arXiv:1011.0275v1 [math.PR] 1 Nov 2010

PARTIAL TRANSPOSITION OF RANDOM STATES AND NON-CENTERED SEMICIRCULAR DISTRIBUTIONS

GUILLAUME AUBRUN

Abstract. Let W be a Wishart random matrix of size $d^2 \times d^2$, considered as a block matrix with $d \times d$ blocks. Let Y be the matrix obtained by transposing each block of W. We prove that the empirical eigenvalue distribution of Y approaches a non-centered semicircular distribution when $d \to \infty$. The proofs are based on the moments method.

This matrix model is relevant to Quantum Information Theory and corresponds to the partial transposition of a random induced state. We show for example that a mixed state on $\mathbf{C}^d \otimes \mathbf{C}^d$, obtained after tracing out a random pure state over some ancilla, is typically PPT (hence undistillable) for large d, whenever the dimension of the ancilla exceeds Cd^2 for some constant C (conjecturally, $C = 4$).

1. Introduction

In the recent years, several connections were established between Random Matrix Theory and Quantum Information Theory. It turns out that random operators, and the random constructions they induce, can be used to construct quantum channels with an unexpected behavior, violating some natural conjectures (the most prominent example being Hastings's counterexample to additivity conjectures [8]). Random matrices appear to be a sharp tool in order to understand the high-dimensional objects from Quantum Information Theory.

In this spirit, we study here a model of random matrices motivated by Quantum Information Theory. We use the most standard method in Random Matrix Theory: the moments method. As in the well-known case of Wigner matrices, the limiting distribution which appears when dimension goes to $+\infty$ is a semicircle distribution. However, some exoticism comes from the fact that this semicircle distribution is *non-centered*.

The model is simple to describe: start from Wishart $n \times n$ random matrices, which is the most natural model of random positive matrices. Assume that their dimension is a square ($n = d^2$). These matrices can be considered as block-matrices, with d^2 blocks, each block being a $d \times d$ matrix. Now our model is obtained by applying the transposition operation inside each block. A equivalent formulation is to consider $d^2 \times d^2$ matrices as operators on the tensor product of two d-dimensional spaces, and to apply to them the *partial transposition* $\mathrm{Id} \otimes T$, where T is the usual transposition. For this model, we prove that the empirical eigenvalue distribution converges towards a non-centered semicircular distribution. We also show a weak form of convergence for the extreme eigenvalues (up to some multiplicative constant).

Since the transposition is not a completely positive map, there is no reason *a priori* for matrices from our model to be positive. However, we show that for some range of the parameters, partially transposed Wishart matrices are typically positive. A threshold occurs when the parameter from the Wishart distribution equals 4.

This research was supported by the *Agence Nationale de la Recherche* grant ANR-08-BLAN-0311-03 and by the *Institut Mittag-Leffler* in Stockholm. I also thank S. Szarek and I. Nechita for useful comments.

The partial transposition appears to play a central role in Quantum Information Theory. An important class of states is the family of states with a Positive Partial Transpose (PPT). Non-PPT states are necessarily entangled [16] and this is the simplest test to detect entanglement. Let us simply mention a related important open problem known as the distillability conjecture [9]: it asks whether, for a state ρ, non-PPT is equivalent to the existence of a protocol which, given many copies of ρ, distills them to obtain Bell singlets—the most useful form of entanglement. A positive answer to the distillability conjecture would give a physical meaning to partial transposition.

The model of Wishart random matrices has also a physical interpretation in terms of open systems: assume the subsystem $\mathbf{C}^d \otimes \mathbf{C}^d$ is coupled with some environment $\mathbf{C}^p$. If the overall system is in a random pure state, the state on $\mathbf{C}^d \otimes \mathbf{C}^d$ obtained by partial tracing over $\mathbf{C}^p$ is distributed as a (normalized) Wishart matrix. Our results can be translated in this language. In particular, a random induced state is typically non-PPT when $p/d^2 < 4$ and is typically PPT when $p/d^2 > C$ for some constant C. The convergence of the smallest eigenvalue (observed numerically) would imply that $C = 4$.

Organization. The paper is organized as follows: Sections 2–6 are written in the language of random matrices and contain the proof of our theorems. Section 2 introduces background and states theorem 1 (convergence towards the non-centered semicircle distribution) and theorem 2 (an approximate version of the convergence for the extreme eigenvalues). Section 3 reminds the reader about non-crossing partitions and the combinatorics behind the moments method for Wishart matrices, on which we rely. Sections 4 and 5 contains the proof of the main lemmas needed in the proof of theorem 1. Section 6 contains the proof of theorem 2. Section 7 connects to Quantum Information Theory. Section 8 contains some general remarks.

2. Background and statement of the main theorem

2.1. **Conventions.** By the letter C we denote an absolute constant, which value may change from occurrence to occurrence. The integer part of a real number x is denoted by $\lfloor x \rfloor$.

2.2. **Semicircular and Marčenko–Pastur distributions.** Let $m \in \mathbf{R}$ and $\sigma > 0$. The *semicircular distribution with mean m and variance σ^2* is the probability distribution $\mu_{SC(m,\sigma^2)}$ with support $[m - 2\sigma, m + 2\sigma]$ and density

$$\frac{d\mu_{SC(m,\sigma^2)}}{dx} = \frac{1}{2\pi\sigma}\sqrt{4\sigma^2 - (x-m)^2}.$$

Lemma 2.1. *The moments of the semicircular distributions are given by the following formula* ($k \in \mathbf{N}$)

$$\int x^k d\mu_{SC(m,\sigma^2)}(x) = \sum_{i=0}^{\lfloor k/2 \rfloor} \frac{1}{i+1}\binom{k}{2i}\binom{2i}{i}\sigma^{2i} m^{k-2i}. \tag{1}$$

Proof. It is well-known ([1], page 7) that if X is a random variable with $SC(0,1)$ distribution, the moments of X are related to the Catalan numbers

$$\mathbf{E}\, X^{2k} = \frac{1}{k+1}\binom{2k}{k}, \qquad \mathbf{E}\, X^{2k+1} = 0.$$

Now, the random variable $Y = m + \sigma X$ has a $SC(m, \sigma^2)$ distribution, and therefore

$$\mathbf{E}\, Y^k = \sum_{j=0}^{k} \binom{k}{j} \mathbf{E}(\sigma X)^j m^{k-j}.$$

The terms with odd j vanish because the random variable X is symmetric. Setting $j = 2i$ gives the result. □

We now introduce the Marčenko–Pastur distributions. First, for $0 < \alpha \leqslant 1$, let f_α be the probability density defined on $[b_-, b_+]$ (where $b_\pm = (1 \pm \sqrt{\alpha})^2$) by

$$f_\alpha(x) = \frac{\sqrt{(x - b_-)(b_+ - x)}}{2\pi x \alpha}.$$

The *Marčenko–Pastur distribution with parameter* α is the following probability distribution $\mu_{MP(\alpha)}$.

- If $\alpha \geqslant 1$, then $\mu_{MP(\alpha)}$ is the probability distribution with density $f_{1/\alpha}$.
- If $0 < \alpha \leqslant 1$, then $d\mu_{MP(\alpha)}(x) = (1 - \alpha)\delta_0 + \alpha df_\alpha(x)$, where δ_0 denotes a Dirac mass at 0.

In particular, note the following fact: if X has a semicircle $SC(0, 1)$ distribution, then X^2 has a Marčenko–Pastur $MP(1)$ distribution.

2.3. **Asymptotic spectrum of Wishart matrices: Marčenko–Pastur distribution.** Define a (n, p)-*Wishart matrix* as a random $n \times n$ matrix W obtained by setting $W = \frac{1}{p} GG^\dagger$, where G is a $n \times p$ matrix with independent (real or complex[1]) $N(0, 1)$ entries. The real case and complex case are completely similar. In the sequel we focus on the complex case, which is relevant to Quantum Information Theory.

Let A be a $n \times n$ Hermitian matrix, and denote $\lambda_1, \ldots, \lambda_n$ the eigenvalues of A. The *empirical eigenvalue distribution* of A, denoted μ_A, is the probability measure on Borel subsets of $\mathbf{R}$ defined as

$$\mu_A = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i}.$$

In other words, $\mu_A(B)$ is the proportion of eigenvalues that belong to the Borel set B. For large sizes, the empirical eigenvalue distribution of a Wishart matrix approaches a *Marčenko–Pastur distribution.*

Theorem (Marčenko–Pastur, [12])**.** *Fix $\alpha > 0$, and for every n, let W_n be $(n, \lfloor \alpha n \rfloor)$-Wishart matrix. Then the sequence of probability distributions (μ_{W_n}) converges $(n \to \infty)$ towards the Marčenko-Pastur distribution $MP(\alpha)$.*

2.4. **Partial transposition.** We now assume that $n = d^2$. One can think of any $n \times n$ matrix A as a block matrix, consisting of $d \times d$ blocks, each block being a $d \times d$ matrix. The entries of the matrix are then conveniently described using 4 indices ranging from 1 to d

$$A = (A_{i,j}^{k,l})_{i,j,k,l}.$$

Here i denotes the block row index, j the block column index, k the row index inside the block (i, j) and l the column index inside the block (i, j). We can then apply to each block of A the transposition operation. The resulting matrix is denoted A^Γ and called the *partial transposition*[2] of A. Using indices, we may write

$$(A^\Gamma)_{i,j}^{k,l} = A_{i,j}^{l,k}. \tag{2}$$

[1] A complex-valued random variable ξ has a complex $N(0, 1)$ distribution if its real and imaginary parts are independent random variables with real $N(0, \frac{1}{2})$ distribution. In particular, $\mathbf{E}\,|\xi|^2 = 1$.

[2] An explanation for the notation is that Γ is "half" of the letter T which denotes the usual transposition.

Such a block matrix A can be naturally seen as an operator on $\mathbf{C}^d \otimes \mathbf{C}^d$. Indeed, a natural basis in this space is the double-indexed family $(e_i \otimes e_k)_{1 \leqslant i,k \leqslant d}$, where (e_i) is the canonical basis of $\mathbf{C}^d$. The action of A on this basis is described as

$$A(e_i \otimes e_k) = \sum_{j,l=1}^{d} A_{i,j}^{k,l} e_j \otimes e_l.$$

We may identify canonically $\mathcal{M}(\mathbf{C}^d \otimes \mathbf{C}^d)$ with $\mathcal{M}(\mathbf{C}^d) \otimes \mathcal{M}(\mathbf{C}^d)$. Via this identification, the matrix A^Γ coincides with $(\mathrm{Id} \otimes T)(A)$, where $T : \mathcal{M}(\mathbf{C}^d) \to \mathcal{M}(\mathbf{C}^d)$ is the usual transposition map. The map T is the simplest example of a map which is positive but not completely positive: $A \geqslant 0$ does not imply $A^\Gamma \geqslant 0$.

2.5. **Asymptotic spectrum of partially transposed Wishart matrices: non-centered semicircular distribution.** Motivated by Quantum Information Theory, we investigate the following question: what does the spectrum of A^Γ look like ? As we will see, the partial transposition dramatically changes the spectrum: the empirical eigenvalue distribution of A^Γ is no longer close to a Marčenko–Pastur distribution, but to a shifted semicircular distribution ! This is our main theorem.

Theorem 1. *Fix $\alpha > 0$. For each positive integer d, let W_d be a $(d^2, \lfloor \alpha d^2 \rfloor)$-Wishart matrix, and let $Y_d = W_d^\Gamma$ be the partial transposition of W_d. Then the (random) sequence of probability distributions $(\mu_{Y_d})_{d \in \mathbf{N}}$ converges weakly, in probability, towards the semicircular distribution $\mu_{SC(1,1/\alpha)}$.*

What we mean by "weakly, in probability" is the following: for any bounded continuous function $f : \mathbf{R} \to \mathbf{R}$ and any $\varepsilon > 0$,

$$\lim_{d \to \infty} \mathbf{P}\left(\left| \int f d\mu_{Y_d} - \int f d\mu_{SC(1,1/\alpha)} \right| > \varepsilon \right) = 0.$$

Note that the trace and the Hilbert–Schmidt norm are obviously invariant under partial transpose. The distributions $MP(\alpha)$ and $SC(1, 1/\alpha)$ (corresponding to eigenvalue distribution before and after applying partial transpose) indeed share the same first and second moments.

The support of the limiting spectral distribution $SC(1, 1/\alpha)$ is the interval $[1 - \frac{2}{\sqrt{\alpha}}, 1 + \frac{2}{\sqrt{\alpha}}]$. If we denote by $\lambda_{\min}(A)$ (resp. $\lambda_{\max}(A)$) the smallest (resp. largest) eigenvalue of a matrix A, theorem 1 implies that for every $\varepsilon > 0$,

$$\lim_{d \to \infty} \mathbf{P}\left(\lambda_{\min}(Y_d) \geqslant 1 - \frac{2}{\sqrt{\alpha}} + \varepsilon \right) = \lim_{d \to \infty} \mathbf{P}\left(\lambda_{\max}(Y_d) \leqslant 1 + \frac{2}{\sqrt{\alpha}} - \varepsilon \right) = 0.$$

As usual in Random Matrix Theory, a natural—and harder—question is the convergence of extreme eigenvalues. Numerical evidence support the following conjecture:

Conjecture. *With the notation from theorem 1, is is true that for any $\varepsilon > 0$,*

$$\lim_{d \to \infty} \mathbf{P}\left(\lambda_{\min}(Y_d) \leqslant 1 - \frac{2}{\sqrt{\alpha}} - \varepsilon \right) = \lim_{d \to \infty} \mathbf{P}\left(\lambda_{\max}(Y_d) \geqslant 1 + \frac{2}{\sqrt{\alpha}} + \varepsilon \right) = 0 \quad ?$$

This would exhibit a threshold for $\alpha = 4$: for $\alpha > 4$, the conjecture would imply that the matrix Y_d is typically positive, while theorem 1 implies that Y_d is typically non-positive for $\alpha < 4$. The property of "having a positive partial transpose" is relevant to Quantum Information Theory (see section 7).

Although we do not prove the conjecture here (it is likely that a more sophisticated version of the moments method can be used to this end), the following theorem is a version "up to multiplicative constant" of the conjecture.

Theorem 2. *There exists a constant C such that the following holds. Fix $\alpha \geqslant 1$, and each positive integer d, let W_d be a $(d^2, \lfloor \alpha d^2 \rfloor)$-Wishart matrix, and $Y_d = W_d^\Gamma$ be the partial transposition of W_d. Then*

$$\lim_{d\to\infty} \mathbf{P}\left(\lambda_{\min}(Y_d) \leqslant 1 - \frac{C}{\sqrt{\alpha}}\right) = \lim_{d\to\infty} \mathbf{P}\left(\lambda_{\max}(Y_d) \geqslant 1 + \frac{C}{\sqrt{\alpha}}\right) = 0.$$

2.6. **Proof of theorem 1: the moments method.** We will prove theorem 1 using the standard *method of moments* going back to Wigner (see [1] for a modern book exposition). For each d, let W_d be a (d^2, p)-Wishart matrix, with $p = \lfloor \alpha d^2 \rfloor$. By definition, $W_d = \frac{1}{p} G_d G_d^\dagger$ where G_d is a $d^2 \times p$ matrix with independent $N(0,1)$ entries. Finally, $Y_d = W_d^\Gamma$ denotes the partial transpose of W_d. The theorem will follow from a couple of lemmas.

Lemma 2.2. *For every $k \in \mathbf{N}$,*

$$\lim_{d\to\infty} \mathbf{E}\, \frac{1}{d^2} \operatorname{Tr} Y_d^k = m_k,$$

where m_k denotes the kth moment of a $SC(1, 1/\alpha)$ random variable, given by formula (1).

Lemma 2.3. *For every $k \in \mathbf{N}$,*

$$\lim_{d\to\infty} \mathbf{Var}\, \frac{1}{d^2} \operatorname{Tr} Y_d^k = 0.$$

We postpone the proof of lemmas 2.2 and 2.3 to sections 4 and 5. It is a standard procedure to show how both lemmas imply theorem 1. We do not reproduce the argument and refer to [1] (section 2.1.2) since the presentation from there (in the setting of Wigner matrices) can be "transposed" verbatim.

3. Non-crossing partitions and combinatorics of Wishart matrices

3.1. **Non-crossing partitions.** Let S be a finite set with a total order $<$. Usually, S equals $[k]$ (the set $\{1, \ldots, k\}$) for some positive integer k. If $i, j \in S$ with $i \neq j$, the *interval* (i, j) is defined as

$$(i,j) = \begin{cases} \{l \in S \text{ s.t } i < l < j\} & \text{if } i < j, \\ \{l \in S \text{ s.t } i < l \text{ or } l < j\} & \text{if } j < i. \end{cases}$$

It is useful to represent elements of S as points on a circle. We introduce the concept of non-crossing partitions and refer to [14] for more information and pictures.

- A *partition* π of S is a family $\{V_1, \ldots, V_p\}$ of disjoint nonempty subsets of S, whose union is S. The sets V_i are called the *blocks* of π. The number of blocks of π is denoted $|\pi|$.
- A partition π of S is said to be *non-crossing* if there does not exist elements $i < j < k < l$ in S such that a block of π contains i and k while another block of π contains j and l. We denote by $NC(S)$ the set of non-crossing partitions of S, and $NC(k) = NC([k])$.
- A *chording* (or a *non-crossing pair partition*) of S is a non-crossing partition of S in which each block contains exactly two elements. Chordings exist only when the cardinal of S is even. We denote by $NC_2(S)$ the set of chordings of S, and $NC_2(k) = NC_2([k])$. Given a chording π, the *partner* of an element $i \in S$ is the unique $j \in S$ such that $\{i, j\}$ is a block of π.
- A *partial chording* of S is a non-crossing partition of S in which each block contains at most two elements.

Counting non-crossing partitions is a well-known combinatorial problem involving Catalan numbers.

Proposition. *([14], Lemma 8.9 and Proposition 9.4) Let $k \in \mathbf{N}^*$. The number of elements in $NC(k)$ and the number of elements in $NC_2(2k)$ are both equal to the kth Catalan number $C_k = \frac{1}{k+1}\binom{2k}{k}$.*

A natural bijection $\Phi : NC_2(2k) \to NC(k)$ can be constructed as follows (see also [14], exercise 9.42). Consider the set $[k]$, with $2k$ additional numbers $1^-, \dots, k^-, 1^+, \dots, k^+$ in the following order:

$$1^- < 1 < 1^+ < 2^- < 2 < 2^+ < \cdots < k^- < k < k^+.$$

Let $\pi \in NC_2(2k) \simeq NC_2(\{1^-, 1^+, \dots, k^-, k^+\})$. It is easily checked (e.g. by induction on k) that every block in π has the form $\{i^-, j^+\}$. Define $\Phi(\pi)$ to be the coarsest partition $\sigma \in NC(k)$ such that $\pi \cup \sigma$ is non-crossing. Equivalently, let $\sim$ be the relation on $[k]$ given by $i \sim j$ when either $\{i^-, j^+\}$ or $\{i^+, j^-\}$ is a block in π. The blocks of $\Phi(\pi)$ are given by the transitive closure of the relation $\sim$.

If $\sigma \in NC(k)$, one can recover $\Phi^{-1}(\sigma)$ as follows: whenever $\{i_1, \dots, i_p\}$ is a block of σ with $i_1 < \cdots < i_p$, then $\{i_1^+, i_2^-\}, \dots, \{i_{p-1}^+, i_p^-\}$ and $\{i_p^+, i_1^-\}$ must be blocks of $\Phi^{-1}(\sigma)$.

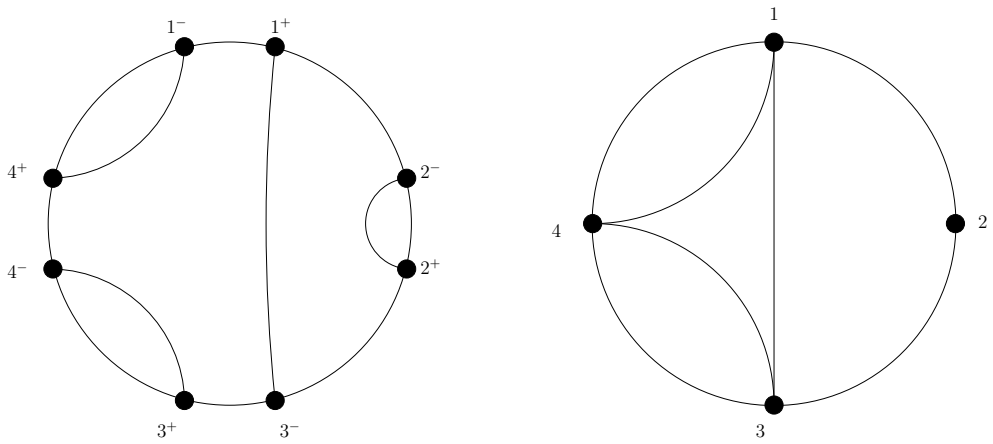

FIGURE 1. An example of a chording $\pi \in NC_2(8)$ on the left, together with the corresponding non-crossing partition $\Phi(\pi) \in NC(4)$ on the right. Visually the action of Φ collapses the points i^- and i^+ to a single point i.

The next simple lemma will play a central role in our proof: it characterizes the non-crossing partitions which are invariant under the map swapping $+$ and $-$.

Lemma 3.1. *Let $S = \{1^-, 1^+, \dots, k^-, k^+\}$, equipped with the order $<$ defined by $1^- < 1^+ < \cdots < k^- < k^+$. Let $\pi \in NC_2(2k) \simeq NC_2(S)$. The following are equivalent:*

(i) *The partition π is also non-crossing when S is equipped with the order $\lessdot$ defined as*

$$1^+ \lessdot 1^- \lessdot 2^+ \lessdot 2^- \cdots \lessdot k^+ \lessdot k^-.$$

(ii) *For every block $\{i^-, j^+\}$ of π, the set $\{i^+, j^-\}$ is also a block of π.*

(iii) *Every block in $\Phi(\pi)$ contains at most two elements.*

Proof. We show that (i) and (ii) are equivalent. Assume (i), and let $\{i^-, j^+\}$ be a block of π. If $i = j$, there is nothing to prove. Otherwise, the partner of i^+ must belong to

- the interval (i^+, j^+) for the order $<$ (since π is non-crossing for $<$),
- the interval (j^+, i^+) for the order $\lessdot$ (since π is non-crossing for $\lessdot$).

The only common point of these intervals is j^-. This shows (ii). The converse implication (ii) $\Rightarrow$ (i) is obvious.

Let us show that (ii) implies (iii). Let $i \in [k]$. If the partner of i^- is i^+, it follows easily from the definition of $\Phi(\pi)$ that the singleton $\{i\}$ is a block of $\Phi(\pi)$. Otherwise, both $\{i^-, j^+\}$ and $\{i^+, j^-\}$ are blocks of π for some $j \neq i$. Similarly, it follows from the definition of $\Phi(\pi)$ that the pair $\{i, j\}$ is a block of $\Phi(\pi)$.

Finally, we check that (iii) implies (ii). Let $\{i^-, j^+\}$ be a block of π. If $j = i$ there is nothing to prove. Otherwise, let l^- be the partner of i^+. Since i and j (resp. i and l) are in the same block of $\Phi(\pi)$, it follows that $l = j$. □

It follows that the number of pair-partitions on $\{1^-, 1^+, \dots, k^-, k^+\}$ which are non-crossing for both orders $<$ and $\lessdot$ is equal to the number of partial chordings of $[k]$, also known as the kth Motzkin number (see [6]). The Motzkin numbers are the moments of the (non-centered) semicircle distribution $SC(1, 1)$. This is already a great step towards theorem 1.

Let us also introduce the *Kreweras complementation* as the map $K : NC(k) \mapsto NC(k)$ defined as follows. For $\pi \in NC(k) \simeq NC(\{1^-, \dots, k^-\})$, $K(\pi)$ is defined as the coarsest partition $\sigma \in NC(\{1^+, \dots, k^+\}) \simeq NC(k)$ such that $\pi \cup \sigma$ is a non-crossing partition of $\{1^-, 1^+, \dots, k^-, k^+\}$. The map K is a bijection on $NC(k)$ (see e.g. [14]).

3.2. **Combinatorics related to moments of Wishart matrices.** In this section we remind the reader about the (standard) proof of the Marčenko–Pastur theorem via the moments method. This proof can be found for example in [11, 15] or the book [4]. Not only our proof will mimic this one, but we will actually strongly recycle most of the combinatorial lemmas.

Let $W_n = (W_{ij})$ be a (n, p)-Wishart matrix, and $k \in \mathbf{N}$. The expansion of $\mathbf{E}\,\frac{1}{n}\operatorname{Tr} W_n^k$ reads

$$\begin{aligned} \mathbf{E}\,\frac{1}{n}\operatorname{Tr} W_n^k &= \frac{1}{n}\sum_{\vec{a}\in[n]^k} \mathbf{E}\, W_{a_1,a_2}W_{a_2,a_3}\cdots W_{a_k,a_1} \\ &= \frac{1}{np^k}\sum_{\vec{a}\in[n]^k,\vec{c}\in[p]^k} \mathbf{E}\, G_{a_1,c_1}\overline{G_{a_2,c_1}}G_{a_2,c_2}\overline{G_{a_3,c_2}}\cdots G_{a_k,c_k}\overline{G_{a_1,c_k}}. \end{aligned} \tag{3}$$

The next task is to analyze which couples $(\vec{a}, \vec{c})$ give dominant contributions to the sum (3) when $n \to \infty$ and $p = \lfloor \alpha n \rfloor$. One argues as follows. First, if one pair (a_i, c_i) or (a_{i+1}, c_i) appears only once in the product, then the contribution is exactly zero (because entries of G are independent and mean zero). This motivates the following definition:

Definition. A couple $(\vec{a}, \vec{c})$ of length k multi-indices satisfies the *Wishart matching condition* if every couple in the following list of $2k$ elements appears at least twice:

$$(a_1, c_1), (a_2, c_1), (a_2, c_2), (a_3, c_2), \dots, (a_k, c_k), (a_1, c_k). \tag{4}$$

Let $\vec{a}$ and $\vec{c}$ be length k multi-indices. We define $d_W(\vec{a}, \vec{c})$ as the number of distinct couples appearing in the list (4), and set $\ell_W(\vec{a}, \vec{c}) = \#\vec{a} + \#\vec{c}$.

Lemma 3.2. *Let $(\vec{a}, \vec{c})$ be a couple of length k multi-indices. Then $\ell_W(\vec{a}, \vec{c}) \leqslant d_W(\vec{a}, \vec{c}) + 1$. Therefore, if $(\vec{a}, \vec{c})$ satisfies the Wishart matching condition, then $\ell_W(\vec{a}, \vec{c}) \leqslant k + 1$.*

Proof. Read the list (4) from left to right, and count how many new indices you read. The first couple (a_1, c_1) brings two new indices, and each subsequent couple that did not appeared previously in the list

(there are $d-1$ such couples) may bring at most one new index (since it shares a common index with the couple just before).

The second part follows easily: if every couple in the list (4) appears at least twice, then this list contains at most k different couples. □

Now, the couples $(\vec{a},\vec{c})$ satisfying $\ell_W(\vec{a},\vec{c}) < k+1$ are easily shown to have a contribution which is asymptotically zero. Let us say that $(\vec{a},\vec{c})$ is *Wishart-admissible* if it satisfies the matching condition, together with the equality $\ell_W(\vec{a},\vec{c}) = k+1$.

Let $\vec{a} = (a_1,\dots,a_k)$ be a multi-index of length k. The *partition induced by* $\vec{a}$ is the partition of $[k]$ defined as follows: i and j belong to the same block if and only if $a_i = a_j$. Two length k multi-indices $\vec{a},\vec{b}$ are *equivalent* ($\vec{a}\sim\vec{b}$) if they induce the same partition of $[k]$. Similarly, a couple $(\vec{a},\vec{c})$ is equivalent to a couple $(\vec{a}\,',\vec{c}\,')$ if $\vec{a}\sim\vec{a}\,'$ and $\vec{c}\sim\vec{c}\,'$. The next proposition (see [11] or [15] for details) characterizes the combinatorial structure of (equivalence classes of) Wishart-admissible couples.

Proposition 3.3. *Let $(\vec{a},\vec{c})$ be a Wishart-admissible couple of length k multi-indices. Then*

(i) *Each couple in the list (4) appears exactly twice. One occurrence is of the form (a_i,c_i) while the other occurrence is of the form (a_{i+1},c_i).*

(ii) *Let π be the pair-partition of $[2k]$ induced by the list (4). Then π is non-crossing.*

(iii) *Let σ (resp. σ') be the partition of $[k]$ induced by $\vec{a}$ (resp. by $\vec{c}$). The partitions σ and σ' are non-crossing, and $\sigma' = K(\sigma)$. In particular, given $\vec{a}$, the multi-index $\vec{c}$ is uniquely determined up to equivalence (and vice-versa). Moreover, $\sigma' = \Phi(\pi)$, where $\Phi : NC_2(2k) \to NC(k)$ is the bijection described in the previous section.*

(iv) *The mapping $(\vec{a},\vec{c}) \mapsto \sigma'$ induces a bijection between the set of equivalence classes of Wishart-admissible couples of length k multi-indices and the set $NC(k)$.*

Example. Let us give an example of a Wishart-admissible couple for $k=4$. Let $\vec{a} = (1,2,2,3)$ and $\vec{c} = (7,3,7,7)$. Then $\ell_W(\vec{a},\vec{c}) = 5$. The list (4) reads as

$$(1,7);(2,7);(2,3);(2,3);(2,7);(3,7);(3,7);(1,7)$$

and each couple appears exactly twice, so that the couple $(\vec{a},\vec{c})$ is indeed Wishart-admissible. The partitions π,σ' and σ' appearing in proposition 3.3 are

$$\pi = \{\{1,8\},\{2,5\},\{3,4\},\{6,7\}\},$$
$$\sigma = \{\{1\},\{2,3\},\{4\}\},$$
$$\sigma' = K(\sigma) = \Phi(\pi) = \{\{1,3,4\},\{2\}\}.$$

The partitions π and σ' were depicted in figure 1.

From proposition 3.3, it is easy to check (if $p \simeq \alpha n$) that $\lim_{n\to\infty} \mathbf{E}\,\frac{1}{n}\operatorname{Tr} W_n^k$ coincides with the kth moment of the Marčenko–Pastur distribution with parameter α. To obtain more information than convergence in expectation, one usually needs also a control of the variance of $\frac{1}{n}\operatorname{Tr} W_n^k$. The next lemma is then relevant. Actually, the stronger conclusion $\ell_W(\vec{a},\vec{c}) + \ell_W(\vec{a}\,',\vec{c}\,') \leqslant 2k$ holds, but we do not need this sophistication here.

Lemma 3.4. *Let $\vec{a}$, $\vec{c}$, $\vec{a}\,'$ and $\vec{c}\,'$ be multi-indices of length k satisfying the following conditions*

(i) *Each couple in the following list of $4k$ elements appears at least twice:*

$$(a_1,c_1),(a_2,c_1),\dots,(a_k,c_k),(a_1,c_k)\,;\,(a'_1,c'_1),(a'_2,c'_1),\dots,(a'_k,c'_k),(a'_1,c'_k). \tag{5}$$

(ii) *At least some couple appears both in the left half and in the right half of the list* (5).

Then $\ell_W(\vec{a},\vec{c}) + \ell_W(\vec{a}\,',\vec{c}\,') \leqslant 2k+1$.

Proof. As before, we read the list (5) and keep track of the number of indices. We first read the left half of the list in its natural order. We then read the right half of the list, starting by a element which already appeared in the left half and reading from left to right—with the convention that (a'_1, c'_1) stands at the right of (a'_1, c_k).

The first element (a_1, c_1) brings two new indices, and each subsequent new couple (there are at most $2k-1$ many, since each couple in the list appears at least twice) brings at most one new index. □

The lower-order contributions to the sum (3) can also be estimated. This is useful to prove the convergence of the extreme eigenvalues. Here is a lemma that we will use.

Lemma 3.5. *The number of equivalence classes of couples* $(\vec{a},\vec{c})$ *which satisfy the Wishart matching condition and such that* $\ell_W(\vec{a},\vec{c}) = \ell$*, is bounded by* $C^k k^{k-\ell}$*, where* C *is an absolute constant.*

Proof. It follows from the combinatorial analysis in [7] (see also the book [4] and references therein) that the number under investigation is bounded by $C^k k^{6(k+1-\ell)}$ (which is not exactly of the form claimed by the lemma, but would anyway be suitable for our purposes). An alternative non-combinatorial argument based on concentration of measure (Lemma 4 from [2]) shows that the number of couples $(\vec{a},\vec{c}) \in [k]^k \times [k]^k$ satisfying the Wishart-matching condition is bounded by $(Ck)^k$ for some constant C. The cardinality of the equivalence class of $(\vec{a},\vec{c})$ in $[k]^k \times [k]^k$ is

$$k(k-1)\cdots(k+1-\#\vec{a}) \cdot k(k-1)\cdots(k+1-\#\vec{c}).$$

For $p \in [k]$, we may use the (crude) lower bound $k(k-1)\cdots(k+1-p) \geqslant k^p e^{-k}$. Therefore, the cardinality of the equivalence class of $(\vec{a},\vec{c})$ is at least $k^{\#\vec{a}+\#\vec{c}} e^{-2k}$. On the other hand, by lemma 3.2, the Wishart matching condition implies that $\#\vec{a} \leqslant k$ and $\#\vec{c} \leqslant k$, so that any couple satisfying the Wishart matching condition is equivalent to a couple in $[k]^k \times [k]^k$. The conclusion is now immediate. □

4. Proof of lemma 2.2: expectation of $\operatorname{Tr} Y_d^k$

Let G_d be a $d^2 \times p$ matrix with independent $N(0,1)$ entries. We denote the entries of G_d as $(G^c_{a,b})$, where $(a,b) \in [d]\times[d]$ denote the row indices and $c \in [p]$ denotes the column index. We label the entries of W_d and Y_d as $(W^{b,b'}_{a,a'})$ and $(Y^{b,b'}_{a,a'})$, where $(a,a',b,b') \in [d]^4$ according to the convention described in section 2.4. Let us consider the expansion of $\mathbf{E}\,\frac{1}{d^2}\operatorname{Tr} Y_d^k$.

$$\begin{aligned}
\mathbf{E}\,\frac{1}{d^2}\operatorname{Tr} Y_d^k &= \frac{1}{d^2} \sum_{\vec{a}\in[d]^k, \vec{b}\in[d]^k} \mathbf{E}\, Y^{b_1,b_2}_{a_1,a_2} \cdot Y^{b_2,b_3}_{a_2,a_3} \cdots Y^{b_{k-1},b_k}_{a_{k-1},a_k} \cdot Y^{b_k,b_1}_{a_k,a_1} \\
&= \frac{1}{d^2} \sum_{\vec{a}\in[d]^k, \vec{b}\in[d]^k} \mathbf{E}\, W^{b_2,b_1}_{a_1,a_2} \cdot W^{b_3,b_2}_{a_2,a_3} \cdots W^{b_k,b_{k-1}}_{a_{k-1},a_k} \cdot W^{b_1,b_k}_{a_k,a_1} \\
&= \frac{1}{d^2 p^k} \sum_{\vec{a}\in[d]^k, \vec{b}\in[d]^k, \vec{c}\in[p]^k} \mathbf{E}\, G^{c_1}_{a_1,b_2} \overline{G^{c_1}_{a_2,b_1}} \cdot G^{c_2}_{a_2,b_3} \overline{G^{c_2}_{a_3,b_2}} \cdots G^{c_{k-1}}_{a_{k-1},b_k} \overline{G^{c_{k-1}}_{a_k,b_{k-1}}} \cdot G^{c_k}_{a_k,b_1} \overline{G^{c_k}_{a_1,b_k}}.
\end{aligned}$$

If $(\vec{a},\vec{b},\vec{c}) \in [d]^k \times [d]^k \times [p]^k$, we define

$$\ell(\vec{a},\vec{b},\vec{c}) := \ell_W(\vec{a},\vec{c}) + \ell_W(\vec{b},\vec{c}) = \#\vec{a} + \#\vec{b} + 2\#\vec{c}.$$

Define also

$$\Pi(\vec{a},\vec{b},\vec{c}) := G^{c_1}_{a_1,b_2}\overline{G^{c_1}_{a_2,b_1}} \cdot G^{c_2}_{a_2,b_3}\overline{G^{c_2}_{a_3,b_2}} \dots G^{c_{k-1}}_{a_{k-1},b_k}\overline{G^{c_{k-1}}_{a_k,b_{k-1}}} \cdot G^{c_k}_{a_k,b_1}\overline{G^{c_k}_{a_1,b_k}}.$$

The expression above takes the more compact form

$$\mathbf{E}\,\frac{1}{d^2}\operatorname{Tr} Y_d^k = \frac{1}{d^2 p^k} \sum_{\vec{a}\in[d]^k,\vec{b}\in[d]^k,\vec{c}\in[p]^k} \mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c}). \tag{6}$$

Let $\mathscr{C}_k$ be the (finite) family of all equivalence classes of triples $(\vec{a},\vec{b},\vec{c})$ of length k multi-indices. Since the quantities $\ell(\vec{a},\vec{b},\vec{c})$ and $\mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c})$ depend only on the equivalence class of the triple $(\vec{a},\vec{b},\vec{c})$, we may abusively write $\ell(C)$ and $\mathbf{E}\,\Pi(C)$, where $C \in \mathscr{C}_k$. A straightforward application of Hölder's inequality shows that $\mathbf{E}\,\Pi(C)$ is bounded by the $2k$th moment of a standard Gaussian variable—a constant depending only on k. We rearrange the sum according to equivalence classes of triples:

$$\mathbf{E}\,\frac{1}{d^2}\operatorname{Tr} Y_d^k = \frac{1}{d^2 p^k} \sum_{C\in\mathscr{C}_k} \#\left\{C \cap [d]^k \times [d]^k \times [p]^k\right\} \mathbf{E}\,\Pi(C).$$

Let $C \in \mathscr{C}_k$ be the equivalence class of a triple $(\vec{a},\vec{b},\vec{c})$. When $d \to \infty$

$$\#\{C \cap ([d]^k \times [d]^k \times [p]^k)\} \sim d^{\#\vec{a}} d^{\#\vec{b}} p^{\#\vec{c}} \sim \alpha^{\#\vec{c}} d^{\ell(\vec{a},\vec{b},\vec{c})}.$$

Consequently, if $\ell(C) < 2k+2$, the contribution of the class C is asymptotically zero. On the other hand, if some factor in the product $\Pi(\vec{a},\vec{b},\vec{c})$ appears only once, then, because of independence of entries of the matrix G, $\mathbf{E}\,\Pi(C) = 0$. This motivates the following definition.

Definition. A triple $(\vec{a},\vec{b},\vec{c})$ of length k multi-indices satisfies the *matching condition* if no factor in $\Pi(\vec{a},\vec{b},\vec{c})$ appears only once. In other words, $(\vec{a},\vec{b},\vec{c})$ satisfies the matching condition if, in the following list of $2k$ triples, each triple appears at least twice:

$$(a_1,b_2,c_1),(a_2,b_1,c_1)\,;\,(a_2,b_3,c_2),(a_3,b_2,c_2)\,;\,\dots\,;\,(a_k,b_1,c_k),(a_1,b_k,c_k). \tag{7}$$

The following easy observation will be used repeatedly

Fact 4.1. *Assume that $(\vec{a},\vec{b},\vec{c})$ satisfies the matching condition. Then both $(\vec{a},\vec{c})$ and $(\vec{b},\vec{c})$ satisfy the Wishart matching condition.*

Together with lemma 3.2, this fact implies that a triple $(\vec{a},\vec{b},\vec{c})$ which satisfies the matching condition must satisfy $\ell(\vec{a},\vec{b},\vec{c}) \leqslant 2k+2$. Let us say that $(\vec{a},\vec{b},\vec{c})$ is *admissible* if it satisfies the matching condition and $\ell(\vec{a},\vec{b},\vec{c}) = 2k+2$. We already noticed that non-admissible triples have an asymptotically zero contribution.

We now simply have to count the number of (equivalence classes of) admissible triples of length k multi-indices. Recall the terminology introduced just before proposition 3.3: $\vec{a} \sim \vec{a}\,'$ means that $\vec{a}$ and $\vec{a}\,'$ induce the same partition, and $(\vec{a},\vec{b},\vec{c}) \sim (\vec{a}\,',\vec{b}\,',\vec{c}\,')$ means $\vec{a} \sim \vec{a}\,'$, $\vec{b} \sim \vec{b}\,'$ and $\vec{c} \sim \vec{c}\,'$. The following proposition makes explicit a bijection with the set of partial chordings of $[k]$.

Proposition 4.2. *Let $(\vec{a},\vec{b},\vec{c})$ be an admissible triple of length k multi-indices. Then*

(i) *Each triple in the list* (7) *appears exactly twice. One occurrence is of the form* (a_i, b_{i+1}, c_i) *while the other occurrence is of the form* (a_{i+1}, b_i, c_i)*. Therefore* $\mathbf{E}\,\Pi(\vec{a}, \vec{b}, \vec{c}) = 1$*, both in the real and the complex cases.*

(ii) *The couples* $(\vec{a}, \vec{c})$ *and* $(\vec{b}, \vec{c})$ *are Wishart-admissible. Therefore* $\vec{a}$ *and* $\vec{b}$ *are uniquely determined (up to equivalence) by* $\vec{c}$ *; in particular* $\vec{a} \sim \vec{b}$*.*

(iii) *The partition induced by* $\vec{c}$ *is a partial chording of* $[k]$*.*

Conversely, if $(\vec{a}, \vec{c})$ *is a Wishart-admissible couple of length k multi-indices, such that the partition induced by* $\vec{c}$ *is a partial chording of* $[k]$*, then* $(\vec{a}, \vec{a}, \vec{c})$ *is admissible.*

Let us postpone the proof of proposition 4.2, and show first how it implies the theorem. For $C \in \mathscr{C}_k$, denote $\gamma(C) = \#\vec{c}$, where $(\vec{a}, \vec{b}, \vec{c})$ is any member of C. We have

$$\begin{aligned}\lim_{d\to\infty} \frac{1}{d^2} \operatorname{Tr} Y^k &= \frac{1}{\alpha^k} \sum_{C \in \mathscr{C}_k, C \text{ admissible}} \alpha^{\gamma(C)} \\ &= \frac{1}{\alpha^k} \sum_{\pi \text{ partial chording of } [k]} \alpha^{|\pi|}.\end{aligned}$$

A partial chording of $[k]$ with i chords is uniquely determined by the data of the $2i$ endpoints of the chords, together with a chording of $[2i]$. Using proposition 3.1, the number of partial chordings of $[k]$ with i chords is therefore equal to

$$\binom{k}{2i} C_i = \frac{1}{i+1} \binom{k}{2i} \binom{2i}{i}.$$

Moreover, such a partial chording consists of $k - i$ blocks (i chords and $k - 2i$ isolated points). Therefore,

$$\lim_{d\to\infty} \frac{1}{d^2} \operatorname{Tr} Y^k = \frac{1}{\alpha^k} \sum_{i=0}^{\lfloor k/2 \rfloor} \frac{1}{i+1} \binom{k}{2i} \binom{2i}{i} \alpha^{k-i} = \sum_{i=0}^{\lfloor k/2 \rfloor} \frac{1}{i+1} \binom{k}{2i} \binom{2i}{i} \alpha^{-i}.$$

By lemma 2.1, the last quantity is exactly the kth moment of the semicircular distribution $SC(1, 1/\alpha)$.

Proof of proposition 4.2. Since the couples $(\vec{a}, \vec{c})$ and $(\vec{b}, \vec{c})$ satisfy the Wishart matching condition, we have $\ell_W(\vec{a}, \vec{c}) \leqslant k+1$ and $\ell_W(\vec{b}, \vec{c}) \leqslant k+1$. This is only possible if both quantities are equal to $k+1$, meaning that the couples $(\vec{a}, \vec{c})$ and $(\vec{b}, \vec{c})$ are Wishart-admissible. Assertion (i) then follows from the corresponding assertion in proposition 3.3. Again by proposition 3.3, one of the multi-indices from a Wishart-admissible couple determines the second multi-index (up to equivalence). This proves that $\vec{a}$ and $\vec{b}$ are equivalent.

Let $(\vec{a}, \vec{c})$ be a Wishart-admissible couple. It remains to show that $(\vec{a}, \vec{a}, \vec{c})$ is admissible if and only if the partition induced by $\vec{c}$ is a partial chording. Let π be the chording of $[2k] \simeq \{1^-, 1^+, \ldots, k^-, k^+\}$ induced by the list (4). A moment's thought reveals that $(\vec{a}, \vec{a}, \vec{c})$ is admissible if and only if

$$\{i^-, j^+\} \in \pi \iff \{i^+, j^-\} \in \pi.$$

By lemma 3.1, this is equivalent to say that $\Phi(\pi)$ is a partial chording. By proposition 3.3, $\Phi(\pi)$ is exactly the partition induced by $\vec{c}$.

□

5. Proof of lemma 2.3: variance of $\operatorname{Tr} Y_d^k$

We expand the trace, as in the proof of lemma 2.2

$$\mathbf{Var}\,\frac{1}{d^2}\operatorname{Tr} Y_d^k = \frac{1}{d^4p^{2k}} \sum_{\vec{a},\vec{b},\vec{c},\vec{a}\,',\vec{b}\,',\vec{c}\,'} \mathbf{E}\left[\Pi(\vec{a},\vec{b},\vec{c})\Pi(\vec{a}\,',\vec{b}\,',\vec{c}\,')\right] - \mathbf{E}\left[\Pi(\vec{a},\vec{b},\vec{c})\right]\mathbf{E}\left[\Pi(\vec{a}\,',\vec{b}\,',\vec{c}\,')\right], \tag{8}$$

where the summation is taken over multi-indices $\vec{a},\vec{b},\vec{a}\,',\vec{b}\,'$ in $[d]^k$, and $\vec{c},\vec{c}\,'$ in $[p]^k$. We first identify the vanishing contributions.

Lemma 5.1. *Let $(\vec{a},\vec{b},\vec{c})$ and $(\vec{a}\,',\vec{b}\,',\vec{c}\,')$ be two triples of length k multi-indices such that*

$$\mathbf{E}\left[\Pi(\vec{a},\vec{b},\vec{c})\Pi(\vec{a}\,',\vec{b}\,',\vec{c}\,')\right] \neq \mathbf{E}\left[\Pi(\vec{a},\vec{b},\vec{c})\right]\mathbf{E}\left[\Pi(\vec{a}\,',\vec{b}\,',\vec{c}\,')\right].$$

Then $\ell(\vec{a},\vec{b},\vec{c}) + \ell(\vec{a}\,',\vec{b}\,',\vec{c}\,') \leqslant 4k+2$.

Proof. The independence of entries of G shows that the following two conditions must hold:

- Each couple in the following list of $4k$ elements appears at least twice:

$$(a_1,b_2,c_1),(a_2,b_1,c_1),\dots,(a_k,b_1,c_k),(a_1,b_k,c_k)\,;\,(a'_1,b'_2,c'_1),(a'_2,b'_1,c'_1),\dots,(a'_k,b'_1,c'_k),(a'_1,b'_k,c'_k). \tag{9}$$

- At least some couple appears both in the left half and in the right half of the list (9). Otherwise, the random variables $\Pi(\vec{a},\vec{b},\vec{c})$ and $\Pi(\vec{a}\,',\vec{b}\,',\vec{c}\,')$ would be independent, and their covariance would be zero.

As is immediately checked, these conditions imply that $\vec{a},\vec{c},\vec{a}\,',\vec{c}\,'$ satisfy the hypotheses of lemma 3.4. Therefore,

$$\ell_W(\vec{a},\vec{c}) + \ell_W(\vec{a}\,',\vec{c}\,') \leqslant 2k+1.$$

Similarly, one may apply lemma 3.4 to $\vec{b},\vec{c},\vec{b}\,',\vec{c}\,'$ to obtain

$$\ell_W(\vec{b},\vec{c}) + \ell_W(\vec{b}\,',\vec{c}\,') \leqslant 2k+1.$$

It remains to add both inequalities. □

We now gather the non-zero terms appearing in the sum (8) according to the equivalence class of $(\vec{a},\vec{b},\vec{c},\vec{a}\,',\vec{b}\,',\vec{c}\,')$. The cardinality of the equivalence class of $(\vec{a},\vec{b},\vec{c},\vec{a}\,',\vec{b}\,',\vec{c}\,')$ is bounded by

$$d^{\#\vec{a}+\#\vec{b}+\#\vec{a}\,'+\#\vec{b}\,'}p^{\#\vec{c}+\#\vec{c}\,'} = O\left(d^{\ell(\vec{a},\vec{b},\vec{c})+\ell(\vec{a}\,',\vec{b}\,',\vec{c}\,')}\right) = O\left(d^{4k+2}\right).$$

The overall factor $1/d^4p^{2k} = O(1/d^{4k+4})$ in front of the sum (8) shows that each class has contribution asymptotically zero. Since the number of equivalence classes depends only on k, this proves the lemma.

6. Proof of theorem 2

Let $W = \frac{1}{p}GG^\dagger$ be a (d^2,p)-Wishart matrix, with $p = \lfloor \alpha d^2 \rfloor$, and $Y = W^\Gamma$. Denote by $\|\cdot\|$ the operator norm. The theorem will follow (using Markov's inequality) if we show that

$$\left(\mathbf{E}\,\|Y - \mathrm{Id}\|^{k_d}\right)^{1/k_d} \leqslant \frac{C}{\sqrt{\alpha}}$$

for some sequence (k_d) tending to $+\infty$. Let $(G_i)_{1\leqslant i\leqslant p}$ denote the columns of G, seen as vectors in $\mathbf{C}^d \otimes \mathbf{C}^d$. We use the Dirac notation $|x\rangle\langle x|$ to denote the rank 1 linear map $y \mapsto \langle x|y\rangle x$. We have

$W = \frac{1}{p}\sum_{i=1}^p |G_i\rangle\langle G_i|$ and $Y = \frac{1}{p}\sum_{i=1}^p |G_i\rangle\langle G_i|^\Gamma$. Let us start with a standard symmetrization argument due to Giné–Zinn.

Lemma 6.1. *For every integer $k \geqslant 1$,*

$$\mathbf{E}\,\|Y - \mathrm{Id}\|^k \leqslant 2^k\,\mathbf{E}\left\|\frac{1}{p}\sum_{i=1}^p \varepsilon_i |G_i\rangle\langle G_i|^\Gamma\right\|^k,$$

where (ε_i) is a sequence of independent ± 1-valued random variables, independent from G, so that $\mathbf{P}(\varepsilon_i = 1) = \mathbf{P}(\varepsilon_i = -1) = 1/2$.

Proof. Let $G' = (G'_i)$ be an independent copy of G. We have $\mathbf{E}\,\frac{1}{p}\sum |G'_i\rangle\langle G'_i|^\Gamma = \mathrm{Id}^\Gamma = \mathrm{Id}$. Applying Jensen's inequality to the expectation over G' and the convex function $x \mapsto \|Y - x\|^k$ yields

$$\mathbf{E}\,\|Y - \mathrm{Id}\|^k \leqslant \mathbf{E}\left\|\frac{1}{p}\sum_{i=1}^p |G_i\rangle\langle G_i|^\Gamma - \frac{1}{p}\sum_{i=1}^p |G'_i\rangle\langle G'_i|^\Gamma\right\|^k = \mathbf{E}\left\|\frac{1}{p}\sum_{i=1}^p |G_i\rangle\langle G_i|^\Gamma - |G'_i\rangle\langle G'_i|^\Gamma\right\|^k.$$

Now, the random variables $|G_i\rangle\langle G_i|^\Gamma - |G'_i\rangle\langle G_i|^\Gamma$ are symmetric and therefore have the same distribution as $\varepsilon_i(|G_i\rangle\langle G_i|^\Gamma - |G'_i\rangle\langle G'_i|^\Gamma)$. The triangle inequality gives

$$\mathbf{E}\,\|Y - \mathrm{Id}\|^k \leqslant \mathbf{E}\left(\left\|\frac{1}{p}\sum_{i=1}^p \varepsilon_i |G_i\rangle\langle G_i|^\Gamma\right\| + \left\|\frac{1}{p}\sum_{i=1}^p \varepsilon_i |G'_i\rangle\langle G'_i|^\Gamma\right\|\right)^k \leqslant 2^k\,\mathbf{E}\left\|\frac{1}{p}\sum_{i=1}^p \varepsilon_i |G_i\rangle\langle G_i|^\Gamma\right\|^k. \qquad \square$$

When k is an even integer, we may use the bound $\|A\|_{op}^k \leqslant \mathrm{Tr}\,A^k$, with $A = \sum \varepsilon_i |G_i\rangle\langle G_i|^\Gamma$. Expanding the trace leads to an expression similar to (6)

$$\mathbf{E}\,\|Y - \mathrm{Id}\|^k \leqslant \frac{2^k}{p^k} \sum_{\vec{a}\in[d]^k, \vec{b}\in[d]^k, \vec{c}\in[p]^k} \mathbf{E}\,\varepsilon_{c_1}\ldots\varepsilon_{c_k}\,\mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c}).$$

Let us say that $\vec{c}$ is *paired* if every element appearing in $\vec{c}$ appears an even number of times. The quantity $\mathbf{E}\,\varepsilon_{c_1}\ldots\varepsilon_{c_k}$ equals 1 if $\vec{c}$ is paired, and 0 otherwise. Therefore we have

$$\mathbf{E}\,\|Y - \mathrm{Id}\|^k \leqslant \frac{2^k}{p^k} \sum_{\vec{a}\in[d]^k, \vec{b}\in[d]^k, \vec{c}\in[p]^k, \vec{c}\text{ paired}} \mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c}). \tag{10}$$

As we already saw, a necessary condition for $\mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c}) \neq 0$ is that $(\vec{a},\vec{b},\vec{c})$ satisfies the matching condition.

Lemma 6.2. *There exists a constant C with the following property: for any triple $(\vec{a},\vec{b},\vec{c})$ satisfying the matching condition, we have*

$$\mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c}) \leqslant (Ck)^{\frac{3}{2}(2k+2-\ell(\vec{a},\vec{b},\vec{c}))}.$$

This bound is sharp since $\mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c}) = 1$ when $(\vec{a},\vec{b},\vec{c})$ is admissible.

Proof. Consider the $2k$ factors in the product $\Pi(\vec{a},\vec{b},\vec{c})$:

$$G^{c_1}_{a_1,b_2}, G^{c_1}_{a_2,b_1}, G^{c_2}_{a_2,b_3}, G^{c_2}_{a_3,b_2}, \cdots, G^{c_{k-1}}_{a_{k-1},b_k}, G^{c_{k-1}}_{a_k,b_{k-1}}, G^{c_k}_{a_k,b_1}, G^{c_k}_{a_1,b_k}. \tag{11}$$

Using the independence of entries of G, we may write $\mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c})$ as a product of quantities of the form $\mathbf{E}(G^c_{a,b})^q$. If G is a $N(0,1)$ random variable, then $\mathbf{E}\,|G|^{2n}$ equals $1\cdot 3\cdot 5\cdots(2n-1)$ in the real case and $n!$ in the complex case. In any case, for some constant C,

$$\mathbf{E}\,|G|^q \begin{cases} = 1 & \text{if } q=2, \\ \leqslant (C\sqrt{q})^q & \text{if } q>2. \end{cases} \tag{12}$$

Let $n_2(\vec{a},\vec{b},\vec{c})$ be the number of indices i such that the ith factor from the list (11) appears exactly twice, and $n_{>2}(\vec{a},\vec{b},\vec{c})$ be the number of indices i such that the ith factor appears 3 times or more. The matching condition implies that $n_2(\vec{a},\vec{b},\vec{c}) + n_{>2}(\vec{a},\vec{b},\vec{c}) = 2k$. Bounding each individual factor according to (12) and using $q \leqslant 2k$ leads to

$$\mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c}) \leqslant (C\sqrt{2k})^{n_{>2}(\vec{a},\vec{b},\vec{c})}.$$

Let also $d(\vec{a},\vec{b},\vec{c})$ be the total number of different terms appearing in the list (11). We have

$$\ell(\vec{a},\vec{b},\vec{c}) \leqslant 2\max(\ell_W(\vec{a},\vec{c}),\ell_W(\vec{b},\vec{c})) \leqslant 2\max(d_W(\vec{a},\vec{c})+1, d_W(\vec{b},\vec{c})+1) \leqslant 2d(\vec{a},\vec{b},\vec{c})+2,$$

where the second inequality follows from lemma 3.2. On the other hand,

$$d(\vec{a},\vec{b},\vec{c}) \leqslant \frac{1}{2}n_2(\vec{a},\vec{b},\vec{c}) + \frac{1}{3}n_{>2}(\vec{a},\vec{b},\vec{c}) = k - \frac{1}{6}n_{>2}(\vec{a},\vec{b},\vec{c}),$$

so that $n_{>2}(\vec{a},\vec{b},\vec{c}) \leqslant 6(k-d(\vec{a},\vec{b},\vec{c}))$. Therefore,

$$\mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c}) \leqslant (C\sqrt{k})^{6(k-d(\vec{a},\vec{b},\vec{c}))} \leqslant (C\sqrt{k})^{3(2k+2-\ell(\vec{a},\vec{b},\vec{c}))}. \qquad \square$$

We gather the terms in the sum (10) according to the value of $\ell = \ell(\vec{a},\vec{b},\vec{c})$. Note that we may drop triples which do no satisfy the matching condition. For remaining triples, the values of ℓ range from 4 to $2k+2$ (see the remark following fact 4.1). Let β_ℓ be the number of equivalence classes of triples $(\vec{a},\vec{b},\vec{c})$ satisfying the matching condition and such that $\ell(\vec{a},\vec{b},\vec{c}) = \ell$. The following lemma gives an upper bound on β_ℓ (we postpone the proof).

Lemma 6.3. *For $4 \leqslant \ell \leqslant 2k+2$, we have the bound $\beta_\ell \leqslant C^k k^{2k+2-\ell}$.*

The number of triples in $[d]^k \times [d]^k \times [p]^k$ which are equivalent to a given triple $(\vec{a},\vec{b},\vec{c})$ is bounded by

$$d^{\#\vec{a}} d^{\#\vec{b}} p^{\#\vec{c}} = d^{\ell(\vec{a},\vec{b},\vec{c})-2\#\vec{c}} p^{\#\vec{c}} \leqslant \alpha^{\#\vec{c}} d^{\ell(\vec{a},\vec{b},\vec{c})}.$$

Also, a multi-index $\vec{c} \in [p]^k$ which is paired must satisfy $\#\vec{c} \leqslant k/2$. Therefore, the number of triples $(\vec{a},\vec{b},\vec{c})$, satisfying the matching condition, with $\ell(\vec{a},\vec{b},\vec{c}) = \ell$ and such that $\vec{c}$ is paired, is bounded by $\beta_\ell \alpha^{k/2} d^\ell$. Using lemma 6.2, we have

$$\mathbf{E}\,\|Y-\mathrm{Id}\|^k \leqslant \frac{2^k}{p^k} \sum_{\ell=4}^{2k+2} \beta_\ell \alpha^{k/2} d^\ell (Ck)^{\frac{3}{2}(2k+2-\ell)} \leqslant \left(\frac{C}{\sqrt{\alpha}}\right)^k d^2 \sum_{\ell=4}^{2k+2} \left(\frac{k^{5/2}}{d}\right)^{2k+2-\ell}.$$

We now choose k to be the largest even integer such that $k^{5/2} \leqslant d$. This gives (since $d^2(2k-1) \leqslant C^k$)

$$\mathbf{E}\,\|Y-\mathrm{Id}\|^k \leqslant \left(\frac{C}{\sqrt{\alpha}}\right)^k.$$

Proof of lemma 6.3. The map

$$(\vec{a},\vec{b},\vec{c}) \mapsto ((\vec{a},\vec{c}),(\vec{b},\vec{c}))$$

is certainly injective. If $(\vec{a},\vec{b},\vec{c})$ is admissible, then $(\vec{a},\vec{c})$ and $(\vec{b},\vec{c})$ are Wishart-admissible. By lemma 3.5, (denoting $\ell_1 = \#\vec{a} + \vec{c}$ and $\ell_2 = \#\vec{b} + \#\vec{c}$), this leads to the bound

$$\beta_\ell \leqslant \sum_{\ell_1+\ell_2=\ell} \left(C^k k^{k-\ell_1}\right) \cdot \left(C^k k^{k-\ell_2}\right) \leqslant C^k k^{2k-\ell}. \qquad \square$$

7. Relevance to Quantum Information Theory

In this section we consider finite-dimensional complex Hilbert spaces. We write $\mathcal{M}(\mathbf{C}^d)$ for the space of linear operators (=matrices) on $\mathbf{C}^d$.

7.1. **PPT states.** A *state* (=density matrix) ρ on $\mathbf{C}^n$ is a positive operator on $\mathbf{C}^n$ with trace 1. We write $\mathcal{D}(\mathbf{C}^n)$ for the set of states on $\mathbf{C}^n$. A *pure state* is a rank one state and is denoted $\rho = |x\rangle\langle x|$, where x is a unit vector in the range of ρ. We typically consider the case $\mathbf{C}^n \simeq \mathbf{C}^d \otimes \mathbf{C}^d$. We have the following canonical identification

$$\mathcal{M}(\mathbf{C}^d \otimes \mathbf{C}^d) \simeq \mathcal{M}(\mathbf{C}^d) \otimes \mathcal{M}(\mathbf{C}^d).$$

A state $\rho \in \mathcal{D}(\mathbf{C}^d \otimes \mathbf{C}^d)$ if called *separable* if it can be written as a convex combination of product states. A state ρ is called PPT ("positive partial transpose") if ρ^Γ is a positive operator (the partial transposition $\rho^\Gamma = (\mathrm{Id} \otimes T)\rho$ was defined in (2)). The partial transposition of a *separable state* ρ is always positive [16] ; however there exist non-separable (=entangled) PPT states. However, checking positivity of partial transpose is the most useful tool to detect entanglement. We refer to the survey [10] for more information about PPT states and entanglement.

7.2. **Random induced states are normalized Wishart matrices.** There is a canonical probability measure on the set of pure states on any finite-dimensional Hilbert space H, obtained by pushing forward the uniform measure on the unit sphere of H under the map $x \mapsto |x\rangle\langle x|$. We define the measure $\mu_{n,p}$ to be the distribution of $\mathrm{Tr}_{\mathbf{C}^p} |x\rangle\langle x|$, where x is uniformly distributed on the unit sphere of $\mathbf{C}^n \otimes \mathbf{C}^p$. The *partial trace* $\mathrm{Tr}_{\mathbf{C}^p}$ is the linear operation

$$\mathrm{Tr}_{\mathbf{C}^p} := \mathrm{Id}_{\mathcal{M}(\mathbf{C}^n)} \otimes \mathrm{Tr} : \mathcal{M}(\mathbf{C}^n \otimes \mathbf{C}^p) \to \mathcal{M}(\mathbf{C}^n),$$

where $\mathrm{Id}_{\mathcal{M}(\mathbf{C}^n)}$ is the identity operation on $\mathcal{M}(\mathbf{C}^n)$ and $\mathrm{Tr} : \mathcal{M}(\mathbf{C}^p) \to \mathbf{C}$ is the usual trace.

The measure $\mu_{n,p}$ is a probability measure on the set $\mathcal{D}(\mathbf{C}^n)$ of pure states on $\mathbf{C}^n$. These family of measures are called *induced measures*; the space $\mathbf{C}^p$ is called the *ancilla space*. This family of measures has a simple physical motivation: they can be used if our only knowledge about a state is the dimensionality of the environment (see [5], Section 14.5 and references therein).

The induced measures are very close to the Wishart distributions. Indeed, if W is a (n,p)-Wishart random matrix, then $\frac{1}{\mathrm{Tr}\, W} W$ is distributed according to $\mu_{n,p}$. Moreover, the random variables $\mathrm{Tr}\, W$ and $\frac{1}{\mathrm{Tr}\, W} W$ are independent (this fact explicitly appears in [13]). Therefore, results about Wishart matrices can be easily translated in the language of induced measures. The special case $p = n$, when the dimension of the ancilla equals the dimension of the system, deserves to be highlighted, thanks to the following proposition [17].

Proposition 7.1. *The measure $\mu_{n,n}$ is equal to the normalized Lebesgue measure restricted to the set $\mathcal{D}(\mathbf{C}^n)$.*

7.3. **Partial transposition of random induced states.** Our main results admits an immediate translation in the language of induced random states. Here is a version of theorem 1 for induced states.

Theorem. *Fix $\alpha > 0$. For each d, let ρ_d be a random mixed state on $\mathbf{C}^d \otimes \mathbf{C}^d$ chosen according to the measure $\mu_{d^2,\lfloor \alpha d^2 \rfloor}$. Then the sequence $\mu_{d^2 \rho_d^\Gamma}$ converges ($d \to \infty$) weakly, in probability, towards the semicircular distribution $SC(1, 1/\alpha)$.*

Proof. One may for example check that lemmas 2.2 and 2.3 also hold for the random induced states. Let W_d be a $(d^2, \lfloor \alpha d^2 \rfloor)$-Wishart random matrix. Then W_d has the same distribution as the product $Z \cdot d^2 \rho_d$, where $Z = \frac{\operatorname{Tr} W_d}{d^2}$ is independent from $d^2 \rho_d$. The random variable Z is especially simple to analyze since it follows a normalized χ^2 distribution. One easily checks (e.g. by expanding moments) that for every $k \in \mathbf{N}$,

$$\lim_{d \to \infty} \mathbf{E}\, Z^k = 1,$$
$$\lim_{d \to \infty} \mathbf{Var}\, Z^k = 0,$$

from which the lemmas (hence the theorem) follow. □

When d is fixed, the induced measures $\mu_{d^2,p}$ concentrate towards the maximally mixed state on $\mathbf{C}^d \otimes \mathbf{C}^d$ when p increases. For small values of p, one expects to get typically very entangled states. Therefore one can consider the critical p for which the property "being PPT" becomes typically true. The next result follows immediately from the analogous statement for Wishart matrices.

Theorem. *Fix $\alpha > 0$, and for each d, let ρ_d be a random mixed state on $\mathbf{C}^d \otimes \mathbf{C}^d$ chosen according to the measure $\mu_{d^2,\lfloor \alpha d^2 \rfloor}$.*

1. *If $\alpha < 4$, then the probability that ρ_d is PPT tends to 0 when d tends to $+\infty$.*
2. *If $\alpha > C$, then the probability that ρ_d is PPT tends to 1 when d tends to $+\infty$,*

where $C \geqslant 4$ is an absolute constant.

A positive answer to conjecture 2.5 would imply that $C = 4$.

8. Miscellaneous remarks

8.1. **Unbalanced bipartite systems.** We may apply partial transposition to any decomposition $\mathbf{C}^{d^2} \simeq \mathbf{C}^{d_1} \otimes \mathbf{C}^{d_2}$, with $d_1 d_2 = d^2$. Provided the ratio d_1/d_2 stays away from 0 and ∞, theorems 1 and 2 remain valid. The point is that the main contributions come from terms in which $\vec{a} \sim \vec{b}$, so that $d_1^{\#\vec{a}} d_2^{\#\vec{b}}$ depends only on the product $d_1 d_2$. The situation is radically different when d_1 or d_2 is fixed and $d \to \infty$.

8.2. **Free probability.** The appearance of the semicircular distribution as a limit object, as well as the use of non-crossing partitions, strongly suggests connections with free probability. Let G be a $d^2 \times \lfloor \alpha d^2 \rfloor$ matrix with independent $N(0,1)$ entries, and let (G_i) be the column vectors of G. Then,

$$(GG^\dagger)^\Gamma = \sum_{i=1}^{\lfloor \alpha d^2 \rfloor} |G_i\rangle\langle G_i|^\Gamma.$$

This is a sum of independent random matrices. Asymptotic freeness of random matrices, together with the free central limit theorem, suggest that the limit should be semicircular. However the situation here does not seem (to my knowledge) to be covered by the standard literature and I would be very interested in a proof of theorem 1 along these lines.

Note that taking partial transpose is fundamental. If we look at $GG^\dagger$ instead of $(GG^\dagger)^\Gamma$, one obtains the Marčenko–Pastur distribution: the (free) Poissonnian approximation is then relevant.

8.3. **Volume of the PPT convex body.** How many states have a positive partial transpose ? This question may be formulated using the Lebesgue measure (or "volume") induced by the Hilbert–Schmidt scalar product, or equivalently (cf proposition 7.1) by the induced measure over an ancilla equal dimension. Let W_d be a (d^2, d^2)-Wishart random matrix. It was shown in [3] (formulated as a lower bound on the volume of the set of PPT states, and using techniques from high-dimensional convexity) that for some constant $c > 0$

$$\mathbf{P}(W_d^\Gamma \geqslant 0) \geqslant \exp(-cd^4). \tag{13}$$

By theorem 1, the probability on the left-hand side tends to 0 when d tends to $+\infty$. How fast it goes to zero is actually a question about large deviations. For standard models of random matrices, very precise results are known about large deviations (see e.g. [1], section 2.6.2), and one may expect the lower bound from (13) to be sharp. This would show that PPT states are not so common in large dimensions.

References

[1] G. Anderson, A. Guionnet and O. Zeitouni, An Introduction to Random Matrices, *Cambridge Studies in Advanced Mathematics* **118** (2009).

[2] G. Aubrun, Sampling convex bodies: a random matrix approach, *Proc. Amer. Math. Soc.* **135** (2007), no. 5, 1293–1303 (electronic).

[3] G. Aubrun and S. Szarek, Tensor product of convex sets and the volume of separable states on N qudits, *Phys. Rev. A* **73** (2006)

[4] Z. Bai and J. Silverstein, Spectral analysis of large dimensional random matrices. Second edition. *Springer Series in Statistics* (2010).

[5] I. Bengtsson and K. Życzkowski, Geometry of quantum states. An introduction to quantum entanglement. Cambridge University Press, Cambridge (2006).

[6] R. Donaghey and L. W. Shapiro, Motzkin numbers, *J. Comb. Th. Ser. A* **23** (1977), no. 3, 291–301.

[7] S. Geman, A limit theorem for the norm of random matrices, *Ann. Prob.* **8** (1980), no. 2, 252–261.

[8] M. B. Hastings, *Superadditivity of communication capacity using entangled inputs.* Nature Physics **5**, 255 (2009).

[9] M. Horodecki, P. Horodecki and R. Horodecki, Mixed-State Entanglement and Distillation: Is there a "Bound" Entanglement in Nature?, *Phys. Rev. Lett.* **80** 5239–5242 (1998).

[10] R. Horodecki, P. Horodecki, M. Horodecki and K. Horodecki, Quantum entanglement, *Rev. Mod. Phys.* **81**, No. 2, pp. 865-942 (2009).

[11] D. Jonsson, Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* **12** (1982), no. 1, 1–38.

[12] V.A. Marčenko and L.A. Pastur, Distribution of eigenvalues in certain sets of random matrices, *Mat. Sb. (N.S.)*, **72** (114), 507–536 1967.

[13] I. Nechita, Asymptotics of random density matrices, *Ann. Henri Poincaré* **8** (2007), no. 8, 1521–1538.

[14] A. Nica and R. Speicher, Lectures on the Combinatorics of Free Probability, London Mathematical Society Lecture Note Series **335** (2006).

[15] F. Oravecz and D. Petz, On the eigenvalue distribution of some symmetric random matrices, *Acta Sci. Math. (Szeged)* **63** (1997), no. 3-4, 383–395.

[16] A. Peres, Separability criterion for density matrices, *Phys. Rev. Lett.* **77**, 1413–1415 (1996).

[17] K. Życzkowski and H.-J. Sommers, Induced measures in the space of mixed quantum states, *J. Phys. A* **34** (2001), no. 35, 7111–7125.

Institut Camille Jordan, Université Claude Bernard Lyon 1, 43 boulevard du 11 novembre 1918, 69622 Villeurbanne CEDEX, France

E-mail address: aubrun@math.univ-lyon1.fr