**Diplomarbeit**

# Free Probability and Random Matrices: Theory and Applications to MIMO Communication Systems

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Diplom–Ingenieurs
unter der Leitung von

Dipl. Ing. Dominik Seethaler
Dipl. Ing. Dr. techn. Harold Artés
Ao. Prof. Dipl. Ing. Dr. techn. Franz Hlawatsch
Institut für Nachrichtentechnik und Hochfrequenztechnik (E 389)

eingereicht an der Technischen Universität Wien
Fakultät für Elektrotechnik und Informationstechnik

von

**Ana Magdalena Skupch, 9625552**
Lerchenfelderstraße 37/11
A-1070 Wien

Wien, Februar 2005 ____________________

**Abstract**

In this thesis we give an overview of free probability, random matrices and their applications for MIMO communication systems. First, Chapter 1 introduces the main concepts, definitions and theorems of free probability theory and compares them to the ones of classical probability theory. In Chapter 2 we present how random matrices can be treated with free probability; how we can embed them into the definitions and how the theorems of free probability can be applied to them. Chapter 3 presents some interesting applications of free probability and random matrices for MIMO communication systems, as the computation of pairwise error probability of space-time codes and the derivation of the asymptotic channel capacity of a correlated MIMO channel by means of the free multiplicative convolution. Finally, Chapter 4 introduces some deeper concepts of classical probability and free probability as the classical and free cumulant sequence and the Fock spaces.

# Contents

# Chapter 1

# Free probability

Free probability theory (FPT) was first introduced by Voiculescu [24], who extended the already existing non-commutative probability theory by adding the concept of the free independence. Non-commutative probability theory [1] was initially introduced to solve problems in quantic mechanics, by considering selfadjoint operators as kinds of random variables and it turned out that the theory became very interesting by itself. In Voiculescu's work in FPT, links with random matrix theory appeared [23] which made the theory interesting for communication engineers. The FPT has its analogy to classical probability theory (CPT) by construction. FPT employs the same concepts as CPT, i.e. this includes the concept of a probability space, of random variables, of distributions, of moments, of independence, and the concept of convolution. Nevertheless it is not a generalization; CPT is not a special case of FPT and neither is FPT a special case of CPT. If we want to consider classical random variables (RVs) in the framework of FPT, we need to embed them into it. There are several ways of doing it; some of them will be presented in the following. The concepts and problems of probability theory are born from and deal with the analysis of random phenomena. This is not the case for FPT, where the background is a very theoretical one. In the following, the main definitions and theorems of FPT will be introduced, i.e. the concept of a non-commutative probability space, the concept of non-commutative random variables and its distributions, the concept of free independence and the free central limit theorem, and finally the concepts of free additive and multiplicative convolution. The analogy to CPT will be pointed out and become more and more evident in the course of this chapter.

## 1.1 Non-commutative probability spaces

In CPT a random variable is a representation of random events, which can be very abstract objects, in the real numbers. Hence a RV is a function that maps random events into the real numbers. The random events are seen as sets grouped together to a probability space $(\Omega, \Sigma, \mathbb{P})$ and a RV is defined as a function on this probability

space. With random variables, we are able to calculate expectations, mean values for random objects and grasp the randomness in a way. Following the definition in [13], the probability space consists of a sure event $\Omega$, which is the set of all elementary events, the $\sigma-$algebra $\Sigma$ of events (in the measure theoretical sense, see Chapter 1 in [13]) and the probability measure $\mathbb{P}$ on $\Sigma$, that fullfills the condition $\mathbb{P}(\Omega) = 1$. An elementary event is then an element belonging to the space $\Omega$ and an event is a set of elementary events. A random variable $X$ maps the random events from the probability space into the real numbers, hence it is a function $X : \Omega \to \mathbb{R}$. In FPT, the definitions of probability spaces and random variables are taken and generalized to a new definition. In contrast to CPT where the concepts of a probability space and random variable have an intuitive background, in FPT these concepts loose their intuitiveness. What remains is only one thing: the fact that random variables form an algebra.
An algebra is defined [12] as a vector space $V$ together with a product $\langle \cdot, \cdot \rangle$:

$$\begin{aligned} \langle \cdot, \cdot \rangle : \quad V \times V &\to V, \\ \langle x, y \rangle &\mapsto xy, \quad \forall x, y \in V. \end{aligned}$$

The product respects the laws of distributivity, hence for $\alpha, \beta \in \mathbb{R}$, $x, y, z \in V$, we have

$$\begin{aligned} (\alpha x + \beta y)z &= \alpha(xz) + \beta(yz), \\ x(\alpha y + \beta z) &= \alpha(xy) + \beta(xz). \end{aligned}$$

Since the sum and the product of classical random variables are again random variables, we can say that they form an algebra. FPT introduces the definition of a non-commutative random variable as a mathematical object that has the property of being part of an algebra [24].

**Definition 1.1.1:** *A* non-commutative probability space $(\mathcal{A}, \varphi)$ *is an algebra* $\mathcal{A}$ *over* $\mathbb{C}$ *with the unit* 1, *endowed with a linear functional* $\varphi : \mathcal{A} \to \mathbb{C}$, *fullfilling* $\varphi(1) = 1$, *called the* state. *Elements* $A \in \mathcal{A}$ *are called* non-commutative random variables.

Recall that a classical random variable (RV) is a function that maps random events into the real numbers [13]. Actually a RV can be regarded as a NCRV, by defining the appropriate non-commutative probability space. How this is done will be explained in the following example. The definition of a classical RV is limited, while the one of a NCRV remains very general. A very wide class of objects can be regarded as non-commutative random variables. Actually anything that is part of a unital algebra (this is an algebra containing the unit). NCRVs were actually conceived to consider operators as RVs and apply a similar probability theory on them. How operators can be embedded into FPT is demonstrated in the Appendix, Section 4.2.1. In this thesis, we will consider most of the time random matrices as NCRVs. In the following, two examples are given of how non-commutative probability spaces and NCRVs can be defined.

EXAMPLE 1.1.1: Classical random variables
There are many ways of putting classical random variables into the context of FPT. The simplest is to consider classical random variables over a specific probability space $(\Omega, \Sigma)$. These RV are measurable functions into the Borel probability space [13] . If $X_1$ and $X_2$ are two RV on $(\Omega, \Sigma)$, then $\alpha X_1 + \beta X_2$ and $\alpha X_1 X_2$ are RV as well. So the space of RVs forms an algebra. The unit of this algebra is the determinate function $I := 1$, that always equals one. As the state $\varphi$ we choose the expectation, so

$$\varphi(X^k) := \mathbb{E}\{X^k\}.$$

The required condition $\varphi(I) = \mathbb{E}\{I\} = 1$ is satisfied. The algebra together with the state form a non-commutative probability space. RVs are elements of it, hence they are NCRVs.

■

EXAMPLE 1.1.2: Random matrices
Analogous, if $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n\times n}$ then $\alpha\mathbf{A} + \beta\mathbf{B}$ and $\alpha\mathbf{AB} \in \mathbb{C}^{n\times n}$, $\forall\alpha, \beta \in \mathbb{C}$. This means, that the space of RMs $\in \mathbb{C}^{n\times n}$ forms an algebra over $\mathbb{C}$ with the unit matrix $\mathbf{I}$. We will call this algebra $\mathcal{M}_n$. In order to get a non-commutative probability space $(\mathcal{M}_n, \varphi)$, any linear functional can be assigned to $\mathcal{M}_n$, it just has to fulfill the condition $\varphi(\mathbf{I}) = 1$. The state that is introduced in this example is the state, that was introduced by Voiculescu in the context of random matrices in free probability [23]:

$$\varphi(\mathbf{A}) = \tau_n(\mathbf{A}) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\{\mathbf{A}_{ii}\}.$$

■

## 1.2 Distributions

As we can see the state in FPT can be compared to the expectation functional from the CPT.In analogy to CPT, in FPT the expressions $\varphi(A^k)$ with $A$ being a NCRV are called moments. Most of the times in CPT the moments give a full description of the RV. The statistical information of a RV in the CPT is summarized under the probability distribution (or its derivation the probability density function (pdf)). With this function all the moments can be calculated. Similarly in FP a functional is introduced, that fully describes the NCRV because all the moments can be obtained with it. The functional applies on the space $\mathbb{C}\langle X\rangle$ of all polynomials with complex coefficients. In the literature this functional is just called distribution, but in order to avoid confusions with the distribution function from CPT, we are defining it as the *distribution functional.*

**Definition 1.2.1:** *The* distribution functional *of a non-commutative random variable $A$ is the linear functional $\mu_A : \mathbb{C}\langle X\rangle \to \mathbb{C}$ that statisfies*

$$\mu_A(x^k) = \varphi(A^k).$$

*For a polynomial $p(x) = \sum_{i=0}^{n} \alpha_i x^i \in \mathbb{C}\langle X\rangle$, we have*

$$\mu_A(p(x)) = \varphi(p(A)) = \sum_{i=0}^{n} \alpha_i \varphi(A^i).$$

The reason why it is a functional and not a function as we are used to in CPT will become clearer, when dealing with more than one NCRV. For a single NCRV a distribution measure can be given; however this is not possible for joint probabilities anymore.

**Definition 1.2.2:** *Under the condition, that the moment problem has a unique solution, a* distribution measure *can be given. It is the distribution $F(x)$ associated to the moment sequence $\varphi(A^k)$ that satisfies:*

$$\varphi(A^k) = \int x^k dF(x).$$

*If $F$ is differentiable, a probability density function $f(x) = \frac{dF(x)}{dx}$ can be evaluated and we can write*

$$\varphi(A^k) = \int x^k f(x) dx.$$

Example 1.2.1: Classical random variables
The distribution functional of a RV $X$, in the context of FPT is:

$$\mu_X(x^k) = \varphi(x^k) = \mathbb{E}\{X^k\}.$$

The associated distribution measure is the distribution of the RV, $F_X(x)$ and its associated pdf is $f_X(x)$, since:

$$\varphi(x^k) = \int x^k dF_X(x) = \int x^k f_X(x) dx = \mathbb{E}\{X^k\}.$$

■

Example 1.2.2: Random matrices
The distribution functional of a random matrix $\mathbf{A}$ using Voiculescu's state is

$$\mu_{\mathbf{A}}(x^k) = \tau_n(\mathbf{A}^k) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left\{(\mathbf{A}^k)_{ii}\right\} = \mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^{n}(\mathbf{A}^k)_{ii}\right\} = \mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^{n}\lambda_i^k\right\},$$

where $\lambda_i$ are the eigenvalues of $\mathbf{A}$.
The distribution measure $F_{\mathbf{A}}(x)$ and its associated pdf $f_{\mathbf{A}}(x)$ are given by

$$F_{\mathbf{A}}(x) = \frac{1}{n}\sum_{i=1}^{n} F_{\lambda_i}(x), \quad f_{\mathbf{A}}(x) = \frac{1}{n}\sum_{i=1}^{n} f_{\lambda_i}(x),$$

with $F_{\lambda_i}(x)$ and $f_{\lambda_i}(x)$ denoting the marginal distributions and densities, respectively, of the eigenvalues $\lambda_i$ of $\mathbf{A}$. More interpretations of these distributions are given in Chapter 2.
■

In CPT we are able to describe how RV are related to each other over the joint distribution function. For two random variables $X_1, X_2$ it is $F_{X_1,X_2}(x_1, x_2)$ that enables us to calculate the joint moments $\mathbb{E}\{X_1^n X_2^m\}$, $n, m \in \mathbb{N}$. Notice that for commutative random variables, $\mathbb{E}\{X_1^n X_2^m\} = \mathbb{E}\{X_1^{n-1} X_2 X_1 X_2^{m-1}\}$; note that this is not necessarily the case for non-commutative ones. It is not possible to give a joint distribution measure for NCRV. This is why a functional is introduced, with which we can get all non-commutative moments. The functional is defined over the space of non-commutative polynomials $\mathbb{C}\langle X_1, X_2\rangle$. We call a non-commutative polynomial $p_n(x_1, x_2) \in \mathbb{C}\langle X_1, X_2\rangle$ of order $n$, depending on two variables $x_1$ and $x_2$, if

$$p_n(x_1, x_2) := \sum_k \alpha_k\, x_1^{i_1}\, x_2^{j_1}\, x_1^{i_2}\, x_2^{j_2}\, \ldots\, x_1^{i_n}\, x_2^{j_n}, \quad \text{with} \quad \sum_{l=1}^{n} i_l \leq n, \quad \sum_{l=1}^{n} j_l \leq n.$$

A non-commutative polynomial includes all non-commutative products of $x_1$ and $x_2$, where every element contains a maximum of $n$ factors $x_1$ and $x_2$, respectively. How a simple non-commutative polynomial is build can be seen in the following example.

EXAMPLE 1.2.3: The general form of a non-commutative polynomial of order 2 is:

$$\begin{aligned} p_2(x_1, x_2) \;=\;& \alpha_1 1 + \alpha_2 x_1 + \alpha_3 x_2 + \alpha_4 x_1 x_2 + \alpha_5 x_2 x_1 + \alpha_6 x_1 x_2 x_1 \\ &+ \alpha_7 x_2 x_1 x_2 + \alpha_8 x_1^2 x_2 + \alpha_9 x_1 x_2^2 + \alpha_{10} x_2 x_1^2 + \alpha_{11} x_2^2 x_1 \\ &+ \alpha_{12} x_1^2 x_2^2 + \alpha_{13} x_2^2 x_1^2 + \alpha_{14} x_1 x_2 x_1 x_2 + \alpha_{15} x_2 x_1 x_2 x_1. \end{aligned}$$

Notice that every factor is composed of maximal two $x_1's$ and $x_2's$, respectively. ■

**DEFINITION 1.2.3:** *The* distribution functional *of two non-commutative random variables $A_1$ and $A_2$ is the linear functional $\mu_{A_1,A_2} : \mathbb{C}\langle X_1, X_2\rangle \to \mathbb{C}$ satisfying*

$$\mu_{A_1,A_2}(\,p\,) = \varphi(\,p(A_1, A_2)\,), \quad \text{for } p \in \mathbb{C}\langle X_1, X_2\rangle.$$

Notice that the space $\mathbb{C}\langle X_1, X_2\rangle$ using as variables the NCRVs $A_1$ and $A_2$ can be interpreted as the subalgebra generated by $A_1$ and $A_2$. Since the subalgebra generated by two elements $A_1$ and $A_2$ needs to contain all products and sums it has to contain all possible non-commutative polynomials in $A_1$ and $A_2$.

Definition 1.2.3 can also be extended to a family of random variables $(X_i)_{i\in\mathcal{I}}$, $\mathcal{I}$ being a set of indices. The obtained functional is defined over the space of non-commutative polynomials $\mathbb{C}\langle X_i | i \in I\rangle$ depending on the variables $(x_i)_{i\in\mathcal{I}}$.

**DEFINITION 1.2.4:** *The* distribution functional of a family of non-commutative random variables $(A_i)_{i\in I} \in (\mathcal{A}, \varphi)$ *is a linear functional* $\mu_{(A_i)_{i\in I}} : \mathbb{C}\langle X_i | i \in I\rangle \to \mathbb{C}$ *defined by*

$$\mu_{(A_i)_{i\in I}}(\,p\,) = \varphi(\,p(\,(A_i)_{i\in I})\,), \qquad \text{for } p \in \mathbb{C}\langle X_i | i \in I\rangle.$$

## 1.3 Free independence

Now that the non-commutative random variables, their probability space and their distributions have been introduced, we are going to introduce the concept of free independence, which is the analogon to independence in the CPT. To understand what free independence is useful to, let us recall what independence means in the classical sense.
When we have two classical random variables, i.e: $X_1$ and $X_2$, the joint distribution function $F_{X_1,X_2}(x_1, x_2)$ and the the joint probability density function $f_{X_1,X_2}(x_1, x_2)$ provide information on how the random variables are related to each other. If $X_1$ and $X_2$ are independent, then the joint pdf is the product of the marginal ones, i.e. $f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$, and all the joint moments $\mathbb{E}\{X_1^n X_2^m\}$ can easily be calculated. Independence is equivalent to the fact that the joint moments of $X_1$ and $X_2$ are the product of the individual moments: $\mathbb{E}\{X_1^n X_2^m\} = \mathbb{E}\{X_1^n\}\mathbb{E}\{X_2^m\}$, $\forall m, n \in \mathbb{N}$.

Does it make sense to define such a property for non-commutative random variables as well? The difference lies in the possible non-commutativity of NCRVs (remember that also commutative RVs can be NCRVs).
Let us assume that we used the classical definition of independence for two NCRVs $A_1$ and $A_2$. For joint moments of two factors, this property would be useful in the same way as in CPT, we could say that $\varphi(A_1^m A_2^n) = \varphi(A_1^m)\varphi(A_2^n)$. But for more complex joint moments the property would not help very much. This can be seen on the simple example of the product $A_1A_2A_1A_2$. The product $A_1A_2$ is not independent of itself and since $A_1A_2A_1A_2$ does not need to equal $A_1^2A_2^2$, the classical independence property would not help to simplify $\varphi(A_1A_2A_1A_2)$, i.e. to write it in form of a product of moments of $A_1$ and $A_2$. The definition of independence for NCRVs, that was introduced for the first time by Voiculescu enables us to compute all non-commutative joint moments of NCRVs. This definition of independence is referred to as *free independence.*

**Definition 1.3.1:** *Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space and $\mathcal{I}$ a set of indices.*

- *A* family of subalgebras $(\mathcal{A}_i)_{i\in\mathcal{I}}$ *is called* free, *if $\varphi(A_1 \dots A_n) = 0$ whenever $A_j \in \mathcal{A}_{i(j)}$ with $i(j) \neq i(j+1)$ $(1 \leq j \leq n-1)$ and $\varphi(A_k) = 0$ $(1 \leq k \leq n)$.*
- *A* family of subsets $(\Omega_i)_{i\in\mathcal{I}}$ *is called* free *if the subalgebras $\mathcal{A}_i$ generated by $\Omega_i$ form a free family of subalgebras.*
- *A* family of random variables $(f_i)_{i\in\mathcal{I}}$ *in $A$, is called* free *if the family of subsets $(\{f_i\})_{i\in\mathcal{I}}$ is free.*

For example let us consider two NCRVs $A_1$ and $A_2$. If $A_1$ and $A_2$ are freely independent, the subalgebras $\mathcal{A}_1$ and $\mathcal{A}_2$, generated by $A_1$ and $A_2$ respectively form a free family of subalgebras. The subalgebra generated by $A_i$ has to include all the powers of $A_i$, i.e. $A_i^k$ , $k \in \mathbb{N}$, and the sums of all products with scalars. Thus, the subalgebras include all the polynomials $p(A_i) = \sum_{j=0}^{n} \alpha_j A_i^j$ with $\alpha_j \in \mathbb{C}$.
The free independence of the subalgebras $\mathcal{A}_2$ and $\mathcal{A}_2$ implies that for all polynomials

$$p_1(A_1) = \sum_{i=0}^{m} \alpha_i A_1^i, \quad p_2(A_2) = \sum_{j=0}^{n} \beta_j A_2^j,$$

if $\varphi(\, p_1(A_1)\,) = \varphi(\, p_2(A_2)\,) = 0$, then $\varphi(\, p_1(A_1)p_2(A_2)\,) = 0$.
It is not obvious at the first sight how this will enable us to compute the joint moments out of the individual ones. For a better understanding of that, let us consider the following examples.

Example 1.3.1: If $A_1$, and $A_2$ are freely independent, then

$$\varphi((A_1 - \varphi(A_1)\cdot 1)(A_2 - \varphi(A_2)\cdot 1)) = 0, \tag{1.2}$$

since $A_i - \varphi(A_i)\cdot 1 \in \mathcal{A}_i$ and $\varphi(A_i - \varphi(A_i)\cdot 1) = \varphi(A_i) - \varphi(A_i)\varphi(1) = 0$.
By developing expression (1.2), we obtain

$$\begin{aligned} \varphi(A_1A_2 - \varphi(A_1)A_2 - \varphi(A_2)A_1 + \varphi(A_1)\varphi(A_2)\cdot 1) &= 0, \\ \varphi(A_1A_2) - \varphi(A_1)\varphi(A_2) - \varphi(A_2)\varphi(A_1) + \varphi(A_1)\varphi(A_2)\varphi(1) &= 0. \end{aligned}$$

And finally

$$\varphi(A_1A_2) = \varphi(A_1)\varphi(A_2).$$

■

Example 1.3.2: For non-commutative products with 3 terms $\varphi(A_1A_2A_1)$, we use again the fact that

$$A_i - \varphi(A_i)\cdot 1 \in \mathcal{A}_i \quad \text{and} \quad \varphi(A_i - \varphi(A_i)\cdot 1) = 0, \quad \forall \in \mathbb{N},$$

$$\varphi((A_1 - \varphi(A_1)\cdot 1)(A_2 - \varphi(A_2)\cdot 1)(A_1 - \varphi(A_1)\cdot 1))) = 0.$$

By simplification we obtain

$$\varphi(A_1A_2A_1) = \varphi(A_1^2A_2),$$

which can be transformed using the result of Example 1.3.1, since $A_1^2$ and $A_2$ are free:

$$\varphi(A_1A_2A_1) = \varphi(A_1^2)\varphi(A_2).$$

■

We get a very similar result, as the ones obtained in CPT for mixed moments of statistically independent RV. This changes for moments of higher order.

Example 1.3.3: For non-commutative products with 4 terms $\varphi(A_1A_2A_1A_2)$, the calculations are done as in the last two examples, Example 1.3.1 and Example 1.3.2.
By developing and simplifying the equation

$$\varphi((A_1 - \varphi(A_1)\cdot 1)(A_2 - \varphi(A_2)\cdot 1)(A_1 - \varphi(A_1)\cdot 1)(A_2 - \varphi(A_2)\cdot 1))) = 0,$$

we obtain

$$\varphi(A_1A_2A_1A_2) = \varphi(A_1)^2\varphi(A_2^2) + \varphi(A_1^2)\varphi(A_2^2) - \varphi(A_1)^2\varphi(A_2)^2.$$

■

Here we see, that the relation between the joint moments of free NCRVs and their individual moments is not such a simple one as it was in CPT. One thing to notice is that Example 1.3.1 and Example 1.3.2 could make us think that free NCRVs could be commutative, since $A_1A_2$ and $A_2A_1$, $A_1A_2A_1$ and $A_1^2A_2$ have the same moments of first order, but Example 1.3.3 shows us that they are not. If we compare the moments of $A_1A_2A_1A_2$ and $A_1^2A_2^2$:

$$\varphi(A_1^2A_2^2) = \varphi(A_1^2)\varphi(A_2^2) \neq \varphi(A_1A_2A_1A_2).$$

The moments would have been equal, if $A_1$ and $A_2$ commute.

## 1.4 Free central limit theorem

In a similar way as random variables, distributions and independence find their analogue in free probability, the central limit theorem has its analogous theorem: the free central limit theorem [23]. Let us recall, what the central limit theorem says [13]:

**Theorem 1.4.1:** *If $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with mean $\mathbb{E}\{X_i\} = 0$ and variance $\mathbb{E}\{X_i^2\} = 1$, then the distribution of*

$$S_n = \frac{X_1 + X_2 + \ldots + X_n}{\sqrt{n}}$$

*tends to the standard Gaussian law, i.e.*

$$\lim_{n\to\infty} f_{S_n}(x) = \frac{1}{2\pi} e^{-\frac{x^2}{2}}.$$

For non-commutative random variables, the free central limit theorem is very similar, the resulting distribution is called the semicircle law.

**THEOREM 1.4.2:** *Let $A_1, A_2, \ldots, A_n$ be a free sequence in the non-commutative probability space $(\mathcal{A}, \varphi)$. Assume that $\varphi(A_i) = 0$ and $\varphi(A_i^2) = 1$. Furthermore, assume that* $\sup_i \mid \varphi(A_i^k) \mid < +\infty$ *for all $k \in \mathbb{N}$. Then the distribution of the sequence*

$$S_n = \frac{(A_1 + A_2 + \ldots + A_n)}{\sqrt{n}}$$

*converges as $n \to \infty$ to the standard semicircle law, i.e.*

$$\lim_{n\to\infty} f_{S_n}(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2}, & \text{for } |x| < 2, \\ 0, & \text{elsewhere,} \end{cases} \tag{1.3}$$

*that is illustrated in CP 1.1.*
*Hence, the moments of $S_n$ tend to the moments of the semicircle law, which are the* catalan numbers*:*

$$\lim_{n\to\infty} \varphi(S_n^m) = \begin{cases} \frac{1}{m/2+1}\binom{m}{m/2}, & m \text{ even,} \\ 0, & m \text{ odd.} \end{cases}$$

In FPT, the semicircle law plays a very similar role as the Gaussian law in CPT.

## 1.5 Free additive convolution

In CPT, when we have two independent RVs $X_1$ and $X_2$, it is possible to get statistical information about the sum, $X_1+X_2$, from the statistical information of $X_1$ and $X_2$. As shown in Section 4.1.1 of the Appendix and in [13], there are several possibilities to do that. For example, we can calculate the moment sequence of $X_1+X_2$ over the moment sequences of $X_1$ and $X_2$ by using the binomial formula. The characteristic function of $X_1+X_2$ is given by $\Psi_{X_1+X_2} = \Psi_{X_1}\Psi_{X_2}$ and hence the pdf of $X_1+X_2$ can be calculated over the convolution $f_{X_1+X_2} = f_{X_1} * f_{X_2}$. In the FPT, we will see that it is possible to use the information about two NCRVs $A_1$ and $A_2$, that are freely independent to get information about $A_1 + A_2$, i.e. the moments or the distribution. This can be easily seen by computing the moment sequence of the sum of two freely independent NCRVs.

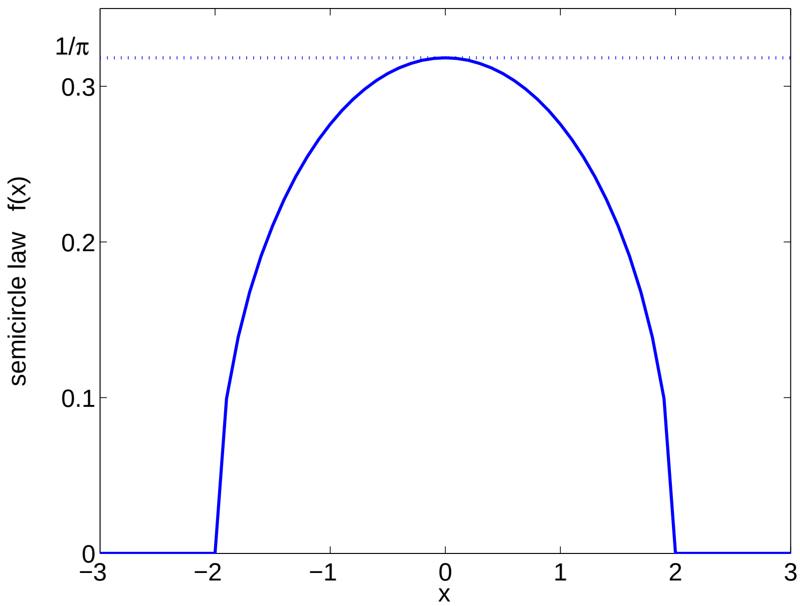



Figure 1.1: Probability density function $f(x)$ of the semicircle law, which is the pdf of the free central limit theorem (see Theorem 1.4.2). The shape ressembles a semicircle, which explains the name.

Remember the binomial formula (see Section 4.1.1) for two independent RVs $X_1$ and $X_2$:

$$\mathbb{E}\{(X_1 + X_2)^n\} = \sum_{i=0}^{n} \binom{n}{i} \mathbb{E}\{X_1^i\}\mathbb{E}\{X_2^{n-i}\}.$$

Does a similar relation exist for free NCRVs? In Example 1.5.1, this is done for the second order moments.

Example 1.5.1: Assume two freely independent NCRVs $A_1$ and $A_2$. For the second order moment, we have

$$\varphi((A_1 + A_2)^2) = \varphi(A_1^2 + A_1A_2 + A_2A_1 + A_2^2) = \varphi(A_1^2) + \varphi(A_1A_2) + \varphi(A_2A_1) + \varphi(A_2^2).$$

If $A_1$ and $A_2$ are free, then, as we know from Example 1.3.1, $\varphi(A_1A_2) = \varphi(A_2A_1) = \varphi(A_1)\varphi(A_2)$ and hence

$$\varphi((A_1 + A_2)^2) = \varphi(A_1^2) + 2\varphi(A_1)\varphi(A_2) + \varphi(A_2^2).$$

■

This computation is not so difficult for small powers, but it gets more and more complicated for higher order moments. Hence, there is no simple binomial formula as in the CPT. In addition, no characteristic function of a NCRV can be defined (see Section

4.1.1), which implies, that we cannot easily get the distribution function of the sum of two RV as in the CPT. Nevertheless a similar concept to the convolution of pdfs is introduced in the FPT, called *free convolution* of NCRV-distributions [24]. Here, no operation is actually done on the distributions. Instead, the distribution functional $\mu_A$ of the sum of two NCRVs $A = A_1 + A_2$ with distributions $\mu_{A_1}$ and $\mu_{A_2}$, respectively, is obtained by adding the R-transforms of $A_1$ and $A_2$, that will be introduced in the following Definition 1.5.1. The distribution functional $\mu_A$ is then simply called the free convolution of $\mu_{A_1}$ and $\mu_{A_2}$. First of all, the R-transform is a power series and has as coefficients the free cumulants, which are introduced in more detail in Section 4.2.2.

**Definition 1.5.1:** *Assume a NCRV $A \in \mathcal{A}$ and $(\alpha_i)_{i\in\mathbb{N}}$ its sequence of free cumulants (see Definition 4.2.6). The R-transform of $A$ is defined as the power series in $(\alpha_i)_{i\in\mathbb{N}}$ , i.e.*

$$R_A(z) := \sum_{i=0}^{\infty} \alpha_{i+1} z^i, \quad \text{with} \quad z \in \mathbb{C}. \tag{1.4}$$

The R-transform and the free cumulants linearize the free convolution similarly to the cumulant sequence in the CPT. This is shown in the following theorem.

**Theorem 1.5.2:** *If two NCRVs $A_1$ and $A_2$ in $(\mathcal{A}, \varphi)$ with distribution functionals $\mu_{A_1}$ and $\mu_{A_2}$ are free, then the R-transform of the NCRV $A_1 + A_2$ is*

$$R_{A_1+A_2}(z) = R_{A_1}(z) + R_{A_2}(z).$$

*In analogy to the CPT, the distribution $\mu$ of $A$ is called the* additive free convolution*, that is denoted as $\mu = \mu_{A_1} \boxplus \mu_{A_2}$.*

Hence, if we have the R-transform $R_{A_1}(z)$ and $R_{A_2}(z)$ of two NCRVs $A_1$ and $A_2$ that are free, we easily get the R-transform of the sum $A_1 + A_2$. In order to actually obtain the R-transform $R_A(z)$ of a NCRV $A$, we could calculate the free cumulants $\alpha_i$ from the moments $m_i = \varphi(A^i)$, and form the power series; this however is rather complicated. Usually the R-transform is computed by means of the Cauchy-transform [24, 10] (sometimes also called Stieltjes-transform [17]), which will be introduced in the following. The Cauchy-transform is a complex integral transformation, that is applied on distribution measures. It can also be written as a power series containing as coefficients the moment sequence of the distribution.

**Definition 1.5.3:** *The* Cauchy-transform *$G(z)$ of a distribution measure $F(x)$ is* *given by the complex integral transform*

$$G(z) := \int_{-\infty}^{\infty} \frac{dF(x)}{z-x}. \tag{1.5}$$

*When the probability density function $f(x)$ associated to the distribution measure is given, then*

$$G(z) = \int_{-\infty}^{\infty} \frac{f(x)}{z-x} dx. \tag{1.6}$$

The Cauchy-transform $G(z)$ can also be written in form of a power series [24, 10], with the moments $m_k = \int x^k dF(x)$ as coefficients:

$$G(z) = z^{-1} + \sum_{k=1}^{\infty} m_k z^{-k-1}. \tag{1.7}$$

This is simply done by replacing $\frac{1}{z-x}$ in (1.6) with $\sum_{k=0}^{\infty} x^k z^{-k-1}$.

For a NCRV $A$ that has the moment sequence $\varphi(A^k) = m_k$, or equivalently the distribution measure $F(x)$, the Cauchy transform can be defined as

$$G_A(z) = z^{-1} + \sum_{k=1}^{\infty} \varphi(A^k) z^{-k-1}.$$

For every distribution measure $F(x)$, the Cauchy-transform can be given [24, 10], and inversely we can recover $F(x)$ by applying

$$F(x) = \lim_{y \to 0^+} -\frac{1}{\pi} \Im\{G_F(x+iy)\} dx, \tag{1.8}$$

which implies for the associated pdf $f(x)$

$$f(x) = \lim_{y \to 0^+} -\frac{1}{\pi} \Im\{G_F(x+iy)\}. \tag{1.9}$$

Example 1.5.2: The Cauchy-tranform of the semicircle law (2.10) is given by [10]

$$G(z) = \frac{1}{2\pi} \int_{-2}^{2} \frac{\sqrt{4-x^2}}{z-x} dx = \frac{1}{2}(z - \sqrt{z^2-4}).$$

■

Example 1.5.3: The Cauchy-tranform of the pdf $f(x) = \delta(x-\lambda)$ is given by

$$G(z) = \int_{-\infty}^{+\infty} \frac{\delta(x-\lambda)}{z-x} dx = \frac{1}{z-\lambda}.$$

■

Notice that the R-transform (1.4) and the Cauchy-transform (1.7) of a NCRV have a very similar form. They both can be described as power series, one having the free cumulants as coefficients, the other the moments of the NCRV. Since the moments and the free cumulants are related to each other (see Theorem 4.2.8 and [10]), the R-transform and the Cauchy-transform are related as well. It is shown in [24] that the relation is of the form

$$R(z) = G^{-1}(z) - z^{-1}, \tag{1.10}$$

where $G^{-1}(z)$ denotes the inverse function of the Cauchy-transform $G(z)$.
On the other side, the Cauchy-transform can be obtained from the R-transform, by calculating

$$G(z) = (R(z) + z^{-1})^{-1}. \tag{1.11}$$

EXAMPLE 1.5.4: The R-transform of the semicircle law (2.10) is computed with the Cauchy-transform obtained in Example 1.5.2:

$$G^{-1}(z) = \frac{(z^2+1)}{z} \quad \Rightarrow \quad R(z) = z.$$

■

EXAMPLE 1.5.5: The R-transform of the pdf $f(x) = \delta(x - \lambda)$ using the result of Example 1.5.3 is given by

$$G^{-1}(z) = \frac{1}{z} + \lambda \quad \Rightarrow \quad R(z) = \lambda.$$

■

## 1.6 Free multiplicative convolution

The free independence of NCRVs enables us not only to compute the distribution or the moments of the sum of two NCRVs but also of the product of two NCRVs. In analogy to the additive convolution, the multiplicative convolution is introduced [24]. In analogy to the additive convolution introduced in the previous section, a transformation is defined, the S-transform. This S-transform is not obtained over coefficients like the free cumulants.

**DEFINITION 1.6.1:** *The S-transform of a NCRV is defined such that for two NCRVs $A_1$ and $A_2$ with distributions functionals $\mu_{A_1}$ and $\mu_{A_2}$ the S-transform of the product $A_1A_2$ satisfies:*

$$S_{A_1A_2}(z) = S_{A_1}(z)S_{A_2}(z). \tag{1.12}$$

*The distribution $\mu$ of $A_1A_2$ is called the free multiplicative convolution and this is denoted as $\mu = \mu_{A_1} \boxtimes \mu_{A_2}$.*

The S-transform, similarly to the R-transform can be computed over the Cauchy-transform. But first another intermediate transformation $\Upsilon(z)$ needs to be defined:

$$\Upsilon(z) = \int \frac{zx}{1-zx} dF(x) = \int \frac{zx}{1-zx} f(x) dx. \tag{1.13}$$

The transform $\Upsilon(z)$ can also be written as a power series:

$$\Upsilon(z) = \sum_{k=1}^{\infty} m_k z^k.$$

The expression is very close to the Cauchy transform, hence a rather simple relation between $\Upsilon(z)$ and $G(z)$ is given by

$$\Upsilon(z) = z^{-1} G(z^{-1}) - 1, \tag{1.14}$$

and inversily

$$G(z) = z^{-1}(\Upsilon(z^{-1}) + 1). \tag{1.15}$$

Finally as shown in [24], the S-transform of a NCRV $A$ with distribution measure $F(x)$ and associated pdf $f(x)$ is given by

$$S(z) = \frac{1+z}{z} \Upsilon^{-1}(z), \tag{1.16}$$

where $\Upsilon^{-1}(z)$ is the inverse function of $\Upsilon(z)$.

Example 1.6.1: For the semicircle law (2.10), the S-transform is computed over the Cauchy-transform obtained in Example 1.5.2:

$$\begin{aligned}
\Upsilon(z) &= z^{-1}(G(z^{-1}) - z) = \frac{1}{2z}\left(\frac{1}{z} - \sqrt{\frac{1}{z^2} - 4} - 2z\right), \\
\Upsilon^{-1}(z) &= \frac{\sqrt{z}}{z+1}, \\
S(z) &= \frac{z+1}{z}\Upsilon^{-1}(z) = \frac{1}{\sqrt{z}}.
\end{aligned}$$

■

Example 1.6.2: The S-transform of the pdf $f(x) = \delta(x - \lambda)$ using the result of

Example 1.5.3 is obtained by calculating

$$\begin{aligned}\Upsilon(z) &= \frac{\lambda z}{1-\lambda z}, \\ \Upsilon^{-1}(z) &= \frac{1}{\lambda}\frac{w}{1+w}, \\ S(z) &= \frac{1}{\lambda}.\end{aligned}$$

■

## 1.7 Summary of free convolutions

### 1.7.1 Additive free convolution

In the following, the whole procedure of the additive free convolution will be shown, step by step. First it will be applied generally on freely independent NCRVs $A_1$ and $A_2$ and then for $A_1$ and $A_2$ having special distributions.

Step 1: **Getting the distribution of $A_1$ and $A_2$** (see Section 1.2):

| | |
|---|---|
| distribution functional $\mu_{A_1}$<br>distribution measure $F_{A_1}(x)$<br>associated pdf $f_{A_1}(x)$ | distribution functional $\mu_{A_2}$<br>distribution measure $F_{A_2}(x)$<br>associated pdf $f_{A_2}(x)$ |

Step 2: **Computing the Cauchy-transform** (see (1.6)):

| | |
|---|---|
| $G_{A_1}(z) = \int \frac{f_{A_1}(x)}{z-x}dx$ | $G_{A_2}(z) = \int \frac{f_{A_2}(x)}{z-x}dx$ |

Step 3: **Computing the R-transform** (see (1.10)):

| | |
|---|---|
| $R_{A_1}(z) = G_{A_1}^{-1}(z) - z^{-1}$ | $R_{A_2}(z) = G_{A_2}^{-1}(z) - z^{-1}$ |

Step 4: **R-transform of** $A = A_1 + A_2$ (see Theorem 1.5.2 ):

$$R_A = R_{A_1} + R_{A_2}$$

Step 5: **Getting back the Cauchy-transform** (see (1.11)):

$$G_A(z) = \left(R_A(z) + z^{-1}\right)^{-1}$$

Step 6: **Getting back the pdf** (see (1.9)):

$$f(x) = \lim_{y \to 0^+} \Im\{G(x+iy)\}$$

EXAMPLE 1.7.1: Assume two NCRVs $A_1$ and $A_2$ as in Example 1.5.2 and Example 1.5.3. As it will be shown in Section 2.2.4, there exist NCRVs with these distributions that are free. To get the distribution of the sum of these NCRVs, $A = A_1 + A_2$, steps 1 to 6 are followed.

Step 1: **Getting the distribution of** $A_1$ **and** $A_2$ (see (2.10)):

| | |
|---|---|
| $f_{A_1}(x) = \delta(x - \lambda)$ | $f_{A_2}(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4 - x^2} & \text{for } \lvert x \rvert < 2 \\ 0 & \text{elsewhere} \end{cases}$ |

Step 2: **Computing the Cauchy-transform** (see Example 1.5.2 and 1.5.3):

| | |
|---|---|
| $G_{A_1}(z) = \frac{1}{z - \lambda}$ | $G_{A_2}(z) = \frac{1}{2}\left(z^2 - \sqrt{z^2 - 4}\right)$ |

Step 3: **Computing the R-transforms** (see Example 1.5.4 and 1.5.5):

| | |
|---|---|
| $R_{A_1}(z) = \lambda$ | $R_{A_2}(z) = z$ |

Step 4: **R-transform of** $A = A_1 + A_2$ :

$$R_A = z + \lambda$$

Step 5: **Getting back the Cauchy-transform**:

$$G_A(z) = \frac{1}{2}\left((z-\lambda) - \sqrt{(z-\lambda)^2 - 4}\right)$$

Step 6: **Getting back the pdf**:

$$f(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-(x-\lambda)^2} & \text{for } |x-\lambda| < 2 \\ 0 & \text{elsewhere} \end{cases}$$

For $\lambda = 1/2$, the resulting pdf $f(x)$ is plotted in Figure 1.7.1. It is a semicircle law, as in Figure 1.1 shifted with a factor $\lambda$.

■

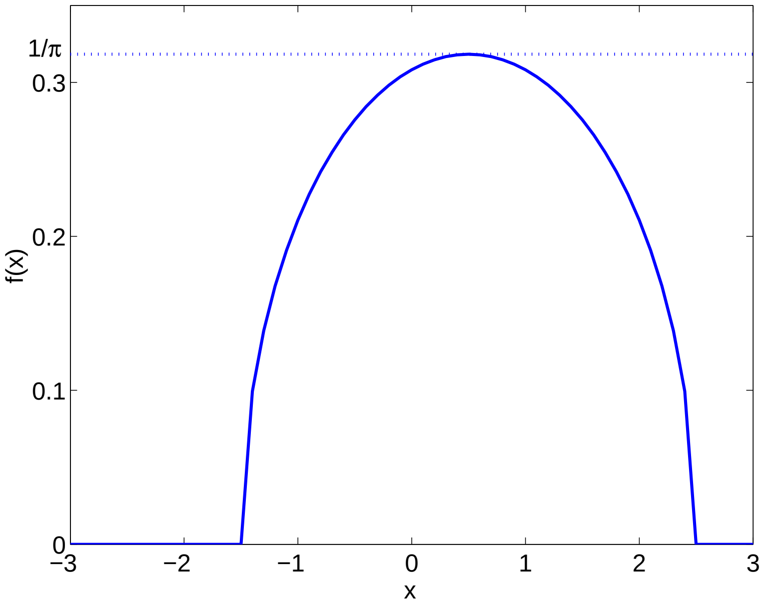



Figure 1.2: Resulting pdf $f(x)$ of Example 1.7.1 for $\lambda = 1/2$. It is the pdf of the semicircle law (see Figure 1.1) shifted with a factor $\lambda$.

### 1.7.2 Multiplicative free convolution

Here the whole procedure of the multiplicative free convolution is shown step by step on general free NCRVs $A_1$ and $A_2$ and then for NCRVs with special distributions.

Step 1: **Getting the distribution of $A_1$ and $A_2$**:

| | |
|---|---|
| distribution functional, $\mu_{A_1}$<br>distribution measure, $F_{A_1}(x)$<br>associated pdf. $f_{A_1}(x)$ | distribution functional, $\mu_{A_2}$<br>distribution measure, $F_{A_2}(x)$<br>associated pdf. $f_{A_2}(x)$ |

Step 2: **Computing the Cauchy-transform**:

| | |
|---|---|
| $G_{A_1}(z) = \int \frac{f_{A_1}(x)}{z-x} dx$ | $G_{A_2}(z) = \int \frac{f_{A_2}(x)}{z-x} dx$ |

Step 3: **Computing the intermediate transform**:

| | |
|---|---|
| $\Upsilon_{A_1}(z) = z^{-1}G_{A_1}(z^{-1}) - 1$ | $\Upsilon_{A_2}(z) = z^{-1}G_{A_2}(z^{-1}) - 1$ |

Step 4: **Computing the S-transforms:**

| | |
|---|---|
| $S_{A_1}(z) = \frac{z+1}{z}\Upsilon_{A_1}^{-1}(z)$ | $S_{A_2}(z) = \frac{z+1}{z}\Upsilon_{A_2}^{-1}(z)$ |

Step 5: **S-transform of $A = A_1A_2$:**

$$S_A = S_{A_1}S_{A_2}$$

Step 6: **Getting back the intermediate transform :**

$$\Upsilon_A(z) = \left(\frac{z}{z+1}S_A(z)\right)^{-1}$$

Step 7: **Getting back the Cauchy-transform:**

$$G_A(z) = z^{-1}(\Upsilon_A(z^{-1}) + 1)$$

Step 8: **Getting back the pdf:**

| |
|---|
| $f(x) = \lim_{y\to 0^+} \Im\{G(x+iy)\}$ |

EXAMPLE 1.7.2: In analogy to Example 1.7.1, assume two NCRVs $A_1$ and $A_2$ as in Example 1.5.2 and Example 1.5.3. In the following, the distribution of the product of these NCRVs, $A = A_1 A_2$, will be obtained by following steps 1 to 8 introduced before.

Step 1: **Getting the distribution of $A_1$ and $A_2$:**

| | |
|---|---|
| $f_{A_1}(x) = \delta(x-\lambda)$ | $f_{A_2}(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2} & \text{for } \lvert x\rvert < 2 \\ 0 & \text{elsewhere} \end{cases}$ |

Step 2: **Computing the Cauchy-transform**:

| | |
|---|---|
| $G_{A_1}(z) = \frac{1}{z-\lambda}$ | $G_{A_2}(z) = \frac{1}{2}(z - \sqrt{z^2-4})$ |

Step 3: **Computing the intermediate transform**:

| | |
|---|---|
| $\Upsilon_{A_1}(z) = \frac{\lambda z}{1-\lambda z}$ | $\Upsilon_{A_2}(z) = \frac{1-\sqrt{1-4z^2}}{2z^2}$ |

Step 4: **Computing the S-transforms**:

| | |
|---|---|
| $S_{A_1}(z) = \frac{1}{\lambda}$ | $S_{A_2}(z) = \frac{1}{\sqrt{z}}$ |

Step 5: **S-transform of** $A = A_1 A_2$:

$$S_A = \frac{1}{\lambda z}$$

Step 6: **Getting back the intermediate transform**:

$$\Upsilon_A(z) = \frac{1-\sqrt{1-4(\lambda z)^2}}{2(\lambda z)^2} - 1$$

Step 7: **Getting back the Cauchy-transform**:

$$G_A(z) = \frac{1}{2\lambda^2}(z - \sqrt{z^2 - 4\lambda^2})$$

Step 8: **Getting back the pdf**:

$$f(x) = \begin{cases} \frac{1}{2\pi\lambda^2}\sqrt{4\lambda^2 - x^2}, & \text{for } |x| < 2\lambda, \\ 0, & \text{elsewhere.} \end{cases}$$

For $\lambda = 1/2$, the resulting pdf $f(x)$ is plotted in Figure 1.7.1. It is the semicircle law (as in Figure 1.1) stretched with a factor $\lambda$.

■

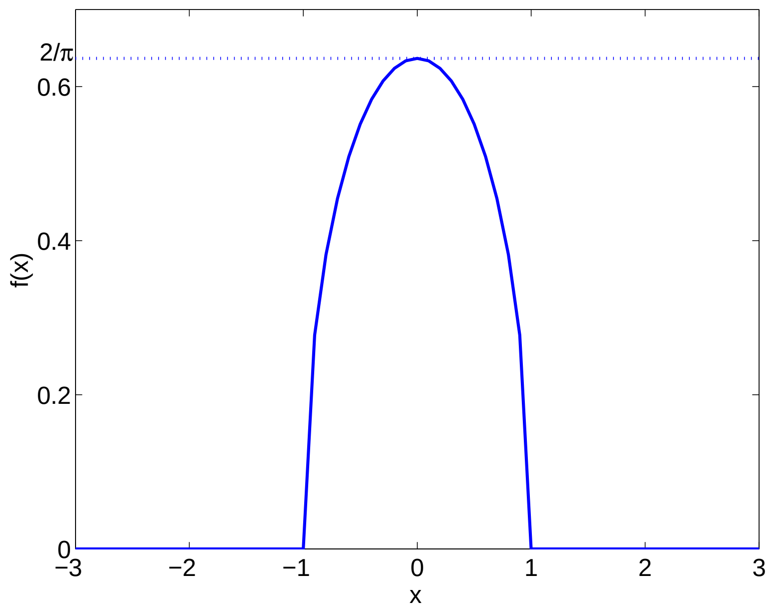



Figure 1.3: Resulting pdf $f(x)$ of Example 1.7.2 for $\lambda = 1/2$. It is the pdf of the semicircle law (see Figure 1.1) stretched with a factor $\lambda$.

## 1.8 Analogy between classical and free probability

| Classical Probability Theory | Free Probability Theory |
|---|---|
| • Probability space<br>$(\Omega, \Sigma)$ | • Non-commutative probability space<br>$(\mathcal{A}, \varphi)$ |
| • Expectation functional<br>$\mathbb{E}\{\cdot\}$ | • State<br>$\varphi(\cdot)$ |
| • Random variable<br>$X : (\Omega, \Sigma) \to (\mathbb{R}, \mathcal{B})$ | • Non-commutative random variable<br>$A \in \mathcal{A}$ |
| | • Distribution functional<br>$\mu_A(x^k) = \varphi(A^k)$ |
| • Distribution<br>$F(x) = \mathbb{P}\{X < x\}$ | • Distribution measure<br>$F_A(x)$ |
| • Pdf of a RV<br>$f(x) = \frac{dF(x)}{dx}$ | • Pdf of a NCRV<br>$f_A(x) = \frac{dF_A(x)}{dx}$ |
| • Moments<br>$\mathbb{E}\{X^k\}$, $k \in \mathbf{N}$ | • Non-commutative moments<br>$\varphi(A^k)$ |
| • Independence | • Free independence |
| • Convolution | • Free convolution<br>Additive and Multiplicative |

# Chapter 2

# Random matrices

Random matrices (RMs) are interesting for communication engineering since they appear in stochastic channel models, for example in MIMO systems. RMs have been topic of research already in the 1920's in the multivariate statistical analysis, where Wishart concentrated himself in fixed size matrices with Gaussian entries [6]. This approach uses the tools of classical probability theory and has lead to interesting results about the joint distribution of the entries of so called Wishart matrices $\mathbf{HH}^H$, where $\mathbf{H}$ is a RM with Gaussian entries, and where $\cdot^H$ denotes the Hermitian operator. Marginal distribution of eigenvalues and condition numbers of Gaussian and of Wishart matrices can be found in Edelman's work of the 1980's [6]. The multivariate statistical analysis approach give a very complete description of the matrix, but it remains very difficult since most of the time we deal with multidimensional integrals (see Section 2.1), which are difficult to solve. Wigner in the 1950's [6], motivated by nuclear physics applications, thought for the first time of evaluating the limiting spectrum of large matrices. He came to very interesting closed results, which in fact founded the so called random matrix theory. From the 1960's to the 1980's the limiting spectrum of certain RMs, specially of Wishart matrices were independently found and proved by several mathematicians. The main ones were Marčenko-Pastur [14] and Silverstein [19]. These results became very important and popular for communication engineers, since the capacity formula derived by Telatar [21] or Foschini [7] includes the eigenvalues of Wishart matrices. The free probability (FPT) approach to random matrices was introduced by Voiculescu in the 1980's, as he realized, that RMs represented asymptotic realizations of non-commutative random variables. He also proved that a large family of RMs, actually the one with iid entries becomes free as the dimension tends to infinity. This approach is a very promising one, since from the already known results about asymptotic spectra of RMs, new results about products and sums of them can be derived. The CPT approach to RMs is presented in Section 2.1, followed by the FPT approach in Section 2.2, which includes the embedding of RMs into the FPT framework and the presentation of some main results of random matrix theory.

## 2.1 Classical probability theory and random matrices

In the following it will be described, how RMs can be treated with classical probability theory. A joint distribution of the entries will be introduced and we will see how eigenvalue distributions can be obtained.

RMs $\mathbf{X} \in \mathbb{C}^{m \times n}$ are composed of random entries; they can be seen as a set, or as a family of random variables, which are the entries of the matrix. Hence, such a random matrix $\mathbf{X}$ can be completely described by the joint pdf of its entries $\mathbf{X}_{ij}$, $1 \leq i \leq m$, $1 \leq j \leq n$.

If the entries $\mathbf{X}_{ij}$ are statistically independent, it is not difficult to obtain the joint pdf of the entries; we just need the individual pdf of every entry and compute the product of them. Let $f_{\mathbf{X}_{11},\ldots,\mathbf{X}_{mn}}(x_{11},\ldots,x_{mn})$ be the joint pdf of the entries and $f_{\mathbf{X}_{ij}}(x_{ij})$ the pdf of one entry $\mathbf{X}_{ij}$. We obtain

$$f_{\mathbf{X}_{11},\ldots,\mathbf{X}_{mn}}(x_{11},\ldots,x_{mn}) = \prod_{1 \leq i \leq m, 1 \leq j \leq n} f_{\mathbf{X}_{ij}}(x_{ij}). \tag{2.1}$$

It gets even simpler, if the entries are independent and identically distributed (iid), like in the following example.

Example 2.1.1: Let $\mathbf{X}$ be a random matrix with iid Gaussian entries, with zero mean and variance $\sigma^2$. The joint distribution of the entries is then given by

$$\begin{aligned} f_{\mathbf{X}_{11},\ldots,\mathbf{X}_{mn}}(x_{11},\ldots,x_{mn}) &= \prod_{1 \leq i \leq m, 1 \leq j \leq n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{x_{ij}^2}{2\sigma^2}\right) \\ &= \frac{1}{(\sqrt{2\pi\sigma^2})^{mn}} \exp\left(\sum_{1 \leq i \leq m, 1 \leq j \leq n} \frac{x_{ij}^2}{2\sigma^2}\right) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{mn}{2}}} \exp\left(\frac{1}{2\sigma^2}\mathrm{trace}(\mathbf{X}\mathbf{X}^T)\right). \end{aligned} \tag{2.2}$$

If we assume that the matrix is self-adjoint and of dimension $n \times n$, the pdf reduces to

$$f_{\mathbf{X}_{11},\ldots,\mathbf{X}_{nn}}(x_{11},\ldots,x_{nn}) = \frac{1}{(2\pi\sigma^2)^{\frac{n^2}{2}}} \exp\left(\frac{1}{2\sigma^2}\mathrm{trace}(\mathbf{X}^2)\right),$$

and thus can be represented as a function $g$ depending only on the eigenvalues $\lambda_i$ of $\mathbf{X}^2$, i.e.

$$f_{\mathbf{X}_{11},\ldots,\mathbf{X}_{nn}}(x_{11},\ldots,x_{nn}) = g(\lambda_1,\lambda_2,\ldots,\lambda_n),$$

with

$$g(\lambda_1,\lambda_2,\ldots,\lambda_n) = \frac{1}{(2\pi\sigma^2)^{\frac{n^2}{2}}} \exp\left(\frac{1}{2\sigma^2}\sum_{i=1}^{n} \lambda_i^2\right).$$

In this case, the joint pdf of the entries depends only on the eigenvalues $\lambda_i$ of $\mathbf{X}$.

Note, that in Example 2.1.1 the argument of the pdf depends only on the trace of $\mathbf{X}\mathbf{X}^T$. It can be easily shown, that this expression is invariant under an orthogonal transformation of $\mathbf{X}$ [10], i.e.

$$\text{trace}(\mathbf{X}'\mathbf{X}'^T) = \text{trace}(\mathbf{O}\mathbf{X}\mathbf{O}^T\mathbf{O}\mathbf{X}^T\mathbf{O}^T) = \text{trace}(\mathbf{O}\mathbf{X}\mathbf{X}^T\mathbf{O}^T) = \text{trace}(\mathbf{X}\mathbf{X}^T), \qquad (2.3)$$

where $\mathbf{O}$ denotes an orthogonal matrix. Hence, the joint pdf of the entries of the RM $\mathbf{X}$ can be said to be invariant under an orthogonal transformation of $\mathbf{X}$. When the Gaussian RM of the example above is squared and self-adjoint, the pdf depends only on the eigenvalues. This result is generalized in [6, 10]; it says that for any self-adjoint matrix, the joint pdf of the entries $f(x_{11}, x_{12}, \dots, x_{1n}, x_{22}, \dots, x_{nn})$ is invariant with respect to (w.r.t) any orthogonal transformation if and only if the pdf can be expressed in terms of the eigenvalues. In other words, that there exist a function $g$, such that:

$$f(x_{11}, x_{12}, \dots, x_{1n}, x_{22}, \dots, x_{nn}) = g(\lambda_1, \lambda_2, \dots, \lambda_n). \qquad (2.4)$$

It is interesting to see, that the eigenvalues come into the scene. The eigenvalues of a RM are random as well and many applications in communications need statistical information about them (see Chapter 3). This statistical information could be for example the joint pdf of the eigenvalues. The last result about the joint pdf of the entries of squared matrices expressed in terms of eigenvalues is used in a theorem in [10, 6], that enables us to compute the joint pdf of the eigenvalues with help of the joint pdf of the entries.

Example 2.1.2: If a selfadjoint matrix $\mathbf{X} \in \mathbb{C}^{n\times n}$ has a joint pdf for the entries, that is invariant under orthogonal transformations, and hence that fullfills condition (2.4), then the joint pdf of the ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ is given by [6, 10]

$$f_{\lambda_1,\lambda_2,\dots,\lambda_n}(\lambda_1, \dots, \lambda_n) = \frac{\pi^{n(n+1)/4}}{\prod_{j=1}^{n}\Gamma(j/2)} g(\lambda_1, \dots, \lambda_n) \prod_{i<j} \mid \lambda_i - \lambda_j \mid .$$

■

Example 2.1.3: For a self-adjoint matrix $\mathbf{X} \in \mathbb{C}^{n\times n}$ with iid Gaussian entries, with zero mean and variance $\sigma^2$, the joint pdf for the increasingly ordered eigenvalues $f_o(\lambda_1, \dots, \lambda_n)$ [6] is

$$f_o(\lambda_1, \dots, \lambda_n) = \frac{\pi^{n(n+1)/4}}{\prod_{j=1}^{n}\Gamma(j/2)} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\lambda_i^2\right) \prod_{i<j} \mid \lambda_i - \lambda_j \mid .$$

For the non-ordered eigenvalues, the pdf $f_n(\lambda_1, \dots, \lambda_n)$ results in

$$f_n(\lambda_1, \dots, \lambda_n) = \frac{\pi^{n(n+1)/4}}{n!\prod_{j=1}^{n}\Gamma(j/2)} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\lambda_i^2\right) \prod_{i<j} \mid \lambda_i - \lambda_j \mid .$$

These joint pdfs of the eigenvalues can now be used to get other kind of statistical information. For example by computing the marginal pdf over the ordered joint distribution, we can get the pdf of the largest eigenvalue $f_{\lambda_1}(\lambda)$,

$$f_{\lambda_1}(\lambda) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_o(\lambda, \lambda_2, \ldots, \lambda_n) d\lambda_2 \ldots d\lambda_n, \tag{2.5}$$

or of the $m$th eigenvalue $m \leq n$,

$$f_{\lambda_m}(\lambda) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_o(\lambda_1, \ldots, \lambda_{m-1}, \lambda, \lambda_{m+1}, \ldots, \lambda_n) d\lambda_1 \ldots \lambda_{m-1}\lambda_{m+1} \ldots d\lambda_n. \tag{2.6}$$

In Edelman's work [6] we can find calculations of these marginal densities. Edelman has found some interesting results about the marginal density of smallest and largest eigenvalues of Wishart matrices, about condition numbers and also some asymptotic results derived from the joint probability density function of the entries of a Wishart matrix. Nevertheless, the results remain very complicated and difficult to process.

## 2.2 Free probability applied on random matrices

In Chapter 1 we have discussed FPT and in Section 2.1 we have demonstrated, RMs can be treated with CPT. In the following we will see how FPT can be used to work with RMs.

### 2.2.1 Random matrices seen as NCRVs and their distributions

We have seen in the previous Section, that in CPT a RM is a family of random variables, which are the entries of the matrix. In FPT a RM is seen as just one random object. By defining the appropriate non-commutative probability space, such that RMs are elements of it, they can be considered as NCRVs. As it has already been mentioned in Example 1.1.2 and Example 1.2.2, for $n \times n$-dimensional random matrices, we introduce the non-commutative probability space, the state, the distribution functional, and the distribution measure.

#### The non-commutative probability space

The set of random matrices $\mathbf{A} \in \mathbb{C}^{n\times n}$ forms an algebra. Hence we can define the *non-commutative probability space* (NCPS, see Definition 1.1.1) as the algebra $\mathcal{M}$ formed by the random matrices $\mathbf{A} \in \mathbb{C}^{n\times n}$. The NCPS can only be formed by squared matrices of a certain dimension, since the conditions of an algebra [12] have to be fullfilled. The unit in this space is the deterministic unit matrix $\mathbf{I}_n$. Following Definition 1.1.1, the NCPS has to be endowed with a state.

#### The state

As introduced in Definition 1.1.1, the *state* is a linear functional $\phi$ defined on the NCPS, that respects $\phi(\mathbf{I}_n) = 1$. Any functional can actually be used, that respects this condition. For RMs one usually uses the normalized trace $\mathrm{tr}_n : \mathcal{M} \to \mathbb{C}$ or the expectation of it $\tau_n : \mathcal{M} \to \mathbb{C}$.

- The normalized trace is defined as

$$\phi(\mathbf{A}) = \mathrm{tr}_n(\mathbf{A}) = \frac{1}{n}\mathrm{trace}(\mathbf{A}) = \frac{1}{n}\sum_{i=0}^{n}\mathbf{A}_{ii} = \frac{1}{n}\sum_{i=0}^{n}\lambda_i,$$

  with $\lambda_i$ denoting the eigenvalues of $\mathbf{A}$.

- Voiculescu introduced the expectation of the normalized trace, $\tau_n : \mathcal{M} \to \mathbb{C}$ as the state in the NCPS of RMs [24]:

$$\phi(\mathbf{A}) = \tau_n(\mathbf{A}) = \mathbb{E}\{\mathrm{tr}_n(\mathbf{A})\} = \frac{1}{n}\sum_{i=0}^{n}\mathbb{E}\{\lambda_i\}.$$

Using $\tau_n$ as the state, Voiculescu proved his theorems for random matrices [23]. Hiai showed [10] that the theorems were also true in expectation for the normalized trace. As we will see in the following Sections, $\mathrm{tr}_n$ leads to results which are easier to interpret for our purposes in digital communications.

EXAMPLE 2.2.1: We take a sample of a random self-adjoint matrix $\mathbf{A} \in \mathbb{C}^{10\times 10}$. The entries are iid complex Gaussian with zero mean and unit variance. For the eigenvalues of $\mathbf{A}$ we obtain:

-4.1223, -3.1749, -1.6832, -1.4228, -0.2236, 0.6959, 0.9750, 1.1724, 3.2743, 4.6783

The moments of $\mathbf{A}$ with respect to the normalized trace are then given by

$$\phi(\mathbf{A}) = 0.1691, \quad \phi(\mathbf{A}^2) = 67.3978, \quad \phi(\mathbf{A}^3) = 30.6543, \quad \phi(\mathbf{A}^4) = 999.4939.$$

These values are random, they vary from sample to sample. ■

### The distribution functional

The *distribution functional* (see Definition 1.2.1) $\mu_{\mathbf{A}} : \mathbb{C}\langle X\rangle \to \mathbb{C}$ of a RM $\mathbf{A}$ is

- given by
$$\mu_{\mathbf{A}}(x^k) = \mathrm{tr}_n(\mathbf{A}^n) = \frac{1}{n}\sum \lambda_i^k,$$
for the normalized trace,

- or
$$\mu_{\mathbf{A}}(x^k) = \tau_n(\mathbf{A}^n) = \frac{1}{n}\sum \mathbb{E}\{\lambda_i^k\},$$
for Voiculescu's state.

### The distribution measure and its pdf

The *distribution measure* (see Definition 1.2.2) $F_{\mathbf{A}}(x)$ of a RM $\mathbf{A}$ and its associated *pdf* $f_{\mathbf{A}}(x)$ are the functions that satisfy

$$\phi(\mathbf{A}^n) = \int x^n dF_{\mathbf{A}}(x) = \int x^n f_{\mathbf{A}}(x) dx. \tag{2.7}$$

- For $\phi = \mathrm{tr}_n$, $F^e_{\mathbf{A}}(x)$ is called the *empirical eigenvalue distribution*, that can be defined as the ratio of the number of eigenvalues that are smaller than $x$ over the total number of eigenvalues, i.e.

$$F^e_{\mathbf{A}}(x) = \frac{1}{n}|\{\lambda_i : \lambda_i < x\}| = \frac{1}{n}\sum_{i=1}^{n} \tau(x - \lambda_i), \tag{2.8}$$

where $\tau(x)$ is the unit step function:

$$\tau(x) = \begin{cases} 1, & \text{for } x \geq 0, \\ 0, & \text{for } x < 0. \end{cases}$$

By differentiating $F^e_{\mathbf{A}}(x)$, we obtain the corresponding pdf:

$$f^e_{\mathbf{A}}(x) = \frac{1}{n}\sum_{i=1}^{n} \delta(x - \lambda_i), \tag{2.9}$$

where $\delta(x)$ is the usual Dirac function. This result is obtained, by replacing $f_{\mathbf{A}}$ with $f^e_{\mathbf{A}}$ in relation (2.7).

- For $\phi = \tau_n$, $F^m_{\mathbf{A}}(x)$ is called the *mean eigenvalue distribution*. With $F_{\lambda_\mathbf{i}}(x)$ and $f_{\lambda_\mathbf{i}}(x)$ denoting the distribution and the pdf of the $i$th eigenvalue $\lambda_i$, respectively, it can be seen, that $F_{\mathbf{A}}(x)$ is the mean of these eigenvalue distributions (this also holds for $f_{\mathbf{A}}(x)$), i.e.

  $$F^m_{\mathbf{A}}(x) = \frac{1}{n}\sum_{i=1}^{n} F_{\lambda_i}(x), \qquad f^m_{\mathbf{A}}(x) = \frac{1}{n}\sum_{i=1}^{n} f_{\lambda_i}(x).$$

  We can be confirm this result by replacing $f_{\mathbf{A}}$ with $f^m_{\mathbf{A}}$ in relation (2.7), by using the fact that

  $$\mathbb{E}\{\delta(x - \lambda_i)\} = \int \delta(x - \lambda_i) f_{\lambda_i}(\lambda_i) d\lambda_i = f_{\lambda_i}(x).$$

  We conclude that the mean eigenvalue distribution and the associated pdf are the expectations of the empirical distribution and the corresponding pdf with respect to the eigenvalues

  $$F^m_{\mathbf{A}}(x) = \mathbb{E}\{F^e_{\mathbf{A}}\}, \qquad f_{\mathbf{A}}(x) = \mathbb{E}\{f^e_{\mathbf{A}}\}.$$

Most of the results, that we are going to use in the following have been proved in the NCPS using $\tau_n$ as the state. But in general they are also valid for the state $\text{tr}_n$. The mean eigenvalue distribution is always continuous, which however is not the case for the empirical distributions: they are random and discrete. For the mean eigenvalue distribution of a random matrix it is difficult to give any interpretation. Thus, we will often consider the empirical eigenvalue distribution, as in the following example.

Example 2.2.2: With the matrix sample from Example 2.2.1, the empirical distribution can be obtained in the following way: $F_{\mathbf{A}}(x)$ is the number of eigenvalues that are smaller than $x$ divided by the total number of eigenvalues $n$. In other words:

$$\text{If } x \in [\lambda_{k-1}, \lambda_k], \qquad \text{then } F_{\mathbf{A}}(x) = \frac{k}{n}.$$

The resulting empirical eigenvalue distribution of the matrix (a $10 \times 10$ RM with iid complex Gaussian entries) is a stairfunction as shown in Figure 2.1. The pdf of this matrix is a sum of dirac pulses weighted with $1/n$, which is illustrated in Figure 2.2.

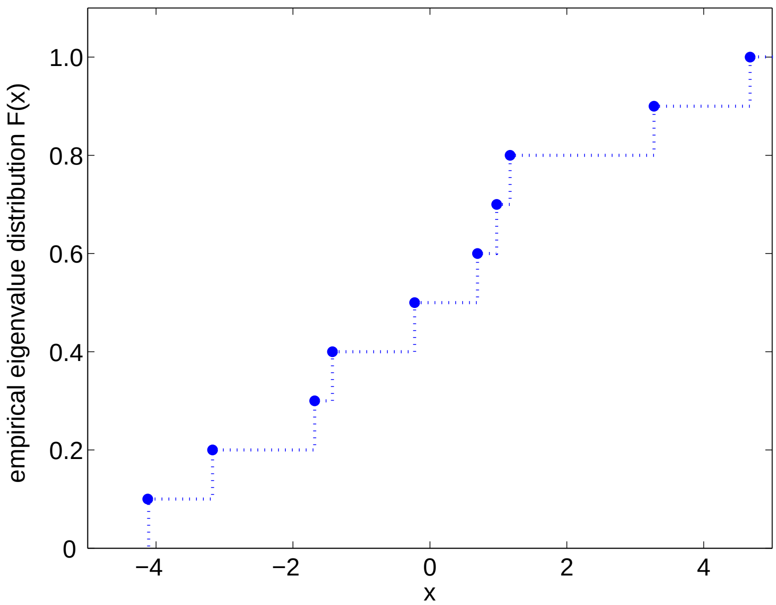



Figure 2.1: Empirical eigenvalue distribution $F(x)$ of a sample of selfadjoint $10 \times 10$ RM with iid complex Gaussian entries with zero mean and unit variance.

One additional plot, that is defined and introduced here is the *normalized histogram* of the eigenvalues of a sample of a random matrix. Consider an intervall $[x_1, x_m]$ and split it into $m$ intervals $[x_{i-1}, x_i]$ of width $\Delta = x_i - x_{i-1}$. If $x$ is in the interval $[x_{i-1}, x_i]$, the normalized histogram $h(x)$ of a $n \times n$ matrix is defined as the number of eigenvalues that lie in this intervall divided by the total number of eigenvalues $n$ times the intervall width $\Delta$:

$$h(x) = \frac{|\{\lambda, x_{i-1} < \lambda < x_i\}|}{n\Delta}, \quad \text{if } x \in [x_{i-1}, x_i].$$

The reason why $h(x)$ is defined this way, is that it yields an approximation for the empirical eigenvalue pdf. Additionally, in Section 2.2.2 it will become clear, that as $n$

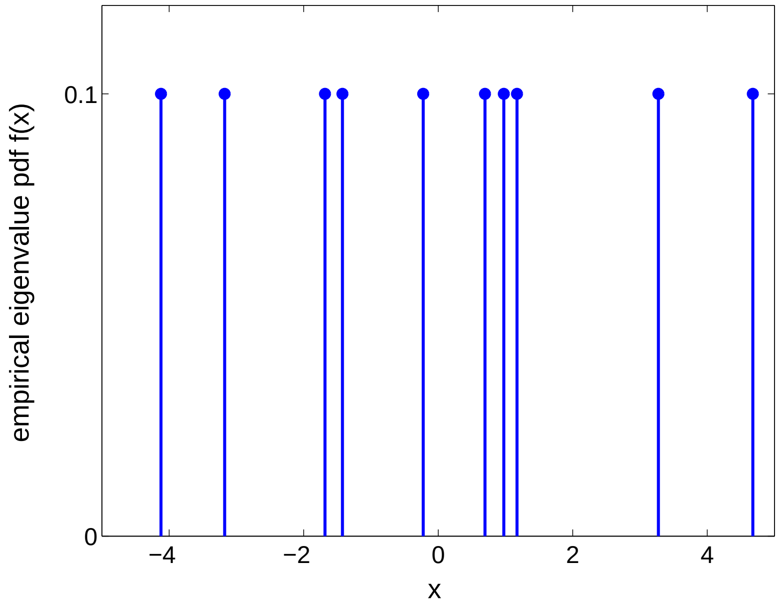



Figure 2.2: Empirical eigenvalue pdf $f(x)$ of a sample of a selfadjoint $10 \times 10$ RM with iid complex Gaussian entries with zero mean and unit variance.

and $\Delta$ tend to infinity, $h(x)$ will tend to $f(x)$. Notice, that $h(x)$ can be expressed in terms of $F(x)$

$$h(x) = \frac{F(x_i) - F(x_{i-1})}{\Delta}, \quad \text{if } x \in [x_{i-1}, x_i],$$

and also in terms of $f(x)$ using the mean value theorem [**?**]:

$$h(x) = \frac{1}{\Delta} \int_{x_{i-1}}^{x_i} f(x)dx \sim f(\xi), \quad \text{with } \xi \in [x_{i-1}, x_i].$$

Hence, $h(x)$ yields an approximation for $f(x)$. A normalized histogram, from the sample of Example 2.2.1 is illustrated in Figure 2.3.

■

### 2.2.2 Random matrices in the large limit

It would be natural to give examples of mean and empirical eigenvalue distributions for some special classes of random matrices. Unfortunately simple results cannot be given. For finite dimensional matrices, the mean eigenvalue distribution of a matrix with iid Gaussian entries could be obtained with methods from classical probability theory [6]. But, as we have seen in Section 2.1, calculations can get very complicated. Additionally, the empirical eigenvalue distribution of a finite dimensional matrix is a random stairfunction, and its empirical eigenvalue pdf is a discrete random function (see

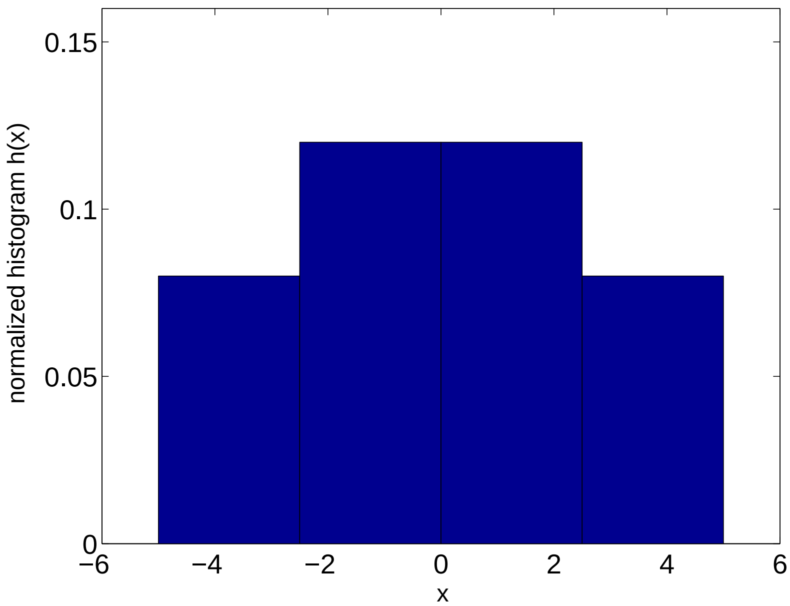



Figure 2.3: Normalized histogram $h(x)$ of the eigenvalues of a sample of a selfadjoint $10 \times 10$ RM with iid complex Gaussian entries with zero mean and unit variance.

Example 2.2.2). We cannot give much information about these random functions. The situation becomes easier, when we consider matrices in their large limit. In 1950 Wigner [?] proved for the first time, that the mean eigenvalue distribution of matrices with iid Gaussian entries tends to a distribution (the semicircle law) as the dimension tends to infinity. More theorems about large RMs have been found by mathematicians in the subsequent years [14, 19]. It is a very important result, that the empirical eigenvalue distribution of certain random matrices tend to a deterministic distribution, actually to the same mean eigenvalue distribution as the dimension of the matrix tends to infinity [10]. In the following, we present these asymptotic results. We start by introducing the concept of the *asymptotic eigenvalue distribution*.

**Definition 2.2.1:** *Let us assume a matrix* $\mathbf{A}$ *of dimension* $n$*, with eigenvalue distribution* $F_n(x)$*. If there exist a limit of* $F_n(x)$*, hence*

$$\text{if } \lim_{n\to\infty} F_n(x) = F(x),$$

*then* $F(x)$ *is called the* asymptotic eigenvalue distribution *of matrix* $\mathbf{A}$*.*
*If* $F(x)$ *is an empirical eigenvalue distribution, then* $F(x)$ *is called the* asymptotic empirical eigenvalue distribution *of* $\mathbf{A}$*. If* $F(x)$ *is the mean eigenvalue distribution, then* $F(x)$ *is called the* asymptotic mean eigenvalue distribution *of* $\mathbf{A}$*.*

We will see that in general the asymptotic empirical and the asymptotic mean eigenvalue distributions happen to converge to the same distribution [10]. The first theorem about the asymptotic eigenvalue distribution is for selfadjoint RMs. This is the semicircle law, which takes the place of the Gaussian distribution in the FPT. It can be considered the main result of large RM. It was proved for the first time by Wigner in the 1950's [?].

**Theorem 2.2.2:** *The semicircle law*
*Let* $\mathbf{S} \in \mathbb{C}^{n \times n}$ *be a self-adjoint random matrix with iid complex entries, with mean* $\mu = 0$ *and variance* $\sigma^2 = \frac{1}{n}$. *When* $n$ *tends to infinity, then the mean eigenvalue distribution and the empirical eigenvalue distribution of this matrix converge to the same asymptotic distribution* $F_{\mathbf{S}}(x)$, *whose asymptotic pdf* $f_{\mathbf{S}}(x)$ *is the semicircle law:*

$$f_{\mathbf{S}}(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4 - x^2}, & \text{for } |x| < 2, \\ 0, & \text{elsewhere }. \end{cases} \tag{2.10}$$

The name semicircle law is due to the shape of the pdf (see Figure 2.5). Recall that this distribution has already been mentioned in the context of the free central limit theorem (see Section 1.4). A very interesting thing to notice is that the distribution does not depend on the distribution of the entries of the matrix. The eigenvalues are identically distributed for any distribution of the entries (e.g. Gaussian, Rayleigh...).

Example 2.2.3: The semicircle law.
The asymptotic distribution can now be used to give an approximation of the eigenvalue distributions of finite dimensional matrices. The approximation can be illustrated by taking a sample of a large matrix, fullfilling the above conditions and plotting the empirical eigenvalue distribution as it was performed in Example 2.2.2. In Figure 2.6 (a), the empirical eigenvalue distribution of a sample of a $40 \times 40$ matrix was plotted, and in Figure 2.6 (b) the empirical eigenvalue distribution of 20 samples of $8 \times 8$ matrices is illustrated. We can already see that the distribution of the semicircle law gives a good approximation even for finite dimensional matrices. The interesting thing is that we come to a very similar result by taking one sample of a large matrix and several samples of smaller matrices.
When the dimension of a random matrix tends to infinity, then the normalized histogram of the eigenvalues, as it has been introduced in Example 2.2.2, tends to the eigenvalue pdf of the matrix. We can show this, by considering that

$$\lim_{\Delta \to 0} h(x) = \lim_{\Delta \to 0} \frac{F(x_i) - F(x_{i-1})}{\Delta} = f(x),$$

since the middle term is nothing else than the derivation of $F(x)$, which is equivalent

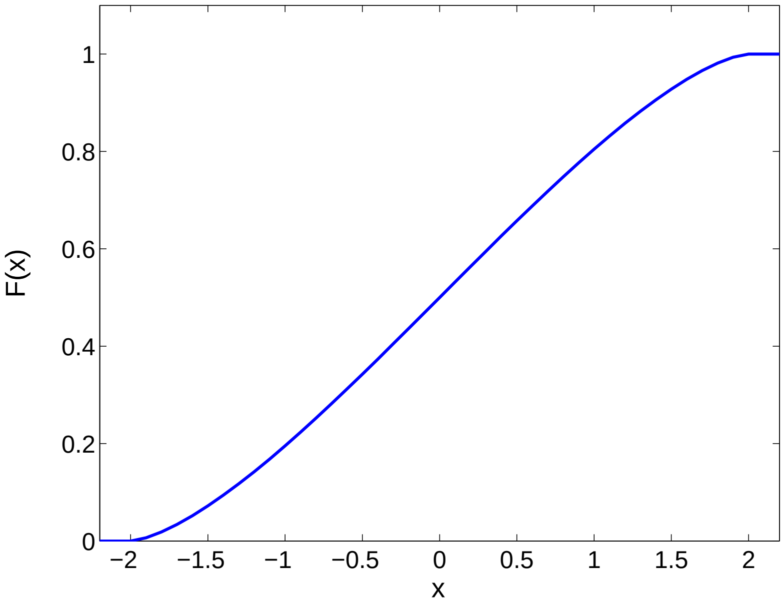



Figure 2.4: The distribution $F(x)$ of the semicircle law is the asymptotic mean eigenvalue distribution and also the asymptotic empirical eigenvalue distribution of $n \times n$ selfadjoint RMs, whose entries are iid with zero mean and variance $1/n$.

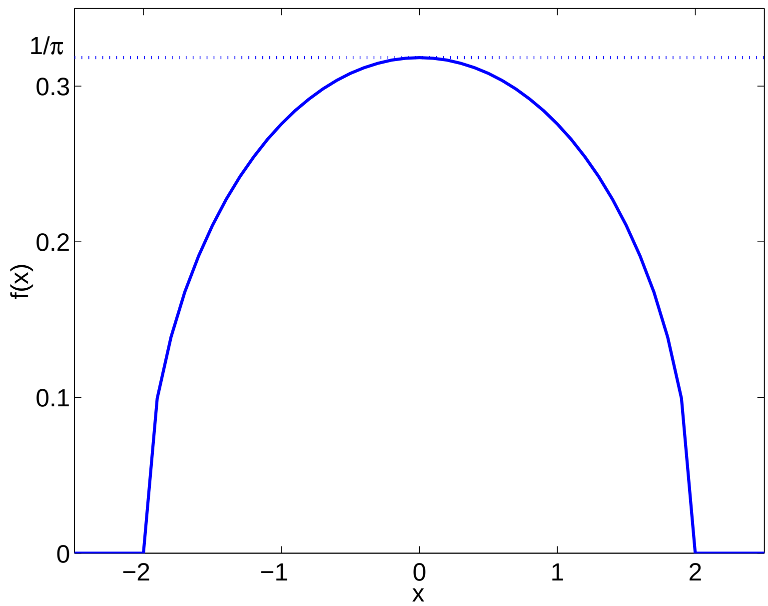



Figure 2.5: The pdf $f(x)$ of the semicircle law is the asymptotic mean eigenvalue pdf and also the asymptotic empirical eigenvalue pdf of $n \times n$ selfadjoint RMs, whose entries are iid with zero mean and variance $1/n$. As we can see in the plot, the shape of the pdf resembles a semicircle.

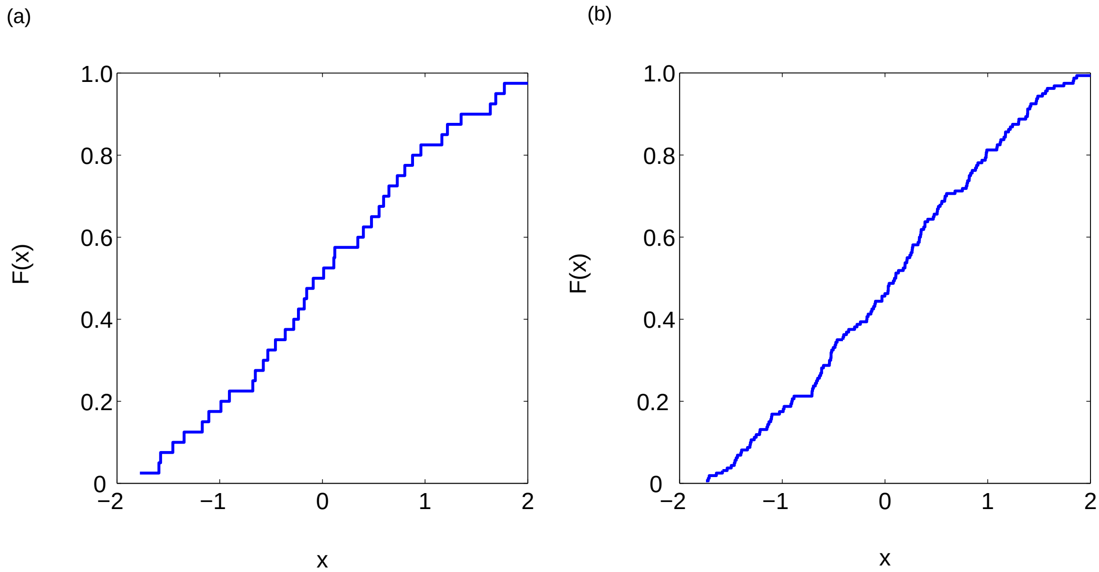



Figure 2.6: Empirical eigenvalue distributions $F(x)$ of $n \times n$ selfadjoint RMs with iid complex Gaussian entries having zero mean and variance $1/n$: (a) empirical eigenvalue distribution of a single $40 \times 40$ selfadjoint RM sample, (b) empirical eigenvalue distribution of 20 $8 \times 8$ selfadjoint RM samples. As expected by Theorem 2.2.2, (a) and (b) are well approximated by the distribution of the semicircle law.

to the pdf $f(x)$. This implies that the asymptotic eigenvalue pdf can be used to get an approximation of the histogram of the eigenvalues:

$$h(x) \sim f(x).$$

This approximation can be illustrated by the normalized histogram of RM samples, fullfilling the conditions of the semicircle law. For the same matrices as in Figure 2.6, the normalized histogram has been plotted in Figure 2.7. We can already imagine the semicircular shape.

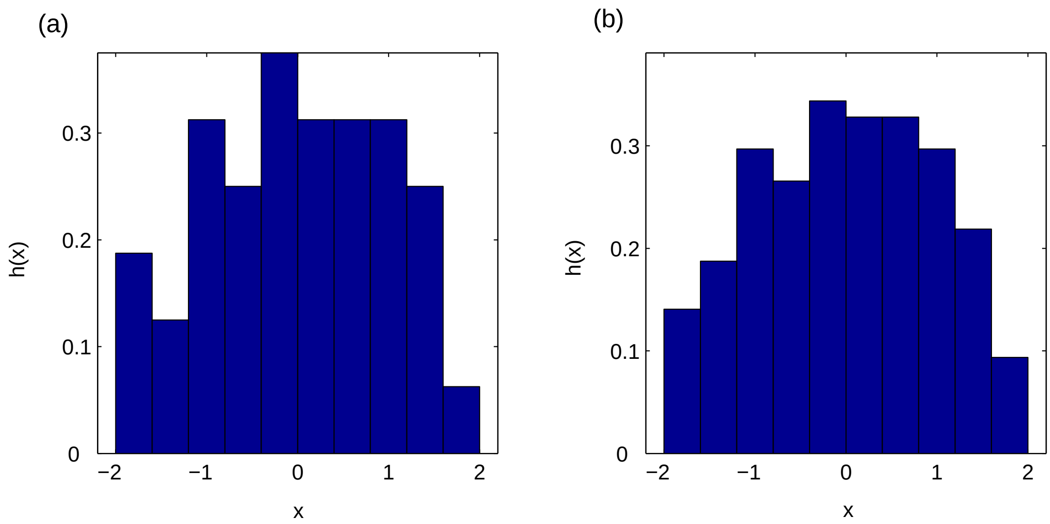



Figure 2.7: Normalized histograms $h(x)$ of $n \times n$ selfadjoint RMs with iid complex Gaussian entries having zero mean and variance $1/n$: (a) normalized histogram of a single $40 \times 40$ selfadjoint RM sample, (b) normalized histogram of 20 $8 \times 8$ selfadjoint RM samples. As expected by Theorem 2.2.2, (a) and (b) are approximated by the pdf of the semicircle law.

■

The next theorem is for large RMs having iid complex entries. It is known as Girko's fullcircle law [**?**].

**THEOREM 2.2.3:** *The fullcircle law.*
*Let* $\mathbf{F} \in \mathbb{C}^{n \times n}$ *be a random matrix with iid complex Gaussian entries, with zero mean and variance* $\sigma^2 = 1/n$. *The matrices* $\mathbf{F}$ *are not self-adjoint, so the eigenvalues are complex. When the dimension of* $\mathbf{F}$ *tends to infinity, the empirical and the mean eigenvalue distribution tend to the asymptotic distribution with the following pdf:*

$$f_{\mathbf{F}}(z) = \begin{cases} \frac{1}{\pi}, & \text{for } |z| < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

This implies that the eigenvalues are equally distributed over a circle of center 0 and radius one.

EXAMPLE 2.2.4: The fullcircle law.
This distribution of the eigenvalues can already be observed for finite dimensions as shown in Figure 2.8, where all the eigenvalues of 20 matrices of dimension $40 \times 40$ with iid complex Gaussian entries have been plotted.

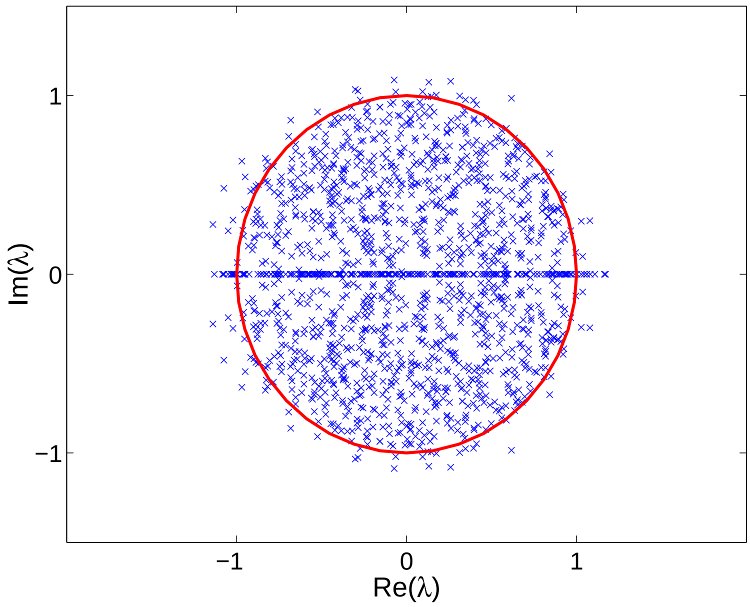



Figure 2.8: Eigenvalues $\lambda$ of 20 samples of a $40 \times 40$ RM having iid complex Gaussian entries with zero mean and unit variance. We observe that the fullcircle law is a good approximation.

■

The following theorem is a very interesting one for communication engineers, since the asymptotic distribution is needed to compute asymptotic capacities, as it will be done in Section 3.2. It was proved several times in the 1960's and 1970's, [14, 19].

**THEOREM 2.2.4:** *Marčenko Pastur law*
*Let a matrix* $\mathbf{M} \in \mathbb{C}^{n \times k}$ *have iid complex entries with zero mean and variance* $\sigma^2 = 1/n$.

*Then, as $n, k \to \infty$ with $\beta = \frac{k}{n}$ fixed, the empirical distribution of the singular values of $\mathbf{M}$, hence, the empirical eigenvalues of the eigenvalues of $\mathbf{D} = \mathbf{M}\mathbf{M}^H$ converge to a distribution, whose pdf is given by:*

$$f_{\mathbf{D}}(x) = \begin{cases} \frac{\sqrt{4\beta-(x-1-\beta)^2}}{2\pi x}, & (1-\sqrt{\beta})^2 < x < (1+\sqrt{\beta})^2, \\ [1-\beta]^+\delta(x), & \text{elsewhere,} \end{cases} \tag{2.11}$$

*where* $[\cdot]^+ = \max(0, \cdot)$.
*The eigenvalue distribution of the matrix* $\tilde{\mathbf{D}} = \mathbf{M}^H\mathbf{M}$ *can be easily derived from (2.11), since* $\mathbf{D}$ *and* $\tilde{\mathbf{D}}$ *have the same non-zero eigenvalues. As* $n \to \infty$,

$$f_{\tilde{\mathbf{D}}}(x) = \frac{1}{\beta} f_{\mathbf{D}}(x) + \left(1 - \frac{1}{\beta}\right)\delta(x). \tag{2.12}$$

This distribution is known as the Marčenko-Pastur law, its density is illustrated in Figure 2.9 for different values of $\beta$.

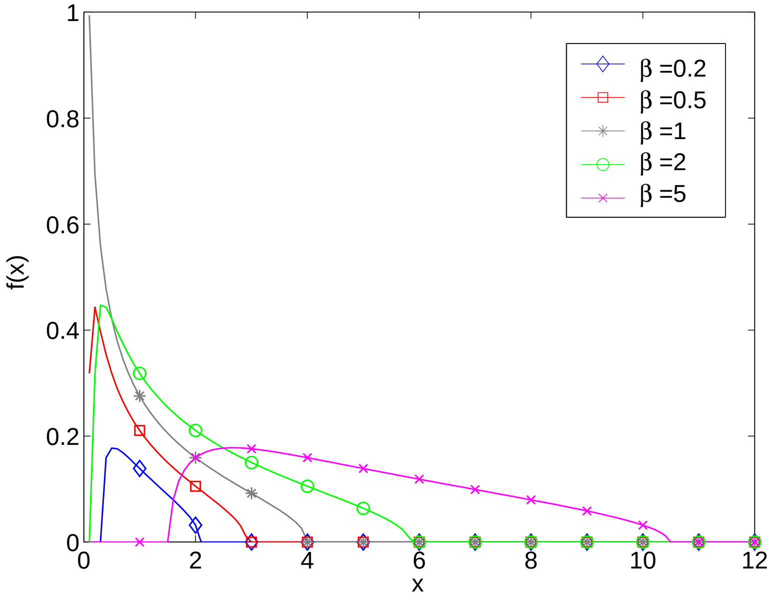



Figure 2.9: Marčenko Pastur law $f(x)$ for different values of $\beta = 0.2\,, 0.5\,, 1\,, 2\,, 5$.

Example 2.2.5: Marčenko-Pastur law.
For 50 samples of a squared RM $\mathbf{C} \in \mathbb{C}^{50\times 50}$ $(\beta = 1)$ with iid Gaussian entries, having zero mean and unit variance, the normalized histogram of the eigenvalues of $\mathbf{D}$ is illustrated in Figure 2.10. It is good approximated by the pdf of the Marčenko-Pastur law illustrated in Figure 2.9.

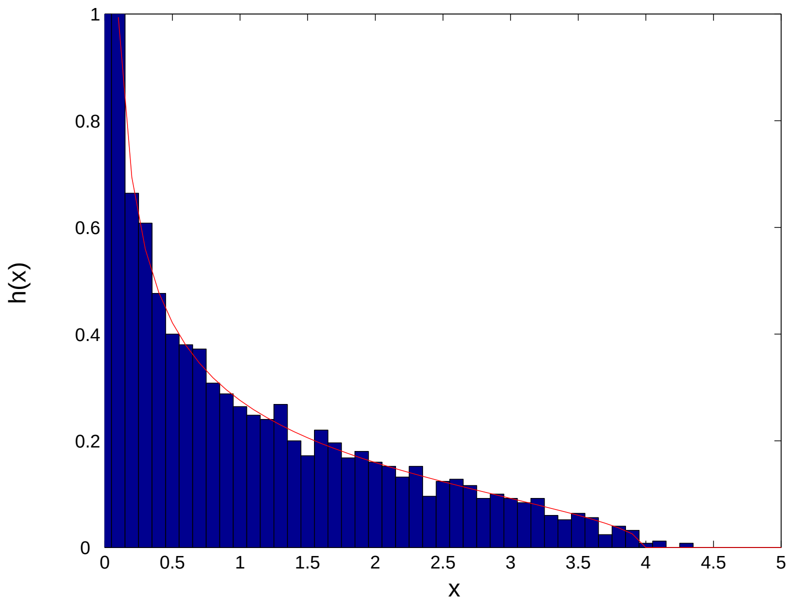



Figure 2.10: Normalized histogram of 50 RM samples of a matrix $\mathbf{D} = \mathbf{M}\mathbf{M}^H$ with $\mathbf{M}$ a squared matrix $\mathbf{M} \in \mathbb{C}^{50\times 50}$ $(\beta = 1)$ with iid Gaussian entries.

■

Constant matrices can also be regarded as NCRVs, since they are just deterministic RM. An asymptotic eigenvalue distribution can also be given for them. It is easy to see that the mean eigenvalue distribution and the empirical eigenvalue distribution are identical for constant matrices. In the following we will present some examples of asymptotic eigenvalue distributions, which are mentioned and used in other sections of this work.

Example 2.2.6: Matrices with all eigenvalues equal $\lambda$.
A simple example is that of the weighted unit matrix $\lambda\mathbf{I}_n$ of dimension $n \times n$. Its asymptotic eigenvalue distribution is $\tau(x-\lambda)$ with $\tau(x)$ denoting the unit step function and thus its asymptotic eigenvalue pdf is given by $\delta(x-\lambda)$. Any matrix with identical eigenvalues $\lambda$, has $\delta(x-\lambda)$ as asymptotic eigenvalue pdf. This can be for example an upper triangular matrix with values $\lambda$ on the diagonal $\mathbf{A}_1$ or a matrix $\mathbf{A}_2$, whose eigenvalue decomposition, gives all eigenvalues $\lambda$

$$\mathbf{A}_1 = \begin{pmatrix} \lambda & x_{12} & \dots & x_{1n} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & x_{(n-1)n} \\ 0 & \dots & 0 & \lambda \end{pmatrix} \quad \text{or} \quad \mathbf{A}_2 = \mathbf{U}^H \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \dots & \dots & 0 & \lambda \end{pmatrix} \mathbf{U},$$

where $\mathbf{U}$ is a matrix containing the eigenvectors of the matrix $\mathbf{A}_2$.

■

The asymptotic distributions of other constant matrices, i.e. Toeplitz matrices of the form

$$\begin{pmatrix} t_0 & t_{-1} & t_{-2} & \dots & t_{-(n-1)} \\ t_1 & t_0 & t_{-1} & \ddots & \vdots \\ t_2 & t_1 & \ddots & \ddots & t_{-2} \\ \vdots & \ddots & \ddots & t_0 & t_{-1} \\ t_{n-1} & \dots & t_2 & t_1 & t_0 \end{pmatrix}$$

will be used in Section 3.2. In the following, we will show how to obtain the empirical eigenvalue distribution of special classes of Toeplitz matrices including a bandmatrix and a Toeplitz matrix, where the coefficients follow some exponential rule.

Example 2.2.7: Bandmatrices.
Consider matrices of the form

$$\mathbf{B} = \begin{pmatrix} 1 & \rho & 0 & \dots & 0 \\ \rho & 1 & \rho & \ddots & \vdots \\ 0 & \rho & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 & \rho \\ 0 & \dots & 0 & \rho & 1 \end{pmatrix}.$$

The eigenvalues $\lambda_k$ of this matrix can be easily calculated [9] by

$$\lambda_k = 1 + 2\rho \cos\left(\frac{k\pi}{n}\right), \qquad k = 0 \dots n-1. \tag{2.13}$$

One example of the eigenvalues of this bandmatrix with parameter $\rho = 0.3$ is shown in Figure 2.11.
In order to get the empirical eigenvalue distribution of a bandmatrix of dimension $n$, we have to compute the inverse of $\lambda_k$ with respect to $k/n$, hence invert (2.13), i.e.

$$\frac{k}{n} = \frac{1}{\pi} \arccos\left(\frac{\lambda_k - 1}{2\rho}\right), \tag{2.14}$$

which results in a stairfunction (obtained as in Example 2.2.2). The resulting distribution of the eigenvalues of Figure 2.11 is illustrated in Figure 2.12.
As we can observe, the empirical eigenvalue distribution tends to a continous distribution as $n \to \infty$. We can say that the bandmatrix $\mathbf{B}$ has an asymptotic empirical eigenvalue distribution $F_{\mathbf{B}}(x)$,

$$F_{\mathbf{B}}(x) = \frac{1}{\pi} \arccos\left(\frac{x-1}{2\rho}\right). \tag{2.15}$$

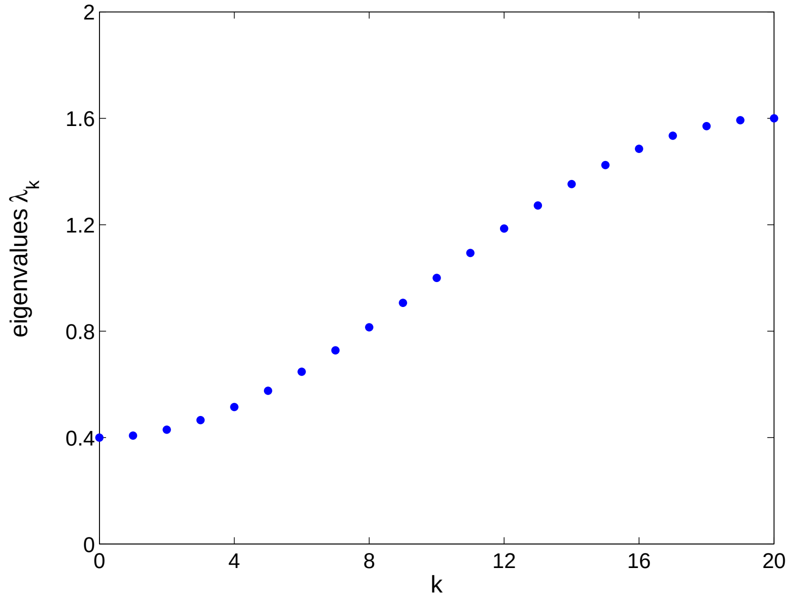



Figure 2.11: Eigenvalues $\lambda_k$ of a bandmatrix of dimension $20 \times 20$ with parameter $\rho = 0.3$ (see (2.13)).

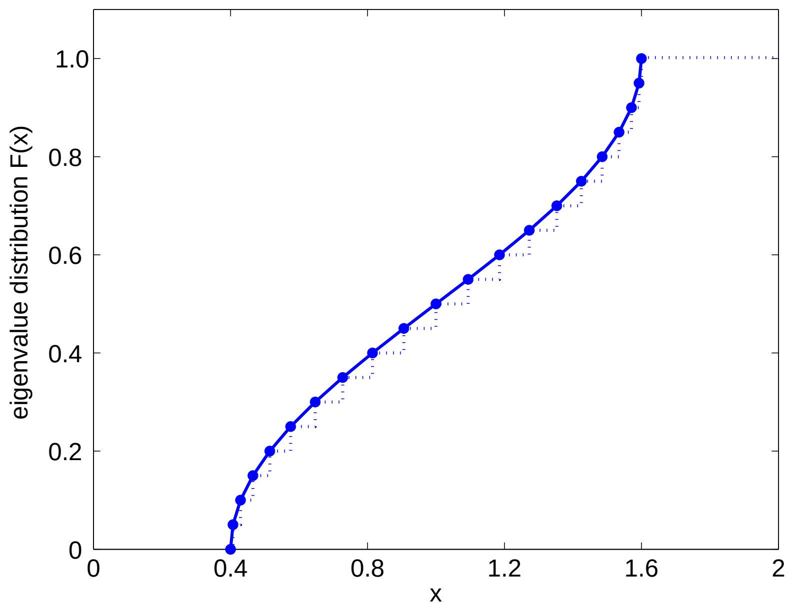



Figure 2.12: Eigenvalue distribution $F(x)$ of a $20 \times 20$ bandmatrix with parameter $\rho = 0.3$ (see (2.15)).

The corresponding pdf $f_{\mathbf{B}}(x)$ is obtained by differentiating $F_{\mathbf{B}}(x)$, i.e.

$$f_{\mathbf{B}}(x) = \frac{dF_{\mathbf{B}}(x)}{dx} = \frac{1}{\pi}\frac{1}{\sqrt{4\rho^2 - (x-1)^2}}, \tag{2.16}$$

which is illustrated in Figure 2.13 for $\rho = 0.3$.

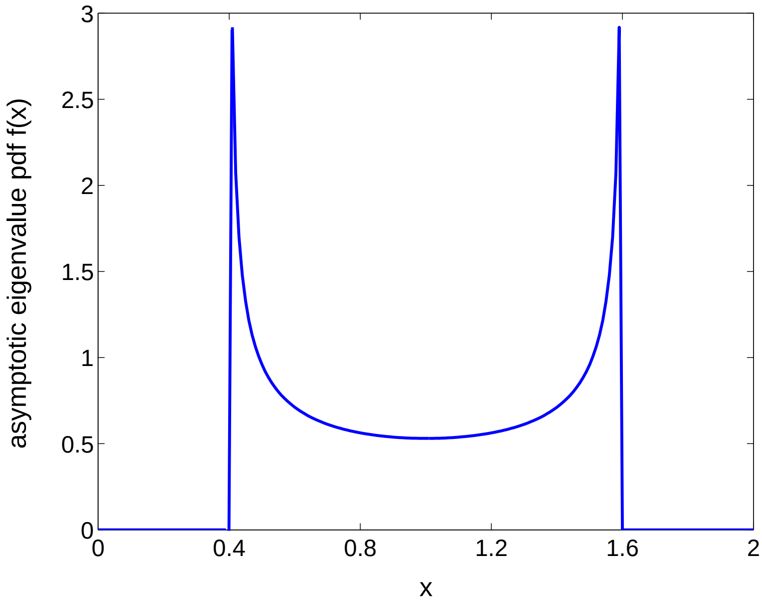



Figure 2.13: Asymptotic eigenvalue pdf $f(x)$ of a bandmatrix with $\rho = 0.3$ (see (2.16).

To illustrate, that the empirical eigenvalue pdf approximates the normalized histogram of the eigenvalues of a bandmatrix, the normalized histogram of a $200 \times 200$ bandmatrix with $\rho = 0.3$ is illustrated in Figure 2.14.

■

Example 2.2.8: Exponential Toeplitz matrix.

We assume a matrix $\mathbf{E}$, which corresponds to a Toeplitz matrix with coefficients $t_j = \rho^{|j|}$:

$$\mathbf{E} = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \ddots & \vdots \\ \rho^2 & \rho & \ddots & \ddots & \rho^2 \\ \vdots & \ddots & \ddots & 1 & \rho \\ \rho^{n-1} & \dots & \rho^2 & \rho & 1 \end{pmatrix}.$$

As we know from Gray [9], the asymptotic eigenvalue distribution of Toeplitz matrices tends to the discrete Fourier-transform of the first row, hence as $n \to \infty$

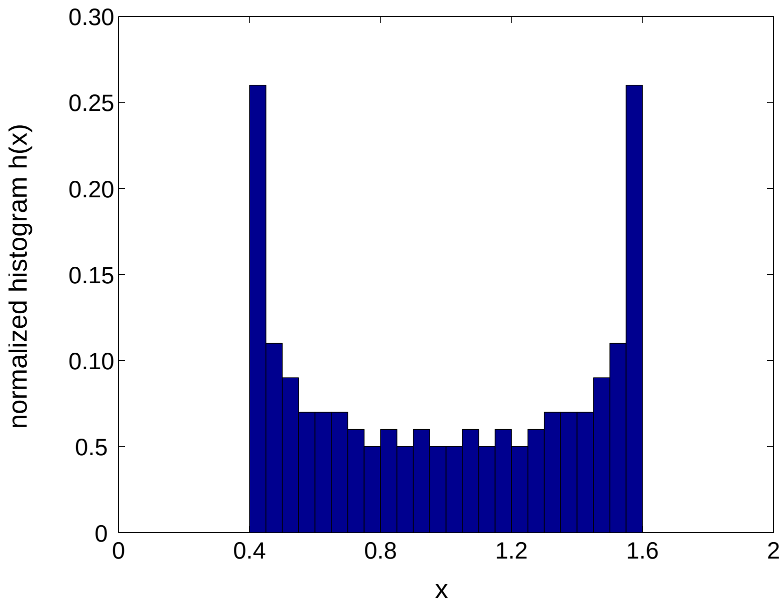



Figure 2.14: Normalized histogram $h(x)$ of the eigenvalues of a $20 \times 20$ bandmatrix with $\rho = 0.3$.

$$\lambda_k = \sum_{l=-n}^{n} \rho^{|l|} \mathrm{e}^{i\,2\pi l\,k/n} = \frac{1-\rho^2}{1+\rho^2+2\rho\,\cos(\pi k/n)}. \tag{2.17}$$

In Figure 2.15, we can see that the eigenvalues tend to this asymptotic distribution as the dimension increases. The eigenvalues are already very close to the asymptotic ones for $n = 50$. The asymptotic empirical eigenvalue distribution $F_{\mathbf{E}}(x)$ can then be obtained by inverting (2.17) with respect to $k/n$. We obtain

$$F_{\mathbf{E}}(x) = \frac{1}{\pi} \arccos\left( \frac{(1-\rho^2)}{2\rho x} - \frac{(1+\rho^2)}{2\rho} \right), \tag{2.18}$$

which is plotted in Figure 2.16. The asymptotic eigenvalue pdf $f_{\mathbf{E}}(x)$ can then be derived by differentiating (2.18), which results in

$$f_{\mathbf{E}}(x) = \frac{1}{\pi x} \frac{1}{\sqrt{-x^2 + 2x\left(\frac{1+\rho^2}{1-\rho^2}\right) - 1}}. \tag{2.19}$$

This pdf $f_{\mathbf{E}}(x)$ is illustrated in Figure 2.17. To illustrate, that the normalized histogram approximates (2.19), the normalized histogram of a $200 \times 200$ matrix $\mathbf{E}$ with parameter $\rho = 0.3$ is illustrated in Figure 2.18.

■

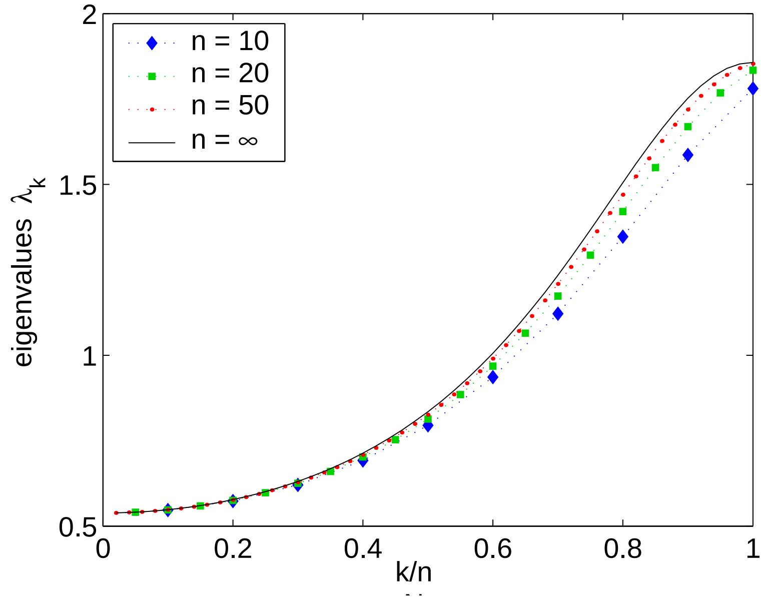



Figure 2.15: Eigenvalues $\lambda_k$ of an exponential Toeplitz matrix with parameter $\rho = 0.3$ for increasing dimensions $n = 10, 20, 50, \infty$. The eigenvalues tend to the asymptotic values given by (2.17).

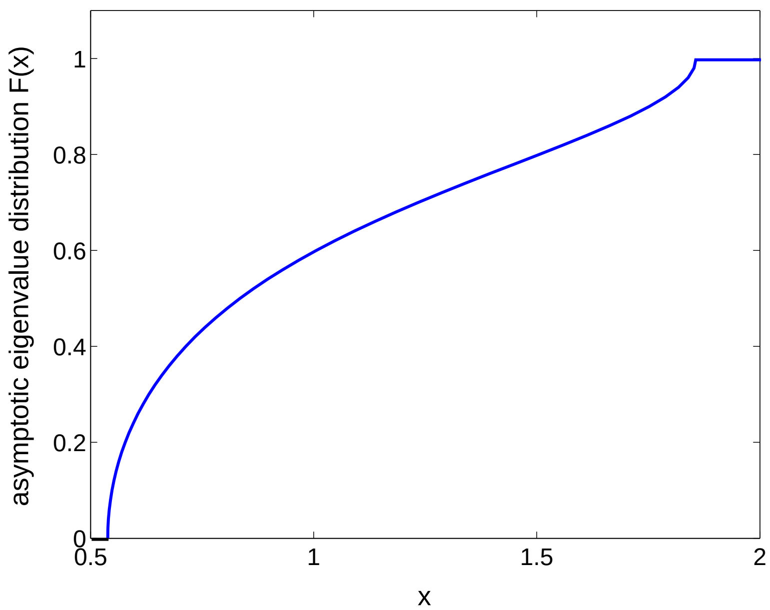



Figure 2.16: Asymptotic eigenvalue distribution of an exponential Toeplitz matrix with parameter $\rho = 0.3$.

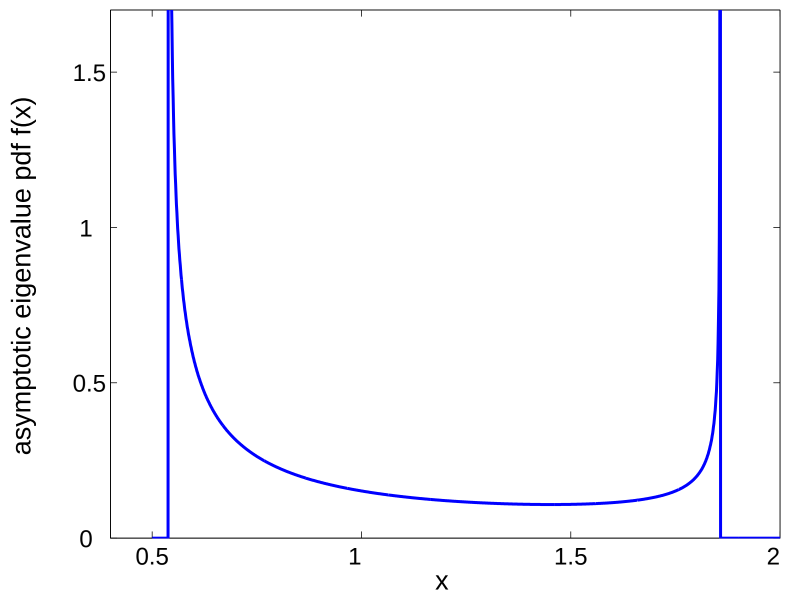



Figure 2.17: Asymptotic eigenvalue pdf of an exponential Toeplitz matrix with parameter $\rho = 0.3$.

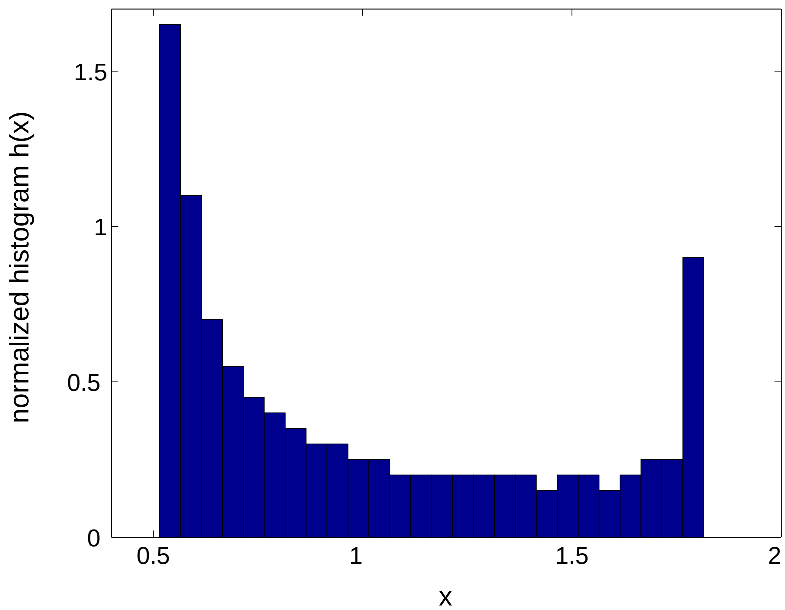



Figure 2.18: Normalized histogram of the eigenvalues of a 200×200 exponential Toeplitz matrix with parameter $\rho = 0.3$. The shape of the histogram meets the shape of the asymptotic eigenvalue pdf obtained in (2.19) (see Figure 2.17).

### 2.2.3 Asymptotic free independence of random matrices

After embedding random matrices into free probability and finding out, that there are classes of random matrices that have asymptotic (empirical and mean) eigenvalue distributions, we might want to know if there exist free families of random matrices. In classical probability theory, it is not very difficult to recognize the independence of random variables. This is unfortunately not the case for the free independence. The definition of free independence (see Definition 1.3.1) implies that there is no easy criteria to find out if NCRVs are free. Although there exist some results about certain classes of NCRVs and especially about RMs. Most of them were discovered by Voiculescu [24]. Here again, the results are not for finite-dimensional matrices but for matrices in their large limit. This is the reason why the concept of free independence is extended to the concept of *asymptotic free independence* [10].

**DEFINITION 2.2.5:** *Let $(A_{i,n})_{i\in\mathcal{I},n\in\mathbb{N}}$ be a sequence of families of NCRVs in the NCPS $(\mathcal{A}_n,\varphi_n)$, that converges to the family $(A_i)_{i\in\mathcal{I}}$ in distribution as $n\to\infty$. $(A_{i,n})_{i\in\mathcal{I},n\in\mathbb{N}}$ is called* asymptotically free*, if its limit $(A_i)_{i\in\mathcal{I}}$ is free.*

Assume $\mathbf{A}_{1,n}$ and $\mathbf{A}_{2,n}$ are RMs of dimension $n\times n$ that converge in distribution to the NCRVs $A_1$ and $A_2$, respectively. In other words, the eigenvalue distributions $F_{1,n}(x)$ and $F_{2,n}(x)$ of $\mathbf{A}_{1,n}$ and $\mathbf{A}_{2,n}$ converge to the distributions $F_1(x)$ and $F_2(x)$, i.e.

$$\lim_{n\to\infty} F_{i,n}(x) = F_i(x), \quad i=1,2.$$

If the NCRVs $A_1$ and $A_2$ are free, then the matrices $\mathbf{A}_{1,n}$ and $\mathbf{A}_{2,n}$ are said to be asymptotically free. This implies also that the results of Examples 1.3.1, 1.3.2 and 1.3.3 can be applied to RMs. For example, if $\mathbf{A}_{1,n}$ and $\mathbf{A}_{2,n}$ are asymptotically free, then as $n\to\infty$

$$\mathrm{tr}_n(\mathbf{A}_{1,n}\mathbf{A}_{2,n}) = \mathrm{tr}_n(\mathbf{A}_{1,n})\mathrm{tr}_n(\mathbf{A}_{2,n}), \tag{2.20}$$

or

$$\mathrm{tr}_n(\mathbf{A}_{1,n}\mathbf{A}_{2,n}\mathbf{A}_{1,n}) = \mathrm{tr}_n(\mathbf{A}_{1,n}^2)\mathrm{tr}_n(\mathbf{A}_{2,n}).$$

The same can be said for the Voiculescu's trace $\tau_n$.
In the following, the main known classes of asymptotically free random matrices are introduced. Proofs can be found in Voiculescu's work [23] and also in [10].

**THEOREM 2.2.6:** *Let $(\mathbf{X}_i)_{i\in\mathcal{I}}$ be an independent family of $n\times n$ random matrices. The entries are iid distributed with mean $\mu=0$ and variance $\sigma^2=1/n$. Let $(\mathbf{D}_j)_{j\in\mathcal{J}}$ be a family of $n\times n$ constant matrices with bounded norm and such that $(\mathbf{D}_j,\mathbf{D}_j^H)_{j\in\mathcal{J}}$ has a limit eigenvalue distribution. Then the family*

$$\left((\{\mathbf{X}_i,\mathbf{X}_i^H\})_{i\in\mathcal{I}},(\{\mathbf{D}_j,\mathbf{D}_j^H\})_{j\in\mathcal{J}}\right)$$

*is asymptotically free as* $n \to \infty$.

Note that the matrices $\mathbf{X}_i$ fullfill the conditions of the fullcircle law, mentioned in Theorem 2.2.3. Hence the families of matrices that have the semicircle law as asymptotic eigenvalue distribution are asymptotically free. The same can be said for RMs fullfilling the fullcircle law (Theorem 2.2.3).

**THEOREM 2.2.7:** *Let* $(\mathbf{H}_i)_{i\in\mathcal{I}}$ *be an independent family of* $n \times n$ *self-adjoint RMs with iid complex entries with mean* $\mu = 0$ *and variance* $\sigma^2 = 1/n$. *Let* $(\mathbf{D}_j)_{j\in\mathcal{J}}$ *be as in Theorem 2.2.6. Then the family*

$$\big((\{\mathbf{H}_i, \mathbf{H}_i^H\})_{i\in\mathcal{I}}, (\{\mathbf{D}_j, \mathbf{D}_j^H\})_{j\in\mathcal{J}}\big)$$

*is asymptotically free as* $n \to \infty$.

A very important result is for Hermitian matrices, since it is very useful for communication engineers.

**THEOREM 2.2.8:** *Let* $(\mathbf{H}_i)_{i\in\mathcal{I}}$ *be an independent family of* $n \times k$ *RMs with iid complex entries with mean* $\mu = 0$ *and variance* $\sigma^2 = 1/n$ *and* $\mathbf{P}_i = \mathbf{H_i}\mathbf{H_i}^H$. *Let* $(\mathbf{D}_j)_{j\in\mathcal{J}}$ *be as in Theorem 2.2.6. Then the family*

$$\big((\{\mathbf{P}_i\})_{i\in\mathcal{I}}, (\{\mathbf{D}_j, \mathbf{D}_j^H\})_{j\in\mathcal{J}}\big)$$

*is asymptotically free as* $n \to \infty$.

Note that the RM $\mathbf{P}_i$ fullfills the conditions of the Marčenko-Pastur law, mentioned in Example 2.2.4. The next theorem is for unitary matrices.

**THEOREM 2.2.9:** *Let be* $(\mathbf{U}_i)_{i\in\mathcal{I}}$ *be an independent family of* $n \times n$ *standard unitary random matrices. Let* $(\mathbf{D}_j)_{j\in\mathcal{J}}$ *be a family of constant matrices such that* $(\mathbf{D}_j, \mathbf{D}_j^H)_{j\in\mathcal{J}}$ *has a limit distribution. Then the family*

$$\big((\{\mathbf{U}_i, \mathbf{U}_i^H\})_{i\in\mathcal{I}}, (\{\mathbf{D}_j, \mathbf{D}_j^H\})_{j\in\mathcal{J}}\big)$$

*is asymptotically free as* $n \to \infty$.

EXAMPLE 2.2.9: We assume a $n \times n$ constant matrix $\mathbf{D}$, having a limit eigenvalue distribution and a $n \times n$ matrix $\mathbf{H}$ that fullfills the conditions of the semicircle law.

Following Theorem 2.2.7, these matrices are in asymptotically free. Assume that $\mathbf{D}$ has a diagonal filled with 3's and thus $\mathrm{tr}_n(\mathbf{D}) = 3$. Furthermore, assume that $\mathbf{H}$ has the semicircle law as asymptotic eigenvalue distribution. If we want to compute the normalized trace of $\mathbf{HDH}$, then we can use the relation

$$\mathrm{tr}_n(\mathbf{HDH}) = \mathrm{tr}_n(\mathbf{H}^2)\mathrm{tr}_n(\mathbf{D}) = 1 \cdot 3 = 3,$$

where $\mathrm{tr}_n(\mathbf{H}^2)$ is the second order catalan number (see Example 2.2.3). By simulation, and for $n = 100$, we obtained the following results

$$\mathrm{tr}_n(\mathbf{HDH}) = 3.0020 + 0.0029 \quad \mathrm{tr}_n(\mathbf{H}^2)\mathrm{tr}_n(\mathbf{D}) = 3.0041,$$

which fit well with the asymptotic results. ■

EXAMPLE 2.2.10: Assume the same constant matrix $\mathbf{D}$, but a $n \times k$ RM $\mathbf{H}$, with iid complex entries with mean $\mu = 0$ and variance $\sigma^2 = 1/n$. Following Theorem 2.2.8, these matrices are asymptotically free. If we compute the normalized trace of $\mathbf{DHH}^H$, we obtain

$$\mathrm{tr}_n(\mathbf{DHH}^H) = \mathrm{tr}_n(\mathbf{D})\mathrm{tr}_n(\mathbf{HH}^H) = 1 \cdot 3 = 3,$$

$\mathrm{tr}_n(\mathbf{HH}^H)$ being the first moment of the Marčenko-Pastur distribution (see Example 2.2.3). By simulation, for $n = 100$ and $k = 200$, we obtained the following results

$$\mathrm{tr}_n(\mathbf{DHH}^H) = 5.9716 - 0.0248i \quad \mathrm{tr}_n(\mathbf{D})\mathrm{tr}_n(\mathbf{HH}^H) = 3 \cdot 1.9918 = 5.9755,$$

which fit with the asymptotic results. ■

### 2.2.4 Asymptotic free convolution

FPT applied on RMs provides us with statistical information about the eigenvalues, i.e. the eigenvalue distribution, the eigenvalue pdf, and the moments that correspond to the normalized trace of the matrix (see Section 2.2.1). Normally, the knowledge of the eigenvalues of two matrices does not help us to compute the eigenvalues of the sum or the product of them. In FPT, if two RMs $\mathbf{A}_1$ and $\mathbf{A}_2$ are (asymptotically) free, then knowing the eigenvalue distributions of $\mathbf{A}_1$ and $\mathbf{A}_2$ enables us to compute the eigenvalue distributions of the sum $\mathbf{A}_1 + \mathbf{A}_2$ and the product $\mathbf{A}_1\mathbf{A}_2$. This is done by means of the free convolution, introduced in Section 1.5. Since we have results about the free independence of certain random matrices, we are able to apply the R-transform and the S-transform on them as it was done in Example 1.7.1 and Example 1.7.2.

#### Additive free convolution

With help of the additive free convolution and the R-transform, the eigenvalue distribution of the sum of two RM, which are (asymptotically) free can be computed. The following Example uses matrices, that have already been mentioned in other sections.

EXAMPLE 2.2.11: Assume two RMs $\mathbf{A}_1$ and $\mathbf{A}_2$, with $\mathbf{A}_1$ a selfadjoint $n \times n$ matrix with iid complex Gaussian entries, with zero mean and variance $1/n$ and $\mathbf{A}_2$ a $n \times n$ constant matrix with $n$ equal eigenvalues $\lambda$. Theorem 2.2.7 states that $\mathbf{A}_1$ and $\mathbf{A}_2$ are asymptotically free. $\mathbf{A}_1$ fullfills the conditions of the semicircle law (see Theorem 2.2.2). Its asymptotic empirical eigenvalue pdf is given by (2.10):

$$\lim_{n\to\infty} f_{\mathbf{A}_1}(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2}, & \text{for } |x| < 2, \\ 0, & \text{elsewhere }. \end{cases}$$

In Example 2.2.6 the asymptotic empirical eigenvalue pdf of a constant matrix with all eigenvalues equal $\lambda$ was introduced as

$$\lim_{n\to\infty} f_{\mathbf{A}_2}(x) = \delta(x-\lambda).$$

Since the RMs $\mathbf{A}_1$ and $\mathbf{A}_2$ are in asymptotically free, the free convolution can be applied on them. By using the R-transforms $R_{\mathbf{A}_1}(z)$ and $R_{\mathbf{A}_2}(z)$, the asymptotic empirical pdf $f_{\mathbf{A}}(z)$ of the matrix $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$ can be computed. This was already done in Example 1.7.1. The resulting empirical eigenvalue pdf was a shifted semicircle law,

$$\lim_{n\to\infty} f_{\mathbf{A}}(x) = \begin{cases} \frac{1}{2\pi}\sqrt{(z-\lambda)^2-4}, & \text{for } |x-\lambda| < 2, \\ 0, & \text{elsewhere }, \end{cases}$$

as illustrated in Figure 1.2.
If we use $\mathbf{A}_2 = \lambda\mathbf{I}$, then it is very intuitive and easy to prove, that the eigenvalue pdf is a shifted semicircle law. In this case $\mathbf{A} = \lambda\mathbf{I} + \mathbf{A}_1$, the eigenvalues of $\mathbf{A}$ are the eigenvalues of $\mathbf{A}_1$ plus a factor $\lambda$. This explains the shift of the resulting pdf. If matrices like the ones introduced in Example 2.2.6 are taken for $\mathbf{A}_2$, then it is not so intuitive anymore how we get the resulting pdf.
In Figure 2.19, the result is illustrated by simulation. The histogram of the eigenvalues of the sum of matrices of the kind of $\mathbf{A}_1$ and $\mathbf{A}_2$ is plotted. For $\mathbf{A}_2$, a matrix of the kind of $\mathbf{M}_2$ was used.

■

### Multiplicative free convolution

With help of the multiplicative free convolution and the S-transform, the eigenvalue distribution of the product of two RMs, which are asymptotically free can be computed. Again, the following Example applies the S-transform on RMs with known eigenvalue distributions.

EXAMPLE 2.2.12: In analogy to Example 2.2.11, here the asymptotic eigenvalue pdf of the product of two RM $\mathbf{A}_1$ and $\mathbf{A}_2$ is computed using the multiplicative free convolution; the same matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ are used. And we recall the result of Example 1.7.2. The asymptotic eigenvalue pdf of $\mathbf{A} := \mathbf{A}_1\mathbf{A}_2$ is a stretched semicircle law.

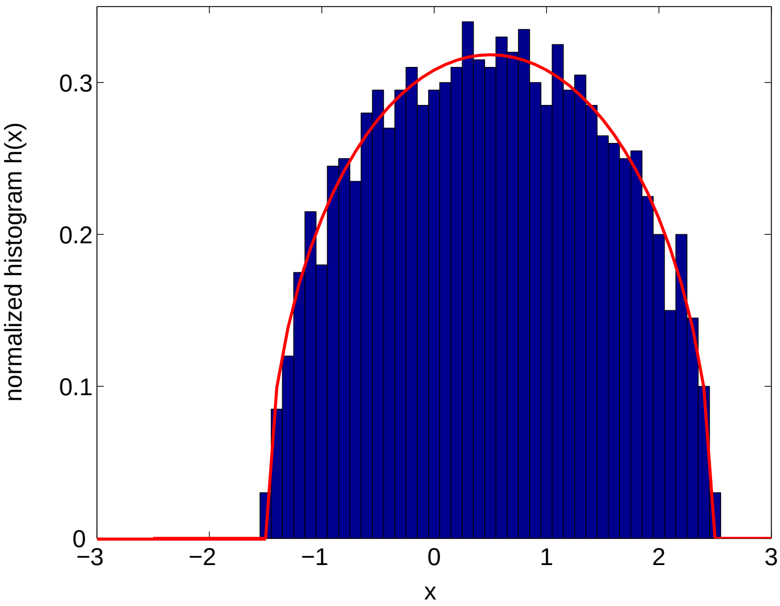



Figure 2.19: To verify the result of Example 2.2.11, the normalized histogram of the eigenvalues of 20 matrices $\mathbf{A} \in \mathbb{C}^{100\times100}$, which are the sum of two matrices, $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$, is plotted here. Matrices $\mathbf{A}_1 \in \mathbb{C}^{100\times100}$ are samples of a self-adjoint RM with complex iid Gaussian entries with mean zero and variance $1/100$. Their asymptotic empirical eigenvalue pdf is the semicircle law. Matrices $\mathbf{A}_2 \in \mathbb{C}^{100\times100}$ are constant matrices with the eigenvalue 0.5 of order 100. Their asymptotic empirical eigenvalue pdf is the dirac function. As it was expected, the histogram is very well approximated by a shifted semicircle law, as calculated in Example 1.7.1.

$$\lim_{n\to\infty} f_{\mathbf{A}}(x) = \begin{cases} \frac{1}{2\pi\lambda^2}\sqrt{4\lambda^2 - x^2}, & \text{for } |x| < 2\lambda, \\ 0, & \text{elsewhere.} \end{cases}$$

Again if we use the matrix $\mathbf{A}_2 = \lambda\mathbf{I}$, then it is very intuitive and easy to prove, that the eigenvalue pdf is a stretched semicircle law. In this case $\mathbf{A} = \lambda\mathbf{A}_1$, the eigenvalues of $\mathbf{A}$ are the eigenvalues of $\mathbf{A}_1$ times a factor $\lambda$ which explains the stretched form of the resulting pdf. In Figure 2.20, the result was illustrated by simulation. The normalized histogram of the eigenvalues of the product of matrices of the kind of $\mathbf{A}_1$ and $\mathbf{A}_2$ is plotted. For $\mathbf{A}_2$, a matrix of the kind of $\mathbf{M}_2$ was used. ■

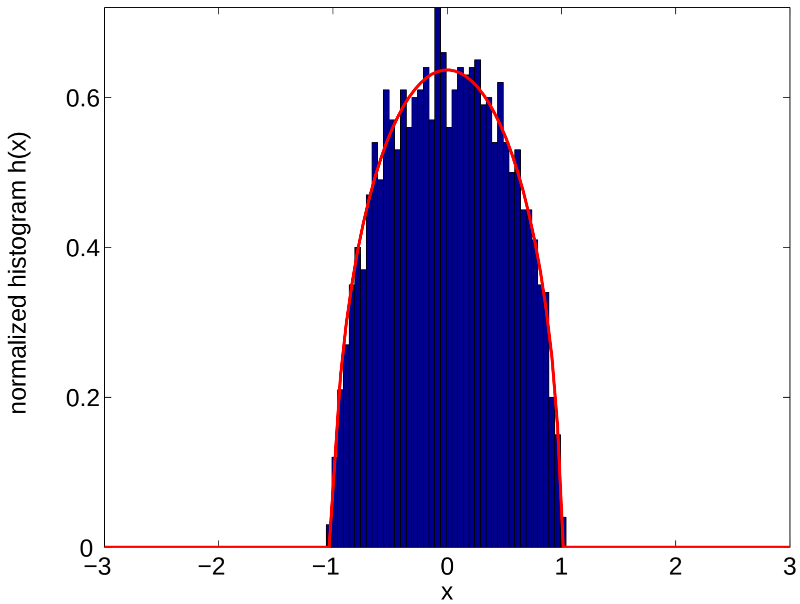



Figure 2.20: To verify the result of Example 2.2.12, the normalized histogram of the eigenvalues of 20 matrices $\mathbf{A} \in \mathbb{C}^{100\times100}$, which are the product of two matrices, $\mathbf{A} = \mathbf{A}_1\mathbf{A}_2$, is plotted here. $\mathbf{A}_1 \in \mathbb{C}^{100\times100}$ are samples of self-adjoint RM with complex iid Gaussian entries with mean zero and variance $1/100$. Their asymptotic empirical eigenvalue pdf is the semicircle law. $\mathbf{A}_2 \in \mathbb{C}^{100\times100}$ is a constant matrix with the eigenvalue 0.5 of order 100. As it was expected, the histogram is very well approximated by a streched semicircle law, calculated in Example 1.7.2

# Chapter 3

# Applications

Now, free probability theory and random matrices have been introduced; thus, we can start to ask where in communications the obtained results can be applied. In general matrices appear in equations, in input-output relations between vectors and also matrices. In general the FPT is used in relation with random matrices, when the information about the eigenvalues or also the trace is needed. This information is used to calculate error probabilities, or capacities. Free probability can then be used to work with the matrices in case. As we have seen in Chapter 1, the most interesting results about free probability applied on random matrices are valid in the case where the dimension of the matrix tends to infinity. The tools of FPT are thus used under this rather unrealistic assumption. The assumption of a infinite dimension has different interpretations; in the case of a MIMO system or when we work with space-time codes it is equivalent to the assumption that the number of transmit and receive antennas tend to infinity. Nevertheless these "asymptotic" results provide good approximations for the finite dimensional case. In the following, Section 3.1 presents calculations of asymptotic pairwise error probabilities for space time codes and Section 3.2 investigates asymptotic channel capacity of correlated MIMO channels.

## 3.1 Performance of space-time codes for a large number of antennas

A very nice application of FPT is illustrated by the work of Biglieri [5]. Here, the error probability of space-time codes (STCs), computed by means of the pair-wise error probability (PEP), is evaluated assuming that the numbers of transmit and receive antennas go to infinity. The asymptotic value of the PEP is used as an approximation for the finite dimensional case. The system model is based on iid Gaussian RMs. The assumption of an infinite dimensional system enables us to use the tools of FPT. Simple properties are used, i.e. the moments of the Marčenko-Pastur law (its pdf being $f_{MP}(x)$) and the property shown in Example 1.3.1 and applied on RMs in (2.20). The PEP is computed for the case of ML detection with no linear interface and with a zero-forcing interface. With help of FPT, the expressions for PEP can be simplified and interesting conclusions can be given.

### 3.1.1 Space-Time Codes

Space-Time Codes were first introduced by Tarokh [20] in order to provide transmit diversity to the multiple antenna fading channel. Space time block codes operate on a block of input symbols, producing a matrix output, whose columns represent time and rows represent antennas. The main feature of these codes is to provide full diversity with a simple decoding scheme.

#### Channel Model

We consider a system with $N_T$ transmit and $N_R$ receive antennas. The transmitted data is encoded into a space-time block codeword. The codeword $x_n[i]$ is transmitted by the $n$th antenna at the discrete time $i = 1 \ldots N$. By stacking the components, we obtain $\mathbf{x}[i] = (x_1[i], x_2[i], \ldots, x_{N_T}[i])^T$, and $\mathbf{X} = (\mathbf{x}[1], \mathbf{x}[2], \ldots, \mathbf{x}[N])$. The input-output relation is then given by

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z}. \tag{3.1}$$

where $\mathbf{Y} \in \mathbb{C}^{N_R \times N}$ represents the receive signal, $\mathbf{N} \in \mathbb{C}^{N_R \times N}$ represents the noise and $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ is the channel matrix. The noise matrix $\mathbf{N}$ has iid complex Gaussian entries with zero mean and variance $N_0$. The noise is spatially and temporally independent, which implies that $\mathbb{E}\{\mathbf{N}\mathbf{N}^H\} = N N_0 \mathbf{I}_{N_R}$. The channel matrix $\mathbf{H}$ has iid complex Gaussian entries with zero-mean and variance $1/N_R$ (for the conditions to apply the Marčenko-Pastur law and the asymptotic free independence, see Theorem 2.2.4 and Theorem 2.2.8).

#### Norm, Scalar Product

The mathematical tools that are going to be used demand the introduction of a scalar product and a norm of the signals.

- The Frobenius norm $\|\cdot\|$ is applied on matrices. For $\mathbf{A} \in \mathbb{C}^{m \times n}$, with elements $\mathbf{A}_{i,j}$, $i = 1 \ldots m$, $j = 1 \ldots n$,

$$\|\mathbf{A}\|^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} |\mathbf{A}_{ij}|^2 = \operatorname{trace}(\mathbf{A}^H \mathbf{A}) = \operatorname{trace}(\mathbf{A}\mathbf{A}^H).$$

- The scalar product can be chosen as

$$\langle \mathbf{A}, \mathbf{B} \rangle = \operatorname{trace}(\Re\{\mathbf{A}^H \mathbf{B}\}) = \operatorname{trace}(\Re\{\mathbf{B}^H \mathbf{A}\}) = \frac{1}{2} \operatorname{trace}(\mathbf{A}^H \mathbf{B} + \mathbf{B}^H \mathbf{A}).$$

- Note that the scalar product and also the norm can be rewritten in a vector notation. If vec(·) states the operator that stacks the colums into a vector [11], we obtain

$$\langle \mathbf{A}, \mathbf{B} \rangle = \Re\{\mathrm{vec}(\mathbf{A})^H \cdot \mathrm{vec}(\mathbf{B})\} \quad \text{and } \|\mathbf{A}\|^2 = \mathrm{vec}(\mathbf{A})^H \cdot \mathrm{vec}(\mathbf{A}).$$

Note that the use of the trace will enable us to apply the FPT tools.

### Detection

Maximum Likelihood (ML) detection is considered, which minimizes the distance $\| \mathbf{Y} - \mathbf{H}\mathbf{X} \|^2$. In the following the PEP of the code using different kinds of interfaces will be computed.

### Error Probability

The error probability $\mathbb{P}(e)$ is approximated using the union bound [5, 8]

$$\mathbb{P}(e) = \frac{1}{\|\chi\|} \sum_{\mathbf{X}\in\chi} \sum_{\hat{\mathbf{X}}\in\chi\setminus\{\hat{\mathbf{X}}\}} \mathbb{P}(\mathbf{X} \rightarrow \hat{\mathbf{X}}),$$

where $\mathbb{P}(\mathbf{X} \rightarrow \hat{\mathbf{X}})$ denotes the PEP, the probability that a wrong codeword $\hat{\mathbf{X}}$ is chosen, while a codeword $\mathbf{X}$ was transmitted.

- In the case of simple ML detection, such an error occurs if the distance $\|\mathbf{Y} - \mathbf{H}\hat{\mathbf{X}}\|^2$ is smaller than $\|\mathbf{Y} - \mathbf{H}\hat{\mathbf{X}}\|^2$, i.e.

$$\mathbb{P}(\mathbf{X} \rightarrow \hat{\mathbf{X}}) \;=\; \mathbb{P}\left(\|\mathbf{Y} - \mathbf{H}\hat{\mathbf{X}}\|^2 < \|\mathbf{Y} - \mathbf{H}\mathbf{X}\|^2\right).$$

This PEP can be reformulated as

$$\mathbb{P}(\mathbf{X} \rightarrow \hat{\mathbf{X}}) = \mathbb{P}\left(\langle \mathbf{N}, \mathbf{H}(\mathbf{X} - \hat{\mathbf{X}})\rangle < -\frac{\|\mathbf{H}(\mathbf{X} - \hat{\mathbf{X}})\|^2}{2}\right).$$

The product $\langle \mathbf{N}, \mathbf{H}(\mathbf{X} - \hat{\mathbf{X}})\rangle$ can also be represented in terms of a vector product using a linear combination of elements of $\mathbf{N}$. This implies that it is Gaussian distributed as well, with zero mean and variance $N_0\|\mathbf{H}(\mathbf{X} - \hat{\mathbf{X}})\|^2$. Using a result in [2], the PEP can then be written under the form

$$\mathbb{P}(\mathbf{X} \rightarrow \hat{\mathbf{X}}) = \mathbb{E}\left\{Q\left(\frac{\|\mathbf{H}(\mathbf{X} - \hat{\mathbf{X}})\|}{\sqrt{2N_0}}\right)\right\}, \tag{3.2}$$

where $Q(\cdot)$ is the $Q$-function. The final expression mainly depends on $\|\mathbf{H}(\mathbf{X} - \hat{\mathbf{X}})\|$.

- If we assume a linear receiver interface $\mathbf{A}$, the processed signal is $\mathbf{AY}$ and the PEP is given by

$$\mathbb{P}(\mathbf{X} \to \hat{\mathbf{X}}) \;=\; \mathbb{P}\left(\|\mathbf{AY} - \hat{\mathbf{X}}\|^2 < \|\mathbf{AY} - \mathbf{X}\|^2\right),$$

  or

$$\mathbb{P}(\mathbf{X} \to \hat{\mathbf{X}}) = \mathbb{P}\left(\langle \mathbf{AN}, \mathbf{X} - \hat{\mathbf{X}}\rangle < -\frac{\|\mathbf{X} - \hat{\mathbf{X}}\|^2 + 2\langle(\mathbf{AH} - \mathbf{I})\mathbf{X}, \mathbf{X} - \hat{\mathbf{X}}\rangle}{2}\right).$$

  Here again, $\langle \mathbf{AN}, \mathbf{X} - \hat{\mathbf{X}}\rangle$ is just a linear combination of elements of $\mathbf{N}$, and hence it is Gaussian distributed with variance $N_0\|\mathbf{A}^H(\mathbf{X} - \hat{\mathbf{X}})\|^2$. This implies for the PEP, that

$$\mathbb{P}(\mathbf{X} \to \hat{\mathbf{X}}) = \mathbb{E}\left\{Q\left(\frac{\|\mathbf{X} - \hat{\mathbf{X}}\|^2 + 2\langle(\mathbf{AH} - \mathbf{I})\mathbf{X}, \mathbf{X} - \hat{\mathbf{X}}\rangle}{2\sqrt{2N_0\|\mathbf{A}^H(\mathbf{X} - \hat{\mathbf{X}})\|^2}}\right)\right\}. \tag{3.3}$$

We now assume that the system dimension tends to infinity, such that we can apply FPT methods.

### 3.1.2 Applying free probability

In the expressions for the PEPs, we always have to consider the norm of certain matrices. The norms are computed by means of the trace. By assuming the dimension of the system being infinite, we can use the tools of FPT, i.e. the result about the asymptotic pdf of the singular values of a matrix with iid Gaussian entries and the properties that we obtained from the asymptotic free independence. The use of FPT requires the introduction of a normalized trace $\mathrm{tr}_n$. Here we will do the normalization with respect to the number of receive antennas $N_R$, i.e.

$$\mathrm{tr}_n(\cdot) = \mathrm{tr}_{N_R}(\cdot) = \frac{\mathrm{trace}(\cdot)}{N_R}.$$

- In the case of ML detection, (3.2) depends only on the norm $\|\mathbf{H}(\mathbf{X} - \hat{\mathbf{X}})\|^2$, which can also be written as

$$\|\mathbf{H}(\mathbf{X} - \hat{\mathbf{X}})\|^2 = \mathrm{trace}\left(\mathbf{H}(\mathbf{X} - \hat{\mathbf{X}})(\mathbf{X} - \hat{\mathbf{X}})^H\mathbf{H}^H\right) \tag{3.4}$$

  Since the trace has the property $\mathrm{trace}(\mathbf{AB}) = \mathrm{trace}(\mathbf{BA})$, we obtain

$$\|\mathbf{H}(\mathbf{X} - \hat{\mathbf{X}})\|^2 = \mathrm{trace}\left((\mathbf{X} - \hat{\mathbf{X}})(\mathbf{X} - \hat{\mathbf{X}})^H\mathbf{HH}^H\right),$$

where the entries of $\mathbf{H}$ are iid with zero mean and variance $1/N_R$, and where $(\mathbf{X}-\hat{\mathbf{X}})(\mathbf{X}-\hat{\mathbf{X}})^H$ is a constant matrix. If we consider the system in the large limit, we can apply Theorem 2.2.8, which states that $\mathbf{H}\mathbf{H}^H$ and $(\mathbf{X}-\hat{\mathbf{X}})(\mathbf{X}-\hat{\mathbf{X}})^H$ are asymptotically free. Hence the property shown in Example 1.3.1 and used for RMs in (2.20) can be applied. As $N_R \to \infty$,

$$\mathrm{tr}_{N_R}\left((\mathbf{X}-\hat{\mathbf{X}})(\mathbf{X}-\hat{\mathbf{X}})^H\mathbf{H}\mathbf{H}^H\right) = \mathrm{tr}_{N_R}\left((\mathbf{X}-\hat{\mathbf{X}})(\mathbf{X}-\hat{\mathbf{X}})^H\right)\mathrm{tr}_{N_R}(\mathbf{H}\mathbf{H}^H). \tag{3.5}$$

Hence, (3.4) reduces to

$$||\mathbf{H}(\mathbf{X}-\hat{\mathbf{X}})||^2 = ||\mathbf{X}-\hat{\mathbf{X}}||^2\mathrm{tr}_{N_R}(\mathbf{H}\mathbf{H}^H).$$

Assuming $N_R \to \infty$, Theorem 2.2.4 can be applied. Here, $\mathbf{H}\mathbf{H}^H$ fullfills the conditions of the Marčenko-Pastur law. Its moments can be computed using the asymptotic empirical eigenvalue pdf $f_{MP}(x)$. And $\mathrm{tr}_{N_R}(\mathbf{H}\mathbf{H}^H)$ is the first moment of this distribution, i.e.

$$\mathrm{tr}_{N_R}(\mathbf{H}\mathbf{H}^H) = \int x f_{MP}(x)dx = 1.$$

This implies that

$$||\mathbf{H}(\mathbf{X}-\hat{\mathbf{X}})||^2 = ||\mathbf{X}-\hat{\mathbf{X}}||^2,$$

and the PEP (3.2) depends, as the dimension of the system tends to infinity, only on the norm $||\mathbf{X}-\hat{\mathbf{X}}||$,

$$\mathbb{P}(\mathbf{X}\to\hat{\mathbf{X}}) = \mathbb{E}\left\{Q\left(\frac{||\mathbf{X}-\hat{\mathbf{X}}||}{\sqrt{2N_0}}\right)\right\}. \tag{3.6}$$

- An analogue procedure can be done for the case, where a linear interface is used. Let us consider a zero-forcing interface, which is modeled by the matrix

$$\mathbf{A} = \mathbf{H}^+, \qquad \text{with } \mathbf{H}^+ = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H.$$

By replacing $\mathbf{A} = \mathbf{H}^+$ can now be replaced in (3.3) we obtain

$$\mathbb{P}(\mathbf{X}\to\hat{\mathbf{X}}) = \mathbb{E}\left\{Q\left(\frac{||\mathbf{X}-\hat{\mathbf{X}}||^2}{2\sqrt{2N_0||\mathbf{H}^{+H}(\mathbf{X}-\hat{\mathbf{X}})||^2}}\right)\right\}. \tag{3.7}$$

The variance can be rewritten as

$$\begin{aligned} ||\mathbf{H}^{+H}(\mathbf{X}-\hat{\mathbf{X}})||^2 &= \mathrm{trace}\left((\mathbf{X}-\hat{\mathbf{X}})^H\mathbf{H}^+\mathbf{H}^{+H}(\mathbf{X}-\hat{\mathbf{X}})\right) \\ &= \mathrm{trace}\left((\mathbf{X}-\hat{\mathbf{X}})^H(\mathbf{H}^H\mathbf{H})^{-1}(\mathbf{X}-\hat{\mathbf{X}})\right). \end{aligned}$$

The property trace(**AB**) = trace(**BA**) implies that

$$\|\mathbf{H}^{+H}(\mathbf{X}-\hat{\mathbf{X}})\|^2 \;=\; \text{trace}\left((\mathbf{H}^H\mathbf{H})^{-1}(\mathbf{X}-\hat{\mathbf{X}})(\mathbf{X}-\hat{\mathbf{X}})^H\right).$$

Here again, by assuming the system dimension being infinite, we can apply the property shown in Example 1.3.1 and similarly to (3.5) we have

$$\|\mathbf{H}^{+H}(\mathbf{X}-\hat{\mathbf{X}})\|^2 \;=\; \text{tr}_{N_R}((\mathbf{H}^H\mathbf{H})^{-1})\|\mathbf{X}-\hat{\mathbf{X}}\|^2,$$

where $\text{tr}_{N_R}((\mathbf{H}^H\mathbf{H})^{-1})$ can be computed using the Marčenko-Pastur law $f_{MP}(x)$, i.e.

$$\text{tr}_{N_R}\left((\mathbf{H}^H\mathbf{H})^{-1}\right) = \int x^{-1} f_{MP}(x)dx = 1-\beta.$$

Here $\beta = \frac{N_T}{N_R}$. Thus the PEP reduces to

$$\mathbb{P}(\mathbf{X}\to\hat{\mathbf{X}}) = \mathbb{E}\left\{Q\left(\sqrt{1-\beta}\frac{\|\mathbf{X}-\hat{\mathbf{X}}\|}{\sqrt{2N_0}}\right)\right\}.$$

We can observe that using elementary properties of FPT, i.e. (2.20) and the Marčenko-Pastur law, the PEP expressions can be strongly simplified in the system limit. The asymptotic results show us for example, that for large number of antennas, the PEP depends only on the error distance $\|\mathbf{X}-\hat{\mathbf{X}}\|$ with a certain height and that using the ZF interface implies an asymptotic power loss of $(1-\beta)$ with respect to the ML interface.

## 3.2 Asymptotic capacity of correlated MIMO channels

Looking forward to the fourth generation of mobile communication systems, it is of great interest to evaluate the performance of MIMO channels under a correlated fading model. The channel models that are often used are stochastic channel models which implies the use of random matrices. In several publications, the asymptotic capacity has been considered and has given a good approximation of the capacity for the finite dimensional case [7, 15, 16, 17]. The assumption of a infinite number of transmit and receive antennas, hence of a large systems enables us to use the theorems for large random matrices (see Section 2.2.2). As we will see in the following, the capacity can be evaluated using the eigenvalue distribution of the Hermitian product of the channel matrix. Some simple results are known about the asymptotic eigenvalue distribution, which helps to evaluate the asymptotic capacity. Additionally, the use of simple models will enable us to apply results from FPT (here we will use the multiplicative free convolution).

### 3.2.1 Capacity of MIMO Channels

Assume that we have a channel with $N_T$ transmit and $N_R$ receive antennas. The corresponding input/output relation is given by a vector matrix equation [21]

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}\,, \tag{3.8}$$

where

- $\mathbf{x}$ is a $N_T \times 1$ vector representing the input signal,
- $\mathbf{y}$ is a $N_R \times 1$ vector representing the output signal,
- $\mathbf{n}$ is a $N_R \times 1$ vector modelling the noise, and
- $\mathbf{H}$ is the $N_R \times N_T$ channel matrix.

The total channel input power is constrained to $P$, i.e.

$$\mathbb{E}\{\mathbf{x}^H\mathbf{x}\} = \sum_i |x_i|^2 = P. \tag{3.9}$$

The noise power is assumed zero mean white Gaussian with variance $\sigma^2$, i.e.

$$\mathbb{E}\{\mathbf{n}\mathbf{n}^H\} = \sigma^2\mathbf{I}_{N_R}.$$

The channel matrix is normalized to

$$\mathbb{E}\{\mathrm{trace}(\mathbf{H}\mathbf{H}^H)\} = \sum_{i,j}\mathbb{E}\{|\mathbf{H}_{i,j}|^2\} = N_T N_R. \tag{3.10}$$

In the following we will assume a constant ratio $\beta$ of transmit to receive antennas

$$\beta \triangleq \frac{N_T}{N_R}.$$

With this channel model, the channel capacity is given by [21]

$$C = \log\left(\det\left(\mathbf{I} + \frac{\mathrm{SNR}}{N_T}\mathbf{H}\mathbf{H}^H\right)\right), \tag{3.11}$$

where the SNR is the ratio between the total input power $P$ and the noise power $\sigma^2$

$$\mathrm{SNR} = \frac{P}{\sigma^2}. \tag{3.12}$$

The relation (3.11) depends on the channel matrix $\mathbf{H}$ but can also be rewritten depending only on the eigenvalues $\lambda_i$ of the matrix $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$ [21],

$$C = \sum_{i=1}^{N_R} \log\left(1 + \frac{\text{SNR}}{\beta}\lambda_i\right). \tag{3.13}$$

This expression can again be transformed and rewritten using the empirical eigenvalue pdf $f^e(x)$ (introduced in Section 2.2.1) of the matrix $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$ by using relation (2.9):

$$\frac{C}{N_R} = \int \log\left(1 + \frac{\text{SNR}}{\beta}x\right) f^e(x)dx. \tag{3.14}$$

If the channel matrix is assumed to be random, the capacity becomes a random value as well. In order to get a non-random value for the capacity of the channel, one usually considers the *ergodic capacity*:

$$C_{\text{erg}} = \mathbb{E}\{C\}. \tag{3.15}$$

The ergodic capacity depends on the mean eigenvalue distribution of $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$ with the pdf $f^m(x)$ (introduced in Section 2.2.1)

$$\frac{C_{\text{erg}}}{N_R} = \int \log\left(1 + \frac{\text{SNR}}{\beta}x\right) f^m(x)dx. \tag{3.16}$$

Thus, the empirical and the mean eigenvalue distributions of RMs can be used to compute the capacity of a MIMO channel. Note that the results about eigenvalue distributions that have been treated until now (see Chapter 2) are those for RMs in the large limit (Section 2.2.2). Under the assumption of a large system, the capacity can be computed using the asymptotic eigenvalue distribution. In the following, the results of large RMs and the theorems of FPT will be used to evaluate the asymptotic capacity of MIMO channels. This asymptotic value is a good approximation of the capacity for the finite dimensional case and it will enable us to study the influence of the correlation on the capacity.

### 3.2.2 Asymptotic capacity for the canonical channel model

The simplest model for a MIMO channel, not taking into account the correlation between the antennas is the so called canonical channel model. Here, $\mathbf{H}$ is assumed to have iid Gaussian entries with zero mean and unit variance. This matrix fullfills the normalization condition (3.10). Under the assumption that the number of antennas on the transmitter and the receiver side tends to infinity with a fixed ratio $\beta = N_T/N_R$, the empirical eigenvalue pdf of the matrix $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$ is known. It is the pdf of the Marčenko-Pastur law (see Theorem 2.2.4). Using this asymptotic eigenvalue pdf and (3.14), the asymptotic capacity of the canonical channel model can be evaluated [22].

$$\begin{aligned}\frac{C}{N_R} &= \beta \log\left(1+\frac{\mathrm{SNR}}{\beta}-\frac{1}{4}\mathcal{F}\left(\frac{\mathrm{SNR}}{\beta},\beta\right)\right) \\ &\quad + \log\left(1+\mathrm{SNR}-\frac{1}{4}\mathcal{F}\left(\frac{\mathrm{SNR}}{\beta},\beta\right)\right) \\ &\quad - \frac{\beta \log e}{4\,\mathrm{SNR}}\mathcal{F}\left(\frac{\mathrm{SNR}}{\beta},\beta\right), \end{aligned} \tag{3.17}$$

with

$$\mathcal{F}\left(\frac{\mathrm{SNR}}{\beta},\beta\right) = \left(\sqrt{\frac{\mathrm{SNR}}{\beta}\left(1+\sqrt{\beta}\right)^2+1} - \sqrt{\frac{\mathrm{SNR}}{\beta}\left(1-\sqrt{\beta}\right)^2+1}\right)^2. \tag{3.18}$$

The resulting capacity per receive antenna is illustrated in Figure 3.1 for different values of $\beta$.

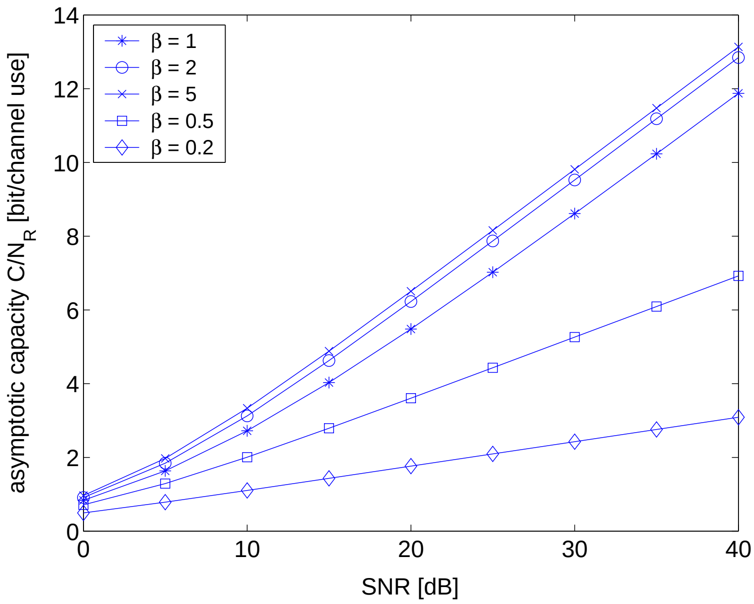



Figure 3.1: Asymptotic capacity per receive antenna versus SNR for the canonical model for different values of $\beta = 0.2, 0.5, 1, 2, 5$.

In Figure 3.2, the capacity of a system with 8 receive antennas for different values of $\beta$ has been computed. We can observe, that the asymptotic capacity gives a good approximation even for small systems.
In Figure 3.3, for a given high SNR (30dB), the asymptotic capacity versus $\beta$ is shown. The resulting curve illustrates a result already found in [25, 7]. The first part of the

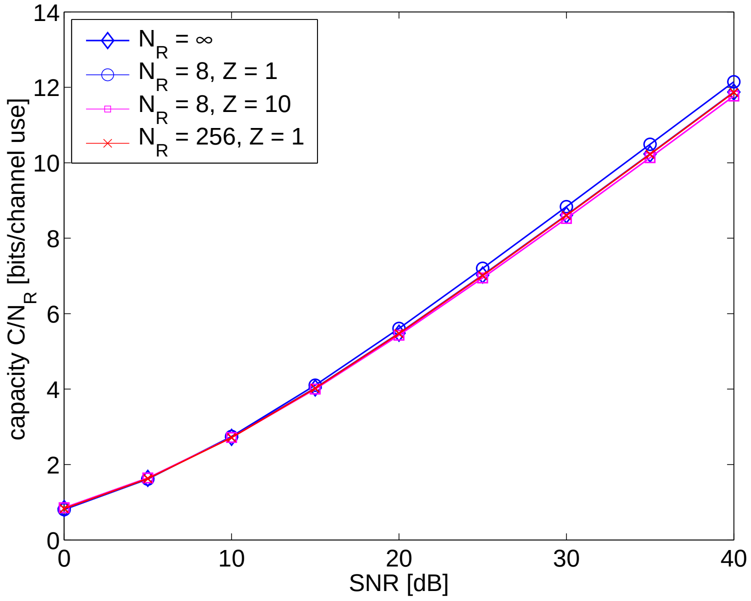



Figure 3.2: Capacity per receive antenna versus SNR for the finite dimensional canonical model for $\beta = 1$ and different sizes $N_R \times N_R$ of matrices and different numbers Z of matrix samples.

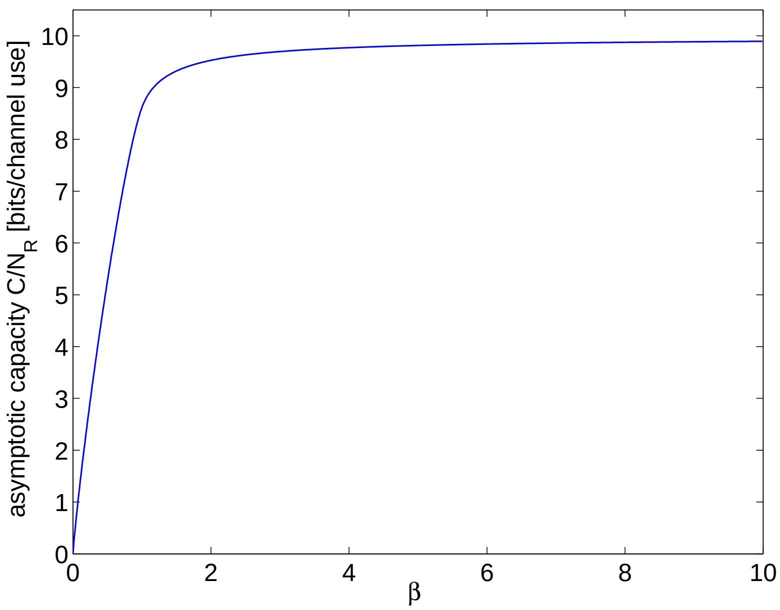



Figure 3.3: Asymptotic capacity per receive antenna versus $\beta$ for a high SNR $=$ 30dB, for the canonical model.

curve grows linearly with $\beta$, hence for a larger number of receive than transmit antennas ($\beta < 1$),

$$\frac{C}{N_R} \propto \beta \quad \Rightarrow \quad C \propto N_T.$$

For $\beta > 1$, hence for a larger number of transmit than receive antennas, the curve gets flat,

$$\frac{C}{N_R} \propto \text{const.} \quad \Rightarrow \quad C \propto N_R.$$

This result for the asymptotic capacity and its relation to the number of antennas, confirms the result for high SNR in [25, 7], that states that the capacity grows linearly with the smaller number of transmit or receive antennas, i.e.

$$C \propto \min\{N_T, N_R\}.$$

### 3.2.3 Asymptotic capacity for correlated MIMO channel

#### The Kronecker model

The canonical model assume that there is no correlation between the fading gains of the channel matrix. This is not a very realistic assumption [3]. The simplest way to model a correlated MIMO channel is the Kronecker Model (used and verified in [3, 15]). In the Kronecker model, the channel matrix is modelled as

$$\mathbf{H} = \mathbf{\Theta}_R^{1/2}\mathbf{G}\mathbf{\Theta}_T^{1/2}, \tag{3.19}$$

with

- $\mathbf{G}$ denoting a $N_R \times N_T$ matrix with iid complex Gaussian entries, with zero mean and unit variance. Here, $\mathbf{G}$ fullfills the normalization condition (3.10).
- $\mathbf{\Theta}_T$ and $\mathbf{\Theta}_R$ denoting $N_T \times N_T$ and $N_R \times N_R$ matrices modelling the correlation of the channel at the transmitter and the receiver side, respectively. The correlation matrices are supposed to respect the following relations:

$$\begin{aligned} N_T\mathbf{\Theta}_R = \mathbb{E}\{\mathbf{H}\mathbf{H}^H\}, \quad N_R\mathbf{\Theta}_T = \mathbb{E}\{\mathbf{H}^T\mathbf{H}^*\} \\ \text{trace}(\mathbf{\Theta}_R) = N_R, \quad \text{trace}(\mathbf{\Theta}_T) = N_T. \end{aligned} \tag{3.20}$$

  The receiver and transmitter correlation matrices are in relation with the correlation matrix $\mathbf{\Theta}_\mathbf{H}$ of the channel matrix by means of the Kronecker product, which explains the name of the model, i.e.

$$\mathbf{\Theta}_\mathbf{H} = \mathbb{E}\{\text{vec}(\mathbf{H})\text{vec}(\mathbf{H})^H\} = \mathbf{\Theta}_T \otimes \mathbf{\Theta}_R.$$

The model implies that the correlation between two transmit antennas is the same regardless of the receive antenna at which the observation made and vice versa, which, in turn means that the correlation process at either end of the link are independent.

### Simplified Kronecker model

In the following, we assume only correlation at the receiver side in order to simplify the calculations and to enable us to use free probability tools. That is, we consider

$$\mathbf{\Theta}_T = \mathbf{I}, \qquad \mathbf{\Theta}_R = \mathbf{\Theta} \quad \Rightarrow \quad \mathbf{H} = \mathbf{\Theta}^{1/2}\mathbf{G}.$$

A similar approach is proposed in [16].

### Asymptotic capacity for simplified Kronecker model

In order to obtain the asymptotic capacity for this model with the expression (3.14), we need the asymptotic empirical eigenvalue distribution of the matrix $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$, that is given by

$$\frac{1}{N_R}\mathbf{H}\mathbf{H}^H = \frac{1}{N_R}\mathbf{\Theta}^{1/2}\mathbf{G}\mathbf{G}^H\mathbf{\Theta}^{H/2}.$$

Recall that the trace of the product of two matrices $\mathbf{A}$ and $\mathbf{B}$ satisfies

$$\text{trace}(\mathbf{AB}) = \text{trace}(\mathbf{BA}),$$

which implies that

$$\text{trace}\left(\frac{1}{N_R}\mathbf{H}\mathbf{H}^H\right) = \text{trace}\left(\frac{1}{N_R}\mathbf{G}\mathbf{G}^H\mathbf{\Theta}\right).$$

Thus, the empirical eigenvalue distribution of $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$ is the same as the empirical eigenvalue distribution of $\frac{1}{N_R}\mathbf{G}\mathbf{G}^H\mathbf{\Theta}$. Note that we need the empirical distribution of the product of two matrices, i.e. $\frac{1}{N_R}\mathbf{G}\mathbf{G}^{\mathbf{H}}$ and $\mathbf{\Theta}$. We know the asymptotic empirical eigenvalue pdf of the matrix $\frac{1}{N_R}\mathbf{G}\mathbf{G}^{\mathbf{H}}$: it is the Marčenko-Pastur law. Furthermore $\mathbf{\Theta}$ is assumed to be a constant matrix with known asymptotic empirical eigenvalue distribution. From Theorem 2.2.3, we know that these two matrices are asymptotically free. Thus, the asymptotic eigenvalue pdf of the product can be computed by means of the asymptotic free convolution (see Section 2.2.4). With this asymptotic eigenvalue pdf and relation (3.14), the asymptotic capacity for channels modeled with the simplified Kronecker model can be obtained. For special models of the channel correlation matrix $\mathbf{\Theta}$, this asymptotic capacity will be computed in the following.

## 3.2.4 Asymptotic capacity for bandmatrix correlation model

At first we investigate the asymptotic capacity using a bandmatrix as correlation matrix (like the one introduced in Example 2.2.7). This corresponds to a system, where the fading gains of neighbored antennas are correlated with a constant factor $\rho$. Thus, the correlation matrix $\mathbf{\Theta}$ is then given by

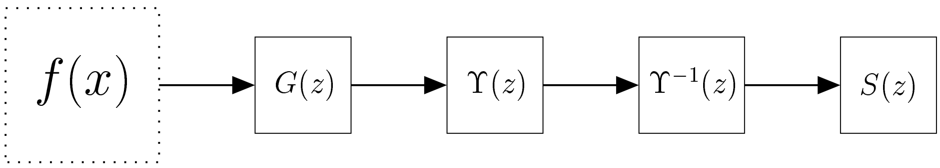
$$\mathbf{\Theta} = \begin{pmatrix} 1 & \rho & 0 & \dots & 0 \\ \rho & 1 & \rho & \ddots & \vdots \\ 0 & \rho & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 & \rho \\ 0 & \dots & 0 & \rho & 1 \end{pmatrix}.$$

In order to evaluate the asymptotic capacity, we first have to calculate the asymptotic eigenvalue pdf of the matrix $\frac{1}{N_R}\mathbf{G}\mathbf{G}^H\mathbf{\Theta}$, which is obtained over the free multiplicative convolution of the asymptotic empirical eigenvalue pdfs of $\mathbf{\Theta}$ and $\frac{1}{N_R}\mathbf{G}\mathbf{G}^H$. The steps to follow are the ones of Example 1.7.2. In the following, we will obtain the S-transform of the matrix $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$, after computing the S-transforms of $\mathbf{\Theta}$ and $\frac{1}{N_R}\mathbf{G}\mathbf{G}^H$ in the asymptotic case.

### S-transform of $\mathbf{\Theta}$

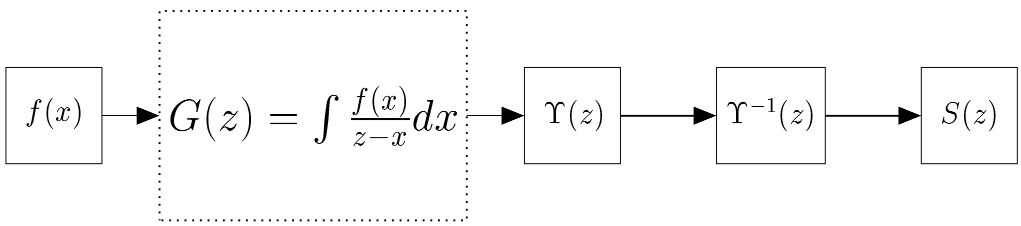
The steps 1-4 of Example 1.7.2 are followed.

- Getting the eigenvalue pdf $f_{\mathbf{\Theta}}(x)$:



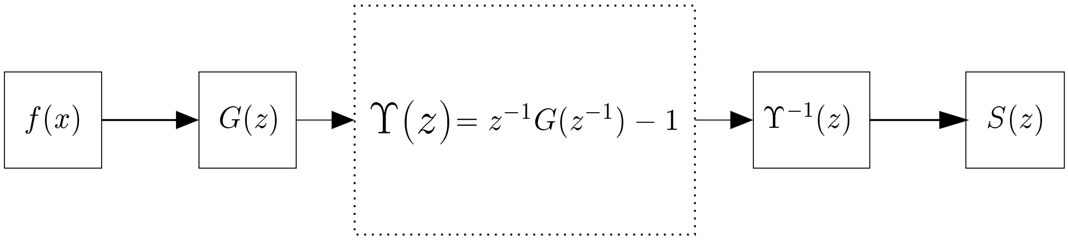

$$f_{\mathbf{\Theta}}(x) = \frac{1}{\pi}\frac{1}{\sqrt{4\rho^2 - (x-1)^2}}.$$

- Computing the Cauchy-transform $G_{\mathbf{\Theta}}(z)$:



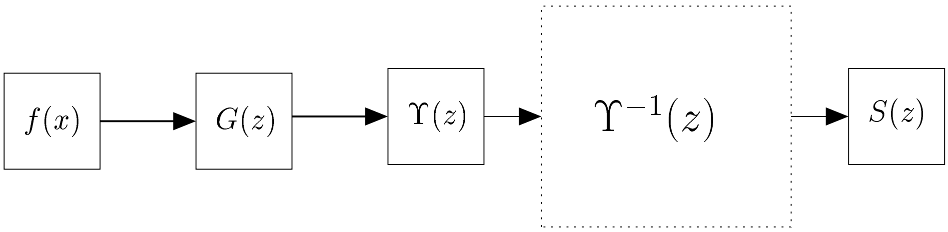

$$G_{\mathbf{\Theta}}(z) = i\frac{1}{\sqrt{4\rho^2 - (z-1)^2}}.$$

- Computing the intermediate transform $\Upsilon_{\mathbf{\Theta}}(z)$:




$$\Upsilon_{\mathbf{\Theta}}(z) = i\frac{1}{\sqrt{4\rho^2 z^2 - (1-z)^2}} - 1.$$

- Computing the inverse $\Upsilon_{\mathbf{\Theta}}^{-1}(z)$:




The inverse of $\Upsilon_{\mathbf{\Theta}}(z)$ is obtained by solving the equation, with respect to $z$

$$i\frac{1}{\sqrt{4\rho^2 z^2 - (1-z)^2}} - 1 = w,$$

which can rewritten as

$$(4\rho^2 - 1)z^2 + 2z + \frac{1}{(w+1)^2} - 1 = 0.$$

The solution of the equation is

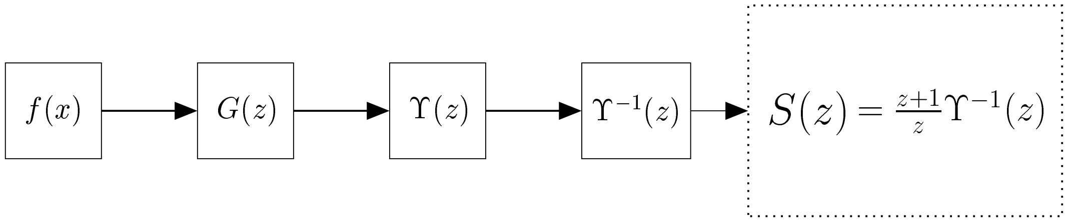
$$z = \frac{-2 \pm \sqrt{4 - 4(4\rho^2 - 1)\left(\frac{1}{(w+1)^2} - 1\right)}}{2(4\rho^2 - 1)}.$$

Hence, the inverse of $\Upsilon_{\mathbf{\Theta}}(z)$ is

$$\Upsilon_{\Theta}^{-1}(z) = \frac{-1 \pm \sqrt{4\rho^2 - \Delta\frac{1}{(z+1)^2}}}{\Delta},$$

with $\Delta = 4\rho^2 - 1$.

- Computing the S-transform $S_{\mathbf{\Theta}}(z)$:

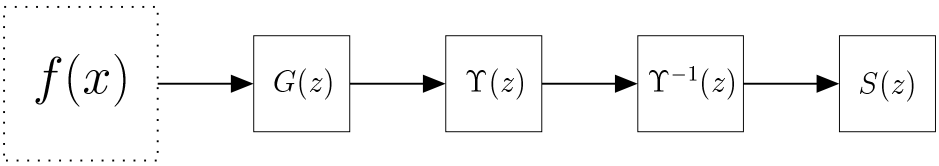



$$S_{\mathbf{\Theta}}(z) = \frac{z+1}{z} \frac{-1 \pm \sqrt{4\rho^2 - \Delta \frac{1}{(z+1)^2}}}{\Delta}$$

**S-transform of $\frac{1}{N_R}\mathbf{G}\mathbf{G}^H$**

The same steps 1-4 of Example 1.7.2 are followed to compute the S-transform of the matrix $\mathbf{D} = \frac{1}{N_R}\mathbf{G}\mathbf{G}^H$. Here, $\mathbf{D}$ fullfills the conditions of the Marčenko-Pastur law (Theorem 2.2.4).

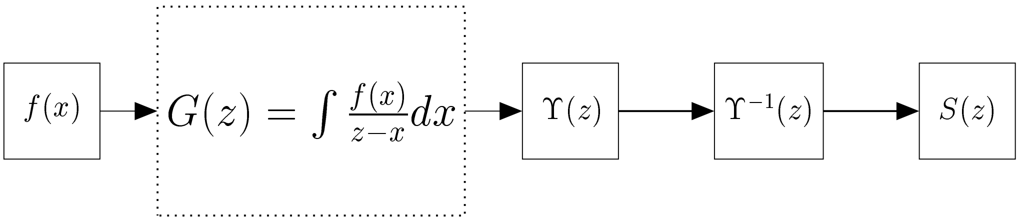
- Getting the eigenvalue pdf $f_{\mathbf{D}}(x)$:




It is the pdf of the Marčenko-Pastur law (2.11), i.e.

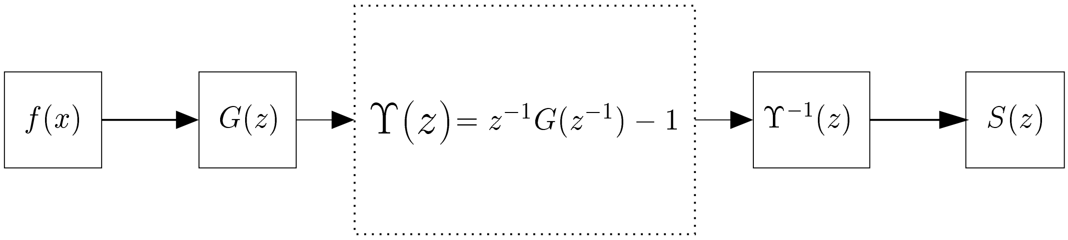
$$f_{\mathbf{D}}(x) = \begin{cases} \frac{\sqrt{4\beta-(x-1-\beta)^2}}{2\pi x}, & (1-\sqrt{\beta})^2 < x < (1+\sqrt{\beta})^2, \\ [1-\beta]^+\delta(x), & \text{elsewhere.} \end{cases}$$

- Computing the Cauchy-transform $G_{\mathbf{D}}(z)$:



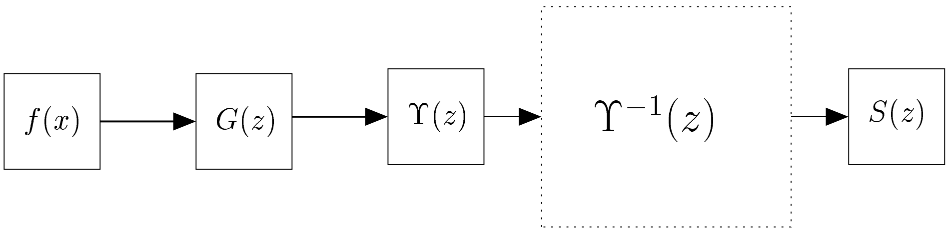

$$G_{\mathbf{D}}(z) = \frac{z+1-\beta \pm \sqrt{z^2 - 2(\beta+1)z + (\beta-1)^2}}{2z}.$$

- Computing the intermediate transform $\Upsilon_{\mathbf{D}}(z)$:



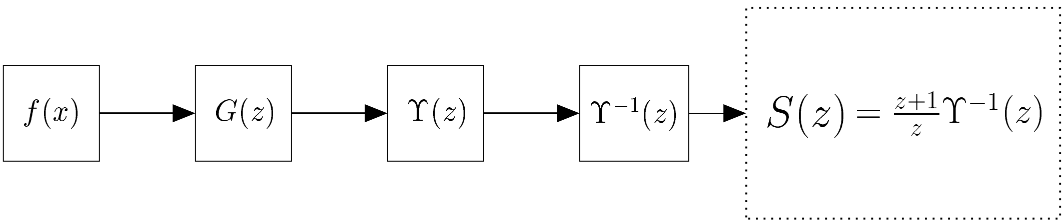

$$\Upsilon_{\mathbf{D}}(z) = \frac{1 - z - \beta z \pm \sqrt{1 - 2(\beta+1)z + (\beta+1)^2}}{2z}.$$

- Computing the inverse $\Upsilon_{\mathbf{D}}^{-1}(z)$:




$$\Upsilon_{\mathbf{D}}^{-1}(z) = \frac{z}{z^2 + z\beta + z + \beta}.$$

- Computing the S-transform $S_{\mathbf{D}}(z)$:




$$S_{\mathbf{D}}(z) = \frac{1}{z+\beta}.$$

### S-transform of $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$

As mentioned in Section 3.2.3, we obtain the S-transform of $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$ by considering that it is the same as $\frac{1}{N_R}\mathbf{G}\mathbf{G}^H\mathbf{\Theta}$. The matrices $\mathbf{\Theta}$ and $\mathbf{D} = \frac{1}{N_R}\mathbf{G}\mathbf{G}^H$ are asymptotically free (see Theorem 2.2.6). Hence, the S-transform $S(z)$ of $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$ is given by

$$\begin{aligned} S(z) &= S_{\mathbf{D}}(z)S_{\mathbf{\Theta}}(z) = \frac{1}{z+\beta}\frac{z+1}{z}\Upsilon_{\mathbf{\Theta}}^{-1}(z) \\ &= \frac{1}{z+\beta}\frac{z+1}{z}\frac{-1 \pm \sqrt{4\rho^2 - \Delta\frac{1}{(z+1)^2}}}{\Delta}. \end{aligned} \tag{3.21}$$

### Eigenvalue distribution of $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$

We have computed the S-transform of the matrix $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H(x)$; thus we are in fact able to obtain the empirical eigenvalue distribution by following the steps back, like it is explained in Example 1.7.2, however, doing it directly demands inverting functions analytically, which can get very complicated. In [4] and [17] an implicite solution is proposed. The relation between the S-transform, the Cauchy-transform and the eigenvalue distribution is used to obtain an equation, which solution leads us to the eigenvalue distribution of $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H(x)$.

- Computing back the inverse intermediate transform $\Upsilon^{-1}(z)$:

$$\Upsilon^{-1}(z) = \frac{1}{z+\beta}\frac{-1 \pm \sqrt{4\rho^2 - \Delta\frac{1}{(z+1)^2}}}{\Delta}.$$

- Computing the intermediate transform $\Upsilon(z)$:
  To obtain $\Upsilon(z)$, we have to invert $\Upsilon^{-1}(z)$, hence to solve the equation

$$\Upsilon^{-1}(z) = w,$$

  which is equivalent to

$$\pm\sqrt{4\rho^2 - \Delta\frac{1}{(z+1)^2}} = \Delta w(z+\beta) + 1.$$

  After several transformations, we obtain

$$\Delta w^2(z+\beta)^2(z+1)^2 + 2w(z+\beta)(z+1)^2 - (z+1)^2 + 1 = 0. \tag{3.22}$$

Recall that we started with

$$w = \Upsilon^{-1}(z), \qquad z = \Upsilon(w)$$

Hence solving Equation (3.22) with respect to $z$, leads us to $\Upsilon(w)$.

- Getting the Cauchy-transform $G(z)$:

  We could solve the equation now, and follow the steps to obtain the Cauchy transform $G(z)$ and the eigenvalue distribution in the usual way as it has been shown in Section 1.7.2. The idea is here not to solve the equation immediately. Instead, $\Upsilon(z)$ is replaced by what it represents in terms of $G(z)$. The equation can then be solved and the eigenvalue pdf $f(x)$ of $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$ is derived from $G(z)$. Equation (3.22) is rewritten by renaming the variables and using relation (1.9)

  $$w = \nu^{-1} \quad z = \Upsilon(\nu^{-1}) = \nu G(\nu) - 1 = \nu G - 1.$$

  We obtain an equation of the fourth order in $G(\nu) = G$ with parameters $\rho$, $\beta$ and $\nu$:

  $$\Delta\beta^2\nu^2G^4 + 2\beta\nu(\nu + \Delta(1-\beta))G^3 + (\Delta(1-\beta)^2 + 2(1-\beta)\nu - \nu^2)G^2 + 1 = 0. \quad (3.23)$$

- Getting the eigenvalue pdf:

  To obtain $f(x)$ again, we use relation (1.9):

  $$f(x) = \lim_{y \to 0^+} \Im\{G(x + iy)\}.$$

  This implies that we have to solve Equation (3.23) for real $\nu$, and the imaginary part of the solution tends then to the asymptotic empirical eigenvalue pdf of $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$.
  The equation can be solved analytically by means of Cardano's formula.

The resulting eigenvalue pdfs for $\beta = 1$ are shown in Figure 3.4 and for $\beta = 0.2$ in Figure 3.5, hence for five times as many receive than transmit antennas.
The shape of the functions remain very similar to the Marčenko-Pastur distribution. For a higher correlation coefficient, the value of the largest possible eigenvalue increases, but we get also with higher probability smaller eigenvalues, which fits to the result, that for higher correlation, the channel matrix is more likely to be ill conditioned. For finite dimensional matrices $\mathbf{H}$, that fullfill the conditions of our simplified Kronecker model with a bandmatrix correlation, the eigenvalue pdf is illustrated for $\beta = 1$ and $\rho = 0.4$ in Figure 3.6 and compared to the asymptotic pdf, we have computed using free multiplicative convolution. We can observe that the asymptotic eigenvalue pdf is a good approximation for the finite dimensional case.

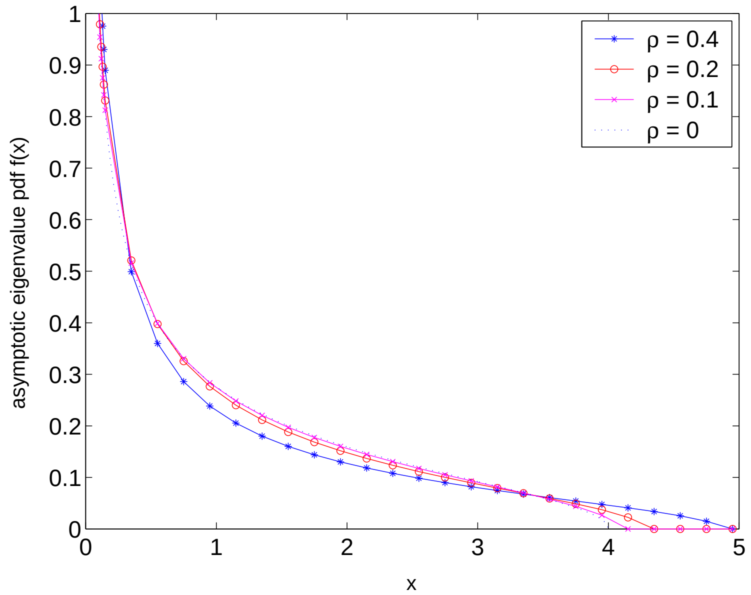



Figure 3.4: Asymptotic eigenvalue pdfs of a simplified Kronecker model matrix with bandmatrix correlation, with fixed $\beta = 1$ and correlation parameter $\rho = 0.4, 0.2, 0.1, 0$.

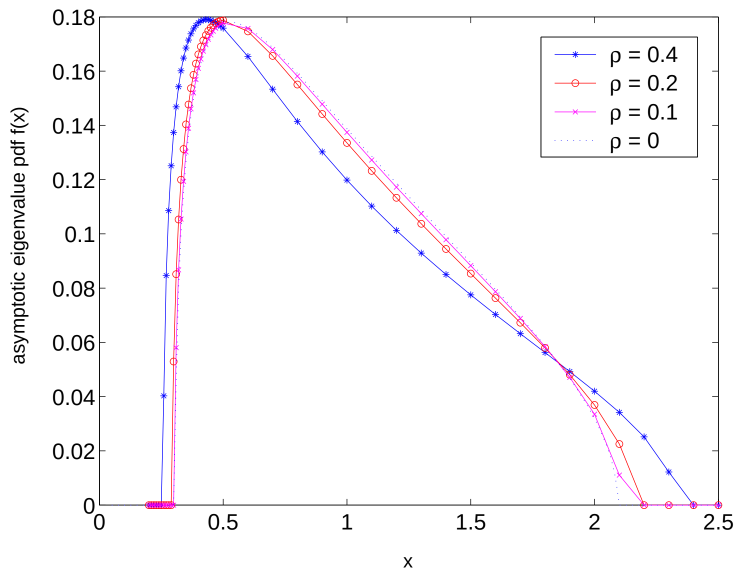



Figure 3.5: Asymptotic eigenvalue pdfs of a simplified Kronecker model matrix with bandmatrix correlation, with fixed $\beta = 0.2$ and correlation parameter $\rho = 0.4, 0.2, 0.1, 0$.

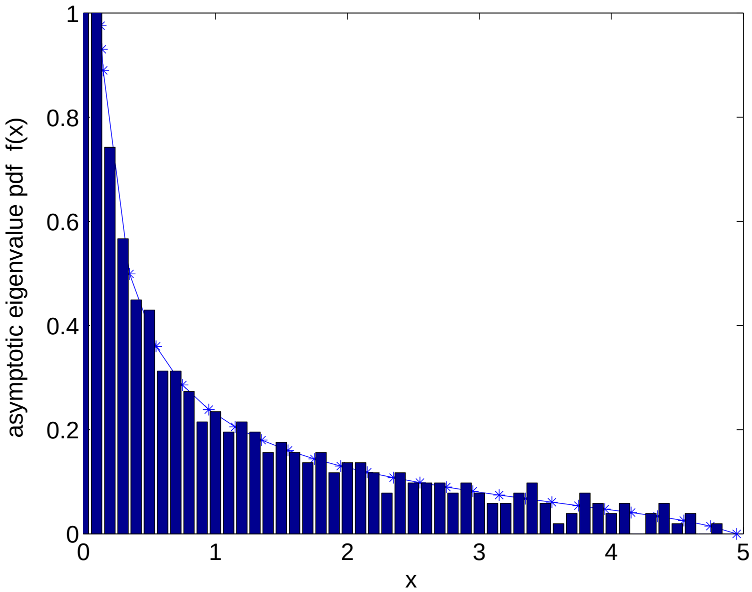



Figure 3.6: Normalized histogram of a simplified Kronecker Model matrix $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$ for $N_R = 512$, with a bandmatrix as correlation matrix with parameter $\rho = 0.4$, $\beta = 0.4$.

#### Calculating the asymptotic capacity

Now that we have the asymptotic empirical eigenvalue pdf for our model, we are able to compute the asymptotic capacity per receive antenna using relation (3.14). The integration is done numerically, the resulting asymptotic capacity is illustrated for different values of the correlation coefficient $\rho$ in Figure 3.7 for $\beta = 1$ and in Figure 3.8 for $\beta = 0.2$. As expected the capacity per receive antenna decreases with the correlation coefficient. The capacity per receive antennas for finite dimensional model matrices has been computed for some samples of the channel matrix and shown in Figure 3.9. We can observe that the asymptotic result is a good approximation. In Figure 3.10, the asymptotic capacity per receive antenna versus $\beta$ is shown for high SNR (30dB) and for a fixed correlation coefficient $\rho = 0.5$. We can observe the same result as for the canonical model. The capacity for a high SNR and for a fixed correlation coefficient is proportional to the minimum number of transmit or receive antennas.

### 3.2.5 Asymptotic capacity for exponential correlation model

A more elaborated model for the correlation matrix is a Toeplitz matrix with exponential coefficients, like the one of Example 2.2.8 with $n = N_R$.

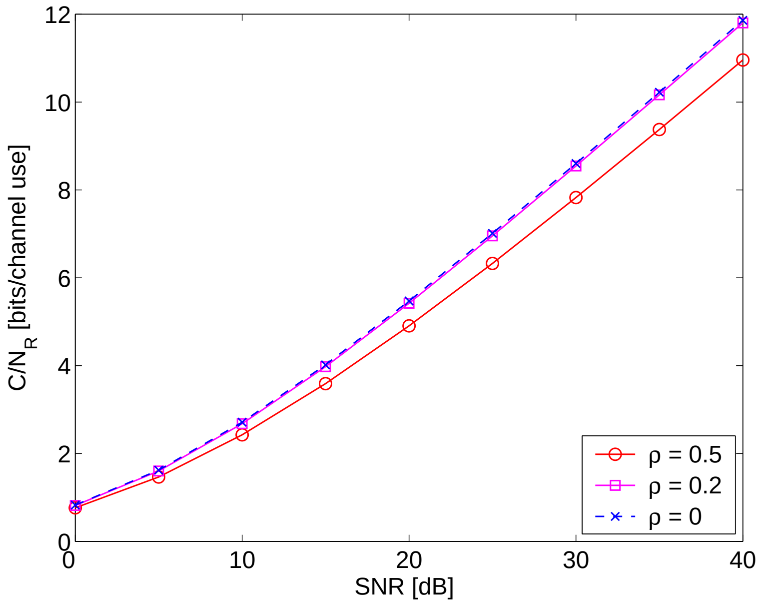



Figure 3.7: Asymptotic capacity for the simplified Kronecker model with bandmatrix correlation for fixed $\beta = 1$ and $\rho = 0.5, 0.2, 0$.

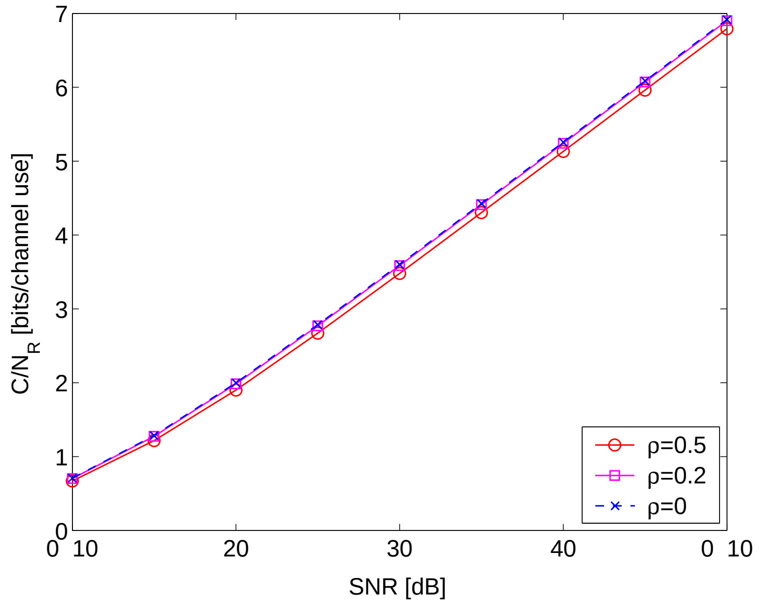



Figure 3.8: Asymptotic capacity for the simplified Kronecker model with bandmatrix correlation for fixed $\beta = 0.5$ and $\rho = 0.5, 0.2, 0$.

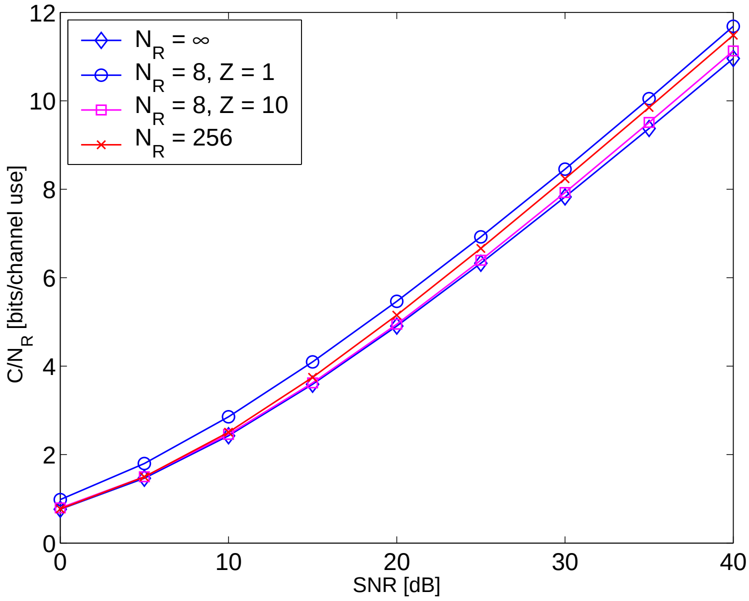



Figure 3.9: Capacity for finite dimensional matrices that fullfill the simplified Kronecker model with bandmatrix correlation matrix, for fixed $\beta = 1$, $\rho = 0.5$ for one $8 \times 8$, ten $8 \times 8$ and one $256 \times 256$ matrices.

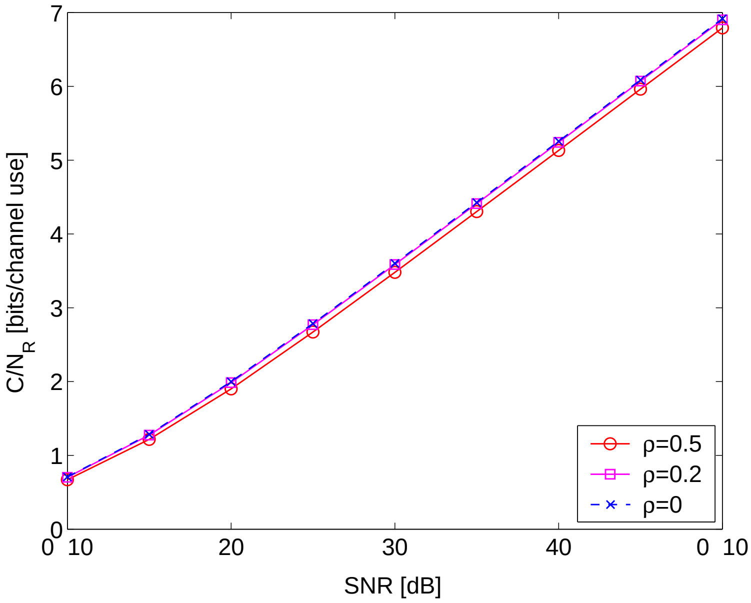


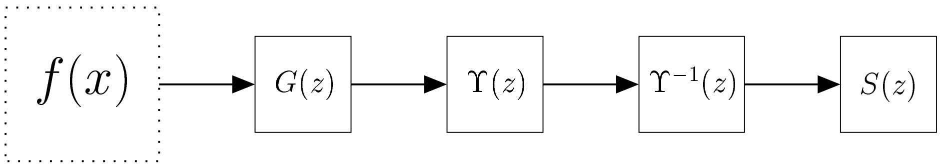

Figure 3.10: Asymptotic capacity versus $\beta$ for high SNR $=$ 30dB and $\rho = 0.5$ for the simplified Kronecker model with bandmatrix correlation.

$$\mathbf{\Theta} = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \ddots & \vdots \\ \rho^2 & \rho & \ddots & \ddots & \rho^2 \\ \vdots & \ddots & \ddots & 1 & \rho \\ \rho^{n-1} & \dots & \rho^2 & \rho & 1 \end{pmatrix}.$$

In the following, the same steps as in Section 3.2.4 are done in order to compute the asymptotic empirical eigenvalue pdf of $\frac{1}{N_R}\mathbf{HH}$ and after that the asymptotic capacity under this channel model.

**S-transform of $\mathbf{\Theta}$**

- Getting the eigenvalue pdf $f_{\mathbf{\Theta}}(x)$:

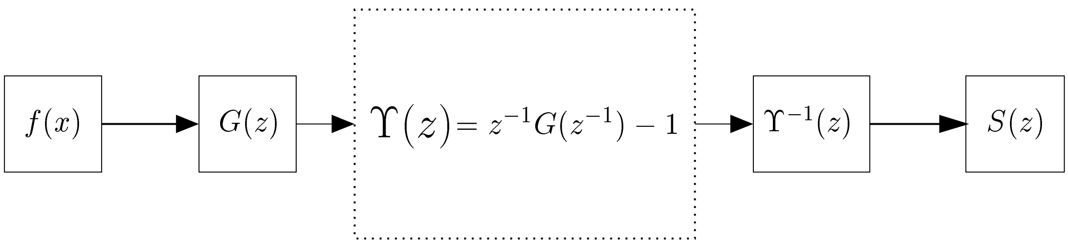



$$f_{\mathbf{\Theta}}(x) = \frac{1}{\pi x}\frac{1}{\sqrt{-x^2+2x\alpha-1}},$$

with

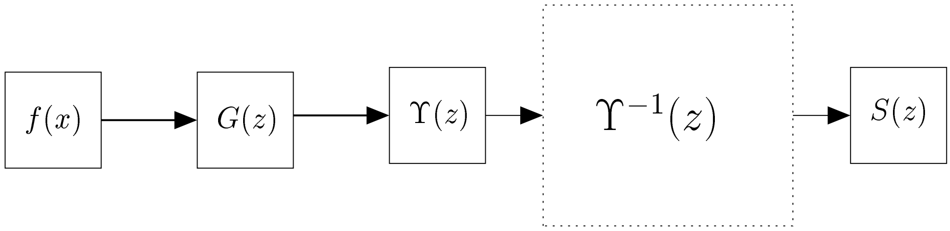
$$\alpha = \left(\frac{1+\rho^2}{1-\rho^2}\right).$$

- Computing the intermediate transform $\Upsilon_{\mathbf{\Theta}}(z)$:




$$\begin{aligned} \Upsilon_{\mathbf{\Theta}}(z) &= \int_{-\infty}^{\infty} \frac{zx}{1-zx} f(x)dx \\ &= \frac{z}{\sqrt{z^2-2z\alpha+1}} \end{aligned} \tag{3.24}$$

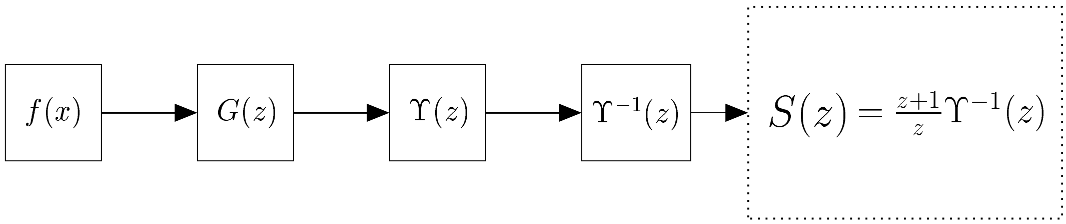
- Computing the inverse $\Upsilon_{\mathbf{\Theta}}^{-1}(z)$:




The function $\Upsilon_{\mathbf{\Theta}}^{-1}(z)$ is obtained by solving the equation $\Upsilon_{\mathbf{\Theta}}(z) = w$, with respect to $z$:

$$\Upsilon_{\mathbf{\Theta}}^{-1}(z) = \frac{\alpha z^2 \pm z\sqrt{\alpha^2 z^2 - (z^2 - 1)}}{(z^2 - 1)}.$$

- Computing the S-transform $S_{\mathbf{\Theta}}(z)$:




$$S_{\mathbf{\Theta}}(z) = \frac{\alpha z \pm \sqrt{\alpha^2 z^2 - (z^2 - 1)}}{(z - 1)}.$$

**S-transform of $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$**

Here again, the matrices $\frac{1}{N_R}\mathbf{G}\mathbf{G}^H$ and $\mathbf{\Theta}$ are asymptotically free. The S-transform of $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$ is the product of the S-transforms $S_{\mathbf{D}}(z)$ and $S_{\mathbf{\Theta}}(z)$.

$$S(z) = \frac{1}{z + \beta}\frac{\alpha z \pm \sqrt{\alpha^2 z^2 - (z^2 - 1)}}{(z - 1)}.$$

**Eigenvalue distribution of $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$**

- Computing back the inverse intermediate transform $\Upsilon^{-1}(z)$

$$\Upsilon^{-1}(z) = \frac{1}{z + \beta}\frac{\alpha z^2 \pm z\sqrt{\alpha^2 z^2 - (z^2 - 1)}}{(z^2 - 1)}.$$

- To obtain $\Upsilon(z)$, we need to solve the equation $\Upsilon^{-1}(z) = w$, with respect to $z$. This equation is equivalent to solving

$$w^2(z^2 - 1)(z + \beta)^2 - z^2(2\alpha w(z + \beta) - 1) = 0. \tag{3.25}$$

- Getting the Cauchy-transform

  Equation (3.25) is rewritten by renaming the variables and using relation (1.9)

$$w = \nu^{-1} \quad z = \Upsilon(\nu^{-1}) = \nu G(\nu) - 1 = \nu G - 1.$$

  We obtain the following equation of fourth order in $G(\nu) = G$ with parameters $\beta, \alpha, \nu$:

$$c_4 G^4 + c_3 G^3 + c_2 G^2 + c_1 G + c_0 = 0. \tag{3.26}$$

  the coefficients $c_i$, $i = 0 \ldots 5$, being

$$\begin{aligned}
c_4 &= \nu^4, \\
c_3 &= \nu^3 \left(2(\beta - 1) - 2 - 2\alpha\nu\right), \\
c_2 &= \nu^2 \left((\beta - 1)^2 - 4(\beta - 1) - 2\alpha(\beta - 1)\nu + 4\alpha\nu + \nu^2\right), \\
c_1 &= \nu \left(-2(\beta - 1)^2 + 4\alpha(\beta - 1)\nu - 2\alpha\nu - 2\nu^2\right), \\
c_0 &= -2\alpha\nu(\beta - 1) + \nu^2.
\end{aligned}$$

- Getting the eigenvalue pdf

  To obtain $f(x)$ again, we use relation (1.9):

$$f(x) = \lim_{y \to 0^+} \Im\{G(x + iy)\}$$

  As for the bandmatrix model (see Section 3.2.4), we have to solve Equation (3.26) for real $\nu$, and the imaginary part of the solution tends then to the asymptotic empirical eigenvalue pdf of $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$.
  The equation can be solved analytically by means of Cardano's formula.

The resulting eigenvalue pdfs are shown in Figure 3.11 for $\beta = 1$, in Figure 3.12 for $\beta = 0.5$, hence for twice as many receive than transmit antennas, and in Figure 3.13 for five times as many, hence for $\beta = 0.2$.
Here again we can observe similar results as for the bandmatrix correlation model (see Section 3.2.4). The shape of the resulting pdf remains very similar to the one of the Marčenko-Pastur distribution, but we can notice the higher influence of the correlation with this model. For a higher correlation coefficient, the value of the largest possible eigenvalue increases, but but we get also more smaller eigenvalues, which fits to the result, that the channel is less well conditioned. For finite dimensional matrices $\mathbf{H}$ that

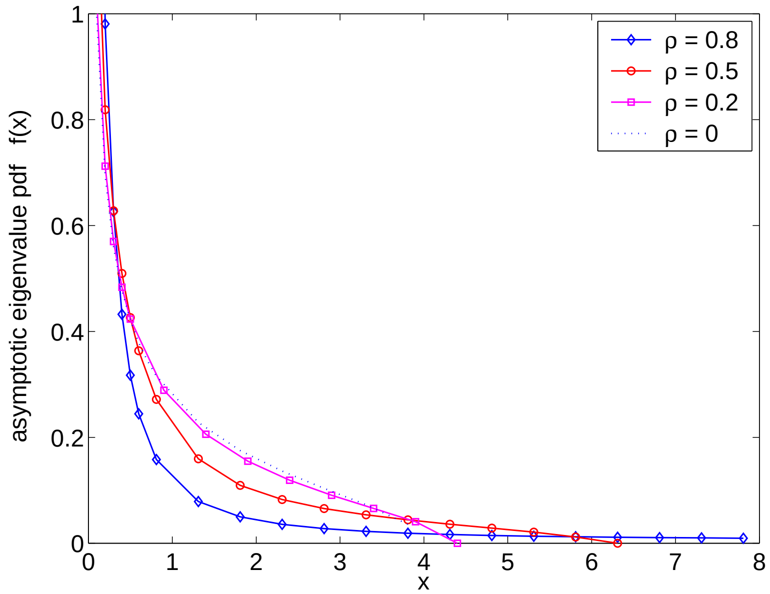



Figure 3.11: Asymptotic empirical eigenvalue pdfs of simplified Kronecker Model matrices $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$, with exponential correlation matrix, with fixed $\beta = 1$ and correlation parameter $\rho = 0.8, 0.5, 0.2, 0$.

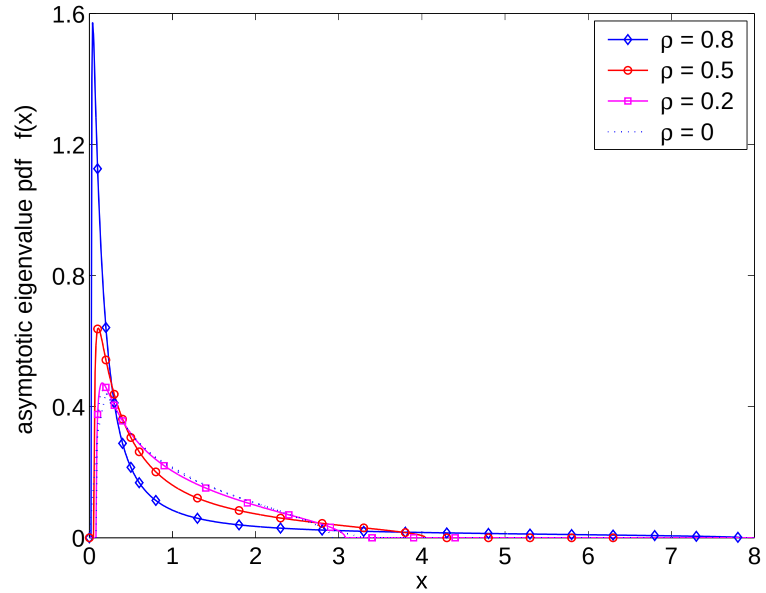



Figure 3.12: Asymptotic empirical eigenvalue pdfs of simplified Kronecker Model matrices $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$, with exponential correlation matrix, with fixed $\beta = 0.5$ and correlation parameter $\rho = 0.8, 0.5, 0.2, 0$.

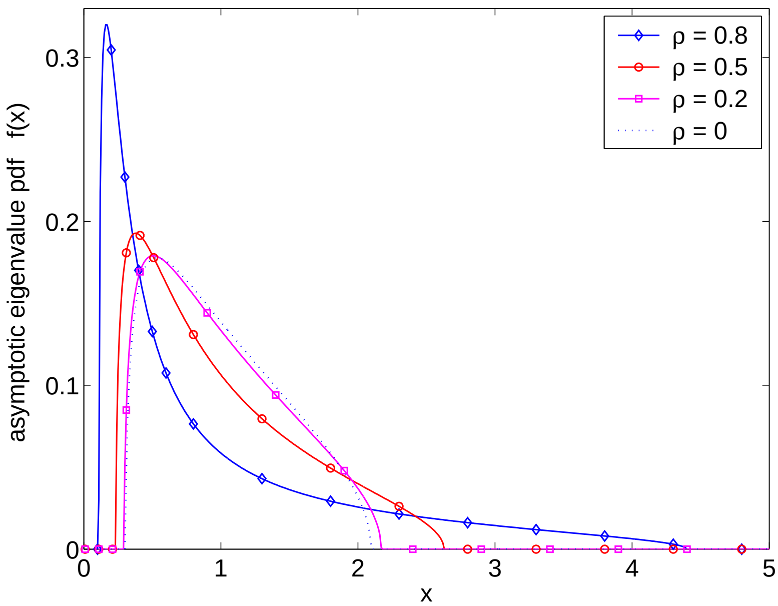



Figure 3.13: Asymptotic empirical eigenvalue pdfs of simplified Kronecker Model matrices $\frac{1}{N_R}\mathbf{H}\mathbf{H}^H$, with exponential correlation matrix, with fixed $\beta = 0.2$ and correlation parameter $\rho = 0.8, 0.5, 0.2, 0$.

fullfill the conditions of our simplified Kronecker model, with exponential correlation matrix, the eigenvalue pdf is illustrated for $\beta = 1$ and $\rho = 0.4$ in Figure 3.6 and compared to the asymptotic pdf we have computed using free multiplicative convolution. Here again, we can observe that the asymptotic eigenvalue pdf is a good approximation for finite dimensional systems.

### Calculating the asymptotic capacity

With relation (3.14) and the asymptotic eigenvalue pdf obtained previously, we are now able to compute the asymptotic capacity per receive antenna. The integration is done numerically, the resulting asymptotic capacity is illustrated for different values of the correlation coefficient $\rho$ in Figure 3.14 for $\beta = 1$ and in Figure 3.15 for $\beta = 0.2$, hence for ten times as many receive as transmit antennas. As expected, the capacity per receive antenna decreases with the correlation coefficient. The capacity per receive antennas for finite dimensional model matrices has been computed for some samples of the channel matrix and shown in Figure 3.9. Again, we can observe that the asymptotic result is a good approximation. In Figure 3.17, the asymptotic capacity per receive antenna versus $\beta$ is shown for high SNR (30dB) and for a fixed correlation coefficient $\rho = 0.5$. The same result as for the canonical model. The capacity is for a high SNR and for a fixed correlation coefficient proportional to the minimum number of transmit or receive antennas.

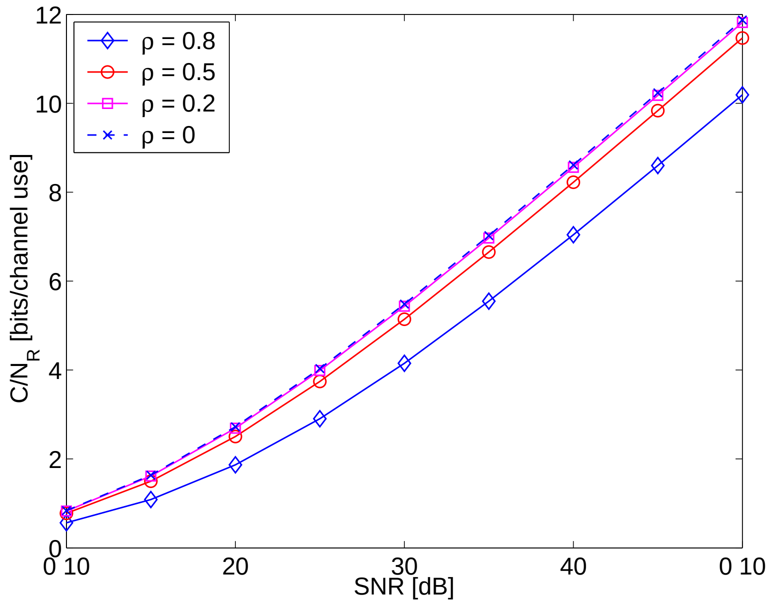



Figure 3.14: Asymptotic capacity for the simplified Kronecker model with exponential correlation matrix, for fixed $\beta = 1$, $\rho = 0.8, 0.5, 0.2, 0$.

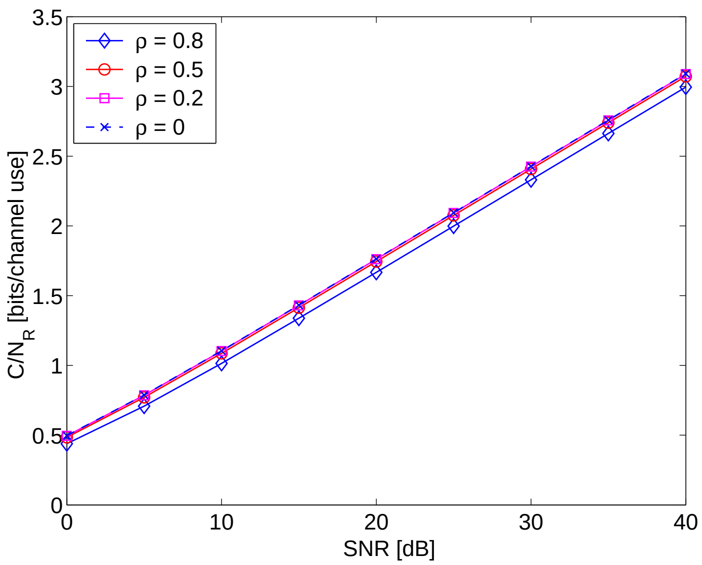



Figure 3.15: Asymptotic capacity for the simplified Kronecker model with exponential correlation matrix, for fixed $\beta = 0.2$, $\rho = 0.8, 0.5, 0.2, 0$.

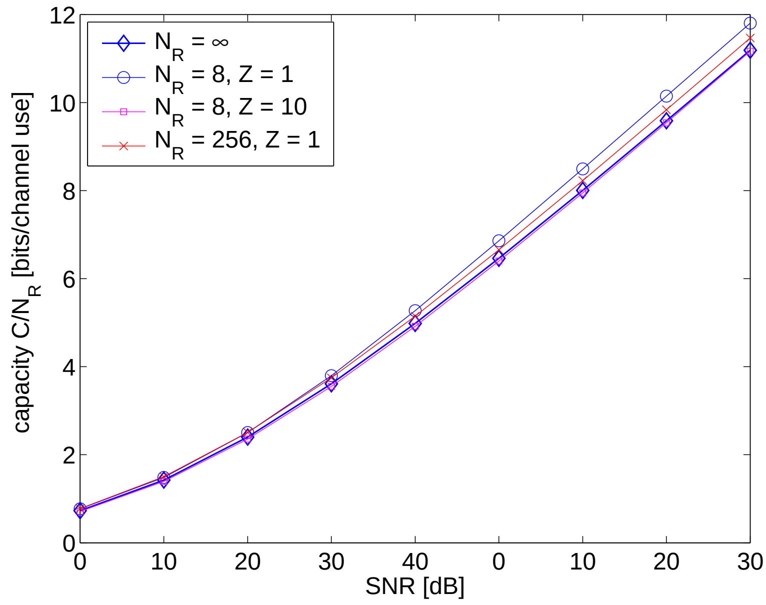



Figure 3.16: Capacity for finite dimensional matrices that fullfill the simplified Kronecker model with exponential correlation matrix, for fixed $\beta = 1$, $\rho = 0.5$ for one $8 \times 8$, ten $8 \times 8$ and one $256 \times 256$ matrices.

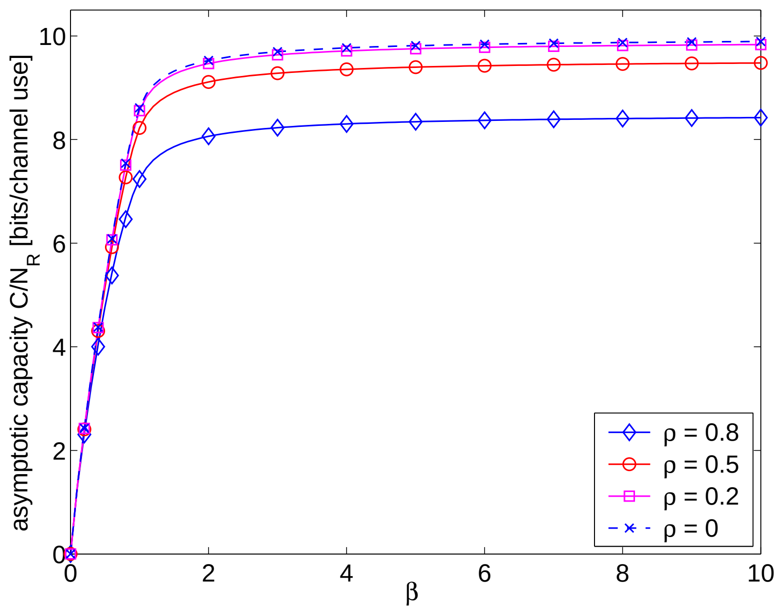



Figure 3.17: Asymptotic capacity versus $\beta$ for high SNR = 30dB and $\rho = 0.5$ for the simplified Kronecker model with exponential correlation matrix.

The results for the asymptotic eigenvalue pdf and the asymptotic capacity for the exponential correlation model are similar to the ones obtained in Section 3.2.4 for the bandmatrix correlation model. Nevertheless the correlation has more influence on the results. The results confirm our assumption, that the exponential matrix model is a more realistic than the bandmatrix model.

# Chapter 4

# Appendix

## 4.1 Fundamentals of classical probability theory

### 4.1.1 Moments and cumulants

About a classical random variable (CRV), we might be interested in many different kinds of information (e.g. the expectation, the variance, the probability of finding this CRV in a certain interval). This infromation can be obtained using either the moment sequence, the distribution or the characteristic function. Any of these descriptions of the CRV contains all the statistical information about the CRV; but one representation might be more practical than another in order to get some specific information about the CRV. For example, if a probability value of the CRV is required, it is the most useful to use the distribution; instead if we are interested in the expectation, the moment sequence will be more practical. Considering two CRV, i.e. $X_1$ and $X_2$, if they are statistically independent, then with the statistical information of $X_1$ and $X_2$, we can obtain the statistical information of $X_1 + X_2$. The relation between $X_1$, $X_2$ and $X_1 + X_2$ can be described by means of the moment sequence, of the distributions or the pdfs, and also by means of the characteristic function. How this is done, is shown in the following. Let

- $X_1$ and $X_2$ be two independent CRVs,
- $F_1(x)$ and $F_2(x)$ their distributions,
- $f_1(x)$ and $f_2(x)$ their pdfs,
- $m_n(X_1)$ and $m_n(X_2)$ their moment sequences, and
- $\Psi_{X_1}(\omega)$ and $\Psi_{X_2}(\omega)$ their characteristic functions.

Let

- $X = X_1 + X_2$, with

- $F(x)$ its distribution,
- $f(x)$ its pdf,
- $m_n(X)$ its moment sequence,
- $\Psi_X(\omega)$ its characteristic function.

The moment sequence $m_n(X)$ can be obtained over the binomial formula [13]

$$\begin{aligned} m_n(X) = \mathbb{E}\{X^n\} = \mathbb{E}\{(X_1 + X_2)^n\} = \mathbb{E}\left\{\sum_{i=0}^{n} \binom{n}{i} X_1^i X_2^{n-i}\right\} \\ = \sum_{i=0}^{n} \binom{n}{i} \mathbb{E}\{X_1^i\} E\{X_2^{n-i}\} = \sum_{i=0}^{n} \binom{n}{i} m_i(X_1) m_{n-i}(X_2)\,. \end{aligned} \quad (4.1)$$

The characteristic function of the sum of two CRVs can be found very easily, it is the product of the individual characteristic functions, i.e.

$$\Psi_X(\omega) = \Psi_{X_1+X_2}(\omega) = \mathbb{E}\{e^{i\omega(X_1+X_2)}\} = \mathbb{E}\{e^{i\omega X_1}\}\mathbb{E}\{e^{i\omega X_2}\} = \Psi_{X_1}(\omega)\Psi_{X_2}(\omega)\,. \quad (4.2)$$

Since the characteristic function is the also the Fourier transform of the pdf of the RV, the pdf of the sum of two CRVs can be obtained by means of the convolution [13], i.e.

$$f_X = f_{X_1+X_2} = f_{X_1} * f_{X_2}. \quad (4.3)$$

The convolution operation can be linearized by applying the logarithmic function on it. Thus, the logarithmic characteristic function of the sum of two CRVs is the sum of the individual logarithmic characteristic functions,

$$\ln \Psi_X(\omega) = \ln \Psi_{X_1+X_2}(\omega) = \ln \Psi_{X_1}(\omega) + \ln \Psi_{X_2}(\omega). \quad (4.4)$$

And finally a relation between the moment sequence and the logarithmic characteristic function can be easily obtained using the Taylor series expansion given by

$$\Psi_X(\omega) = \sum_{n=0}^{\infty} \frac{(i\omega)^n}{n!} m_n(X).$$

The logarithmic characteristic function can also be represented as a Taylor function. Its coefficients are interesting for us since, as $\log \Psi_X$, they linearize the convolution as well. These coefficients are called cumulants.

**Definition 4.1.1:** *The* cumulant sequence $s_n(X)$ *of a RV* $X$ *satisfies:*

$$\log \Psi_X(\omega) = \sum_{n=1}^{\infty} \frac{(i\omega)^n}{n!} s_n(X). \quad (4.5)$$

Example 4.1.1: For a Gaussian CRV $X$ with zero mean and variance $\sigma^2$, the characteristic function and the logarithmic characteristic function are given by

$$\Psi_X(\omega) = e^{-\frac{\sigma^2\omega^2}{2}} \quad \Rightarrow \quad \log \Psi_X(\omega) = -\frac{\sigma^2\omega^2}{2}.$$

Hence, the cumulant sequence of $X$ is given by

$$s_2(X) = \sigma^2, \quad s_n(X) = 0, \text{for } n \neq 2$$

■

It can be easily seen that the cumulant sequence of the sum $X = X_1 + X_2$ of two independent CRVs $X_1$ and $X_2$ is the sum of the individual cumulant sequences, i.e.

$$s_n(X) = s_n(X_1 + X_2) = s_n(X_1) + s_n(X_2). \tag{4.6}$$

Example 4.1.2: Assume two Gaussian RVs $X_1, X_2$, with zero-mean and variances $\sigma_1^2, \sigma_2^2$. The sum of these RVs $X = X_1 + X_2$ is again a Gaussian RV with zero mean and variance $\sigma_1^2 + \sigma_2^2$ [13]. This is equivalent to summing up the cumulant sequences.

■

The cumulant sequence contains again all the statistical information of the RV. It can be obtained over the distribution, the characteristic function and it is related to the moment sequence by a closed formula (see also (4.7)). Before we show this formula, it is nevertheless necessary to introduce some combinatoric concepts.

**Definition 4.1.2:** *A* partition *of a* set $\{1 : n\} := \{1, 2, \ldots, n\}$ *is a sequence of nonempty, pairwise disjoint blocks* $V_1, V_2, \ldots, V_s$ *such that*

$$\bigcup_{i=1}^{s} V_i = \{1 : n\}$$

Example 4.1.3: One possible partition of the set $\{1 : 7\}$ is $V_1 = \{1, 5, 6\}, V_2 = \{2, 3\}, V_3 = \{4, 7\}$. ■

The definition of a partition enables us to write the following theorem.

**Theorem 4.1.3:** *The moment sequence* $m_n(X)$ *of a CRV* $X$ *is related to its cumulant sequence* $s_n(X)$ *over the formula [10]:*

$$m_n(X) = \sum_i \prod_j s_{|V_j^{(i)}|}(X). \tag{4.7}$$

*The summation is over all partitions* $\nu^{(i)} = \{V_1^{(i)}, V_2^{(i)}, \ldots, V_s^{(i)}\}$ *of the set* $\{1 : n\}$. *Here* $|V_j^{(i)}|$ *denotes the number of elements of* $V_j^{(i)}$.

Example 4.1.4:

- For n = 1 , there is just one partition with one element $\nu^{(1)} = V_1^{(1)} = 1$

$$m_1 = s_1.$$

- For n = 2, there are two different possible partitions of $\{1 : 2\}$,

$$\begin{aligned} \nu^{(1)} =& \{V_1^{(1)}\}, & \text{with} \quad & V_1^{(1)} = \{1, 2\}, \\ \nu^{(2)} =& \{V_1^{(2)}, V_2^{(2)}\}, & \text{with} \quad & V_1^{(2)} = \{1\}, V_2^{(2)} = \{2\}. \end{aligned}$$

Hence $$m_2 = s_{|V_1^{(1)}|} + s_{|V_1^{(2)}|} s_{|V_2^{(2)}|},$$

$$m_2 = s_2 + s_1^2.$$

- For n=3, there are five different possible partitions of $\{1 : 3\}$:

$$\begin{aligned} \nu^{(1)} =& \{V_1^{(1)}\} & \text{with} \quad & V_1^{(1)} = \{1, 2, 3\}, \\ \nu^{(2)} =& \{V_1^{(2)}, V_2^{(2)}\} & \text{with} \quad & V_1^{(2)} = \{1\}, V_2^{(2)} = \{2, 3\}, \\ \nu^{(3)} =& \{V_1^{(3)}, V_2^{(3)}\} & \text{with} \quad & V_1^{(3)} = \{2\}, V_2^{(3)} = \{1, 3\}, \\ \nu^{(4)} =& \{V_1^{(4)}, V_2^{(4)}\} & \text{with} \quad & V_1^{(4)} = \{3\}, V_2^{(4)} = \{1, 2\}, \\ \nu^{(5)} =& \{V_1^{(5)}, V_2^{(5)}, V_3^{(5)}\} & \text{with} \quad & V_1^{(5)} = \{1\}, V_2^{(5)} = \{2\}, V_3^{(5)} = \{3\}. \end{aligned}$$

$$\begin{aligned} m_3 &= s_{|V_1^{(1)}|} + s_{|V_1^{(2)}|} s_{|V_2^{(2)}|} + s_{|V_1^{(3)}|} s_{|V_2^{(3)}|} + s_{|V_1^{(4)}|} s_{|V_2^{(4)}|} + s_{|V_1^{(5)}|} s_{|V_2^{(5)}|} s_{|V_3^{(5)}|} \\ m_3 &= s_3 + 3 s_2 s_1 + s_1^3 \end{aligned}$$

■

Example 4.1.5: For a Gaussian RV with zero mean and variance $\sigma^2$, we can verify Equation (4.7):

$$m_1 = s_1 = 0, \quad m_2 = s_2 = \sigma^2, \quad m_3 = s_3 = 0, \ldots$$

■

## 4.2 Fundamentals of free probability theory

### 4.2.1 Operators seen as NCRVs

FPT enables us to consider a very wide class of objects NCRVs. Once they fullfill the primary conditions (see Chapter 1) we can apply the theorems and results of FPT on them. To become a NCRV, an object has just to be embedded in a NCPS, hence, it has to be an element of a unital algebra endowed with a state (see Definition 1.1.1). The idea of seeing an operator as a NCRV arised from quantum theory [1]. The operators considered there are bounded operators on a Hilbert space $\mathcal{H}$, that is endowed with the scalar product $\langle \cdot, \cdot \rangle$. The sum and the product of bounded operators are again bounded operators. Thus, we can consider the algebra $B(\mathcal{H})$ of bounded operators acting on $\mathcal{H}$. In order that an operator $\boldsymbol{A} : \mathcal{H} \to \mathcal{H}$ can be regarded as a NCRV, it needs to be an element of a NCPS. In this case, the NCPS will be $(B(\mathcal{H}), \varphi)$ with the unital algebra $B(\mathcal{H})$ and the functional $\varphi : B(\mathcal{H}) \to \mathbb{C}$, defined by

$$\varphi(\boldsymbol{A}) = \langle \boldsymbol{A}\xi, \xi \rangle,$$

with $\xi$ denoting a unit vector in $\mathcal{H}$. The functional $\varphi$ verifies the condition of the state, since

$$\varphi(\mathbf{1}) = \langle \xi, \xi \rangle = 1,$$

where $\mathbf{1}$ denotes the identity operator. Since the operator is now defined as a NCRV, its distribution functional $\mu_{\boldsymbol{A}} : \mathbb{C}\langle X \rangle \to \mathbb{C}$ can be given:

$$\mu_{\boldsymbol{A}}(x^k) = \varphi(\boldsymbol{A}^k) = \langle \boldsymbol{A}^k \xi, \xi \rangle.$$

When the operator is selfadjoint, then the spectral theorem [1, 18] says that:

$$\langle \boldsymbol{A}^k \xi, \eta \rangle = \int x^k d\langle E_x \xi, \eta \rangle, \quad \forall \xi, \eta \in \mathcal{H},$$

where $E_x$ denotes the spectral decomposition of $\boldsymbol{A}$. Hence

$$\varphi(\boldsymbol{A}^k) = \langle \boldsymbol{A}^k \xi, \xi \rangle = \int x^k d\langle E_x \xi, \xi \rangle,$$

which implies that we can define a distribution measure $F(x)$ of the operator $\boldsymbol{A}$, i.e.

$$F(x) = \langle E_x \xi, \xi \rangle.$$

### 4.2.2 Fock spaces

Operators on any Hilbert space can be seen NCRVs. A special case of a Hilbert space is the Fock space, which will be introduced in the following. On this space two operators will be defined, that can describe any NCRV. They will give interesting results and enable to prove certain theorems.

**Definition 4.2.1:** *Let $\mathcal{H}$ be a Hilbert space and let $\mathcal{H}^{\otimes n}$ be the n-fold tensor product $\mathcal{H} \otimes \mathcal{H} \otimes \ldots \otimes \mathcal{H}$. Then*

$$\mathcal{F}(\mathcal{H}) := \bigoplus_{n=0}^{\infty} \mathcal{H}^{\otimes n} \tag{4.8}$$

*is called the* full Fock space *with $\mathcal{H}^{\otimes 0} = \mathbb{C}\Phi$, where the unit vector $\Phi$ is referred to as the* vacuum vector.

On the Fock space a scalar product is defined. Since one element $\eta$ of $\mathcal{F}(\mathcal{H})$ is the sum of elements of the powers $\mathcal{H}^{\otimes n}$, $\eta$ can be written as

$$\eta = \sum_{n=0}^{\infty} \eta^{(n)} \quad \text{with} \quad \eta^{(n)} \in \mathcal{H}^{\otimes n}.$$

The element $\eta^{(n)}$ of the tensor product $\mathcal{H}^{\otimes n} = \mathcal{H} \otimes \mathcal{H} \otimes \ldots \otimes \mathcal{H}$ can be written as

$$\eta^{(n)} = \sum_i \eta_{i,1} \otimes \cdots \otimes \eta_{i,n} \quad \text{with} \quad \eta_{i,j} \in \mathcal{H},$$

Since two elements $\eta^{(n)}$ and $\eta^{(m)}$ are orthogonal for $m \neq n$, we just need the scalar product for two vectors $\eta_n$ and $\xi_n$ out of the same space $\mathcal{H}^{\otimes n}$:

$$\langle \eta_1 \otimes \cdots \otimes \eta_n, \xi_1 \otimes \cdots \otimes \xi_n \rangle = \langle \eta_1, \xi_1 \rangle \cdots \langle \eta_n, \xi_n \rangle. \tag{4.9}$$

The generalized form of the scalar product for any elements $\eta, \xi \in \mathcal{F}(\mathcal{H})$ can easily be derived due to linearity. Now that we have introduced the Fock space and its scalar product two main operators can be defined.

**Definition 4.2.2:** *For each $h \in \mathcal{H}$, we can define a bounded operator $l(h)$ on $\mathcal{F}(\mathcal{H})$*

$$l(h)\eta = \begin{cases} h, & \text{if } \eta = \Phi, \\ h \otimes \eta, & \text{if } \langle \eta, \Phi \rangle = 0, \end{cases} \tag{4.10}$$

*which is called the* left creation operator.

**Definition 4.2.3:** *Its adjoint $l(h)^*$ is called the* annihilation operator *given by*

$$l(h)^* \eta_1 \otimes \eta_2 \otimes \ldots \otimes \eta_k = \langle \eta_1, h \rangle \eta_2 \otimes \ldots \otimes \eta_k, \qquad l(h)^* \Phi = 0. \tag{4.11}$$

The fact that $l(h)^*$ is the adjoint operator of $l(h)$ can be confirmed by showing that

$$\langle l(h)\eta, \xi \rangle = \langle \eta, l(h)^* \xi \rangle, \quad \forall \xi, \eta \in \mathcal{H}.$$

In order to compute the scalar product, $l(h)\eta$ and $\xi$ need to be from the same space $\mathcal{H}^{\otimes n}$. This is true for example for $\eta = \eta_1 \otimes \cdots \otimes \eta_{n-1}$ and $\xi_1 \otimes \cdots \otimes \xi_n$.

$$\begin{aligned}\langle l(h)\eta_1 \otimes \cdots \otimes \eta_{n-1}, \xi_1 \otimes \cdots \otimes \xi_n \rangle &= \langle h \otimes \eta_1 \otimes \cdots \otimes \eta_{n-1}, \xi_1 \otimes \cdots \otimes \xi_n \rangle \\ &= \langle h, \xi_1 \rangle \langle \eta_1, \xi_2 \rangle \cdots \langle \eta_{n-1}, \xi_n \rangle \\ &= \langle \eta_1 \otimes \cdots \otimes \eta_{n-1}, h\xi_1 \otimes \xi_2 \otimes \cdots \otimes \xi_n \rangle \\ &= \langle \eta_1 \otimes \cdots \otimes \eta_{n-1}, l(h)^* \xi_1 \otimes \cdots \otimes \xi_n \rangle\end{aligned}$$

There are interesting results about these operators, which will be very useful in the following.
Consider the operators as NCRVs, elements of the NCPS $(B(\mathcal{F}(\mathcal{H})), \varphi)$, $B(\mathcal{F}(\mathcal{H}))$ being the algebra of bounded operators on the Fock space and the vacuum state $\varphi(\cdot) = \langle \cdot \Phi, \Phi \rangle$.

**THEOREM 4.2.4:** *The NCRV $l(h)^* + l(h)$, has the semicircle law as its distribution, i.e.*

$$w_r(x) = \begin{cases} \frac{2}{\pi r^2}\sqrt{r^2 - x^2}, & \text{if } -r \leq x \leq -r, \\ 0, & \text{otherwise .} \end{cases} \qquad r = 2\|h\|.$$

The proof, that uses combinatorics can be found in [10]. Recall that for $r = 1$, we obtain the same function as the semicircle law mentioned in (2.10).

**THEOREM 4.2.5:** *If $h_1$,$h_2 \in \mathcal{H}$ and if $l(h_i)$ denote the corresponding left creation operators on $\mathcal{F}(\mathcal{H})$, $\{l(h_1), l(h_1)^*\}$ and $\{l(h_2), l(h_2)^*\}$ are free with respect to the vacuum state if and only if $\langle h_1, h_2 \rangle = 0$.*

Consider $l = l(h)$ as a creation operator with $\|h\| = 1$, and consider the sequences of the form

$$l^* + \sum_{k=0}^{\infty} \alpha_{i+1} l^k.$$

These series converge in norm and are bounded operators on the full Fock space. They can be regarded as NCRVs, elements of the NCPS $(\mathcal{E}, \varphi)$, where $\mathcal{E}$ denotes the corresponding unital algebra with the state $\varphi(\cdot) = \langle \cdot \Phi, \Phi \rangle$.
Interesting to note is that for every distribution measure with compact support a sequence $l^* + \sum_{k=0}^{\infty} \alpha_{i+1} l^k$ exists, that has this distribution. This sequence is called the canonical NCRV associated to $F(x)$.

**Definition 4.2.6:** *A* canonical non-commutative random variable *of the distribution $F(x)$ is the sequence*

$$l^* + \sum_{k=0}^{\infty} \alpha_{i+1} l^k$$

*with $(\alpha_i)_{i\in\mathbb{N}}$ such that*

$$\int x^k dF(x) = \langle \left( l^* + \sum_{k=0}^{\infty} \alpha_{i+1} l^k \right)^n \Phi, \Phi \rangle.$$

*The sequence $(\alpha_i)_{i\in\mathbb{N}}$ is called the* free cumulant sequence*. For every NCRV A, whose distribution measure is $F(x)$, $(\alpha)_{i\in\mathbb{N}}$ is called the free cumulant sequence associated to A.*

Hence, any NCRV has a representation in terms of a canonical NCRV.

Example 4.2.1: From Theorem 4.2.4, we know, that the canonical NCRV associated to the semicircle law $w_r$ is $l^* + l$. The free cumulant sequence is then

$$\alpha_1 = 1, \quad \alpha_i = 0 \quad \text{for} \quad i \neq 1.$$

The name free cumulant sequence has been chosen in analogy to the cumulant concept in CPT. As the cumulant sequence does it in CPT (see Section 4.1.1), the free cumulant sequence linearizes the additive convolution operation in FPT. The free cumulant sequence is used to define the R-transform (see Definition 1.5.1):

$$R(z) = \sum_{i=0}^{\infty} \alpha_i z^i.$$

It is rather difficult to obtain the free cumulant sequence, so the R-transform is usually obtained by means of the Cauchy-transform (see Section 1.5).

Example 4.2.2: The R-transform of the semicircle law has been computed in Example 1.5.5:

$$R(z) = z.$$

This verifies the result of Example 4.2.1 ■

The similarity between the free cumulant sequence and the cumulant sequence known from the CPT lies also in the relation between the cumulant and the moment sequence. It is again a relation that involves combinatorics, but here non-crossing partitions have to be used [10].

**DEFINITION 4.2.7:** *A partition* $\nu = \{V_1, V_2, \dots, V_s\}$ *of the set* $[n]$ *is called a* non-crossing partition *[10] if for*

$$V_i = \{v_1, v_2, \dots, v_p\} \quad \textit{and} \quad V_j = \{w_1, w_2, \dots, w_q\}$$

*we have*

$$w_m < v_1 < w_{m+1} \quad \text{if and only if} \quad w_m < v_p < w_{m+1}, \quad (m = 1, 2, \dots, q-1)$$

This definition enables us to give the relation between the moment sequence of a NCRV and its free cumulant sequence.

**THEOREM 4.2.8:** *For a NCRV A, the moment sequence and the free cumulant sequence are related to each other over*

$$m_n(A) = \sum_{\nu \in NC(n)} \prod_i \alpha_{|V_i|}(A). \tag{4.12}$$

In the FPT many of the results have been proved with combinatorics (for example Theorem 1.4.2, which has been proved using a interesting results about the moments of the semicircle and the number of non-crossing partitions of a set). The proofs of the results can be found in [10].

**THEOREM 4.2.9:** *The number of non-crossing partitions of a set* $\{1 : n\}$ *equals the nth moment of the standard semicircle law.*

# Bibliography


[1] P. Biane. Calcul stochastique non-commutatif. In P. Bernard, editor, *Lectures on Probability Theory an Statistics, Ecole d'Eté de Probabilités de Saint Flour XXIII*, pages 279–349. Springer, 1993.

[2] E. Biglieri and G. Taricco. Exact pairwise error probability of space-time codes. *IEEEIT*, 48:510–513, Feb. 2002.

[3] Chen-Nee Chuah, D.N.C. Tse, J.M. Kahn, and R.A. Valenzuela. Capacity Scaling in MIMO Wireless Systems Under Correlated Fading. *IEEE Transactions on Information Theory*, 48(3):637–650, Mar. 2002.

[4] M. Debbah, W. Hachem, P. Loubaton, and M. de Courville. Analysis of certain large isometric random precoded systems. *IEEE Trans. Inf. Theory*, 49(5):1293–1311, May 2003.

[5] E. Biglieri, G. Taricco, and A. Tulino. Performance of space-time codes for a large number of antennas. *IEEEIT*, 48(9):1794–1803, July 2002.

[6] A. Edelman. *Eigenvalues and Condition Numbers of Random Matrices.* PhD thesis, MIT, MA, 1989.

[7] G. J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Labs Tech. J.*, 1(2):41–59, 1996.

[8] B. Friedrichs. *Kanalcodierung.* Springer, 1996.

[9] R. M. Gray. On the asymptotic eigenvalue distribution of toeplitz matrices. *IEEEIT*, 18(6):725–730, Nov. 1972.

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

[11] K. Schmidt and G. Trenkler. *Moderne Matrix-Algebra.* Springer, Berlin-Heidelberg, 1998.

[12] M. Koecher. *Lineare Algebra und Analytische Geometrie.* Springer, Berlin-Heidelberg, 1992.

[13] M. Loève. *Probability Theory.* Van Nostrand, Princeton (NJ), 3rd edition, 1962.

[14] V. Marcenko and L. Pastur. Distributions of eigenvalues for some sets of random matrices. *Math. USSR-Sbornik 1*, pages 457–483, 1967.

[15] C. Martin and B. Ottersten. Asymptotic eigenvalue distributions and capacity for mimo channels under correlated fading. *IEEEWC*, 3(4):1350–1359, July 2004.

[16] X. Mestre and J. R. Fonollosa. Capacity of mimo channels: Asymptotic evaluation under correlated fading. *IEEEJSAC*, 21(5):829–838, June 2003.

[17] R. R. Müller. A random matrix model of communication via antenna arrays. *IEEEIT*, 48(9):2495–2506, Sept. 2002.

[18] W. Rudin. *Functional analysis.* Tata McGraw-Hill, New Delhi, 1988.

[19] J. W. Silverstein. On the empirical distribution of eigenvalues of a class of large dimenstional random matrices. *JMVA*, 54:175–192, 1995.

[20] V. Tarokh, N. Seshadri, and A. R. Calderbank. Space-time codes for high data rate wireless communications: Performance criterion and code construction. *IEEE Trans. Inf. Theory*, 44(2):744–765, March 1998.

[21] I. E. Telatar. Capacity of multi-antenna gaussian channels. Technical Report BL0112170-950615-07TM, AT&T Bell Laboratories, June 1995.

[22] S. Verdú and S.Shamai. Spectral efficiency of cdma with random spreading. *IEEE Trans. Inf. Theory*, 45(2):622–640, 1999.

[23] D. Voiculescu. Limit laws for random matrices and free products. *Invent. Math.*, 104:201–220, July 1991.

[24] D. Voiculescu. Lectures on free probability theory. In P. Bernard, editor, *Lectures on Probability Theory an Statistics, Ecole d'Eté de Probabilités de Saint Flour XXVIII*, pages 279–349. Springer, 1998.

[25] L. Zheng and D. Tse. Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels. *IEEE Trans. Inf. Theory*, 49(5):1073–1096, May 2003.