimes \mu_c$ is equivalent to (after rearranging (4.4))

$$-\frac{1}{\eta_\mu^{-1}(z)} = -\frac{1}{(1 - c + cz)\eta_\mu^{-1}(z)}$$

Applying $\hat{\theta}$ on both sides and using (5.6) yields that this is equivalent to

(5.7) $$-\frac{1}{\eta_{\mu_{R_n}}^{-1}(z)} = \hat{\theta} \left(-\frac{1}{(1 - c + cz)\eta_{\mu_{R_n}}^{-1}(z)}\right)$$

Observe now that (5.7) and (5.5) are related in the following way: If we substitute $z = \frac{1}{c} \left(\frac{\hat{\theta}(w)}{w} - (1 - c)\right)$ into (5.7), the argument on the right hand side can be rewritten using (5.5) to

$$-\frac{1}{(1 - c + \frac{\hat{\theta}(z)}{z} - (1 - c)) \left(-\frac{1}{\hat{\theta}(z)}\right)} = z,$$

so that (5.7) is nothing but a restatement of (5.5), at least on values of the form $z = \frac{1}{c} \left(\frac{\hat{\theta}(w)}{w} - (1 - c)\right)$. If these values take on an open set of real values, equality in (5.7) follows for all $z$ by analytic continuation. This happens when $\hat{\theta}(z) \neq k z$ for some constant $k$. If $\hat{\theta}(z) = k z$, then $\eta_{\mu_{R_n}}$ is seen to be constant, which only happens in trivial cases. Thus we have that $\mu_{R_n} = \mu \otimes \mu_c$, and we are done.

Several remarks concerning theorem 5.1 are in place. First of all, the $G^2$-estimator has a much shorter expression in terms of multiplicative free deconvolution, which also places it as an ingredient in theorem 1.1. The theorem is nice to combine with continuity results for free convolution. Voiculescu has proved such results when convergence is in the weak-$\ast$ topology [1]. This enables us in many cases to conclude that

$$\lim_{n \to \infty} G^2_n(z) = \lim_{n \to \infty} m_{\mu_{\Gamma_n} \boxplus \mu_c} = m_{\mu_{\Gamma} \boxplus \mu_c}$$

for some probability measure $\mu_{\Gamma}$. Secondly, [14] expresses the exact same estimator, i.e.

$$\lim_{n \to \infty} \mu_{\Gamma_n} \boxplus \mu_c = \lim_{n \to \infty} \mu_{\Theta_n}$$

in the case of Gaussian systems. Theorem 5.1 can be seen as a way of generalizing from the Gaussian case.
6. Further work. The concept of freeness and free convolution can be extended to unbounded random variables and general probability measures. In [1] it is shown how this can be done in the context of unbounded operator spaces, and certain regularity properties are proved. For instance, if $\mu_n \to \mu$ and $\nu_n \to \nu$ in the weak-$*$ topology with both $\mu \neq \delta_0$ and $\nu \neq \delta_0$, then $\mu_n \boxplus \nu_n \to \mu \boxplus \nu$ in the weak-$*$ topology also. It is possible that applying such extensions together with the methods applied here can extend the results to the same generality as those in [3]. This may be addressed in a future paper.

The $G^2$-estimator is just one of many estimators introduced by Girko. He has estimators for many other quantities also [4], like for the square root and the moments of covariance matrices. Certain of these estimators may also have alternative expressions in terms of free probability constructs.

References.

[1] H. Bercovici and D. V. Voiculescu. Free convolution of measures with unbounded support. Indiana Univ. Math. J., 42(3):733–774, 1993.
[2] J-P Bouchaud and M. Potters. Theory of Financial Risks-From Statistical Physics to Risk Management. Cambridge University Press, Cambridge, 2000.
[3] B. Dozier and J.W. Silverstein. On the empirical distribution of eigenvalues of large dimensional information-plus-noise type matrices. Submitted, 2004. http://www4.ncsu.edu/~jack/infnoise.pdf.
[4] V. L. Girko. Ten years of general statistical analysis. http://general-statistical-analysis.girko.freewebspace.com/chapter14.pdf.
[5] V. L. Girko. Statistical Analysis of Observations of Increasing Dimension. Kluwer Academic Publishers, 1995.
[6] T. Guhr, A. Müller-Groeling, and H.A. Weidenmüller. Random Matrix Theories in Quantum Physics: Common Concepts. Physica Rep., pages 190–, 299 1998.
[7] F. Hiai and D. Petz. The Semicircle Law, Free Random Variables and Entropy. American Mathematical Society, 2000.
[8] M.L. Mehta. Random Matrices. Academic Press, New York, 2nd edition, 1991.
[9] X. Mestre. Designing good estimators for low sample sizes: random matrix theory in array processing applications. In 12th European Signal Processing Conference, (EUSIPCO’2004), Sept. 2004.
[10] A. Nica. R-transforms of free joint distributions, and non-crossing partitions. J. Funct. Anal., 135(2):271–297, 1996.
[11] A. Nica and R. Speicher. On the multiplication of free $n$-tuples of noncommutative random variables. Amer. J. Math., 118(4):799–837, 1996.
[12] A. Nica and R. Speicher. R-diagonal pairs - a common approach to haar unitaries and circular elements. In D. V. Voiculescu, editor, Free Probability Theory, pages 149–188. American Mathematical Society, 1997.
[13] A. Nica and R. Speicher. Lectures on the Combinatorics of Free Probability. Cambridge University Press, 2006.
[14] N.R. Rao and A. Edelman. Free probability, sample covariance matrices and signal processing. ICASSP, pages 1001–1004, 2006.
[15] S. Thorbjørnsen. Mixed moments of voiculescu’s gaussian random matrices. J. Funct. Anal., 176(2):213–246, 2000.
[16] Ø. Ryan. *Computational tools for free convolution*, 2007. http://ifi.uio.no/~oyvindry/freedeconvaapapps/.

[17] Ø. Ryan. Implementation of free deconvolution. *Planned for submission to IEEE Trans. Signal Process.*, 2007. http://www.ifi.uio.no/~oyvindry/freedeconvsigprocessing.pdf.

[18] Ø. Ryan and M. Debbah. Free deconvolution for signal processing applications. *Submitted to IEEE Trans. Inform. Theory*, 2007. arxiv.org/abs/cs.IT/0701025.

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

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

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