1-BOUNDED ENTROPY AND APPLICATIONS I

SRIVATSAV KUNNAWALKAM ELAYAVALLI

Abstract

This is a series of talks on Ben Hayes' work on the above subject and his paper. In the first lecture we build up to define Voiculescu's microstates free entropy. We also define free entropy dimension in the contexts of Voiculescu and Kenley Jung. Caution, this part is a bit long considering the background that's necessary.

Voiculescu invented free probability to study the isomorphism question of free group factors. The free product construction gives rise to the consideration of "free independence", and motivates us to consider an analogous probability theory to the classical Kolmogorov probability theory which is based on tensor products and tensor independence. In this first part, we try to retrace Voiculescu's steps in building such a theory and defining important analogues of entropy and minkowski dimension among others.

1 Free Probability

Free probability, as one would expect, is indeed a vast generalization of classical probability. There are several ways of motivating free probability, but in this exposition, we introduce free probability rather immediately, and then in the next section we'll talk about random matrices which are really the true motivations of free probability.

The material in this section is based on Dimitri Shlyakhtenko's notes on free probability [Shl04], and Djalil Chafai's survey paper on entropy titled "From Boltzmann to random matrices and beyond" [Cha15].

Definition 1.1 (Probability spaces)**.** Recall that a classical probability space is a measure space (X, B, μ). Here B is a sigma algebra, μ is a probability measure, i.e, $\mu(X) = 1$. The set X should be thought of as a set of events, and for $Y \in B$, the measure $\mu(Y)$ is the probability of an event occurring in the set Y.

Definition 1.2 (Classical Random Variables)**.** An alternative pount of view on probability theory involves considering classical random variables, i.e, measurable functions $f : X \to \mathbb{C}$. One can think of a classical random variable as a measurement, which assigns to each event $x \in X$ a value $f(x)$. The probability of the value of f lying in a set $A \subset \mathbb{C}$ is exactly $\mu(f^{-1}(A)) = (f_*\mu)(A)$.

Definition 1.3 (The expectation E)**.** Let us say that $f \in L^\infty(X, \mu)$ is an essentially bounded classical random variable. Then the integral

$$E(f) = \int f(x) d\mu(x)$$

has the meaning of the expected value of f. The linear functional given by this: $E : L^\infty(X, \mu) \to \mathbb{C}$ is called an expectation. E satisfies $E(1) = 1$ and positivity. The reason for the term expectation is obviously because the interpretation matches with the usual meaning of the word "expectation".

We now discuss the notion of non-commutative probability:

1.0.1 Non Commutative Probability Spaces

Definition 1.4. An algebraic non-commutative probability space is a pair (A, ϕ) consisting of a unital algebra A and a linear functional $\phi : A \to \mathbb{C}$ such that $\phi(1) = 1$.

Remark here that A acts as our space of random variables, and ϕ acts as our expectation functional. One should note that knowing about a probability space X is the same as knowing about the space of classical random variables, $L^\infty(X)$, which is the kind of approach we adopt here. Also observe that $L^\infty(X)$ is all the abelian von Neumann algebras, so in our constructions, classical probability will drop out in the abelian case.

In this exposition, we shall not consider very general algebras A, but stick to A being a finite von Neumann algebra, and ϕ being the accompanying trace-state.

Definition 1.5 (Laws of Random Variables)**.** Recall that we assigned to a classical random variable f it's law μ_f. In the non commutative case, suppose $a \in A$ is a self adjoint random variable, note from the spectral theorem, we have a spectral measure ν_a on $\mathbb{R}$ so that

$$\phi(a) = \int t d\nu_a(t)$$

This is going to be our Law of a, or distribution of a. The distribution function F_a is given by $F_a(x) = \nu_a(-\infty, x)$.

Definition 1.6 (Moments)**.** We define the moments of a self adjoint random variable a to be as follows

$$\mathbb{E}(a^p) = \phi(a^p) = \int x^p d\mu_a(x)$$

For a family F of variables $(a_1, \ldots, a_n \in A)$, we say that an expression of the form $\phi(a_{i_1} \ldots a_{i_p})$ is the $i_1, \ldots, i_p$-th moment of the family f. The collection of all moments can be thought of as a linear functional μ_F defined on the algebra of polynomials in n indeterminates $t_1, \ldots, t_n$ by

$$\mu_F(p) = \phi(p(a_1, \ldots, a_n))$$

1.0.2 Independence: classical and free

Definition 1.7 (Classical independence)**.** Two random variables f and g in $L^\infty(X, \mu)$ are called independent if $E(f^n g^m) = E(f^n)E(g^m)$ for all $n, m \geq 0$. A more useful definition (for the sake of motivating the condition for free independence) is as follows: f and g are independent if $E(FG) = 0$ whenever $E(F) = E(G) = 0$ and F is in the algebra $W^*(f)$ generated by f, while $G \in W^*(g)$.

Definition 1.8 (Free Independence)**.** Let $F_1, F_2 \subset (A, \phi)$ be two families of non-commutative random variables. We say that F_1 and F_2 are freely independent if

$$\phi(a_1 \dots a_n) = 0$$

whenever $a_j \in Alg(1, F_{i(j)})$, where $i(1) \neq i(2)$, $i(2) \neq i(e)$ and so on; and $\phi(a_1) = \phi(a_2) = \dots = 0$. There are several equivalent definitions of free independence, but this one is the simplest.

The natural product structure associated with this independence is called the free product. As a product on von Neumann algebras M and N, it is denoted $M * N$, and obeys the same kind of independence result in the non-commutative analogue.

Proposition 1.9. *Let M, N be finite von Neumann algebras, and let $M * N$ denote their free product with respect to their traces. Let π_M and π_N be the natural embeddings of M and N into the free product. Then, for any subalgebras $A \subseteq M$ and $B \subseteq N$, we have $\pi_M(A)$ and $\pi_N(B)$ are freely independent in $M * N$.*

There are some important properties of this free product on von Neumann algebras. For instance, it is compatible with the free product on groups:

Proposition 1.10. *Let Γ_1 and Γ_2 be locally compact groups. Then*

$$L(\Gamma_1) * L(\Gamma_2) \cong L(\Gamma_1 * \Gamma_2)$$

At first sight it may be difficult to come up with examples of freely independent algebras that do not arise from groups, so we provide a very important example here:

1.0.3 Free Fock Spaces

Let $\mathcal{H}$ be a Hilbert space, and let Ω be a vector (called the vacuum vector). We define

$$F(H) := \mathbb{C}\Omega \oplus \mathcal{H} \oplus (\mathcal{H} \otimes \mathcal{H}) \oplus \dots$$

The inner product ϕ will be the state with respect to the vacuum vector. For $h \in \mathcal{H}$ consider the left creation operator

$$l(h) : F(\mathcal{H}) \to F(\mathcal{H})$$

given by

$$l(h)h_1 \otimes \dots \otimes h_n = h \otimes h_1 \otimes \dots \otimes h_n$$

(here, $h \otimes \Omega = h$ by convention). Then $l(h)^*$ exists and is given by

$$l(h)^* h_1 \otimes \dots \otimes h_n = \langle h, h_1 \rangle h_2 \otimes \dots \otimes h_n$$

and $l(h)^*\Omega = 0$. The operator $l(h)^*$ is called the left annihilation operator.

These operators are interesting because they have a very interesting combinatorial information. Let $h_1, h_2, \dots h_n \in \mathcal{H}$. Suppose one is in interested in the joint $*-$ distribution of the l_i's, i.e,

$$\phi(l_{i(1)}^{g(1)} \dots l_{i(k)}^{g(k)})$$

where $i(j) \in \{1, \dots, n\}$ and $g(j) \in \{., *\}$, $j = 1, \dots, k$. It is fascinating that $\phi(l_{i(1)}^{g(1)} \dots l_{i(k)}^{g(k)}) = 1$ if and only if this describes a non-crossing pairings of $[k]$, (in a way we shall not describe in detail. We refer the reader to Roland Speicher's notes [NS06] which talks in detailed about the combinatorics of free probability), and 0 otherwise. Thus we have,

$$\phi((l_1 + l_1^*)^k) = C_k$$

for k even,where C_k is the $k/2^{th}$ Catalan number. Since we are dealing with an operator with a compact spectrum, this information suffices to describe the distribution of $(l_1 + l_1^*)$, and this is precisely the semicircular law, given by

$$d\mu_s = \frac{\sqrt{4 - x^2}}{2\pi} \chi_{[-2,2]]} dm$$

This semicircular distribution will become the center piece of free probability as we shall see soon.

Now let $A = C^*(l(h) : h \in \mathcal{H})$ and let $\phi : A \to \mathbb{C}$ as usual be given by $\phi(a) = \langle \Omega, a\Omega \rangle$. It is not hard to show that if $\mathcal{H}_1 \perp \mathcal{H}_2$ are two subspaces of $\mathcal{H}$, then the algebras

$$C^*(l(h) : h \in \mathcal{H}_1) \text{ and } C^*(l(h) : h \in \mathcal{H}_2)$$

are freely independent in (A, ϕ) Now we talk about one of the main results in classical probability, and its free analogue.

1.0.4 CLT and Gauss Laws

The classical CLT- Central Limit Theorem is the corner stone of probability. The result was one of the earliest with the flavor of "universality", i.e, a universal object amid randomness.

Theorem 1.11 (Classical CLT)**.** *Let a_i be centered, variance 1, IID random variables. Then*

$$\mu\left(\frac{a_1 + \dots + a_n}{\sqrt{n}}\right) \to \mu_g$$

where g is the normal distribution given by it's Radon-Nikodym derivation

$$d\mu_g = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dm$$

The law that satisfies the central limit theorem is given the name Gauss law.

Theorem 1.12 (Free CLT)**.** *Let a_i be freely independent non commutative random variables in (M, τ) with mean 0 and variance 1. Then,*

$$\mu\left(\frac{a_1 + \ldots + a_n}{\sqrt{n}}\right) \to \mu_s$$

where μ_s is the semicircular distribution as seen before.

The marvelous thing about the free CLT is that it is almost unchanged except the Gauss law is now the semi-circular law. Before one attempts to prove this result, one must know about the distribution of a sum $a + b$ of freely independent random variables a and b (the non-commutative analogue of convolution). This begins the conversation of "free harmonic analysis", which we shall not enter into. But we will just say that such an analogue does exist, and involves the usage of Voiculescu's R-transform (analogous to the Fourier transform). It turns out that the proof of the CLT becomes rather easy once the machinery is established. Here is an important result to ponder about.

Theorem 1.13. *$L(\mathbb{F}_n) \cong W^*(s_1, \ldots, s_n)$ where s_i are freely independent semicirculars in a large algebra M.*

Classical Probability	**Free Probability**
Bounded random variable on $\mathbb{C}$	$a \in (M, \tau)$ non commutative probability space
Mixed moments $\mathbb{E}(X_1^{m_1} X_2^{m_2} \ldots X_n^{m_n})$	Mixed moments $\tau(a_{i_1} \ldots a_{i_k})$
Classical independence, and tensor product	Free independence and free product
CLT, Normal distribution	Free CLT, Semicircular distribution
Convolution $\mu_X * \mu_Y$	Free convolution $\mu_a \boxplus \mu_b$
Boltzmann entropy	Voiculescu's microstates free entropy

2 Random Matrices

Random matrix theory is a full fledged subject of its own. For the purposes of this series of talks, we need to discuss two of the great many results which provide us insight in understanding free probability, in particular, free entropy. They are Wigner's theorem, and Voiculescu's asymptotic freeness which can be thought of a vast generalization of Wigner's theorem.

Definition 2.1 (Random matrix)**.** An $n \times n$ random matrix is a matrix whose entries are random variables. It can also be thought of as a matrix valued random variable.

Definition 2.2 (Expected distributions)**.** Let X_N be a self adjoint random matrix of size $N \times N$. We think of X_N as a function $X_N : \Sigma \to M_N(\mathbb{C})$ on some probability space (Σ, σ). Integration with respect to σ has the meaning of taking the expected value, and will be denoted by E.

We are mainly interested in the spectral measure of these matrices, i.e, the the expected proportion of the eigenvalues of X_N that lie in a given interval $[a, b]$:

$$\Lambda_N([a, b]) = \frac{1}{N} E(\#\{\text{eigenvalues of } X_N(t) \text{ in } [a, b]\})$$

Let $\lambda_1(t), \ldots, \lambda_n(t)$ be the eigenvalues of $X(t)$, listed with multiplicity and viewed as random variables. Let

$$\nu_N^t = \frac{1}{N}\sum_{j=1}^{N} \delta_{\lambda_j}(t)$$

be a random measure associated with the list of eigenvalues. Then

$$\Lambda_N([a,b]) = E(\nu_N^t([a,b]))$$

is the expected value of ν_N. Setting $\mu_N = E(\nu_N^t)$ we see that ν_N^t is the distribution of X_N^t when viewed as a random variable in $(M_N(\mathbb{C}), \frac{1}{N}Tr)$. We are interested in studying this quantity (often times called the empirical spectral distribution).

2.0.1 Asymptotics of Random Matrices

Let X_N be a self adjoint random matrix, whose entries are G_{ij} determined as follows: The variables $\{g_{ij} : i \leq j\}$ are independent; if $i < j$ then g_{ij} is a centered complex Gaussian random variable of variance $1/N$. If $i = j$ then g_{ij} is a centered real Gaussian random variable of variance $2/N$. Finally, if $i > j$, $g_{ij} = \overline{g_{ji}}$.

Let μ_N be the expected value of the distribution of X_N. Then,

Theorem 2.3 (Wigner '58 [Wig58]). *as $N \to \infty$, we have the following weak convergence:*

$$\mu_N \to \mu_s$$

where μ_s is the usual semicircular distribution.

Families of certain $N \times N$ random matrices behave as free random variables in large N asymptotics. For each N, let D_N be a diagonal matrix. Assume that the operator norms $\|D_N\|$ are uniformly bounded and assume that the distribution of D_N converges in moments to a limit measure ν. Let $X_N^{(1)}, \ldots, X_N^{(k)}$ be random matrices described as follows: Let $\Sigma = M_N(\mathbb{C})^k$ with the measure σ given by

$$d\sigma(A_1, \ldots, A_k) = C_{N,k} e^{-\frac{1}{N}Tr(A_1^* A_1 + \ldots + A_k^* A_k)} dA_1 \ldots dA_k$$

for a suitable constant $C_{N,k}$. Then $X_N^{(p)}$ is the map

$$X_N^{(p)} : (A_1, \ldots, A_k) \mapsto \frac{A_p + A_p^*}{2}$$

More explicitly, if we denote by $g_{ij}^{(p)}$ the $i, j-th$ entry of $X_N^{(p)}$, then $\{g_{ij}^{(p)} : 1 \leq i \leq j \leq N, 1 \leq p \leq k\}$ form a family of independent centered Gaussian random variables, so that: $g_{ij}^{(p)}$ is a complex Gaussian of variance $1/N$ if $i < j$; $g_{ii}^{(p)}$ is real Gaussian of variance $2/N$ and $g_{ij}^{(p)} = \overline{g_{ji}^{(p)}}$ if $i > j$.

The family $(X_N^{(1)}, \dots, X_N^{(k)})$ is sometimes called the Gaussian Unitary Ensemble, or GUE, because of the obvious invariance of their joint distribution under conjugation by k unitaries.

Let μ_N be the distribution of the family $(D_N, X_N^{(1)}, \dots, X_N^{(k)})$, viewed as a linear functional on the space of polynomials in $k+1$ indeterminates. Then Voiculescu proved the following:

Theorem 2.4 (Voiculescu '91 [Voi91]). *Let $(d, x_1, x_2, \dots, x_k)$ be a family of free random variables in a non-commutative probability space (A, ϕ), so that d has distribution ν, and $x_1, \dots, x_k$ have semicircular distribution. Let μ be the distribution of this family, and let μ_N be the distribution of $(D_N, X_N^{(1)}, \dots, X_N^{(k)})$ as described above. Then as $N \to \infty$, $\mu_N \to \mu$ in moments. In other words. for any t and any $j_1, \dots, j_t \in \{1, \dots, k\}$, $n_0, \dots, n_t \in \{0, 1, 2, \dots\}$ one has*

$$\lim_{N\to\infty} E\left(\frac{1}{N} Tr(D_N^{n_0} X_N^{(j_1)} D_N^{n_1} \dots X_N^{(j_t)} D_N^{n_t})\right) = \phi(d^{n_0} x_{j_1} d^{n_1} \dots x_{j_t} d^{n_t})$$

Note that in particular we have D_N and $X_N^{(1)}, \dots, X_N^{(k)}$ are asymptotically free. One also recovers Wigner's theorem as a mere consequence of this. This asymptotic freeness result is very important to the definition of free entropy as we shall see, because free entropy is defined using matrix microstates in the asymptotics. The proof of asymptotic freeness is not covered here, and can be found in the paper. We are now ready to discuss free entropy.

3 Free Entropy (Caution: some minor edits pending in this section

3.1 Motivation

We will need to understand the foundations of Entropy, to really motivate free entropy. Chafai has written a beautiful survey [Cha15] on the history of entropy, and we refer the reader to that paper for a detailed list of references in the literature.

3.1.1 Boltzmann

Boltzmann defined Entropy because he wanted to model disorder in a particle system. We provide this very simple model of combinatorial disorder to motivate the concept of entropy.

Consider a system of n distinguishable particles, each of them being in one of the r possible states (typically energy levels). We have $n = n_1 + \dots + n_r$ where n_i is the number of particles in state i. The vector $(n_1, \dots, n_r)$ encodes the macroscopic state of the system, while the microscopic state of the system is encoded by the vector $(b_1, \dots, b_n) \in \{1, \dots, r\}^n$ where b_i is the state of the i-th particle. The number of microscopic states compatible with a fixed macroscopic state $(n_1, \dots, n_r)$ is given by the multinomial coefficient (encoding the occurrence of each face of an r-faces dice thrown n times):

$$\frac{n!}{(n_1! \dots n_r!)}$$

This integer measures the microscopic degree of freedom given the macroscopic state. As a consequence, the additive degree of freedom per particle is then naturally given by

$$\frac{1}{n} log \left(\frac{n!}{n_1! \ldots n_r!} \right)$$

. Now we are interested in the limiting behavior of this quantity. Let us suppose simply that when n tends to ∞ we have $\frac{n_i}{n} \to p_i$, (encoding a discrete probability distribution), for every $1 \leq i \leq r$. Then, thanks to our favorite Stirling's formula, we get, denoting $p := (p_1, \ldots, p_r)$,

$$\mathcal{S}(p) := \lim_{n \to \infty} \frac{1}{n} log \left(\frac{n!}{n_1! \ldots n_r!} \right) = - \sum_{i=1}^{r} p_i log(p_i)$$

Now we say that this quantity $\mathcal{S}(p)$ is the Boltzmann entropy of the discrete probability distribution p.

The right way to think about Boltzmann entropy is as an asymptotic additive degree of freedom per particle in a system. When the system is directly described by a probability density function $f : \mathbb{R}^d \to \mathbb{R}_+$ instead of a discrete probability measure, we may analogously define by passage to the limit, the continuous Boltzmann entropy of f, denoted $\mathcal{S}(f)$ or $-H(f)$ in the old terminology of Boltzmann himself,

$$\mathcal{S}(f) = - \int_{\mathbb{R}^d} f(x) \log(f(x)) dx$$

More generally, when X is a random variable, denote by $S(X)$ the entropy of its law. One should not be bogged down by the formula of entropy insofar as one forgets the formulation of it involving the volume of the microstates of the system.

Boltzmann was seeking to identify a probability density f_* that maximizes the linear functional $f \mapsto \mathcal{S}(f)$ over a convex class $\mathcal{C}$ formed by a set of constraints on f. One usually only cares about a fixed second moment condition. Indeed, it was proved that the Gaussian distribution (normal distribution) $e^{-1/2x^2}/\sqrt{2x}$ maximizes Boltzmann entropy under the constraint of fixed variance equal to 1.

3.1.2 Shannon

The Boltzmann entropy plays also a fundamental role in communication theory, founded in the 1940's by Claude Elwood Shannon (1916 to 2001), where it is known as "Shannon entropy". It has a deep interpretation in terms of uncertainty and information in relation with coding theory. For example the discrete Boltzmann entropy $\mathcal{S}(p)$ computed with a logarithm in base 2 is the average number of bits per symbol needed to encode a random text with frequencies of symbols given by the law p. This plays an essential role in lossless coding, and the Huffman algorithm for constructing Shannon entropic codes is probably one of the most used basic algorithm (data compression is everywhere)

We focus on a link suggested by Shannon between the Boltzmann entropy and the central limit theorem. Recall the CLT:

$$S_n := \frac{X_1 + \ldots + X_n}{\sqrt{n}} \to^d_{n\to\infty} \frac{e^{\frac{-1}{2}x^2}}{\sqrt{2\pi}}$$

where X_i are IID centered random variables with unit variance. Shannon observed that the entropy $\mathcal{S}$ is monotonic along the CLT when n is a power of w, in other words $\mathcal{S}(S_{2^{m+1}}) \geq \mathcal{S}(S_{2^m})$ for every integer $m \geq 0$, and this naively follows from:

$$\mathcal{S}(\frac{X_1 + X_2}{\sqrt{2}}) = \mathcal{S}(S_2) \geq \mathcal{S}(S_1) = \mathcal{S}(X_1)$$

A rigorous proof of this is due to Stam in [Sta59].

By analogy with the Boltzmann's H-theorem (see Chafai [Cha15]), a conjecture attributed to Shannon says that the Boltzmann entropy $\mathcal{S}$ is monotonic along the CLT for any n, more precisely

$$\mathcal{S}(X_1) = \mathcal{S}(S_1) \leq \ldots \leq \mathcal{S}(S_{n+1}) \ldots \leq \mathcal{S}(G)$$

The idea of proving the CLT using the Boltzmann entropy is very old and goes back to Linnik, but one should note that proving convergence differs from proving monotonicity, even if these two aspects are obviously linked.

3.1.3 "free" "entropy"

We have, at this point, seen free probability and entropy, and are thus ready to make sense of "free" "entropy". It is obviously an analogue of entropy in free probability. It must be noted that such an analogue is by no means straightforward to conjure. Voiculescu spent years until he could perfect his definition in [Voi94]. In fact, even the first paper he published [Voi93] in his series had a possible definition that was later proved incorrect by himself.

The free entropy (say χ) must necessarily (but not necessarily sufficiently) satisfy certain important properties, as we have noted before in this section. As we have already discussed, the analogue of the Gaussian in free probability theory is the semicircular law. Therefore, as we have noted in section 3.1.1, we should have that the free analogue of the Gaussian maximizes the free entropy under the constraint of a fixed second moment. That is,

$$\text{“}\chi(s) \geq \chi(X) \quad \text{for all centered non commutative RV's } X \text{ with } \mathbb{E}(X^2) = 1''$$

Spoiler alert, but this is indeed going to be true, and is proved in the upcoming sections. There is one more necessary property, called Shannon monotonicity, which is going to be the monotonicity of entropy along the CLT, in our case of free entropy along the free CLT.

$$\text{“}\chi(S_1) \leq \ldots \leq \chi(S_n) \to \chi(s) \quad \text{for } S_1, S_2, \ldots \text{ partial sums along the free CLT}''$$

Spoiler alert, this is also going to be true, except that the proof is this result is not discussed here. Shlyakhtenko in his wonderful paper [Shl07] proves this using an analogue of Fisher's information measure which Voiculescu introduced in [Voi98a], and in fact also gives an elegant proof of the classical Shannon monotonicity problem.

We would also like free entropy to satisfy other useful properties that entropy satisfies, like subadditivity, additivity for independent random variables (in our case, freely independent) and so on. But, while have a list of demands we have made for free entropy, we have no clue as to how define it. This is where we must dwell on the philosophical roots of entropy, which dates back to Boltzmann. We will need a free entropy functional maximized by the semicircular law at a fixed second moment, and which is monotonic along the free CLT. Inspired by the micro-macro construction of Boltzmann entropy, one considers an approximation at the level of moments of a non commutative random variable, by finite dimensional Hermitian matrices. The volume of this set of all "micro-states", living an appropriate measure space, will be the object of interest.

3.2 Definition of Free Entropy via Microstates

3.2.1 Setup

Unless otherwise stated, everything that follows is due to Voiculescu, primarily in [Voi94]. Let (M, τ) be a finite W^* probability space. Here M is a finite von Neumann algebra and τ is the unique faithful normal tracial state. Throughout these notes, we will almost always use self adjoint operators unless stated otherwise. Let $X_j \in M$ be self adjoint operators. We will denote by $\mathfrak{M}_k$ the $k \times k$ matrix algebra over $\mathbb{C}$, and τ_k the unique normalized trace on these matrices. $\mathfrak{M}_k^{SA}$ will stand for the $k \times k$ self adjoint matrices. Here we consider the measure λ to be the lebesgue measure on $(\mathfrak{M}_k^{SA})^n$ corresponding to the Euclidean norm given by the Hilbert-Schmidt norm:

$$\big\|(A_1, \dots, A_n)\big\|_{HS}^2 = Tr(A_1^2 + \dots + A_n^2)$$

So here we have a measure on the space of n tuples of self adjoint $k \times k$ matrices, which is precisely the lebesgue measure on $\mathbb{R}^{k(k+1)n}$ viewing these n-tuples as vectors in $\mathbb{R}^{k(k+1)n}$. Note that the Hilbert Schmidt norm here is the sum of the squares of all the entries of the matrices, which is the Euclidean norm of the n-tuple viewed as a vector in $\mathbb{R}^{k(k+1)n}$. Now, we define $\Gamma_R(X_1, \dots X_n; m, k, \epsilon)$ to be the set of n tuples of self adjoint $k \times k$ matrices $(A_1, \dots A_n) \in (\mathfrak{M}_k^{SA})^n$ such that $\|A_i\| \leq R$ and $|\tau_k(A_{i_1}, \dots A_{i_p}) - \tau(X_{i_1} \dots X_{i_p})| < \epsilon$ for all $(i_1, \dots, i_p) \in (1, \dots, n)^p$ where $1 \leq p \leq m$. In words, the set $\Gamma_R(X_1, \dots, X_n; m, k, \epsilon)$ takes in a parameter R that bounds the operator norms of the individual matrix approximants, an integer m that acts as the maximum degree of the monomials that are approximated in trace, and k which determines the dimension of the matrices. The set $\Gamma_R(X_1, \dots, X_n; m, k, \epsilon)$ is the candidate we are interested in because it contains the set of "microstates" that approximate our operators in distribution. So on our way to define the entropy, we try and remove the dependance of $\Gamma_R(X_1, \dots, X_n; m, k, \epsilon)$ on these parameters one by one through the process of finding the "asymptotic additive degree of freedom". In the future, in case we write Γ_R without writing its parameters, we mean $\Gamma_R(X_1, \dots, X_n; m, k, \epsilon)$.

3.2.2 Definition of $\chi(X_1, \ldots, X_n)$

Definition 3.1. For self adjoint variables $X_1, \ldots X_n$ in (M, τ) we define the following:

$$\chi_R(X_1, \ldots, X_n; m, k, \epsilon) = \log \lambda(\Gamma_R(X_1, \ldots, X_n; m, k, \epsilon))$$

$$\chi_R(X_1, \ldots, X_n; m, \epsilon) = \limsup_{k\to\infty}(k^{-2}\chi_R(X_1, \ldots, X_n; m, k, \epsilon) + (n \log k)/2)$$

$$\chi_R(X_1, \ldots, X_n) = \inf_{m\in\mathbb{N}, \epsilon>0} \chi_R(X_1, \ldots, X_n; m, \epsilon)$$

$$\chi(X_1, \ldots, X_n) = \sup_{R>0} \chi_R(X_1, \ldots, X_n)$$

This $\chi(X_1, \ldots, X_n)$ is defined to be the Microstates Free Entropy.

Here are a few remarks about the above definition. Firstly, observe that in the third and fourth lines, one can replace the inf/sup with a limit along the appropriate net. This is because $\chi_R(X_1, \ldots, X_n; m, k, \epsilon)$ is a decreasing function with $m, R \uparrow \infty$, and also decreasing as $\epsilon \downarrow 0$.

3.3 Immediate observations and properties

3.3.1 Boundedness on fixed 2nd moment

Although the removal of constraints is clear throughout the definition of χ, there is a lack of motivation for the normalization constants in line 2. Here we present a boundedness result that provides the above motivation. We remark that the true reason why one wants this is because in order for free entropy to be the right analogue of entropy, as said before, we must have that the free entropy is maximized by the semicircular law. Indeed to this end, we shall see that the free entropy of the semicircular law is indeed the one we see in the below proposition.

Proposition 3.2. *For a fixed second moment, $C^2 = \tau(X_1^2 + \ldots + X_n^2)$, we have the following:*

$$\chi_R(X_1, \ldots, X_n; m, k, \epsilon) \leq \frac{nk(k+1)}{2}\left(\log\left(\frac{2\pi e\left(C^2 + n\epsilon\right)}{n}\right) - \log k\right)$$

Proof. We use Shannon's p-dimensional inequality here. It states the following:

$$-\int f \log f d\lambda_p \leq \frac{p\log(2\pi e a^2/p)}{2}$$

Where f is a probability density function on the space $\mathbb{R}^p$ and λ_p is the Lebesgue measure on the space, and $a^2 = \int(x_1^2 + \ldots + x_p^2) f d\lambda_p$. Now, we consider our p to be $k(k+1)n$ which is now the same as replacing $\mathbb{R}^p$ by $(\mathfrak{M}_k^{SA})^n$. We define f to be the indicator density function on our set of matrix approximants Γ. So, define

$$f(A_1, \ldots, A_n) = \frac{1}{\lambda(\Gamma_R(X_1, \ldots, X_n; m, k, \epsilon))} \text{ when } (A_1, \ldots, A_n) \in \Gamma_R(X_1, \ldots, X_n; m, k, \epsilon)$$

And 0 otherwise. This is indeed the setup we desire. We have

$$-\int f\log(f)d\lambda_p = \frac{1}{\lambda_p(\Gamma_R)}\int_{\Gamma_R}\chi_R(X_1,\ldots,X_n;m,k,\epsilon)d\lambda_p = \chi_R(X_1,\ldots,X_n;m,k,\epsilon)$$

Now, applying Shannon's inequality in our setup, we have

$$\chi_R(X_1,\ldots,X_n;m,k,\epsilon) \leq \frac{nk(k+1)\left(\log(2\pi e a^2) - \log(nk(k+1))\right)}{2}$$
$$\leq \frac{nk(k+1)\left(\log(2\pi e a^2/k) - \log(nk)\right)}{2}$$

It suffices to prove now that $a^2/k \leq C^2 + n\epsilon$, i.e,

$$\int \frac{1}{k}Tr(A_1^2 + \ldots A_n^2)fd\lambda_p = \frac{1}{\lambda_p(\Gamma_R)}\int_{\Gamma_R}\frac{1}{k}Tr(A_1^2+\ldots A_n^2)fd\lambda_p \leq \sum_{j=1}^{n}\tau(X_j^2) + n\epsilon$$

But now we are only interested in $(A_1,\ldots,A_n) \in \Gamma_R$, and these are precisely the n-tuples that approximate our $X_1,\ldots,X_n$ in distribution by ϵ. So, we have that

$$\left|\frac{1}{k}Tr\left(A_1^2+\ldots+A_n^2\right) - \tau\left(\sum_{i=1}^{n}X_i^2\right)\right| \leq n\epsilon$$

This gives us the result. □

We have as an immediate corollary, the entropy of n self adjoint variables with a given second moment C^2 is bounded above.

Corollary 3.3. *Let* $C^2 = \tau(\sum_{i=1}^{n} X_i^2)$. *We have if* $m \geq 2$,

$$\chi(X_1,\ldots,X_n) \leq \frac{n\log(2\pi e C^2/n)}{2}$$

Proof. Following our definition of free entropy, which was indeed designed to make the above corollary work, and Proposition 2.2, we have

$$\chi_R(X_1,\ldots,X_n;m,\epsilon) = \limsup_{k\to\infty}(k^{-2}\chi_R(X_1,\ldots,X_n;m,k,\epsilon) + (n\log k)/2)$$
$$\leq \frac{n}{2}\log\left(\frac{2\pi e(C^2+n\epsilon)}{n}\right)$$

Now taking $\inf$ along the appropriate nets, we have our result:

$$\chi(X_1,\ldots,X_n) \leq \frac{n\log(2\pi e C^2/n)}{2}$$

□

3.3.2 Subadditivity

This result will be important for us later: we shall be looking at when additivity occurs, and show that it happens only when the random variables are freely independent, thereby showing that the free independence is the right kind of independence for free entropy as one would expect (for classical entropy is additive if and only if the random variables are classically independent).

Proposition 3.4. *If $1 \leq p < n$, then we have*

$$\chi_R(X_1, \dots, X_n; m, k, \epsilon) \leq \chi_R(X_1, \dots, X_p; m, k, \epsilon) + \chi_R(X_{p+1}, \dots, X_n; m, k, \epsilon)$$

Proof. Suppose $(A_1, \dots, A_n) \in \Gamma_R(X_1, \dots, X_n; m, k, \epsilon)$, we have simply from the definition that for any monomial P taking p variables with degree $\leq m$,

$$\left|Tr_k(P(A_1, \dots, A_p)) - \tau(P(X_1, \dots X_p))\right| \leq \epsilon$$

and for any monomial Q taking $n-p$ variables with degree $\leq m$, we have

$$\left|Tr_k(Q(A_{p+1}, \dots, A_n)) - \tau(P(X_{p+1}, \dots X_n))\right| \leq \epsilon$$

This shows that

$$\Gamma_R(X_1, \dots, X_n; m, k, \epsilon) \subseteq \Gamma_R(X_1, \dots, X_p; m, k, \epsilon) \times \Gamma_R(X_{p+1}, \dots, X_n; m, k, \epsilon)$$

Taking the product measure and log, we have our result. □

The subadditivity result is weaker than the above.

3.3.3 Saturation of χ_R Based on Operator Norms

The boundedness condition on the norms of matrix approximants is relaxed (as in, we take $R \to \infty$ in the definition), but we would like to know more about this limit. Here we show that it indeed saturates when R is bigger than all of the operator norms of the bounded self adjoint random variables.

Theorem 3.5. *If $\rho = max_{1 \leq j \leq n}\left\|X_j\right\|$ and $R > \rho$, then*

$$\chi_R(X_1, \dots, X_n) = \chi(X_1, \dots, X_n)$$

Proof. We prove an equivalent statement: If $\rho < R_1 < R_2$, then $\chi_{R_1}(X_1, \dots, X_n) = \chi_{R_2}(X_1, \dots, X_n)$. Since we know that χ is an increasing function with R, it suffices to prove that

$$\chi_{R_2}(X_1, \dots, X_n) \leq \chi_{R_1}(X_1, \dots, X_n)$$

The strategy we shall adopt is to define a map from Γ_{R_2} to Γ_{R_1} and show that the Jacobian of this map is bounded from below. Fix an $\rho < R_0 < R_1$. Now, we let $g : [-R_2, R_2] \to \mathbb{R}$ be the function which is linear on the intervals $[-R_2, -R_0]$, $[-R_0, R_0]$, $[R_0, R_2]$, and such that $g(-R_2) = -R_1$, $g(-R_0) = -R_0$, $g(R_0) = R_0$, and $g(R_2) = R_1$. Define $G(A_1, \dots, A_n) = (g(A_1), \dots, g(A_n))$

where the $g(A_i)$ is nothing but the functional calculus for $g|_{Spec(A)}$ applied to the matrix A_i.

Now, we contest that this is the map we desire. We claim that given $m \in \mathbb{N}$ and $\epsilon > 0$ there are $m_1 > m$ and $\epsilon_1 < \epsilon$ such that

$$G(\Gamma_{R_2}(X_1, \dots, X_n; m_1, k, \epsilon_1)) \subseteq \Gamma_{R_1}(X_1, \dots, X_n; m, k, \epsilon) \tag{1}$$

Immediately we see that each coordinate in $G(A_1, \dots, A_n)$ has operator norm less than R_1 by definition of the functional calculus. To prove the above inclusion, it suffices to prove that for a fixed $\delta > 0$, $\left\|g(A_j) - A_j\right\|_1 = Tr_k(\left|g(A_j) - A_j\right|) < \delta$.
If we show that

$$\tau_k(E(A_j; [-R_2, -R_0] \cup [R_0, R_2])) \leq \delta$$

i.e, the trace of the spectral projection of A_j corresponding to the above set on $\mathbb{R}$ (intersection with the spectrum) is less than δ, then we are done. Here we give a sketch of a proof of the claim, and request the reader to work out the formal details:

This is because on $[-R_0, R_0]$, $g(A_j)$ and A_j agree (with respect to the spectral distribution), and we only care about the part of the spectrum that is in the above set. Observe that $\tau_k(E(A_j; [-R_2, -R_0] \cup [R_0, R_2]))$ is simply recording the (normalized) number of eigenvalues of A_j in that range. Intuitively, if there were many eigenvalues in this range, then, for a sufficiently small ϵ_1 and large m_1, it would be hard for A_j to approximate X_j (whose spectrum is contained in $[-\rho, \rho]$) in distribution. Note, the role that ϵ_1 and m_1 play is as follows: Since $A_j \in \Gamma_{R_2}(X_1, \dots, X_n; m_1, k, \epsilon_1)$ we know that $\left|\tau_k(A_j^p)\right| \leq \tau(X_j)^p + \epsilon_1 \leq \rho^p + \epsilon_1$, so for very small ϵ_1 and large m_1, we have that the spectrum of A_j is "very similar" to the spectrum of X_j, (since the polynomials are dense- the case when $m_1 \to \infty$ and the approximations are very close for $\epsilon_1 \to 0$).

Now that we have proved (1), our goal is to investigate the Jacobian of this transformation. Through a quick computation, we see that the absolute value of the Jacobian of the map $A \to G(A)$, where $A \in \mathfrak{M}_k^{SA}$ is given by

$$\left|\mathfrak{J}(G)\right| = \prod_{i \neq j} \frac{g(\lambda_i) - g(\lambda_j)}{\lambda_i - \lambda_j} g'(\lambda_1) \dots g'(\lambda_k)$$

where A has eigenvalues $\lambda_1 < \dots < \lambda_n$, and $\lambda_j \neq R_0$ (this can be guaranteed from the choice of R_0). Now, we assume that $\tau_k(E(A; [-R_2, -R_0] \cup [R_0, R_2])) < \delta$, which is what our matrices in Γ_{R_2} satisfied. Firstly, observe that for all $\lambda_{i,j}$, we have $\left|g(\lambda_i) - g(\lambda_j)\right| \leq R_1 - R_0$, and similarly $\left|\lambda_i - \lambda_j\right| < R_2 - R_0$. Recall that $\tau_k(E(A; [-R_2, -R_0] \cup [R_0, R_2])) < \delta$ just tells us that $k(1-\delta)$ eigenvalues of A lie inside $[-R_0, R_0]$. Now using this, we get the evident "very weak" lower bound:

$$\left(\frac{R_1 - R_0}{R_2 - R_0}\right)^{k(k+1) - (k(1-\delta))^2} \leq \left|\mathfrak{J}(G)\right|$$

Hence, directly applying this inequality, we have

$$\chi_{R_1}(X_1, \dots, X_n; m, k, \epsilon) \geq \log \lambda(G(\Gamma_{R_2}(X_1, \dots, X_n; m_1, k, \epsilon_1)))$$

Simplifying using the jacobian, we have

$$\chi_{R_1}(X_1,\ldots,X_n;m,k,\epsilon) \geq \chi_{R_2}(X_1,\ldots,X_n;m_1;k,\epsilon_1) + n(k+k^2(2\delta-\delta^2))\log\frac{R_1-R_0}{R_2-R_0}$$

Then, removing k, we have

$$\chi_{R_1}(X_1,\ldots,X_n;m,\epsilon) \geq \chi_{R_2}(X_1,\ldots,X_n;m_1,\epsilon_1) + 2n\delta\log\frac{R_1-R_0}{R_2-R_0}$$

But, we see that δ can be made arbitrarily small using m_1 and ϵ_1, so we have

$$\chi_{R_1}(X_1,\ldots X_n) \leq \chi_{R_2}(X_1,\ldots,X_n)$$

which gives us our result. □

The above theorem can be made more refined by looking at the following proposition.

Proposition 3.6. *Define $\Gamma_{R_1,\ldots R_n}(X_1,\ldots,X_n;m,k,\epsilon)$ to be defined like our original Γ_R, except we require that $\|A_j\| \leq R_j$. Then, for $\|X_j\| < R_j < R'_j$ we have*

$$\chi_{R_1,\ldots R_n}(X_1,\ldots,X_n) = \chi_{R'_1,\ldots,R'_n}(X_1,\ldots,X_n)$$

Proof. The proof is a run through of exactly the same idea as before. Except now, define $\|X_i\| < R_{i_0} < R_i$, and $g_i : [-R'_i, R'_i] \to \mathbb{R}$ is piecewise linear between $[-R'_i, -R_{i_0}]$, $[-R_{i_0}, R_{i_0}]$, $[R_{i_0}, R'_i]$ and such that $g(\pm R'_i) = \pm R_i$, $g(\pm R_{i_0}) = \pm R_{i_0}$. The rest of the proof is immediate as we follow the same strategy but with this setup. □

3.3.4 Limits in Distribution

We consider a notion of convergence in operators, called convergence in distribution, and observe the connections with the entropy of the limiting operators. We remind the reader about another one of the important properties we want free entropy to possess: Monotonicity along the free central limit theorem. This proposition will be towards that direction, although the real problem at hand is very hard and requires tools beyond the scope of this presentation. See [Shl07]

Proposition 3.7. *Let $(X_1,\ldots,X_n)$ and $(X_1^{(p)},\ldots,X_n^{(p)})$ $(p \in \mathbb{N})$ be n-tuples of self adjoint random variables in (M,τ), so that $(X_1^{(p)},\ldots,X_n^{(p)})$ converges in distribution to $(X_1,\ldots,X_n)$, i.e,*

$$\lim_{p\to\infty}\tau(X_{i_1}^{(p)}\ldots X_{i_m}^{(p)}) = \tau(X_{i_1},\ldots,X_{i_m})$$

for all $1 \leq i_j \leq n$, $1 \leq j \leq m$, $m \in \mathbb{N}$, then one has

$$\limsup_{p\to\infty}\chi_R(X_1^{(p)},\ldots,X_n^{(p)}) \leq \chi_R(X_1,\ldots,X_n)$$

Proof. Suppose $(A_1, \ldots, A_n) \in \Gamma_R(X_1^{(p)}, \ldots, X_n^{(p)}; m, k, \epsilon)$, we observe that for a sufficiently large p, for all monomials P, we have from the convergence in distributions of $(X_1^{(p)}, \ldots, X_n^{(p)})$ to $(X_1, \ldots, X_n)$, that $\left|\tau(P(X_1^{(p)}, \ldots, X_n^{(p)}) - \tau(P(X_1, \ldots, X_n))\right| < \epsilon$, so we have

$$\left|\tau_k(P(A_1, \ldots, A_n)) - \tau(P(X_1, \ldots, X_n))\right| < 2\epsilon$$

Hence, $(A_1, \ldots, A_n) \in \Gamma_R(X_1, \ldots, X_n; m, k, 2\epsilon)$. Therefore we have

$$\Gamma_R(X_1^{(p)}, \ldots, X_n^{(p)}; m, k, \epsilon) \subseteq \Gamma_R(X_1, \ldots, X_n; m, k, 2\epsilon)$$

Now, we see that for sufficiently large p,

$$\chi_R(X_1^{(p)}, \ldots, X_n^{(p)}) \leq \chi_R(X_1, \ldots, X_n; m, 2\epsilon)$$
$$\Rightarrow \limsup_{p\to\infty} \chi_R(X_1^{(p)}, \ldots, X_n^{(p)}) \leq \chi_R(X_1, \ldots, X_n; m, 2\epsilon)$$
$$\Rightarrow \limsup_{p\to\infty} \chi_R(X_1^{(p)}, \ldots, X_n^{(p)}) \leq \chi_R(X_1, \ldots, X_n)$$

as required. □

Corollary 3.8. *If, moreover* $\sup_{p\in\mathbb{N}} \left\|X_j^{(p)}\right\| < \infty$*, for* $1 \leq j \leq n$*, then*

$$\limsup_{p\to\infty} \chi(X_1^{(p)}, \ldots, X_n^{(p)}) \leq \chi(X_1, \ldots, X_n^{(p)})$$

Proof. Let $R > 0$ be such that $R > \left\|X_j^{(p)}\right\|$ for all p and $1 \leq j \leq n$. Then from Theorem 2.6 and Proposition 2.8 we see that

$$\limsup_{p\to\infty} \chi_R(X_1^{(p)}, \ldots, X_n^{(p)}) = \limsup_{p\to\infty} \chi(X_1^{(p)}, \ldots, X_n^{(p)}) \leq \chi(X_1, \ldots, X_n) = \chi_R(X_1, \ldots, X_n)$$

as required. □

3.3.5 Conditioning

As a quick note, we take a look at conditional free entropy, a quantity that might be of use. We consider our usual setup of a non commutative probability space (M, τ).

Definition 3.9. Let $X_1, \ldots, X_n$, $Y_1, \ldots, Y_n$ be self adjoint random variables in (M, τ). Suppose $\chi(Y_1, \ldots, Y_n) > -\infty$, then the conditional free entropy is given by:

$$\chi(X_1, \ldots, X_n | Y_1, \ldots, Y_n) = \chi(X_1, \ldots, X_n, Y_1, \ldots, Y_n) - \chi(Y_1, \ldots, Y_n)$$

Proposition 3.10. *If* $\chi(Y_1, \ldots, Y_m) > \infty$ *then*

$$\chi(X_1, \ldots, X_n | Y_1, \ldots Y_m) \leq \chi(X_1, \ldots X_n)$$

Proof. This is just a direct application of subadditivity. □

3.4 How to compute Free Entropy?

We begin with the computation of free entropy for one variable. We seek herein a formula similar to the "logarithmic energy" integral we see in Boltzmann theory. We will need a couple of crucial lemmas. Before we start the hard work, we remark we are dealing in general with volumes, and so we will be on the lookout for familiar subsets whose volumes can be computed, that approximate our microstate space.

3.4.1 A nice formula for $\chi(X)$

We begin with a very powerful lemma in linear algebra 101.

Lemma 3.11 (Hoffman Wielandt inequality)**.** *Let* $A, B \in \mathfrak{M}_k^{sa}$ *and let* $\lambda_1 \leq \lambda_2 \ldots \leq \lambda_k$ *and* $\mu_1 \leq \ldots \leq \mu_k$ *be their eigenvalues. Then*

$$\sum_{1 \leq j \leq k} (\lambda_j - \mu_j)^2 \leq Tr(A-B)^2$$

Proof. An involved proof of this is provided in [AGZ09] in chapter 2, section 2.1.5. □

Now, in the following lemma, we describe the volume of a set that could prove to be useful in approximating the microstate space.

Lemma 3.12. *Let* $A \in \mathfrak{M}_k^{sa}$ *have eigenvalues* $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_k$ *and let* $\epsilon > 0$ *and* $0 < \alpha < 1/2$. *If*

$$\Omega = \{B \in \mathfrak{M}_k^{sa} | Tr(B - UAU^*)^2 \leq k\epsilon^2\theta \quad \textit{for some } U \textit{ unitary}\}$$

where $\theta = (\alpha + 2\alpha^2)(\alpha+2)^{-1}$, *then*

$$\lambda(\Omega) = k^{k/2} \cdot \epsilon^k \cdot \big(\Gamma(k/2+1)\big)^{-1} \cdot (1+2\alpha)^{k(k-1)/2} \cdot e^{2k^2\epsilon} \cdot \pi^{k^2/2} \cdot 2^{k(k-1)/2} \cdot \left(\prod_{1 \leq j \leq k} j!\right)^{-1} \cdot \prod_{1 \leq i < j \leq k} ((\mu_i - \mu_j)^2 + \epsilon)$$

Before we begin the proof, we remark that Ω is the set of matrices that approximate some element in the unitary orbit of the matrix A. This is useful because one can approximate the measure of the unitary orbit, as we shall see.

Proof. From the Hoffman Wielandt inequality, we see that any $B \in \Omega$, with eigenvalues $\lambda_1 \leq \lambda_2 \ldots \leq \lambda_k$ satisfies $\sum_j (\lambda_j - \mu_j)^2 \leq k\epsilon^2\theta$. Also, in the other direction, any matrix with eigenvalues satisfying the above condition will be in Ω, with U being the unitary that diagonalizes A. Hence, we have

$$\Omega = \{B \in \mathfrak{M}_k^{sa} : (\lambda_1, \ldots, \lambda_k) \in D, \lambda_1 \leq \ldots \leq \lambda_k \text{ the eigenvalues of B}\}$$

where

$$D = \{(\lambda_1, \ldots, \lambda_k) \in \mathbb{R}^k : \sum_j (\lambda_j - \mu_j)^2 \leq k\epsilon^2, \lambda_1 \leq \ldots \leq \lambda_k\}$$

The k-dimensional volume of D is bounded by the volume of the k-dimensional ball of radius $\sqrt{k\epsilon^2\theta}$ (this is evident because the constraint is precisely meaning that $(\lambda_j)_{j=1,2,\ldots,k}$ sits inside the ball of such radius centered at $(\mu_j)_{j=1,2,\ldots,k}$), which is

$$\pi^{k/2} \cdot \left(\Gamma(k/2+1)^{-1}\right) (k\theta\epsilon^2)^{k/2}$$

where Γ is the Γ-function. This is just the formula for the volume of the k-dimensional ball. Now, we use a theorem from [Meh91] (theorem 3.3.1), to get the following bound:

$$\lambda(\Omega) \leq \pi^{k/2}\Gamma(k/2+1)^{-1}(k\theta\epsilon^2)^{k/2} \cdot C_k \cdot \mathcal{J}$$

where

$$C_k = (2\pi)^{k(k-1)/2} \left(\prod_{1\leq i\leq k} j!\right)^{-1}$$

and

$$\mathcal{J} = \sup\left(\prod_{1\leq i<leqk} (\lambda_i - \lambda_j)^2 : (\lambda_1, \ldots, \lambda_n) \in D\right)$$

We are left to find an appropriate upper bound for $\mathcal{J}$. If $(\lambda_1, \ldots, \lambda_n) \in D$, let $\delta_j = \epsilon_j\theta^{-1/2}$ be defined so that $\sum_j \delta_j^2 \leq k\epsilon^2$ and

$$(\lambda_i - \lambda_j)^2 \leq (1+2\alpha)(\mu_i - \mu_j)^2 + (1+2\alpha^{-1})(\epsilon_i^2 + \epsilon_j^2)$$

after substitution and simplification, we have

$$\leq (1+2\alpha)((\mu_i - \mu_j)^2 + \delta_i^2 + \delta_j^2)$$

Plugging this into the product, we now have

$$\prod_{1\leq i<j\leq k} (\lambda_i - \lambda_j)^2 \leq (1+2\alpha)^{k(k-1)/2} \cdot \prod_{1\leq i<j\leq k} \left((\mu_i - \mu_j)^2 + \epsilon + \delta_i^2 + \delta_j^2\right)$$

Note, we added the epsilon to try and massage out the δ's. Now we provide an approximation to the term on the right. We have,

$$\sum_{1\leq i<j\leq k} (log((\mu_i - \mu_j)^2 + \epsilon + \delta_i^2 + \delta_j^2) - log((\mu_i - \mu_j)^2 + \epsilon))$$

$$= \sum_{1\leq i<j\leq k} log\left(1 + \frac{\delta_i^2 + \delta_j^2}{(\mu_i - \mu_j)^2 + \epsilon}\right) \leq \sum_{1\leq i<j\leq k} \frac{\delta_i^2 + \delta_j^2}{(\mu_i - \mu_j)^2 + \epsilon}$$

$$\leq \sum_{1\leq i<j\leq k} \frac{\delta_i^2 + \delta_j^2}{\epsilon} \leq 2k \cdot \frac{1}{\epsilon} \cdot k\epsilon^2 = 2k^2\epsilon$$

Hence, we have,

$$\prod_{1\le i<j\le k} (log((\mu_i-\mu_j)^2+\epsilon+\delta_i^2+\delta_j^2) \le e^{2k^2\epsilon} \prod_{1\le i<j\le k} ((\mu_i-\mu_j)^2+\epsilon)$$

We therefore have,

$$\mathcal{J} \le \prod_{1\le i<j\le k} ((\mu_i-\mu_j)^2+\epsilon)(1+2\alpha)^{k(k-1)/2} e^{2k^2\epsilon}$$

Combining all the terms, gives us the required conclusion. □

One should not tremble at the horrendous formula we have provided, as asymptotically most terms will vanish. Now, we will look at the true reason why we are considering this set.

Lemma 3.13. *Let $\epsilon > 0$ be given. Then, there exists is a $\omega > 0$, and an $N \in \mathbb{N}$ so that the following is satisfied: Given $A, B \in \mathfrak{M}_k^{sa}$ and $\|A\| \le 1$,*

$$\left|\tau_k(A^p) - \tau_k(B^p)\right| < \omega \quad (1 \le p \le N)$$

implies that there exists a unitary $U \in \mathfrak{M}_k^{sa}$ so that

$$\tau_k((B - UAU^*)^2) < \epsilon$$

The lemma says that the microstate space will be a subset of the set Ω_k, in an informal way. In order to prove this we will have to develop some machinery.

Proof. Firstly, we show that it suffices to assume that $\|B\| \le 1$. This follows from a continuity type argument as follows. Suppose for given $\epsilon > 0$, we have N and ω as in the lemma, if $B \in \mathfrak{M}_k^{sa}$, $\|B\| > 1$ such that for all $1 \le p \le N$, we have

$$\left|\tau_k(A^p) - \tau_k(B^p)\right| < \omega$$

implies the existence of a $U \in \mathfrak{U}(\mathfrak{M}_k^{sa})$ so that

$$\tau_k((B - UAU^*)^2) < \epsilon$$

Consider $\rho : \mathbb{R} \to [-1, 1]$ to be the continuous retraction of $\mathbb{R}$ to $[-1, 1]$ fixing the interval, and sending $x \mapsto \frac{x}{|x|}$. Now consider the functional calculus $\rho(B)$, and note that what ρ does is to cut off the spectrum of B beyond $[-1, 1]$, so that $\left\|\rho(B)\right\| \le 1$. Now, from the fact that B approximates A nicely in moments $(1 \le p \le N)$ where $\|A\| \le 1$, we have that there are $\delta > 0$ and $M \in \mathbb{N}$ so that for all $k \in \mathbb{N}$, we have $\tau_k(B^{2M}) < 1 + \delta$ further implying that

$$\left|\tau_k(B^p) - \tau_k(\rho(B)^p)\right| < \omega$$

One also has in particular, the second moment of their difference is small

$$\tau_k((B - \rho(B))^2) < \epsilon$$

Hence, we can replace B by $\rho(B)$, thereby proving that it suffices to consider $\|B\| \leq 1$.

Now, for $A \in \mathfrak{M}_k^{sa}$ let μ_A and $\mathcal{F}_A$ denote the distribution of A with respect to τ_n and the distribution function of μ_A respectively, that is,

$$\mathcal{F}_A(t) = \mu_A((-\infty, t)) = \tau_n(E(A; (-\infty, t)))$$

where recall that the notation $E(A, \Omega)$ is the spectral measure with respect to A. Now, the Levy-metric on distribution functions is given by the following:

$$d(\mathcal{F}_1, \mathcal{F}_2) = \inf\{\epsilon > 0 : \mathcal{F}_1(t - \epsilon) - e\epsilon \leq \mathcal{F}_2(t) \leq \mathcal{F}_1(t + \epsilon) + \epsilon \text{ for all } t \in \mathbb{R}\}$$

Firstly, observe that this metric induces the weak topology on the space of probability measures on $[-1, 1]$ and also that this space of measures is a compact space. Secondly, as a consequence of the fact that the following metric:

$$r(\mu_\nu) = \sum_{p=1}^{\infty} 2^{-p} \left| \int_{-1}^{1} t^p d\mu(t) - \int_{-1}^{1} t^p d\nu(t) \right|$$

also induces the same weak topology on the space of measures, we have the following: Given $\delta > 0$ there are $\omega > 0$ and $N \in \mathbb{N}$ so that for all $k \in \mathbb{N}$, $A, B \in \mathfrak{M}_k^{sa}$, $\|A\| \leq 1$, $\|B\| \leq 1$, if

$$\left|\tau_k(A^p) - \tau_k(B^p)\right| < \omega \quad (1 \leq p \leq N)$$

then

$$d(\mathcal{F}_{\mu_A}, \mathcal{F}_{\mu_B}) < \delta$$

Now we begin with some estimates. Choose an appropriate N and ω such that

$$d(\mathcal{F}_{\mu_A}, \mathcal{F}_{\mu_B}) < (10M)^{-3}$$

for some given $M \in \mathbb{N}$ satisfying $10^3/M < \epsilon$.

The main idea of the proof is to locate the unitary U, which will extend from a sum of partial isometries which each transport the spectral projections from A to B, maintaining some close approximations.

Begin by partitioning the ambient spectrum $[-1, 1]$ into $2M^3$ small intervals depending on $-M^3 \leq a < M^3$, $I(a) = [a(M^{-3}), (a + 1)(M^{-3}))$. We also want the very last of these intervals, i.e, $a = M^3 - 1$ to be closed on the right. Now, for each $-M \leq j < M$, pick a_j between $jM^2 \leq a_j < (j + 1)M^2$, such that $\mu_A(I(a_j) \leq M^{-2}$. Note that choosing M sufficiently large will mean that this can definitely be accomplished because the eigenvalues of a matrix are discrete points on the spectrum. From the bound for the Levy distance, we infer the following inequality, which is basically informing us of the closeness of the distributions, when restricted to some special intervals:

$$\mu_A([(a_j + 1)M^{-3}, a_{j+1}M^{-3})) - \mu_B([(a_j + 1/2)M^{-3}, (a_{j+1} + 1/2)M^{-3})) \leq \frac{1}{10M^3}$$

(changing the right endpoints appropriately for $j = -M-1$ and $j = M$ so that we don't cross the limits). Let U_j be partial isometries that transport the part of the spectrum of A to that of B, as described in the above inequality: (here, $-M-1 \leq j \leq M$)

$$U_j^* U_j \leq E(A; [(a_j+1)M^{-3}, a_{j+1}M^{-3}))$$

and

$$U_j U_j^* \leq E(B; [(a_j+1/2)M^{-3}, (a_{j+1}M^{-3})))$$

From the estimates, we have

$$\tau_n(E(A, [(a_j+1)M^{-3}, a_{j+1}M^{-3}))) \leq \frac{1}{10M^3}$$

Therefore, expanding,

$$\sum_j \tau_n(U_j^* U_j) \geq 1 - \frac{(M+2)}{10M^3} - M \times M^{-2} \geq 1 - \frac{2}{M}$$

We also have a norm estimate below:

$$\left\| U_j A - B U_j \right\| \leq \left\| U_j A - \frac{j}{M} U_j \right\| + \left\| \frac{j}{M} U_j - B U_j \right\| \leq \frac{8}{M}$$

Now, let $W = \sum_j U_j$. W is a partial isometry, and there is a unitary extending W, let it be U. This will be the unitary that we seek in the lemma. Indeed,

$$\|WA - BW\| \leq \frac{8}{M}$$

and

$$\tau_n((U-W)^*(U-W)) \leq \frac{2}{M}$$

so that

$$\begin{aligned} \tau_n((B - UAU^*)^2) &= \tau_n((BU - UA)^*(BU - UA)) \\ \leq \sqrt{\tau_n((BW - WA)^*(BW - WA))} &+ 2\sqrt{\tau_n((U-W)^*(U-W))} \\ \leq (\frac{8}{M} + \frac{4}{M^{1/2}})^2 &\leq \frac{10^3}{M} < \epsilon \end{aligned}$$

Thus, we have the result. □

We will be needing the following deterministic result which will show up as an additive constant in our formula for free entropy.

Lemma 3.14.

$$\lim_{k \to \infty} \left(-\frac{log \prod_{1 \leq j \leq k} j!}{k^2} + \frac{log(k)}{2} \right) = \frac{3}{4}$$

Proof. From Stirling's formula, we have

$$\frac{1}{k^2}\left(log(\prod_{1\leq j\leq k} j!) - log(\prod_{1\leq j\leq k} j^j)\right) = -\frac{1}{2}$$

From the fact that $\prod_{1\leq j\leq k} j! = \prod_{1\leq j\leq k} j^{k-j}$ we have by liberally using Stirling's formula,

$$\lim_{k\to\infty}\left(-\frac{1}{k^2}log(\prod_{1\leq j\leq k} j!) + \frac{log(k)}{2}\right) = \frac{1}{4} + \lim_{k\to\infty}\left(-\frac{1}{2k^2}log(\prod_{1\leq j\leq k} j^k) + \frac{log(k)}{2}\right)$$

$$= \frac{1}{4} + \lim_{k\to\infty}\left(-\frac{log(k!)}{2k} + \frac{log(k)}{2}\right) = \frac{3}{4}$$

□

We are now ready to state and prove the main result of this section.

Theorem 3.15. *Let X be a self adjoint random variable, and μ be its distribution. Then,*

$$\chi(X) = \int\int log|s-t|d\mu(s)d\mu(t) + \frac{3}{4} + \frac{log(2\pi)}{2}$$

We split the proof in two, (1) being the proof of LHS $\leq$ RHS, and (2) being the other direction.

Proof. (1): This direction is the relatively easier direction. Assume without loss of generality that $supp(\mu) \subseteq [-1,1]$ and let $A_k \in \mathfrak{M}_k^{sa}$ with $\|A_k\| \leq 1$ and

$$A_k \in \Gamma_1\left(X; N, k, \frac{w}{2}\right)$$

where N, w are like in the setting of lemma 3.14. As we have been seeking we get from applying the lemma,

$$\Gamma_2(X; N, k, \frac{w}{2}) \subset \Omega_k$$

, where Ω_k is the set

$$\Omega_k = \{B \in \mathfrak{M}_k^{sa} : Tr(B - UA_kU^*)^2 \leq k\epsilon^2\theta, \text{ for some } U \in \mathfrak{U}(\mathfrak{M}_k)\}$$

Now just using the estimates we found for this volume, and the deterministic result, we have

$$\chi(X; N, \frac{w}{2}) \leq \limsup_{k\to\infty}(k^{-2}log(\Omega_k) + log(k)/2)$$

$$\leq log(1+2\alpha)/2 + 2\epsilon + log(2\pi)/2 + 1/2\left(\int\int log((s-t)^2+\epsilon)d\mu(s)d\mu(t)\right) + 3/4$$

Hence, by taking limits across the appropriate nets, we have the result:

$$\chi(X) \leq \frac{log(2\pi)}{2} + 3/4 + \int\int log|s-t|d\mu(s)d\mu(t)$$

□

Proof. (2): First, we remark that it suffices to prove the inequality under the assumption that $supp(\mu) = [a, b]$ and μ has a well defined density ρ on $[a, b]$ which is positive and is in C^∞. Indeed, it suffices to prove the reduced statement, for if μ is a weak limit of probability measures μ_n, each having support in $[a, b]$ and density in C^∞ and

$$\lim_{n\to\infty} \int\int log|s-t|d\mu_n(s)d\mu_n(s) = \int\int log|s-t|d\mu(s)d\mu(t)$$

then our required inequality (for $-\mu$) follows from a direct application of Corollary 3.9.

Now, first we are reduced to finding such μ_n as above. This is seen as follows. Define $\tilde{\mu_n} = \mu * P_{1/n}$, where P_ϵ is the Poisson kernel. The logarithmic energies (given by our proposed formula for free entropy) of μ_n will converge to the logarithmic energy of μ, because of the observation that $P_\epsilon * log|\cdot| = log|\cdot + i\epsilon|$. Now we can then choose $\mu_n = (1+\epsilon_n)\phi(\tilde{\mu_n})$ where ϕ is the characteristic function of a sufficiently large interval that $[a, b]$ and ϵ_n converges down to 0.

Now, the proof is complete is we show the result for μ satisfying these additional constraints we have established. Define

$$a < a_1^{(k)} < b_1^{(k)} < a_2^{(k)} < \ldots < a_k^{(k)} < b_k^{(k)} = b$$

so that every part of this partition has μ- measure (integral of the density) $(2k)^{-1}$. From a continuity-type argument for the density, we can say that there is a $\delta > 0$ such that the lengths of the intervals of the partition are bigger than δ/k. Now, define the following set of matrices (which will nicely approximate X, and thereby becoming a subset of the microstate space):

$$\Xi_k = \{A \in \mathfrak{M} : a_j^{(k)} \leq \lambda_j(A) \leq b_j^{(k)}, 1 \leq j \leq k\}$$

where $\lambda_1(A) \leq \ldots \leq \lambda_n(A)$ are the eigenvalues of A. By squinting at the definition, one can spot that

$$\Xi_k \subset \Gamma(X; m_k, k, \epsilon_k)$$

for some $m_k \to \infty$ and $\epsilon_k \to 0$ as $k \to \infty$. Applying the limiting arguments, we have

$$\chi(X) \geq \limsup_{k\to\infty} \left(\frac{log\lambda(\Xi_k)}{k^2} + \frac{log(k)}{2} \right)$$

What now stands between us and the end of the proof is some routine manipulation and substitutions. From the bound we had established earlier on the size of a set such as Ξ_k, we have

$$\lambda(\Xi_k) \geq (2k)^{k(k-1)/2} \cdot \left(\prod_{1\leq j\leq k} j! \right)^{-1} \cdot (\delta k^{-1})^k \cdot \prod_{1\leq p<q\leq k} (a_q^{(k)} - b_p^{(k)})^2$$

Substituting this into the above inequality and applying the deterministic calculation we made in Lemma 3.15, we have

$$\chi(X) \geq \frac{2\pi}{2} + \frac{3}{4} + \limsup_{k\to\infty} k^{-2} log \prod_{1\leq p<q=k} (a_q^{(k)} - b_p^{(k)})^2$$

We aim to rewrite the above expression, especially the last limsup. Let $g : [0,1] \to [a,b]$ is the inverse of the function

$$t \mapsto \int_a^t d\mu(s)$$

Then, we have the last limsup in the above expression becomes

$$\limsup_{k\to\infty} k^{-2} \sum_{1\le p<q=k} log\big|g(2qk^{-1}) - g((2p+1)k^{-1})\big|^2$$

Now, with a bit of realignment, we have that the above expression is equal to

$$\int\int_{0\le s<t\le 1} log\big|g(t)-g(s)\big|^2 dsdt = \int_0^1\int_0^1 log\big|g(t)-g(s)\big|dsdt = \int\int log|s-t|d\mu(s)d\mu(t)$$

This finishes the proof. $\square$

3.4.2 Some computations

We compute the free entropy for the semicircular distribution, and thereby show that it has the highest entropy for a fixed variance. We also show that a distribution admits atoms only if the free entropy of the distribution is $-\infty$. To show that the other direction of the result is false, i.e, one can have an atomless distribution with free entropy being $-\infty$.

While the fact that full free entropy dimension is a strictly weaker condition than finite free entropy is very well known, an explicit example in the literature only appeared recently in [NCon]. Here is the example they constructed.

Example 1. Let $I_n \subset [0,1]$ be a disjoint sequence of intervals such that the Lebesgue measure $\lambda(I_n) < e^{-12^n}$, so that these interval thicknesses are rapidly decaying. Define a function f as follows:

$$f : \mathbb{R} \to \mathbb{R}_{\ge 0}$$
$$t \mapsto \sum_{n=1}^{\infty} 2^{-n}\frac{\chi_{I_n}(t)}{\lambda(I_n)}$$

By construction, f is non negative and integrable, and is also a probability density function. Let μ be the measure with density given by f. We claim that the (negative) logarithmic energy of μ is infinite. Indeed,

$$\int\int_{\mathbb{R}^2} \log|x-y|d\mu(x)d\mu(y) \le \sum_{n=1}^{\infty}\int\int_{I_n^2} \log|x-y|d\mu(x)d\mu(y)$$

$$\le \sum_{n=1}^{\infty} log(r^{-12^n})4^{-n} = -\infty$$

3.4.3 Additivity of Free Entropy

In [Voi91] Voiculescu proved an incredibly powerful result (which we have discussed here in Theorem 2.61) and this turns out to be the essential ingredient in this section. We are interested in showing that $\chi(X_1, \ldots, X_n) = \chi(X_1) + \ldots + \chi(X_n)$ if $X_1, \ldots, X_n$ are freely independent. Essentially, we want to be showing that the n-tuples of microstates of $X_1, X_2, \ldots, X_n$ individually, asymptotically become microstates of $X_1, \ldots, X_n$. But from Voiculescu's asymptotic freeness, we know that independent random matrix tuples become asymptotically free in the large N limit. It is therefore encouraging to pursue free additivity from this point of view.

We first fix up some notation. Denote by $U(k)$ the unitary group $\mathfrak{U}(\mathfrak{M}_k)$ and by μ the corresponding Haar probability measure. If $A \in \mathfrak{M}_k^{sa}$, let $O(A)$ denote its unitary orbit $\{UAU^* : U \in U(k)\}$. Also denote by $\Gamma_R(X_1, \ldots, X_n; Y_1, \ldots, Y_p; m, k, \epsilon)$ the set of $(A_1, \ldots, A_n; B_1, \ldots, B_p) \in (\mathfrak{M}_k^{sa})^n \times (\mathfrak{M}_k)^p$ such that monomials of these matrices approximate the corresponding monomials, i.e, τ of a *-monomial in $X_1, \ldots, Y_1, \ldots$ of degree $\leq m$ is within ϵ of τ_k of the corresponding *-monomial in $A_1, \ldots, B_1, \ldots$, and the operators norms are bounded by R. If we remove the R in the notation, just denote the set without the bounds on the uniform norms.

Now we record some variants of theorems in [Voi91].

Lemma 3.16. *Let $U_0, \ldots, U_n$ be unitary, with Haar distributions and *-free in (M, τ) and let $N \in \mathbb{N}$ and $\epsilon > 0$ be given. Further, define $W(k)$ to be a special unitary, which is the diagonal unitary $diag(exp(2\pi ij/k) : 0 \leq j \leq k-1)$. Let ω_k be the following set:*

$$\omega_k = \{(V_1, \ldots, V_n) \in (U(k))^n : (W(k), V_1, \ldots, V_n) \in \Gamma_1(U_0, \ldots, U_n; N, k, \epsilon)\}$$

Then, we have

$$\lim_{k\to\infty} \mu^{\otimes n}(\omega_k) = 1$$

Now, we modify this for a slightly stronger asymptotic freeness result:

Lemma 3.17. *Let $U_1, \ldots, U_n$ be unitaries with Haar distributions and $X = X^*$ with distribution as Lebesgue measure on $[0, 1]$ and assume $X, U_1, \ldots, U_n$ are *-free in (M, τ). Let further $N \in \mathbb{N}$ and $\epsilon > 0$ be given and let $h_j : [0, 1] \to \mathbb{R}$ $(1 \leq j \leq n)$ be continuous functions. As before, if $D_j(k) = diag(h_j(s/k) : 0 \leq s \leq k-1)$ and if*

$$\omega_k = \{(V_1, \ldots, V_n) \in U(k)^n : (D_1(k), \ldots, V_1, \ldots) \in \Gamma(h_1(x), \ldots; U_1, \ldots; N, k, \epsilon)\}$$

Then, we have

$$\lim_{k\to\infty} \mu^{\otimes n}(\omega_k) = 1$$

Proof. If $h_j(t) = P_j(exp(2\pi it))$ for Laurent polynomials P_j, the lemma follows immediately from the previous lemma (changing ϵ, N) (why?). The general case is then obtained by a routine approximation argument, approximating the h_j by $P_j(exp(2\pi it))$ in L^2 norm. $\square$

The following lemma is the crucial step in the proof of free additivity.

Lemma 3.18. *Let $U_1, \ldots, U_n, X$ be like in the previous lemma. Assume that $h_j : [0,1] \to \mathbb{R}$ are C^1 functions and $h_j'(t) > 0$ for all $t \in [0,1]$. Define further,*

$$\Omega(h_j; k) = \{A \in \mathfrak{M}_k^{sa} : h_j(2s/2k) \leq \lambda_{s+1}(A) \leq h_j((2s+1)/2k), 0 \leq s \leq k-1\}$$

where $\lambda_1(A) \leq \ldots \leq \lambda_k(A)$ are the eigenvalues of A. If

$$\Theta(k) = \{(A_1, \ldots, A_n) \in \prod_{1 \leq j \leq n} \Omega(h_j; k) : (A_1, \ldots, A_n) \in \Gamma(U_1 h_1(X) U_1^*, \ldots, U_n h_n(X) U_n^*); N, k, \epsilon\}$$

Then,

$$\lim_{k \to \infty} \frac{\lambda(\Theta(k))}{\lambda(\Omega(h_1; k) \times \ldots \times \Omega(h_n; k))} = 1$$

Proof. Let ω_k be the set defined in lemma 3.18, except with N_1, ϵ_1 replacing N, ϵ. Firstly, remark the following: If $A \in \mathfrak{M}_k^{sa}$ is such that $h_j(2s/2k) \leq \lambda_j(A) \leq h_j((2s+1)/2k)$ and if $D_j = diag(h_j(s/k) : 0 \leq s \leq k-1)$ then there is a unitary U, and a C independent of k such that

$$\left|A - U D_j U^*\right|_2 \leq \frac{C}{k}$$

Now, if we define the set of diagonal matrices with such restricted eigenvalues:

$$\Delta_k = \{diag(\lambda_1^{(j)}, \ldots, \lambda_k^{(j)})_{1 \leq j \leq n} :$$

$$h_j(s/k) \leq \lambda_{s+1}^{(j)} \leq h_j((2s+1)/2k), 1 \leq j \leq n, 0 \leq s \leq k-1\}$$

, then we simply have that the set $\prod_{1 \leq j \leq n} \Omega(h_j, k)$ is the disjoint union of products of the unitary orbits of n-tuples in Δ_k, i.e, $\prod_{1 \leq j \leq n} O(T_j)$, with $(T_1, \ldots, T_n) \in \Delta_k$. Now that we have described this disintegration, the lemma will follow if we prove that there is η_k a lower bound for the ratio of the volume of $(\prod_{1 \leq j \leq n} O(T_j)) \cap \Theta(k)$ to the volume of $\prod_{1 \leq j \leq n} O(T_j)$, uniform in $(T_1, \ldots, T_n) \in \Delta_k$ such that $\eta_k \to 1$ as $k \to \infty$.

Now suppose m is the map $m((V_1, \ldots, V_n)) = (V_1 T_1 V_1^*, \ldots, V_n T_n V_n^*)$. Observe that m is $(U(k))^n$-equivariant and $(\mathfrak{U}(k))^n$ and $\prod_{1 \leq j \leq n} O(T_j)$ have unique $(U(k))^n$ invariant probability measures. It is easy to see that

$$m(\omega_k) \subset \left(\prod_{1 \leq j \leq n} O(T_j)\right) \cap \Theta(k)$$

We now set

$$\eta_k = \mu^{\otimes n}(\omega_k) \leq \mu^{\otimes n}(m^{-1}(m(\omega_k))) = (m_* \mu^{\otimes n})(m(\omega_k))$$

Now, from the strengthened asymptotic freeness result, we have the conclusion directly follows. $\square$

We are now ready to prove the main result of this section.

Theorem 3.19. *If the self adjoint random variables $X_1, \ldots, X_n$ are free, then*

$$\chi(X_1, \ldots, X_n) = \chi(X_1) + \ldots + \chi(X_n)$$

Proof. Firstly, note that it suffices to prove the $\geq$ inequality, since we already established the other direction for any general $n-tuple$. One can also say the above statement for operators $X_1, \ldots, X_n$ such that there are free $X_1^{(m)}, \ldots, X_n^{(m)}$ such that $\lim_{m\to\infty} \chi(X_j^{(m)}) = \chi(X_j)$ and the distribution of $X_j^{(m)}$ converges to the distribution of X_j.

We can reduce the proof to the case when the measure μ_j which is the distribution of X_j, has support an interval $[a, b]$ and has a C^∞ density on $[a, b]$ which is > 0 on $[a, b]$. We can also assume additionally that with $X, U_1, \ldots, U_n$ like in the previous lemma,

$$X_j = U_j h_j(X) U_j^*$$

where $h_j : [0, 1] \to \mathbb{R}$ is C^∞ and $h_j'(t) > 0$ for all $t \in [0, 1]$. From the bound on the ratio we had from the previous lemma, we have

$$\lim_{k\to\infty} (log\lambda(\Theta(k)) - log\lambda(\Omega(h_1, k) \times \ldots \times \Omega(h_n, k))) = 0$$

Now applying the routine calculations,

$$\chi(X_1, \ldots, X_n; N, \epsilon) \geq \sum_{1\leq j\leq} \limsup_{k\to\infty} (k^{-2} log\lambda(\Omega(h_j, k)) + log(k)/2)$$

From the calculation we made previously in the section on computing free entropy of a single variable, we have

$$\limsup_{k\to\infty} (k^{-2} log\lambda(\Omega(h_j, k)) + log(k)/2) = \chi(h_j(X)) = \chi(X_j)$$

Therefore, we have the desired inequality:

$$\chi(X_1, \ldots, X_n; N, \epsilon) \geq \chi(X_1) + \ldots + \chi(X_n)$$

and thus,

$$\chi(X_1, \ldots, X_n) \geq \chi(X_1) + \ldots + \chi(X_n)$$

□

3.4.4 Non commutative power series transformations

We have seen so far how to compute free entropy for a single variable, how to compute free entropy for multiple variables in the case that they are free. But what if they are not free? This situation is quite intractable because it is closely tied with the Connes' embedding problem, which is far from solved. In this section we consider the effect of analytic noncommutative functional calculus on free entropy. This should be viewed as an infinitesimal change of variables formula. This is a very potent result, as it can be used as a tool to compute free entropy. Given that we know the free entropy of n free semicirculars, we can easily find using this formula, the free entropy of a nice large subset of the von Neumann algebra generated by the free semicirculars.

We begin by discussing some structure that is required to define a non commutative Jacobian, and an analytical functional calculus.

Let $M \otimes_\pi M^{op}$ be the projective tensor product of M with its opposite algebra (with its natural involutive Banach-algebra structure). Let $\alpha : M \otimes_\pi M^{op} \to B(M)$ be the contractive homomorphism given by

$$\alpha(a \otimes b) = L_a R_b$$

where L_a and R_b are the left multiplication and right multiplication operators. We shall denote by $LR(M)$ the algebra $\alpha(M \otimes_\pi M^{op})$. We have the following inequality,

$$\left|\alpha(x)m\right|_p = \|x\|_\pi |m|_p$$

for $1 \leq p \leq \infty$, hence, $LR(M)$ acts boundedly on $L^p(M)$. For $p = 2$, we have the map

$$\beta : LR(M) \to C^*(M, M')$$

where M and M' are with respect to the standard form of M on $L^2(M)$.Furthermore, let γ be the *-homomorphism

$$\gamma : C^*(M, M') \to M \overline{\otimes} M^{op}$$

Note also that $M \otimes M^{op}$ is a finite factor with the cannonical trace $\tau \otimes \tau$.

Now that we have laid out the foundations, let us define a noncommutative determinant called the Kadison Fuglede determinant. Consider a typical element $T \in \mathfrak{M}_n \otimes LR(M)$. If T is an invertible element, define its positive determinant $|det|$ to be the Kadison Fuglede determinant of $(id_n \otimes (\gamma \circ \beta))(T)$ in $\mathfrak{M}_n \otimes (M \otimes M^{op})$. That is, we have

$$|det|(T) = exp(\frac{Tr_n \otimes (\tau \otimes \tau)}{2})log((id_n \otimes (\gamma \circ \beta))(T^*T))$$

We leave it to the reader to verify that $|det|(I) = 1$ and $|det|(T_1T_2) = |det|(T_1)|det|(T_2)$. Note that the chosen normalization $(\tau \otimes Tr_n)(I) = n$ so $\tau \otimes Tr_n$ is not a state if $n \geq 2$.

Now we introduce the notion of a non commutative power series with a multiradius of convergence. Let $t_1, t_2, \ldots, t_n$ be non commuting indeterminates and let

$$F(t_1, \ldots, t_n) = \sum_{k=0}^{\infty} \sum_{1 \leq i_1 \ldots i_k \leq n} c_{i_1 \ldots i_k} t_{i_1} \ldots t_{i_k}$$

be a non-commutative power series with complex coefficients. For $R_j \geq 0$, $(1 \leq j \leq n)$ are real numbers, we say that $(R_1, \ldots, R_n)$ is a multiradius of convergence of F if

$$\sum_{k=0}^{\infty} \sum_{1 \leq i_1, \ldots, i_k \leq n} \left|c_{i_1 \ldots i_k}\right| R_{i_1} \ldots R_{i_k} < \infty$$

For convenience denote this number by $M(F;R_1,\ldots,R_n)$. Suppose quite generally, $X_1,\ldots,X_n$ are elements of a Banach algebra and $(\|X_1\|,\ldots,\|X_n\|)$ is a multiradius of convergence, then it easy to see that $F(X_1,\ldots,X_n)$ is well defined. Also clearly,

$$M(F;\|X_1\|,\ldots,\|X_n\|) \geq \big\|F(X_1,\ldots,X_n)\big\|$$

Now, if $(R_1,\ldots,R_n)$ is a multiradius of convergence of F then the map taking $(X_1,\ldots,X_n)$ to $F(X_1,\ldots,X_n)$ is an analytic function on

$$\prod_{1\leq j\leq n} \{X_j \in M : \big\|X_j\big\| < R_j\}$$

with values in M. By abusing notation, denote this analytic function by F. Then the 'differential' of F at $(X_1,\ldots,X_n)$ is defined to be

$$DF(X_1,\ldots,X_n) = (D_1F(X_1,\ldots,X_n),\ldots,D_jF(X_1\ldots,X_n)\ldots,D_nF(X_1,\ldots,X_n))$$

where the $D_jF(X_1,\ldots,X_n)$ is given by

$$D_jF(X_1,\ldots,X_n) = \sum_{k=1}^{\infty} \sum_{1\leq i_1,\ldots,i_k\leq n} c_{i_1\ldots i_k} \sum_{i_s=j} L_{X_{i_1}\ldots X_{i_{s-1}}} R_{X_{i_{s+1}}\ldots X_{i_n}}$$

Note that by our definition, $D_j(F(X_1,\ldots,X_n)) \in LR(M)$. Now, more generally, suppose $F_1,\ldots,F_n$ are non-commutative power series and $(R_1,\ldots,R_n)$ is a common multiradius of convergence we get an analytic function F which maps

$$M^n \supset \prod_{j=1}^{n} \{X_j \in M : \big\|X_j\big\| < R_j\} \mapsto M^n$$

where $F(X_1,\ldots,X_n) = (F_1(X_1,\ldots,X_n),\ldots,F_n(X_1,\ldots,X_n))$. The differential $DF(X_1,\ldots,X_n)$ is an element of $LR(M)\otimes\mathfrak{M}_k$, and we can write $DF = \big(D_iF_j\big)_{1\leq i,j\leq n}$. The "positive Jacobian" of F at $(X_1,\ldots,X_n)$ is then defined by

$$|\mathcal{J}|(F)(X_1,\ldots,X_n) = |det|(DF(X_1,\ldots,X_n))$$

This is very analogous to the classical Jacobian as we clearly see. In fact, if $G = (G_1,\ldots,G_n)$ is another such analytic map and $(M(F_1;R_1,\ldots,R_n),\ldots,M(F_n;R_1,\ldots,R_n))$ is a common multiradius of convergence for $G_1,\ldots,G_n$, then the composition $G\circ F$ is well defined and we leave it to the reader to verify the chain rule:

$$|\mathcal{J}|(G\circ F)(X_1,\ldots,X_n) = |\mathcal{J}|(G)(F(X_1,\ldots,X_n))\cdot|\mathcal{J}|(F)(X_1,\ldots,X_n)$$

. One more small thing to note before we proceed is that we would like to have $F(X_1,\ldots,X_n)$ to be self adjoint (because we're only dealing with self adjoint operators), and so we define the power series F^* by

$$F^*(t_1,\ldots,t_n) = \sum_{k=0}^{\infty} \sum_{1\leq i_1\ldots i_k\leq n} \overline{c_{i_1\ldots i_k}} t_{i_k}\ldots t_{i_1}$$

so that if X_j are self adjoint,

$$F^*(X_1,\dots,X_n) = (F(X_1,\dots,X_n))^*$$

In particular, if $F = F^*$, then $F(X_1,\dots,X_n)$ is self adjoint. We are now ready to address the main theorem of this section.

3.4.5 Infinitesimal change of variables formula

We present a formula for the free entropy of a transformation of the n- self adjoint random variables. However, this transformation satisfies some constraints as we shall see. It admits noncommutative power series of $X_1,\dots,X_n$ that are, as noncommutative polynomials in n variables, invertible, in an appropriate sense, and "stable" under small perturbations.

Proposition 3.20. *Let X_j $(1 \le j \le n)$ be self adjoint random variables in (M,τ). Let $F_1,\dots,F_n$ and $G_1,\dots,G_n$ be self adjoint noncommutative power series, in the sense that we have seen before. Let $(\|X_1\|+\epsilon,\dots,\|X_n\|+\epsilon)$ is a multiradius of convergence for the $F_j's$ for some $\epsilon > 0$ and*

$$\big(M(F_1;\|X_1\|+\epsilon,\dots,\|X_n\|+\epsilon),\dots,M(F_n:\|X_1\|+\epsilon,\dots,\|X_n\|+\epsilon)\big)$$

is a multiradius of convergence for the G_j's. Assume moreover the left invertibility condition:

$$G_j(F_1(t_1,\dots,t_n),\dots,F_n(t_1,\dots,t_n)) = t_j \quad (1 \le j \le n)$$

Then, we have

$$\chi(F_1(X_1\dots,X_n),\dots,F_n(X_1,\dots,X_n)) \ge log|\mathcal{J}|(F_1,\dots,F_n)(X_1,\dots,X_n) + \chi(X_1,\dots,X_n)$$

Proof. Firstly, fix $\|X_j\| < R_j < \|X_j\|+\epsilon$ and $M(F_j;\|X_1\|+\epsilon,\dots,\|X_n\|+\epsilon) > \rho_j > M(F_j;R_1,\dots,R_n)$. From this information, we immediately get the following: given m,ϵ there are some $m_1 \ge m$ and $0 < \epsilon_1 < \epsilon$ such that

$$(A_1,\dots,A_n) \to F(A_1,\dots,A_n) = \big(F_1(A_1,\dots,A_n),\dots,F_n(A_1,\dots,A_n)\big)$$

will provide an embedding of $\Gamma_{R_1,\dots,R_n}(X_1,\dots,X_n;m_1,k,\epsilon_1)$ into

$$\Gamma_{\rho_1,\dots,\rho_n}(F_1(X_1,\dots,X_n),\dots,F_n(X_1,\dots,X_n);m,k,\epsilon)$$

Note also that the map $G = (G_1,\dots,G_n)$ has a multiradius of convergence strictly larger than $(\rho_1,\dots,\rho_n)$. Recall that in the classical case, if f is holomorphic on an an open ball, it's derivative is also holomorphic on the same interior. By an easy analogue of this statement (which we leave to the reader to write down and check), we have $\|DG(A_1,\dots,A_n)\| \le C$ whenever $\|A_j\| \le \rho_j$, where C is independent of k. Note that in the above statement, the norm of the differential is the norm of the operator on $L^2(\mathfrak{M}_k,\tau_k)$. Similarly, we have $\|DF(A_1,\dots,A_n)\| \le C$ whenever $DG(F_1(A_1,\dots,A_n),\dots,F_n(A_1,\dots,A_n))DF(A_1,\dots,A_n) = Id_{(\mathfrak{M}_k)^n}$. From the above two statements, we have that the spectrum of $(DF(A_1,\dots,A_n))^*(DF(A_1,\dots,A_n))$ is contained in $[C^{-2},C^2]$.

Our strategy from this point onwards is to deal with $DF(A_1, \ldots, A_n)$. Our usage of the left invertibility condition has provided us with a bound on the spectrum. Since the log of a differential operator is almost always hard to deal with, we first approximate it with a polynomial. Given $\delta > 0$, there is an $n \times n$ matrix P with noncommutative polynomial entries such that

$$\left\| DF(A_1, \ldots, A_n) - P(L_{A_1}, \ldots, L_{A_n}, R_1, \ldots, R_{A_n}) \right\| < \delta$$

Now, if $C^{-1} - \delta > 0$, choosing a polynomial $Q(t)$ such that $\left|Q(t) - log(t)\right| < \delta$ on $[(C^{-1} - \delta)^2, (C + \delta)^2]$ we have the following inequality:

$$\left\| log((DF(A_1, \ldots, A_n))^* DF(A_1, \ldots, A_n)) - Q((P(L_{A_1}, \ldots, R_{A_n}))^* P(L_{A_1}, \ldots, R_{A_n})) \right\|$$
$$\leq \delta + 2\delta(C + \delta)(C^{-1} - \delta)^{-2}$$

Adding and subtracting $log((P(L_{A_1}, \ldots, R_{A_n}))^* P(L_{A_1}, \ldots, R_{A_n}))$ inside the norm, and applying functional calculus, we get the δ term. The other term is got by applying mean value theorem (the idea is that if $0 \leq X \leq Y, \implies log(X) \leq log(Y)$) (Note that the exactness of these soft estimates doesn't matter too much to the proof, as they will disappear in the limit). Now, observe in the formula for the Kadison-Fuglede determinant,

$$(Tr_n \otimes \tau_{k^2})(id_n \otimes (\gamma \circ \beta))(Q(P(L_{A_1}, \ldots, R_{A_n})))$$
$$= (Tr_n \otimes \tau_k \otimes \tau_k)(Q(P(A_1 \otimes I, \ldots, A_n \otimes I, I \otimes A_1, \ldots, I \otimes A_n)))$$

is a polynomial of degree ≤ 2 in the noncommutative moments $\tau_k(A_{t_1}, \ldots, A_{t_m})$. This is very useful for us, because being a microstate precisely means that approximation occurs at such a (polynomial) level. Note also that the we can replace the A_i's with X_i in (M, τ) and the same facts will hold true. Now, choosing ϵ_1 sufficiently small, and m_1 sufficiently large, if $(A_1, \ldots, A_n) \in \Gamma_{R_1,\ldots,R_n}(X_1, \ldots, X_n; m_1, k, \epsilon_1)$ then,

$$(Tr_n \otimes \tau_k \otimes \tau_k)(id_n \otimes (\gamma \circ \beta))(DF(A_1, \ldots, A_n)^* DF(A_1, \ldots, A_n))$$
$$-(Tr_n \otimes \tau \otimes \tau)(id_n \otimes (\gamma \circ \beta))(DF(X_1, \ldots, X_n)^* DF(X_1, \ldots, X_n))$$
$$\leq 3\delta + 4\delta(C + \delta)(C^{-1} - \delta)^{-2}$$

for all $k \in \mathbb{N}$. On the other hand, we have from the definition of the Jacobian on the transformation on the microstate space,

$$\chi_{R_1,\ldots,R_n}(X_1, \ldots, X_n; m_1, k, \epsilon_1) + inf\{log|\mathcal{J}|(F)(A_1, \ldots, A_n) :$$
$$(A_1, \ldots, A_n) \in \Gamma_{R_1,\ldots,R_n}(X_1, \ldots, X_n; m_1, k_1, \epsilon_1)\}$$
$$\leq \chi_{\rho_1,\ldots,\rho_n}(F_1(X_1, \ldots, X_n), \ldots, F_n(X_1, \ldots, X_n); m, k, \epsilon)$$

From what we have derived, we can replace the Jacobian term with what we require, as follows: Given $m \in \mathbb{N}$, $\epsilon > 0$ and $\eta > 0$ there are $m_1 \in \mathbb{N}$, $0 < \epsilon_1 < \epsilon$ such that

$$\chi_{R_1,\ldots,R_n}(X_1, \ldots, X_n; m_1, k, \epsilon_1) + log|\mathcal{J}|(F)(X_1, \ldots, X_n) - \eta$$
$$\leq \chi_{\rho_1,\ldots,\rho_n}(F_1(X_1, \ldots, X_n), \ldots, F_n(X_1, \ldots, X_n); m, k, \epsilon)$$

for all $k \in \mathbb{N}$. Routinely removing the m, k, ϵ, we have the required result. $\square$

Theorem 3.21. *Let the setup be as defined in the previous proposition. If moreover,*

$$(M(G_1; \|F_1(X_1, \ldots, X_n)\| + \epsilon, \ldots, \|F_n(X_1, \ldots, X_n)\| + \epsilon), \ldots,$$
$$M(F_n; \|F_1(X_1, \ldots, X_n)\| + \epsilon, \ldots, \|F_n(X_1, \ldots, X_n)\| + \epsilon))$$

is a multiradius of convergence for the F_j's, then

$$\chi(F_1(X_1, \ldots, X_n), \ldots, F_n(X_1, \ldots, X_n)) = log|\mathcal{J}|((F_1, \ldots, F_n))(X_1, \ldots, X_n) + \chi(X_1, \ldots, X_n)$$

Proof. Note that in order to get the equality, from the previous proposition, it suffices to show that the assumptions imply

$$F_j(G_1(t_1, \ldots, t_n), \ldots, G_n(t_1, \ldots, t_n)) = t_j$$

Note that DF at 0 in M^n, is a scalar $n \times n$ matrix and therefore $G \circ F = Id$ implies that $DF(0)$ is invertible. By the inverse function theorem, we infer that $F \circ G = id$ in a neighborhood of $F(0)$. By analyticity of $F \circ G$, we have $F \circ G = id$ on $\{(T_1, \ldots, T_n) \in M^n : \|T_j\| < \rho_j, 1 \leq j \leq n\}$. Choosing M to be a sufficiently large implies that $F \circ G = id$ at the level of noncommutative power series. □

3.4.6 Examples and corollaries

The formula that we derived in the above section has several important applications. First of all, it provides a source for numerous examples in free entropy. It is also noteworthy to remark here that such a transformation formula is not known to hold in the non-microstates free entropy case [Voi98a], which is also called infinitesimal free entropy. This non-microstates free entropy was constructed using a free analogue of the Fisher's information measure, and is arguably the more friendlier definition of free entropy today. It is unfortunate that such transformational formulas are still a mystery here.

Proposition 3.22.

1. *If $c_1, \ldots, c_n \in \mathbb{R}$ then $\chi(X_1 + c_1 I, \ldots, X_n + c_n I) = \chi(X_1, \ldots, X_n)$*
2. *If $A = (a_{ij})_{1 \leq i,j \leq n}$ is an invertible real $n \times n$ matrix, then*
$$\chi\left(\sum_j a_{1j} X_j, \ldots, \sum_j a_{nj} X_j\right) = \chi(X_1, \ldots, X_n) + log\left|det(A)\right|$$
3. *If $X_1, \ldots, X_n$ are linearly dependent, then $\chi(X_1, \ldots, X_n) = -\infty$.*

Proof. Recall that the inverse of an affine transformation is also an affine transformation. Hence, the transformations from (1) and (2) have inverses, and satisfy the conditions of the theorem. Therefore we have the results. For part (3), suppose $X_1, \ldots, X_n$ are linearly dependent, then there is $A \in GL(m, \mathbb{R})$ with $\left|det(A)\right| < 1$ so that

$$\sum_j a_{kj} X_j = X_k \quad (1 \leq k \leq n)$$

Now applying (b) to this, we have $\chi(X_1, \ldots, X_n) = \pm\infty$. But since free entropy is always $< \infty$, we have the result. □

Lemma 3.23. *Let $X_1, \dots, X_n$, $Y_1, \dots, Y_n$ be self adjoint random variables in (M, τ) and assume that $Y_1 = X_1$, $Y_2 = X_2 + P_2(X_1)$, …, $Y_j = X_j + P_j(X_1, \dots, X_{j-1})$, …, $Y_n = X_n + P_n(X_1, \dots, X_{n-1})$ where $P_2, \dots, P_n$ are noncommutative polynomials. Then $\chi(X_1, \dots, X_n) = \chi(Y_1, \dots, Y_n)$.*

Proof. Firstly remark that by some algebraic manipulation, one can derive polynomials R_j, such that $X_j = Y_j + R_j(Y_1, \dots, Y_{j-1})$. Assume that $P_j = P_j^*$ and $R_j = R_j^*$. Defining $F_j(X_1, \dots, X_n) = X_j + P_j(X_1, \dots, X_n)$ and $G_j(Y_1, \dots, Y_n) = Y_j + R_j(Y_1, \dots, Y_{j-1})$ we may apply the main theorem to get $\chi(X_1, \dots, X_n) \leq \chi(Y_1, \dots, Y_n)$. The reverse inequality follows by symmetry, i.e, replacing the roles of X_j and Y_j. $\square$

The following proposition offers a more generalized version of the previous lemma.

Proposition 3.24. *Let $X_1, \dots, X_n$, $Y_1, \dots, Y_n$ be self adjoint random variables in (M, τ) and assume $X_1 = Y_1$ and $Y_j - X_j \in W^*(X_1, \dots, X_{j-1})$ if $2 \leq j \leq n$. Then $\chi(X_1, \dots, X_n) = \chi(Y_1, \dots, Y_n)$.*

Proof. Like in the previous lemma, there is some symmetry going on between the X's and the Y's, sicne the assumptions imply that $X_j - Y_j \in W^*(Y_1, \dots, Y_{j-1})$, $(2 \leq j \leq n)$. It will suffice to then show that $\chi(X_1, \dots, X_n) \leq \chi(Y_1, \dots, Y_n)$. Let $P_j^{(t)}(X_1, \dots, X_{j-1})$, $(t \in \mathbb{N}, 2 \leq j \leq n)$ be noncommutative polynomials, such that

$$X_j + P_j^{(t)}(X_1, \dots, X_{j-1}) \to Y_j$$

strongly, as $t \to \infty$. By the previous lemma, and an old result Prop 3.8, we have

$$\begin{aligned} \chi(Y_1, \dots, Y_n) &\geq \limsup_{t \to \infty} \chi(X_1, X_2 + P_2^{(t)}(X_1), \dots, P_n^{(t)}(X_1, \dots, X_{n-1})) \\ &= \chi(X_1, \dots, X_n) \end{aligned}$$

Hence proved. $\square$

We end this section with a useful proposition that is in the same flavor as above.

Proposition 3.25. *Let $Y_1, \dots, Y_{m+n}$ be self adjoint free random variables so that $\chi(Y_j) > -\infty$, $1 \leq j \leq m + n$. Let further $X_j = X_j^* \in W^*(Y_{m+1}, \dots, Y_{m+n})$, $1 \leq j \leq m$. Then*

$$\chi(X_1 + Y_1, \dots, X_n + Y_n) \geq \chi(Y_1, \dots, Y_n)$$

Proof. We have

$$\chi(X_1 + Y_1, \dots, X_n + Y_n) + \chi(Y_{m+1}, \dots, Y_{m+n}) \geq \chi(X_1 + Y_1, \dots, X_m + Y_m, Y_{m+1}, \dots, Y_{m+n})$$

But from the above proposition, we have

$$\chi(X_1 + Y_1, \dots, X_m + Y_m, Y_{m+1}, \dots, Y_{m+n}) - \chi(Y_{m+1}, \dots, Y_{m+n}) =$$
$$\chi(Y_1, \dots, Y_m, Y_{m+1}, \dots, Y_{m+n}) - \chi(Y_{m+1}, \dots, Y_{m+n})$$

And now from the mutual freeness of the Y_i's, we have simply the above expression is equal to

$$\sum \chi(Y_i) = \chi(Y_1, \dots, Y_n)$$

which concludes our result. $\square$

These above results indicate to us that free entropy is a quantity that speaks more about the von Neumann subalgebra generated by the set, and not plainly about the set itself. This agrees more with the intuition of treating free entropy as a measure quantity. (Next when we define free entropy dimension, it makes more sense to define such a quantity for an algebra, rather than a set of random variables.)

3.5 Some remarks on free entropy

Here we do not discuss the "non-microstates free entropy" which is greatly discussed in Voiculescu's 6th paper on Fisher's Information [Voi98a]. This is a kind of infinitesimal free entropy derived from a non commutative analogue of the fisher's information measure. A very important open problem in the literature now is to decide if microstates free entropy is equal to the non microstates free entropy. In an important paper [BCG03] Biane, Capitaine and Guionnet prove that the microstates free entropy is bounded from above by the non microstates free entropy. This in particular has several consequences. Most importantly, upper bounds on the non microstates case (which are usually more easy to find) can easily translate to the microstates case.

On another note, many have worked on tweaking the definitions of the above microstates free entropy, which already seems too "magical", in the sense that everything seems to work despite having to go through such a complicated construction. Yoann Dabrowski has recently posted a paper on arxiv [Dab16] starting a program to prove that one can replace the limsup with liminf in the definition of free entropy. He has not completely proved it but has settled a few cases. Belinschi and Berkovici showed in 2003 [BB03] that one can stop worrying about bounding the operator norm in the definition, i.e, the quantity χ_∞ defined exactly like χ, except ignoring the bound R on the operator norm of the matrices, is equal to the quantity χ.

4 Free Entropy Dimension (Caution: Some edits pending in this section

Voiculescu in [Voi94] introduces free entropy dimension as a coarser measure of the information content of an n-tuple of self adjoint random variables. The real reason for defining this quantity was to draw a parallel between the relationship of Lebesgue measure (vol) with the Minkowski dimension. The Minkowski dimension of a subset of $\mathbb{R}^n$ reveals the so called "intrinsic" dimension of the subset, and therefore, an analogous quantity defined on an n-tuple of self adjoint random variables could perhaps shed light on the von Neumann algebra they generate, in particular, the potential of it to be a free group factor on m-generates. It is many people's hopes and dreams that this analogue (free entropy dimension) would be a von Neumann algebra invariant, i.e, if $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ generate the same von Neumann algebra, then their free entropy dimensions align. This statement, if it were true, would actually resolve the free group factor isomorphism problem, as we shall see. But it turns out this is a farfetched dream. In this section, it is shown that free entropy dimension is an algebra invariant. It is also shown that it is a noncommutative "smooth" functional calculus invariant. The notion of smoothness is defined and described, and the noncommutative power series we defined in the earlier section satisfies this criterion. However, it is possible to improve the result with slightly better algebras. In fact, in a later work of Kenley Jung [Jun02b], it is shown that it is in fact an invariant if the von Neumann algebra generated is hyperfinite. The techniques used in Jung's work will be discussed in this section as they are extremely useful in matters related to the microstates approach. Unfortunately, the general von Neumann algebra case is hopeless, even the C^* algebra case for that matter.

4.1 Motivation for defining Free Entropy Dimension

We studied the motivation for defining free entropy, very well. It was a kind of "lebesgue measure" quantity, precisely the limiting logarithm of the lebesgue measure of a subset of $\mathbb{R}^n$. Therefore, the focus is really on the "microstate space" of the operator, and free entropy gives us a measure of such a space. Ideally, the driving goal of free probability is arguably to learn more about the operator itself from it's distribution, and in particular in our case the microstate space. To just throw an example where the importance of this is seen, consider the Connes' embedding problem. It asks whether every type II_1 factor on a separable Hilbert space can be embedded into the ultrapower of the hyperfinite type II_1 factor by a free ultrafilter. This problem can be resolved if one settles the following problem. Suppose $X_1, \ldots, X_n$ are self adjoint random variables in (R, τ) where R is the hyperfinite II_1 factor, can $\chi(X_1, \ldots, X_n) > -\infty$ (obviously we want $n \geq 2$, else we can just take an atomic distribution). For results about the relation of this problem to other questions in operator algebra and Banach space theory, see [Kir93]. There has been a lot of progress in the area of Connes' embedding, but the main problem remains far from solved.

Voiculescu saw potential in the notion of Minkowski dimension which is a kind of fractal dimension, to produce a new quantity of use in free probability. [1] Before defining Minkowski dimension, consider the following example. Say $0 \leq d \leq n$ and E is a bounded subset of a d-dimensional subspace of $\mathbb{R}^n$. For instance let $E = B^d(0,1) \times \{0\}^{n-d}$, where $B^d(0,1)$ is the d-dimensional unit ball. Define E_δ to be the δ neighborhood of E, i,e, $\bigcup_{x \in E} B(x, \delta)$. From the triangle inequality, one has

$$B^d(0,1) \times B^{n-d}(0,\delta) \subseteq E_\delta \subseteq B^d(0,2) \times B^{n-d}(0,\delta)$$

for all $0 \leq \delta \leq 1$, with the appropriate identifications in $\mathbb{R}^n$. From this and the formula for n-dimensional volume of the unit ball, we have the existence of constants c, C such that

$$c\delta^{n-d} \leq vol_n(E_\delta) \leq C\delta^{n-d}$$

In particular, taking logarithms,

$$\lim_{\delta \to 0} \frac{log(vol_n(E_\delta))}{log(\delta)} = n - d$$

[1] The material about Minkowski dimension is along the lines of Terry Tao's blog post on Hausdorff dimension https://terrytao.wordpress.com/2009/05/19/245c-notes-5-hausdorff-dimension-optional/. The reader is advised to read it for a complete understanding of the topic

written in a better way,

$$\lim_{\delta \to 0} n - \frac{log(vol_n(E_\delta))}{log(\delta)} = d$$

This rather trivial exercise actually motivates us to define the following quantity:

Definition 4.1 (Minkowski dimension)**.** Define the upper Minkowski dimension to be

$$\overline{dim}(E) = \limsup_{\delta \to 0} n - \frac{log(vol_n(E_\delta))}{log(\delta)}$$

Define the lower Minkowski dimension to be

$$\underline{dim}(E) = \liminf_{\delta \to 0} n - \frac{log(vol_n(E_\delta))}{log(\delta)}$$

and if $\overline{dim}(E) = \underline{dim}(E)$, say the set E has Minkowski dimension $dim(E) = \overline{dim}(E) = \underline{dim}(E)$.

Observe that in line with the initial example discussed, Minkowski dimension in $\mathbb{R}^n$ reveals the "intrinsic" dimension of the $\mathbb{R}$-vector space the set is living in. To shed more light on this, we now provide some very illuminating examples.

Example 2. As a singularity, we see directly from applying the formula that $dim(\emptyset) = -\infty$.

Example 3. The unit ball in R^n has full Minkowski dimension. A collection of discrete points in R^n have null Minkowski dimension. We've also seen how subsets can have integral Minkowski dimension between 0 and n (as an exercise, one can prove that the Minkowski dimension of a nonempty subset has to be between this interval). Now, the question is about non integral Minkowski dimension. Let C be the cantor set whose elements are of the form $\sum_{i=1}^{\infty} a_i 4^{-i}$ where $a_i \in \{0, 3\}$. Then, the Minkowski dimension of this set is 1/2.

Example 4. It is easy to see that any bounded set $E \subset \mathbb{R}^n$ with $dim(E) < n$ has lebesgue measure (in $\mathbb{R}^n$) 0. In other words, every positive measured set has full Minkowski dimension. But it is not true that a null lebesgue measured set does not have full Minkowski dimension. We would like the reader to remember this example, as it would provide a very useful analogy. It is also a good exercise to prove the above two claims.

Example 5. This example provides a more concrete picture of a non integral Minkowski dimension set. Consider the topologists' sine curve, i.e, $f(x) = \sin(1/x)$. The graph of this function (a bounded segment of it including the unit square centered at the origin) has Minkowski dimension $3/2$.

Example 6 (Problem)**.** Let P be a polynomial in n variables, and let A be a 'nice' (?) bounded subset of the algebraic variety of this polynomial (the set of roots in $\mathbb{R}^n$). Then, the Minkowski dimension of A is at most $n - 1$.

Now, we do not get deeper into the study of Minkowski dimension, which involves studying Hausdorff measures and so on, although a deeper understanding of Minkowski dimension is sure to provide more insight on free entropy dimension. For our purposes, it is quite sufficient to be well aware of the above material. After Voiculescu provided his free entropy dimension, Jung found a more cleaner way to re derive it, and we shall be discussing this point of view too, mainly because of the great utility it has found in modern research in the area. Jung used some alternative formulations of Minkowski dimension as his foundation.

Definition 4.2. Packing numbers:

- $N_\delta^{ext}(E)$: The minimum number of δ balls with centers in $\mathbb{R}^n$ that cover the set E.
- $N_\delta^{int}(E)$: The minimum number of δ balls with centers in E that cover E.
- $N_\delta^{net}(E)$: (also called δ-metric entropy) cardinality of the largest δ net in E. (δ net just means a collection of sets each of whose diameter is less than δ).
- $N_\delta^{pack}(E)$: The largest number of δ balls that are disjoint, with centers in E.

On first observation, one can easily see that these are closely related to each other and to the volume, $vol_n(E_\delta)$.

The following lemma precisely relates these quantities.

Lemma 4.3. *For any bounded set $E \subseteq \mathbb{R}^n$ and any $\delta > 0$,*

$$N_{2\delta}^{net} = N_\delta^{pack} \leq \frac{vol_n(E_\delta)}{vol_n(B_n(0,\delta))} \leq 2^n N_\delta^{ext}$$

and

$$N_\delta^{ext} \leq N_\delta^{int} \leq N_\delta^{net}$$

The proof of this result is not tedious, but simply a routine application of triangle inequality. Consequently (from the above lemma), we have

$$\overline{dim}(E) = \limsup_{\delta \to 0} \frac{log(N_\delta^*(E))}{log(1/\delta)}$$

where $*$ is any of ext, int, net, pack. This alternative formula for Minkowski dimension is of a very different flavor. It involves more of a discrete/counting approach, rather than a geometric volume based approach. Some situations can benefit with the usage of the original definition, while some other can benefit with this alternative "packing" definition.

Remark 1. We want to remark here that the logarithm of the packing and covering numbers are sometimes referred to as entropy numbers, and are somewhat analogues to the concepts of thermodynamic entropy and information theoretic entropy, in that they measure the amount of disorder in the metric space or fractal at scale ϵ, and also measure how many bits or digits one would need to specify a point of the space to accuracy ϵ.

Remark 2. The first construction of Minkowski dimension we have seen involves the usage of notions of volume, which do not necessarily generalize to arbitrary metric spaces. However, the intrinsic notion of a Minkowski dimension does indeed generalize to the arbitrary metric space, and one can use the second approach we have provided to make sense of this, as there is no usage of volumes and measures.

4.2 Definition

The first definition (presented in [Voi94]) is the one we will primarily use in this paper. It is based on the first kind of Minkowski dimension we saw, using ϵ-neighborhood volumes. We instead compute the entropy of a small perturbation of the operator by a ϵ scaled semicircular.

Definition 4.4. Let $X_1, \ldots, X_n$ be self adjoint random variables in (M, τ). The free entropy dimension is defined by

$$\delta(X_1, \ldots, X_n) = n + \limsup_{\epsilon \to 0} \frac{\chi(X_1 + \epsilon S_1, \ldots, X_n + \epsilon S_n)}{\left|log(\epsilon)\right|}$$

where $\{S_1, \ldots, S_n\}$ is a semicircular family and the algebras generated by $\{X_1, \ldots X_n\}$ and $\{S_1, \ldots, S_n\}$ are free.

One sees immediately:

Remark 3. The above definition is well defined. In other words, it is independent of the particular choice of $S_1, \ldots, S_n$

Proof. This is true because the free entropy only cares about the distribution, and since the $S_i's$ are free, the distribution is given by the additive convolution. □

Subadditivity of δ is also immediate from the subadditivity of χ. Now we look at the alternate definition of Jung [Jun02a] using the packing formulation of Minkowski dimension. We will need to set up some notation to define a suitable packing number. Instead of the notation we have used let us instead say $K_\epsilon(X, d)$ be the covering number (which we called N_δ^{ext}) and $P_\epsilon(X, d)$ be the packing number (which we called N_δ^{pack}). Here we are working on (X, d), any metric space. In our situation, say (M, τ) is a non commutative probability space, $(\mathfrak{M}_k^{sa})^n$ is the space of n-tuples of self adjoint $k \times k$ matrices with tr_k is the normalized trace state. We have a metric ρ_k on $X = (\mathfrak{M}_k^{sa})^n$ induced by the normalized Hilbert Schmidt norm, $\frac{\|.\|_2}{k^{1/2}}$. We shall be using our packing numbers on this metric space (X, ρ_k).

Definition 4.5. For any $k, m \in \mathbb{N}$, $R, \gamma, \epsilon > 0$, define

$$\mathbb{P}_{\epsilon,R}(a_1, \ldots, a_n; m, k, \gamma) = P_\epsilon(\Gamma_R(a_1, \ldots, a_n; m, k, \epsilon), \rho_k)$$

$$\mathbb{P}_{\epsilon,R}(a_1, \ldots, a_n; m, \gamma) = \limsup_{k \to \infty} \frac{1}{k^2} log(\mathbb{P}_{\epsilon,R}(a_1, \ldots, a_n; m, k, \gamma))$$

$$\mathbb{P}_{\epsilon,R}(a_1, \ldots, a_n) = \inf_{m \in \mathbb{N}, \gamma > 0} \{\mathbb{P}_{\epsilon,R}(a_1, \ldots, a_n; m, \gamma)\}$$

$$\mathbb{P}_\epsilon(a_1, \ldots, a_n) = \sup_{R > 0} \{\mathbb{P}_{\epsilon,R}(a_1, \ldots, a_n)\}$$

Similarly define $\mathbb{K}_\epsilon(a_1, \dots, a_n)$ replacing the packing number by the covering number.

Proposition 4.6. *If $h_1, \dots h_n$ is a set of free semicirculars, freely independent from $a_1, \dots, a_n$, then*

$$\delta(a_1, \dots, a_n) = n + \limsup_{\epsilon \to 0} \frac{\chi(a_1 + \epsilon h_1, \dots, a_n + \epsilon h_n)}{\left|log(\epsilon)\right|} = \limsup_{\epsilon \to 0} \frac{\mathbb{K}_\epsilon(a_1, \dots, a_n)}{log(\epsilon)}$$

$$= \limsup_{\epsilon \to 0} \frac{\mathbb{P}_\epsilon(a_1, \dots, a_n)}{\left|log(\epsilon)\right|}$$

From the inequality

$$P_\epsilon(X, d) \geq K_{2\epsilon}(X, d) \geq P_{4\epsilon}(X, d)$$

it suffices to just prove the last equality. Now, in [Voi98b] Voiculescu provides a strengthened asymptotic freeness result which is an improvement of what we saw in section 2. Using this, Jung derives the left hand side inequality by finding bounds on the volume of the microstate space. The other direction is a rather easy argument. See [Jun02a] for a detailed proof.

Kenley Jung remarks in the end of his paper that in the case of freeness or in the case of a single random variable, the ϵ neighborhood definition is more convenient, while on the other hand the packing formulation is more fruitful when providing formulas for generators of M when M has a simple algebraic decomposition into a tensor product of a von Neumann algebra N with the $k \times k$ matrices or into a direct sum of algebras.

4.3 Computing Free Entropy Dimension

We have built some intuitive reasons as to why free entropy dimension should be an important quantity. Here we provide a formula for the free entropy dimension of one variable, and notice several consequences of this formula.

4.3.1 Formula for $\delta(X)$

Theorem 4.7. *Let X be a self adjoint random variable in (M, tau) and let μ be the spectral distribution of X. Then,*

$$\delta(X) = 1 - \sum_{t \in \mathbb{R}} (\mu(\{t\}))^2$$

Proof. This is quite a long proof. The goal of this proof would be to evaluate a certain integral. We already know the formula of free entropy, so we basically need to evaluate:

$$\lim_{\epsilon \to 0} \frac{\chi(X + \epsilon S)}{\left|log(\epsilon)\right|} = \frac{\int log|s - t| d\mu_\epsilon(s) d\mu_\epsilon(t) + \frac{3}{4} + \frac{log(2\pi)}{2}}{\left|log(\epsilon)\right|} = \frac{\int log|s - t| d\mu_\epsilon(s) d\mu_\epsilon(t)}{\left|log(\epsilon)\right|}$$

where S is a semicircular free from X, and μ_ϵ is the spectral distribution of the perturbation $X + \epsilon S$. First of all, if the distribution of such a perturbation has any atoms, then we run into a problem.

But the beauty of these small perturbations is the fact that the distribution is indeed absolutely continuous with respect to the lebesgue measure and its density ν_ϵ is in L^p $(1 \leq p \leq \infty)$ and

$$\|\nu_\epsilon\| \leq \epsilon^{-1+1/p} C$$

for a constant C independent of p and ϵ. The above is a consequence of a proposition (Prop 4.7) in Voiculescu's previous paper [Voi93]. Now, the above integral is problematic at the diagonal (on $\mathbb{R}^2$), which is the behavior we are mainly interested in, therefore we make a small change that we hope shouldn't affect our computation. We have

$$0 \leq \int\int \nu_\epsilon(s)\nu_\epsilon(t) log|s-t+i4\epsilon| dsdt - \int\int \nu_\epsilon(s)\nu_\epsilon(t) log|s-t| dsdt$$
$$= \int\int \nu_\epsilon(s)\nu_\epsilon(t) log\left|1 + i\frac{4\epsilon}{s-t}\right|$$

We show below that the above integral is bounded, thereby providing a valid alternative for our computation, as we wanted. Let $a > 0$ be a constant such that $a > 4(\|X\| + \|S\|)$, and thus satisfying $supp(\mu_\epsilon) \subset [-a/4, a/4]$ for $0 \leq \epsilon \leq 1$ (from the spectral theorem). Define the function

$$g_\epsilon(x) := log\left|1 + i\epsilon x^{-1}\right|$$

for $-a \leq x \leq a$. Then, from a basic computation, we know that $g_\epsilon \in L^p(-a,a)$ for $1 \leq p < \infty$ and

$$\|g_\epsilon\|_p \leq C\epsilon^{1/p}$$

Let

$$h_\epsilon(s) = \int log\left|1 + i\frac{4\epsilon}{s-t}\right| \nu_\epsilon(t) dt$$

From the previous estimates, we have a constant K such that

$$\|h_\epsilon\|_\infty \leq \|\nu_\epsilon\|_2 \|g_{4\epsilon}\|_2 \leq K$$

Hence, we have

$$\int\int \nu_\epsilon(s)\nu_\epsilon(t) log\left|1 + i\frac{4\epsilon}{s-t}\right| dsdt \leq \|\nu_\epsilon\|_1 \|h_\epsilon\|_\infty \leq K$$

so that

$$\lim_{\epsilon\to 0} \frac{\left|\int\int log|s-t+14\epsilon| d\mu_\epsilon(s) d\mu_\epsilon(t) - \int\int log|s-t| d\mu_\epsilon(s) d\mu_\epsilon(t)\right|}{\left|log(\epsilon)\right|} = 0$$

Hence it will suffice to prove that

$$\lim_{\epsilon\to 0} \frac{\int\int log|s-t+i4\epsilon| d\mu_\epsilon(s) d\mu_\epsilon(t)}{\left|log(\epsilon)\right|} = -\sum_{t\in\mathbb{R}} (\mu(\{t\}))^2$$

From this point on, we will need a crucial observation. After the observation, we will split our domain of the integral ($\mathbb{R}^2$) into two parts which are each easy to compute.

From the Courant minimax principle which gives the eigenvalues of a symmetric matrix, we see that upon normalizing S so that $\|S\| = 1$,

$$\tau(E(X + \epsilon S; (-\infty, b - \epsilon))) \leq \tau(E(X; (-\infty, b))) \leq \tau(E(X + \epsilon S; (-\infty, b + \epsilon)))$$

Hence

$$\mu_\epsilon(-\infty, b - \epsilon) \leq \mu(-\infty, b) \leq \mu_\epsilon(-\infty, b + \epsilon)$$

More generally, we have that if $b_1 \leq b_2$,

$$\mu_\epsilon([b_1, b_2]) \leq \mu([b_1 - \epsilon, b_2 + \epsilon])$$

$$\mu([b_1, b_2]) \leq \mu_\epsilon([b_1 - \epsilon, b_2 + \epsilon])$$

We now look at the region of $\mathbb{R}^3$ given by $|s - t| > \epsilon^\delta$ for some $0 < \delta < 1$ and $0 < \epsilon < 1$. Here,

$$\frac{\left|\int\int_{|s-t|>\epsilon^\delta} log|s - t + i4\epsilon| d\mu_\epsilon(s)\mu_\epsilon(t)\right|}{\left|log(\epsilon)\right|} \leq \frac{log(1 + a + 4\epsilon)}{\left|log(\epsilon)\right|} + \delta$$

therefore we get the bound

$$\limsup_{\epsilon \to 0} \frac{\left|\int\int_{|s-t|>\epsilon^\delta} log|s - t + i4\epsilon| d\mu_\epsilon(s) d\mu_\epsilon(t)\right|}{\left|log(\epsilon)\right|} \leq \delta$$

We now have a clue that the diagonal is the part of the integral that matters. Denote by $\Delta(a, b)$ the set $\{(s, t) \in \mathbb{R}^2 : a \leq |s - t| \leq b\}$ and by Δ denote the diagonal of $\mathbb{R}^2$. Observe that $(\mu \otimes \mu)(\Delta(0, r)) \downarrow (\mu \otimes \mu)(\Delta)$ if $r \downarrow 0$ and hence if $0 < r_1 < r_2$ and $r_2 \downarrow 0$, then from the previous computation,

$$(\mu \otimes \mu)(\Delta(r_1, r_2)) \downarrow 0$$

We now try to compute what happens on the diagonal.

$$\mu_\epsilon \otimes \mu_\epsilon(\Delta(0, r)) \leq (\mu \otimes \mu_\epsilon) \left(\bigcup_{n \in \mathbb{Z}} [(n-1)\epsilon - r, (n+2)\epsilon + r) \times [n\epsilon, (n+1)\epsilon) \right)$$

$$\leq (\mu \otimes \mu_\epsilon)(\Delta(0, r + \epsilon)) \leq (\mu \otimes \mu) \left(\bigcup_{n \in \mathbb{Z}} [n\epsilon, (n+1)\epsilon) \times [(n-2)\epsilon - r, (n + 3\epsilon + r)) \right)$$

$$\leq (\mu \otimes \mu)(\Delta(0, r + 2\epsilon))$$

Applying similar steps using the inequalities (we found from applying minimax principles), we have

$$(\mu \otimes \mu)(\Delta(0, r)) \leq (\mu_\epsilon \otimes \mu_\epsilon)(\Delta(0, r + 2\epsilon))$$

In particular this implies that if $2\epsilon < r_1(\epsilon) < r_2(\epsilon)$ and $r_2 \downarrow 0$, then $(\mu_\epsilon \otimes \mu_\epsilon)(\Delta(r_1(\epsilon), r_2(\epsilon))) \downarrow 0$. Also from the sandwiching $\mu \otimes \mu(\Delta(0, r))$ we observed earlier,

$$(\mu_\epsilon \otimes \mu_\epsilon)(\Delta(0, r(\epsilon))) \downarrow (\mu \otimes \mu)(\Delta)$$

Now, we get down to evaluating the interval. On $\Delta(4\epsilon, \epsilon^\delta)$, we have

$$\big|log|s - t + i4\epsilon|\big| \leq \big|log(\epsilon)\big|$$

and on $\Delta(0, 4\epsilon)$ we have $log|s - t + i4\epsilon| \in [log(4\epsilon), log(8\epsilon)]$, and hence

$$\frac{log|s - t + i4\epsilon|}{\big|log(\epsilon)\big|} \in \left[-1, -1 + \frac{log(8)}{\big|log(\epsilon)\big|}\right] \rightarrow_{\epsilon \rightarrow 0} \{-1\}$$

Now evaluating the integral,

$$\lim_{\epsilon \rightarrow 0} \int \int_{\Delta(0,4\epsilon)} log|s - t + i4\epsilon| d\mu_\epsilon(s) d\mu_\epsilon(t) = - \lim_{\epsilon \rightarrow 0} (\mu_\epsilon \otimes \mu_\epsilon)(\Delta(0, 4\epsilon)) = -(\mu \otimes \mu)(\Delta)$$

Thus, we have shown that

$$\limsup_{\epsilon \rightarrow 0} \left| \big|log(\epsilon)\big|^{-1} \int \int log|s - t + i4\epsilon| d\mu_\epsilon(s) d\mu_\epsilon(t) + (\mu \otimes \mu)(\Delta) \right| \leq \delta$$

and since δ was arbitrary, we have that

$$\delta(X) = 1 - \mu \otimes \mu(\Delta) = - \sum_{r \in \mathbb{R}} (\mu(\{t\})^2$$

Hence proved. ☐

4.3.2 For multiple variables

Just like in the situation of free entropy, one cannot provide a direct formula for the free entropy dimension of an $n - tuple$ of operators, unless of course in special case. This is partly because of the dependance of free entropy dimension on the actual free entropy and the lack of such formulae therein. Here, we prove some inequalities under less general circumstances.

Proposition 4.8. *Let $(X_1, \ldots, X_n)$ be a free family of self adjoint random variables in (M, τ) and let μ_j be the distribution of X_j. Then*

$$\delta(X_1, \ldots, X_n) = \delta(X_1) + \ldots + \delta(X_n)$$

Proof. Firstly, let a, b, c, d be random variables such that a and b are freely independent, and c and d are freely independent, then, $a + c$ and $b + d$ are freely independent. Hence, from the definition 4.4, we directly have the result. ☐

One also has the general inequality that works (whose equality is satisfied precisely when the variables are free).

Proposition 4.9. *For $X_1, X_2, \ldots, X_n$ as usual, we have*

$$\delta(X_1, \ldots, X_n) \leq \sum_{i=1}^{n} \delta(X_i)$$

Proof. This is straightforward from the subadditivity of free entropy. ☐

4.3.3 Significance and relevance of this quantity

Suppose a non commutative random variable X in (M, τ) has no atoms, i.e, it is diffuse, then through a functional transformation, one can obtain the semicircular distribution, and therefore the von Neumann algebra generated by X will be isomorphic to the free group factor. The isomorphism $W^*(s) \cong LF_1 = L(\mathbb{Z})$ follows from the fact that the pontryagin dual of $\mathbb{Z}$ is indeed $\mathbb{T}$ which is isomorphic as measure space to the spectrum of the semicircular operator, $[0, 1]$. Therefore, the absence of atoms is precisely the necessary and sufficient condition for an operator to generate a free group factor. This absence of atoms is particularly characterized, as we have seen, by possessing full free entropy dimension. In a more appetizing language, we can now say in the one variable case, that the free entropy dimension gives us a free-group factoresque dimension of the von Neumann algebra generated by X. Ofcourse, we will be discussing this in more rigorous terms in the next section.

It is important to discuss the relationship between free entropy and free entropy dimension. Recall that we had stated that the Minkowski dimension of a subset of $\mathbb{R}^n$ is less than n only if the lebesgue measure of that subset is 0. An analogous statement can be said in the free case. The free entropy dimension of an n-tuple of operators is less than n only if the free entropy of the n-tuple is $-\infty$ (not 0 as in the other case, because we consider a log of the lebesgue measure). Indeed, one has $\chi(X_1, \ldots, X_n) > -\infty$ implies the absence of atoms in the distributions of each X_i. But more generally, one has

$$\limsup_{\epsilon \to 0} \frac{\chi(X_1 + \epsilon S_1, \ldots, X_n + \epsilon S_n)}{\left|log(\epsilon)\right|} = 0$$

since the numerator is bounded. Thus, full free entropy dimension is attained. Other similar analogues can be stated and proved in a similar manner.

Lastly note that free entropy dimension is a positive quantity, simply because the measures we are dealing with are probability measures, so the sum of the squares of the atomic masses cannot be greater than 1.

4.4 The free group factor problem

4.4.1 Revisiting free group factors

In 2.3.6, we defined the group von Neumann algebras. In the special case of $\Gamma = \mathbb{F}_n$ free group on n generators we have the free group factors (an exercise is to check why these are factors. In general, any group with infinitely long conjugacy classes will yield a factor). When $n = 1$, it is easy to identify what the von Neumann algebra is. It has to be abelian first of all, so that tells us that it is an L^∞ space. Consider $L(\mathbb{Z}) = L(\mathbb{F}_1)$. For the action of $L(\mathbb{Z})$ on $l^2(\mathbb{Z})$, δ_0 is a cyclic vector. Let x be the unitary operator corresponding to $1 \in \mathbb{Z}$. Recall that $\mathbb{Z}$ and $\mathbb{T}$ are Pontryagin duals to each other via the map $\mathbb{Z} \times \mathbb{T} \ni (n, \zeta) \mapsto \zeta^n$. Using this duality, define a unitary $U : l^2(\mathbb{Z}) \to L^2(\mathbb{T}, m)$ (where m is the normalized lebesgue measure), such that $[U(\xi)](\zeta) = \sum_{n \in \mathbb{Z}} \xi(n)\zeta^n$. One can easily check that $UxU^* = f$ where f on $L^\infty(\mathbb{T})$ is the identity function $f(\xi) = \xi$. Note that a net $(f_\alpha) \subset L^\infty(X, \mu)$ converges in the weak* topology if and only if the multiplication operators

$(m_{f_\alpha}) \subset \mathcal{B}(L^2(X,\mu))$ converges in the WOT. Using this fact, and the density of polynomials, we can conclude

$$UL(\mathbb{Z})U^* = U\overline{\mathbb{C}\langle x\rangle}_{WOT}U^* = \overline{U\mathbb{C}\langle x\rangle U^*}_{WOT} = \overline{\mathbb{C}\langle f\rangle}_{WOT} = \overline{\mathbb{C}\langle f\rangle}_{wk^*} = L^\infty(\mathbb{T}, m)$$

This gives us $L(\mathbb{Z}) \cong L^\infty(\mathbb{T}, m)$.

Now, we'd like to know more about the case when $n > 2$. It is here that we define the notion of a free product. It will require some work to define this, because we want important properties like those around Proposition 2.5.3, to hold. Now, the notion of free product of von Neumann algebras existed long ago, as one can see in the paper [Chi73]. But there has been much development in notation and technology since the introduction of the study of free group factors by Voiculescu. We present here a relatively modern construction, as given in [PAD10].

Let M_1 and M_2 be two tracial von Neumann algebras (the reader is advised to think of group von Neumann algebras, in particular the free group von Neumann algebras). Set for $i = 1, 2$, $\mathcal{H}_i = L^2(M_i, \tau_i)$, the GNS constructions with respect to ϕ, and $\xi_i = \hat{1}_{M_i}$. First we construct a free product on the level of Hilbert spaces, so that we can represent M_i on that. Denote by $\mathcal{H}_i^\circ$ the orthogonal complement of $\mathbb{C}\xi_i$ in $\mathcal{H}_i$. The Hilbert space free product $(\mathcal{H}_1, \xi_1) * (\mathcal{H}_2, \xi_2)$ is $(\mathcal{H}, \xi)$ given by the direct Hilbertian sum:

$$\mathcal{H} = \mathbb{C}\xi \oplus \bigoplus_{n\geq 1}\left(\bigoplus_{i_1\neq i_2\neq \ldots \neq i_n} \mathcal{H}_{i_1}^\circ \otimes \ldots \otimes \mathcal{H}_{i_n}^\circ\right)$$

where ξ is a unit vector. We set

$$\mathcal{H}_l(i) = \mathbb{C}\xi \oplus \bigoplus_{n\geq 1}\left(\bigoplus_{i_1\neq i_2\neq \ldots \neq i_n, i_1\neq i} \mathcal{H}_{i_1}^\circ \otimes \ldots \otimes \mathcal{H}_{i_n}^\circ\right)$$

Now define the unitary operator $V_i : \mathcal{H}_i \otimes \mathcal{H}_l(i) \to \mathcal{H}$ as follows:

$$\xi_i \otimes \xi \mapsto \xi$$

$$\xi_i \otimes \eta \mapsto \eta, \quad \forall \eta \in \mathcal{H}_{i_1}^\circ \ldots \otimes \mathcal{H}_{i_n}^\circ, i_1 \neq i$$

$$\eta \otimes \xi \mapsto \eta, \quad \forall \eta \in \mathcal{H}_i^\circ$$

$$\eta \otimes \eta' \mapsto \eta \otimes \eta', \quad \forall \eta \in \mathcal{H}_i^\circ, \eta' \in \mathcal{H}_{i_1}^\circ \otimes \ldots \otimes \mathcal{H}_{i_n}^\circ, i_1 \neq i$$

Similarly, set

$$\mathcal{H}_r(i) = \mathbb{C}\xi \oplus \bigoplus_{n\geq 1} \bigoplus_{i_1\neq i_2\neq \ldots \neq i_n, i_n\neq i} \mathcal{H}_{i_1}^\circ \otimes \ldots \otimes \mathcal{H}_{i_n}^\circ$$

and define the corresponding unitary operator $W_i : \mathcal{H}_r(i) \otimes \mathcal{H}_i \to \mathcal{H}$. Now finally, we are ready to faithfully represent M_i on $\mathcal{H}$, by the following:

$$\forall x \in M_i, \quad \lambda_i(x) = V_i(x \otimes Id_{\mathcal{H}_l(i)})V_i^*$$

We now call

$$(M_1, \tau_1) * (M_2, \tau_2) := ((\lambda_1(M_1) \cup \lambda_2(M_2))'', \omega_\xi)$$

where ω_ξ is the vector state associated to ξ. The same construction can be generalized for an arbitrary free product of tracial von Neumann algebras.

Free group factors were studied for many reasons. They were viewed as exotic creatures before Voiculescu's seminal work. There are a couple of problems that people were interested in greatly, in the operator algebras.

Example 7 (Free group factor isomorphism)**.** For $n \neq m$, is $L(\mathbb{F}_n)$ not isomorphic to $L(\mathbb{F}_m)$?

Remark 4. Note that this problem is easy at the level of groups. The non isomorphism is guaranteed because of different abelianizations. At the level of C^* algebras, the non isomorphism is also achieved, and this is a highly non trivial work due to Voiculescu and Pimnser, and involves computation of K-theory of these free group factors.

Example 8 (Absence of Cartan Subalgebras)**.** Does there exist a seperable $II - 1$ factor which does not have a Cartan subalgebra?

Remark 5. A Cartan subalgebra is a MASA whose normalizer generates the whole von Neumann algebra. The diagonal algebra in the matrix algebras is a good example to consider. In fact, this is a very central example, because it was proven by Feldman and Moore that a $II - 1$ factor contains a Cartan subalgebra if and only if it arises from a measurable ergodic equivalence relation. Refer [FM77]

Example 9 (Prime $II - 1$ factors)**.** Are the free group factors prime?

Remark 6. Primeness is an essential property. A $II - 1$ factor is prime if it cannot be expressed as a W^* tensor product of infinite dimensional von Neumann algebras.

These problems (except the first one) were knocked out by Voiculescu with the introduction of free entropy dimension.

4.4.2 Generating a free group factor

We already saw that $\delta(s) = 1$ for a semicircular. Also, $W^*(s) \cong L(\mathbb{F}_1)$, so it is a natural question to wonder if free entropy dimension is the quantity that could possibly determine the rank of the free group factor that the set generates. In essence, we wonder if the free entropy dimension captures the "degree of freeness" of the variables.

Proposition 4.10. *Let X_j, $1 \leq j \leq n$ be self adjoint random variables in (M, τ). Then one has $\delta(X_1, \ldots, X_n) \leq n$. If moreover, $X_j \in M_1$, $1 \leq j < n$ where M_1 is a unital sub von Neumann algebra of M, that is isomorphic to a free group factor, then $\delta(X_1, \ldots, X_n) \geq 0$.*

Proof. The first claim of the proposition is obvious because of the fact that the sum of squares of the atoms is a positive number. Secondly, we simply apply the proposition 3.26 to our definition, and get

$$\chi(X_1 + \epsilon S_1, \ldots, X_n + \epsilon S_n) \geq \chi(\epsilon S_1, \ldots, \epsilon S_n) =$$
$$-n\big|log(\epsilon)\big| + \chi(S_1, \ldots, S_n)$$

From just plugging this into the definition and recognizing that the free entropy of a bunch of semicirculars is positive, one has the result. □

Now, we would like to recall a fact that we have already discussed, which is an immediate corollary to the above lemmas.

Corollary 4.11. *If $\delta(X_1, \ldots, X_n) = n$ then $Ker(X_j - tId) = 0$, in particular, each X_j has an atomless distribution.*

In fact, if one starts with X_i freely independent, with free entropy greater than $-\infty$, one has that $\delta(X_1, \ldots, X_n) = \delta(X_1) + \ldots + \delta(X_n) = n$. More generally, we have $W^*(X_1, \ldots, X_n) = W^*(X_1) * \ldots * W^*(X_n) \cong W^*(s_1) * \ldots * W^*(s_n) \cong L(\mathbb{F}_n)$ where s_i are free semicirculars.

It is a very important open problem whether $W^*(X_1, \ldots, X_n) \cong L(\mathbb{F}_{\delta(X_1,\ldots,X_n)})$. Well, this doesn't make sense if the free entropy dimension is not an integer. Questions of these nature motivated Ken Dykema (and independently, Florin Radulescu in [Rad94]) to define interpolated free group factors in [Dyk94]. The interpolated free group factors $L(\mathbb{F}_t)$ precisely are "free group factors of non integer rank" in the sense that the following properties are satisfied:

1. $L(\mathbb{F}_t) \cong L(\mathbb{F}_n)$ for $t = n$ and $n \in \mathbb{N}$.
2. $L(\mathbb{F}_{r_1}) * L(\mathbb{F}_{r_2}) = L(\mathbb{F}_{r_1+r_2})$
3. $L(\mathbb{F}_r)_\gamma = L\left(\mathbb{F}_{\left(1+\frac{r-1}{\gamma^2}\right)}\right)$

where for a II_1 factor M, M_γ means the algebra defined as follows: for $0 < \gamma \leq 1$, $M_\gamma = pMp$ where p is a self adjoint projection of trace γ. Upon viewing the definition of these interpolated free group factors, Voiculescu observed the following immediate corollary:

Corollary 4.12. *If $X_1, \ldots, X_n$ generate a II_1 factor and are free then*

$$W^*(X_1, \ldots, X_n) \cong L(\mathbb{F}_{\delta(X_1,\ldots,X_n)})$$

At first sight, this seems like a good generalization of the observation we made earlier. But what will be an incredible result is one that somehow relaxes the freeness condition. From our change of variables formulae, we find that free entropy dimension is an algebra invariant. In other words, if two sets of random variables generate the same algebra, then their free entropy dimension is the same. In our effort to try to expand this invariant, we need to record this basic result about the impact of adding random variables from the generated von Neumann algebra on free entropy.

Proposition 4.13. *Let $(X_1, \dots, X_n)$ be a free family of self adjoint random variables in (M, τ) and assume $\chi(X_1, \dots, X_n) > -\infty$ and $Y_1, \dots, Y_m \in W^*(X_1, \dots, X_n)$. Then*

$$n = \delta(X_1, \dots, X_n) \leq \delta(X_1, \dots, X_n, Y_1, \dots, Y_m)$$

Proof. Let $(S_1, \dots S_{m+n})$ be a free semicircular family in (M, τ) which is free from $(X_1, \dots, X_n)$. Now, let us try and compute the right hand side, in particular:

$$\chi(X_1 + \epsilon S_1, \dots, X_n + \epsilon S_n, Y_1 + \epsilon S_{n+1}, \dots, Y_m + \epsilon S_{n+m})$$

we know that this will be greater than when we condition with the free semicirculars $S_1, \dots, S_n$ (see proposition 3.11) .

$$\chi(X_1 + \epsilon S_1, \dots, X_n + \epsilon S_n, Y_1 + \epsilon S_{n+1}, \dots, Y_m + \epsilon S_{n+m}) \leq$$

$$\chi(X_1 + \epsilon S_1, \dots, X_n + \epsilon S_n, Y_1 + \epsilon S_{n+1}, \dots, Y_m + \epsilon S_{n+m} | S_1, \dots, S_n)$$

From proposition 3.25, one has

$$= \chi(X_1, \dots, X_n, Y_1 + \epsilon S_{n+1}, \dots, Y_m + \epsilon S_{n+m} | S_1, \dots, S_n)$$

and since $Y_i \in W^*(X_1, \dots, X_n)$, we again have

$$= \chi(X_1, \dots, X_n, \epsilon S_{n+1}, \dots, \epsilon S_{n+m} | S_1, \dots, X_n)$$

Now applying the formula for condition entropy, we have

$$= \chi(X_1, \dots, X_n, \epsilon S_1, \dots, \epsilon S_n) - \chi(S_1, \dots, S_n)$$

using additivity of free entropy, we have □

4.4.3 An important smoothness condition

It turns out that for considering a transformation of $X_1, X_2, \dots, X_n$ to get $Y_1, \dots, Y_m$, in order to tame the free entropy dimension, we require this following crucial smoothness condition, that seems to be unfortunately quite necessary for our purposes (as we shall see later).

Proposition 4.14. *Let $X_1, \dots, X_n$, $Y_1, \dots, Y_m$ be self adjoint random variables in (M, τ) such that $Y_j \in W^*(X_1, \dots, X_n)$. Let $(S_1, \dots, S_n)$ be a semicircular family in (M, τ) free with respect to $(X_1, \dots, X_n)$ and let*

$$d_2(Y_j; X_1, \dots, X_n)(\epsilon) = \inf\{\|Y_j - T\|_2, T \in W^*(X_1 + \epsilon S_1, \dots, X_n + \epsilon S_n)\}$$

We assume that $d_2(Y_j; X_1, \dots, X_n)(\epsilon) = \mathcal{O}(\epsilon^s)$ for all $0 < s < 1$ and $1 \leq j \leq m$. Then we have

$$\delta(X_1, \dots, X_n, Y_1, \dots, Y_m) \leq \delta(X_1, \dots, X_n)$$

Proof. Let us without loss of generality, consider an extended semicircular family $S_1, \dots, S_{n+m}$ free from $X_1, \dots, X_n$. We have from proposition 3.25

$$\chi(X_1 + \epsilon S_1, \dots, X_n + \epsilon S_n, Y_1 + \epsilon S_{n+1}, \dots, Y_m + \epsilon S_{n+m})$$

$$= \chi(X_1 + \epsilon S_1, \dots, X_n + \epsilon S_n, Y_1 - T_1 + \epsilon S_{n+1}, \dots, Y_m - T_m + \epsilon S_{n+m})$$

where T_j is the conditional expectation of Y_j onto the subspace $W^*(X_1 + \epsilon S_1, \dots, X_n + \epsilon S_n)$. Now, we use subadditivity to remove out the X_i's, and remove a factor of ϵ^s by adding the Jacobian cost (of this trivial transformation):

$$\leq \chi(X_1 + \epsilon S_1, \dots, X_n + \epsilon S_n)$$

$$+\chi(\epsilon^{-s}(Y_1 - T_1) + \epsilon^{1-s} S_{n+1}, \dots, \epsilon^{-s}(Y_m - T_m) + \epsilon^{1-s} S_{n+m}) + mlog(\epsilon^s) \tag{2}$$

Now, since $d_2(Y_j; X_1, \dots, X_n)(\epsilon) = \mathcal{O}(\epsilon^s)$, for s strictly less than 1, we have $Y_j - T_j$ is strictly 2-norm bounded by a factor of ϵ^s, so the free entropy of the perturbation of S_i by these $Y_j - T_j$ will have bounded free entropy independent of ϵ. Indeed,

$$Var^2 = \tau\left(\sum_{j=1}^{m} (\epsilon^{-s}(Y_j - T_j) + \epsilon^{1-s} S_{j+m})^2\right) = \sum_{j=1}^{m} \tau(\epsilon^{-s}(Y_j - T_j) + \epsilon^{1-s} S_{j+m})^2$$

$$= \sum_{j=1}^{m} \left\|(\epsilon^{-s}(Y_j - T_j) + \epsilon^{1-s} S_{j+m})\right\|_2^2 \leq \sum_{j=1}^{m} \left(\epsilon^{-s}\left\|(Y_j - T_j)\right\|_2 + \epsilon^{1-s}\left\|S_{j+m}\right\|_2\right)^2$$

$$\leq \sum_{j=1}^{m} \left(\epsilon^{-s}\epsilon^{s}\mathcal{O}(1) + \left\|S_{j+m}\right\|_2\right)^2 = \mathcal{O}(1)$$

Now, it follows from this that expression (2) is

$$\leq \chi(X_1 + \epsilon S_1, \dots, X_n + \epsilon S_n) + K + mslog(\epsilon)$$

dividing by $\left|log(\epsilon)\right|$, taking limsup and adding n to the LHS and $m + (n - m)$ to the RHS, we get the result:

$$\delta(X_1, \dots, X_n, Y_1, \dots, Y_m) \leq \delta(X_1, \dots, X_n) + m(1 - s)$$

Letting s be arbitrarily close to 1, we have the result we seek. □

Observe that this result gives the reverse inequality of Proposition 4.13. In particular, we can get closer to some kind of invariant result for free entropy dimension, provided we have this notion of smoothness $d_2(Y_j : X_1, \dots, X_n)(\epsilon)$ is $\mathcal{O}(1)$. This smoothness condition we have described is a kind of Holder condition, and it is immediate that this smoothness holds in the case of non commutative polynomials. We can now employ this weapon in the context of non commutative power series transformations, if we can verify that it is smooth in the above spirit. But by no means is this non commutative power series the most general situation we can hope for. But unfortunately, getting to L^∞ functional transformations (the best situation one can hope for) is not possible.

4.4.4 Smoothness of non commutative power series transformations

Let us set up our construction. Let $t_1, \ldots, t_n$ be non commuting indeterminates and let

$$F(t_1, \ldots, t_n) = \sum_{k=0}^{\infty} \sum_{1 \le i_1 \ldots i_k \le n} C_{i_1 \ldots i_k} t_{i_1, \ldots i_k}$$

be a non commutative power series. Let $R_j \ge 0$ $(1 \le j \le n)$ be real numbers. We recall that $(R_1, \ldots, R_n)$ is a multi radius of convergence if

$$\sum_{k=0}^{\infty} \sum_{1 \le i_1 \ldots i_k \le n} \left| C_{i_1 \ldots i_k} \right| R_{i_1} \ldots R_{i_k} < \infty$$

Proposition 4.15. *Let $X_1, \ldots, X_n$, Y be non-commutative random variables in (M, τ) and assume that there is a non-commutative power series F such that $(R_1, \ldots, R_n)$ is a multi radius of convergence for some $R_j > \left\| X_j \right\|$ and $F(X_1, \ldots, X_n) = Y$. If $T_1, \ldots, T_n$ are non-commutative random variables in (M, τ) and*

$$\phi(\epsilon) = inf\{\|Y - T\| : T \in C^*(X_1 + \epsilon T_1, \ldots, X_n + \epsilon T_n)\}$$

Then, $\phi(\epsilon) = \mathcal{O}(\epsilon)$

Proof. First remark that this result will be particularly useful because, $\phi(\epsilon) = \mathcal{O}(\epsilon)$ implies that $d_2(Y; X_1, \ldots, X_n)(\epsilon) = \mathcal{O}(\epsilon^s)$ for all $0 < s < 1$. Indeed, $\|Y - T\|_2 \le \|Y - T\|$, and the C^* algebra generated is norm dense in the W^* algebra generated.

Now, we have from the analyticity of the transformation, we have F defines an analytic function

$$\Phi : \prod_{j=1}^{n} \{Z \in M, \|Z\| < R_j\} \to M$$

given by $\Phi(Z_1, \ldots, Z_n) = F(Z_1, \ldots, Z_n)$. Then, we have

$$G(\epsilon) = \Phi(X_1 + \epsilon T_1, \ldots, X_n + \epsilon T_n) - \Phi(X_1, \ldots, X_n)$$

defines an analytic function on ϵ in some neighborhood of $0 \in \mathbb{C}$. If one is familiar with analytic functions around 0, it is easy to see that $\left\| G(\epsilon) \right\| = \mathcal{O}(\|\epsilon\|)$ as $\epsilon \to 0$. And also, $\left\| G(\epsilon) \right\| \ge \phi(\epsilon)$ by construction. Hence we have the result. $\square$

4.4.5 Free entropy dimension as a ‘smooth’ algebra invariant

Corollary 4.16. *Let $(X_1, \ldots, X_n)$ be a free family of self adjoint random variables in (M, τ) such that $\chi(X_j) > -\infty$, $(1 \le j \le n)$. Let $Y_k = Y_k^* = F_k(X_1, \ldots, X_n)$, $(1 \le k \le n)$ where the F_k's are non commutative power series for which there is a multi radius of convergence $(R_1, \ldots, R_n)$ such that $\left\| X_j \right\| < R_j$ $(1 \le j \le n)$. Then*

$$n = \delta(X_1, \ldots, X_n) = \delta(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$$

Proof. From Proposition 4.15, one has that the hypothesis of 4.14 is satisfied. But also observe that the hypothesis of 4.13 is also satisfied independently. Thus, from 4.14 and 4.13, one has the desired equality. $\square$

Finally, we have the main result:

Corollary 4.17. *(Voiculescu '94) Let $X_1, \dots, X_n$, $Y_1, \dots, Y_m$ be self adjoint random variables in (M, τ) such that $W^*(X_1, \dots, X_n) = W^*(Y_1, \dots, Y_m)$. Assume that $Y_k = F_k(X_1, \dots, X_n)$ where the F_k's are non commutative power series with $(\|X\| + \epsilon, \dots, \|X_n\| + \epsilon)$ a multi radius of convergence for some $\epsilon > 0$. Assume also that $(Y_1, \dots, Y_m)$ is a free family and $\chi(Y_j) > -\infty$ for all j. Then $\delta(X_1, \dots, X_n) \geq m$ and in particular, $n \geq m$.*

Proof. We know that $\delta(X_1, \dots, X_n, Y_1, \dots, Y_m) \geq \delta(Y_1, \dots, Y_m)$, and we also know that by free additivity $\delta(Y_1, \dots, Y_m) = m$. However, by 4.14, we also have

$$\delta(X_1, \dots, X_n) \geq \delta(X_1, \dots, X_n, Y_1, \dots, Y_m)$$

From the definition we also have that $n \geq \delta(X_1, \dots, X_n)$. Combining these, we have $n \geq m$. $\square$

In particular, one has, suppose $L(\mathbb{F}_n) \cong L(\mathbb{F}_m)$ (without loss of generality, $n \leq m$), then considering $X_1, \dots, X_n$ to be a free semicircular family generating $L(\mathbb{F}_n)$, and suppose there is $Y_1, \dots, Y_m$ which are a 'nice' smooth function of the X_i's, and are free and generate $L(\mathbb{F}_m)$, then one has $n = m$. Note that by replacing the 'smooth' with L^∞, we solve the free group factor isomorphism problem. Indeed, since X_i's and Y_j's generate the same von Neumann algebras, one can guarantee to get the Y_j's by an L^∞ map of the X_i's.

It is important to point here the following remark, which could attempt to salvage this approach.

Remark 7. If δ satisfies the following semicontinuity property: if $X_j^{(p)}$ strongly converges to X_j as $p \to \infty$ for $1 \leq j \leq n$ in (M, τ) then

$$\liminf_{p \to \infty} \delta(X_1^{(p)}, \dots, , X_n^{(p)}) \geq \delta(X_1, \dots, X_n)$$

then it would follow that one can remove the assumption $d_2(Y_j, X_1, \dots, X_n)(\epsilon) = \mathcal{O}(\epsilon)$ can be removed from the proposition 4.15. This is extremely strong as it would further imply that one can just use $Y_k \in W^*(X_1, \dots, X_n)$ and therefore imply the non isomorphism of the free group factors.

References

[AGZ09] G. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to random matrices*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2009.

[BB03] S. T. Belinschi and H. Berkovici. A property of free entropy. *Pacific J. Math*, 211(1):35–40, 2003.

[BCG03] P. Biane, M. Capitaine, and A. Guionnet. Large deviation bounds for matrix brownian motion. *Invent. math*, 152(2):433–459, 2003.

[Cha15] Djalil Chafai. From Boltzmann to random matrices and beyond. *Annales de la Faculté de Sciences de Toulouse*, 12(4):641–689, 2015.

[Chi73] Wai-Mee Ching. Free products of von Neumann algebras. *Trans. Amer. Math. Soc*, 178:147–163, 1973.

[Dab16] Yoann Dabrowski. A Laplace Principle for Hermitian Brownian Motion I: the convex functional case. *arXiv:1604.06420*, 2016.

[Dyk94] Kenneth J. Dykema. Interpolated free group factors. *Pacific J. Math*, 163:123–135, 1994.

[FM77] J. Feldman and C. C. Moore. Ergodic equivalence relations, cohomology and von Neumann algebras, I, II. *Trans. Amer. Math. Soc*, 234:289–359, 1977.

[Jun02a] Kenley Jung. A free entropy dimension lemma. *Pac. J. Math*, 211(2):265–271, 2002.

[Jun02b] Kenley Jung. A hyperfinite inequality for von neumann algebras. *Pac. J. Math*, 211(2):265–271, 2002.

[Kir93] E. Kirchberg. On non-semisplit extensions, tensor products and exactness of group c^* algebras. *Invent. Math.*, 112:449–489, 1993.

[Meh91] M. L. Mehta. *Random Matrices*. revised and enlarged. San Diego: Academic Press, 2 edition, 1991.

[NCon] Brent Nelson and Ian Charlesworth. *Free Stein Discrepency as a Regularity Condition*. In Preparation.

[NS06] Alexandru Nica and Roland Speicher. *Lectures on the combinatorics of free probability*. Number 335 in London Mathematical Society Lecture Note Series. Cambridge University Press, 2006.

[PAD10] Sorin Popa and Claire Anantharaman-Delaroche. An introduction to ii_1 factors. link: http://www.math.ucla.edu/~popa/Books/IIun-v10.pdf, 2010.

[Rad94] Florin Radulescu. Random matrices, amalgamated free products and subfactors of the von neumann algebra of a free group of noninteger index. *Invent. math*, 115:347–389, 1994.

[Shl04] Dimitri Shlyakhtenko. Free probability theory. link: http://lanl.arxiv.org/abs/math/0504063, 2004.

[Shl07] Dimitri Shlyakhtenko. A free analogue of shannon's problem on monotonicity of entropy. *Adv. Math.*, 208(2):824–833, 2007.

[Sta59] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2(2):101–112, 1959.

[Voi91] Dan-Virgil Voiculescu. Limit laws for random matrices and free products. *Invent. math*, 104:201–220, 1991.

[Voi93] Dan-Virgil Voiculescu. The analogues of entropy and of Fisher's information measure in free probability theory I. *Commun. Math. Phys*, 155:71–92, 1993.

[Voi94] Dan-Virgil Voiculescu. The analogues of entropy and of fisher's information measure in free probability theory II. *Invent. math*, 118:411–440, 1994.

[Voi98a] Dan-Virgil Voiculescu. The analogues of entropy and of fisher's information measure in free probability theory V. *Invent. math*, 132:289–227, 1998.

[Voi98b] Dan-Virgil Voiculescu. A strengthened asymptotic freeness result for random matrices with applications to free entropy. *Int. Math. Res. Not.*, (1):41–64, 1998.

[Wig58] Eugene P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math.*, 67(2):325–327, 1958.