USING RANDOM MATRICES IN QUANTUM INFORMATION THEORY

ION NECHITA

Abstract. The goal of this series of lectures is to present some recent results in quantum information theory which make use of random matrices. After an introduction to random matrix theory, I will present the method of moments, one of the most successful methods used to study the spectra of large random matrices. This will be the occasion to discuss integration over Gaussian spaces. On the quantum information side, I will focus on two main topics, random quantum states and random quantum channels. I will then prove two recent results, one on the asymptotic eigenvalue distribution of the partial transposition of random quantum states, and another on the output set of random quantum channels. Both will require some terminology and results from free probability, which will also be discussed in detail.

Contents

1. Lecture 1 — Generalities. Wishart matrices 2
1.1. Introduction 2
1.2. Wishart matrices and their limit distribution 2
1.3. Graphical notation for tensors 4
1.4. Wick formula, algebraic and diagrammatic formulations 5
1.5. Non-crossing partitions and permutations 6
1.6. Proof of the Marcenko-Pastur theorem 7
2. Lecture 2 — Partial transposition of random quantum states. Free probability 9
2.1. Some elements of free probability theory 9
2.2. The full Fock space, free semicircular random variables 11
2.3. The partial transpose of random quantum states 12
3. Lecture 3 — Random quantum channels and their minimum output entropy 14
3.1. Quantum channels, minimum output entropies, additivity 14
3.2. Random quantum channels 16
3.3. Strong convergence and the (t)-norm 16
3.4. The MOE of a (large) random quantum channel 18
References 19

Date: February 27, 2016.

1. Lecture 1 — Generalities. Wishart matrices

1.1. **Introduction.** The birth of random matrix theory can be traced to statistics and physics. Wishart introduced the distribution that bears his name in the 1920's [Wis28], in order to explain the discrepancy between the eigenvalues of a measured covariance matrix, and an expected covariance matrix. Later, Wigner was studying nuclear physics when he introduced [Wig55] the semi-circle distribution. Since then, random matrix theory has played a role in many fields of mathematics and science, including operator algebras [VDN92], combinatorics, complex analysis, theoretical physics and telecommunication theory, just to cite a few. Quantum information theory is definitely one of the most recent of fields of application; for more on this, we direct the interested reader to the recent review [CN16].

In quantum information theory, randomness is built in, by the axioms of quantum mechanics. Since quantum states are modeled by (unit trace, positive semidefinite) matrices, it is clear that the two fields intersect. However, we can see two more reasons for the use of random matrices in quantum information. First, we would like to understand the *typical properties* of quantum states and channels, relative to tasks ans paradigms in quantum information theory. Very early, properties such as the average entanglement of quantum states were studied [Pag93], and several probability distribution over the set of quantum states were introduced [ŻS01]. Second, it turns out that some problems – in particular the minimum output entropy additivity problem, which we discuss at length here – did not have an obvious non-random answer, therefore it became not only natural, but also important, to consider random quantum objects.

One paper which popularized the use of random techniques in quantum information was [HLSW04]. This work pointed out that some well-established techniques in the mathematics of random matrices – measure concentration in this case – could be of use in quantum information.

Let us now gather here some basic definitions from quantum information theory and set up some notation.

A *quantum state* is a positive semidefinite matrix of unit trace. The set of all quantum states is a convex body denoted by

$$\mathcal{M}_d^{1,+}(\mathbb{C}) := \{\rho \in \mathcal{M}_d(\mathbb{C}) \,:\, \rho \geq 0 \text{ and } \operatorname{Tr}\rho = 1\}.$$

The extremal points of $\mathcal{M}_d^{1,+}(\mathbb{C})$ are the rank one projectors xx^* ($x \in \mathbb{C}^d$, $\|x\| = 1$), and they are called *pure states*.

Of particular interest are states of multiple quantum systems, which are quantum states acting on the *tensor power* of the corresponding Hilbert spaces. Of particular importance are the *separable states*, which in the bipartite case can be described as

$$\mathcal{SEP}_{d_1,d_2} := \operatorname{conv}\{\rho_1 \otimes \rho_2\}_{\rho_i \in \mathcal{M}_{d_i}^{1,+}(\mathbb{C})}.$$

Non-separable states are called *entangled*, and among those, of particular importance is the *maximally entangled state* $d^{-1}\Omega_d \in \mathcal{M}_{d^2}^{1,+}(\mathbb{C})$, where

$$\Omega_d = \sum_{i=1}^{d} e_i \otimes e_i, \tag{1.1}$$

where $\{e_i\}$ is an orthonormal basis of $\mathbb{C}^d$.

1.2. **Wishart matrices and their limit distribution.** The probability density of the *normal distribution* is:

$$f(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Here, μ is the *mean*. The parameter σ is its *standard deviation* with its variance then σ^2. A random variable with a Gaussian distribution is said to be normally distributed.

Suppose X and Y are random vectors in $\mathbb{R}^k$ such that (X, Y) is a $2k$-dimensional normal random vector. Then we say that the complex random vector $Z = X + iY$ has the *complex normal distribution*. The normal distribution (resp. random vector) are also called *Gaussian distribution* (resp. random vectors).

Historically the first ensemble of random matrices having been studied is the Wishart ensemble [Wis28], see [BS10, Chapter 3] or [AGZ10, Section 2.1] for a modern presentation.

Definition 1.1. *Let $G \in \mathcal{M}_{d\times s}(\mathbb{C})$ be a random matrix with complex, standard, i.i.d. Gaussian entries. The distribution of the positive-semidefinite matrix $W = GG^* \in \mathcal{M}_d(\mathbb{C})$ is called a* Wishart distribution *of parameters (d, s) and is denoted by $\mathcal{W}_{d,s}$.*

The study of the asymptotic behavior of Wishart random matrices is due to Marčenko and Pastur [MP67], while the stronger convergence results have been proved by analytic tools such as determinantal point processes; one can also recover the stronger forms of the theorem as direct consequences of the much more general results [Mal12]. Since we aim at giving complete proofs of our results, we state it here in a rather week form: the convergence in moments.

Definition 1.2. *A sequence of random matrices X_d is said to* converge in moments *to a probability distribution ν if for all positive integers p, we have*

$$\lim_{d\to\infty} \mathbb{E} \int t^p d\mu_{X_d} = \mathbb{E}\frac{1}{d}\operatorname{Tr}(X_d^p) = \int t^p d\nu,$$

where μ_{X_d} is the empirical eigenvalue distribution *of X_d*

$$\mu_{X_d} = \frac{1}{d}\sum_{i=1}^{d} \delta_{\lambda_i(X_d)}.$$

Theorem 1.3. *Consider a sequence s_d of positive integers which behaves as $s_d \sim cd$ as $d \to \infty$, for some constant $c \in (0, \infty)$. Let W_d be a sequence of positive-semidefinite random matrices such that W_d is distributed according to $\mathcal{W}_{d,s_d}$. Then, the sequence W_d converges in moments to the* Marčenko-Pastur distribution *π_c given by*

$$\pi_c = \max(1 - c, 0)\delta_0 + \frac{\sqrt{(b - x)(x - a)}}{2\pi x} \mathbf{1}_{(a,b)}(x)\, dx, \tag{1.2}$$

where $a = (1 - \sqrt{c})^2$ and $b = (1 + \sqrt{c})^2$.

The Marčenko-Pastur distribution π_c is sometimes called the *free Poisson distribution*, see [NS06, Proposition 12.11]. We plotted in Figure 1 its density in the cases $c = 1$ and $c = 4$.

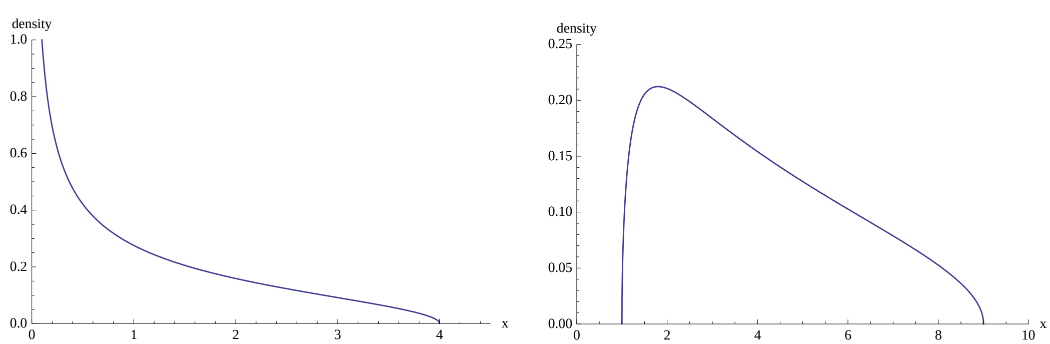

FIGURE 1. The density of the Marčenko-Pastur distributions π_1 (left) and π_4 (right).

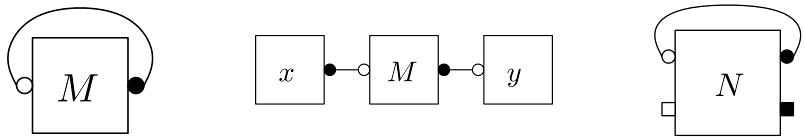

FIGURE 2. Some simple diagrams

Remark 1.4. *The Dirac mass appearing in* (1.2) *is due to the fact that if $c < 1$, the matrix W_d is rank deficient. Since $cd < d$, a fraction $1 - c$ of the eigenvalues of W_d are null, yielding the Dirac mass at zero.*

We postpone the proof of Theorem 1.3 to Section 1.6.

We end this section by the statement of the so-called Carleman condition, which ensures that a sequence of moments defines a unique probability measure.

Proposition 1.5. *Let μ be a probability measure on $\mathbb{R}$ having finite moments*

$$m_n = \int t^p d\mu(t)$$

which satisfy

$$\sum_{n=1}^{\infty} m_{2n}^{-1/(2n)} = +\infty.$$

Then, μ is the only measure on $\mathbb{R}$ having the sequence (m_n) as moments.

1.3. **Graphical notation for tensors.** Most operations from linear and multilinear algebra (composition, tensor product, (partial) traces) can be efficiently represented graphically. The leading idea is that a string in a diagram means a tensor contraction. Many graphical theories for tensors and linear algebra computations have been developed in the literature [Pen05, Coe10]. Although they are all more or less equivalent, we will stick to the one introduced in [CN10b], as it allows to compute the expectation of random diagrams in a diagrammatic way subsequently. For more details on this method, we refer the reader to the paper [CN10b] and to other work which make use of this technique [CN11a, CN10a, CNŻ10, CNŻ13, FŚ13, CGGPG13, Lan15]

In the graphical calculus, matrices (or, more generally, tensors) are represented by boxes. Each box has differently shaped symbols, where the number of different types of them equals that of different spaces (exceptions are mentioned below). Those symbols are empty (white) or filled (black), corresponding to primal or dual spaces. Wires connect these symbols, corresponding to tensor contractions. A diagram is a collection of such boxes and wires and corresponds to an element of an abstract tensor product space. Rather than going through the whole theory, we focus next on a few key examples.

Suppose that each diagram in Figure 2 comes equipped with two vector spaces V_1 and V_2 which we shall represent respectively by circle and square shaped symbols. In the first diagram, M is a tensor (or a matrix, depending on which point of view we adopt) $M \in V_1^* \otimes V_1$, and the wire applies the contraction $V_1^* \otimes V_1 \to \mathbb{C}$ to M. The result of the diagram $\mathcal{D}_a$ is thus $T_{\mathcal{D}_a} = \mathrm{Tr}(M) \in \mathbb{C}$. In the second diagram, again there are no free decorations, hence the result is the complex number $T_{\mathcal{D}_b} = \langle y, Mx \rangle$. Finally, in the third example, N is a $(2, 2)$ tensor or a linear map $N \in \mathrm{End}(V_1 \otimes V_2, V_1 \otimes V_2)$. When one applies to the tensor N the contraction of the couple (V_1, V_1^*), the result is the partial trace of N over the space V_1: $T_{\mathcal{D}_c} = \mathrm{Tr}_{V_1}(N) \in \mathrm{End}(V_2, V_2)$. We depict in Figure 3 the maximally entangled (un-normalized) state Ω_d from (1.1), as well as its partial trace, $[\mathrm{id} \otimes \mathrm{Tr}](\Omega_d) = I_d$.

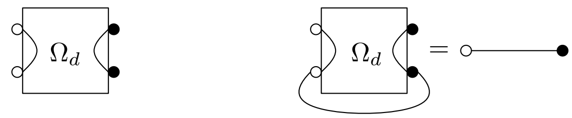

FIGURE 3. The maximally entangled state (left) and its partial trace (right).

1.4. **Wick formula, algebraic and diagrammatic formulations.** The following theorem is the link between combinatorics and probability theory for Gaussian vectors: it allows to compute moments of any Gaussian vector thanks to its covariance matrix. A *Gaussian space* V is a real vector space of random variables having moments of all orders, with the property that each of these random variables has centered Gaussian distributions. In order to specify the covariance information, such a Gaussian space comes with a positive symmetric bilinear form $(x, y) \to \mathbb{E}[xy]$. Gaussian spaces are in one-to-one correspondence with Euclidean spaces. In particular, the Euclidean norm of a random variable determines it fully (via its variance) and if two random variables are given, their joint distribution is determined by their angle. The following is usually called the Wick Lemma:

Theorem 1.6. *Let V be a Gaussian space and $x_1, \ldots, x_k$ be elements in V. If $k = 2l + 1$ then $\mathbb{E}[x_1 \cdots x_k] = 0$ and if $k = 2l$ then*

$$\mathbb{E}[x_1 \cdots x_k] = \sum_{\substack{p=\{\{i_1,j_1\},\ldots,\{i_l,j_l\}\} \\ \textit{pairing of } \{1,\ldots,k\}}} \prod_{m=1}^{l} \mathbb{E}[x_{i_m} x_{j_m}]. \tag{1.3}$$

In particular it follows that if $x_1, \ldots, x_p$ are independent standard Gaussian random variables, then

$$\mathbb{E}[x_1^{2k_1} \ldots x_p^{2k_p}] = \prod_{i=1}^{p} (2k_i)!!.$$

The main difference between the real case discussed above and the complex case is that one has to pair Gaussian variables to their conjugates in the complex situation. This follows from the fact that if Z is a standard complex Gaussian random variable,

$$\mathbb{E}[Z^2] = \mathbb{E}[\bar{Z}^2] = 0, \quad \text{while} \quad \mathbb{E}[Z\bar{Z}] = 1.$$

We shall now recast the Wick formula above in the graphical formalism described in the previous section. Consider a diagram which containts a new special box G corresponding to a *Gaussian random matrix*. We shall compute the expected value of a random diagram with respect to the Gaussian probability measure; as we shall see, this operation will consist of *expanding* the diagram, by erasing the Gaussian boxes and replacing them with wires.

To start, consider $\mathcal{D}$ a diagram which contains, amongst other constant tensors, boxes corresponding to independent Gaussian random matrices of *covariance one* (identity). One can deal with more general Gaussian matrices by multiplying the standard ones with constant matrices. Note that a box can appear several times, adjoints of boxes are allowed and the diagram may be disconnected. Also, Gaussian matrices need not be square.

The expectation value of such a random diagram $\mathcal{D}$ can be computed by a *removal* procedure as in the unitary case. Without loss of generality, we assume that we do not have in our diagram adjoints of Gaussian matrices, but instead their complex conjugate box. This assumption allows for a more straightforward use of the Wick formula from Theorem 1.6. We can assume that $\mathcal{D}$ contains only one type of random Gaussian box G; other independent random Gaussian matrices are assumed constant at this stage as they can be removed in the same manner afterwards.

A removal of the diagram $\mathcal{D}$ is a pairing between *Gaussian boxes* G and their conjugates $\bar{G}$. The set of removals is denoted by $\mathrm{Rem}_G(\mathcal{D})$ and it may be empty: if the number of G boxes is different

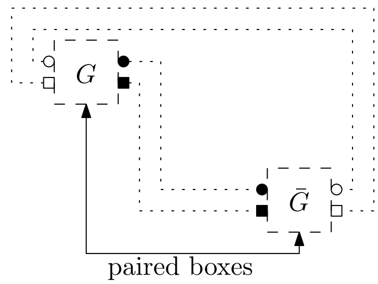

FIGURE 4. Pairing of boxes in the Gaussian case

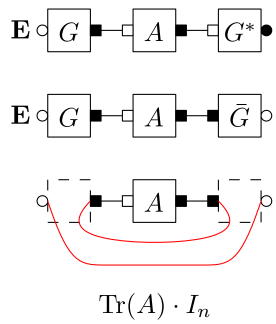

FIGURE 5. Applying Theorem 1.7 to compute $\mathbb{E}[GAG^*]$.

from the number of $\bar{G}$ boxes, then $\mathrm{Rem}_G(\mathcal{D}) = \emptyset$ (since no pairing between matrices and their conjugates can exist). Otherwise, a removal r can identified with a permutation $\alpha \in \mathcal{S}_p$, where p is the number of G and $\bar{G}$ boxes. In the Gaussian/Wick calculus, one pairs conjugate boxes: white and black decorations are paired in an identical manner, hence only one permutation is needed to encode the removal.

To each removal r associated to a permutation $\alpha \in \mathcal{S}_p$ corresponds a removed diagram $\mathcal{D}_r$ constructed as follows. One starts by erasing the boxes G and $\bar{G}$, but keeps the decorations attached to these boxes. Then, the decorations (white *and* black) of the i-th G box are paired with the decorations of the $\alpha(i)$-th $\bar{G}$ box in a coherent manner, see Figure 4.

The graphical reformulation of the Wick formula from Theorem 1.6 becomes the following theorem, which we state without proof.

Theorem 1.7. *The following holds true:*

$$\mathbb{E}_G[\mathcal{D}] = \sum_{r \in \mathrm{Rem}_G(\mathcal{D})} \mathcal{D}_r.$$

In Figure 5, we present an example of application of the theorem above. We consider, on the first row, the diagram corresponding to $\mathbb{E}[GAG^*]$, where $G \in \mathcal{M}_{n\times k}(\mathbb{C})$ is a $n \times k$ Gaussian matrix, and $A \in \mathcal{M}_k(\mathbb{C})$ is a square, deterministic matrix. The first row contains the diagram $\mathcal{D}$ associated to the algebraic expression. In the second row, we rewrite the same diagram, replacing G^* by $\bar{G}^\top$, in order to be able to apply Theorem 1.7. The third row contains the result of the application: we erase the $G/\bar{G}$ boxed and we add the wires corresponding to the permutation $(1) \in \mathcal{S}_1$ (in red). We recognize the diagrams for the identity matrix and for the trace of A: $\mathbb{E}[GAG^*] = \mathrm{Tr}(A)I_n$.

1.5. **Non-crossing partitions and permutations.** For a permutation $\sigma \in \mathcal{S}_p$, denote by $\#\sigma$ the number of its cycles, including the trivial ones (fixed points). Denote also by $|\sigma|$ its *length*, i.e. the minimum number of transposition which multiply to σ. It is well known that for all permutations

FIGURE 6. A non-crossing partition $\{\{1,2,4\},\{3\}\}$ (left) vs. a crossing one $\{\{1,3\},\{2,4\}\}$ (right).

$\sigma \in \mathcal{S}_p$,

$$\#\sigma + |\sigma| = p.$$

The set of *non-crossing partitions* will play a crucial role in what follows. Recall that a partition π of $[p] := \{1, 2, \ldots, p\}$ is called non-crossing if there are now quadruples (a, b, c, d) such that a, b (resp. c, d) belong to the same block of π, and $a < c < b < d$; see Figure 6 for some examples. The are supremum and infimum operations on $NC(p)$, which turn it into a lattice, see [NS06, Lecture 9]. The number of elements in the set $NC(p)$ is the *Catalan number*

$$\mathrm{Cat}_p = \frac{1}{p+1}\binom{2p}{p}.$$

These numbers satisfy the recurrence relation

$$\mathrm{Cat}_p = \sum_{i=1}^{p} \mathrm{Cat}_{i-1}\mathrm{Cat}_{p-i},$$

and thus their generating series is given by

$$M(z) = \sum_{p=0}^{\infty} \mathrm{Cat}_p z^p = \frac{1-\sqrt{1-4z}}{2z}.$$

We collect now a some properties of the distance function over the symmetric group, which allow us to bijectively identify a subset of $\mathcal{S}_p$ with $NC(p)$. This result can be traced back to [Bia97].

Lemma 1.8. *The function $d(\sigma, \tau) = |\sigma^{-1}\tau|$ is an integer valued distance on $\mathcal{S}_p$. Besides, it has the following properties:*

- *the diameter of $\mathcal{S}_p$ is $p-1$;*
- *$d(\cdot,\cdot)$ is left and right translation invariant;*
- *for three permutations $\sigma_1, \sigma_2, \tau \in \mathcal{S}_p$, the quantity $d(\tau,\sigma_1) + d(\tau,\sigma_2)$ has the same parity as $d(\sigma_1,\sigma_2)$;*
- *the set of* geodesic *points (elements which saturate the triangular inequality) between the identity permutation* id *and some permutation $\sigma \in \mathcal{S}_p$ is in bijection with the set of non-crossing partitions smaller than π, where the partition π encodes the cycle structure of σ. Moreover, the preceding bijection preserves the lattice structure.*

1.6. **Proof of the Marcenko-Pastur theorem.** Proof of the Marčenko-Pastur theorem We have now all the elements to present a short and elegant proof of Theorem 1.3.

Proof of Theorem 1.3. The proof will consist of three independent steps: computing the moments, at fixed d, of the random matrix W_d, letting $d \to \infty$ and computing the limiting moments, and finally identifying the probability measure having precisely these moments.

Step 1. Moment formula

We are interested, for any fixed dimensions d, s, in computing the p-th moment of the random matrix $W_d = GG^*$, where G is a $d \times s$ matrix with i.i.d. complex standard Gaussian random entries. To do this, we consider the diagram $\mathcal{D}$ corresponding to the random variable $\mathrm{Tr}(W_d^p)$. This diagram contains p pairs $(G, \bar{G})$ of Gaussian boxes, which are connected as in Figures 7 and 8. More precisely, the label corresponding to $\mathbb{C}^d$ which is attached to the i-th G-box is connected to the

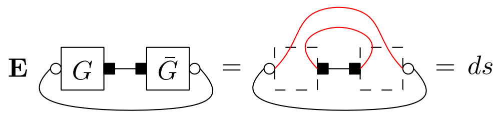

FIGURE 7. The first moment of a Wishart matrix using the graphical Wick calculus from Theorem 1.7. Round labels correspond to $\mathbb{C}^d$, while square labels correspond to $\mathbb{C}^s$.

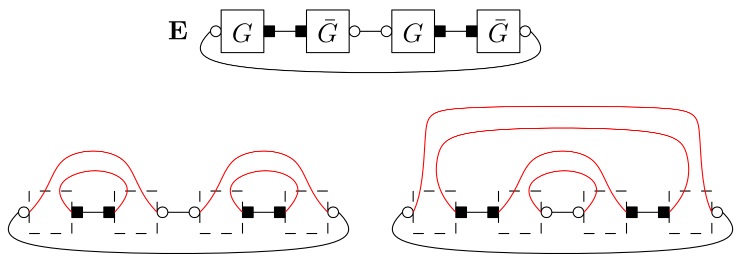

FIGURE 8. The second moment of a Wishart matrix using the graphical Wick calculus. On the top row, the diagram for $\mathbb{E}\,\mathrm{Tr}(W_d^2)$. On the bottom row, the two diagrams corresponding to the permutations id $= (1)(2)$, on the left, and (12), on the right. Their values are respectively ds^2 and d^2s.

corresponding label attached to the $(i-1)$-th $\bar{G}$-box. On the other hand, the label corresponding to $\mathbb{C}^s$ which is attached to the i-th G-box is connected to the corresponding label attached to the i-th $\bar{G}$-box. Using the graphical Wick formula from Theorem 1.7, we have

$$\mathbb{E}\,\mathrm{Tr}(W_d^p) = \mathbb{E}\mathcal{D} = \sum_{\alpha \in \mathcal{S}_p} \mathcal{D}_\alpha,$$

where $\mathcal{D}_\alpha$ is the removal diagram obtained by deleting the $G/\bar{G}$ boxed and connecting the labels according to the permutation α. It is clear that each diagram $\mathcal{D}_\alpha$ consists only of loops of two types: ones coming from round labels corresponding to $\mathbb{C}^d$ spaces, and others coming from square labels corresponding to $\mathbb{C}^s$ spaces. The number of loops of each type is the number of cycles in the permutation $\beta^{-1}\alpha$, where β encodes the initial wiring of the labels of each type; see Figures 7 and 8 for some examples. In conclusion, we have

$$\mathbb{E}\,\mathrm{Tr}(W_d^p) = \sum_{\alpha \in \mathcal{S}_p} d^{\#(\gamma^{-1}\alpha)} s^{\#\alpha}. \tag{1.4}$$

In the formula above, $\#(\cdot)$ is the number of cycles function, and γ is the full cycle permtuation

$$\gamma = (p\,(p-1)\cdots 3\,2\,1) \in \mathcal{S}_p.$$

Step 2. Asymptotic moments

Let us now consider the asymptotic regime we are interested in, $d \to \infty$ and $s \sim cd$, for some fixed parameter $c \in (0,\infty)$. Since the terms in (1.4) are all positive, we have

$$\mathbb{E}\,\mathrm{Tr}(W_d^p) \sim \sum_{\alpha \in \mathcal{S}_p} c^{\#\alpha} d^{\#(\gamma^{-1}\alpha)+\#\alpha}.$$

The dominating terms in the sum above are those maximizing the quantity $\#(\gamma^{-1}\alpha) + \#\alpha$ over the symmetric group. Using the properties of the distance function $|\cdot|$ on permutations from Lemma 1.8, we have

$$\#(\gamma^{-1}\alpha) + \#\alpha = 2p - (|\alpha| + |\gamma^{-1}\alpha|) \le 2p - |\gamma| = p+1,$$

where equality is attained iff α is a *geodesic* permutation (it saturates the triangle inequality $|\mathrm{id}^{-1}\alpha| + |\alpha^{-1}\gamma| \geq |\mathrm{id}^{-1}\gamma|$). We conclude that

$$\mathbb{E}\,\mathrm{Tr}(W_d^p) \sim d^{p+1} \sum_{\sigma \in NC(p)} c^{\#\sigma}.$$

Notice that considering only the dominating terms from the sum (1.4), indexed over all permutations, selects the ones for which the permutations are non-crossing partitions.

Step 3. The Marčenko-Pastur distribution

We are going to treat here the case $c = 1$; the general case is similar. We can rewrite the asymptotic moment formula as

$$\lim_{d\to\infty} \mathbb{E}\frac{1}{d}\,\mathrm{Tr}\left[(d^{-1}W_d)^p\right] = \mathrm{Cat_p}.$$

We claim that the unique probability measure μ having the Catalan numbers as moments is the one from (1.2):

$$\pi_1 = \frac{\sqrt{x(4-x)}}{2\pi x}\,\mathbf{1}_{(0,4)}(x)\,dx.$$

To show this, recall that the generating function of the Catalan number must be the moment generating function of μ:

$$M_\mu(z) = \sum_{p=0}^{\infty} z^p \int t^p d\mu = \frac{1-\sqrt{1-4z}}{2z},$$

where the relation above holds formally (as a power series in z), and analytically, in a small neighborhood of 0. The *Cauchy transform* of μ reads now

$$G_\mu(z) = \int \frac{1}{z-t} d\mu(t) = z^{-1}M_\mu(z^{-1}) = \frac{1-\sqrt{1-4z^{-1}}}{2},$$

which holds now on a neighborhood of the infinity in the complex plane. One recovers the density of μ via the *Stieltjes inversion formula*, which says that if we denote by

$$h_\varepsilon(t) := -\frac{1}{\pi}\Im G_\mu(t+i\varepsilon),$$

then

$$\frac{d\mu}{dt} = \lim_{\varepsilon\to 0} h_\varepsilon(t).$$

In our case, we recover $\mu = \pi_1$.

The uniqueness clame comes from the fact that π_1 is compactly supported, hence it satisfies the Carleman condition from Proposition 1.5. □

2. Lecture 2 — Partial transposition of random quantum states. Free probability

2.1. **Some elements of free probability theory.** We have studied random matrices in the previous lecture by their moments: the only properties of the ambient probability space we have used were the fact that the random variables have an algebra structure, and the existence of the expectation functional. We abstract out these notions in the following definition [NS06, Lecture 1].

Definition 2.1. *A* non-commutative probability space *is an algebra $\mathcal{A}$ with unit endowed with a tracial state φ. An element of $\mathcal{A}$ is called a (non-commutative) random variable.*

In these lectures we have already encountered the non-commutative probability space of *random matrices* $(\mathcal{M}_d(L^{\infty-}(\Omega,\mathbb{P})), \mathbb{E}[d^{-1}\,\mathrm{Tr}(\cdot)])$, where we use the standard notation $L^{\infty-}(\Omega,\mathbb{P}) = \cap_{p\geq 1}L^p(\Omega,\mathbb{P})$; the $L^{\infty-}$ space contains all random variables with moments of all orders. We shall encounter another example in Section 2.2.

In classical probability theory, the notion of *independence* of random variables plays a very important role; in particular, it allows to compute the joint distribution of independent random variables in terms of the marginal distributions (i.e. the distributions of the individual random variables). The notion of freeness is a non-commutative alternative to classical independence.

Definition 2.2. *Let $\mathcal{A}_1, \ldots, \mathcal{A}_k$ be subalgebras of $\mathcal{A}$ having the same unit as $\mathcal{A}$. They are said to be* free *if for all $a_i \in \mathcal{A}_{j_i}$ $(i = 1, \ldots, k)$ such that $\varphi(a_i) = 0$, one has*

$$\varphi(a_1 \cdots a_k) = 0$$

as soon as $j_1 \neq j_2$, $j_2 \neq j_3, \ldots, j_{k-1} \neq j_k$. Collections $S_1, S_2, \ldots$ of random variables are said to be free if the unital subalgebras they generate are free.

Let $(a_1, \ldots, a_k)$ be a k-tuple of selfadjoint random variables and let $\mathbb{C}\langle X_1, \ldots, X_k\rangle$ be the free $*$-algebra of non commutative polynomials on $\mathbb{C}$ generated by the k indeterminates $X_1, \ldots, X_k$. The *joint distribution* of the family $\{a_i\}_{i=1}^k$ is the linear form

$$\begin{aligned} \mu_{(a_1,\ldots,a_k)} : \mathbb{C}\langle X_1, \ldots, X_k\rangle &\to \mathbb{C} \\ P &\mapsto \varphi(P(a_1, \ldots, a_k)). \end{aligned}$$

In the case of a single, self-adjoint random variable x, if the moments of x coincide with those of a compactly supported probability measure μ, i.e.

$$\forall p \geq 1, \qquad \varphi(x^p) = \int t^p d\mu(t),$$

we say that x has distribution μ. The most important distribution in free probability theory is the semicircular distribution

$$\mu_{SC(0,1)} = \frac{\sqrt{4-x^2}}{2\pi} \mathbf{1}_{[-2,2]}(x)dx,$$

which is, for reasons we will not get into, the free world equivalent of the Gaussian distribution in classical probability (see [NS06, Lecture 8] for the details). A random variable x having distribution $\mu_{SC(0,1)}$ has the Catalan number for moments:

$$\varphi(x^p) = \begin{cases} \mathrm{Cat}_p := \frac{1}{p+1}\binom{2p}{p}, & \text{if } p \text{ is even} \\ 0, & \text{if } p \text{ is odd.} \end{cases}$$

More generally, if x has distribution $\mu_{SC(0,1)}$, we say that $y = \sigma x + m$ has distribution

$$\mu_{SC(m,\sigma^2)} = \frac{\sqrt{4\sigma^2 - (x-m)^2}}{2\pi\sigma^2} \mathbf{1}_{[m=2\sigma, m+2\sigma]}(x)dx. \tag{2.1}$$

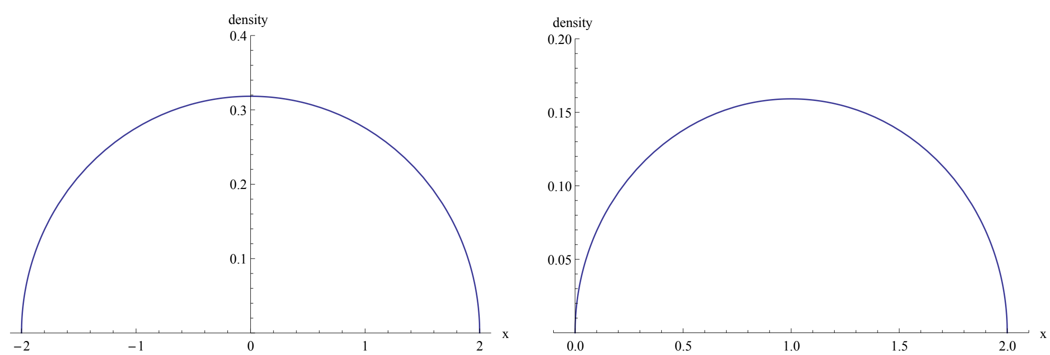

FIGURE 9. The density of the semicircular distributions $\mu_{SC(0,1)}$ (left) and $\mu_{SC(1,1/4)}$ (right).

Remark 2.3. *If the non-commutative random variable x has (standard) semicircular distribution, then x^2 has a free Poisson (or Marchenko-Pastur distribution) of parameter $c = 1$.*

Given a k-tuple $(a_1, \ldots, a_k)$ of free random variables such that the distribution of a_i is μ_{a_i}, the joint distribution $\mu_{(a_1,\ldots,a_k)}$ is uniquely determined by the μ_{a_i}'s. A family $(a_1^n, \ldots, a_k^n)_n$ of k-tuples of random variables is said to *converge in distribution* towards $(a_1, \ldots, a_k)$ iff for all $P \in \mathbb{C}\langle X_1, \ldots, X_k\rangle$, $\mu_{(a_1^n,\ldots,a_k^n)}(P)$ converges towards $\mu_{(a_1,\ldots,a_k)}(P)$ as $n \to \infty$. Sequences of random variables $(a_1^n)_n, \ldots, (a_k^n)_n$ are called *asymptotically free* as $n \to \infty$ iff the k-tuple $(a_1^n, \ldots, a_k^n)_n$ converges in distribution towards a family of free random variables.

Given two free random variables $a, b \in \mathcal{A}$, the distribution μ_{a+b} is uniquely determined by μ_a and μ_b. The free additive convolution of μ_a and μ_b is defined by $\mu_a \boxplus \mu_b = \mu_{a+b}$. When $x = x^* \in \mathcal{A}$, we identify μ_x with the spectral measure of x with respect to τ. The operation $\boxplus$ induces a binary operation on the set of probability measures on $\mathbb{R}$. Similarly, we write $\mu_a \boxminus \mu_b = \mu_{a-b}$.

2.2. **The full Fock space, free semicircular random variables.** We discuss now a more abstract non-commutative probability space, in which freeness appears naturally.

Definition 2.4. *Let H be a complex Hilbert space. The* full Fock space *over H is defined to be*

$$\mathcal{F}(H) = \bigoplus_{n=0}^{\infty} H^{\otimes n} = \mathbb{C}\Omega \oplus \bigoplus_{n=1}^{\infty} H^{\otimes n}.$$

The bounded operators on $\mathcal{F}(H)$, together with the vacuum state

$$\tau(X) = \langle \Omega, X\Omega\rangle$$

form a non-commutative probability space. We also define, for a vector $f \in H$, the creation and annihilation operators $\ell(h)$ and $\ell(h)^*$, defined as follows:

$$\ell(f)\Omega = f$$
$$\ell(f) f_1 \otimes \cdots \otimes f_n = f \otimes f_1 \otimes \cdots \otimes f_n$$

and

$$\ell(f)^*\Omega = 0$$
$$\ell(f)^* f_1 = \langle f, f_1\rangle \Omega$$
$$\ell(f)^* f_1 \otimes \cdots \otimes f_n = \langle f, f_1\rangle f_2 \otimes \cdots \otimes f_n.$$

The following theorem is taken from [NS06, Section 7], where it is proven in a more general form.

Theorem 2.5. *Let $f, g \in H$ be two orthogonal vectors. Then the non-commutative random variables $x = \ell(f) + \ell(f)^*$ and $y = \ell(g) + \ell(g)^*$ are semicircular and free.*

Proof. Let us first show that both x and y have semicircular distributions; moreover, without loss of generality, let us assume that $\|f\| = 1$, and task to show that x has $\mu_{SC(0,1)}$ distribution.

To do this, fix some moment order p, and consider $\tau(x^p)$:

$$\tau(x^p) = \sum_{w:[p]\to\{1,*\}} \langle \Omega, \ell(f)^{w(p)} \ell(f)^{w(p-1)} \cdots \ell(f)^{w(2)} \ell(f)^{w(1)} \Omega\rangle.$$

For each choice of the function w, the scalar product above is either 0 or 1; we have thus to count how many choices of w give 1. It is clear that a function w gives 1 iff $p = 2q$ is even, and the lattice path induced by w is a *Dyck* path. Recall that a Dyck path is a path in the lattice $\mathbb{Z}^2$, starting at $(0, 0)$, ending at $(p = 2q, 0)$, having $(1, \pm 1)$ steps, and, importantly, staying above the x-axis at all times; see Figure 10 for an example. The number of such paths is given by the Catalan numbers, and the first part of the proof is complete.

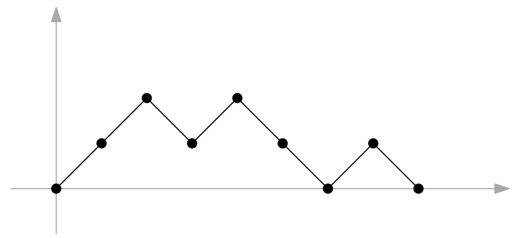

FIGURE 10. A Dyck path.

Let us now show that x and y are free. Let us first identify which elements in the algebra generated by $\{1, \ell(f)\}$ are traceless. It is easy enough to see that, after some cancellations of the form $\ell(f)^*\ell(f) = \|f\|^2$, the only such elements are of the form

$$\ell(f) \cdots \ell(f)\ell(f)^* \cdots \ell(f)^*,$$

where the product above is non empty. The conclusion follows by considering arbitrary alternating products of the above type for f and g, and by noting that whenever $\ell(f)^*\ell(g)$ appears, the end result is zero; hence, the $*$-algebras generated by $\ell(f)$ and $\ell(g)$ are free. The conclusion follows. $\square$

2.3. **The partial transpose of random quantum states.** We study here the asymptotical eigenvalue distribution of the partial transposition of random quantum states. Here, by random, we mean the probability distributions on $\mathcal{M}_d^{1,+}(\mathbb{C})$ known as the *induced measures*, which were introduced in [ŻS01]. A random quantum state ρ having the induced measure of parameters (d, s) is simply a normalized Wishart matrix of the same parameters, see also [Nec07, ŻPNC11]

$$\rho = \frac{W}{\operatorname{Tr} W} = \frac{GG^*}{\operatorname{Tr}(GG^*)}.$$

However, since we are just interested in the positivity of certain operators, it is enough to work with the cone of positive semidefinite matrices (and the Wishart matrices) instead of working with quantum states. Recall that the cone of separable matrices is defined as

$$\mathcal{SEP}_{d,n} = \{A \in \mathcal{M}_{dn} \ : \ A = \sum_i B_i \otimes C_i, \text{ where } B_i, C_i \geq 0\} \subseteq \mathcal{PSD}_{dn}.$$

The question whether a given mixed quantum state is separable or entangled has been proven to be an NP-hard one [Gur03]. To circumvent this worst-case intractability, *entanglement criteria* are used. These are efficiently computable conditions which are necessary for separability; in other words, an entanglement criterion is a (usually convex) super-set $\mathcal{X}_d$ of the set of separable states, for which the membership problem is efficiently solvable (see [AS15] for the number of such criteria needed to obtain a good approximation of the set of separable states). As in the previous section, from a probabilistic point of view, estimating the probability that a random quantum state (sampled from the induced ensemble) is an element of $\mathcal{X}_d$ is central.

In what follows we shall tackle this problem for one entanglement criterion in the framework of *thresholds*. Given a family $G_d \subseteq \mathcal{PSD}_d$ of convex cones, a pair of functions (s'_d, s''_d) is called a threshold for the family G_d if the following two properties are satisfied:

(1) If W_d is a sequence of Wishart random matrices of parameters (d, s_d) with $s_d \geq s''_d$, then

$$\lim_{d\to\infty} \mathbb{P}[W_d \in G_d] = 1.$$

(2) If W_d is a sequence of Wishart random matrices of parameters (d, s_d) with $s_d \leq s'_d$, then

$$\lim_{d\to\infty} \mathbb{P}[W_d \in G_d] = 0.$$

Let us start with the most used example, the *positive partial transpose* criterion (PPT). The PPT criterion has been introduced by Peres in [Per96]: if a positive semidefinite matrix $A \in \mathcal{M}_d \otimes \mathcal{M}_n$ is separable, then

$$A^\Gamma := [\mathrm{id} \otimes \mathrm{transp}](A) \geq 0.$$

Note that the positivity of A^Γ is equivalent to the positivity of $A^\daleth = [\mathrm{transp} \otimes \mathrm{id}](A)$, so it does not matter on which tensor factor the transpose application acts. We denote by $\mathcal{PPT}_{d,n}$ the PPT cone

$$\mathcal{PPT}_{d,n} := \{A \in \mathcal{M}_{dn} \,:\, A^\Gamma \geq 0\} \supseteq \mathcal{SEP}_{d,n}.$$

This necessary condition for separability has been shown to be also sufficient for qubit-qubit and qubit-qutrit systems ($dn \leq 6$) in [HHH96]; the result was a simple consequence of the fact that all the positive application from $\mathcal{M}_2$ to $\mathcal{M}_{2,3}$ are decomposable. These non trivial facts are due to Woronowocz [Wor76]. The PPT criterion for random quantum states has first been studied numerically in [ŽPBC07]. The analytic results in the following proposition are from [Aub12] (in the balanced case) and from [BN13] (in the unbalanced case); see also [FŚ13] for some improvements in the balanced case and the relation to meanders.

Proposition 2.6. *Consider a sequence $W_d \in \mathcal{M}_{dn_d}$ of random Wishart matrices of parameters (dn_d, cdn_d), where n_d is a function of d and c is a positive constant.*

In the balanced regime $n_d = d$, the (properly rescaled) empirical eigenvalue distribution of the matrices W_d^Γ converges to a semicircular measure $\mu_{SC(1,1/c)}$ of mean 1 and variance $1/c$, see (2.1). In particular, the threshold for the sets $\mathcal{PPT}_{d,d}$ ($d \to \infty$) is $c_0 = 4$.

In the unbalanced regime $n_d = n$ fixed, the (properly rescaled) empirical eigenvalue distribution of the matrices $d^{-1}W_d^\Gamma$ converges to a free difference of free Poisson distributions (see Section 2.1 for the definitions)

$$\pi_{cn(n+1)/2} \boxminus \pi_{cn(n-1)/2}.$$

In particular, the threshold for the sets $\mathcal{PPT}_{d,n}$ (n fixed, $d \to \infty$) is

$$c_0 = 2 + 2\sqrt{1 - \frac{1}{n^2}}.$$

Proof. We are going to sketch the proof of the convergence result in the unbalanced case; for the balanced case, see [Aub12] and for the threshold in the unbalanced case, see [BN13, Section 6].

Using again the graphical Wick formula, one can find the following expression for the (unnormalized) moments of W_d^Γ:

$$\mathbb{E}\,\mathrm{Tr}[({W_d}^\Gamma)^p] = \sum_{\alpha \in \mathcal{S}_p} s^{\#\alpha} d^{\#(\gamma^{-1}\alpha)} n^{\#(\gamma\alpha)}.$$

Using the fact that, for every noncrossing partition $\sigma \in NC(p)$, denoting by $e(\sigma)$ the number of blocks of even size of σ, we have $1 + \mathrm{e}(\sigma) = \#(\sigma\gamma)$, we arrive at the formula

$$\begin{aligned}
\mathbb{E}(dn)^{-1}\,\mathrm{Tr}[(d^{-1}{W_d}^\Gamma)^p] &\sim \sum_{\sigma \in NC(p)} n^{\#\sigma + e(\sigma)} c^{\#\sigma} \\
&\sim \sum_{\sigma \in NC(p)} \prod_{b \in \sigma} cn^{1 + \mathbf{1}_{|b| \text{ is even}}} \\
&\sim \sum_{\sigma \in NC(p)} \prod_{b \in \sigma} \left(\frac{cn(n+1)}{2} + \frac{cn(n-1)}{2}(-1)^{|b|} \right).
\end{aligned}$$

We can now identify the free difference of free Poisson operators using the free cumulant approach of [NS06]: the free cumulant of order p of the limiting measure is

$$\frac{cn(n+1)}{2} + \frac{cn(n-1)}{2}(-1)^{|b|}.$$

□

Remark 2.7. *The computation of the limiting distribution of in the unbalanced case performed above was done using the method of moments. A more general approach, allowing to answer the same question for general maps and general matrix distributions, was provided in* [ANV16] *using operator valued free probability theory.*

Remark 2.8. *The value of the threshold in the theorem above has a practical significance: if one considers a random pure quantum state on $\mathcal{H} = \mathbb{C}^d \otimes \mathbb{C}^n \otimes \mathbb{C}^{cdn}$, takes the partial trace on the third subsystem, and the partial transposition on the second subsystem, then the resulting matrix is positive semidefinite if $c > c_0$, and has negative eigenvalues if $c < c_0$, with large probability as n is fixed and $d \to \infty$.*

3. Lecture 3 — Random quantum channels and their minimum output entropy

3.1. **Quantum channels, minimum output entropies, additivity.** In Quantum Information Theory, a *quantum channel* is the most general transformation of a quantum system. Quantum channels generalize the unitary evolution of isolated quantum systems to *open quantum systems*. Mathematically, we recall that a quantum channel is a linear completely positive trace preserving map Φ from $\mathcal{M}_n(\mathbb{C})$ to itself. The trace preservation condition is necessary since quantum channels should map density matrices to density matrices. The complete positivity condition can be stated as

$$\forall d \geq 1, \quad \Phi \otimes I_d : \mathcal{M}_{nd}(\mathbb{C}) \to \mathcal{M}_{nd}(\mathbb{C}) \text{ is a positive map.}$$

The following three characterizations of quantum channels turn out to be very useful; they are due to Stinespring [Sti55] and Choi [Cho75].

Proposition 3.1. *A linear map $\Phi : \mathcal{M}_n(\mathbb{C}) \to \mathcal{M}_n(\mathbb{C})$ is a quantum channel if and only if one of the following three equivalent conditions holds.*

(1) (Stinespring dilation) *There exists a finite dimensional Hilbert space $\mathcal{K} = \mathbb{C}^d$, a density matrix $Y \in \mathcal{M}_d^{1,+}(\mathbb{C})$ and an unitary operator $U \in \mathcal{U}_{nd}$ such that*

$$\Phi(X) = \operatorname{Tr}_{\mathcal{K}} \left[U(X \otimes Y)U^* \right], \quad \forall X \in \mathcal{M}_n(\mathbb{C}). \tag{3.1}$$

(2) (Kraus decomposition) *There exists an integer k and matrices $L_1, \ldots, L_k \in \mathcal{M}_n(\mathbb{C})$ such that*

$$\Phi(X) = \sum_{i=1}^k L_i X L_i^*, \quad \forall X \in \mathcal{M}_n(\mathbb{C}).$$

and

$$\sum_{i=1}^k L_i^* L_i = I_n.$$

(3) (Choi matrix) *The following matrix, called the* Choi matrix *of Φ*

$$\mathcal{M}_{n^2}(\mathbb{C}) \ni C_\Phi = [\mathrm{id} \otimes \Phi](\Omega_d) = \sum_{i,j=1}^n E_{ij} \otimes \Phi(E_{ij}) \tag{3.2}$$

is positive-semidefinite and satisfies $[\mathrm{id} \otimes \operatorname{Tr}](C_\Phi) = I_n$.

It can be shown that the dimension of the ancilla space $\mathcal{K}$ in the Stinespring dilation theorem can be chosen $d = \dim \mathcal{K} = n^2$ and that the state Y can always be considered to be a rank one projector. A similar result holds for the number of Kraus operators: one can always find a decomposition with $k = n^2$ operators.

As in classical information theory [Sha48], entropic quantities play a very important role in quantum information theory. We define next the quantities of interest for the current work. Let

$\Delta_k = \{x \in \mathbb{R}_+^k \mid \sum_{i=1}^k x_i = 1\}$ be the $(k-1)$-dimensional probability simplex. For a positive real number $p \in (0,1) \cup (1,\infty)$, define the *Rényi entropy of order* p of a probability vector $x \in \Delta_k$ to be

$$H_p(x) = \frac{1}{1-p} \log \sum_{i=1}^k x_i^p.$$

Since $\lim_{p\to 1} H_p(x)$ exists, we define the *Shannon entropy* of x to be this limit, namely:

$$H(x) = H_1(x) = -\sum_{i=1}^k x_i \log x_i.$$

We also define the values for the parameters $p = 0, \infty$:

$$\begin{aligned} H_0(x) &= \log \#\{i \,:\, x_i \neq 0\} \\ H_\infty(x) &= -\log \|x\|_\infty. \end{aligned}$$

We extend these definitions to density matrices by functional calculus: for $\rho \in \mathcal{M}_n^{1,+}(\mathbb{C})$, we put

$$\begin{aligned} H_0(\rho) &= \log \operatorname{rk}(\rho) \\ H_p(\rho) &= \frac{1}{1-p} \log \operatorname{Tr} \rho^p \qquad p \in (0,1) \cup (1,\infty) \\ H(\rho) &= H_1(\rho) = -\operatorname{Tr} \rho \log \rho \\ H_\infty(\rho) &= -\log \|\rho\|_\infty. \end{aligned}$$

Of special interest for the computation of capacities of quantum channels to transmit classical information are the following quantities, called the *minimum output entropies* of the channel. As the Rényi entropies, they are indexed by some positive real parameter p

$$H_p^{\min}(\Phi) = \min_{\rho \in \mathcal{M}_n^{1,+}(\mathbb{C})} H_p(\Phi(\rho)). \tag{3.3}$$

The following theorem summarizes some of the most important breakthroughs in quantum information theory in the last decade. It is based in particular on the papers [Has09, HW08], and concerns the minimum output entropies of quantum channels, defined in (3.3). The result came as a surprise to the community, since additivity (i.e. equality in (3.4)) was shown to hold for many examples of quantum channels.

Theorem 3.2. *For every $p \in [1,\infty]$, there exist quantum channels Φ and Ψ such that*

$$H_p^{\min}(\Phi \otimes \Psi) < H_p^{\min}(\Phi) + H_p^{\min}(\Psi). \tag{3.4}$$

Except for some particular cases ($p > 4.79$, [WH02] and $p > 2$, [GHP10]), the proof of this theorem uses the random method, i.e. the channels Φ, Ψ are random channels, and the above inequality occurs with non-zero probability. At this moment, we are not aware of any explicit, non-random choices for Φ, Ψ in the case $1 \leq p \leq 2$.

The additivity property for the minimum output entropy $H_{\min}(\cdot)$ was related in [Sho04] to the additivity of another important entropic quantity, the *Holevo quantity*

$$\chi(\Phi) = \max_{\{p_i, X_i\}} \left[H\left(\sum_i p_i \Phi(X_i)\right) - \sum_i p_i H(\Phi(X_i))\right].$$

The regularized Holevo quantity provides [Hol98, SW97] the *classical capacity* of a quantum channel Φ, i.e. the maximum rate at which classical information can be reliably sent through the noisy channel. The importance of the additivity question stems mainly from the difficulty in computing the above regularized quantity. Indeed, if additivity holds, there would be no need for regularization, and the classical capacity of the channel Φ would simply be equal to the Holevo (or one-shot) capacity $\chi(\Phi)$.

3.2. **Random quantum channels.** As discussed before, all the known counter-examples from Theorem 3.2, at least in the case $p = 1$ or p close to 1, come from random constructions. We will define now what we mean by a *random quantum channel*. Note that the model described below is just one of many possible. It has the merit of providing the *largest* violations for the inequality in Theorem 3.2, as well as the lowest output dimensions, see [BCN16].

We consider the probability distributions on quantum channels (i.e. trace preserving, completely positive maps)

$$\Phi : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_k(\mathbb{C}), \quad \Phi(X) = [\mathrm{id}_k \otimes \mathrm{Tr}_n](VXV^*), \tag{3.5}$$

where $V : \mathbb{C}^d \to \mathbb{C}^k \otimes \mathbb{C}^n$ is a *random Haar isometry*; note that Φ is a quantum channel, by the Stinespring dilation result from Proposition 3.1. By a random Haar isometry we mean the unique invariant probability measure on the Stiefel manifold $\mathcal{V}_d(\mathbb{C}^{nk})$. A random isometry V can also be seen as a truncation of a random Haar unitary $U \in \mathcal{U}_{nk}$.

It has been shown by many authors [FK10, ASW11, CN11b, BCN16] that sequences (Φ_n) of random quantum channels violate asymptotically, with probability one, the additivity relation from Theorem 3.2; most of the examples use the asymptotic regime $d_n \sim tkn$ for suitable fixed output dimension k and input space ratio $t \in (0, 1)$.

All the counter examples use the so-called Hayden-Winter trick, that is letting $\Psi = \bar{\Phi}$ and lower bounding the left hand side of the additivity relation (3.4) by the output of the maximally entangled state Ω_d. The details of this lower bound can be found in [CN10b], we shall not discuss them here. We study the behavior of the right hand side of (3.4) in the next subsection.

3.3. **Strong convergence and the (t)-norm.** Strong convergence and the (t)-norm

We have seen in Section 2.2 how freeness appears naturally in the full Fock space setting. Another very important situation where freeness manifests itself is the asymptotic theory of random matrices. The following result was one of Voiculescu's breakthroughs [Voi98].

Theorem 3.3. *Let (A_n) and (B_n) be sequences of $n \times n$ matrices such that A_n and B_n converge in distribution (with respect to $n^{-1}\mathrm{Tr}$) for $n \to \infty$. Furthermore, let (U_n) be a sequence of Haar unitary $n \times n$ random matrices. Then, A_n and $U_n B_n U_n^*$ are* asymptotically free *for $n \to \infty$.*

If A_n, B_n are matrices of size n, whose spectra converge towards μ_a, μ_b, the spectrum of $A_n + U_n B_n U_n^*$ converges to $\mu_a \boxplus \mu_b$; for the definition of the free additive convolution $\boxplus$, see Section 2.1. Similarly, If A_n, B_n are matrices of size n such that $A_n \geq 0$, whose spectra converge towards μ_a, μ_b, the spectrum of $A_n^{1/2} U_n B_n U_n^* A_n^{1/2}$ converges to $\mu_a \boxtimes \mu_b$; the operation $\boxtimes$ is called the *free multiplicative convolution*.

Actually, if the matrices A_n and B_n have well-behaved eigenvalues, not only do the moments of $A_n + U_n B_n U_n^*$ converge to those of $a + b$ with a, b free, but we also have a norm convergence, called *strong convergence*

$$\text{almost surely, } \lim_{n \to \infty} \|A_n + U_n B_n U_n^*\|_\infty = \|a + b\|.$$

Note that in the above setting, we need to consider $a, b \in (\mathcal{A}, \tau)$ a C^* non-commutative probability space. This result has been shown for GUE matrices in [HT05], and further extended in [Mal12, CM14].

Let us now consider an example, the truncation of random matrices. Let $P_n \in \mathcal{M}_n$ a projection of rank $n/2$; its eigenvalues are 0 and 1, with multiplicity $n/2$. Hence, the distribution of P_n converges, when $n \to \infty$, to the Bernoulli probability measure $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. Let $C_n \in \mathcal{M}_{n/2}$ be the top $n/2 \times n/2$ corner of $U_n P_n U_n^*$, with U_n a Haar random unitary matrix. Up to zero blocks, $C_n = Q_n(U_n P_n U_n^*)Q_n$, where Q_n is the diagonal orthogonal projection on the first $n/2$ coordinates of $\mathbb{C}^n$. The distribution of Q_n converges to $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. A direct computation in free probability theory tells us that the distribution of C_n will converge to

$$\left[\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1\right] \boxtimes \left[\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1\right] = \frac{1}{\pi\sqrt{x(1-x)}}\mathbf{1}_{[0,1]}(x)dx,$$

which is the arcsine distribution.

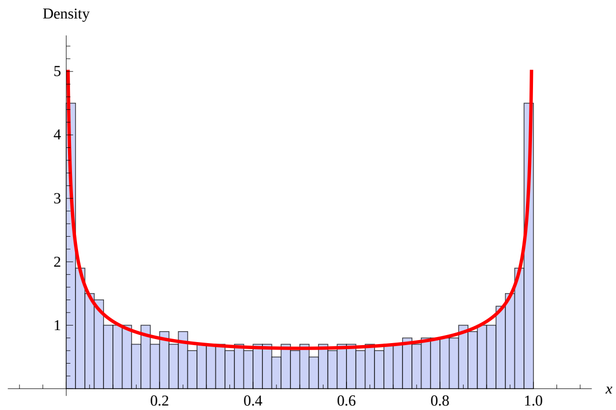

Figure 11. Histogram of eigenvalues of a truncated randomly rotated projector of relative rank 1/2 and size $n = 4000$; in red, the density of the arcsine distribution.

We introduce now a new norm, which will play a crucial role in computing MOE for random quantum channels.

Definition 3.4. *For a positive integer k, embed $\mathbb{R}^k$ as a self-adjoint real subalgebra $\mathcal{R}$ of a C^*-ncps $(\mathcal{A}, \tau)$, so that $\tau(x) = (x_1 + \cdots + x_k)/k$. Let p_t be a projection of rank $t \in (0, 1]$ in $\mathcal{A}$, free from $\mathcal{R}$. On the real vector space $\mathbb{R}^k$, we introduce the following norm, called the (t)-norm:*

$$\|x\|_{(t)} := \|p_t x p_t\|_\infty,$$

where the vector $x \in \mathbb{R}^k$ is identified with its image in $\mathcal{R}$.

One can show that $\|\cdot\|_{(t)}$ is indeed a norm, which is permutation invariant. When $t > 1 - 1/k$, $\|\cdot\|_{(t)} = \|\cdot\|_\infty$ on $\mathbb{R}^k$, and we can show that $\lim_{t\to 0^+} \|x\|_{(t)} = k^{-1}|\sum_i x_i|$.

For vectors x with non-negative elements, $\|x\|_{(t)}$ is the right end of the support of the probability measure

$$\left[\sum_i \delta_{x_i}\right] \boxtimes \left[(1-t)\delta_0 + t\delta_1\right].$$

Let us now look at an example of a computation for the t-norm, in the case where x has just components taking two values.

Theorem 3.5. *In $\mathbb{C}^n$, choose at random according to the Haar measure two independent subspaces V_n and V'_n of respective dimensions $q_n \sim sn$ and $q'_n \sim tn$ where $s, t \in (0, 1]$. Let P_n (resp. P'_n) be the orthogonal projection onto V_n (resp. V'_n). Then,*

$$\lim_n \|P_n P'_n P_n\|_\infty = \varphi(s,t) = \sup\operatorname{supp}((1-s)\delta_0 + s\delta_1) \boxtimes ((1-t)\delta_0 + t\delta_1),$$

with

$$\varphi(s,t) = \begin{cases} s + t - 2st + 2\sqrt{st(1-s)(1-t)} & \text{if } s + t < 1; \\ 1 & \text{if } s + t \geq 1. \end{cases} \tag{3.6}$$

Hence, we can compute

$$\|\underbrace{1, \cdots, 1}_{j \text{ times}}, \underbrace{0, \cdots, 0}_{k-j \text{ times}}\|_{(t)} = \varphi(j/k, t).$$

3.4. **The MOE of a (large) random quantum channel.** From now on, we shall abuse notation: recall that we are interested in random isometries $V : \mathbb{C}^d \to \mathbb{C}^k \otimes \mathbb{C}^n$. Since the quantities $H_p^{\min}$ only depend on the range of V, also write $V = \operatorname{ran} V$. For a subspace $V \subset \mathbb{C}^k \otimes \mathbb{C}^n$, define

$$H_p^{\min}(V) = \min_{y \in V, \|y\|=1} H_p(y),$$

the minimal p-entropy of vectors in V; for a channel as in (3.5), we have

$$H_p^{\min}(V) = H_p^{\min}(\Phi).$$

For a subspace $V \subset \mathbb{C}^k \otimes \mathbb{C}^n$ of dimension $\dim V = d$, define the set eigen-/singular values or Schmidt coefficients

$$K_V = \{\lambda(x) \,:\, x \in V, \|x\| = 1\}.$$

Our goal is to understand K_V, and thus, the particular statistic $H_p^{\min}(V)$. The set K_V is a compact subset of the ordered probability simplex $\Delta_k^{\downarrow}$, having the following properties

- Local invariance: $K_{(U_1 \otimes U_2)V} = K_V$, for unitary matrices $U_1 \in \mathcal{U}(k)$ and $U_2 \in \mathcal{U}(n)$.
- Monotonicity: if $V_1 \subset V_2$, then $K_{V_1} \subset K_{V_2}$.
- Recovering minimum entropies:

$$H_p^{\min}(\Phi) = H_p^{\min}(V) = \min_{\lambda \in K_V} H_p(\lambda).$$

Example 3.6. *The anti-symmetric subspace provides the (explicit) counter-example for the additivity of the p-Rényi entropy* [GHP10]*. Let $k = n$ and put $V = \Lambda^2(\mathbb{C}^n)$. The subspace V is almost half of the total space:* $\dim V = n(n-1)/2$*. Antisymmetric vectors in V are typically*

$$V \ni x = \frac{1}{\sqrt{2}}(e \otimes f - f \otimes e).$$

Since singular values of vectors in V come in pairs, the least entropy vector in V is as above, with $e \perp f$ and $H(x) = \log 2$. Thus, $H^{\min}(V) = \log 2$ and one can show that

$$K_V = \{(\lambda_1, \lambda_1, \lambda_2, \lambda_2, \ldots) \in \Delta_n \,:\, \lambda_i \geq 0, \sum_i \lambda_i = 1/2\}.$$

Problem 3.7. *Find explicit (i.e. non-random) examples of subspaces $V \subset \mathbb{C}^k \otimes \mathbb{C}^n$ with*

- *large* $\dim V$*;*
- *large $H^{\min}(V)$.*

We state now the main result of this section, a characterization of the set K_V, for a random isometry V of large size [BCN12]. Recall that we are interested in random isometries/subspaces in the following asymptotic regime: k fixed, $n \to \infty$, and $d \sim tkn$, for a fixed parameter $t \in (0, 1)$.

Theorem 3.8. *For a sequence of uniformly distributed random subspaces V_n, the set K_{V_n} of singular values of unit vectors from V_n converges (almost surely, in the Hausdorff distance) to a deterministic, convex subset $K_{k,t}$ of the probability simplex Δ_k*

$$K_{k,t} := \{\lambda \in \Delta_k \mid \forall x \in \Delta_k, \langle \lambda, x \rangle \leq \|x\|_{(t)}\}.$$

The theorem above allows for the computation of the asymptotic behavior of channel statistics which are related to the output set of random quantum channels. The problem is reduced to the analogous question on the deterministic set $K_{k,t}$. The case of the MOE was treated in [BCN16], where the following corollary was proved.

Corollary 3.9. *For all $p \geq 1$,*

$$\lim_{n \to \infty} H_p^{min}(\Phi) = \min_{\lambda \in K_{k,t}} H_p(\lambda) = H_p(a, b, b, \ldots, b),$$

where a, b do not depend on p, $b = (1-a)/(k-1)$ and $a = \varphi(1/k, t)$, where the function φ was defined in (3.6).

Proof of Theorem 3.8. The statement in the proof concerns the duals of the set $K_{k,t}$, so we are going to consider how far does the set K_{V_n} extend into some given direction $a \in \Delta_k^{\downarrow}$.

Let us start by considering the direction $a = (1, 0, \ldots, 0)$. We would like to compute the largest maximal singular value $\max_{x \in V, \|x\|=1} \lambda_1(x)$ of vectors from the subspace V ?

$$\begin{aligned}
\max_{x \in V, \|x\|=1} \lambda_1(x) &= \max_{x \in V, \|x\|=1} \|[\mathrm{id}_k \otimes \mathrm{Tr}_n] P_x\|_\infty \\
&= \max_{x \in V, \|x\|=1} \max_{y \in \mathbb{C}^k, \|y\|=1} \mathrm{Tr}\left[([\mathrm{id}_k \otimes \mathrm{Tr}_n] P_x) \cdot P_y\right] \\
&= \max_{x \in V, \|x\|=1} \max_{y \in \mathbb{C}^k, \|y\|=1} \mathrm{Tr}\left[P_x \cdot P_y \otimes \mathrm{I}_n\right] \\
&= \max_{y \in \mathbb{C}^k, \|y\|=1} \max_{x \in V, \|x\|=1} \mathrm{Tr}\left[P_x \cdot P_y \otimes \mathrm{I}_n\right] \\
&= \max_{y \in \mathbb{C}^k, \|y\|=1} \|P_V \cdot P_y \otimes \mathrm{I}_n \cdot P_V\|_\infty .
\end{aligned}$$

For fixed y, P_V and $P_y \otimes \mathrm{I}_n$ are independent projectors of relative ranks t and $1/k$ respectively. Thus, almost surely,

$$\begin{aligned}
\|P_V \cdot P_y \otimes \mathrm{I}_n \cdot P_V\|_\infty &\to \| \left((1-t)\delta_0 + t\delta_1\right) \boxtimes \left((1-1/k)\delta_0 + 1/k\delta_1\right) \| \\
&= \varphi(t, 1/k) = \|(1, 0, \ldots, 0)\|_{(t)} .
\end{aligned}$$

The computation above shows that the asymptotic behavior of $\|P_V \cdot P_y \otimes \mathrm{I}_n \cdot P_V\|_\infty$ is independent of y. We can thus take the max over y at no cost, by considering a finite net of y's, since k is fixed.

To get the full result $\limsup_{n\to\infty} K_{V_n} \subset K_{k,t}$, we have to consider $\langle \lambda, a \rangle$ for all directions a; the computations are similar.

The inclusion $\liminf_{n\to\infty} K_{V_n} \supset K_{k,t}$, is much easier, and follows from the convergence in distribution/moments.

□

Acknowledgments. These notes were prepared for a series of lectures I gave at Seoul National University — I would like to thank Hun Hee Lee for the invitation and for his hospitality. I would also like to thank Seung-Hyeok Kye for the invitation to a conference in Daejeon, one week prior to the lectures. The author's research has been supported by the ANR projects RMTQIT ANR-12-IS01-0001-01 and StoQ ANR-14-CE25-0003-01.

References

[AGZ10] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices.* Number 118. Cambridge University Press, 2010. 3

[ANV16] Octavio Arizmendi, Ion Nechita, and Carlos Vargas. On the asymptotic distribution of block-modified random matrices. *Journal of Mathematical Physics*, 57(1):015216, 2016. 14

[AS15] Guillaume Aubrun and Stanislaw Szarek. Dvoretzky's theorem and the complexity of entanglement detection. *arXiv preprint arXiv:1510.00578*, 2015. 12

[ASW11] Guillaume Aubrun, Stanisław Szarek, and Elisabeth Werner. Hastings's additivity counterexample via Dvoretzky's theorem. *Communications in mathematical physics*, 305(1):85–97, 2011. 16

[Aub12] Guillaume Aubrun. Partial transposition of random states and non-centered semicircular distributions. *Random Matrices: Theory and Applications*, 1(02):1250001, 2012. 13

[BCN12] Serban Belinschi, Benoît Collins, and Ion Nechita. Eigenvectors and eigenvalues in a random subspace of a tensor product. *Inventiones mathematicae*, 190(3):647–697, 2012. 18

[BCN16] Serban T Belinschi, Benoit Collins, and Ion Nechita. Almost one bit violation for the additivity of the minimum output entropy. *Communications in Mathematical Physics*, 341(3):885–909, 2016. 16, 18

[Bia97] Philippe Biane. Some properties of crossings and partitions. *Discrete Mathematics*, 175(1):41–53, 1997. 7

[BN13] Teodor Banica and Ion Nechita. Asymptotic eigenvalue distributions of block-transposed wishart matrices. *Journal of Theoretical Probability*, 26(3):855–869, 2013. 13

[BS10] Zhidong Bai and Jack W Silverstein. *Spectral analysis of large dimensional random matrices*, volume 20. Springer, 2010. 3

[CGGPG13] Benoît Collins, Carlos E Gonzalez Guillen, and David Pérez García. Matrix product states, random matrix theory and the principle of maximum entropy. *Communications in Mathematical Physics*, 320(3):663–677, 2013. 4

[Cho75] Man-Duen Choi. Completely positive linear maps on complex matrices. *Linear algebra and its applications*, 10(3):285–290, 1975. 14

[CM14] Benoit Collins and Camille Male. The strong asymptotic freeness of haar and deterministic matrices. *Annales scientifiques de l'Ecole Normale Supérieure*, 47:147–163, 2014. 16

[CN10a] Benoît Collins and Ion Nechita. Eigenvalue and entropy statistics for products of conjugate random quantum channels. *Entropy*, 12(6):1612–1631, 2010. 4

[CN10b] Benoît Collins and Ion Nechita. Random quantum channels I: graphical calculus and the Bell state phenomenon. *Communications in Mathematical Physics*, 297(2):345–370, 2010. 4, 16

[CN11a] Benoît Collins and Ion Nechita. Gaussianization and eigenvalue statistics for random quantum channels (III). *The Annals of Applied Probability*, pages 1136–1179, 2011. 4

[CN11b] Benoît Collins and Ion Nechita. Random quantum channels II: Entanglement of random subspaces, Rényi entropy estimates and additivity problems. *Advances in Mathematics*, 226(2):1181–1201, 2011. 16

[CN16] Benoit Collins and Ion Nechita. Random matrix techniques in quantum information theory. *Journal of Mathematical Physics*, 57(1), 2016. 2

[CNŻ10] Benoît Collins, Ion Nechita, and Karol Życzkowski. Random graph states, maximal flow and fuss–catalan distributions. *Journal of Physics A: Mathematical and Theoretical*, 43(27):275303, 2010. 4

[CNŻ13] Benoît Collins, Ion Nechita, and Karol Życzkowski. Area law for random graph states. *Journal of Physics A: Mathematical and Theoretical*, 46(30):305302, 2013. 4

[Coe10] Bob Coecke. Quantum picturalism. *Contemporary physics*, 51(1):59–83, 2010. 4

[FK10] Motohisa Fukuda and Christopher King. Entanglement of random subspaces via the hastings bound. *Journal of Mathematical Physics*, 51(4):042201, 2010. 16

[FŚ13] Motohisa Fukuda and Piotr Śniady. Partial transpose of random quantum states: Exact formulas and meanders. *Journal of Mathematical Physics*, 54(4):042202, 2013. 4, 13

[GHP10] Andrzej Grudka, Michał Horodecki, and Łukasz Pankowski. Constructive counterexamples to the additivity of the minimum output rényi entropy of quantum channels for all p¿ 2. *Journal of Physics A: Mathematical and Theoretical*, 43(42):425304, 2010. 15, 18

[Gur03] Leonid Gurvits. Classical deterministic complexity of edmonds' problem and quantum entanglement. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 10–19. ACM, 2003. 12

[Has09] Matthew B Hastings. Superadditivity of communication capacity using entangled inputs. *Nature Physics*, 5(4):255–257, 2009. 15

[HHH96] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223(1):1–8, 1996. 13

[HLSW04] Patrick Hayden, Debbie Leung, Peter W Shor, and Andreas Winter. Randomizing quantum states: Constructions and applications. *Communications in Mathematical Physics*, 250(2):371–391, 2004. 2

[Hol98] Alexander S Holevo. The capacity of quantum channel with general signal states. *IEEE Trans. Inform. Theory*, 44(1):269 273, 1998. 15

[HT05] Uffe Haagerup and Steen Thorbjørnsen. A new application of random matrices: : $ext(c^*_{red}(\mathbb{F}_2))$ is not a group. *Annals of Mathematics*, pages 711–775, 2005. 16

[HW08] Patrick Hayden and Andreas Winter. Counterexamples to the maximal p-norm multiplicativity conjecture for all p¿ 1. *Communications in mathematical physics*, 284(1):263–280, 2008. 15

[Lan15] Cécilia Lancien. k-extendibility of high-dimensional bipartite quantum states. *arXiv preprint arXiv:1504.06459*, 2015. 4

[Mal12] Camille Male. The norm of polynomials in large random and deterministic matrices. *Probability Theory and Related Fields*, 154(3-4):477–532, 2012. 3, 16

[MP67] Vladimir A Marčenko and Leonid Andreevich Pastur. Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics*, 1(4):457–483, 1967. 3

[Nec07] Ion Nechita. Asymptotics of random density matrices. *Annales Henri Poincaré*, 8(8):1521–1538, 2007. 12

[NS06] Alexandru Nica and Roland Speicher. *Lectures on the combinatorics of free probability*, volume 13. Cambridge University Press, 2006. 3, 7, 9, 10, 11, 13

[Pag93] Don N Page. Average entropy of a subsystem. *Physical review letters*, 71(9):1291, 1993. 2

[Pen05] Roger Penrose. *The road to reality, Alfred A*. Knopf, New York, 2005. 4

[Per96] Asher Peres. Separability criterion for density matrices. *Physical Review Letters*, 77(8):1413, 1996. 13

[Sha48] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. 14

[Sho04] Peter W Shor. Equivalence of additivity questions in quantum information theory. *Communications in Mathematical Physics*, 246(3):453–472, 2004. 15

[Sti55] W Forrest Stinespring. Positive functions on c*-algebras. *Proceedings of the American Mathematical Society*, 6(2):211–216, 1955. 14

[SW97] Benjamin Schumacher and Michael D Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56(1):131, 1997. 15

[VDN92] Dan V Voiculescu, Ken J Dykema, and Alexandru Nica. *Free random variables*. Number 1. American Mathematical Soc., 1992. 2

[Voi98] Dan Voiculescu. A strengthened asymptotic freeness result for random matrices with applications to free entropy. *International Mathematics Research Notices*, 1998(1):41–63, 1998. 16

[WH02] Reinhard F Werner and Alexander S Holevo. Counterexample to an additivity conjecture for output purity of quantum channels. *Journal of Mathematical Physics*, 43(9):4353–4357, 2002. 15

[Wig55] Eugene P Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, pages 548–564, 1955. 2

[Wis28] John Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, pages 32–52, 1928. 2, 3

[Wor76] Stanisław Lech Woronowicz. Positive maps of low dimensional matrix algebras. *Reports on Mathematical Physics*, 10(2):165–183, 1976. 13

[ŽPBC07] Marko Žnidarič, Tomaž Prosen, Giuliano Benenti, and Giulio Casati. Detecting entanglement of random states with an entanglement witness. *Journal of Physics A: Mathematical and Theoretical*, 40(45):13787, 2007. 13

[ŻPNC11] Karol Życzkowski, Karol A Penson, Ion Nechita, and Benoit Collins. Generating random density matrices. *Journal of Mathematical Physics*, 52(6):062201, 2011. 12

[ŻS01] Karol Życzkowski and Hans-Jürgen Sommers. Induced measures in the space of mixed quantum states. *Journal of Physics A: Mathematical and General*, 34(35):7111, 2001. 2, 12

CNRS, Laboratoire de Physique Théorique, IRSAMC, Université de Toulouse, UPS, F-31062 Toulouse, France

E-mail address: nechita@irsamc.ups-tlse.fr