数学研究及评论
Mathematical Research with Reviews
Issue 2 (2019) Art.22

● 课题视角 ●

A Simple Introduction to Free Probability Theory and Its Application to Random Matrices

作者： Xiang-Gen Xia (University of Delaware, Newark, DE, USA)
Email: xxia@ee.udel.edu
收稿日期：2019年10月13日

Abstract: Free probability theory started in the 1980s has attracted much attention lately in signal processing and communications areas due to its applications in large size random matrices. However, it involves massive mathematical concepts and notations, and is really hard for a general reader to comprehend. The main goal of this paper is to briefly describe this theory and its application in random matrices as simple as possible so that it is easy to follow. Applying free probability theory, one is able to calculate the distributions of the eigenvalues/singular-values of large size random matrices using only the second order statistics of the matrix entries. One of such applications is the mutual information calculation of a massive multi-input multi-output (MIMO) system.

Keywords: probability theory, free random variables, massive MIMO, random matrices, and semicircular distributions.

1. INTRODUCTION

Free probability theory was started by Voiculescu in the 1980's [1, 2, 3]. It is about calculating moments (or distributions) of non-commutative random variables, such as, random matrices where the matrix entries are classical random variables.

In classical probability theory, random variables are usually real-valued and can be extended to be complex-valued. For convenience, let us say that they are real-valued. Therefore, they are commutative. For example, assume x_1, x_2 are two independent non-zero random variables and E denotes the expectation. Then,

$$E(x_1x_2x_1x_2) = E(x_1^2x_2^2) = E(x_1^2)E(x_2^2) > 0, \tag{1}$$

no matter whether x_1 and/or x_2 have 0 mean or not, which is because x_1 and x_2 are commutative.

However, if x_1 and x_2 are not commutative, then, the property (1) may not hold and two natural questions are as follows. What will happen to (1)? What does the independence mean to non-commutative random variables?

Free probability theory addresses the above two questions. It introduces freeness between non-commutative random variables, which is analogous to the independence between classical commutative random variables. It basically says that although $E(x_1x_2x_1x_2)$ may not be equal to $E(x_1^2x_2^2)$, it is 0 if x_1 and x_2 are free and both have mean 0.

With this freeness, when a large number of free random variables are summed with proper weights, it converges to the classical semicircular distribution. This is the free central limit theorem similar to the classical central limit theorem, where the Gaussian distribution corresponds to semicircular distribution. Note that the eigenvalue distribution of a random matrix with entries of independent Gaussian random variables (for simplicity, the matrix symmetricity is not specified here) goes to semicircular distribution as well when the matrix size goes to infinity. This suggests a connection between free random variables and large size random matrices. Free probability theory says that, it indeed has a strong connection, i.e., random matrices of independent Gaussian random variables become free when the matrix size goes to infinity. In other words, when the size of matrices is large, these matrices are approximately free.

Furthermore, the entries in random matrices can be replaced by free semicircular random variables (called deterministic equivalent). With the replacement, all the joint moments or cumulants of random matrices can be calculated, which may lead to the calculations of the distributions of the eigenvalues of the functions of these random matrices.

This is the reason why free probability theory has attracted much attention in wireless communications and signal processing areas. Massive MIMO systems have been identified as potential candidates in future wireless communications systems, where the number of inputs and/or the number of outputs are large. In massive MIMO systems, their channel matrices are random of large sizes. Therefore, it is natural to apply free probability theory to do some of the difficult calculations, such as, channel capacity [14, 16, 18]. It is particularly interesting when some statistics of a channel matrix of large size, such as, the first two moments (covariances) of the channel coefficients, are known, how we calculate the channel performance without performing Monte Carlo simulations that may be hard to do in practice when the channel matrix size is large, such as, a massive MIMO channel.

The main goal of this tutorial paper is to briefly introduce free probability theory and its application to large size random matrices so that an ordinary researcher in signal processing and communications areas can easily understand.

In the following, we adopt most of the notations in Speicher [4, 5, 6, 7]. All the results described below are from [4, 5, 6, 7] as well. The remainder of this paper is organized as follows. In Section 2, we describe the basics of free random variables and the free central limit theorem without proof. In Section 3, we describe the calculations/relations of joint moments, cumulants, and distributions of multiple free random variables. In Section 4, we describe random matrices and the approximate distributions of their eigenvalues. In Section 5, we describe free deterministic equivalents for random matrices. We also describe how to calculate the Cauchy transforms of random matrices using the second order statistics of their entries. In Section 6, we conclude this paper.

2. Free Random Variables

For convenience, in the following we will use as simple notations as possible, which may be too simplified in terms of mathematical rigorousness.

Let $x_1, x_2, ..., x_n$ be n elements that may not be commutative, and let E be a linear functional on these elements so that $E(1) = 1$. Examples of these elements are matrices and E is like the expectation of a classical random variable.

Definition 2.1. Elements (or random variables) $x_1, x_2, ..., x_n$ are called free or freely independent, if for any m polynomials $p_k(x)$, $1 \leq k \leq m$, with $m \geq 2$,

$$E(p_1(x_{i_1})p_2(x_{i_2})\cdots p_m(x_{i_m})) = 0, \tag{2}$$

when $E(p_k(x_{i_k})) = 0$ for all k, $1 \leq k \leq m$, and any two neighboring indices i_l and i_{l+1} are not equal, i.e., $1 \leq i_1 \neq i_2 \neq \cdots \neq i_m \leq n$.

From (2), if x_1 and x_2 are free, then $E(x_1x_2x_1x_2) = 0$ when $E(x_1) = E(x_2) = 0$, where $m = 4$, $i_1 = 1, i_2 = 2, i_3 = 1, i_4 = 2$, and polynomials $p_k(x) = x$ for $1 \leq k \leq 4$. Comparing with (1) in the classical commutative case, independent real-valued random variables are not free. The terminology "free" comes from the concept of free groups, where there is no any nontrivial relation between any generating elements of a free group.

One might want to ask why, in the above definition, polynomials of the random variables x_k are used. It is for the convenience later in calculating their joint moments. Note that in free probability theory context, it is not convenient to directly define density functions (or distribution functions) for noncommutative random variables. However, as we can recall, in the classical probability theory, if all the moments of a random variable are known, its characteristic function can be often determined and therefore, its density function can be often determined as well. Thus, calculating all the joint moments of free random variables may be sufficient for their joint distributions. Its details will be described in Section 3.

The set $\mathcal{A}_k$ of all polynomials $p(x_k)$ of x_k including the identity element $1 = x_k^0$ is called the subalgebra generated by element x_k for $1 \leq k \leq n$. Subalgebras $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n$ are called free if and only if elements $x_1, x_2, ..., x_n$ are free. Clearly, when elements $x_1, x_2, ..., x_n$ are free, for any n polynomials $p_1(x), ..., p_n(x)$, elements $p_1(x_1), ..., p_n(x_n)$ are free as well.

If elements $x_1, x_2, \cdots, x_n$ are free, they are called free random variables. With the above freeness definition, although one may construct abstract free random variables using possibly many mathematical concepts, it is not easy to show concrete examples of free random variables at this moment.

Two sets $\mathcal{S}_1$ and $\mathcal{S}_2$ are called free if any element in $\mathcal{S}_1$ and any element in $\mathcal{S}_2$ are free. With property (2), when $\{x_1, x_3\}$ and x_2 are free, it is easy to check that $E(x_1x_2) = E(x_1)E(x_2)$ and $E(x_1x_2x_3) = E(x_1x_3)E(x_2)$.

In many practical applications, we may need to deal with complex-valued random variables, such as, complex Gaussian, where the complex conjugation $*$ is usually used. In correspondence with the complex conjugation, the above freeness becomes $*$-freeness. We call that $x_1, x_2, \cdots, x_n$ are $*$-free, if (2) holds when the polynomials $p_k(x)$ in Definition 2.1 are changed to polynomials $p_k(x, x*)$ of two variables. If $x = x^*$, element x is called self-adjoint. For example, when x is a matrix and $*$ is the complex conjugate transpose operation, if x is Hermitian, then x is self-adjoint. In this case, x can be diagonalized by a unitary matrix and all its eigenvalues are real-valued.

Definition 2.2. 1) When two random variables x_1 and x_2 have all the moments the same, i.e., $E(x_1^m) = E(x_2^m)$ for all positive integers m, they are called identically distributed or having the same distribution.

2) For a sequence of random variables x_n, $n = 1, 2, ...$, we call x_n converges to x in distribution when n goes to infinity, if all the moments of x_n converge to the moments of x as n goes to infinity, i.e., for any positive integer m,

$$\lim_{n\to\infty} E(x_n^m) = E(x^m),$$

which is denoted as $\lim_{n\to\infty} x_n \stackrel{distr}{=} x$ or $x_n \stackrel{distr}{\longrightarrow} x$ as $n \to \infty$.

3) Let I be an index set. For each $i \in I$, let $x_n^{(i)}$, $n = 1, 2, ...$, be a sequence of random variables. We call that $(x_n^{(i)})_{i\in I}$ converges to $(x^{(i)})_{i\in I}$ in distribution, if

$$\lim_{n\to\infty} E(x_n^{(i_1)} \cdots x_n^{(i_k)}) = E(x^{(i_1)} \cdots x^{(i_k)})$$

for all positive integers k and all $i_1, ..., i_k \in I$, which is denoted as

$$\lim_{n\to\infty} (x_n^{(i)})_{i\in I} \stackrel{distr}{=} (x^{(i)})_{i\in I} \text{ or } (x_n^{(i)})_{i\in I} \stackrel{distr}{\longrightarrow} (x^{(i)})_{i\in I} \text{ as } n \to \infty.$$

The definition in 2) is about the convergence in distribution for a single sequence of random variables and the definition in 3) is about the convergence in distribution for multiple sequences of random variables jointly.

One of the most important results in classical probability theory is the central limit theorem. It says that the summation of independent random variables of a totally fixed variance converges to a Gaussian random variable, when the number of the independent random variables goes to infinity. For free random variables, it has the following free central limit theorem.

Theorem 2.1. *Let x_k, $k = 1, 2, ...$, be a sequence of self-adjoint, freely independent, and identically distributed random variables with $E(x_k) = 0$ and $E(x_k^2) = \sigma^2$. For a positive integer n, let*

$$S_n = \frac{x_1 + x_2 + \cdots + x_n}{\sqrt{n}}.$$

Then, S_n converges in distribution to a semicircular element s of variance σ^2 as $n \to \infty$, i.e.,

$$\lim_{n\to\infty} E(S_n^i) = \begin{cases} \sigma^i C_{i/2}, & \text{if } i \text{ is even}, \\ 0, & \text{if } i \text{ is odd}, \end{cases} \tag{3}$$

where C_k is the Catalan number and the $(2k)$th moment of the semicircular distribution:

$$C_k = \frac{1}{2\pi}\int_{-2}^{2} t^{2k}\sqrt{4-t^2}dt = \frac{1}{k+1}\binom{2k}{k}.$$

The random variable s in Theorem 2.1 is called a semicircular element in this context and it, after divided by σ, has the same distribution as the classical semicircular random variable of density function

$$q(t) = \begin{cases} \frac{1}{2\pi}\sqrt{4-t^2}, & \text{if } |t| < 2, \\ 0, & \text{otherwise}. \end{cases} \tag{4}$$

Its moment of an even order has the form in (3) and an odd order is always 0.

Note that semicircular distributions are the asymptotic distributions of the eigenvalues of Hermitian Gaussian random matrices when the matrix size goes to infinity, which is called Wigner's semi-circle law and will be discussed in more details in Section 4 later.

3. Moments, Cumulants, and Cauchy Transforms

As mentioned earlier, it is not convenient to directly define a density function or probability measure for a noncommutative random variable, and instead its all moments are defined and the freeness is to simplify the joint moments between free random variables.

In order to see how moments are related to distributions of free random variables, let us first see how in classical probability theory, a probability measure and its moments are related.

Let $\mu(t)$ be a probability measure on the real line $\mathbb{R}$. Assume its all moments are finite and let m_i be its ith moment for a positive integer i and $\varphi(t)$ be its characteristic function, i.e.,

$$m_i = \int_{\mathbb{R}} t^i d\mu(t), \text{ and } \varphi(t) = \int_{\mathbb{R}} e^{\mathbf{i}\tau t} d\mu(\tau),$$

where $\mathbf{i} \triangleq \sqrt{-1}$. Then, it is easy to see

$$m_i = \mathbf{i}^{-i}\varphi^{(i)}(0), \text{ and } \varphi(t) = \sum_{i=0}^{\infty} m_i \frac{(\mathbf{i}t)^i}{i!}, \tag{5}$$

where $\varphi^{(i)}(t)$ stands for the ith derivative of $\varphi(t)$. Furthermore, we can write

$$\log(\varphi(t)) = \sum_{i=1}^{\infty} k_i \frac{(\mathbf{i}t)^i}{i!} \text{ with } k_i = \mathbf{i}^{-i} \frac{d^i}{dt^i} \log(\varphi(t))\bigg|_{t=0}, \tag{6}$$

where k_i are called the cumulants of $\mu(t)$. We will call them the classical cumulants. The moment sequence $\{m_i\}_{i\geq 0}$ and the cumulant sequence $\{k_i\}_{i\geq 1}$ can be determined from each other:

$$m_n = \sum_{\substack{1\cdot r_1+\cdots+n\cdot r_n=n \\ r_1,\ldots,r_n\geq 0}} \frac{n!}{(1!)^{r_1}\cdots(n!)^{r_n} r_1!\cdots r_n!} k_1^{r_1}\cdots k_n^{r_n} \tag{7}$$

$$k_n = \sum_{\substack{1\cdot r_1+\cdots+n\cdot r_n=n \\ r_1,\ldots,r_n\geq 0}} \frac{(-1)^{r_1+\cdots+r_n-1}(r_1+\cdots r_n-1)!n!}{(1!)^{r_1}\cdots(n!)^{r_n} r_1!\cdots r_n!} m_1^{r_1}\cdots m_n^{r_n}. \tag{8}$$

Sometimes, cumulants may be easier to obtain than moments. In this case, one may first obtain cumulants and then moments.

Since for noncommutative random variables, we start with their moments as we have seen so far, it is very important to investigate moment and cumulant sequences for further calculations. Before going to more details, let us see some basic concepts about partitions of an index set, which plays an important role in free probability theory.

3.1. **Partitions, Non-crossing Partitions, and Free-Cumulants.** For a positive integer n, we denote $[n] \triangleq \{1, 2, ..., n\}$. A partition π of set $[n]$ means $\pi = \{V_1, ..., V_k\}$ such that $V_1, ..., V_k \subset [n]$ with $V_i \neq \emptyset$, $V_i \cap V_j = \emptyset$ for all $1 \leq i \neq j \leq n$, and $V_1 \cup \cdots \cup V_k = [n]$. Subsets $V_1, ..., V_k$ are called the blocks of π and $\#(\pi)$ denotes the number of the blocks of π. $\mathcal{P}(n)$ denotes the set of all the partitions of $[n]$. A partition is called a pairing if its each block has size 2 and the set of all the pairings of $[n]$ is denoted by $\mathcal{P}_2(n)$.

Let $\pi \in \mathcal{P}(n)$ and $\{l_i\}_i$ be a sequence. We denote $l_\pi \triangleq l_1^{r_1} l_2^{r_2} \cdots l_n^{r_n}$ where r_i is the number of blocks of π of size i. Then, the determination formulas in (7)-(8) of moments and cumulants can be re-formulated as

$$m_n = \sum_{\pi\in\mathcal{P}(n)} k_\pi, \tag{9}$$

$$k_n = \sum_{\pi\in\mathcal{P}(n)} (-1)^{\#(\pi)-1}(\#(\pi)-1)! m_\pi. \tag{10}$$

For $\pi \in \mathcal{P}(n)$, denote the moment of n random variables $x_1, ..., x_n$ with partition π as

$$E_\pi(x_1, ..., x_n) \triangleq \prod_{\substack{V\in\pi \\ V=(i_1,...,i_l)}} E(x_{i_1} \cdots x_{i_l}),$$

where $V = (i_1, ..., i_l)$ means that set V has l distinct elements with increasing order as $i_1 < i_2 < \cdots < i_l$.

When $\pi \in \mathcal{P}_2(2k)$, i.e., π is a pairing of $[2k]$, we have

$$E_\pi(x_1, ..., x_{2k}) = \prod_{(i,j)\in\pi} E(x_i x_j).$$

With this notation, for Gaussian random variables $X_1, X_2, ..., X_n$, we have the following Wick's formula:

$$E(X_{i_1} \cdots X_{i_{2k}}) = \sum_{\pi\in\mathcal{P}_2(2k)} E_\pi(X_{i_1}, ..., X_{i_{2k}}),$$

where $i_1, ..., i_{2k} \in [n]$.

Let $\pi \in \mathcal{P}(n)$. If there exist $i < j < k < l$ such that i and k are in one block V of π, and j and l in another block W of π, we call that V and W cross. If one cannot find any pair of blocks in π that cross, partition π is called non-crossing. Denote the set of all non-crossing partitions of $[n]$ by $NC(n)$ and the set of all non-crossing pairings of $[n]$ by $NC_2(n)$.

The partition set $\mathcal{P}(n)$ of $[n]$ is partially ordered via

$\pi_1 \leq \pi_2$ if and only if each block of π_1 is contained in a block of π_2.

With this order, $NC(n)$, as a subset of $\mathcal{P}(n)$, is also partially ordered. The largest and the smallest partitions in both $\mathcal{P}(n)$ and $NC(n)$ are $[n]$ and $\{\{1\}, \{2\}, ..., \{n\}\}$, denoted as 1_n and 0_n, respectively.

Definition 3.1. The following free cumulants $\kappa_n(x_1, ..., x_n)$ are defined inductively in terms of moments by the moment-cumulant formula:

$$E(x_1 \cdots x_n) = \sum_{\pi\in NC(n)} \kappa_\pi(x_1, ..., x_n), \tag{11}$$

where

$$\kappa_\pi(x_1, ..., x_n) \triangleq \prod_{\substack{V\in\pi \\ V=(i_1,...,i_l)}} \kappa_l(x_{i_1}, ..., x_{i_l}).$$

The above inductive definition is not hard to implement as follows.

For $n = 1$, we have $E(x_1) = \kappa_1(x_1)$. Thus, $\kappa_1(x_1) = E(x_1)$.

For $n = 2$, we have

$$E(x_1x_2) = \kappa_{(1,2)}(x_1, x_2) + \kappa_{(1),(2)}(x_1, x_2) = \kappa_2(x_1, x_2) + \kappa_1(x_1)\kappa_1(x_2).$$

Thus,

$$\kappa_2(x_1, x_2) = E(x_1x_2) - E(x_1)E(x_2),$$

etc.

Let $\mu(\pi_1, \pi_2)$ be the Möbius function on $\mathcal{P}(n)$ [7, 8, 11] that has a recursion formula to calculate. Then, we also have the following Möbius inversion formula:

$$\kappa_n(x_1, ..., x_n) = \sum_{\pi \in NC(n)} \mu(\pi, 1_n) E_\pi(x_1, ..., x_n). \tag{12}$$

The moment-cumulant formulas (11) and (12) for momemts and free-cumulants for noncummtative random variables are in analogous to (9) and (10) (or (7) and (8)) for classical random variables in classical probability theory.

Theorem 3.1. *Random variables $x_1, ..., x_n$ are free if and only if all mixed cumulants of $x_1, ..., x_n$ vanish. In other words, $x_1, ..., x_n$ are free if and only if, for any $i_1, ..., i_p \in [n] = \{1, 2, ..., n\}$ with $i_j \neq i_l$ for some $j, l \in [p]$, we have $\kappa_p(x_{i_1}, ..., x_{i_p}) = 0$.*

The result in the above theorem significantly simplifies the calculations of the free cumulants of multiple free random variables and therefore, helps to calculate the joint moments of multiple free random variables. For example, if x and y are free, then we have

$$\begin{aligned} \kappa_n^{x+y} &\triangleq \kappa_n(x+y, ..., x+y) \\ &= \kappa_n(x, ..., x) + \kappa_n(y, ..., y) + (\text{mixed cumulants in } x, y) \\ &= \kappa_n^x + \kappa_n^y. \end{aligned} \tag{13}$$

Definition 3.2. Let I be an index set. A self-adjoint family $(s_i)_{i \in I}$ is called a semicircular family of covariance matrix $C = (c_{ij})_{i,j \in I}$ if C is non-negative definite and for any $n \geq 1$ and any n-tuple $i_1, ..., i_n \in I$ we have

$$E(s_{i_1} \cdots s_{i_n}) = \sum_{\pi \in NC_2(n)} E_\pi(s_{i_1}, ..., s_{i_n}),$$

where

$$E_\pi(s_{i_1}, ..., s_{i_n}) = \prod_{(p,q) \in \pi} c_{i_p, i_q}.$$

If C is diagonal, then $(s_i)_{i\in I}$ is a free semicircular family.

The above formula is the free analogue of Wick's formula for Gaussian random variables. If we let $X_1, ..., X_r$ be $N \times N$ matrices of all entries in all matrices i.i.d. Gaussian random variables, then they jointly converge in distribution to a free semicircular family $s_1, ..., s_r$ of covariance matrix $(c_{ij})_{1\leq i,j\leq r} = \mathrm{I}_r$ where I_r is the identity matrix of size r, as N goes to infinity. More details on random matrices will be seen in Section 4.

3.2. **Cauchy Transforms and R-Transforms.** As we have seen earlier, for classical random variables, their distributions or density functions can be determined by their moment sequences or cumulant sequences as shown in (5) and (6). To further study noncommutative random variables, their moment and cumulant sequences similarly lead to their analytic forms as follows.

Let x be a noncommutative random variable and $m_n^x = E(x^n)$ and κ_n^x be its moments and free cumulants, respectively. Their power series (moment and cumulant generating functions) in an indeterminate z are defined by

$$M(z) = 1 + \sum_{n=1}^{\infty} m_n^x z^n \text{ and } C(z) = 1 + \sum_{n=1}^{\infty} \kappa_n^x z^n.$$

Then, the following identity holds:

$$M(z) = C(zM(z)).$$

The Cauchy transform of x is defined by

$$G(z) \triangleq E\left(\frac{1}{z-x}\right) = \sum_{n=0}^{\infty} \frac{E(x^n)}{z^{n+1}} = \sum_{n=0}^{\infty} \frac{m_n^x}{z^{n+1}} = z^{-1}M(z^{-1}),$$

and the R-transform of x is defined by

$$R(z) \triangleq \frac{C(z)-1}{z} = \sum_{n=0}^{\infty} \kappa_{n+1}^x z^n.$$

If we let $K(z) \triangleq R(z) + z^{-1}$, then $K(G(z)) = z$, i.e., $K(z)$ is the inverse of the Cauchy transform $G(z)$.

If we let $G_x(z)$ and $R_x(z)$ denote the Cauchy transform and the R-transform of random variable x, respectively, then, for two free random variables x and y, from (13) we have

$$R_{x+y}(z) = R_x(z) + R_y(z).$$

In case not both $R_x(z)$ and $R_y(z)$ are well-defined on a region of z, one may be able to find the Cauchy transform $G_{x+y}(z)$ of $x+y$ for free random variables x and y from the Cauchy transforms $G_x(z)$ and $G_y(z)$ of x and y as follows.

We shall see soon below that when z is in the upper complex plane $\mathbb{C}^+ \triangleq \{c \in \mathbb{C} | \text{Im}(c) > 0\}$ where $\mathbb{C}$ stands for the complex plane and Im stands for the imaginary part of a complex numnber, a Cauchy transform is well-defined.

For an $z \in \mathbb{C}^+$, solve the following system of two equations for two unknown functions $\omega_x(z)$ and $\omega_y(z)$:

$$G_x(\omega_x(z)) = G_y(\omega_y(z)) \text{ and } \omega_x(z) + \omega_y(z) - \frac{1}{G_x(\omega_x(z))} = z.$$

Then,

$$G_{x+y}(z) = G_x(\omega_x(z)) = G_y(\omega_y(z)).$$

If noncommutative random variable x is self-adjoint, then it has a spectral measure ν on $\mathbb{R}$ such that the moments of x are the same as the conventional moments of the probability measure ν. One can simply see it when x is a Hermitian matrix and then x can be diagonalized by a unitary matrix and has real-valued eigenvalues. These real-valued eigenvalues are the spectra of x that are discrete for a finite matrix but may become continuous when x is a general operator over an infinite dimensional space. In this case, we say that random variable x has distribution ν.

Then, the Cauchy transform $G(z)$ of x can be formulated as

$$G(z) = \int_{\mathbb{R}} \frac{1}{z-t} d\nu(t), \tag{14}$$

and $G(z)$ is also called the Cauchy transform of ν.

One can clearly see from (14) that Cauchy transform $G(z)$ is well-defined when $z \in \mathbb{C}^+$. In fact, $G(z)$ is analytic in $\mathbb{C}^+$, i.e., it exists derivatives of all orders for any $z \in \mathbb{C}^+$. Furthermore, $G(z) \in \mathbb{C}^-$, the lower complex plane similarly defined as $\mathbb{C}^+$. In other words, a Cauchy transform $G(z)$ maps $\mathbb{C}^+$ to $\mathbb{C}^-$.

From (14), one can also see that the Cauchy transform excludes the real axis $\mathbb{R}$ for z, which is because when $z \in \mathbb{R}$, the integration may not exist. After saying so, it may exist in the generalized function sense as if $z \in \mathbb{R}$, the Cauchy transform (14) becomes the Hilbert transform of $d\nu(t)/dt$.

When probability measure ν is compactly supported, i.e., it is supported on a finite interval, not only its Cauchy transform is analytic in $\mathbb{C}^+$, but also its R-transform is analytic on some disk centered at the origin. This, however, may not be true for a general probability measure ν. For more details, see [6].

With a Cauchy transform $G(z)$, its corresponding probability measure can be formulated by the Stieltjes inversion formula as follows.

Theorem 3.2. *Let ν be a probability measure on $\mathbb{R}$ and $G(z)$ be its Cauchy transform. For $a < b$, we have*

$$-\lim_{\tau\to 0^+}\frac{1}{\pi}\int_a^b Im(G(t+j\tau))dt = \nu((a,b)) + \frac{1}{2}\nu(\{a,b\}),$$

where $\nu((a,b))$ and $\nu(\{a,b\})$ are the continuous and the discrete parts of the measure ν, respectively. If ν_1 and ν_2 are two probability measures on $\mathbb{R}$ with equal Cauchy transforms, i.e., $G_{\nu_1}(z) = G_{\nu_2}(z)$, then $\nu_1 = \nu_2$.

This result tells us that Cauchy transforms and probability measures (distributions or random variables) are one-to-one corresponding to each other.

If x and y are two free self-adjoint random variables with distributions ν_x and ν_y, respectively. The distribution of $x + y$ is called the free convolution of those of x and y, which is denoted by $\nu_x \boxplus \nu_y$.

As an example of Cauchy transform, when ν is semicircular with density function $q(t)$ in (4), its Cauchy transform [6] is

$$G_s(z) = \frac{z - \sqrt{z^2 - 4}}{2}. \tag{15}$$

4. Application in Random Matrices

As mentioned in Introduction, random matrices with entries of complex Gaussian random variables are often used in wireless communications and signal processing. In particular, their singular value (eigenvalue) distributions play an important role in analyzing wireless communications systems. This section is on applying free probability theory to random matrices of large sizes. It tells us how to use the second order statistics of the entries of random matries to calculate their asymptotic eigenvalue distributions.

4.1. **GUE Random Matrices and Wigner's Semi-Circle Law.** Let X_N be an $N \times N$ matrix with complex random variables $a_{ij} = x_{ij} + \mathbf{i}y_{ij}$ as entries such that x_{ij} and y_{ij} are real Gaussian random variables, $\sqrt{N}a_{ij}$ is a standard complex random variable, i.e., $E(a_{ij}) = 0$ and $E(|a_{ij}|^2) = 1/N$ and

1) $a_{ij} = a_{ji}^*$,
2) $\{x_{ij}\}_{i\geq j} \cup \{y_{ij}\}_{i>j}$ are i.i.d.

In this case, X_N is Hermitian, i.e., self-adjoint. X_N is called a Gaussian unitary ensemble (GUE) random matrix. The following theorem is Wigner's semi-circle law.

Theorem 4.1. *If $\{X_N\}_N$ is a sequence of GUE random matrices, then, for any positive integer k,*

$$\begin{aligned}\lim_{N\to\infty} E(tr(X_N^k)) &= \frac{1}{2\pi}\int_{-2}^{2} t^k\sqrt{4-t^2}dt \\ &= \begin{cases} \frac{1}{l+1}\binom{2l}{l}, & \text{if } k=2l \text{ for some positive integer } l, \\ 0, & \text{if } k \text{ is odd.}\end{cases}\end{aligned}$$

where tr stands for the normalized matrix trace, i.e., $tr(\cdot) \triangleq Tr(\cdot)/N$ with the conventional matrix trace Tr.

Since X_N is Hermitian, it has spectra (eigenvalues) ν_N that is a random variable as well. Since $tr(X_N^k) = tr(\nu_N^k)$, we have

$$\lim_{N\to\infty} E(tr(X_N^k)) = \lim_{N\to\infty}\int_{\mathbb{R}} t^k d\nu_N(t).$$

Thus, the above theorem says that the eigenvalues of X_N converge in distribution to the semicircular random variable. In fact, the convergence in distribution can be made stronger to the almost surely convergence.

4.2. **Asymptotic Freeness of GUE Random Matrices.** For random matrices X as noncommuntative random variables, their linear functional E used in Section 2 is defined as $E(tr(X))$, i.e., $E(\cdot)$ used before for a noncommutative random variable x corresponds to $E(tr(\cdot))$ for a random matrix X in what follows.

Definition 4.1. Let $\{X_N\}_N$ and $\{Y_N\}_N$ be two sequences of $N\times N$ matrices. We say that X_N and Y_N are asymptotically free if they converge in distribution to two free random variables x and y, respectively, as N goes to infinity.

From Definitions 2.2 and 4.1, X_N and Y_N are asymptotically free, if for any positive integer m and non-negative integers $p_1, q_1, ..., p_m, q_m$ we have

$$\lim_{N\to\infty} E(tr(X_N^{p_1}Y_N^{q_1}\cdots X_N^{p_m}Y_N^{q_m})) = E(x^{p_1}y^{q_1}\cdots x^{p_m}y^{q_m}),$$

for two free random variables x and y.

For a sequence of $N\times N$ deterministic matrices $\{D_N\}_N$, if $\lim_{N\to\infty} tr(D_N^m)$ exists for every non-negative integer m, we say D_N converges to d in distribution, where d is a noncommutative random variable and its mth moment is the same as the limit. We also write it as $\lim_{N\to\infty} D_N \stackrel{distr}{=} d$ or $D_N \stackrel{distr.}{\longrightarrow} d$.

With the above notations, the following theorem of Voiculescu improves Wigner's semi-circle law.

Theorem 4.2. *Assume* $X_N^{(1)}, ..., X_N^{(p)}$ *are* p *independent* $N \times N$ *GUE random matrices and* $D_N^{(1)}, ..., D_N^{(q)}$ *are* q *deterministic* $N \times N$ *matrices such that*

$$D_N^{(1)}, ..., D_N^{(q)} \xrightarrow{distr} d_1, ..., d_q \text{ as } N \to \infty.$$

Then,

$$X_N^{(1)}, ..., X_N^{(p)}, D_N^{(1)}, ..., D_N^{(q)} \xrightarrow{distr} s_1, ..., s_p, d_1, ..., d_q \text{ as } N \to \infty,$$

where each s_i *is semicircular and* $s_1, ..., s_p, \{d_1, ..., d_q\}$ *are free. The convergence above also holds almost surely.*

This result tells that independent GUE random matrices $X_N^{(1)}, ..., X_N^{(p)}$, $\{D_N^{(1)}$, ...,$D_N^{(q)}\}$ are asymptotically free when N is large. Furthermore, $X_N^{(1)}, ..., X_N^{(p)}$ asymptotically have the same distributions as free semicircular elements $s_1, ..., s_p$ do, and this is still true even when they are mixed with deterministic matrices.

4.3. **Asymptotic Freeness of Haar Distributed Unitary Random Matrices.** For a general Hermtian random matrix, it can be diagonalized by a unitary matrix and in this case, the unitary matrix is random as well. Therefore, it is also important to study unitary random matrices.

Let $\mathcal{U}(N)$ denote the group of $N \times N$ unitary matrices U, i.e., $UU^* = U^*U = \mathrm{I}_N$. Since $\mathcal{U}(N)$ is bounded (compact), it has Haar meansure dU with $\int_{\mathcal{U}(N)} dU = 1$. Thus, dU is a probability measure (it can be understood as a uniform distribution). A Haar distributed unitary random matrix is a matrix U_N randomly chosen in $\mathcal{U}(N)$ with respect to Haar measure. One method to construct Haar unitary matrices is as follows. First, take an $N \times N$ random matrix whose entries are the independent standard complex Gaussian random variables. Then, use the Gram-Schmidt orthogonalization procedure to make it unitary.

A noncommutative random variable u is called Haar unitary if it is unitary, i.e., $uu^* = u^*u = 1$ and $E(u^m) = \delta_{0,m}$, i.e., 0 when $m > 0$. A Haar unitary random matrix is Haar unitary, i.e., if $U \in \mathcal{U}(N)$, then $E(tr(U^m)) = 0$ for $m > 0$ [6].

Theorem 4.3. *Assume* $U_N^{(1)}, ..., U_N^{(p)}$ *are* p *independent* $N \times N$ *Haar unitary random matrices and* $D_N^{(1)}, ..., D_N^{(q)}$ *are* q *deterministic* $N \times N$ *matrices such that*

$$D_N^{(1)}, ..., D_N^{(q)} \xrightarrow{distr} d_1, ..., d_q \text{ as } N \to \infty.$$

Then, as $N \to \infty$,

$$U_N^{(1)}, U_N^{(1)*}, ..., U_N^{(p)}, U_N^{(p)*}, D_N^{(1)}, ..., D_N^{(q)} \stackrel{distr}{\longrightarrow} u_1, u_1^*, ..., u_p, u_p^*, d_1, ..., d_q,$$

where each u_i is Haar unitary and $\{u_1, u_1^*\}, ..., \{u_p, u_p^*\}, \{d_1, ..., d_q\}$ are free. The convergence above also holds almost surely.

A more special case is as follows.

Theorem 4.4. *Let $\{A_N\}_N$ and $\{B_N\}_N$ be two sequences of deterministic $N \times N$ matrices with $\lim_{N\to\infty} A_N \stackrel{distr}{=} a$ and $\lim_{N\to\infty} B_N \stackrel{distr}{=} b$. Let $\{U_N\}_N$ be a sequence of $N \times N$ Haar unitary random matrices. Then,*

$$A_N, U_N B_N U_N^* \stackrel{distr}{\longrightarrow} a, b \text{ as } N \to \infty,$$

where a and b are free. This convergence also holds almost surely.

The above theorem says that A_N and $U_N B_N U_N^*$ are asymptotically free when N is large.

4.4. **Aymptotic Freeness of Wigner Random Matrices.** Let μ be a probability measure on $\mathbb{R}$ and a_{ij} with $i \le j$ be i.i.d. real random variables with distribution μ. Let $a_{ij} = a_{ji}$ for $i > j$, and

$$A_N = \frac{1}{\sqrt{N}}(a_{ij})_{1\le i,j\le N},$$

which is self-adjoint (symmetry) and called Wigner random matrix (ensemble).

Theorem 4.5. *Let $\mu_1, ..., \mu_p$ be probability measures on $\mathbb{R}$ with all moments exist and 0 mean. Assume $A_N^{(1)}, ..., A_N^{(p)}$ are p independent $N \times N$ Wigner random matrices with entry distributions $\mu_1, ..., \mu_p$, respectively, and $D_N^{(1)}, ..., D_N^{(q)}$ are q deterministic $N \times N$ matrices such that*

$$D_N^{(1)}, ..., D_N^{(q)} \stackrel{distr}{\longrightarrow} d_1, ..., d_q \text{ as } N \to \infty,$$

and

$$\sup_{r,N} \|D_N^{(r)}\| < \infty.$$

Then, as $N \to \infty$,

$$A_N^{(1)}, ..., A_N^{(p)}, D_N^{(1)}, ..., D_N^{(q)} \stackrel{distr}{\longrightarrow} s_1, ..., s_p, , d_1, ..., d_q,$$

where each s_i is semicircular and $s_1, ..., s_p, \{d_1, ..., d_q\}$ are free.

As a special case, $A_N D_N A_N, E_N \stackrel{distr}{\longrightarrow} sds, e$, where s is semicircular, sds and e are free, and e can be arbitrary.

5. Free Deterministic Equivalents and Random Matrix Singular Value Distribution Calculations

Let H be an $N \times M$ wireless channel matrix, which is usually modelled as a random matrix, with additive white Gaussian noise (AWGN) of variance σ. Then, its mutual information is

$$C(\sigma) = \frac{1}{N} E\left[\log\det\left(\mathrm{I}_N + \frac{HH^*}{\sigma}\right)\right], \tag{16}$$

where * stands for Hermitian operation. Let $\nu(\lambda)$ denote the eigenvalue distribution (or spectra, or probability measure) of matrix HH^*. Then, when N is large,

$$C(\sigma) = \int_0^\infty \log\left(1 + \frac{\lambda}{\sigma}\right) d\nu(\lambda).$$

On the other hand, the Cauchy transform of the probability measure ν and matrix HH^* is

$$G(z) = \int_0^\infty \frac{1}{z-\lambda} d\nu(\lambda) = E(tr(z\mathrm{I}_N - HH^*)^{-1}),$$

where $z \in \mathbb{C}^+$. Assume that $G(z)$ exists as Im$(z) \to 0^+$, whose limit is denoted by $G(\omega)$ with $\omega =$Re(z). For semicircular distribution, from (15) one can see that $G(\omega)$ exists when $\omega =$Re$(z) > 2$. Then, [15],

$$C(\sigma) = \int_\sigma^\infty \left(\frac{1}{\omega} - G(-\omega)\right) d\omega.$$

The above formula tells us that, to calculate the mutual information of the channel with channel matrix H, we only need to calculate the Cauchy transform of matrix HH^*.

As an example, if HH^* is a GUE random matrix, then, when N is large, it is approximately semicircular and its Cauchy transform has the form of (15) with a proper normalization. Thus, its mutual information can be calculated. However, in applications, HH^* may not be a GUE matrix. We next introduce free deterministic equivalents to help to calculate the Cauchy transforms of large random matrices, such as the above HH^*, based on Speicher [4, 6, 9].

5.1. **Matrix-Wise Free Deterministic Equivalents.** From Section 4, we know that when the entries X_{ij}, $i \geq j$, of an $N \times N$ self-adjoint (symmetry for real-valued or Hermitian for complex-valued) matrix X are i.i.d. random variables, when N is large, it is approximately semicircular. It is also true for multiple such random matrices and multiple deterministic matrices jointly.

For a non-adjoint random matrix X of i.i.d. Gaussian entries, it can be made into two independent self-adjoint GUE matrices as $Y_1 = (X + X^*)/\sqrt{2}$ and $Y_2 =$

$-\mathbf{i}(X - X^*)/\sqrt{2}$. Then, $X = (Y_1 + \mathbf{i}Y_2)/\sqrt{2}$. In this case, X converges in distribution to $s = (s_1 + \mathbf{i}s_2)/\sqrt{2}$ for two free semicircular elements s_1 and s_2 with the same distribution. While s_1 and s_2 are semicircular, we call s circular.

In [9, 6] it is proposed to replace these random matrices by semicircular and circular elements etc. Consider the following collections of $N \times N$ matrices, where for each random matrix, its entries of different random variables are i.i.d.:

$$\begin{aligned}
\mathbf{X} = \{X_1, ..., X_{n_1}\} &: \quad \text{independent self-adjoint matrices,} \\
\mathbf{Y} = \{Y_1, ..., Y_{n_2}\} &: \quad \text{independent non-self-adjoint matrices,} \\
\mathbf{U} = \{U_1, ..., U_{n_3}\} &: \quad \text{independent Haar distribued unitary matrices,} \\
\mathbf{D} = \{D_1, ..., D_{n_4}\} &: \quad \text{deterministic matrices.}
\end{aligned}$$

Let

$$\begin{aligned}
\mathbf{s} = \{s_1, ..., s_{n_1}\} &: \quad \text{free semicircular,} \\
\mathbf{c} = \{c_1, ..., c_{n_2}\} &: \quad \text{free circular,} \\
\mathbf{u} = \{u_1, ..., u_{n_3}\} &: \quad \text{free Haar unitary,} \\
\mathbf{d} = \{d_1, ..., d_{n_4}\} &: \quad \text{abstract elements.}
\end{aligned}$$

Assume that the joint distribution of $\mathbf{D}$ is the same as that of $\mathbf{d}$, and $\mathbf{X}, \mathbf{Y}, \mathbf{U}$ are independent among each other. Also assume that $\mathbf{s}, \mathbf{c}, \mathbf{u}$ have their each individual distribution asymptotically the same as that of $\mathbf{X}, \mathbf{Y}, \mathbf{U}$, respectively.

Let P_N be a multi-variable polynomial of $\mathbf{X}, \mathbf{Y}, \mathbf{U}, \mathbf{D}$. Then, when N is large,

$$P_N = P(X_1, ..., X_{n_1}, Y_1, ..., Y_{n_2}, U_1, ..., U_{n_3}, D_1, ..., D_{n_4})$$

can be replaced by

$$P_N^{\square} = P(s_1, ..., s_{n_1}, c_1, ..., c_{n_2}, u_1, ..., u_{n_3}, d_1, ..., d_{n_4})$$

and $P_N^{\square}$ is called the (matrix-wise) free deterministic equivalent of P_N. Then, we have, for any positive integer k,

$$\lim_{N\to\infty} E(tr(P_N^k)) = E((P_N^{\square})^k).$$

Now let us go back to the matrix HH^* in (16). Although matrix H is not self-adjoint itself, but if we follow [4] and [6] and let

$$T = \begin{pmatrix} 0 & H \\ H^* & 0 \end{pmatrix},$$

then, matrix T is self-adjoint. Furthermore,

$$T^2 = \begin{pmatrix} HH^* & 0 \\ 0 & H^*H \end{pmatrix},$$

which includes HH^* as a diagonal block. Using operator-valued free probability theory [9, 6], it can be similarly treated as what is done in the previous sections. Note that T^2 is just a polynomial of T but unfortunately not all entries in matrix T have the same distribution, which makes the above matrix-wise free deterministic equivalent approach difficult to use. In order to deal with this problem, we next consider component-wise free deterministic equivalents.

5.2. **Component-Wise Free Deterministic Equivalents and Cauchy Transform Calculation of Random Matrices.** This part is mainly from [4]. We consider $N \times N$ random matrices $X = (X_{ij})$ where X_{ij} are complex Gaussian random variables with $E(X_{ij}) = 0$ and $E(X_{ij}X_{ij}^*) = \sigma_{ij}/N$, where σ_{ij} are independent of N. Now we replace all entries X_{ij} in X by (semi)circular elements c_{ij} such that

$$E(c_{ij}c_{ij}^*) = E(X_{ij}X_{ij}^*) = \sigma_{ij}/N$$

where if X_{ij} is real-valued (or complex-valued), then, c_{ij} is semicircular (or circular) with mean 0; if X_{ij} and X_{kl} are independent, then c_{ij} and c_{kl} are free; if $X_{ij} = X_{kl}$, then $c_{ij} = c_{kl}$; and $E(X_{ij}X_{kl}^*) = E(c_{ij}c_{kl}^*) = (E(c_{ij}^*c_{kl}))^*$. Then, we form an $N \times N$ matrix of (semi)circular elements as $c = (c_{ij})$. Matrix c is called the component-wise free deterministic equivalent of matrix X.

Let $X_1, ..., X_{n_1}$ be n_1 random matrices, each of which is specified above, where all entries of each of these matrices are independent from all entries of all the remaining matrices. Let $c_1, ..., c_{n_1}$ be the component-wise deterministic equivalents of $X_1, ..., X_{n_1}$, and all elements in c_i are free from all elements in c_j when $i \neq j$. Let $D_1, ..., D_{n_2}$ be n_2 deterministic matrices. Assume P_N is a multi-variable polynomial and

$$\begin{aligned} P_N &= P(X_1, ..., X_{n_1}, D_1, ..., D_{n_2}), \\ P_N^{\square} &= P(c_1, ..., c_{n_1}, D_1, ..., D_{n_2}). \end{aligned}$$

We say that $P_N^{\square}$ is the component-wise free deterministic equivalent of P_N.

5.2.1. *Independent Cases.* Consider the case when every matrix X_i is Hermitian/self-adjoint and entries X_{ij} for $i \geq j$ are all independent. It is explicitly shown in [18] that $\lim_{N\to\infty}(P_N - P_N^{\square}) = 0$, i.e., the matrices $X_1, ..., X_{n_1}, D_1, ..., D_{n_2}$ have the same joint distribution as matrices $c_1, ..., c_{n_1}, D_1, ..., D_{n_2}$ do. Thus, $c_1, ..., c_{n_1}$ may be used to calculate the Cauchy transforms of $X_1, ..., X_{n_1}$ when only the variances of the entries in matrices X_i are used, as N is large.

We now consider a special example shown in [4]. Let $X = (X_{ij})$ be an $N \times N$ Hermitian/self-adjoint Gaussian random matrix with $E(X_{ij}) = 0$ and $E(X_{ij}X_{ij}^*) = \sigma_{ij}/N$, and let $c = (c_{ij})$ be its component-wise deterministic equivalent, i.e., $E(c_{ij}) = 0$ and $E(c_{ij}c_{ij}^*) = \sigma_{ij}/N$. Note that since $X_{ij} = X_{ji}^*$, we have $c_{ij} = c_{ji}^*$ as well. Let A be an $N \times N$ deterministic matrix.

Consider the matrix sum $Y = A + X$. We next show how to calculate the Cauchy transform of Y by calculating that of $T = A + c$.

For an $N \times N$ deterministic matrix $B = (B_{ij})$, define a mapping η that maps B to another $N \times N$ deterministic matrix $\eta(B)$ with its (i, j)th component as

$$[\eta(B)]_{ij} \triangleq E(cBc) = \sum_{k,l} E(c_{ik}B_{kl}c_{lj}) = \sum_{k,l} E(c_{ik}c_{jl}^*)B_{kl} = \delta_{i,j} \sum_{k} \sigma_{ik}B_{kk}, \quad (17)$$

which shows that $\eta(B)$ is a diagonal matrix. Then, the Cauchy transform $g_T(z)$ of T can be determined by solving the following fixed point equation [10], [4]:

$$g_T(z) = tr(G_T(z)), \quad (18)$$

$$G_T(z) = E\left(\frac{1}{z - \eta(G_T(z)) - A}\right), \quad (19)$$

where $E(B) \triangleq (E(B_{ij}))$. It is shown in [10] that there is exactly one solution of the above fixed point equation with the proper positivity constriant.

We next consider the case when X is not Hermitian, such as the channel matrix H in (16), where all entries of X are independent. In this case, consider $Y = A + X$ and we next calculate the Cauchy transform of YY^*. To do so, define

$$T = \begin{pmatrix} 0 & Y \\ Y^* & 0 \end{pmatrix} = \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix} + \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix}.$$

Then,

$$T^2 = \begin{pmatrix} YY^* & 0 \\ 0 & Y^*Y \end{pmatrix}.$$

Since the eigenvalue distributions of YY^* and Y^*Y are the same, the Cauchy transform of YY^* is the same as that of T^2. It is presented in [4] as follows.

For an $M \times M$ matrix $B = (B_{ij})$, define

$$E_{D_M}(B) \triangleq \text{diag}(E(B_{11}), \cdots, E(B_{MM})),$$

where diag stands for the $M \times M$ diagonal matrix with its arguments as its diagonal elements, and also define

$$\eta_1(B) \triangleq E(cBc^*) \text{ and } \eta_2(B) \triangleq E(c^*Bc).$$

Note that since all the entry elements in matrix c are free from each other, $E(cBc^*) = E_{D_N}(cBc^*)$ and $E(c^*Bc) = E_{D_N}(c^*Bc)$ as what is shown for η in (17).

Then, the Cauchy transform $g_{T^2}(z)$ of T^2 or YY^* is $g_{T^2}(z) = tr(G_{T^2}(z))$ and $G_{T^2}(z)$ can be obtained by solving the following fixed point equations [13], [4]:

$$zG_{T^2}(z^2) = G_T(z) = E_{D_{2N}}\left[\begin{pmatrix} z - z\eta_1(G_2(z^2)) & -A \\ -A^* & z - z\eta_2(G_1(z^2)) \end{pmatrix}^{-1}\right], \tag{20}$$

where

$$zG_1(z) = E_{D_N}\left[\left(1 - \eta_1(G_2(z)) + A\frac{1}{z - z\eta_2(G_1(z))}A^*\right)^{-1}\right], \tag{21}$$

$$zG_2(z) = E_{D_N}\left[\left(1 - \eta_2(G_1(z)) + A\frac{1}{z - z\eta_1(G_2(z))}A^*\right)^{-1}\right]. \tag{22}$$

5.2.2. *Correlated Cases and Summary.* When the entries in matrix X are correlated, similar treatment as the above can be done [4]. One can still get the Cauchy transform of Y when X is Hermitain by solving the fixed point equation (18)-(19) and the Cauchy transform of YY^* when X is not Hermitian by solving the fixed point equation (20)-(22), where $\eta(B) = E(cBc)$ may not be diagonal as what is calculated in (17), and $\eta_1(B)$ and $\eta_2(B)$ may not be diagonal either. An example of correlated entries in X is that each column vector (or row vector) of X is a linear transform of a vector of independent Gaussian random variables.

A simpler example of correlated cases is when random matrix $X_1 = BX$ where B is a deterministic matrix and X is a random matrix of independent entries. In this case, X_1 can be treated as a product of two matrices of B and X and thus, was covered previously.

The above Cauchy transform calculation is only based on the covariances (the second order statistics) of the entries of random matrix X. As we mentioned easlier, in this case one does not need to implement Monte-Carlo simulations to do the calculations that may be not convenient in practice when X has a large size.

Going back to the mutual information in the beginning of this section, we can just let $A = 0$ in the above to get the Cauchy transform of $HH^* = YY^*$.

As a remark, the deterministic equivalents defined above are from [9, 6, 4], which we refer to for any difference with those appeared in [12, 13].

6. Conclusions

As mentioned in the beginning of this paper, the main goal here is to introduce free probability theory and its application to random matrices as simple as possible. It is intended for a non-mathematics major researcher in, for example, communications and signal processing areas. This paper is mainly based on [4, 5, 6, 7].

Free probability theory is about noncommutative elements or random variables, such as, random matrices, in contrast to the conventional (real-valued or complex-valued) commutative random variables in the classical probability theory. The freeness significantly simplifies the calculations of the moments and therefore the distributions, and interestingly, random matrices, when their size is large, do have the freeness asymptotically. Therefore, free probability theory is naturally applied to calculate the asymptotic distributions of the eigenvalues/singular-values of random matrices when their size is large, such as wireless channel matrices in massive MIMO systems. It is particularly interesting that the calculation only needs the second order statistics of the matrix entries.

This paper is based on the author's own understanding on free probability theory and by no means the material covered in this paper is complete. More complete materials on this topic are referred to [4, 5, 6, 7, 8, 11, 15, 17].

Acknowledgment

The author would like to thank Dr. Roland Speicher for his free online video lectures [7] and for his helps to my questions and Dr. Anan Lu for his numerous discussions on free probability theory. He also would like to thank the anonymous reviewers for their helpful comments.

References

1. D. Voiculescu, *Symmetries of some reduced free products C^*-algebras*, Operator Algebras and Their Connections with Topology and Ergordic Theory (Busteni, 1983), Lecture Notes in Mathematics, vol. 1132, pp. 556-588, Springer, Berlin, 1985.
2. D. Voiculescu, *Addition of certain noncommuting random variables*, J. Funct. Anal., vol. 66, no. 3, pp. 323-346, 1986.
3. D. Voiculescu, *Multiplication of certain non-commuting random variables*, J. Oper. Theory, vol. 18, no. 2, pp. 223-235, 1987.

4. R. Speicher, *Free probability theory and random matrices*, Lecture Slides Presented at Workshop Random Matrix Theory Wireless Commun., Boulder, CO, USA, July 2008, pp. 1-57. Available online: http://www.mast.queensu.ca/~speicher/papers/Boulder.pdf.
5. R. Speicher, *Free probability theory and its avatars in representation theory, random matrices, and operator algebras; also featuring: non-commutative distributions*, Jahresber Dtsch Math-Ver, vol. 119, no. 1, pp. 3-30, 2017.
6. J. A. Mingo, R. Speicher, *Free Probability and Random Matrices*, Springer, New York, 2017.
7. R. Speicher, *26 Online Lectures*. Available online: https://www.math.uni-sb.de/ag/speicher/web_video/freeprobws1819/lec1.html, 2019.
8. A. Nica, R. Speicher, *Lectures on the Combinatorics of Free Probability*, London Mathematical Society Lecture Note Series, vol. 335, Cambridge University Press, Cambridge, 2006.
9. R. Speicher, A. Vargas, *Free deterministic equivalents, rectangular random matrix models, and operator-valued free probability theory*, Random Matrices: Theory Appl., vol. 1, no. 02, Art. No. 1150008, 2012.
10. J. W. Helton, R. R. Far, R. Speicher, *Operator-valued semicircular elements: solving a quadratic matrix equation with positivity constraints*, Int. Math. Res. Notices IMRN 2007 (22), 15, 2007. Art. ID rnm 086.
11. F. Hiai, D. Petz, *The Semicircle Law, Free Random Variables, and Entropy*, Mathematical Surveys and Monographs, vol. 77, American Mathematical Society. Providence, RI, 2000.
12. V. L. Girko, *Theory of Stochastic Canonical Equations*, Kluwer, Dordrecht, 2001.
13. W. Hachem, P. Loubaton, J. Najim, *Deterministic equivalents for certain functionals of large random matrices*, Ann. Appl. Probab., vol. 17, no. 3, pp. 875-930, Jun. 2007.
14. A. M. Tulino, S. Verdú, *Impact of antenna correlation on the capacity of multiantenna channels*, IEEE Trans. Inform. Theory, vol. 51, no. 7, pp. 2491-2509, Jul. 2005.
15. A. M. Tulino, S. Verdú, *Random Matrix Theory and Wireless Communications*, Norwell, MA, USA: NOW Publisher, 2004.
16. R. Couillet, M. Debbah, J. W. Silverstein, *A deterministic equivalent for the analysis of correlated MIMO multiple access channels*, IEEE Trans. Inform. Theory, vol. 57, no. 6, pp. 3493-3514, Jun. 2011.
17. R. Couillet, M. Debbah, *Random Matrix Methods for Wireless Communications*, Cambridge University Press, Cambridge, U.K., 2011.
18. A.-A. Lu, X. Q. Gao, C. Xiao, *Free deterministic equivalent for the analysis of MIMO multiple access channel*, IEEE Trans. Inform. Theory, vol. 62, no. 8, pp. 4604-4629, Aug. 2016.