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 $d \to \infty$. We also show the convergence of extreme eigenvalues towards the edge of the expected spectrum. 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. A natural question is: "When does a random state have a positive partial transpose (PPT)?". We answer this question and exhibit a strong threshold when the parameter from the Wishart distribution equals 4. When d gets large, a random state on $\mathbf{C}^d \otimes \mathbf{C}^d$ obtained after partial tracing a random pure state over some ancilla of dimension αd^2 is typically PPT when $\alpha > 4$ and typically non-PPT when $\alpha < 4$.

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 [11]). 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. 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, the empirical eigenvalue distribution converges towards a *non-centered* semicircular distribution, and the extreme eigenvalues converge towards the edge of the spectrum. These results were observed numerically by Žnidarič et al. [23]. The aim of the present paper is to give a complete proof of these facts. We rely on a standard tool from Random Matrix Theory: the method of moments.

The fact that the limiting distribution is semicircular is not a complete surprise. In the context of free probability, semicircular distributions are the non-commutative analogue of Gaussian distributions,

1991 *Mathematics Subject Classification.* 60B20, 81P45.

Key words and phrases. Partial transposition, Wishart matrices, semicircular distribution, PPT states.

This research was supported by the *Agence Nationale de la Recherche* grants ANR-08-BLAN-0311-03 and ANR-2011-BS01-008-01, and by the *Institut Mittag-Leffler* in Stockholm. I also thank S. Szarek and I. Nechita for useful comments.

and therefore one expects their appearance in limit theorems. For example Wigner's celebrated theorem identifies the centered semicircular distribution as the limit distribution of eigenvalues of random Hermitian matrices. However, other limiting distributions do appear in the theory: for example, the Wishart matrices themselves (i.e., *without* the partial transposition) converge to the so-called Marčenko–Pastur law (see section 2.3). Moreover, our model brings some additional exoticism since the limiting distribution is non-centered.

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.

The partial transposition appears to play a central role in Quantum Information Theory and is closely related to the concept of entanglement. An important class of states is the family of states with a Positive Partial Transpose (PPT). Non-PPT states are necessarily entangled [22] and this is the simplest test to detect entanglement. Let us simply mention a related important open problem known as the distillability conjecture [13]: 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. Early notable works about entanglement of random states include [16] and [12]. 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 $p/d^2 > 4$. This shows that a threshold for the PPT property occurs at $p = 4d^2$.

Organization. The paper is organized as follows: Sections 2–7 are written in the language of Random Matrix Theory and contain the proof of our theorems. Section 2 introduces the model and states Theorem 1 (convergence towards the non-centered semicircle distribution) and Theorem 2 (convergence of 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 heavily. Section 4 shows how to derive Theorem 1 from moment estimates ; the proof of these estimates (the heart of the moments method) is deferred to Sections 5 and 6. Section 7 contains the proof of Theorem 2. Section 8 connects to Quantum Information Theory. Section 9 contains some general remarks and possible variations on the model. A high-level non-technical overview of the result of this paper and of a related article [3] can be found in [4].

2. Background and statement of the main theorem

2.1. **Conventions.** By the letters $C, C_0, c, \ldots$ we denote absolute constants, whose value may change from occurrence to occurrence. The integer part of a real number x is denoted by $\lfloor x \rfloor$. We denote by $[k]$ the set $\{1, \ldots, k\}$. Addition in $[k]$ is understood modulo k. We denote by $\vec{a}, \vec{b}, \vec{c}, \ldots$ multi-indices which are elements of $\mathbf{N}^k$ for some integer k. The coordinates of $\vec{a}$ are denoted $(a_1, \ldots, a_k)$.

When $\vec{a} \in \mathbf{N}^k$, we denote by $\#\vec{a}$ the number of distinct elements which appear in the set $\{a_1, \ldots, a_k\}$. For example, $\#(1, 4, 1, 2) = 3$. The cardinality of a set A is denoted $\operatorname{card} A$. The notation $\mathbf{1}_E$ denotes a quantity which equals 1 when the event E is true, and 0 otherwise.

By $\|A\|_\infty$ or simply $\|A\|$ we denote the operator norm of a matrix A.

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

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

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 $C_k = \frac{1}{k+1}\binom{2k}{k}$,

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

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*, $\mu_{MP(\alpha)}$, is the following probability distribution

- 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 distributions.** 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. Our results are valid for both, although only the complex case 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 N_A, is the probability measure on Borel subsets of $\mathbf{R}$ defined as

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

In other words, $N_A(B)$ 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, [18])**.** *Fix $\alpha > 0$. For every n, let W_n be a $(n, \lfloor \alpha n \rfloor)$-Wishart matrix. Then the empirical eigenvalue distribution of W_n approaches a Marčenko–Pastur distribution $MP(\alpha)$ in the following sense. For every interval $I \subset \mathbf{R}$ and any $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbf{P}(|N_{W_n}(I) - \mu_{MP(\alpha)}(I)| > \varepsilon) = 0.$$

[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, $\mathbf{E}\,|\xi|^2 = 1$.

2.4. **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

$$(A^\Gamma)_{i,j}^{k,l} = A_{i,j}^{l,k}. \tag{1}$$

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 every 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 empirical eigenvalue distribution of Y_d approaches the semicircular distribution $\mu_{SC(1,1/\alpha)}$ in the following sense. For every interval $I \subset \mathbf{R}$ and any $\varepsilon > 0$,*

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

Recall that $N_{Y_d}(I)$ is the proportion of eigenvalues of the matrix Y_d that belong to the interval I.

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 $[1 - \frac{2}{\sqrt{\alpha}}, 1 + \frac{2}{\sqrt{\alpha}}]$. Denote by $\lambda_{\min}(A)$ (resp. $\lambda_{\max}(A)$) the smallest (resp. largest) eigenvalue of a matrix A. A natural (and harder) question is whether the extreme eigenvalues of Y_d converge towards $1 \pm \frac{2}{\sqrt{\alpha}}$. We show that this is indeed the case:

Theorem 2. *Fix $\alpha > 0$. For every 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, for every $\varepsilon > 0$,*

$$\lim_{d \to \infty} \mathbf{P}\left(\left| \lambda_{\max}(Y_d) - (1 + 2/\sqrt{\alpha}) \right| > \varepsilon \right) = 0,$$

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

$$\lim_{d\to\infty} \mathbf{P}\left(\left|\lambda_{\min}(Y_d) - (1 - 2/\sqrt{\alpha})\right| > \varepsilon\right) = 0.$$

2.6. **Almost sure convergence.** In Random Matrix Theory, it is customary to work with the stronger notion of almost sure convergence. This requires to define all the objects on a single probability space. Such a construction is not natural from a Quantum Information Theory point of view, which usually "avoids infinity" and prefers to work in a fixed (but large) dimension.

However, from a mathematical point of view, it is interesting to note that the results presented here also hold for almost sure convergence. One needs to check that the proof gives enough concentration in order to use the Borel–Cantelli lemma. A key point is the $O(1/d^2)$ estimate for the variance from Proposition 4.2.

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, and additions in $[k]$ are understood modulo k. It is useful to represent elements of S as points on a circle. We introduce the concept of non-crossing partitions and refer to [20] 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 $|\pi|$. We denote $\sim_\pi$ the equivalence relation on S induced by π: $i \sim_\pi j$ means that i and j belong to the same block.
- 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 $i \sim_\pi k, j \sim_\pi l$ and $i \not\sim_\pi j$. 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 $NC_2(k) = NC_2([k])$.

Counting non-crossing partitions is a well-known combinatorial problem involving Catalan numbers (see [20], Lemma 8.9 and Proposition 9.4).

Lemma 3.1. *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 $C_k = \frac{1}{k+1}\binom{2k}{k}$.*

Let us also introduce the *Kreweras complementation* as the map $K : NC(k) \mapsto NC(k)$ defined as follows. For $\pi \in NC(\{1^-, \ldots, k^-\}) \simeq NC(k)$, $K(\pi)$ is defined as the coarsest partition $\sigma \in NC(\{1^+, \ldots, k^+\}) \simeq NC(k)$ such that $\pi \cup \sigma$ is a non-crossing partition of $\{1^-, 1^+, \ldots, k^-, k^+\}$, equipped with the order

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

The map K is bijective. Moreover, given $\sigma \in NC(\{1^+, \ldots, k^+\}) \simeq NC(k)$, one can recover $K^{-1}(\sigma)$ as the coarsest partition $\pi \in NC(\{1^-, \ldots, k^-\}) \simeq NC(k)$ such that $\pi \cup \sigma$ is a non-crossing partition of $\{1^-, 1^+, \ldots, k^-, k^+\}$. See [20] for more details.

The following lemma will be used in connection to partial transposition.

Lemma 3.2. *Let $\pi \in NC(k)$ a non-crossing partition and $K(\pi)$ its Kreweras complement. Then,*

(1) *For every index $i \in [k]$,*

$$\text{The singleton } \{i\} \text{ is a block in } K(\pi) \iff i \sim_\pi i + 1.$$

(2) *For every distinct indices* $i, j \in [k]$,

$$\text{The pair } \{i,j\} \text{ is a block in } K(\pi) \iff i \sim_\pi j+1 \text{ and } i+1 \sim_\pi j \text{ and } i \not\sim_\pi j.$$

Proof. This is geometrically obvious. □

3.2. **Combinatorics related to Wishart matrices.** We now remind the reader about the (standard) proof of the Marčenko–Pastur theorem via the moments method. This proof can be found for example in [15, 21] or the book [5]. 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

$$\begin{aligned} \mathbf{E}\,\frac{1}{n}\operatorname{Tr} W_n^k &= \frac{1}{n}\sum_{\vec{a}\in[n]^k} \mathbf{E}\, W_{a_1,a_2}W_{a_2,a_3}\cdots W_{a_k,a_1} \\ &= \frac{1}{np^k}\sum_{\vec{a}\in[n]^k,\vec{c}\in[p]^k} \mathbf{E}\, G_{a_1,c_1}\overline{G_{a_2,c_1}}G_{a_2,c_2}\overline{G_{a_3,c_2}}\cdots G_{a_k,c_k}\overline{G_{a_1,c_k}}. \end{aligned} \tag{2}$$

The next task is to analyze which couples $(\vec{a},\vec{c})$ give dominant contributions to the sum (2) when $n \to \infty$ and $p = \lfloor \alpha n \rfloor$. One argues as follows. First, if one couple (a_i, c_i) or (a_{i+1}, c_i) appears a odd number of times in the product, then the contribution is exactly zero (because entries of G are independent and symmetric). This motivates the following definition:

Definition. A couple $(\vec{a},\vec{c}) \in \mathbf{N}^k \times \mathbf{N}^k$ satisfies the *Wishart matching condition* if every couple in the following list of $2k$ elements appears an even number of times:

$$(a_1,c_1),(a_2,c_1),(a_2,c_2),(a_3,c_2),\ldots,(a_k,c_k),(a_1,c_k). \tag{3}$$

Let $(\vec{a},\vec{c}) \in \mathbf{N}^k \times \mathbf{N}^k$. We define $d_W(\vec{a},\vec{c})$ as the number of distinct couples appearing in the list (3), and set $\ell_W(\vec{a},\vec{c}) = \#\vec{a} + \#\vec{c}$. We also denote $n_2(\vec{a},\vec{c})$ the number of indices i such that the ith element appears exactly twice in the list (3), and $n_+(\vec{a},\vec{c})$ the number of indices i such that the ith element appears at least 4 times. Note that $n_2(\vec{a},\vec{c}) + n_+(\vec{a},\vec{c}) = 2k$. These parameters satisfy some inequalities:

Lemma 3.3. *Let* $(\vec{a},\vec{c}) \in \mathbf{N}^k \times \mathbf{N}^k$ *satisfy the Wishart matching condition. Then*

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

Moreover, $n_+(\vec{a},\vec{c}) \leqslant 4(k+1-\ell_W(\vec{a},\vec{c}))$.

Proof. Read the list (3) 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 appear previously in the list (there are $d_W(\vec{a},\vec{c})-1$ such couples) may bring at most one new index (since it shares a common index with the couple just before). This shows that $\ell_W(\vec{a},\vec{c}) \leqslant d_W(\vec{a},\vec{c}) + 1$.

The inequality $d_W(\vec{a},\vec{c}) \leqslant k$ is easy: if every couple in the list (3) appears at least twice, then this list contains at most k different couples.

For the last claim, note that

$$d_W(\vec{a},\vec{c}) \leqslant \frac{1}{2}n_2(\vec{a},\vec{c}) + \frac{1}{4}n_+(\vec{a},\vec{c}) = k - \frac{1}{4}n_+(\vec{a},\vec{c}),$$

with equality iff no element in the list (3) appears 6 times or more. □

Now, the couples $(\vec{a},\vec{c})$ satisfying $\ell_W(\vec{a},\vec{c}) < k+1$ are easily shown to have a contribution to the sum (2) 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 $\ell_W(\vec{a},\vec{c}) = k+1$.

If $\vec{a} \in \mathbf{N}^k$, the *partition induced by* $\vec{a}$, denoted $\pi(\vec{a})$, is the partition of $[k]$ defined as follows: i and j belong to the same block if and only if $a_i = a_j$. We say that $\vec{a},\vec{b} \in \mathbf{N}^k$ are *equivalent* ($\vec{a} \sim \vec{b}$) if $\pi(\vec{a}) = \pi(\vec{b})$. 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 [15] or [21] for details) characterizes the combinatorial structure of (equivalence classes of) Wishart-admissible couples.

Proposition 3.4. *For every integer k,*

(a) *If $(\vec{a},\vec{c}) \in \mathbf{N}^k \times \mathbf{N}^k$ is Wishart-admissible, then*
  (i) *Each couple in the list* (3) *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). Moreover, the pair-partition of $[2k]$ induced by the list* (3) *is non-crossing.*
  (ii) *The partitions $\pi(\vec{a})$ and $\pi(\vec{c})$ are non-crossing, and Kreweras-complementary: $\pi(\vec{c}) = K(\pi(\vec{a}))$. In particular, $\vec{a}$ is determined by $\vec{c}$ up to equivalence.*

(b) *The mapping $(\vec{a},\vec{c}) \mapsto \pi(\vec{c})$ induces a bijection between the set of equivalence classes of Wishart-admissible couples in $\mathbf{N}^k \times \mathbf{N}^k$ 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 (3) reads as

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

Indeed, each couple appears exactly twice. The partition induced by this list is

$$\{\{1,8\},\{2,5\},\{3,4\},\{6,7\}\}$$

while the partitions induced by $\vec{a}$ and $\vec{c}$ are

$$\pi(\vec{a}) = \{\{1\},\{2,3\},\{4\}\},$$
$$\pi(\vec{c}) = K(\pi(\vec{a})) = \{\{1,3,4\},\{2\}\}.$$

From Proposition 3.4, it is easy to check (if $p \sim \alpha n$) 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 $1/\alpha$. 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 3.5. *Let $(\vec{a},\vec{c})$ and $(\vec{a}\,'\vec{c}\,')$ be two couples in $\mathbf{N}^k \times \mathbf{N}^k$ satisfying the following conditions*

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

$$(a_1,c_1),(a_2,c_1),\ldots,(a_k,c_k),(a_1,c_k)\,;\,(a'_1,c'_1),(a'_2,c'_1),\ldots,(a'_k,c'_k),(a'_1,c'_k). \tag{4}$$

(ii) *At least some couple appears both in the left half and in the right half of the list* (4).

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

Proof. As before, we read the list (4) 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 an 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. □

If we want to prove estimates on the extreme eigenvalues of Wishart matrices, we also have to analyze lower-order contributions. We here follow the terminology from [10]. Let $(\vec{a}, \vec{c}) \in \mathbf{N}^k \times \mathbf{N}^k$ satisfy the Wishart matching condition. The elements from the list (3) fall into one of the following categories.

type 1: innovations for $\vec{a}$.
type 2: innovations for $\vec{c}$.
type 3: first repetitions of an innovation.
type 4: other elements.

The ith element in the list (3) is an *innovation* if it contains one index which did not appear already in the list. When $i = 2p$ is even, the ith element is an *innovation for* $\vec{a}$ if $a_{p+1} \notin \{a_j \ : \ j < p\}$. When $i = 2p-1$ is odd, the ith element is an *innovation for* $\vec{c}$ if $c_p \notin \{c_j \ : \ j < p\}$. In particular, the first element of the list (3) is always an innovation for $\vec{c}$.

The ith element is the *first repetition* of an innovation if there is a unique $j < i$ such that the jth element from the list (3) equals the ith element, and moreover this jth element is an innovation.

The following lemma asserts that there are few different couples satisfying the Wishart matching condition which have the same types of elements at the same positions. We refer to [10] for a proof.

Lemma 3.6. *Let $T = (t_1, \dots, t_{2k}) \in \{1,2,3,4\}^{2k}$, and let $U = \mathrm{card}\{i \in [2k] \ : \ t_i = 4\}$. Say that $(\vec{a}, \vec{c})$ is of type T if, for every $i \in [2k]$, the ith element in the list (3) has type t_i. Then, the number of equivalence classes of couples satisfying the Wishart matching condition which are of type T is bounded by k^{3U}.*

3.3. **Diagonal elements of Wishart matrices are close to 1.** We will use the following simple fact in our proof.

Lemma 3.7. *Let $W = (W_{ij})$ be a (n,p)-Wishart matrix. Then, for any $\varepsilon \in (0,1)$, we have*

$$\mathbf{P}\left(1 - \varepsilon \leqslant \inf_{1 \leqslant i \leqslant n} W_{ii} \leqslant \sup_{1 \leqslant i \leqslant n} W_{ii} \leqslant 1 + \varepsilon\right) \geqslant 1 - Cn\exp(-cp\varepsilon^2),$$

where $C, c > 0$ are absolute constants.

Proof. Recall that $W = \frac{1}{p} GG^\dagger$, where $G = (G_{ij})$ is a $n \times p$ matrix with independent $N(0,1)$ entries, so that the diagonal terms of W_n follow a χ^2 distribution

$$W_{ii} = \frac{1}{p} \sum_{j=1}^{p} |G_{ij}|^2.$$

The next fact shows that such distributions enjoy very strong concentration properties.

Fact 3.8. *Let $g_1, \dots, g_p$ denote independent (real or complex) $N(0,1)$ random variables, and X be the Euclidean norm of the vector $(g_1, \dots, g_p)$. Then for every $t > 0$,*

$$\mathbf{P}\left(|X - \sqrt{p}| > t\right) \leqslant C' \exp(-c' t^2).$$

Fact 3.8 can be proved by direct calculation or follows from concentration of measure (see e.g.[17]). Indeed, the Euclidean norm is a 1-Lipschitz function and the expectation of X satisfies the inequalities $\sqrt{p-1} \leqslant \mathbf{E}\, X \leqslant \sqrt{p}$. Lemma 3.7 follows from Fact 3.8 via the union bound. □

4. Proof of Theorem 1

For an integer d and $p = \lfloor \alpha d^2 \rfloor$, let G_d be a $d^2 \times p$ matrix with independent $N(0,1)$ entries, $W_d = \frac{1}{p} G_d G_d^\dagger$ and $Y_d = W_d^\Gamma$. We denote the entries of G_d as $(G_{i,j}^k)$, where $(i,j) \in [d] \times [d]$ denote the row indices and $k \in [p]$ denotes the column index. We label the entries of W_d and Y_d as $(W_{i,i'}^{j,j'})$ and $(Y_{i,i'}^{j,j'})$, where $(a,a',b,b') \in [d]^4$ according to the convention described in Section 2.4.

We have to show that N_{Y_d}, the empirical eigenvalue distribution of Y_d, approaches a non-centered semicircular distribution $SC(1,\alpha)$. To handle a more symmetric situation (involving a *centered* semicircular distribution), we will rather consider $Y_d - \mathrm{Id}$. By Lemma 3.7, this matrix is very close to $Z_d = Y_d - \mathrm{diag}(Y_d)$. The latter behaves in a nicer way with respect to moments combinatorics. We label the entries of Z_d as $(Z_{i,i'}^{j,j'})_{i,i',j,j' \in [d]}$. We have

$$Z_{i,i'}^{j,j'} = Y_{i,i'}^{j,j'} \mathbf{1}_{(i,j) \neq (i',j')}.$$

The following proposition is central to our work. We defer the proof (the combinatorial part of the moments method) to the next section.

Proposition 4.1. *For every fixed integer k, we have*

$$\lim_{d \to \infty} \mathbf{E}\left(\frac{1}{d^2} \operatorname{Tr}(Z_d^k) \right) = \begin{cases} \alpha^{-k/2} C_{k/2} & \text{if } k \text{ is even,} \\ 0 & \text{otherwise.} \end{cases}$$

We also show that the variance goes to zero—this is actually simpler.

Proposition 4.2. *For every fixed integer k, we have*

$$\lim_{d \to \infty} \mathbf{Var}\left(\frac{1}{d^2} \operatorname{Tr}(Z_d^k) \right) = 0.$$

The proofs of Proposition 4.1 and 4.2 appear in Sections 5 and 6, respectively.

Proof of Theorem 1 (assuming Propositions 4.1 and 4.2). We claim that for any interval $I \subset \mathbf{R}$ and $\varepsilon > 0$,

$$\lim_{d \to \infty} \mathbf{P}\left(\left| N_{Z_d}(I) - \mu_{SC(0,1/\alpha)}(I) \right| > \varepsilon \right) = 0. \tag{5}$$

Deriving this from Propositions 4.1 and 4.2 is a completely standard procedure. We only sketch the proof and refer to [1] (pp 10-11) for more details. Recall that the Catalan numbers C_k satisfy $C_k \leqslant 4^k$, and that the support of the $SC(0,1/\alpha)$ distribution is $[-2/\sqrt{\alpha}, 2/\sqrt{\alpha}]$. We first check that the proportion of eigenvalues outside $J = [-3/\sqrt{\alpha}, 3/\sqrt{\alpha}]$ is asymptotically zero. For every $\varepsilon > 0$ and even integer k,

$$\begin{aligned} \limsup_{d \to \infty} \mathbf{P}\left(N_{Z_d}(J^c) > \varepsilon \right) &\leqslant \frac{1}{\varepsilon} \limsup_{d \to \infty} \mathbf{E}\, N_{Z_d}(J^c) \\ &\leqslant \frac{1}{\varepsilon} \limsup_{d \to \infty} \mathbf{E} \int x^k (\sqrt{\alpha}/3)^k dN_{Z_d} \\ &\leqslant \frac{1}{\varepsilon} (\sqrt{\alpha}/3)^k C_{k/2} \alpha^{-k/2} \\ &\leqslant \frac{1}{\varepsilon} (2/3)^k, \end{aligned}$$

where the second inequality follows from $\mathbf{1}_{J^c}(x) \leqslant x^k (\sqrt{\alpha}/3)^k$. Since k is arbitrarily large, we obtain that $\mathbf{P}(N_{Z_d}(J^c) > \varepsilon)$ tends to 0.

Therefore, to prove (5), we may assume $I \subset J$. Using the Weierstrass approximation theorem, we may find a polynomial $Q \geqslant \mathbf{1}_I$ such that $\int Q d\mu_{SC(0,1/\alpha)} \leqslant \mu_{SC(0,1/\alpha)}(I) + \varepsilon/2$. It follows from Proposition 4.1 that

$$\lim_{d\to\infty} \mathbf{E} \int Q dN_{Z_d} = \int Q d\mu_{SC(0,1/\alpha)}.$$

$$\lim_{d\to\infty} \mathbf{Var} \int Q dN_{Z_d} = 0.$$

For d large enough, $|\mathbf{E}\int Q dN_{Z_d} - \int Q d\mu_{SC(0,1/\alpha)}| < \varepsilon/4$. Then

$$\begin{aligned} \mathbf{P}\left(N_{Z_d}(I) \geqslant \mu_{SC(0,1/\alpha)}(I) + \varepsilon\right) &\leqslant \mathbf{P}\left(\int Q dN_{Z_d} \geqslant \mathbf{E}\int Q dN_{Z_d} + \varepsilon/4\right) \\ &\leqslant \frac{16}{\varepsilon^2}\,\mathbf{Var}\int Q dN_{Z_d} \end{aligned}$$

and this quantity tends to zero. This is only half of (5). The other half follows by noticing that

$$\mathbf{P}\left(N_{Z_d}(I) \leqslant \mu_{SC(0,1/\alpha)}(I) - \varepsilon\right) \leqslant \mathbf{P}\left(N_{Z_d}(J\setminus I) \geqslant \mu_{SC(0,1/\alpha)}(J\setminus I) + \varepsilon/2\right) + \mathbf{P}\left(N_{Z_d}(J^c) \geqslant \varepsilon/2\right)$$

and applying the previous argument to $J \setminus I$.

We now argue that the empirical eigenvalue distribution is stable under small perturbations. Indeed, for any interval $[a,b]$ and any self-adjoint matrix Δ_d with operator norm smaller than δ,

$$N_{Z_d+\Delta_d}([a+\delta, b-\delta]) \leqslant N_{Z_d}([a,b]) \leqslant N_{Z_d+\Delta_d}([a-\delta, b+\delta]). \tag{6}$$

This is a consequence of the minimax formula for eigenvalues (see e.g. [8], Chapter III). We apply (6) with $\Delta_d = \mathrm{diag}(Y_d) - \mathrm{Id}$. By Lemma 3.7, for every $\varepsilon > 0$, $\mathbf{P}(\|\Delta_d\| > \varepsilon)$ tends to 0 when d tends to infinity. We easily derive from (5) and (6) that, for any interval I,

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

This is clearly equivalent to Theorem 1. □

5. Proof of Proposition 4.1

We expand $\mathbf{E}\,\frac{1}{d^2}\,\mathrm{Tr}(Z_d^k)$ and analyze the underlying combinatorics.

$$\begin{aligned} \mathbf{E}\,\frac{1}{d^2}\,\mathrm{Tr}(Z_d^k) &= \frac{1}{d^2} \sum_{\vec{a}\in[d]^k, \vec{b}\in[d]^k} \mathbf{E}\, Z^{b_1,b_2}_{a_1,a_2} \cdot Z^{b_2,b_3}_{a_2,a_3} \dots Z^{b_{k-1},b_k}_{a_{k-1},a_k} \cdot Z^{b_k,b_1}_{a_k,a_1} \\ &= \frac{1}{d^2} \sum_{\vec{a}\in[d]^k, \vec{b}\in[d]^k} \left(\prod_{i=1}^k \mathbf{1}_{(a_i,b_i)\neq(a_{i+1},b_{i+1})}\right) \mathbf{E}\, Y^{b_1,b_2}_{a_1,a_2} \cdot Y^{b_2,b_3}_{a_2,a_3} \dots Y^{b_{k-1},b_k}_{a_{k-1},a_k} \cdot Y^{b_k,b_1}_{a_k,a_1} \\ &= \frac{1}{d^2} \sum_{\vec{a}\in[d]^k, \vec{b}\in[d]^k} M(\vec{a},\vec{b})\, \mathbf{E}\, W^{b_2,b_1}_{a_1,a_2} \cdot W^{b_3,b_2}_{a_2,a_3} \dots W^{b_k,b_{k-1}}_{a_{k-1},a_k} \cdot W^{b_1,b_k}_{a_k,a_1} \\ &= \frac{1}{d^2 p^k} \sum_{\vec{a}\in[d]^k, \vec{b}\in[d]^k, \vec{c}\in[p]^k} M(\vec{a},\vec{b})\, \mathbf{E}\, \Pi(\vec{a},\vec{b},\vec{c}). \end{aligned}$$

where we have defined

$$M(\vec{a},\vec{b}) = \prod_{i=1}^{k} \mathbf{1}_{(a_i,b_i)\neq(a_{i+1},b_{i+1})}$$

and

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

We introduce some definitions in order to restrict ourselves to triples for which both $M(\vec{a},\vec{b})$ and $\mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c})$ are nonzero.

Definition. A couple $(\vec{a},\vec{b}) \in \mathbf{N}^k \times \mathbf{N}^k$ is said to be *non-repeating* if $M(\vec{a},\vec{b}) = 1$. In other words, $(\vec{a},\vec{b})$ is non-repeating if for every $i \in [k]$, either $a_i \neq a_{i+1}$ or $b_i \neq b_{i+1}$.

Because the entries of G_d are independent, we may factorize $\mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c})$ as a product of quantities of the form $\mathbf{E}(G^k_{i,j})^q(\overline{G^k_{i,j}})^r$. Such a quantity is zero unless $q = r$, and $\mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c})$ is zero whenever one of these factors is zero.

Definition. A triple $(\vec{a},\vec{b},\vec{c}) \in \mathbf{N}^k \times \mathbf{N}^k \times \mathbf{N}^k$ satisfies the *matching condition* if, in the following list of $2k$ triples, each triple appears an even number of times

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

Therefore, if a triple $(\vec{a},\vec{b},\vec{c})$ does not satisfy the matching condition, then $\mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c}) = 0$ both in the real and in the complex cases. The following easy observation will be used repeatedly.

Fact 5.1. *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.*

Recall the definition of equivalence introduced just before Proposition 3.4: $\vec{a} \sim \vec{a}\,'$ 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 $\vec{c} \sim \vec{c}\,'$. Let C be the equivalence class of a triple $(\vec{a},\vec{b},\vec{c})$. When $d \to \infty$

$$\mathrm{card}\{C \cap ([d]^k \times [d]^k \times [p]^k)\} \sim d^{\#\vec{a}} d^{\#\vec{b}} p^{\#\vec{c}} \sim \alpha^{\#\vec{c}} d^{\ell(\vec{a},\vec{b},\vec{c})} \tag{8}$$

where we have defined

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

Together with Lemma 3.3, Fact 5.1 implies that whenever $(\vec{a},\vec{b},\vec{c})$ satisfies the matching condition,

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

Let $\mathscr{C}_k$ be the (finite) family of all equivalence classes of triples $(\vec{a},\vec{b},\vec{c}) \in \mathbf{N}^k \times \mathbf{N}^k \times \mathbf{N}^k$ which satisfy the matching condition. Since the quantities $M(\vec{a},\vec{b})$, $\mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c})$ and $\ell(\vec{a},\vec{b},\vec{c})$ depend only on the equivalence class $C \in \mathscr{C}_k$ of the triple $(\vec{a},\vec{b},\vec{c})$, we may abusively write $M(C)$, $\mathbf{E}\,\Pi(C)$ and $\ell(C)$. We also write $\gamma(C)$ to denote $\#\vec{c}$. Note that these quantities do not depend on the dimension d. We rearrange the sum according to equivalence classes of triples:

$$\lim_{d\to\infty} \frac{1}{d^2}\,\mathbf{E}\,\mathrm{Tr}\,Z_d^k \;=\; \frac{1}{\alpha^k} \sum_{C\in\mathscr{C}_k} M(C)\,\mathbf{E}\,\Pi(C) \lim_{d\to\infty} \frac{1}{d^{2k+2}}\,\mathrm{card}\{C \cap ([d]^k \times [d]^k \times [p]^k)\}. \tag{9}$$

Definition 5.2. Let us say that a triple $(\vec{a},\vec{b},\vec{c})$ is *admissible* if the following three conditions are satisfied

(1) $(\vec{a}, \vec{b}, \vec{c})$ satisfies the matching condition,
(2) $(\vec{a}, \vec{b})$ is non-repeating,
(3) $\ell(\vec{a}, \vec{b}, \vec{c}) = 2k + 2$.

Denote by $\mathscr{C}_k^{\mathrm{adm}} \subset \mathscr{C}_k$ the set of equivalence classes of admissible triples.

Equation (9) implies that

$$\lim_{d\to\infty} \frac{1}{d^2} \mathbf{E} \operatorname{Tr} Z_d^k = \frac{1}{\alpha^k} \sum_{C \in \mathscr{C}_k^{\mathrm{adm}}} M(C) \mathbf{E}\, \Pi(C) \alpha^{\gamma(C)}. \tag{10}$$

Proposition 5.3. *If $(\vec{a}, \vec{b}, \vec{c}) \in \mathbf{N}^k \times \mathbf{N}^k \times \mathbf{N}^k$ is admissible, then*

(1) $M(\vec{a}, \vec{b}) = 1$,
(2) $\mathbf{E}\, \Pi(\vec{a}, \vec{b}, \vec{c}) = 1$
(3) *k is even*
(4) $\#\vec{c} = k/2$.

Moreover, the number of equivalence classes of admissible triples in $\mathbf{N}^k \times \mathbf{N}^k \times \mathbf{N}^k$ is equal to the Catalan number $C_{k/2}$.

Once Proposition 5.3 is proved, Proposition 4.1 is immediate from (10).

Proof of Proposition 5.3. The fact that $M(\vec{a}, \vec{b}) = 1$ is just a reformulation of the non-repeating condition. We now check that $\mathbf{E}\, \Pi(\vec{a}, \vec{b}, \vec{c}) = 1$. Indeed, since $(\vec{a}, \vec{c})$ is Wishart-admissible, every element in the list (3) appears exactly twice, once at an odd position and once at an even position. But the same must be true for the list (7), and therefore $\mathbf{E}\, \Pi(\vec{a}, \vec{b}, \vec{c}) = 1$. To check the last two conditions, we rely on the following lemma

Lemma 5.4. *Let $(\vec{a}, \vec{b}, \vec{c}) \in \mathbf{N}^k \times \mathbf{N}^k \times \mathbf{N}^k$ which satisfies the matching condition and such that $(\vec{a}, \vec{b})$ is non-repeating. Then*

(1) *No index in $\vec{c}$ appears only once, and therefore $\#\vec{c} \leqslant \lfloor k/2 \rfloor$,*
(2) $\#\vec{a} + \#\vec{b} \leqslant 2(\lfloor k/2 \rfloor + 1)$.

Proof. By contraposition, suppose that some index c_i appears only once in $\vec{c}$, i.e. that $c_j \neq c_i$ for every $j \neq i$. The matching condition imposes the equality

$$(a_{i+1}, b_i, c_i) = (a_i, b_{i+1}, c_i)$$

which in turn implies $(a_i, b_i) = (a_{i+1}, b_{i+1})$, contradicting the non-repeating property. For the second part of the lemma, we argue differently according to the parity of k

(k odd) Define $(\vec{x}, \vec{y}) \in \mathbf{N}^k \times \mathbf{N}^k$ as follows

$$\vec{x} = (a_1, a_3, \ldots, a_{k-2}, a_k, a_2, a_4, \ldots, a_{k-1}), \qquad \vec{y} = (b_2, b_4, \ldots, b_{k-1}, b_1, b_3, \ldots, b_{k-2}, b_k).$$

The matching condition implies that $(\vec{x}, \vec{y})$ is Wishart-admissible. Therefore, by Lemma 3.3, we have $\#\vec{x} + \#\vec{y} \leqslant k + 1$. Since $\vec{x}$ (resp. $\vec{y}$) is a permutation of $\vec{a}$ (resp. $\vec{b}$), we have

$$\#\vec{a} + \#\vec{b} \leqslant k + 1 = 2(\lfloor k/2 \rfloor + 1).$$

(k even) Define $(\vec{x}_1, \vec{y}_1)$ and $(\vec{x}_2, \vec{y}_2) \in \mathbf{N}^{k/2} \times \mathbf{N}^{k/2}$ as follows

$$\vec{x}_1 = (a_1, a_3, \ldots, a_{k-1}), \qquad \vec{y}_1 = (b_2, b_4, \ldots, b_k),$$
$$\vec{x}_2 = (a_2, a_4, \ldots, a_k), \qquad \vec{y}_2 = (b_3, b_5, \ldots, b_{k-1}, b_1).$$

Then both $(\vec{x}_1, \vec{y}_1)$ and $(\vec{x}_2, \vec{y}_2)$ are Wishart-admissible. Therefore, using Lemma 3.3, we obtain

$$\#\vec{a} + \#\vec{b} \leqslant \#\vec{x}_1 + \#\vec{x}_2 + \#\vec{y}_1 + \#\vec{y}_2 \leqslant 2(k/2+1).$$

In both cases we proved $\#\vec{a} + \#\vec{b} \leqslant 2(\lfloor k/2 \rfloor + 1)$. □

We continue the proof of Proposition 5.3. If $(\vec{a}, \vec{b}, \vec{c})$ is admissible, Lemma 5.4 implies that $2k+2 = \ell(\vec{a}, \vec{b}, \vec{c}) \leqslant 4\lfloor k/2 \rfloor + 2$. Therefore, k must be even, and necessarily $\#\vec{c} = k/2$ and each index in $\vec{c}$ appears exactly twice.

To prove the last statement in Proposition 5.3, we are going to show that the following map Θ

$$\begin{array}{rcl} \mathscr{C}_k^{\mathrm{adm}} & \to & NC_2(k) \\ (\vec{a}, \vec{b}, \vec{c}) & \mapsto & \pi(\vec{c}) \end{array}$$

is bijective. First, the partition induced by $\vec{c}$ is indeed a chording of $[k]$ (this partition is non-crossing since $(\vec{a}, \vec{c})$ is Wishart-admissible). Because an element of a Wishart-admissible couple is determined (up to equivalence) by the other one, it follows that the map Θ is injective.

We now show that this map is onto. Given a partial chording $\pi \in NC_2(k)$, there is a Wishart-admissible couple $(\vec{a}, \vec{c}) \in \mathbf{N}^k \times \mathbf{N}^k$ such that $\pi(\vec{c}) = \pi$. It remains to check that $(\vec{a}, \vec{a}, \vec{c})$ is admissible.

- *The couple $(\vec{a}, \vec{a})$ is non-repeating.* Otherwise, one would have $a_i = a_{i+1}$ for some index $i \in [k]$. Since $\pi(\vec{c}) = K(\pi(\vec{a}))$, this would imply by Lemma 3.2 that $\{i\}$ is a block in $\pi(\vec{c})$, which is not possible if $\pi(\vec{c})$ is a chording.
- *The triple $(\vec{a}, \vec{a}, \vec{c})$ satisfies the matching condition.* Since we already know that $(\vec{a}, \vec{c})$ satisfies the Wishart matching condition, we have to check the following: whenever $(a_i, c_i) = (a_{j+1}, c_j)$, we have $a_{i+1} = a_j$. Suppose $(a_i, c_i) = (a_{j+1}, c_j)$. Since $(\vec{a}, \vec{a})$ is non repeating, we have $i \neq j$. This implies that $\{i, j\}$ must be a block in $\pi(\vec{c})$ and the result now follows from the second part of Lemma 3.2.

Therefore, the map Θ is bijective, and the cardinal of $\mathscr{C}_k^{\mathrm{adm}}$ equals the cardinal of $NC_2(k)$, which by Lemma 3.1 is the Catalan number $C_{k/2}$. □

6. Proof of Proposition 4.2

Start with a formula from the previous section

$$\frac{1}{d^2} \operatorname{Tr}(Z_d^k) = \frac{1}{d^2 p^k} \sum_{\vec{a} \in [d]^k, \vec{b} \in [d]^k, \vec{c} \in [p]^k} M(\vec{a}, \vec{b}) \Pi(\vec{a}, \vec{b}, \vec{c}).$$

The covariance of two random variables X, Y is defined as $\mathbf{Cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}\, X \cdot \mathbf{E}\, Y$. We have

$$\mathbf{Var}\, \frac{1}{d^2} \operatorname{Tr} Y_d^k = \frac{1}{d^4 p^{2k}} \sum_{\vec{a}, \vec{b}, \vec{c}, \vec{a}\,', \vec{b}\,', \vec{c}\,'} M(\vec{a}, \vec{b}) M(\vec{a}\,', \vec{b}\,') \, \mathbf{Cov}(\Pi(\vec{a}, \vec{b}, \vec{c}), \Pi(\vec{a}\,', \vec{b}\,', \vec{c}\,')), \tag{11}$$

where the summation is taken over 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 6.1. *Let $(\vec{a},\vec{b},\vec{c})$ and $(\vec{a}\,',\vec{b}\,',\vec{c}\,')$ be two triples in $\mathbf{N}^k\times\mathbf{N}^k\times\mathbf{N}^k$ such that*

$$\mathbf{Cov}(\Pi(\vec{a},\vec{b},\vec{c}),\Pi(\vec{a}\,',\vec{b}\,',\vec{c}\,'))\neq 0.$$

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_d shows that the following two conditions must hold:

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

$$(a_1,b_2,c_1),(a_2,b_1,c_1),\dots,(a_k,b_1,c_k),(a_1,b_k,c_k)\,;\,(a'_1,b'_2,c'_1),(a'_2,b'_1,c'_1),\dots,(a'_k,b'_1,c'_k),(a'_1,b'_k,c'_k). \tag{12}$$

- At least some couple appears both in the left half and in the right half of the list (12). 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 3.5. Therefore,

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

Similarly, one may apply Lemma 3.5 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 (11) 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 (11) shows that each class has contribution asymptotically zero. Since the number of equivalence classes depends only on k, this proves the lemma.

7. Convergence of extreme eigenvalues: proof of Theorem 2

Let G_d be a $d^2\times p$ matrix with independent $N(0,1)$ entries, $W_d=\frac{1}{p}G_dG_d^\dagger$, $Y_d=W_d^\Gamma$ and $Z_d=Y_d-\mathrm{diag}(Y_d)$. Assume that $p=\lfloor\alpha d^2\rfloor$.

Half of Theorem 2 can be deduced from Theorem 1. Indeed, for every $\varepsilon>0$, let I be the interval $[1+2/\sqrt{\alpha}-\varepsilon,1+2/\sqrt{\alpha}]$. Since $\mu_{SC(1,1/\alpha)}(I)>0$, Theorem 1 implies that, with probability tending to 1, $N_{Y_d}(I)>0$, which means $\lambda_{\max}(Y_d)\geqslant 1+2/\sqrt{\alpha}-\varepsilon$. A similar argument shows that $\lambda_{\min}(Y_d)\leqslant 1-2/\sqrt{\alpha}+\varepsilon$ with probability tending to 1.

To prove the other half of Theorem 2 (the hard part), we are going to give an upper bound on $\mathbf{E}\,\mathrm{Tr}(Z_d^k)$ which holds in any fixed dimension (as opposed to asymptotic estimates from the previous sections).

Proposition 7.1. *There is a polynomial Q such that, for any integer k,*

$$\mathbf{E}\,\mathrm{Tr}(Z_d^k)\leqslant(2/p)^k(d+Q(k))^{k+2}(\sqrt{p}+Q(k))^k.$$

Assume for the moment that Proposition 7.1 is true. We claim that it implies that for every $\varepsilon>0$,

$$\lim_{d\to\infty}\mathbf{P}\left(\|Y_d-\mathrm{Id}\|\geqslant 2/\sqrt{\alpha}+\varepsilon\right)=0,$$

from which Theorem 1 follows. Indeed, choose $k = k(d)$ an even integer such that $Q(k) = o(d)$ and $\log d = o(k)$. Then, when $d \to \infty$, Proposition 7.1 implies

$$\mathbf{E}\,\|Z_d\|^k \leqslant \mathbf{E}\,\mathrm{Tr}(Z_d^k) \leqslant \left(\frac{2d}{\sqrt{p}} + o(1)\right)^k = \left(\frac{2}{\sqrt{\alpha}} + o(1)\right)^k.$$

Therefore, it follows from Markov's inequality that for every $\varepsilon > 0$,

$$\mathbf{P}\left(\|Z_d\| \geqslant 2/\sqrt{\alpha} + \varepsilon\right) \leqslant \left(\frac{2}{\sqrt{\alpha}} + o(1)\right)^k \left(\frac{2}{\sqrt{\alpha}} + \varepsilon\right)^{-k} \longrightarrow 0.$$

On the other hand, by Lemma 3.7,

$$\mathbf{P}\left(\|\operatorname{diag}(Y_d) - \mathrm{Id}\| \geqslant \varepsilon\right) \leqslant d^2 \exp(-cp\varepsilon^2) \longrightarrow 0.$$

This completes the proof of Theorem 2 since

$$\mathbf{P}\left(\|Y_d - \mathrm{Id}\| \geqslant 2/\sqrt{\alpha} + \varepsilon\right) \leqslant \mathbf{P}\left(\|\operatorname{diag}(Y_d) - \mathrm{Id}\| \geqslant \varepsilon/2\right) + \mathbf{P}\left(\|Z_d\| \geqslant 2/\sqrt{\alpha} + \varepsilon/2\right).$$

Proof of Proposition 7.1. Recall the computation from Section 5

$$\mathrm{Tr}(Z_d^k) = \frac{1}{p^k} \sum_{\vec{a}\in[d]^k, \vec{b}\in[d]^k, \vec{c}\in[p]^k} M(\vec{a},\vec{b})\Pi(\vec{a},\vec{b},\vec{c}). \tag{13}$$

We first give an upper bound on $\mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c})$.

Lemma 7.2. *Let $(\vec{a},\vec{b},\vec{c}) \in \mathbf{N}^k \times \mathbf{N}^k \times \mathbf{N}^k$ satisfy the matching condition, and denote*

$$\Delta = 2k + 2 - \ell(\vec{a},\vec{b},\vec{c}).$$

Note that $\Delta \geqslant 0$. Then

1. *The number N of indices $i \in [2k]$ such that the ith term in the list (7) appears 4 times or more is bounded by 2Δ,*
2. *We have $\mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c}) \leqslant (C_0 k)^{\Delta}$, where C_0 is an absolute constant.*

Proof. At least one of the numbers $k+1-\ell_W(\vec{a},\vec{c})$ and $k+1-\ell_W(\vec{b},\vec{c})$ must be smaller than $\Delta/2$, since their sum equals Δ. Without loss of generality, we may assume that $k+1-\ell_W(\vec{a},\vec{c}) \leqslant \Delta/2$. Then, Lemma 3.3 implies that $n_+(\vec{a},\vec{c}) \leqslant 2\Delta$. Since $N \leqslant n_+(\vec{a},\vec{c})$, the first part of the lemma follows.

For the second part, we use independence to write $\mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c})$ as a product of quantities of the form $\mathbf{E}(G_{i,j}^k)^{q_1}(\overline{G_{i,j}^k})^{q_2} \leqslant \mathbf{E}\,|G_{i,j}^k|^{q_1+q_2}$. 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 both cases, for some constant C_0,

$$\mathbf{E}\,|G|^q \begin{cases} = 1 & \text{if } q = 2, \\ \leqslant (C_0\sqrt{q})^q & \text{if } q > 2. \end{cases} \tag{14}$$

Bounding each individual factor according to (14) and using $q \leqslant 2k$ leads to

$$\mathbf{E}\,\Pi(\vec{a},\vec{b},\vec{c}) \leqslant (C_0\sqrt{2k})^N$$

and the second part of the lemma follows. □

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

$$d(d-1)\cdots(d-\#\vec{a}+1)\cdot d(d-1)\cdots(d-\#\vec{b}+1)\cdot p(p-1)\cdots(p-\#\vec{c}+1) \leqslant d^{\#\vec{a}+\#\vec{b}}p^{\#\vec{c}}.$$

Therefore, it is convenient to rearrange the sum (13) according to the values of $\#\vec{a}+\#\vec{b}$ and $\#\vec{c}$. We denote by m_{ℓ_1,ℓ_2} the number of equivalence classes of triples $(\vec{a}, \vec{b}, \vec{c}) \in \mathbf{N}^k \times \mathbf{N}^k \times \mathbf{N}^k$ which satisfy the matching condition, with $(\vec{a}, \vec{b})$ non-repeating, $\#\vec{a}+\#\vec{b} = \ell_1$ and $\#\vec{c} = \ell_2$. It follows from the analysis above that

$$\mathbf{E}\,\mathrm{Tr}(Y^k) \leqslant \frac{1}{p^k}\sum_{\ell_1,\ell_2} d^{\ell_1}p^{\ell_2}m_{\ell_1,\ell_2}(C_0k)^{2k+2-\ell_1-2\ell_2}. \tag{15}$$

By Lemma 5.4, $m_{\ell_1,\ell_2} = 0$ if either $\ell_1 > k+2$ or $\ell_2 > k/2$. It remains to give a bound on the number m_{ℓ_1,ℓ_2}. This is the content of the following proposition (we postpone the proof to the end of the section).

Proposition 7.3. *There is a polynomial P such that the following holds. Denote by N_Δ the number of equivalence classes of triples $(\vec{a}, \vec{b}, \vec{c}) \in \mathbf{N}^k \times \mathbf{N}^k \times \mathbf{N}^k$ which satisfy the matching condition, with $(\vec{a}, \vec{b})$ non-repeating and $\ell(\vec{a}, \vec{b}, \vec{c}) = 2k+2-\Delta$. We have the bound*

$$N_\Delta \leqslant 2^k P(k)^\Delta. \tag{16}$$

Remark. *The bound given in* (16) *is quite sharp. Indeed, for $\Delta = 0$, it gives $N_0 \leqslant 2^k$. But N_0 is exactly the number of equivalence classes of admissible triples considered in Section 5, where this number was shown to equal the Catalan number $C_{k/2}$, only slightly smaller that 2^k.*

We continue the proof of Proposition 7.1. We have

$$m_{\ell_1,\ell_2} \leqslant N_{2k+2-\ell_1-2\ell_2} \leqslant 2^k P(k)^{2k+2-\ell_1-2\ell_2}.$$

Plugging this into (15) and denoting Q the polynomial $Q(k) = C_0kP(k)$,

$$\begin{aligned}
\mathbf{E}\,\mathrm{Tr}(Z_d^k) &\leqslant \frac{2^k}{p^k}\sum_{\ell_1=2}^{k+2}\sum_{\ell_2=1}^{k/2} d^{\ell_1}p^{\ell_2}Q(k)^{2k+2-\ell_1-2\ell_2} \\
&= \frac{2^k}{p^k}\left(\sum_{\ell_1=2}^{k+2} d^{\ell_1}Q(k)^{k+2-\ell_1}\right)\left(\sum_{\ell_2=1}^{k/2}(\sqrt{p})^{2\ell_2}Q(k)^{k-2\ell_2}\right) \\
&\leqslant (2/p)^k(d+Q(k))^{k+2}(\sqrt{p}+Q(k))^k.
\end{aligned}$$

This completes the proof of Proposition 7.1. □

Proof of Proposition 7.3. For $(\vec{a}, \vec{b}, \vec{c}) \in \mathbf{N}^k \times \mathbf{N}^k \times \mathbf{N}^k$, let $I = I(\vec{a}, \vec{b}, \vec{c}) \subset [k-1]$ be the subset of indices i such that the following condition holds

(1) $a_{i+1} \notin \{a_j \ : \ j < i+1\}$ — one says that a_{i+1} is an innovation,
(2) $b_{i+1} \notin \{b_j \ : \ j < i+1\}$ — one says that b_{j+1} is an innovation,
(3) $c_i \notin \{c_j \ : \ j < i\}$ — one says that c_j is an innovation.

The next lemma shows that the set $I(\vec{a}, \vec{b}, \vec{c})$ is large when Δ is small. We postpone the proof.

Lemma 7.4. *If $(\vec{a},\vec{b},\vec{c}) \in \mathbf{N}^k \times \mathbf{N}^k \times \mathbf{N}^k$ satisfies the matching condition with $(\vec{a},\vec{b})$ non-repeating, then*

$$\operatorname{card} I(\vec{a},\vec{b},\vec{c}) \geqslant k/2 - \Delta.$$

where $\Delta = (2k+2) - \ell(\vec{a},\vec{b},\vec{c})$.

Let A, C be subsets of $[k]$. A couple $(\vec{a},\vec{c})$ satisfying the Wishart matching condition is said to be compatible with (A,C) if

(1) for every $i \in A$, the index a_i is an innovation, i.e. $a_i \notin \{a_j \ : \ j < i\}$,
(2) for every $i \in C$, the index c_i is an innovation, i.e. $c_i \notin \{c_j \ : \ j < i\}$.

Note that if a Wishart-admissible couple $(\vec{a},\vec{c})$ is compatible with (A,C), then by arguing as in the proof of Lemma 3.3, we have

$$\operatorname{card} A + \operatorname{card} C \leqslant d_W(\vec{a},\vec{c}) + 1 \leqslant k+1.$$

Let us state one more lemma, postponing the proof.

Lemma 7.5. *Let A, C be subsets of $[k]$, and $\delta = k + 1 - \operatorname{card} A - \operatorname{card} C$. The number of equivalence classes of couples $(\vec{a},\vec{c}) \in \mathbf{N}^k \times \mathbf{N}^k$ which satisfy the Wishart matching condition and are compatible with (A,C) is bounded by $(2k)^{9\delta}$.*

The number N_Δ is the number (up to equivalence) of triples $(\vec{a},\vec{b},\vec{c})$ which satisfies the matching condition, with $(\vec{a},\vec{b})$ non repeating, and $\ell(\vec{a},\vec{b},\vec{c}) = (2k+2) - \Delta$. To bound N_Δ, we first choose a set $I \subset [k-1]$ of cardinal larger than $k/2 - \Delta$. The number of possibilities for I is bounded by 2^k. Now, given I, let I^+ be the subset of $[k]$ defined as

$$j \in I^+ \iff j = 1 \text{ or } j - 1 \in I.$$

If $(\vec{a},\vec{b},\vec{c})$ satisfies the matching condition with $I(\vec{a},\vec{b},\vec{c}) = I$, then it is easily checked that both couples $(\vec{a},\vec{c})$ and $(\vec{b},\vec{c})$ are compatible with (I^+, I). We have $\operatorname{card}(I^+) + \operatorname{card}(I) = 2\operatorname{card}(I) + 1 \geqslant k + 1 - 2\Delta$. By Lemma 7.5, the number of admissible couples compatible with (I^+, I) is bounded by $(2k)^{18\Delta}$. Therefore the number of possible triples $(\vec{a},\vec{b},\vec{c})$ is bounded by $(2k)^{36}$. This yields the bound

$$N_\Delta \leqslant 2^k (2k)^{36\Delta}.$$

This proves Proposition 7.3 with $P(k) = (2k)^{36}$. □

Proof of Lemma 7.4. For each index $i \in [k]$, one of the following possibility occurs

$P_1(i)$: The indices a_{i+1}, b_{i+1} and c_i are innovations. Necessarily the triples (a_i, b_{i+1}, c_i) and (a_{i+1}, b_i, c_i) are innovations[3].
$P_2(i)$: The triples (a_i, b_{i+1}, c_i) and (a_{i+1}, b_i, c_i) are innovations, but at least one of a_{i+1}, b_{i+1} and c_i is not an innovation.
$P_3(i)$: Only one of the triples (a_i, b_{i+1}, c_i) and (a_{i+1}, b_i, c_i) is an innovation.
$P_4(i)$: Neither (a_i, b_{i+1}, c_i) nor (a_{i+1}, b_i, c_i) is an innovation.

For $j \in \{1,2,3,4\}$, let n_j be the number of indices $i \in [k]$ such that $P_j(i)$ holds in the above alternative. With this notation, $n_1 = \operatorname{card} I(\vec{a},\vec{b},\vec{c})$. The numbers n_1, n_2, n_3, n_4 satisfy the following relations

$$n_1 + n_2 + n_3 + n_4 = k, \tag{17}$$

[3]We say that a triple at jth position from the list (7) is an innovation if it does not coincide with a triple at ith position for $i < j$.

$$n_3 + 2n_4 \geqslant k, \tag{18}$$

$$4n_1 + 3n_2 + n_3 \geqslant 2k - \Delta. \tag{19}$$

- Equation (17) is obvious since possibilities $P_1(i), \dots, P_4(i)$ are mutually exclusive.
- There must be at least k elements in the list (7) which are not innovations, since every element must appear at least twice. But the number of non-innovations in the list (7) is equal to $n_3 + 2n_4$, hence the equation (18).
- For each i, let Z_i be the number

$$Z_i = \mathbf{1}_{\{a_{i+1} \text{ is an innovation}\}} + \mathbf{1}_{\{b_{i+1} \text{ is an innovation}\}} + 2 \cdot \mathbf{1}_{\{c_i \text{ is an innovation}\}}.$$

  The value of Z_i depends on which of $P_1(i), P_2(i), P_3(i), P_4(i)$ occurs. If $P_1(i)$ occurs, then $Z_i = 4$. If $P_4(i)$ occurs, then $Z_i = 0$. If $P_2(i)$ occurs, then $Z_i \leqslant 3$. If $P_3(i)$ occurs, then $Z_i \leqslant 1$. This last point deserves some explanation.
  - If (a_i, b_{i+1}, c_i) is not an innovation, then certainly b_{i+1} and c_i cannot be innovations.
  - If instead (a_{i+1}, b_i, c_i) is not an innovation, then a_{i+1} cannot be an innovation. We claim that c_i is also not an innovation. Indeed, if c_i was an innovation, then necessarily (a_{i+1}, b_i, c_i) would be equal to (a_i, b_{i+1}, c_i) which would contradict the non-repeating property.

  This shows that $\sum Z_i \leqslant 4n_1 + 3n_2 + n_3$. On the other hand, we have

$$\sum_{i=1}^{k} Z_i = \#\vec{a} - 1 + \#\vec{b} - 1 + 2\#\vec{c} = 2k - \Delta.$$

  Therefore, the above discussion implies equation (19).

Adding (19) and twice (18), we obtain

$$4n_1 + 3n_2 + 3n_3 + 4n_4 \geqslant 4k - \Delta.$$

Together with (17), this implies that $n_2 + n_3 \leqslant \Delta$. Since $n_3 \geqslant 0$, this in turn implies $3n_2 + n_3 \leqslant 3\Delta$. Combined with (19), we obtain $4n_1 \geqslant 2k - 4\Delta$, hence $n_1 \geqslant k/2 - \Delta$ as claimed. □

Proof of Lemma 7.5. Given a couple $(\vec{a}, \vec{c}) \in \mathbf{N}^k \times \mathbf{N}^k$ satisfying the Wishart matching condition, there is a partition of $[2k]$ as

$$[2k] = T_1 \cup T_2 \cup T_3 \cup T_4 \tag{20}$$

where T_i denotes the set of indices j such that the jth element in the list (3) is of type i (the four possible types have been defined in Section 2). If the couple $(\vec{a}, \vec{c})$ is compatible with (A, C), then necessarily $T_1^* \subset T_1$ and $T_2^* \subset T_2$, where

$$T_1^* = \{2(i-1) \ : \ i \in A, i \neq 1\},$$
$$T_2^* = \{2i - 1 \ : \ i \in C\}.$$

We claim that the number of partitions (20) satisfying these constraints is bounded by $(2k)^{3\delta}$. Indeed, we first have to enlarge T_1^* into T_1 and T_2^* into T_2. Since $\mathrm{card}(T_1^* \cup T_2^*) = k - \delta$ and $\mathrm{card}(T_1 \cup T_2) \leqslant k$, the number of possible ways to perform these enlargements in at most $(2k)^{\delta}$.

Since $\mathrm{card}(T_3) = \mathrm{card}(T_1) + \mathrm{card}(T_2)$, we have $\mathrm{card}(T_4) \leqslant 2\delta$. Therefore the number of possible choices for T_4 is bounded by $(2k)^{2\delta}$. Once T_1, T_2 and T_4 are chosen, the set T_3 consists of the remaining indices. Hence the claim on the number of possible partitions.

Now, by Lemma 3.6, the number of equivalence classes of couples satisfying the Wishart matching condition with a given partition (20) is bounded by

$$(2k)^{3\operatorname{card} T_4} \leqslant (2k)^{6\delta}.$$

Finally, the total number of equivalence classes satisfying the Wishart matching condition and compatible with (A, C) is bounded by $(2k)^{9\delta}$. □

8. Relevance to Quantum Information Theory

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

8.1. **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 (1)). The partial transposition of a *separable state* ρ is always positive [22]; however there exist non-separable (=entangled) PPT states. For many purposes, checking positivity of the partial transpose is the most efficient tool to detect entanglement. We refer to the survey [14] for more information about PPT states and entanglement.

8.2. **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 $\mathcal{D}(\mathbf{C}^n)$, the set of mixed states on $\mathbf{C}^n$. A random state with distribution $\mu_{n,p}$ is called an *induced state*; 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 [7], Section 14.5 and references therein).

Induced states are closely related to Wishart distributions. Indeed, if W is a (n, p)-Wishart random matrix, then $\frac{1}{\mathrm{Tr}\, W} W$ is a random state with distribution $\mu_{n,p}$. Moreover, the random variables $\mathrm{Tr}\, W$ and $\frac{1}{\mathrm{Tr}\, W} W$ are independent (this fact explicitly appears in [19]). Therefore, results about Wishart matrices can be easily translated in the language of induced states. 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 [24].

Proposition 8.1. *The measure $\mu_{n,n}$ is equal to the normalized Lebesgue measure restricted to the set $\mathcal{D}(\mathbf{C}^n)$.*

Proposition 8.1 follows from a more general fact [24]: whenever $p \geqslant n$, the density of the measure $\mu_{n,p}$ with respect to the Lebesgue measure on $\mathcal{D}(\mathbf{C}^n)$ is proportional to $\det(\rho)^{p-n}$.

8.3. **Partial transposition of random induced states.** Our main results admit an immediate translation in the language of random induced states. Here is a version of Theorem 1 for induced states.

Theorem 3. *Fix $\alpha > 0$. For each d, let ρ_d be a random state on $\mathbf{C}^d \otimes \mathbf{C}^d$ chosen according to the measure $\mu_{d^2,\lfloor \alpha d^2 \rfloor}$. Then for every interval $I = [a,b] \subset \mathbf{R}$ and $\varepsilon > 0$,*

$$\lim_{d\to\infty} \mathbf{P}\left(\left| N_{d^2 \rho_d^\Gamma}(I) - \mu_{SC(1,1/\alpha)}(I) \right| > \varepsilon \right) = 0.$$

Recall that $N_{d^2\rho_d^\Gamma}(I)$ is the proportion of eigenvalues of the matrix ρ_d^Γ that belong to the interval $[a/d^2, b/d^2]$.

Proof. If W is a (d^2,p)-Wishart matrix, then $\frac{W}{\operatorname{Tr} W}$ has distribution as $\mu_{d^2,p}$. Therefore,

$$N_{d^2\rho_d^\Gamma}([a,b]) = N_{\frac{d^2}{\operatorname{Tr} W} W^\Gamma}([a,b]) = N_{W^\Gamma}\left(\left[\frac{\operatorname{Tr} W}{d^2} a, \frac{\operatorname{Tr} W}{d^2} b \right] \right).$$

The distribution of $\frac{\operatorname{Tr} W}{d^2}$ is proportional to a χ^2 distribution. Using Fact 3.8 to quantify its concentration, we obtain that for any $\eta > 0$,

$$\mathbf{P}\left(\left| \frac{\operatorname{Tr} W}{d^2} - 1 \right| > \eta \right) \leqslant C \exp(-c d^2 p \eta^2). \tag{21}$$

When $\left| \frac{\operatorname{Tr} W}{d^2} - 1 \right| \leqslant \eta$, we may use the inclusions

$$[(1+\eta)a, (1-\eta)b] \subset \left[\frac{\operatorname{Tr} W}{d^2} a, \frac{\operatorname{Tr} W}{d^2} b \right] \subset [(1-\eta)a, (1+\eta)b]$$

to show that Theorem 1 implies Theorem 3. □

If 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 following theorem shows that a threshold occurs when $p = 4d^2$.

Theorem 4. *For every $\varepsilon > 0$, there exist positive constants $c(\varepsilon), C(\varepsilon)$ such that the following holds. If ρ is a random state on $\mathbf{C}^d \otimes \mathbf{C}^d$ chosen according to the measure $\mu_{d^2,p}$, then*

1. *If $p \leqslant (4-\varepsilon)d^2$, then*
$$\mathbf{P}(\rho \text{ is PPT}) \leqslant C(\varepsilon) \exp(-c(\varepsilon) p).$$
2. *If $p \geqslant (4+\varepsilon)d^2$, then*
$$\mathbf{P}(\rho \text{ is PPT}) \geqslant 1 - C(\varepsilon) \exp(-c(\varepsilon) p).$$

Proof. We only show the proof of (1), the proof of (2) being similar. We are going to use a concentration argument from [3], where the same question is studied for separability instead of PPT. We start by a lemma that compares the probability that a random state is PPT, for different dimensions.

Lemma 8.2. *Let d_1, d_2, d_1', d_2' and p be integers, with $d_1' \leqslant d_1$ and $d_2' \leqslant d_2$. Let ρ be a random state on $\mathbf{C}^{d_1} \otimes \mathbf{C}^{d_2}$ with distribution $\mu_{d_1 d_2, p}$, and let ρ' be a random state on $\mathbf{C}^{d_1'} \otimes \mathbf{C}^{d_2'}$ with distribution $\mu_{d_1' d_2', p}$. Then*

$$\mathbf{P}(\rho \text{ is PPT}) \leqslant \mathbf{P}(\rho' \text{ is PPT}).$$

Proof. It is enough to prove the lemma in the special case $d_2 = d_2'$ (since both factors play the same role, the full version follows by applying twice this special case).

We construct a coupling between both distributions as follows. Identify $\mathbf{C}^{d_1'}$ as a subspace of $\mathbf{C}^{d_1}$, and let $Q : \mathbf{C}^{d_1} \to \mathbf{C}^{d_1'}$ be the orthogonal projection. Then, $\mathbf{C}^{d_1'} \otimes \mathbf{C}^{d_2} \subset \mathbf{C}^{d_1} \otimes \mathbf{C}^{d_2}$ is the range of the projection $P = Q \otimes \mathrm{Id}$. Let W be a $(d_1 d_2, p)$-Wishart matrix, seen as an operator on $\mathbf{C}^{d_1} \otimes \mathbf{C}^{d_2}$. The random operator PWP, when seen as an operator on $\mathbf{C}^{d_1'} \otimes \mathbf{C}^{d_2}$, has the distribution of a $(d_1' d_2, p)$-Wishart matrix. Therefore, the states

$$\rho = \frac{W}{\operatorname{Tr} W},$$

$$\rho' = \frac{P\rho P}{\operatorname{Tr} P\rho P} = \frac{PWP}{\operatorname{Tr} PWP},$$

have respective distributions $\mu_{d_1 d_2, p}$ and $\mu_{d_1' d_2, p}$. To prove the lemma it remains to check that

$$\rho \text{ is PPT} \Longrightarrow \rho' \text{ is PPT}.$$

This implication holds because $(P\rho P)^\Gamma = P\rho^\Gamma P$. □

Fix $\varepsilon > 0$. As a consequence of Lemma 8.2, it is enough to prove Theorem 4, for every given p, when d is minimal such that $p \leqslant (4-\varepsilon)d^2$ (from now one, we assume that d and s are related by this condition).

Denote by $\|\cdot\|_{\mathrm{PPT}}$ the gauge associated to the convex body of all PPT states. This gauge is defined as follows, for any state ρ on $\mathbf{C}^d \otimes \mathbf{C}^d$

$$\begin{aligned} \|\rho\|_{\mathrm{PPT}} &= \inf\left\{ t \geqslant 0 \ : \ \frac{\mathrm{Id}}{d^2} + \frac{1}{t}\left(\rho - \frac{\mathrm{Id}}{d^2}\right) \text{ is PPT} \right\} \\ &= 1 - d^2 \lambda_{\min}(\rho^\Gamma). \end{aligned}$$

Note in particular that ρ is PPT if and only if $\|\rho\|_{\mathrm{PPT}} \leqslant 1$. Let $\rho_{d^2,p}$ be a random state with distribution $\mu_{d^2,p}$, and denote by $M_{d^2,p}$ the median of the random variable $\|\rho_{d^2,p}\|_{\mathrm{PPT}}$. By applying Proposition 4.2 from [3], we obtain the following inequality: there are absolute constants c, C such that for any $\eta > 0$,

$$\mathbf{P}\left(\left| \|\rho\|_{\mathrm{PPT}} - M_{d^2,p} \right| \geqslant \eta \right) \leqslant C\exp(-cp) + C\exp(-cp\eta^2). \tag{22}$$

Let $W_{d^2,p}$ be a (d^2, p)-Wishart matrix. It follows from Theorem 2 that $\lambda_{\min}(W_{d^2,p}^\Gamma)$ converges in probability towards $1 - 2/\sqrt{4-\varepsilon}$ when d, p tend to infinity. By (21), $\operatorname{Tr} W_{d^2,p}/d^2$ converges in probability to 1. Since $W_{d^2,p}/\operatorname{Tr} W_{d^2,p}$ has distribution $\mu_{d^2,p}$, it follows that $\|\rho_{d^2,p}\|_{\mathrm{PPT}}$ converges to $\frac{2}{\sqrt{4-\varepsilon}}$. In particular,

$$\lim_{p,d\to\infty} M_{d^2,p} = \frac{2}{\sqrt{4-\varepsilon}} > 1.$$

We now choose η such that $2/\sqrt{4-\varepsilon} > 1 + \eta$. For d, p large enough, we have $M_{d^2,p} > 1 + \eta$, and we can apply (22) to obtain

$$\mathbf{P}(\rho \text{ is PPT}) = \mathbf{P}(\|\rho\|_{\mathrm{PPT}} \leqslant 1) \leqslant C\exp(-cp) + C\exp(-cp\eta^2).$$

This concludes the proof of Theorem 4 (small dimensions can be taken into account by adjusting the constants). □

9. Miscellaneous remarks

9.1. **Partial transposition of a random pure state.** Another natural question from the point of view of Quantum Information Theory is to study the partial transposition of random *pure* states (as opposed to random *mixed* states considered here). In that direction, one may prove the following result

Proposition 9.1. *For every d, let ρ_d be a random pure state on $\mathbf{C}^d \otimes \mathbf{C}^d$, with uniform distribution. Then, when d tends to infinity, the empirical eigenvalue distribution of $d\rho_d^\Gamma$ approaches a deterministic distribution which can be described as the distribution of the product of two independent $SC(0,1)$ random variables.*

Remark. *The notion of convergence used is the same as in Theorem 3. The limiting distribution appearing in Proposition 9.1 has vanishing odd moments and even moments equal to the square of Catalan numbers. Such a distribution has been studied recently in* [9]*, where a closed formula for the density (involving special functions) is derived.*

Proof of Proposition 9.1 (sketch). If $\rho = |\psi\rangle\langle\psi|$ is a pure state on $\mathbf{C}^d \otimes \mathbf{C}^d$, the eigenvalues of ρ^Γ can be described from the Schmidt coefficients of ψ (Schmidt coefficients for tensors correspond to singular values for matrices, and are therefore governed by the Marčenko–Pastur distribution). Indeed, given a Schmidt decomposition

$$\psi = \sum_{i=1}^{d} \sqrt{\lambda_i} e_i \otimes f_i,$$

for some orthonormal bases $(e_i), (f_i)$, one checks that

$$|\psi\rangle\langle\psi|^\Gamma = \sum_{i,j=1}^{d} \sqrt{\lambda_i \lambda_j} |e_i \otimes f_j\rangle\langle e_j \otimes f_i|.$$

It follows that the eigenvalues of $|\psi\rangle\langle\psi|^\Gamma$ are

$$\begin{cases} \lambda_i & \text{for every } 1 \leqslant i \leqslant d, \\ \pm\sqrt{\lambda_i \lambda_j} & \text{for every } 1 \leqslant i < j \leqslant d. \end{cases}$$

Eigenvalues of the first category do not contribute to the limit distribution, and the result follows with little effort. □

9.2. **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 $d_1 d_2$.

9.3. **Connexions to free probability.** The same model of partially transposed Wishart matrices has been considered recently by Banica and Nechita [6] in a different asymptotic regime (when d_1 is fixed and d_2 goes to infinity). For that regime the picture is different: they obtain that the limit spectral distribution can be described as the difference of two freely independent random variables with Marčenko–Pastur distributions. The shifted semicircle distribution appears then as a limit case. We refer to [6] for more information.

9.4. **Uniform mixtures of random pure states.** There is another popular model of random states which is very similar to the model of random induced states considered in Section 8, for which our results are also valid. Let $(\psi_i)_{1\leqslant i\leqslant p}$ be unit vectors in $\mathbf{C}^n$, chosen independently according to the uniform probability measure on the the unit sphere. Then we consider the random state

$$\rho = \frac{1}{p}\sum_{i=1}^{p} |\psi_i\rangle\langle\psi_i|.$$

Denote by $\nu_{n,p}$ the distribution of ρ. This model of random states has been considered for example in [23]. When n, p are large, the probability measures $\mu_{n,p}$ and $\nu_{n,p}$ behave similarly. It can be shown that Theorems 3 and 4 remain valid when the probability measures $\mu_{n,p}$ are substituted by the probability measures $\nu_{n,p}$.

9.5. **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 8.1) by the induced measure over an ancilla of equal dimension. Let W_d be a (d^2, d^2)-Wishart random matrix. It was shown in [2] (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 $C > 0$

$$\mathbf{P}(W_d^\Gamma \geqslant 0) \geqslant \exp(-Cd^4). \tag{23}$$

By Theorem 2, 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 (23) to be sharp.

Conjecture. *There is an absolute constant $c > 0$ such that, whenever W_d is a (d^2, d^2)-Wishart matrix,*

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

This would quantify precisely how (un)common are PPT states 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 and S. Szarek, Tensor product of convex sets and the volume of separable states on N qudits, *Phys. Rev. A* **73** (2006)

[3] G. Aubrun, S. Szarek and D.Ye, Entanglement thresholds for random induced states, arxiv:1106.2264

[4] G. Aubrun, S. Szarek and D.Ye, Phase transitions for random states and a semi-circle law for the partial transpose, arxiv:1011.0275

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

[6] T. Banica and I. Nechita, Asymptotic eigenvalue distributions of block-transposed Wishart matrices, arxiv:1105.2556

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

[8] R. Bhatia, Matrix analysis. *Graduate Texts in Mathematics* **169**. Springer-Verlag, 1997.

[9] A. Bostan, P. Flajolet and K. Penson, Combinatorial sequences and moment representations, unpublished.

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

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

[12] P. Hayden, D. Leung and A. Winter, Aspects of generic entanglement, *Comm. Math. Phys.* **265** (2006) 95-117.

[13] 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).
[14] R. Horodecki, P. Horodecki, M. Horodecki and K. Horodecki, Quantum entanglement, *Rev. Mod. Phys.* **81**, No. 2, pp. 865-942 (2009).
[15] D. Jonsson, Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* **12** (1982), no. 1, 1–38.
[16] V. Kendon, K. Życzkowski and W. Munro, Bounds on entanglement in qudit subsystems, *Phys. Rev. A* **66**, 062310 (2002)
[17] M. Ledoux, The concentration of measure phenomenon. *Mathematical Surveys and Monographs*, **89**. American Mathematical Society, Providence, RI, 2001
[18] V. Marčenko and L. Pastur, Distribution of eigenvalues in certain sets of random matrices, *Mat. Sb. (N.S.)*, **72** (114), 507–536 1967.
[19] I. Nechita, Asymptotics of random density matrices, *Ann. Henri Poincaré* **8** (2007), no. 8, 1521–1538.
[20] A. Nica and R. Speicher, Lectures on the Combinatorics of Free Probability, London Mathematical Society Lecture Note Series **335** (2006).
[21] 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.
[22] A. Peres, Separability criterion for density matrices, *Phys. Rev. Lett.* **77**, 1413–1415 (1996).
[23] M. Žnidarič, T. Prosen, G. Benenti and G. Casati, Detecting entanglement of random states with an entanglement witness, *J. Phys. A* **40** (2007) 13787-13798
[24] K. Życzkowski and H.-J. Sommers, Induced measures in the space of mixed quantum states, *J. Phys. A* **34** (2001), 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