Marchenko-Pastur Theorem and Bercovici-Pata bijections for heavy-tailed or localized vectors

Florent Benaych-Georges, Thierry Cabanal-Duvillard

MARCHENKO-PASTUR THEOREM AND BERCOVICI-PATA BIJECTIONS FOR HEAVY-TAILED OR LOCALIZED VECTORS

FLORENT BENAYCH-GEORGES AND THIERRY CABANAL-DUVILLARD

ABSTRACT. The celebrated Marchenko-Pastur theorem gives the asymptotic spectral distribution of sums of random, independent, rank-one projections. Its main hypothesis is that these projections are more or less uniformly distributed on the first grassmannian, which implies for example that the corresponding vectors are *delocalized*, *i.e.* are essentially supported by the whole canonical basis. In this paper, we propose a way to drop this delocalization assumption and we generalize this theorem to a quite general framework, including random projections whose corresponding vectors are *localized*, *i.e.* with some components much larger than the other ones. The first of our two main examples is given by heavy tailed random vectors (as in the model introduced by Ben Arous and Guionnet in [5] or as in the model introduced by Zakharevich in [32] where the moments grow very fast as the dimension grows). Our second main example, related to the continuum between the classical and free convolutions introduced in [11], is given by vectors which are distributed as the Brownian motion on the unit sphere, with localized initial law. Our framework is in fact general enough to get new correspondences between classical infinitely divisible laws and some limit spectral distributions of random matrices, generalizing the so-called *Bercovici-Pata bijection*.

1. INTRODUCTION

In 1967, Marchenko and Pastur introduced a successful matrix model inspired by the elementary fact that each Hermitian matrix is the sum of orthogonal rank one homotheties. Substituting orthogonality with independence, they considered in their seminal paper [23] the $N \times N$ random matrix defined by

$$(1) \qquad \frac{1}{N} \sum_{i=1}^{p} X^i \cdot U_N^i (U_N^i)^*,$$

where $(X^i)_{i \geq 1}$ is an i.i.d. sequence of real valued random variables and $(U_N^i)_{i \geq 1}$ is an i.i.d. sequence of N-dimensional column vectors, whose conjugate transpose are denoted $(U_N^i)^*$,

independent of $(X_i)_{i\geq 1}$. As a main result, they proved that the empirical spectral measure of this matrix converges to a limit with an explicit characterization under the following assumptions:

   (a) N, p tend to infinity in such a way that $p/N \longrightarrow \lambda > 0$;
   (b) the first four joint moments of the entries of U_N^1 are not too far, roughly speaking, from the ones of the entries of a standard Gaussian vector.

In the special case where all X^i's are equal to one, the matrix (1) reduces to a so-called *empirical covariance matrix*, and its limit spectral distribution is none other than the well-known *Marchenko-Pastur distribution with parameter λ*.

It has to be noticed that even in the general case, the limit spectral distribution does not depend on the particular choice of the U_N^i's, granted they satisfy the above hypothesis. For example, one can choose U_N^1 to have uniform distribution on the sphere of $\mathbb{R}^N$ or $\mathbb{C}^N$ with radius $\sqrt{N}$, or to be a standard gaussian: such a vector is said to be *delocalized*, which means that with large probability, $\|U_N^1\|_\infty/\|U_N^1\|_2$ is small; more specifically:

$$\frac{\|U_N^1\|_\infty}{\|U_N^1\|_2} \approx \left(\frac{\log(N)}{N}\right)^{1/2}.$$

After the initial paper of Marchenko and Pastur, a long list of further-reaching results about limit spectral distribution of empirical covariance matrices have been obtained, by Yin and Krishnaiah [31], Götze and Tikhomirov [20, 21], Aubrun [3], Pajor and Pastur [25], Adamczak [1]. All of them are devoted to the empirical covariance matrix of more or less delocalized vectors, with limit spectral distribution being the Marchenko-Pastur distribution (except in the case treated in [31], but there the vector are still very delocalized).

In this paper, our goal is to drop the delocalization assumption and to be able to deal with *localized* U_N^i's, i.e. with some entries much larger than the other ones. For example, the applications of our main theorem include the case where the entries of U_N^i have heavy tails, but also in some other examples of localized vectors, such as the one where the law of U_N^i results from a Brownian motion with localized initial condition.

This approach is based on our preceding works [6, 16] on the Bercovici-Pata bijection (see also [7, 26, 10]). This bijection, that we denote by Λ, is a correspondence between the probability measures on the real line that are infinitely divisible with respect to the classical convolution $*$ and the ones which are infinitely divisible with respect to the free convolution $\boxplus$. In [6, 16], we constructed a set of matrix ensembles which produces a quite natural interpretation of Λ. This construction is easy to describe for compound Poisson laws and makes the connection with Marchenko-Pastur's model quite clear. Let $(X^i)_{i\geq 1}$ be still an i.i.d. sequence of real valued random variables, $(U_N^i)_{i\geq 1}$ i.i.d. column vectors uniformly distributed on the sphere of $\mathbb{C}^N$ with radius $\sqrt{N}$, and $(P(\lambda), \lambda > 0)$ a standard

Poisson process, $(X^i)_{i \geq 1}$, $(U_N^i)_{i \geq 1}$ and $(P(\lambda), \lambda > 0)$ being independent. For each $N \geq 1$, we defined the random matrix

$$(2) \qquad \frac{1}{N} \sum_{i=1}^{P(N\lambda)} X^i \cdot U_N^i (U_N^i)^*$$

and we proved that its empirical spectral law converges to $\Lambda(\mu)$ when N goes to infinity, where μ is the compound Poisson law of

$$\sum_{i=1}^{P(\lambda)} X^i,$$

The link with Marchenko-Pastur's model of (1) is now obvious and it is easy to verify that the empirical spectral laws of (1) and (2) have same limit if $p \sim N\lambda$. Hence, our previous works [6, 16] could be viewed as another insight on Marchenko-Pastur's results, partly more restricted, since we only considered uniformly distributed random vectors U_N^i, partly more general, since our construction extended to all infinitely divisible laws. In fact, the main advantage of our matricial model, which has also been studied in [26], over Marchenko-Pastur's one is to be infinitely divisible. It allows us to derive simpler proofs, using appropriate tools as cumulant computations or semi-groups.

In the present paper, we extend our construction to a larger class of U_N^i's, while continuing to benefit of the infinitely divisible framework. Roughly speaking, we are able to prove the convergence of the empirical spectral law if we suppose only that the entries of U_N^i are exchangeable and have a moment of order k growing as $N \to \infty$ at most in $N^{\frac{k}{2}-1}$, for any fixed k (*cf.* Theorem 2.6).

Then, approximations allow to extend the result to U_N^i with *heavy tailed* entries (see Theorem 3.1). When the X^i's are constant, we recover a result by obtained by Belinschi, Dembo and Guionnet in [4]. Our result is more general from a certain point of view (we allow some random X^i's), but less explicit since we characterize the limit spectral distribution as the weak limit of a sequence of probability distributions with calculable moments and not with a functional equation, as in [4]. We also state (Theorem 3.2) a "covariance matrices version" of Zakharevich's generalization of Wigner's theorem of [32], which is a direct consequence of our key result Theorem 2.6.

We also devote a particular interest to the special case where the column vectors U_N^i's are copies of

$$(3) \qquad U_N(t) = \sqrt{N} e^{-\frac{t}{2}} X_N + \sqrt{1 - e^{-t}} G_N,$$

where X_N is uniformly distributed on the canonical basis $(1, 0, \ldots, 0)^T, \ldots, (0, \ldots, 0, 1)^T$, independent of $G_N := (z_1, \ldots, z_N)^T$, with $z_1, z_2, z_3, \ldots$ independent standard Gaussian variables. Let us emphasized that $U_N(t)$ is a quite typically *localized* vector since it has

exactly one entry which is much bigger than the $N - 1$ others. Precisely, we have:

$$\frac{\|U_N(t)\|_\infty}{\|U_N(t)\|_2} \approx e^{-t/2}.$$

This is the reason why the $N \to \infty$ limits differ from the ones of the classical matrix models. For example, the classical Marchenko-Pastur Theorem about empirical covariance matrices is not true anymore for such vectors. More specifically, we have:

Theorem 1.1. *Let $U_N^1, U_N^2, \ldots$ be i.i.d. copies of the vector $U_N(t) \in \mathbb{C}^N$ defined in (3). Let*

$$(4) \qquad\qquad M = \frac{1}{N} \sum_{i=1}^p U_N^i (U_N^i)^*$$

be the (dilated) empirical covariance matrix of the sample $U_N^1, \ldots, U_N^p$. As $N, p \to \infty$ with $p/N \to \lambda > 0$, the empirical spectral law of M converges to a limit law $P_{\lambda,t}$ with unbounded support which is characterized by its moments, given by the formula

$$(5) \qquad\qquad \int x^k P_{\lambda,t}(\mathrm{d}x) = \sum_{\pi \in \mathrm{Part}(k)} e^{-\kappa(\pi)t} \lambda^{|\pi|}$$

with $\mathrm{Part}(k)$ the set of partitions of $\{1, \ldots, k\}$ and κ defined by:

- $\kappa(\pi) = \kappa(\pi_1) + \cdots + \kappa(\pi_q)$ *if $\pi_1, \ldots, \pi_q$ are the connected components of π;*

- *if π is connected, $\kappa(\pi)$ is the number of times one changes from one block to another when running through π in a cyclic way.*

Note that the law $P_{\lambda,t}$ is the Poisson law with parameter λ if $t = 0$ and tends to the free Poisson law (also called Marchenko-Pastur law) as t tends to $+\infty$: the weight $e^{-\kappa(\pi)t}$ penalizes *crossings* in partitions (indeed, $\kappa(\pi) = 0$ if and only if π is non crossing). An illustration is given in Figure 1 below. In fact, Formula (5) can be generalized into a more general moments-cumulants formula which provides a new continuous interpolation between classicaly and freely infinitely divisible laws. This interpolation is related to the notion of t-freeness developed by the first named author and Lévy in [11] and is based on a progressive penalization of the crossings in the moments-cumulants formula (19).

The paper is organized as follows. Next section describes the main results; Section 3 provides some applications; the following sections are devoted to the proofs; the last section is an appendix where we recall some facts on the infinitely divisible laws, the Bercovici-Pata bijection and (hyper)graphs, for the easyness of the reader.

As this preprint was published, Victor Pérez-Abreu informed us that he is working on a close subject with J. Armando Domínguez-Molina and Alfonso Rocha-Arteaga in the forthcoming preprint [18].

(a) Case where $t = 0.01$

(b) Case where $t = 0.1$

(c) Case where $t = 1$
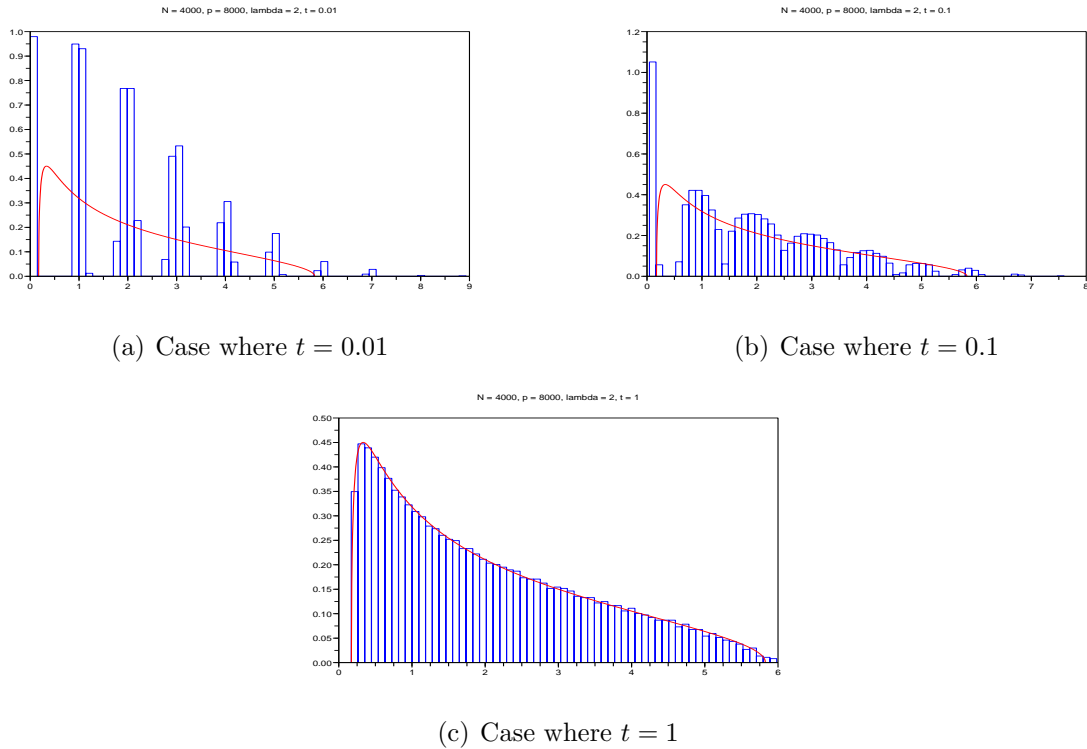
FIGURE 1. The empirical spectral distribution of the matrix M of (4) for $N = 4.10^3$, $p = 8.10^3$ at several values of t. We see that this distribution, which is approximately the law $P_{2,t}$, is close to the **Poisson law** with parameter 2 for t close to zero, and that it converges to the **Marchenko-Pastur law** with parameter 2 (whose density is plotted by a smooth continuous line) as t grows.

2. The main results

2.1. **A general family of matrix ensembles.** The basic facts on the *-*infinitely divisible* laws are recalled in the appendix. For μ such a law, its *Lévy exponent* is the function Ψ_μ such that the Fourier transform of μ is $e^{\Psi_\mu(\cdot)}$.

2.1.1. *Case of compound Poisson laws.* Such a law μ is a one of

$$\sum_{i=1}^{P(\lambda)} X_i,$$

where $P(\lambda)$ is a random variable with Poisson law with expectation λ and the X_i's are i.i.d. random variables, independent of $P(\lambda)$. In this case, if ν denotes the law of the X_i's, the Lévy exponent of μ is $\Psi_\mu(\xi) = \lambda \int (e^{it\xi} - 1) d\nu(t)$. For $N \geq 1$, let $(U_N^i)_{i \geq 1}$ be a sequence of i.i.d. copies of a random column vector $U_N \in \mathbb{R}^{N \times 1}$ or $\mathbb{C}^{N \times 1}$. Then we define $\mathbb{P}_{U_N}^{(\mu)}$ to be the law of

$$\frac{1}{N} \sum_{i=1}^{P(N\lambda)} X_i \cdot U_N^i (U_N^i)^*,$$

where $P(N\lambda)$ is a random variable with Poisson law with expectation $N\lambda$, the X_i's are i.i.d. random variables with law ν and the U_N^i's are i.i.d. copies of U_N (whose conjugate transpose are denoted by $(U_N^i)^*$), all being independent. Let us notice that $\mathbb{P}_{U_N}^{(\mu)}$ is still a compound Poisson law and that its Lévy exponent is given by:

$$A \mapsto N\mathbb{E}\left[\Psi_\mu(U_N^* A U_N / N)\right]$$

for any $N \times N$ Hermitian matrix A.

2.1.2. *General case.* Here, we shall extend the previous construction for μ a general infinitely divisible law. In the following theorem and in the rest of the paper, the spaces of probability measures are endowed with the weak topology. $\mathbb{K}$ denotes either $\mathbb{R}$ or $\mathbb{C}$.

Theorem 2.1. *Let μ be an infinitely divisible law on $\mathbb{R}$, let us fix $N \geq 1$ and let $U_N \in \mathbb{K}^{N \times 1}$ be a random column vector such that $\mathbb{E}\left[\|U_N\|_2^4\right] < +\infty$. Let $(U_N^i)_{i \geq 1}$ be a sequence of i.i.d. copies of U_N and, for each $n \geq 1$, let $X_n^1, X_n^2, \ldots, X_n^{Nn}$ be i.i.d. $\mu^{*\frac{1}{n}}$-distributed random variables. Then the sequence of $N \times N$ Hermitian matrices*

$$(6) \qquad \frac{1}{N} \sum_{i=1}^{N \times n} X_n^i \cdot U_N^i (U_N^i)^*$$

converges in distribution, as $n \longrightarrow \infty$ (N being fixed), to a probability measure $\mathbb{P}_{U_N}^{(\mu)}$ on the space of $N \times N$ Hermitian matrices, whose Fourier transform is given by

$$(7) \qquad \int e^{i \operatorname{Tr}(AM)} d\mathbb{P}_{U_N}^{(\mu)}(M) = \exp\left\{N\mathbb{E}\left[\Psi_\mu(U_N^* A U_N / N)\right]\right\}$$

for any $N \times N$ Hermitian matrix A.

The following proposition extends the theorem to a quite more general framework, where the X_n^i's are i.i.d. but not necessarily distributed according to $\mu^{*\frac{1}{n}}$ and only satisfy the limit theorem

$$\text{law of } (X_n^1 + \cdots\cdots + X_n^{k_n}) \xrightarrow[n \to \infty]{} \mu$$

for a sequence $k_n \to \infty$.

Proposition 2.2. *a) If the law of U_N is compactly supported, then to any limit theorem*

$$\underbrace{\nu_n * \cdots \cdots * \nu_n}_{k_n \ times} \xrightarrow[n\to\infty]{} \mu,$$

there corresponds a limit theorem in the space of $N \times N$ Hermitian matrices

$$(8) \qquad \frac{1}{N} \sum_{i=1}^{N \times k_n} X_n^i \cdot U_N^i (U_N^i)^* \xrightarrow[n\to\infty]{} \mathbb{P}_{U_N}^{(\mu)},$$

where $X_n^1, X_n^2, \ldots, X_n^{Nk_n}$ are i.i.d. ν_n-distributed random variables, $(U_N^i)_{i\geq 1}$ is a sequence of i.i.d. copies of U_N, independent of the X_n^i's.

b) In the case where the law of U_N is not compactly supported, (8) stays true as long as one supposes that $k_n \times \int |t| \nu_n(\mathrm{d}t)$ is bounded uniformly in n.

Remark 2.3. Such a construction has been generalized to the more general setting of Hopf algebras by Schürmann, Skeide and Volkwardt in [29].

2.2. Convergence of the empirical spectral law.

The *empirical spectral law* of a matrix is the uniform probability measure on its eigenvalues (see Equation (44) in the appendix). When $U_N/\sqrt{N}$ is uniformly distributed on the unit sphere, we proved that the empirical spectral law associated to $\mathbb{P}_{U_N}^{(\mu)}$ converges when the size N tends to infinity (*cf.* [6, 16]). In order to obtain other convergences, we shall first make the following assumptions on the random vector $U_N \in \mathbb{K}^{N\times 1}$. We denote the entries of U_N by $U_N(1), \ldots, U_N(N)$. Roughly speaking, these assumptions mean that the $U_N(i)$'s are exchangeable and that the moment of order k of $U_N(1)$ grows at most in $N^{\frac{k}{2}-1}$, for all k. These assumptions will be weakened in the next section to consider heavy tailed variables.

Hypothesis 2.4. *a) For each $N \geq 1$, the entries of U_N are exchangeable and have moments of all orders.*

b) As N goes to infinity, we have:

- *for each $p \geq 1$, for each $i_1, \ldots, i_{2p} \geq 1$,*

$$(9) \qquad \mathbb{E}[U_N(i_1) \cdots U_N(i_p) \overline{U_N(i_{p+1})} \cdots \overline{U_N(i_{2p})}] = O(N^{p-|\{i_1,\ldots,i_{2p}\}|}),$$

- *for each $k \geq 1$, for all positive integers $n_1, \ldots, n_k$, there exists $\Gamma(n_1, \ldots, n_k)$ finite such that*

$$(10) \qquad \mathbb{E}\left[\frac{|U_N(1)|^{2n_1}}{N^{n_1-1}} \cdots \cdots \frac{|U_N(k)|^{2n_k}}{N^{n_k-1}} \right] \xrightarrow[N\to\infty]{} \Gamma(n_1, \ldots, n_k).$$

Moreover, there is a constant C such that for all $k \geq 1$, for all $n_1, \ldots, n_k$,

$$(11) \qquad \Gamma(n_1, \ldots, n_k) \leq C^{n_1+\cdots+n_k}.$$

Remark 2.5. Let define the random column vector $\tilde{U}_N = \mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_N)U_N$, with $\varepsilon_1, \ldots, \varepsilon_N$ some i.i.d. variables uniformly distributed on $\{z \in \mathbb{K} \,;\, |z| = 1\}$, independent of U_N. Note that if U_N satisfies Hypothesis 2.4, so does $\tilde{U}_N$, with the same function Γ. Moreover, the expectation of (9) is null for $\tilde{U}_N$ as soon as the multisets $\{i_1, \ldots, i_k\}$ and $\{i_{k+1}, \ldots, i_{2k}\}$ are not equal (in the complex case) and as soon as an element in the multiset $\{i_1, \ldots, i_{2k}\}$ appear an odd number of times (in the real case).

The following theorem is the key result of the paper. For its second part, we endow the set of functions on multisets of integers (it is the set that Γ belongs to) with the product topology.

Theorem 2.6. *We suppose that Hypothesis 2.4 holds.*

a) For any $*$-infinitely divisible distribution μ, the empirical spectral distribution of a $\mathbb{P}_{U_N}^{(\mu)}$-distributed random matrix converges almost surely, as $N \longrightarrow \infty$, to a deterministic probability measure $\Lambda_\Gamma(\mu)$, which depends only on μ and on the function Γ of Equation (10).

b) The probability measure $\Lambda_\Gamma(\mu)$ depends continuously on the pair (μ, Γ).

c) The moments of $\Lambda_\Gamma(\mu)$ moments can be computed when μ has moments to all orders via the following formula,

$$(12) \qquad \int x^k \Lambda_\Gamma(\mu)(\mathrm{d}x) = \sum_{\pi \in \mathrm{Part}(k)} f_\Gamma(\pi) \prod_{J \in \pi} c_{|J|}(\mu), \qquad (k \geq 1)$$

where the non negative numbers $f_\Gamma(\pi)$, which factorize along the connected components of π, are given at Lemma 6.3 and the numbers $c_{|J|}(\mu)$ are the cumulants of π (whose definition is recalled in the appendix).

d) If the Lévy measure of μ has compact support, then $\Lambda_\Gamma(\mu)$ admits exponential moments of all orders.

The following proposition allows to assert that many limit laws obtained in Theorem 2.6 have unbounded support. Recall that a law is said to be *non degenerate* if it is not a Dirac mass.

Proposition 2.7. *If the function Γ of Hypothesis 2.4 is such that $\inf \Gamma(n)^{1/n} > 0$, then for any non degenerate $*$-infinitely divisible distribution μ whose classical cumulants are all non negative, the law $\Lambda_\Gamma(\mu)$ has unbounded support.*

3. APPLICATIONS AND EXAMPLES

3.1. **Heavy-tailed Marchenko-Pastur theorem.** In this section, we use Theorem 2.6 to extend the theorem of Marchenko and Pastur described in the introduction to vectors

with heavy-tailed entries. This also extends Theorem 1.10 of Belinschi, Guionnet and Dembo in [4] (their theorem corresponds to the case $X_i = 1$), but we do not provide the explicit characterization of the limit that they propose. Also, our result is valid in the complex as in the real case, whereas the one of [4] is only stated in the real case (but we do not know whether this is an essential restriction of the approach of [4]).

Let us fix $\alpha \in (0, 2)$ and let $(Y_{ij})_{i,j \geq 1}$ be an infinite array of i.i.d. $\mathbb{K}$-valued random variables such that the function

$$(13) \qquad L(t) := t^{\alpha} \mathbb{P}(|Y_{11}| \geq t)$$

has slow variations as $t \to \infty$. For each j, define the column vector $V^j := (Y_{ij})_{i=1}^N \in \mathbb{K}^{N \times 1}$ (V^j depends implicitly on N). Let us define $a_N := \inf\{t \, ; \, t^{-\alpha} L(t) \leq 1/N\}$.

Theorem 3.1. *For any set $(X_j)_{j \geq 1}$ of i.i.d. real random variables (with any law, but that does not depend on N) independent of the Y_{ij}'s and any fixed $\lambda > 0$, the empirical spectral law of the random matrix*

$$M_{N,p} := \frac{1}{a_N^2} \sum_{j=1}^p X_j \cdot V^j (V^j)^*$$

converges almost surely, as $N, p \to \infty$ with $p/N \to \lambda$, to a deterministic probability measure which depends only on α, on λ and on the law of the X_i's. This limit law depends continuously of these three parameters.

3.2. Covariance matrices with exploding moments.

The following theorem is a direct consequence of Theorem 2.6 and Proposition 2.7. It is the "covariance matrices version" of Zakharevich's generalization of Wigner's theorem to matrices with exploding moments (see [32] and also the recent work [22] by Male).

Theorem 3.2. *Let $M_{N,p} = [X_{ij}]$ be a complex $N \times p$ random matrix with i.i.d. centered entries whose distribution might depend on N and p. We suppose that there is a sequence $\mathbf{c} = (c_k)_{k \geq 2}$ such that $c_k^{1/k}$ is bounded and such that for each fixed $k \geq 2$,*

$$\frac{\mathbb{E}[|X_{11}|^k]}{N^{\frac{k}{2}-1}} \longrightarrow c_k \qquad \qquad as \ N, p \to \infty \ with \ p/N \to \lambda > 0.$$

Then the empirical spectral law of

$$\frac{1}{N} M_{N,p} M_{N,p}^*$$

converges, as $N, p \to \infty$ with $p/N \to \lambda > 0$, to a probability measure $\mu_{\lambda,\mathbf{c}}$ which depends continuously on the pair $(\lambda, \mathbf{c})$. If $c_k = 0$ for all $k \geq 3$, $\mu_{\lambda,\mathbf{c}}$ is Marchenko-Pastur distribution with parameter λ dilated by a coefficient $\sqrt{c_2}$. Otherwise, $\mu_{\lambda,\mathbf{c}}$ has unbounded support but admits exponential moments of all orders.

3.3. **Non centered Gaussian vectors, Brownian motion on the unit sphere and a continuum of Bercovici-Pata's bijections.** In this section, we give examples of vectors U_N which satisfy Hypothesis 2.4 for a certain function Γ. These examples will allow to construct the continuum of Bercovici-Pata bijections $(\Lambda_t)_{t \geq 0}$ mentioned in the introduction.

Let us fix $t \geq 0$ and consider $U_N = U_{N,t}$ defined thanks to one of the two (or three, actually) following definitions

(a) *Gaussian and uniform cases*:

$$(14) \qquad U_{N,t} := e^{-\frac{t}{2}} \sqrt{N} X_N + \sqrt{1 - e^{-t}} G_N,$$

where X_N is uniformly distributed on the canonical basis $(1, 0, \ldots, 0)^T, \ldots (0, \ldots, 0, 1)^T$ of $\mathbb{K}^{N \times 1}$, independent of $G_N := (g_1, \ldots, g_N)^T$, with $g_1, g_2, g_3, \ldots$ i.i.d. variables whose distribution is either the centered Gaussian law on $\mathbb{K}$ with variance 1 or the uniform law on $\{z \in \mathbb{K} \, ; \, |z| = 1\}$,

(b) *Brownian motion*: $U_{N,t}/\sqrt{N}$ is a *Brownian motion on the unit sphere of* $\mathbb{C}^{N \times 1}$ taken at time t, whose distribution at time zero is the uniform law on the canonical basis of $\mathbb{C}^{N \times 1}$. Such a process is a strong solution of the SDE

$$(15) \qquad \mathrm{d}U_{N,t} = (\mathrm{d}K_t)U_{N,t} - \frac{1}{2}U_{N,t}\mathrm{d}t,$$

where $K = (K_t)_{t \geq 0}$ is a standard Brownian motion on the Euclidian space of $N \times N$ skew-Hermitian matrices, endowed with the scalar product $A \cdot B = N \operatorname{Tr}(A^*B)$.

Of course, these models make sense for $t = +\infty$: in the first model, it means only that $U_{N,t} = G_N$ and the second model, at $t = +\infty$, can be understood as its limit in law, *i.e.* a random vector with uniform law on the sphere with radius $\sqrt{N}$. For $t = +\infty$, all formulas below make sense (and stay true) by taking their $t \to +\infty$ limits.

Thanks to the results of [9], it can be seen that the Gaussian model and the Brownian one have approximately the same finite-dimensional marginals, as $N \to \infty$. The following proposition, whose proof is postponed to Section 9, makes this analogy stronger.

Proposition 3.3. *For $U_N = U_{N,t}$ as defined according to any of the above models, Hypothesis 2.4 holds for $\Gamma := \Gamma_t$ given by the following formulas:*

$$(16) \qquad \Gamma_t(\underbrace{1, \ldots, 1}_{k \ times}) = (1 - e^{-t})^k + ke^{-t}(1 - e^{-t})^{k-1} \quad \text{for every } k \geq 1,$$

$$(17) \qquad \Gamma_t(n, \underbrace{1, \ldots, 1}_{k \ times}) = e^{-nt}(1 - e^{-t})^k \quad \text{for every } n \geq 2 \text{ and } k \geq 0,$$

$$(18) \qquad \Gamma_t(n_1, \ldots, n_k) = 0 \quad \text{if there is } i \neq j \text{ such that } n_i \geq 2, \, n_j \geq 2.$$

Let us now consider the family of transforms $(\Lambda_{\Gamma_t}, t \geq 0)$ (denoted by Λ_t in the introduction).

In both models, when $t = 0$, the rank-one random projector $U_{N,t}U_{N,t}^*$ is a diagonal matrix with unique non zero entry uniformly distributed on the diagonal and equal to N. In such a case, it can be readily seen that $\mathbb{P}_{U_{N,0}}^{(\mu)}$ is the law of a diagonal matrix with i.i.d. entries with law μ. Owing to the law of large numbers, Λ_{Γ_0} is obviously the identity map. Moreover, it has been seen in [6, 16] that for $t = \infty$, Λ_{Γ_∞} is the Bercovici-Pata bijection.

Therefore, $(\Lambda_{\Gamma_t}, t \in [0, +\infty])$ provides a continuum of maps passing from the identity ($t = 0$) to the Bercovici-Pata bijection ($t = +\infty$). This continuum is related to the notion of t-freeness developed by the first named author and Lévy in [11]. The maps Λ_{Γ_t} are made explicit (at least at the moments level) by the following proposition, whose proof, based on the explicitation of the functions f_{Γ_t}, is postponed to Section 10.

Proposition 3.4. *Let μ be an infinitely divisible law with moments of all orders. Then for each $t \geq 0$,*

$$(19) \qquad \int x^k \Lambda_{\Gamma_t}(\mu)(\mathrm{d}x) = \sum_{\pi \in \mathrm{Part}(k)} e^{-\kappa(\pi)t} c_\pi(\mu) \qquad (k \geq 1)$$

with κ defined by:

- *$\kappa(\pi) = \kappa(\pi_1) + \cdots + \kappa(\pi_q)$ if $\pi_1, \ldots, \pi_q$ are the connected components of π;*

- *if π is connected, $\kappa(\pi)$ is the number of times one changes from one block to another when running through π in a cyclic way.*

Example of computation of $\kappa(\pi)$. For instance, let us consider the partition $\pi = \{\{1, 8, 10\}, \{2, 4\}, \{3, 5\}, \{6, 7, 9\}\} \in \mathrm{Part}(10)$ (*cf.* Figure 2). Then the connected components of π are the partitions π_1 and π_2 induced by π on the sets $\{1, 6, 7, 8, 9, 10\}$ and $\{2, 3, 4, 5\}$. Since $\kappa(\pi_1) = 5$ and $\kappa(\pi_2) = 4$, $\kappa(\pi) = 9$.
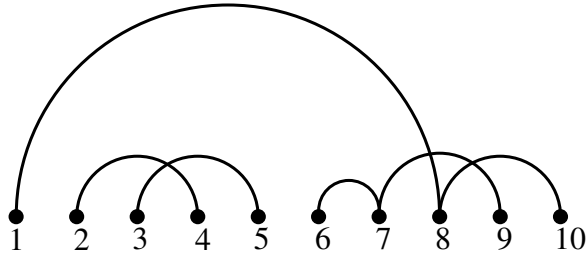

FIGURE 2. The partition $\pi = \{\{1, 8, 10\}, \{2, 4\}, \{3, 5\}, \{6, 7, 9\}\} \in \mathrm{Part}(10)$.

Let us now say a few words about the bijection Λ_{Γ_t}, as a function of $t \in [0, +\infty)$.

If $t = 0$, then:

$$\int x^k \Lambda_{\Gamma_0}(\mu)(\mathrm{d}x) = \sum_{\pi \in \mathrm{Part}(k)} c_\pi(\mu) = \int x^k \mu(\mathrm{d}x).$$

This corroborates the fact that Λ_{Γ_0} is the identity map.

If t tends to infinity, then:

$$\lim_{t \to +\infty} \int x^k \Lambda_{\Gamma_t}(\mu)(\mathrm{d}x) = \sum_{\pi \in \mathrm{NC}(k)} c_\pi(\mu) = \int x^k \Lambda(\mu)(\mathrm{d}x)$$

where $\mathrm{NC}(k)$ denotes the set of non-crossing partitions of $\{1, \ldots, k\}$ (see Section 5.1 for a precise definition). Indeed, $\kappa(\pi)$ is positive unless π is non-crossing. Since the continuity of Λ_Γ w.r.t. the weak topology is uniform in Γ, as it appears from the proof of Proposition 6.5 b), this proves that Λ_{Γ_t} tends to the Bercovici-Pata bijection Λ when t goes to infinity, as expected.

3.4. **The distribution $\mathbb{P}_{U_N}^{(\mu)}$ made more explicit.** When μ is a compound Poisson law, the definition of $\mathbb{P}_{U_N}^{(\mu)}$ has been made explicit in Section 2.1.1. In this section, we explicit $\mathbb{P}_{U_N}^{(\mu)}$ for μ a Dirac mass or a Gaussian laws (the column vector U_N underlying the definition of $\mathbb{P}_{U_N}^{(\mu)}$ staying as general as possible). Since any infinitely divisible law is a weak limit of convolutions of such laws and obviously, by the Formula (7) of the Fourier transform of $\mathbb{P}_{U_N}^{(\mu)}$, the law $\mathbb{P}_{U_N}^{(\mu)}$ depends continuously on μ and satisfies

$$\mathbb{P}_{U_N}^{(\mu)} * \mathbb{P}_{U_N}^{(\nu)} = \mathbb{P}_{U_N}^{(\mu*\nu)},$$

this gives a good idea of what a random matrix distributed according to $\mathbb{P}_{U_N}^{(\mu)}$ looks like. Moreover, we also consider the case where μ is a Cauchy law, where a surprising behaviour w.r.t. the convolution appears.

First, it can easily be seen that if μ is the Dirac mass at $\gamma \in \mathbb{R}$, then $\mathbb{P}_{U_N}^{(\mu)}$ is the Dirac mass at

$$\gamma \mathbb{E}[U_N U_N^*] = \gamma \left(\mathbb{E}\left[|U_N(1)|^2\right] - \mathbb{E}\left[U_N(1)\overline{U_N(2)}\right] \right) + \text{ rank-one matrix.}$$

Hence, due to Hypothesis 2.4 and Lemma 12.2, $\Lambda_\Gamma(\delta_\gamma) = \delta_{\gamma'}$ for $\gamma' = \gamma \Gamma(1)$.

Suppose that now μ is the standard Gaussian law $\mathcal{N}(0,1)$ and that Hypotheses 2.4 holds. Let $\tilde{U}_N$ be as in Remark 2.5. Then the distribution $\mathbb{P}_{\tilde{U}_N}^{(\mu)}$ only depends on $\alpha := \mathbb{E}[|U_N(1)|^2|U_N(2)|^2]$, $\gamma := \mathbb{E}\left[|U_N(1)|^4\right]$ and $\beta := \dim_{\mathbb{R}}(\mathbb{K})$: when $\gamma \geq \alpha$, $\mathbb{P}_{U_N}^{(\mu)}$ is the law of

$$\sqrt{\frac{\gamma - 2\alpha}{N}} \begin{pmatrix} g_1 & & \\ & \ddots & \\ & & g_N \end{pmatrix} + \sqrt{\frac{\alpha}{N}} \begin{pmatrix} g & & \\ & \ddots & \\ & & g \end{pmatrix} + \sqrt{\frac{2\alpha}{N\beta}} \text{GO(U)E},$$

where $g_1, \ldots, g_N, g$ are standard Gaussian variables and GO(U)E designs a GOE or GUE matrix (according to weither $\mathbb{K} = \mathbb{R}$ of $\mathbb{C}$) as defined p. 51 of [2] (*i.e.* a standard Gaussian vector on the space of real symmetric or Hermitian matrices endowed with the scalar product $A \cdot B := \frac{\beta}{2} \operatorname{Tr} AB$). As a consequence, since when $N \to \infty$, $\alpha \sim \Gamma(1,1)$ and $\gamma \sim N\Gamma(2)$, the spectral law of a $\mathbb{P}_{\tilde{U}_N}^{(\mu)}$-distributed matrix converges to

$$\Lambda_\Gamma(\mu) = \mathcal{N}(0, \Gamma(2)) \boxplus (\text{Semicircle law with variance } 2\Gamma(1,1)/\beta).$$

In the next proposition, we consider the case where μ is a *Cauchy law with paramater t*:

$$\mu = \mathcal{C}_t(\mathrm{d}x) := \frac{t\mathrm{d}x}{\pi(t^2 + x^2)}.$$

Let P_t be the associated kernel, defined by

$$P_t(f)(x) = \int f(x + y)\mathcal{C}_t(\mathrm{d}y).$$

Proposition 3.5. *We suppose that* $\mathbb{E}[U_N U_N^*] = I_N$. *For* M_t *a* $\mathbb{P}_{U_N}^{(\mathcal{C}_t)}$-*distributed random matrix and* $f : \mathbb{R} \to \mathbb{R}$ *bounded, for any* $N \times N$ *Hermitian matrix* A,

$$\mathbb{E}[f(A + M_t)] = P_t(f)(A),$$

real functions being applied to Hermitian matrices via the functional calculus.

The following Corollary follows easily, using the vector $V_N := \tilde{U}_N/\sqrt{\Gamma(1)}$ instead of U_N (with $\tilde{U}_N$ as defined in Remark 2.5).

Corollary 3.6. *Under Hypothesis 2.4, the set of Cauchy laws is invariant by* Λ_Γ : *for any* $t > 0$, $\Lambda_\Gamma(\mathcal{C}_t) = \mathcal{C}_{t'}$, *for* $t' = \Gamma(1)t$.

4. Proof of Theorem 2.1 and Proposition 2.2

One proves the theorem and the proposition in the same time, by showing that under the hypotheses of the theorem or of a) or b) of the remark, the Fourier transform of the left hand term of (8), that we denote by M_n, converges pointwise to the right hand term of (7). Indeed, for any Hermitian matrix A,

$$\mathbb{E}[e^{i \operatorname{Tr}(AM_n)}] = \left(1 + \frac{\mathbb{E}[\Psi_n(U_N^* A U_N/N)]}{k_n}\right)^{k_n},$$

where for any $x \in \mathbb{R}$, $\Psi_n(x) := k_n \int (e^{itx} - 1)\mathrm{d}\nu_n(t)$. The function Ψ_n converges to Ψ_μ as $n \longrightarrow \infty$, uniformly on every compact subset of $\mathbb{R}$ (this follows from [27, Lem. 3.1] and from the fact that $\hat{\nu}_n^{k_n} \longrightarrow \hat{\mu}$ uniformly on every compact set). It is enough to prove the result in the case where the law of U_N is compactly supported. To conclude in the case where $\nu_n = \mu^{*\frac{1}{k_n}}$ (resp. in the case where $\int |t|\nu_n(\mathrm{d}t) = O(1/k_n)$), one needs to argue that there is a constant C such that for all $t > 0$, $\Psi_{\mu^t}(\lambda) \leq tC(|\lambda| + \lambda^2)$ (resp. that $|\Psi_n(x)| \leq k_n \int |t|\nu_n(\mathrm{d}t)$).

5. Preliminaries for the proof of Theorem 2.6 : partitions and graphs

The aim of this section is to introduce the combinatorial definitions which will be used in the next section in order to prove Theorem 2.6. The conventions we chose for the definitions of partitions, graphs, and for the less well-known notion of hypergraph are presented in Section 12.3 of the appendix.

5.1. Partitions.

- Let us recall that we denote by $\mathrm{Part}(k)$ the set of partitions of $V = \{1, \ldots, k\}$, and by $\mathrm{NC}(k)$ the set of non-crossing partitions of $\{1, \ldots, k\}$ (a partition π of $\{1, \ldots, k\}$ is said to be *non-crossing* if there does not exist $x < y < z < t$ such that $x \overset{\pi}{\sim} z \overset{\pi}{\nsim} y \overset{\pi}{\sim} t$).
- For any given partition $\pi \in \mathrm{Part}(k)$, we denote by $\mathrm{nc}(\pi)$ be the minimal non-crossing partition which is above π for the refinement order; the partitions induced by π on the blocks of $\mathrm{nc}(\pi)$ are called the *connected components* of π.
- If $\mathrm{nc}(\pi)$ has only one block, then π is said to be *connected*.
- We define $\mathrm{thin}(\pi)$ to be the partition induced by π on the subset of $\{1, \ldots, k\}$ obtained by erasing ℓ whenever $\ell + 1 \overset{\pi}{\sim} \ell$ (with the convention $k + 1 = 1$).
- A partition π is said to be *thin* if $\pi = \mathrm{thin}(\pi)$.

For instance, for $\pi = \{\{1, 8, 10\}, \{2, 4\}, \{3, 5\}, \{6, 7, 9\}\} \in \mathrm{Part}(10)$ the partition $\mathrm{nc}(\pi)$ is $\{\{1, 6, 7, 8, 9, 10\}, \{2, 3, 4, 5\}\}$ and Figure 3 illustrates of the operation $\pi \mapsto \mathrm{thin}(\pi)$.
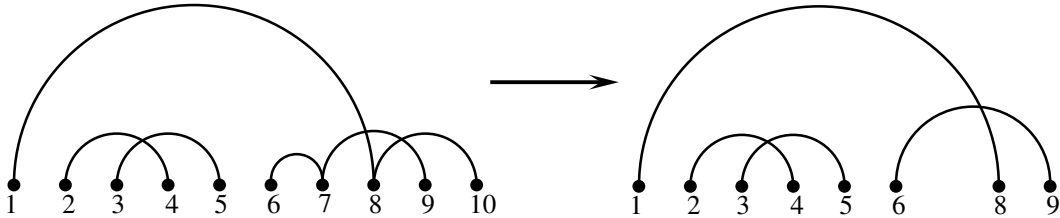
FIGURE 3. Illustration of the operation $\pi \mapsto \mathrm{thin}(\pi)$

For any function f defined on a set E, we denote by $\ker f$ the partition of E whose blocks are the level sets of f. For any partition $\pi \in \mathrm{Part}(k)$, for each $\ell \in \{1, \ldots, k\}$, we denote by $\pi(\ell)$ the index of the class of ℓ in π, after having ordered the classes according to the order of their first element (for the example given Figure 2, we have $\pi(1) = 1$, $\pi(2) = 2$, $\pi(3) = 3$ and $\pi(4) = 2$, $\pi(5) = 3$, $\pi(6) = 4$,...). Moreover, 0 is identified with k and $k + 1$ is identified with 1, so that $\pi(0) = \pi(k)$ and $\pi(k + 1) = \pi(1)$.

5.2. Graphs.
Let $G = (V, E = (e_i, i \in I))$ be a graph, and π a partition of V. Let $\tilde{\pi}$ be the canonical surjection from V onto π. The *quotient graph* $G_{/\pi}$ is the graph with π as set

of vertices and with edges $(e_i^{(\pi)}, i \in I)$ such that $e_i^{(\pi)}$ is the edge between $\tilde{\pi}(v)$ and $\tilde{\pi}(w)$ if e_i is an edge between v and w, with same direction if the graph is directed. Note that the quotient graph can have strictly less vertices than the initial one, but it has as much edges as the initial one. Note also that the quotient of a circuit is still a circuit.

For instance, let us consider the preceding partition π and G be the cyclic graph with vertex set $V = \{1, \ldots, 10\}$ and edges $1 \to 2 \to \cdots \to 10 \to 1$ (*cf.* Figure 4(a) where we draw the partition π with dashed lines). The quotient graph $G_{/\pi}$ is then given by Figure 4(b) with $V1 = \{1, 8, 10\}$, $V2 = \{2, 4\}$, $V3 = \{3, 5\}$ and $V4 = \{6, 7, 9\}$.

(a) A cyclic graph G and a partition π of its vertices

$\longrightarrow$

(b) The quotient $G_{/\pi}$ of the left graph G by the partition π
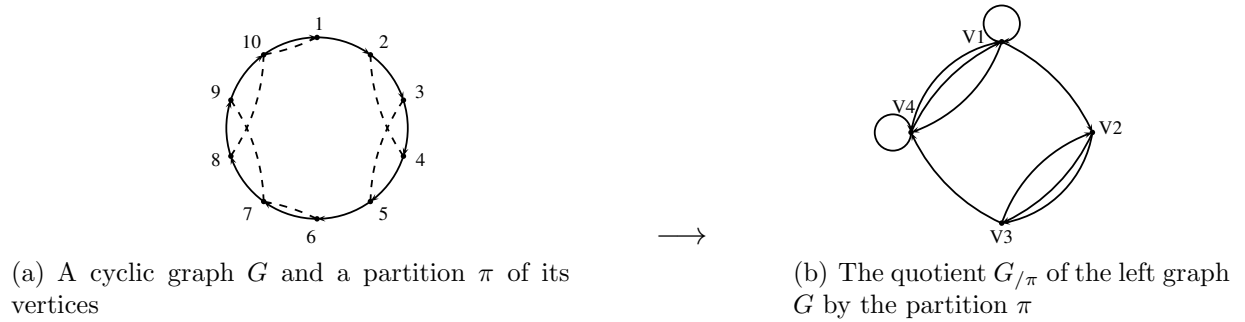
FIGURE 4. A cyclic graph and one of its quotients

Let $G = (V, E)$ a directed graph. For any vertex $v \in V$, we denote by $E^+(v)$ the subset of out-going edges from v, and by $E^-(v)$ the subset of in-going edges to v:

$$\begin{aligned} E^+(v) &= \{e \in E \,;\, (\exists v' \in \pi)(e = (v, v'))\} \\ E^-(v) &= \{e \in E \,;\, (\exists v' \in \pi)(e = (v', v))\} \end{aligned}$$

Notice that $E^+(v)$ and $E^-(v)$ are not necessarily disjoints.

Let us consider $\{1, \ldots, N\}^E$ as a set of "colorings" of the edges of G by the colors $1, \ldots, N$. For any coloring $c \in \{1, \ldots, N\}^E$, let $n_c^+(v, i) := |E^+(v) \cap c^{-1}(\{i\})|$ and $n_c^-(v, i) := |E^-(v) \cap c^{-1}(\{i\})|$. If for each vertex v and each color i, $n_c^+(v, i) = n_c^-(v, i)$, the coloring is called *admissible*.

Figure 5 presents an instance of an admissible coloring of $G_{/\pi}$ with three colors.

A partition τ of the edges of G will be said *admissible* if it is the kernel of an admissible coloring. In this case, for any vertex $v \in G$, we define

(20) $$n_\tau(v) := (n_\tau(v, 1), n_\tau(v, 2), \ldots \ldots)$$

to be the common value of the decreasing reordering of the families $(n_c^+(v, 1), n_c^+(v, 2), \ldots \ldots)$ for c coloring such that $\ker c = \tau$.
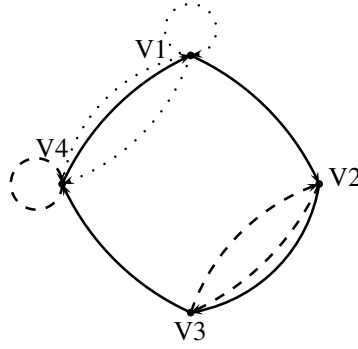
FIGURE 5. An admissible coloring

5.3. **Hypergraph associated to a partition of the edges of a graph.** Let G be a graph with vertex set V, π be a partition of V and τ be a partition of the edge set of G. Then one can define $H(\pi, \tau)$ to be the hypergraph with the same vertex set as $G_{/\pi}$ (i.e. π) and with edges $(E_W, W \in \tau)$, where each edge E_W is the set of blocks $J \in \pi$ such that at least one edge of $G_{/\pi}$ starting or ending at J belongs to W.

For instance, with π, $G_{/\pi}$ and τ as given by the preceding example Figure 5, $H(\pi, \tau)$ is the hypergraph with three edges drawn in Figure 6. Another example, where $H(\pi, \tau)$ has no cycle, is given at Figure 7.
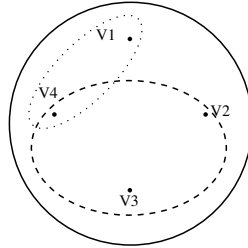
FIGURE 6. The graph $H(\pi, \tau)$ with π, $G_{/\pi}$ and τ as given by the preceding example Figure 5

The next proposition will be used in the following. For τ a partition of the edges of a graph, a cycle of the graph is said to be τ-*monochromatic* if all the edges it visits belong to the same block of τ.

Proposition 5.1. *Let $G = (V, E)$ be a directed graph, which is a circuit. Fix $\pi \in \mathrm{Part}(V)$ and τ be a partition of the edge set of G such that $H(\pi, \tau)$ has no cycle. Then $G_{/\pi}$ is a disjoint union of τ-monochromatic cycles. As a consequence, τ is admissible.*

Proof. If τ has more than one block (the case with one bloc being obvious), since $G_{/\pi}$ is a circuit, one can find a closed path

$$\gamma = (v_0, e_1, v_1, e_2, \dots, e_k, v_k) \qquad (\text{with } v_k = v_0)$$

of $G_{/\pi}$ which visits each edge of $G_{/\pi}$ exactly once and such that e_1 and e_k do not belong to the same block of τ. For each edge e of $G_{/\pi}$, let $E(e)$ be the edge of $H(\pi, \tau)$ consisting of all vertices of $G_{/\pi}$ which are the beginning or the end of an edge of $G_{/\pi}$ with the same color as e (i.e. in the same block of τ as e). Then $(v_0, E(e_1), v_1, E(e_2), \ldots, E(e_k), v_k)$ is a path of $H(\pi, \tau)$, i.e. for all $\ell = 1, \leq, k$, $v_{\ell-1}, v_\ell \in E(e_\ell)$. Let us introduce $i_0 = 0 < i_1 < \cdots < i_p = k$ such that $i_1, \ldots, i_p$ are the instants where the color of the edges change in γ (i.e. for all $s = 1, \ldots, p$, $e_{i_{s-1}+1}, \ldots, e_{i_s}$ have the same color, and for all $s = 1, \ldots, p-1$, e_{i_s} and e_{i_s+1} do not have the same color. Then

$$\bar{\gamma} := (v_{i_0}, E(e_{i_1}), v_{i_1}, \ldots \ldots, v_{i_{p-1}}, E(e_{i_p}), v_{i_p})$$

is a also a path of $H(\pi, \tau)$.

Since $H(\pi, \tau)$ has no cycle, hence is linear, $v_{i_r} = E(e_{i_r}) \cap E(e_{i_{r+1}})$ for $r = 1, \ldots, p-1$ and $v_{i_0} = v_{i_p} = E(e_{i_1}) \cap E(e_{i_p})$. Let $v_{i_r}, v_{i_{r+1}}, \ldots, v_{i_s}$ a sequence of pairwise distinct vertices, with $v_{i_r} = v_{i_s}$ and $r < s$ (this exists since $v_{i_0} = v_{i_p}$). Then $s = r + 1$ otherwise $\left(v_{i_r}, E(e_{i_{r+1}}), v_{i_{r+1}}, \ldots, E(e_{i_s}), v_{i_s}\right)$ would be a cycle of $H(\pi, \tau)$. Let us consider the path $(v_{i_r}, e_{i_r+1}, v_{i_{r+1}}, \ldots, e_{i_s}, v_{i_s})$. All the edges have the same color, and this is a disjoint union of τ-monochromatic cycles.

Let us now apply the same trick to the reduced path

$$(v_{i_0}, E(e_{i_1}), v_{i_1}, \ldots, v_{i_r}, E(e_{i_{s+1}}), \ldots, v_{i_{p-1}}, E(e_{i_p}), v_{i_p}).$$

This allows us to achieve by induction the proof of the proposition.

$\square$

An example of τ such that $H(\pi, \tau)$ has no cycle is given at Figure 7 below (another example of hypergraph with no cycle is given in Figure 10(b) of the appendix).
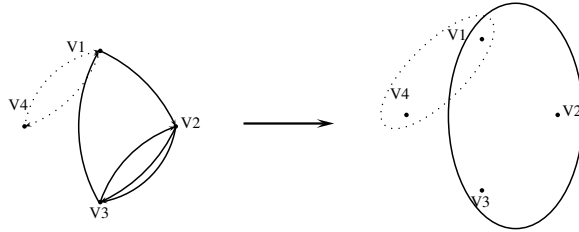

FIGURE 7. On the left: an admissible partition τ of the edges of $G_{/\pi}$ whose corresponding hypergraph $H(\pi, \tau)$ (on the right) has no cycle.

6. Proof of Theorem 2.6

The proof of Theorem 2.6 will follow from Propositions 6.1 and 6.5 and Lemma 6.7 below. We suppose that Hypothesis 2.4 holds.

6.1. **Convergence of the mean spectral distribution in the case with moments.** In this section, we prove the following proposition. The parametrization of infinitely divisible laws via formula $\mu = \nu_*^{\gamma,\sigma}$ is introduced in Section 12.2 of the appendix.

Proposition 6.1. *Consider an infinitely divisible law $\mu = \nu_*^{\gamma,\sigma}$ such that σ has compact support. Define the mean spectral distribution $\bar{\Lambda}_N^{(\mu)}$ of a $\mathbb{P}_{U_N}^{(\mu)}$-distributed random matrix M by*

$$\int g(t)\bar{\Lambda}_N^{(\mu)}(\mathrm{d}t) = \mathbb{E}\left[\frac{1}{N}\operatorname{Tr} g(M)\right] \qquad \text{(for any Borel function } g\text{)}.$$

Then $\bar{\Lambda}_N^{(\mu)}$ converges weakly, as $N \longrightarrow \infty$, to a probability measure $\Lambda_\Gamma(\mu)$ which depends only on μ and on the function Γ introduced in Hypothesis 2.4. This distribution admits exponential moments of any order, hence is characterized by its moments, given by Formula (12) in Theorem 2.6.

Let us introduce a $\mathbb{P}_{U_N}^{(\mu)}$-distributed random matrix M and fix $k \geq 1$. Proposition 6.1 follows directly from Lemmas 6.2, 6.3 and 6.4 below.

Lemma 6.2. *For any $N \geq 1$ and $k \geq 1$, there is a collection $(f_N(\pi))_{\pi\in\mathrm{Part}(k)}$ of numbers depending only on the distribution of U_N such that*

$$\mathbb{E}\left[\frac{1}{N}\operatorname{Tr} M^k\right] = \sum_{\pi\in\mathrm{Part}(k)} f_N(\pi)c_\pi(\mu).$$

Moreover, for any π, $f_N(\pi)$ is given by the formula

$$(21) \qquad f_N(\pi) = N^{|\pi|-1-k}\,\mathbb{E}\left[\prod_{\ell=1}^{k} U_N^{\pi(\ell)*} U_N^{\pi(\ell+1)}\right],$$

where $(U_N^i)_{i\geq 1}$ is a sequence of i.i.d. copies of U_N.

Proof.   For each $n \geq 1$, set $\nu_n = \mu^{*\frac{1}{n}}$ and let $X_n^1, X_n^2, \dots$ be a sequence of i.i.d. ν_n-distributed random variables, independent of the U_N^i's, so that according to Theorem 2.1, for

$$M_n := \frac{1}{N}\sum_{i=1}^{N\times n} X_n^i \cdot U_N^i U_N^{i*},$$

the law of M_n converges weakly to $\mathbb{P}_{U_N}^{(\mu)}$. Then it is immediate that for n large enough,

$$\mathbb{E}\left[\frac{1}{N}\operatorname{Tr} M_n^k\right] = \frac{1}{N^{k+1}} \sum_{\pi \in \operatorname{Part}(k)} \frac{(Nn)!}{(Nn-|\pi|)!} m_\pi(\nu_n) \mathbb{E}\left[\operatorname{Tr} \prod_{\ell=1}^k U_N^{\pi(\ell)} U_N^{\pi(\ell)*}\right],$$

where for any π,

$$m_\pi(\nu_n) := \mathbb{E}\left[\prod_{\ell=1}^k X_n^{\pi(\ell)}\right] = \prod_{J \in \pi} \int t^{|J|} d\nu_n(t).$$

Note that since the U_N^i's are column vectors, $\operatorname{Tr} \prod_{\ell=1}^k U_N^{\pi(\ell)} U_N^{\pi(\ell)*} = \prod_{\ell=1}^k U_N^{\pi(\ell)*} U_N^{\pi(\ell+1)}$. Moreover, by [6, Th. 1.6], $n^{|\pi|} m_\pi(\nu_n) \xrightarrow[n\to\infty]{} c_\pi(\mu)$, which allows to conclude. □

Let us now consider the large N limit of $f_N(\pi)$. Let G be the cyclic graph with vertex set $V = \{1, \ldots, k\}$ and edges $1 \to 2 \to \cdots \to k \to 1$. The hypergraph $H(\pi, \tau)$ is defined in Section 5.3 and the operator thin on partitions is defined in Section 5.1.

Lemma 6.3. *For each $\pi \in \operatorname{Part}(k)$, there is $f_\Gamma(\pi) \geq 0$ such that $f_N(\pi) \xrightarrow[N\to\infty]{} f_\Gamma(\pi)$. The number $f(\pi)$ factorizes along the connected components of π and is given by:*

$$(22) \qquad f_\Gamma(\pi) = \sum_{\substack{\tau \text{ partition of the edges of} \\ G_{/\pi} \text{ s.t. } H(\pi,\tau) \text{ has no cycle}}} \prod_{J \in \pi} \Gamma(n_\tau(J)).$$

If, moreover, U_N is a unitary vector, for any π, $f_\Gamma(\pi) = f_\Gamma(\operatorname{thin}(\pi))$ and for π non crossing $f_\Gamma(\pi) = 1$.

Before proving the lemma, let us give two examples to make the set of partitions τ such that $H(\pi, \tau)$ has no cycle a bit more intuitive.

a) Consider for example the case $k = 4$ and $\pi = \{\{1\}, \{2\}, \{3\}, \{4\}\}$ (trivial partition). Then it can easily be seen that there is only one such τ: it is the trivial partition with only one block.

b) Consider now the partition $\pi = \{\{1\}, \{2, 4\}, \{3\}\}$. Then there are two such τ's: the first one is the one with only one block, containing all vertices; the second one is $\{\{(1, 2), (4, 1)\}, \{(2, 3), (3, 4)\}\}$.

Proof. • Let us first prove that $f_N(\pi)$ has a limit $f_\Gamma(\pi)$ given by Formula (22) as $N \to \infty$. We start with Formula (21), expand the matrix product, and use the independence of the

U_N^i's and the exchangeability of the entries $U_N(1), \ldots, U_N(N)$ of U_N:

$$
\begin{aligned}
f_N(\pi) &= N^{|\pi|-1-k} \, \mathbb{E}\left[\prod_{\ell=1}^{k} U_N^{\pi(\ell)*} U_N^{\pi(\ell+1)}\right] \\
&= N^{|\pi|-1-k} \sum_{1 \leq i_1,\ldots,i_k \leq N} \mathbb{E}\left[\prod_{\ell=1}^{k} \overline{U_N^{\pi(\ell)}(i_\ell)} U_N^{\pi(\ell+1)}(i_\ell)\right] \\
&= N^{|\pi|-1-k} \sum_{1 \leq i_1,\ldots,i_k \leq N} \prod_{J\in\pi} \mathbb{E}\left[\prod_{\ell\in J} \overline{U_N(i_\ell)} \prod_{\ell' \text{ s.t. } \ell'+1\in J} U_N(i_{\ell'})\right] \\
&= N^{|\pi|-1-k} \sum_{\substack{\tau \text{ partition} \\ \text{of the edges of } G_{/\pi}}} A_N^{|\tau|} \prod_{J\in\pi} \mathbb{E}\left[\prod_{\ell\in J} \overline{U_N(i_l^\tau)} \prod_{\ell' \text{ s.t. } \ell'+1\in J} U_N(i_{l'}^\tau)\right],
\end{aligned}
$$

where for all τ, $A_N^{|\tau|} = N(N-1)\cdots(N-|\tau|+1)$ and i^τ is a coloring of the edges of $G_{/\pi}$ with kernel τ.

Now, for each τ, for each J, let $a_\tau(J)$ be the number of blocks of τ containing an edge having J as an extremity. By Formula (9) of Hypothesis 2.4, as $N \to \infty$, the term associated to τ is

$$
O(N^{|\pi|-1+|\tau|-\sum_{J\in\pi} a_\tau(J)}),
$$

and that in the particular case where τ is admissible, this term is equivalent to

$$
N^{|\pi|+|\tau|-1-\sum_{J\in\pi} a_\tau(J)} \prod_{J\in\pi} \Gamma(n_\tau(J)).
$$

Let us consider the hypergraph $H(\pi,\tau)$ defined at Section 5.3, with edges $(E_W, W \in \tau)$. Notice that each edge of $G_{/\pi}$ is included in one edge of $H(\pi,\tau)$. This implies that $H(\pi,\tau)$ is connected. Notice also that:

$$
\sum_{J\in\pi} a_\tau(J) = \sum_{\substack{J\in\pi \\ W\in\tau}} 1_{J\in E_W} = \sum_{W\in\tau} |E_W|
$$

Hence, using Property 12.5, we know that

$$
|\pi| + |\tau| - 1 - \sum_{J\in\pi} a(J) = |\pi| + |\tau| - 1 - \sum_{W\in\tau} |E_W|
$$

is non positive, and vanishes if and only if $H(\pi,\tau)$ has no cycle. By Proposition 5.1, this proves that $f_N(\pi)$ converges when N goes to infinity, with a limit given by (22).

• Let us now prove that $f_\Gamma(\pi)$ factorizes along the connected components of π.

Suppose that π has at least two connected components, and let G be the cyclic graph on $\{1,\ldots,k\}$ introduced in the statement of the previous lemma and $G_{/\pi} = (\pi, E)$ be the associated quotient graph. There exists at least one connected component $\pi_{|A}$ of π whose

support A is an interval $[l, l']$. The corresponding subgraph of $G_{/\pi}$ is connected to the rest of the graph through two directed edges, $\varepsilon = (V_1, V_2)$ and $\varepsilon' = (V_3, V_4)$ with $l - 1 \in V_1$, $l \in V_2$, $l' \in V_3$ and $l' + 1 \in V_4$ (with the usual identifications $0 = k$ and $1 = k + 1$).

For instance, if we still consider the partition $\pi = \{\{1, 8, 10\}, \{2, 4\}, \{3, 5\}, \{6, 7, 9\}\} \in$ Part(10), then the restriction of π to the interval $\{2, 3, 4, 5\}$ is a connected component of π, and the corresponding subgraph of $G_{/\pi}$ (*cf.* Figure 5 page 16) is its restriction to the set of vertices $\{V2, V3\}$, and is connected to the rest of the graph by the edges $\varepsilon = (V1, V2)$ and $\varepsilon' = (V3, V4)$:
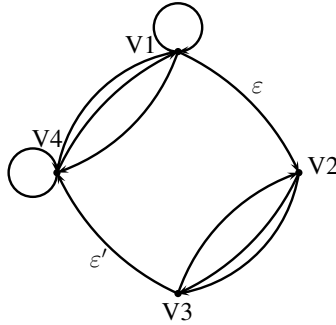

FIGURE 8. The quotient graph $G_{/\pi}$, with the edges ε and ε'

Let us define the graph $T\left(G_{/\pi}\right)$ obtained from $G_{/\pi}$ by replacing the edges ε and ε' by the edges $\varepsilon_0 = (V_3, V_2)$ and $\varepsilon_1 = (V_1, V_4)$. This graph has two connected components, one on the subset of vertices $\pi_{|A}$, the other on the complementary $\pi_{|A^c}$. The first (resp. second) one, that we denote by G_1 (resp. G_0), is the restriction of $G_{/\pi}$ to A (resp. A^c), plus the edge ε_1 (resp. ε_0).

For instance, $T(G_{/\pi})$ is drawn for the preceding example in Figure 9(a).

Let us consider a partition τ of the edges of $G_{/\pi}$ such that $H(\pi, \tau)$ has no cycle. Since, by Proposition 5.1, $G_{/\pi}$ is a union of pairwise disjoint τ-monochromatic cycles, ε and ε' belong to the same block of τ. Let us define the partition $T(\tau)$ of $T\left(G_{/\pi}\right)$ deduced from τ by replacing ε and ε' by ε_0 and ε_1 in their block of τ (see Figure 9(b)). It is easy to see that T is a bijection between the set of partitions τ of the edges of $G_{/\pi}$ such that $H(\pi, \tau)$ has no cycle and the set of pairs of such partitions of the two connected components of $T(G_{/\pi})$. Moreover, $n_\tau(J) = n_{T(\tau)}(J)$ for any $J \in \pi$. The conclusion follows directly.

• Under the additional hypothesis that U_N is a unitary vector, it follows directly from Formula (21) that $f_N(\pi) = f_N(\text{thin}(\pi))$, hence $f_\Gamma(\pi) = f_\Gamma(\text{thin}(\pi))$. It also follows from Formula (21) that for $k = 1$, $f_N(\{\{1\}\}) = 1$. If π is non crossing, then the partition induced by π on its connected components are all some one-bloc partitions. Hence the fact that f_Γ factorizes on its connected components and the identities $f_\Gamma(\pi) = f_\Gamma(\text{thin}(\pi))$ and $f_\Gamma(\{\{1\}\}) = 1$ iterated imply that $f_\Gamma(\pi) = 1$. $\qquad\square$

(a) The graph $T(G_{/\pi})$ (b) The coloring $T(\tau)$
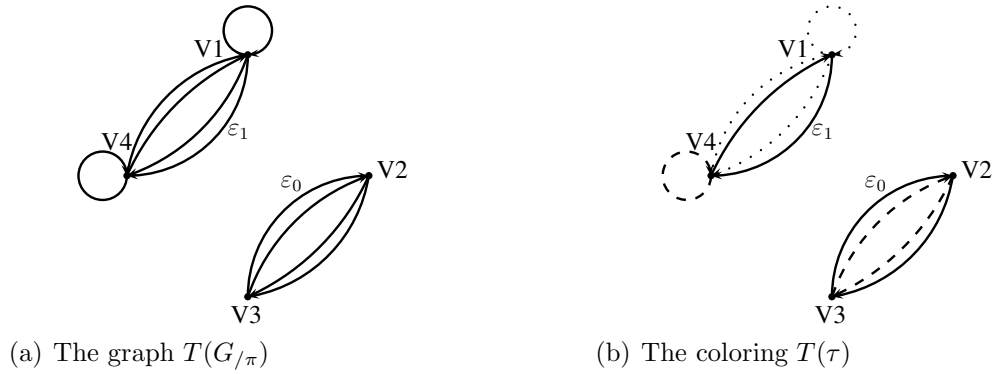
FIGURE 9. The graph $T(G_{/\pi})$ and the coloring $T(\tau)$

Lemma 6.4. *Let us define $m_0 = 1$ and, for each $k \geq 1$,*

$$m_k = \sum_{\pi \in \mathrm{Part}(k)} f_\Gamma(\pi) c_\pi(\mu).$$

Then the radius of convergence of the series $\sum_{k \geq 0} \frac{m_k}{k!} z^k$ is infinite.

Proof. Claim: it suffices to prove the result under the additional hypothesis that $\Gamma(\cdot) = 1$.

Let us prove the claim. Recall that γ, σ are such that $\mu = \nu_*^{\gamma, \sigma}$. Let $\tilde{\sigma}$ be the push-forward of the measure σ by the map $t \mapsto |t|$ and define $\tilde{\mu} := \nu_*^{|\gamma|, \tilde{\sigma}}$. Note that by Formula (50) of the appendix, for any k, $|c_k(\mu)| \leq c_k(\tilde{\mu})$. Let us define, the function $\tilde{\Gamma}(\cdot)$ by $\tilde{\Gamma}(n_1, \ldots, n_k) := 1$ for any integers $n_1, \ldots, n_k$. For C the constant of Equation (11), we have, for any $k \geq 1$ and any $\pi \in \mathrm{Part}(k)$,

$$0 \leq f_\Gamma(\pi) \leq C^k f_{\tilde{\Gamma}}(\pi).$$

As a consequence, if one defines, for each $k \geq 0$,

$$\tilde{m}_k := \sum_{\pi \in \mathrm{Part}(k)} f_{\tilde{\Gamma}}(\pi) c_\pi(\tilde{\mu}),$$

we have $|m_k| \leq C^k \tilde{m}_k$, and it suffices to prove the lemma for $\tilde{m}_k$ instead of m_k. We have proved the claim.

So from now on, we suppose that $\Gamma(\cdot) = 1$. It allows us to choose a particular model for the random column vector U_N: we choose U_N such that its entries $U_N(1), \ldots, U_N(N)$ are i.i.d. random variables whose law is the one of $\sqrt{N} B e^{i\Theta}$, where B has law $(1 - 1/N)\delta_0 + (1/N)\delta_1$, and is independent of Θ, who has uniform law on $[0, 2\pi]$. This vector obviously satisfies Hypothesis 2.4 with $\Gamma(\cdot) = 1$.

Let us define the function $g_\theta(A) := \frac{1}{N} \mathrm{Tr} \, e^{\theta A}$ on Hermitian matrices. To prove the lemma, it suffices to prove that for any $\theta \in \mathbb{R}$, the sequence $\mathbb{P}_{U_N}^{(\mu)}(g_\theta)$ stays bounded as $N \to \infty$,

which, by definition of the law $\mathbb{P}_{U_N}^{(\mu)}$, will be implied by the fact that

$$\limsup_{N\to\infty} \limsup_{n\to\infty} \frac{1}{N} \mathbb{E}[\operatorname{Tr} e^{\frac{\theta}{N}\sum_{i=1}^{N\times n} X_n^i \cdot U_N^i U_N^{i*}}] < \infty,$$

where for each n, the X_n^i's are i.i.d. with law $\mu^{*\frac{1}{n}}$, independent of the U_N^i's, some i.i.d. copies of U_N.

The Golden-Thompson inequality (*cf.* [30, Sect. 3.2]) states that for A_1, A_2 Hermitian matrices, $\operatorname{Tr} e^{A_1+A_2} \leq \operatorname{Tr} e^{A_1} e^{A_2}$. In the case where $A_1, \ldots, A_k$ are random, independent, such that for all i, $\mathbb{E}[e^{A_i}] = m_i I_N$ with $m_i \geq 0$, we can deduce, by induction on k, that

$$(23) \qquad \frac{1}{N}\mathbb{E}[\operatorname{Tr} e^{A_1+\cdots+A_k}] \leq m_1 \cdots m_k.$$

Let us now compute $\mathbb{E}[e^{\frac{\theta}{N}X_n \cdot U_N U_N^*}]$ for X_n a $\mu^{*\frac{1}{n}}$-distributed variable independent of U_N. Note first that since the Lévy measure σ of μ has compact support, by for instance [28, Lem. 25.6], for all n, the Laplace transform of $\mu^{*\frac{1}{n}}$ is defined as an entire function on $\mathbb{C}$ and given by the fonction $e^{\frac{1}{n}\phi(\cdot)}$, where

$$\phi(\lambda) = \lambda\gamma + \int \underbrace{(e^{\lambda t} - 1 - \frac{\lambda t}{1+t^2})\frac{1+t^2}{t^2}}_{:=\frac{\lambda^2}{2}\text{ for } t=0} \sigma(\mathrm{d}t).$$

Since U_N is a column vector, we have

$$(24) \qquad e^{\frac{\theta}{N}X_n \cdot U_N U_N^*} = I_N + \frac{e^{\frac{\theta}{N}X_n\|U_N\|^2} - 1}{\|U_N\|^2} U_N U_N^*,$$

with the convention $(e^t - 1)/t = 1$ for $t = 0$. For any diagonal matrix D with diagonal entries on the unit circle, DU_N has the same law as U_N, hence the expectation of the right-hand term of (24) is a diagonal matrix. Moreover, by exchangeability, for any f,

$$\mathbb{E}[f(\|U_N\|^2)|U_N(k)|^2] = \frac{1}{N}\mathbb{E}[f(\|U_N\|^2)\|U_N\|^2] \qquad (1 \leq k \leq N),$$

so

$$\mathbb{E}[e^{\frac{\theta}{N}X_n \cdot U_N U_N^*}] = I_N + \frac{1}{N}\mathbb{E}[e^{\frac{\theta}{N}X_n\|U_N\|^2} - 1]I_N = I_N + \frac{n\mathbb{E}[e^{\frac{1}{n}\phi(\theta\|U_N\|^2/N)} - 1]}{Nn}I_N.$$

Hence by (23), we have

$$\frac{1}{N}\mathbb{E}[\operatorname{Tr} e^{\frac{\theta}{N}\sum_{i=1}^{N\times n} X_n^i \cdot U_N^i U_N^{i*}}] \leq \left(1 + \frac{n\mathbb{E}[e^{\frac{1}{n}\phi(\|U_N\|^2)} - 1]}{Nn}\right)^{Nn}.$$

As $n \to \infty$ with N fixed, the function $n(e^{\frac{1}{n}\phi(\cdot)} - 1)$ converges weakly to $\phi(\cdot)$. Moreover, the random variable $\|U_N\|$ takes a finite set of values, hence $n\mathbb{E}[e^{\frac{1}{n}\phi(\|U_N\|^2)} - 1] \to \mathbb{E}[\phi(\|U_N\|^2)]$,

so that

$$\limsup_{n\to\infty} \frac{1}{N}\mathbb{E}[\operatorname{Tr} e^{\frac{\theta}{N}\sum_{i=1}^{N\times n} X_n^i \cdot U_N^i U_N^{i*}}] \leq e^{\mathbb{E}[\phi(\theta\|U_N\|^2/N)]}.$$

To conclude, it suffices to verify that $\mathbb{E}[\phi(\theta\|U_N\|^2/N)]$ stays bounded as $N \to \infty$. It follows from the fact that the law of $\|U_N\|^2/N$ is the one of $\sum_{i=1}^{N} B_i$, where the B_i's are i.i.d. with law $(1 - 1/N)\delta_0 + (1/N)\delta_1$. $\qquad\square$

6.2. Convergence of the mean spectral distribution: extension to the case without moments and continuity of the limit.

Proposition 6.5. *a) For any $*$-infinitely divisible law μ, the mean spectral distribution $\bar{\Lambda}_N^{(\mu)}$ of a $\mathbb{P}_{U_N}^{(\mu)}$-distributed random matrix M, defined by*

$$\int g(t)\bar{\Lambda}_N^{(\mu)}(\mathrm{d}t) = \mathbb{E}\left[\frac{1}{N}\operatorname{Tr} g(M)\right] \qquad\qquad \text{(for any function } g\text{)},$$

converges weakly, as $N \longrightarrow \infty$, to a probability measure $\Lambda_\Gamma(\mu)$ which depends only on μ and on the function Γ introduced at Equation (10).

b) The limit measure $\Lambda_\Gamma(\mu)$ depends continuously on the pair (μ, Γ) for the weak topology.

Proof. a) Let (γ, σ) be the Lévy pair of μ. Note that if σ is compactly supported, the result has been established in Proposition 6.1, so we shall prove the result thanks to Lemma 12.2.

Let $\varepsilon > 0$ and $M > 1$ such that $\lambda_\infty := \sigma(\mathbb{R}\backslash[-M, M]) < \varepsilon$. Let σ_ε (resp. σ_∞) the restriction of σ to $[-M, M]$ (resp. $\mathbb{R}\backslash[-M, M]$). Let μ_∞ be the compound Poisson law with Fourier transform:

$$\forall\theta \in \mathbb{R}, \quad \int e^{i\theta t}\mathrm{d}\mu_\infty(t) = \exp\left(\int_{\mathbb{R}} \left(e^{i\theta x} - 1\right)\frac{1 + x^2}{x^2}\sigma_\infty(dx)\right)$$

The Lévy pair of μ_∞ is $(\gamma_\infty, \sigma_\infty)$ for some $\gamma_\infty \in \mathbb{R}$. Let $\gamma_\varepsilon = \gamma - \gamma_\infty \in \mathbb{R}$, and let μ_ε the infinitely divisible law with Lévy pair $(\gamma_\varepsilon, \sigma_\varepsilon)$. The law μ_∞ is a compound Poisson law whose underlying Poisson variable has a parameter less than $2\lambda_\infty$, so by construction of the law $\mathbb{P}_{U_N}^{(\mu)}$ for μ a compound Poisson given at Section 2.1.1, the expectation of the rank of a $\mathbb{P}_{U_N}^{(\mu_\infty)}$-distributed matrix is not larger than $2\varepsilon N$.

Since $\mu = \mu_\varepsilon * \mu_\infty$, we have $\mathbb{P}_{U_N}^{(\mu)} = \mathbb{P}_{U_N}^{(\mu_\varepsilon)} * \mathbb{P}_{U_N}^{(\mu_\infty)}$ by (7). The mean spectral law of a $\mathbb{P}_{U_N}^{(\mu_\varepsilon)}$-distributed matrix converges for any ε, so the result holds by application of Lemma 12.2.

b) Let μ_n be a sequence of infinitely divisible probability measures that converges to a probability measure μ (by [28, Lem. 7.8], we know that μ is infinitely divisible). It is

known (see [19, Th. 1, §19]) that if (γ_n, σ_n) denotes the Lévy pair of μ_n and (γ, σ) the one of μ, the weak convergence of μ_n to μ is equivalent to the fact that

(25) $\qquad \gamma_n \longrightarrow \gamma$ and $\sigma_n(g) \longrightarrow \sigma(g)$ for any continuous bounded function g.

Let us also consider, for each $n \geq 1$, a sequence $(U_N^{(n)})_{N \geq 1}$ of random column vectors which satisfies Hypothesis 2.4. We suppose that there is another sequence $(U_N)_{N \geq 1}$ of random column vectors which satisfies Hypothesis 2.4 such that the function $\Gamma^{(n)}$ associated to $(U_N^{(n)})_{N \geq 1}$ converges pointwise to the function Γ associated to the sequence $(U_N)_{N \geq 1}$.

Let us now prove that the sequence $\Lambda_{\Gamma^{(n)}}(\mu_n)$ converges weakly to $\Lambda_\Gamma(\mu)$.

Let us first suppose that there exists $M > 0$ such that

$$\forall n \geq 1, \quad \sigma_n([-M, M]^c) = 0.$$

Then, since the cumulants of μ_n are more or less the moments of σ_n (see Formula (50) for more details), we have convergence of the cumulants : for all k, $c_k(\mu_n) \underset{n \to \infty}{\longrightarrow} c_k(\mu)$. Hence by Formula (12), the moments of $\Lambda_{\Gamma^{(n)}}(\mu_n)$ tend to the ones of $\Lambda_\Gamma(\mu)$. It implies the weak convergence, since $\Lambda_\Gamma(\mu)$ is determined by its moments, as stated in Proposition 6.1.

Let us now deal with the general case. It suffices to prove that the Cauchy transform of $\Lambda_{\Gamma^{(n)}}(\mu_n)$ converges pointwise to the one of $\Lambda_\Gamma(\mu)$. So let us fix $z \in \mathbb{C}^+$ and $\varepsilon > 0$. Define the function $g(t) := \frac{1}{z-t}$. As in the proof of a), one can find M such that $\sigma(\mathbb{R} \backslash [-M, M]) \leq \varepsilon$ and for all n, $\sigma_n(\mathbb{R} \backslash [-M, M]) \leq \varepsilon$. Hence with some decompositions $\mu = \mu_\varepsilon * \mu_\infty$ and $\mu_n = \mu_{n,\varepsilon} * \mu_{\infty,\varepsilon}$ as above with μ_ε and $\mu_{n,\varepsilon}$ having some Lévy measures supported on $[-M, M]$ and μ_∞ and $\mu_{n,\infty}$ some compound Poisson measures, we have

(26) $\qquad |\Lambda_\Gamma(\mu)(g) - \Lambda_\Gamma(\mu_\varepsilon)(g)| \leq \frac{4\varepsilon}{\Im z}, \qquad |\Lambda_\Gamma(\mu_n)(g) - \Lambda_\Gamma(\mu_{n,\varepsilon})(g)| \leq \frac{4\varepsilon}{\Im z}.$

This is a consequence of the resolvent identity (see Equation (46), Lemma 12.2 in the Appendix). Moreover, by what (25), $\mu_{n,\varepsilon}$ converges weakly to μ_ε as $n \to \infty$. So by what precedes about the compact case, $\Lambda_{\Gamma^{(n)}}(\mu_{n,\varepsilon})$ converges weakly to $\Lambda_\Gamma(\mu_\varepsilon)$. Joining this to (26), we get the desired result. $\qquad \square$

6.3. **Concentration of measure and almost sure convergence.** To conclude the proof of Theorem 2.6, we still need to prove that that not only the *mean spectral law* of a $\mathbb{P}_{U_N}^{(\mu)}$-distributed random matrix M converges to $\Lambda_\Gamma(\mu)$ as $N \longrightarrow \infty$, but also, almost surely, the *empirical spectral law*. By the Borel-Cantelli lemma, it follows directly from what precedes and the following lemma.

Remark 6.6 (The associated Lévy process). There exists a Lévy process $(M_s)_{s \geq 0}$ with values in the space of $N \times N$ Hermitian matrices such that M_s has distribution $\mathbb{P}_{U_N}^{(\mu^{*s})} \equiv (\mathbb{P}_{U_N}^{(\mu)})^{*s}$ for every $s \geq 0$ (this is due to the fact that the distribution $\mathbb{P}_{U_N}^{(\mu)}$ is infinitely

divisible, see *e.g.* [28, Cor. 11.6]). Let $\left(P_s^{(N,\mu)}, s \geq 0 \right)$ be the corresponding semi-group: for every Hermitian matrix A, every bounded function f:

$$P_s^{(N,\mu)}(f)(A) = \mathbb{E}\left[f\left(M_s + A \right) \right].$$

Let $\mathcal{A}^{(N,\mu)}$ the infinitesimal generator associated to this semi-group. Its domain contains the set of twice continuously differentiable functions on $N \times N$ Hermitian matrices vanishing at infinity and if (γ, σ) denotes the Lévy pair of μ (*cf.* Proposition-Definition 12.3), for any Hermitian matrix A, we have, by [28, Th. 31.5],

$$
\begin{aligned}
\mathcal{A}^{(N,\mu)}(f)(A) \;\; = \;\; & \gamma \mathrm{d}f(A)\left[D_N \right] \\
& + \int \left\{ N \left[\mathbb{E}\left[f\left(A + x U_N U_N^* \right) \right] - f(A) \right] - \frac{x}{1+x^2} \mathrm{d}f(A)\left[D_N \right] \right\} \frac{1+x^2}{x^2} \sigma(\mathrm{d}x),
\end{aligned}
$$

where $\mathrm{d}f$ denotes the derivative of f and $D_N = \mathbb{E}\left[U_N U_N^* \right]$. Note that for $f : \mathbb{R} \to \mathbb{R}$ extended to Hermitian matrices by spectral calculus, we have $\mathrm{d}f(A)\left[I_N \right] = f'(A)$.

Lemma 6.7. *Consider μ an infinitely divisible law and $g : \mathbb{R} \to \mathbb{R}$ a Lipschitz function with finite total variation. Then for every $\varepsilon > 0$, there exists $\delta > 0$ such that for all $N \geq 1$,*

$$\mathbb{P}\left(\left| \frac{1}{N} \operatorname{Tr} g(M) - \mathbb{E}\left[\frac{1}{N} \operatorname{Tr} g(M) \right] \right| > \varepsilon \right) \leq 2 e^{-N\delta}$$

with M a $\mathbb{P}_{U_N}^{(\mu)}$-distributed random matrix.

Proof.   This is an extension of Theorem III.4 in [16] which is established for a special case of U_N. Its proof only uses the fact that $\| U_N / \sqrt{N} \|_2 = 1$ (one has to notice that in the present paper, what plays the role of U_N in [16] is $U_N / \sqrt{N}$). Theorem III.4 in [16] can be readily extended to the case when $\mathbb{E}\left[\| U_N / \sqrt{N} \|_2^4 \right]$ is bounded w.r.t. N. Then, we conclude noticing that by exchangeability and Equation (10), $\mathbb{E}\left[\| U_N / \sqrt{N} \|_2^4 \right]$ converges to $\Gamma(2) + \Gamma(1,1)$. Indeed,

$$
\begin{aligned}
\mathbb{E}\left[\| U_N / \sqrt{N} \|_2^4 \right] \;\; = \;\; & N^{-2} \sum_i \mathbb{E}[|U_N(i)|^4] + N^{-2} \sum_{i \neq j} \mathbb{E}[|U_N(i)|^2 |U_N(j)|^2] \\
= \;\; & N^{-1} \mathbb{E}[|U_N(1)|^4] + (1 - N^{-1}) \mathbb{E}[|U_N(1)|^2 |U_N(2)|^2]
\end{aligned}
$$

$\square$

7. Proof of Proposition 2.7

For each $n \geq 1$, let m_n be the nth moment of $\Lambda_\Gamma(\mu)$. Let $c = \inf \Gamma(n)^{1/n}$. It suffices to prove that there is $\varepsilon > 0$ such that for all n even,

$$m_n \geq \varepsilon^n \times n^{th} \text{ moment of a standard Gaussian variable,}$$

i.e. that for all n even, $m_n \geq \varepsilon^n |\operatorname{Part}_2(n)|$, where $\operatorname{Part}_2(n)$ is the number of pairings of $\{1, \ldots, n\}$. The formula of m_n is given by (12), where each term is positive. Moreover, we know that $c_2(\mu) = \operatorname{Var}(\mu) > 0$. Hence it suffices to notice that for any $k \geq 1$ and any $\pi \in \operatorname{Part}_2(k)$, $f_\Gamma(\pi) \geq c^k$, which follows from the expression of $f_\Gamma(\pi)$ as a sum of positive terms in (22), where the term associated to the trivial partition τ with one block is

$$\prod_{J \in \pi} \Gamma(|J|) \geq \prod_{J \in \pi} c^{|J|} = c^k.$$

8. Proof of Theorem 3.1

First, in the case where $\alpha > 1$, we can suppose the Y_{ij}'s to be centered. Indeed, replacing the Y_{ij}'s by $Y_{ij} - \mathbb{E}[Y_{ij}]$ is a rank-two perturbation of $M_{N,p}$ and by Lemma 12.2, a rank-two perturbation of an Hermitian matrix has no influence on the weak convergence of its spectral law.

Let $(P(t), t \geq 0)$ be a standard Poisson process, independent of the other variables. For each $B > 0$, let us introduce both following approximations of the random matrix $M_{N,p}$:

$$\widehat{M}_{N,p} := \frac{1}{a_N^2} \sum_{j=1}^{p} X_j \cdot \widehat{V}^j (\widehat{V}^j)^* \qquad \text{and} \qquad \widetilde{M}_N := \frac{1}{a_N^2} \sum_{j=1}^{P(N\lambda)} X_i \cdot \widehat{V}^j (\widehat{V}^j)^*$$

with $\widehat{V}^j := (Y_{ij} 1_{|Y_{ij}| \leq Ba_N})_{j=1}^{N} \in \mathbb{K}^{N \times 1}$ (this column vector depends implicitly on N and on the cutoff parameter B). It is noticed in Sections 1 and 9 of [5] that for $Z_N^{(B)} := \frac{\sqrt{N}}{a_N} Y_{11} 1_{|Y_{11}| \leq Ba_N}$, we have, for each $n \geq 1$,

$$\frac{\mathbb{E}[|Z_N^{(B)}|^{2n}]}{N^{n-1}} \xrightarrow[N \to \infty]{} \frac{\alpha}{2n - \alpha} B^{2n - \alpha}$$

and $\sqrt{N} \mathbb{E}[Z_N^{(B)}] = O(1)$ (this is where the recentering is necessary when $\alpha > 1$). So Hypothesis 2.4 holds for U_N distributed as the $\widehat{V}^j$'s, with

$$(27) \qquad \Gamma(n_1, \ldots, n_k) = \Gamma^{(B)}(n_1, \ldots, n_k) := B^{2(n_1 + \cdots + n_k) - k\alpha} \prod_{\ell=1}^{k} \frac{\alpha}{2n_\ell - \alpha}.$$

As a consequence, by Theorem 2.6, the empirical spectral law of $\widetilde{M}_N$ converges almost surely, as $N \to \infty$, to the law $\Lambda_{\Gamma^{(B)}}(\mu)$, with μ the (compound Poisson) law of $\sum_{j=1}^{P(\lambda)} X_j$. By Lemma 12.2 (applied with $M_N = \widehat{M}_{N,p}$ and and $M_N^\varepsilon = \widetilde{M}_N$ for all $\varepsilon > 0$), the same holds for $\widehat{M}_{N,p}$, because by the Law of Large Numbers, as $N, p \to \infty$, with $p/N \to \lambda$,

$$\frac{\operatorname{rank}(\widetilde{M}_N - \widehat{M}_{N,p})}{N} \longrightarrow 0 \qquad \text{(almost surely)}.$$

To prove the convergence of the empirical spectral law of $M_{N,p}$, by Lemma 12.2, it suffices to prove that for any $\varepsilon > 0$, if B is large enough, then almost surely, for N, p large enough,

$$(28) \qquad\qquad \operatorname{rank}(M_{N,p} - \widehat{M}_{N,p}) \leq N\varepsilon.$$

But we have

$$\operatorname{rank}(M_{N,p} - \widehat{M}_{N,p}) \leq \sum_{j=1}^{p} 2\mathbb{1}_{V^j \neq \widehat{V}^j}.$$

The above right-hand-term is a sum of p independent Bernoulli variables with parameter

$$q_N := 1 - (\mathbb{P}(|Y_{11}| \leq Ba_N))^N.$$

We have, for L the slow variations function introduced at (13),

$$\mathbb{P}(|Y_{11}| \leq Ba_N) = 1 - B^{-\alpha}a_N^{-\alpha}L(Ba_N) = 1 - B^{-\alpha}a_N^{-\alpha}L(a_N) \times \frac{L(Ba_N)}{L(a_N)} = 1 - \frac{B^{-\alpha}(1 + \epsilon_N)}{N},$$

where ϵ_N is a sequence tending to zero. As a consequence, as $N \to \infty$, q_N tends to $1 - e^{-B^{-\alpha}}$. By some Laplace transform estimates like p. 722 of [5], there is a constant c such that

$$\mathbb{P}(\operatorname{rank}(M_{N,p} - \widehat{M}_{N,p}) \geq 4pq_N) \leq e^{-cpq_N},$$

hence by Borel-Cantelli's lemma, almost surely, for N, p large enough (N and p grow together in such a way that $p/N \to \lambda$), $\operatorname{rank}(M_{N,p} - \widehat{M}_{N,p}) \leq 4pq_N$. But $1 - e^{-B^{-\alpha}} \sim B^{-\alpha}$ as $B \to \infty$, so for any fixed ε, one can choose B such that almost surely, for N, p large enough, (28) holds. It concludes the proof of the convergence.

Let us now denote by ν the law of the X_i's and by $\Lambda_\alpha(\nu, \lambda)$ the limit of the empirical spectral law of $M_{N,p}$. We shall prove that this is a continuous function of (ν, α, λ). By the formula given for $\Gamma^{(B)}$ at (27), for any fixed cutoff B, the limit spectral law of $\widehat{M}_{N,p}$ depends continuously on the parameters (ν, α, λ). Since the cutoff necessary to obtain a right-hand-term $N\varepsilon$ in (28) can be chosen uniformly on α and λ as soon as they vary in compact sets which do not contain zero (regardless to the law of the X_j's), this proves the continuity.

9. Proof of Proposition 3.3

We first treat (together) the Gaussian and uniform cases. Exchangeability is obvious. Let us now prove that for any diagonal matrix $D = \operatorname{diag}(\varepsilon_1, \ldots, \varepsilon_N)$ with diagonal entries in $\{z \in \mathbb{K} \,;\, |z| = 1\}$,

$$(29) \qquad\qquad U_N U_N^* \stackrel{\text{law}}{=} D U_N U_N^* D^*.$$

It will imply that the expectation of (9) is null as soon as the condition on multisets given at Remark 2.5 is satisfied. Let $\{e_1, \ldots, e_N\}$ denote the canonical basis of $\mathbb{C}^{N \times 1}$. For each $\ell = 1, \ldots, N$, let us denote by $\mathcal{L}_\ell$ the law of U_N conditionally to the event $X_N = e_\ell$ (X_N

is the random variable with uniform law on $\{e_1, \ldots, e_N\}$ that was introduced after (14)). We know that the law $\mathcal{L}$ of U_N is given by

$$\text{(30)} \qquad \mathcal{L}(B) = \frac{1}{N} \sum_{\ell=1}^{N} \mathcal{L}_\ell(B)$$

for any Borel set $B \subset \mathbb{K}^N$. Hence it suffices to prove that for all ℓ, an $\mathcal{L}_\ell$-distributed vector V satisfies (29). So let us fix $\ell \in \{1, \ldots, N\}$, consider such a vector $V = (V(1), \ldots, V(N))^T$ and such a matrix $D = \text{diag}(\varepsilon_1, \ldots, \varepsilon_N)$. Note that the entries of V are independent and that all of them except the ℓ-th one are invariant, in law, by multiplication by any ε_i. Hence for $D' = \varepsilon_\ell^{-1} I_N$, we have

$$(D'D)V \overset{\text{law}}{=} V.$$

As a consequence, we have

$$(D'D)VV^*(D'D)^* \overset{\text{law}}{=} VV^*.$$

But we also have $(D'D)VV^*(D'D)^* = DVV^*D^*$, hence (29) is proved. It remains to prove (10) and more precisely Formulas (16), (17) and (18). So let us fix $k \geq 1$, $n_1, \ldots, n_k \geq 1$ and find out the limit, as $N \to \infty$, of

$$N^k \mathbb{E}[|U_N(1)|^{2n_1} \cdots |U_N(k)|^{2n_k}].$$

Note that by (30), we have

$$\text{(31)} \qquad N^{k-(n_1+\cdots+n_k)} \mathbb{E}[|U_N(1)|^{2n_1} \cdots |U_N(k)|^{2n_k}] =$$

$$\text{(32)} \qquad \left(1 - \tfrac{k}{N}\right) N^{k-(n_1+\cdots+n_k)} (1-e^{-t})^{n_1+\cdots+n_k} \prod_{j=1}^{k} \mathbb{E}[|g_1|^{2n_j}] +$$

$$\text{(33)} \quad \sum_{j_0=1}^{k} N^{k-(n_1+\cdots+n_k)+n_{j_0}-1} (1-e^{-t})^{n_1+\cdots+n_k-n_{j_0}} \mathbb{E}[|e^{-\frac{t}{2}} + \tfrac{g_1}{\sqrt{N}}|^{2n_{j_0}}] \prod_{\substack{j=1 \\ j \neq j_0}}^{k} \mathbb{E}[|g_1|^{2n_j}].$$

First, as soon as one of the n_j's is ≥ 2, the term of (32) vanishes as $N \to \infty$. The same happens for each of the terms of the sum (33) if two of the n_j's are ≥ 2. Equation (10) follows easily.

Let us now consider the case where $U_N = U_t$, for $(U_t/\sqrt{N})_{t \geq 0}$ a solution of (15) whose initial law is the uniform law on the canonical basis. Let us introduce the $N \times N$ matrix P_t (the dependence on N is implicit in P_t and in U_t)

$$P_t := U_t U_t^*/N.$$

By a direct application of the matricial Itô calculus (see [9, Sect. 2.1]), the process $(P_t)_{t \geq 0}$ satisfies the SDE

$$\text{(34)} \qquad dP_t = (dK_t)P_t - P_t(dK_t) + \left(\frac{1}{N}I - P_t\right) ds.$$

First, from this SDE, it follows easily that $\text{Tr}\, P_t = 1$ for all t, so that $\|U_t\|^2 = N$. Second, it is easy to see that the law of K is invariant under conjugation by any unitary matrix. By uniqueness, in law, of the solutions of (15) and (34), given the initial conditions, it follows that for any permutation matrix Q and any matrix D as in (29), $QU_t \overset{\text{law}}{=} U_t$ and

$DU_t D^* \overset{\text{law}}{=} P_t$. It remains to prove (10) and Formulas (16), (17) and (18). For each $n_1, \ldots, n_k \geq 0$, let us define

$$\tilde{\Gamma}_t(n_1, \ldots, n_k) := N(N-1)\cdots(N-p+1)\mathbb{E}\left[|U_t(1)|^{2n_1}\cdot\cdots\cdot|U_t(k)|^{2n_k}\right] N^{-(n_1+\cdots+n_k)},$$

where $p = |\{\ell\,;\, n_\ell \neq 0\}|$ (the dependence of $\tilde{\Gamma}_t$ on N is implicit). We have to prove that

$$\tilde{\Gamma}_t(n_1, \ldots, n_k) \underset{N\to\infty}{\longrightarrow} \Gamma_t(n_1, \ldots, n_k).$$

Due to exchangeability and to $\|U_t\| = N$, we have $\tilde{\Gamma}_t(1) = 1$ and
(35)
$$\tilde{\Gamma}_t(n_1, \ldots, n_k) = \tilde{\Gamma}_t(n_1+1, n_2, \ldots, n_k) + \cdots + \tilde{\Gamma}_t(n_1, \ldots, n_{k-1}, n_k+1) + \tilde{\Gamma}_t(n_1, \ldots, n_k, 1),$$

and since all $|U_t(i)|$'s are ≤ 1, if one considers two families $n_1, \ldots, n_k \geq 1$ and $n'_1, \ldots, n'_k \geq 1$ such that $n_j \leq n'_j$ for all j, then

$$(36) \qquad\qquad \tilde{\Gamma}_t(n_1, \ldots, n_k) \geq \tilde{\Gamma}_t(n'_1, \ldots, n'_k).$$

It is easy to see that thanks to (35) and (36) and to the obvious formula $\tilde{\Gamma}_t(1) = 1$, it suffices to prove that

$$(37) \qquad \tilde{\Gamma}_t(2, 2) \underset{N\to\infty}{\longrightarrow} 0 \qquad \text{and} \qquad \tilde{\Gamma}_t(n) \underset{N\to\infty}{\longrightarrow} e^{-nt} \qquad \text{for all } n \geq 2.$$

For any multiset $\{k_1, \ldots, k_n\}$ of integers in $\{1, \ldots, N\}$, let us define

$$\gamma_t(k_1, \ldots, k_n) = N^{-n}|U_t(k_1)|^2 \cdots |U_t(k_n)|^2$$

(the dependence of $\gamma_t(\cdots)$ in N is implicit). By exchangeability, $\mathbb{E}\left[\gamma_t(k_1, \ldots, k_n)\right]$ only depends on the level sets partition of the map $j \mapsto k_j$. To prove (37), we have to prove

$$(38) \qquad\qquad N^2\mathbb{E}\left[\gamma_t(1, 1, 2, 2)\right] \underset{N\to\infty}{\longrightarrow} 0$$

and

$$(39) \qquad\qquad N\mathbb{E}\left[\gamma_t(\underbrace{1, \ldots, 1}_{n \text{ times}})\right] \underset{N\to\infty}{\longrightarrow} e^{-nt} \qquad \text{for all } n \geq 2.$$

Recall that $\{e_1, \ldots, e_N\}$ denotes the canonical basis of $\mathbb{C}^{N\times 1}$. For each $k \geq 1$,

$$d\gamma_t(k) = e_k^*(dK_t)P_t e_k - e_k^* P_t dK_t e_k + \left(\frac{1}{N} - \gamma_t(k)\right) dt.$$

As an application of the matricial Itô calculus (see [9, Sect. 2.1]), we get the quadratic variation dynamics

$$d\langle \gamma_t(k), \gamma_t(\ell)\rangle = -\frac{2}{N}\left[\gamma_t(k)\gamma_t(\ell) - 1_{k=\ell}\gamma_t(k)\right] dt.$$

By Itô's formula for products of semi martingales, it follows that $\mathbb{E}\left[\gamma_t(k_1,\ldots,k_n)\right]$ is a smooth function of t which satisfies the differential system

$$
\begin{aligned}
(40)\quad \partial_t \mathbb{E}\left[\gamma_t(k_1,\ldots,k_n)\right] \;=\; & -n\left(1+\frac{n-1}{N}\right)\mathbb{E}\left[\gamma_t(k_1,\ldots,k_n)\right] \\
& +\frac{1}{N}\sum_{j=1}^{n}(2\,|\{i\,;\,i<j,k_i=k_j\}|+1)\mathbb{E}\left[\gamma_t(k_1,\ldots,\hat{k}_j,\ldots,k_n)\right],
\end{aligned}
$$

where we used the convention $\gamma_t(k_1,\ldots,k_n)=1$ if $n=0$. Let us now prove (39). For each $n\geq 1$, we define

$$
q_t(n) = Ne^{nt}\mathbb{E}\left[\gamma_t(\underbrace{1,\ldots,1}_{n\text{ times}})\right]
$$

(the dependence of $q_t(n)$ on N is implicit). Note first that since $\|U_t\|^2 = N$, for all $n\geq 1$,

$$
\mathbb{E}\left[|U_t(1)|^{2n}\right]N^{-n+1}\leq \mathbb{E}\left[|U_t(1)|^2\right]=1,
$$

so that $q_t(n)\leq e^{nt}$. Moreover, by the differential system of (40), we have, for all $n\geq 2$,

$$
\partial_t q_t(n) = -\frac{n-1}{N}q_t(n) + \frac{1}{N}\sum_{i=1}^{n}(2j-3)q_t(n-1).
$$

It follows that $\partial_t q_t(n)$ converges to zero uniformly on every compact subset of $\mathbb{R}_+$ as $N\longrightarrow\infty$. Since for $t=0$, $q_t(n)=1$ for each n, we get that for each $n\geq 2$,

$$
q_t(n)\underset{N\to\infty}{\longrightarrow}1,
$$

so that (39) is proved. Let us at last prove (38). Thanks to the differential system of (40) and to the fact that $\mathbb{E}\left[\gamma_t(1,1,2)\right]=\mathbb{E}\left[\gamma_t(1,2,2)\right]$, we have

$$
\partial_t \mathbb{E}\left[\gamma_t(1,1,2,2)\right] = -4\left(1+\frac{3}{N}\right)\mathbb{E}\left[\gamma_t(1,1,2,2)\right] + \frac{8}{N}\mathbb{E}\left[\gamma_t(1,1,2)\right]
$$

Due to exchangeability and to $\|U_t\|=N$, both $\mathbb{E}\left[\gamma_t(1,1,2)\right]$ and $\mathbb{E}\left[\gamma_t(1,1,2)\right]$ are not greater than $\frac{1}{N(N-1)}$. Therefore, we get:

$$
\left|\partial_t\left(N^2 e^{4t}\mathbb{E}\left[\gamma_t(1,1,2,2)\right]\right)\right|\leq \frac{11}{N}
$$

Since $\gamma_0(1,1,2,2)=0$, this proves (38). $\qquad\square$

10. Proof of Proposition 3.4

Let $\pi\in\operatorname{Part}(k)$ and $\Gamma:=\Gamma_t$; we have to prove that $f_{\Gamma_t}(\pi)=e^{\kappa(\pi)t}$. By the very definition of κ, it is obvious that $\kappa(\pi)=\kappa(\operatorname{thin}(\pi))$ and that κ is additive on the connected components of π. Therefore, due to Lemma 6.3, it is enough to prove that $f_{\Gamma_t}(\pi)=e^{\kappa(\pi)t}$ for π thin and connected. Note that in this case, $\kappa(\pi)$ is equal to k.

We start from Formula (22) for $f_{\Gamma_t}(\pi)$ and are going to prove that the only non-vanishing term in this formula corresponds to the trivial partition τ with only one block. This will imply:

$$f_{\Gamma_t}(\pi) = \prod_{J \in \pi} \Gamma_t(|J|) = \prod_{J \in \pi} e^{-|J|t} = e^{-kt} = e^{-\kappa(\pi)t}$$

and this will prove the proposition.

For $k = 1$, it is trivial. Let us suppose $k > 1$. Since π is thin and connected, it has a crossing (and this implies in fact $k \geq 4$). Let us consider a non trivial τ (i.e. with at least two blocks) such that $H := H(\pi, \tau)$ has no cycle. Let c be a coloring of the edges of $G_{/\pi}$ with kernel τ.

Since H is connected, there exists two blocks E and E' of H with a non-empty intersection. Let $J_0 \in \pi$ such that $\{J_0\} = E \cap E'$ (since H has no cycle, this intersection is a singleton). Since π is thin, by Proposition 5.1, E is not a singleton. Let $J_0' \in E \setminus \{J_0\}$. Since the graph $G_{/\pi}$ is a circuit, there exist at least one path from J_0 to J_0' and one path from J_0' to J_0. But if they were the only ones, there would be no crossing between J_0 and the others blocks of E, and π would not be connected (*cf.* for instance the (counter-)example page 15). Therefore there exist at least two paths from J_0 to J_0' and two paths from J_0' to J_0. Hence we deduce that $n_c^+(J_0, i) \geq 2$ for a certain colour i (see page 15 for the definition of $n_c^+(J_0, i)$). With the same argument, we get $n_c^+(J_0, j) \geq 2$ for a certain $j \neq i$. This implies that

$$\Gamma(n_\tau(J_0)) = \Gamma(n_c^+(J_0, \ell), \ell \geq 1) = 0.$$

Hence, $\prod_{J \in \pi} \Gamma(n_\tau(J)) = 0$, and this implies that the only τ with a non-vanishing contribution to the computation of $f_{\Gamma_t}(\pi)$ is the trivial one with only one block. $\qquad\square$

11. Proof of Proposition 3.5

Lemma 11.1. *For any Hermitian matrices A, B with B positive, any column vector V, any constant $c \in \mathbb{R}$:*

(41) $$\mathbb{E}\left[(A + iB - cX_t VV^*)^{-1}\right] = (A + iB + it|c|VV^*)^{-1}$$

with X_t a random variable with law $\mathcal{C}_t$.

Proof.   On can suppose V to have unit norm. First, for f a rational function bounded with non positive degree whose poles are all in $\mathbb{C}^+ := \{z \in \mathbb{C} \,;\, \Im(z) > 0\}$, $\mathcal{C}_t(f) = f(-ti)$ (it can be proved via the residue formula if $f(z) = \frac{1}{z-a}$ and can then be generalized by density). Second, for Q the orthogonal projector on $V^\perp$,

$$\det(A + iB - cxVV^*) = (\langle V, (A + iB)V \rangle - cx)\det_{|Q}(Q(A + iB)Q)$$

with $\det_{|Q}$ the determinant on the space $V^\perp$. If we suppose $c > 0$, this implies that the entries of $(A + iB - cxVV^*)^{-1}$ are rational functions of x with poles in $\mathbb{C}^+$. The equality

(41) can then be deduced from these two remarks.   □

Proof of Proposition 3.5. Let $X_n^1, \ldots, X_n^{Nn}$, $U_N^1, U_N^2, \ldots$ be as in Theorem 2.1, with $X_n^i \sim \mathcal{C}_{\frac{t}{n}}$. If we apply repeatedly the preceding equality, we obtain that for any Hermitian matrix A, $z \in \mathbb{C}^+$, $c \in \mathbb{R}$:

$$\mathbb{E}\left[(zI_N - A - (X_n^1 U_N^1 U_N^{1*} + \cdots + X_n^{Nn} U_N^{Nn} U_N^{Nn*})/N)^{-1}|U_N^1, \ldots, U_N^{Nn}\right]$$

$$= \left(zI_N + it\frac{U_N^1 U_N^{1*} + \cdots + U_N^{Nn} U_N^{Nn*}}{nN} - A\right)^{-1}$$

When $n \longrightarrow \infty$, $\frac{U_N^1 U_N^{1*} + \cdots + U_N^{Nn} U_N^{Nn*}}{nN}$ tends a.s. to $\mathbb{E}[U_N U_N^*] = I_N$ by the law of large numbers, and $(X_n^1 U_N^1 U_N^{1*} + \cdots + X_n^{Nn} U_N^{Nn} U_N^{Nn*})/N$ tends in law to $\mathbb{P}_{U_N}^{(\mathcal{C}_t)}$ following Theorem 2.1. Therefore, it is now easy to deduce that:

$$\mathbb{E}\left[(zI_N - A - M_t)^{-1}\right] = (zI_N + itI_N - A)^{-1} = P_t(f_z)(A)$$

with $f_z(x) = (z - x)^{-1}$. This establishes the proposition for $f = f_z$. Using routine density and linearity arguments, this is enough to prove the proposition.   □

12. Appendix

12.1. Matrix approximations in the sense of the rank.

The *Cauchy transform* of a finite measure μ on $\mathbb{R}$ is $G_\mu(z) := \int \frac{\mu(\mathrm{d}t)}{z-t}$, for $z \in \mathbb{C}^+$.

Lemma 12.1. *Let μ_N be a sequence of random probability measures on the real line such that for any $\varepsilon > 0$, there is another sequence μ_N^ε converging weakly to a deterministic probability measure μ^ε and such that almost surely, for N large enough, for any $z \in \mathbb{C}^+$,*

$$(42) \qquad \left|G_{\mu_N}(z) - G_{\mu_N^\varepsilon}(z)\right| \le \frac{\varepsilon}{\Im z}.$$

Then μ_N converges weakly to a deterministic probability measure $\mu = \lim_{\varepsilon \to 0} \mu^\varepsilon$.

Proof. Almost surely, by (42), for any $z \in \mathbb{C}^+$, the sequence $G_{\mu_N}(z)$ is a Cauchy sequence, hence converges to a deterministic limit $G(z)$ such that for any ε,

$$(43) \qquad |G_\mu(z) - G_{\mu^\varepsilon}(z)| \le \frac{\varepsilon}{\Im z}.$$

So by [2, Th. 2.4.4 b)], μ_N converges vaguely to a measure μ with total mass ≤ 1 such that $G_\mu = G$ and $\mu = \lim_{\varepsilon \to 0} \mu^\varepsilon$. To see that μ is a probability measure, it suffices to notice that for any measure ν, $iy\Im(G_\nu(iy))$ increases to $\nu(\mathbb{R})$ as $y \uparrow \infty$, and then to use (43).   □

For M a random $N \times N$ Hermitian matrix with eigenvalues $\lambda_1, \ldots, \lambda_N$, recall that the *empirical spectral law* of M is the random probability measure μ_M defined by

$$(44) \qquad \mu_M := \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}$$

and that the *mean spectral law* of M is the deterministic probability measure $\overline{\mu}_M$ defined, for any bounded Borel function f, by

$$(45) \qquad \overline{\mu}_M(f) := \mathbb{E}(\mu_M(f)),$$

with μ_M as in (44).

Lemma 12.2. *For each N, let M_N be an $N \times N$ random Hermitian matrix. Suppose that for any $\varepsilon > 0$, there is a sequence M_N^ε of random Hermitian matrices whose empirical (resp. mean) spectral law converges almost surely (resp. converges) to a deterministic probability measure μ^ε and such that almost surely, for N large enough, $\mathrm{rank}(M_N - M_N^\varepsilon) \leq N\varepsilon$ (resp. such that for N large enough, $\mathbb{E}[\mathrm{rank}(M_N - M_N^\varepsilon)] \leq N\varepsilon$). Then the empirical (resp. mean) spectral law of M_N converges almost surely (resp. converges) to a deterministic probability measure $\mu = \lim_{\varepsilon \to 0} \mu^\varepsilon$*

Proof.   Let us first prove the *almost sure version* of the lemma. For any Hermitian matrix H and any $z \in \mathbb{C}^+$, we have $\|(z - H)^{-1}\|_{\mathrm{op}} \leq (\Im z)^{-1}$. So by the formula

$$(z - B)^{-1} - (z - A)^{-1} = (z - B)^{-1}(B - A)(z - A)^{-1},$$

we have

$$(46) \qquad \left| \mathrm{Tr}(z - M_N)^{-1} - \mathrm{Tr}(z - M_N^\varepsilon)^{-1} \right| \leq 2(\Im z)^{-1} \times \mathrm{rank}(M_N - M_N^\varepsilon)$$

and for any ε, with probability one, for N large enough, for all $z \in \mathbb{C}^+$

$$\left| \frac{1}{N} \mathrm{Tr}(z - M_N)^{-1} - \frac{1}{N} \mathrm{Tr}(z - M_N^\varepsilon)^{-1} \right| \leq \frac{2\varepsilon}{\Im z}, \quad i.e. \quad |G_{\mu_{M_N}}(z) - G_{\mu_{M_N^\varepsilon}}(z)| \leq \frac{2\varepsilon}{\Im z}.$$

It follows that the sequences $\mu_N := \mu_{M_N}$ and $\mu_N^\varepsilon := \mu_{M_N^\varepsilon}$ satisfy the hypotheses of the previous lemma, which allows to conclude.

Let us now prove the *expectations version* of the lemma. Taking the expectation of (46), we get

$$|G_{\overline{\mu}_{M_N}}(z) - G_{\overline{\mu}_{M_N^\varepsilon}}(z)| \leq \frac{2\varepsilon}{\Im z}.$$

It follows that the (non random) sequences $\mu_N := \overline{\mu}_{M_N}$ and $\mu_N^\varepsilon := \overline{\mu}_{M_N^\varepsilon}$ satisfy the hypotheses of the previous lemma, which allows to conclude. $\square$

12.2. **Infinitely divisible distributions and cumulants.**

12.2.1. *The classical case.* Let us recall the definition of infinitely divisible laws with respect to the classical convolution $*$ (see [27]).

Proposition-Definition 12.3. *Let μ be a probability measure on $\mathbb{R}$. The we have equivalence between :*

(i) *there is a sequence (k_n) of positive integers tending to $+\infty$ and a sequence (ν_n) of probability measures on the real line such that as $n \longrightarrow \infty$,*

(47)
$$\underbrace{\nu_n * \cdots * \nu_n}_{k_n \ times} \longrightarrow \mu,$$

(ii) *there is a continuous semigroup $(\mu^{*t})_{t \in [0,\infty)}$ for the convolution $*$ starting at δ_0 such that $\mu^{*1} = \mu$,*

(iii) *the Fourier transform of μ has the form $\int_{t \in \mathbb{R}} e^{it\xi} \mathrm{d}\mu(t) = e^{\Psi_\mu(\xi)}$, with*

(48)
$$\Psi_\mu(\xi) = i\gamma\xi + \int_{\mathbb{R}} \underbrace{(e^{it\xi} - 1 - \frac{it\xi}{t^2+1}) \frac{t^2+1}{t^2}}_{:= -\frac{\xi^2}{2} \ for \ t = 0} \mathrm{d}\sigma(t),$$

γ being a real number and σ being a finite positive measure on $\mathbb{R}$.

In this case, μ is said to be $*$-*infinitely divisible*, the pair (γ, σ), called the *Lévy pair* of μ, is unique, and μ will be denoted by $\nu_*^{\gamma,\sigma}$.

Part (i) of the above definition characterizes such laws as the limit laws of sums of i.i.d. random variables, Part (ii) expresses them as the distributions of one-dimensional marginals of Lévy processes, and Part (iii) is known as the *Lévy-Kinchine formula*. The pair (γ, σ) can be interpreted as follows: γ is a drift, $\sigma(\{0\})$ represents the brownian component of the Lévy process associated to μ, and the measure $1_{x \neq 0} \frac{1+x^2}{x^2} \mathrm{d}\sigma(x)$, when it is finite, represents its Poisson compound part (when this measure is not finite, the Lévy process can be understood as a limit of such decompositions).

12.2.2. *The free case.* In [13, 14], Bercovici and Voiculescu have proved that Proposition-Definition 12.3 stays true if one replaces the classical convolution $*$ by the free additive convolution $\boxplus$ and Formula (48) by the following formula for the R-transform of μ:

(49)
$$R_\mu(z) = \gamma + \int_{\mathbb{R}} \frac{z+t}{1-tz} \mathrm{d}\sigma(t)$$

(recall that the R-transform of μ is defined by the formula $R_\mu(z) = G_\mu^{-1}(z) + \frac{1}{z}$, see [24, 2]). In this case, μ is said to be $\boxplus$-*infinitely divisible* and is denoted by $\nu_\boxplus^{\gamma,\sigma}$.

The map $\Lambda : \nu_*^{\gamma,\sigma} \in \{*\text{-infinitely divisible laws}\} \to \nu_\boxplus^{\gamma,\sigma} \in \{\boxplus\text{-infinitely divisible laws}\}$ is called the *Bercovici-Pata bijection*. In [12] (see also [17]), Bercovici and Pata proved that

this bijection preserves limit theorems, i.e. that for any sequence (k_n) of positive integers tending to infinity, for any sequence (ν_n) of laws on $\mathbb{R}$, for any $*$-infinitely divisible law μ, we have

$$\underbrace{\nu_n * \cdots * \nu_n}_{k_n \text{ times}} \longrightarrow \mu \iff \underbrace{\nu_n \boxplus \cdots \boxplus \nu_n}_{k_n \text{ times}} \longrightarrow \Lambda(\mu).$$

12.2.3. *The cumulants point of view.* Let μ be a probability measure on $\mathbb{R}$ whose Laplace transform is defined in a neighborhood of zero. Its *classical cumulants* are the numbers $(c_n(\mu))_{n \geq 1}$ defined by the formula

$$\log \int_{t \in \mathbb{R}} e^{\xi t} \mathrm{d}\mu(t) = \sum_{n \geq 1} \frac{c_n(\mu)}{n!} \xi^n \qquad (\xi \in \mathbb{C} \text{ small enough}).$$

In the same way, for μ a compactly supported probability measure on $\mathbb{R}$, the *free cumulants* of μ are the numbers $(k_n(\mu))_{n \geq 1}$ defined by the formula

$$R_\mu(z) = \sum_{n \geq 1} k_n(\mu) z^{n-1} \qquad (z \in \mathbb{C} \text{ small enough}).$$

It can easily be seen (see e.g. [8, Eq. (2.1)]) that for any Lévy pair (γ, μ), the classical (resp. free) cumulants of $\nu_*^{\gamma,\sigma}$ (resp. of $\nu_\boxplus^{\gamma,\sigma}$) are the given by the formula

$$(50) \qquad k_n(\nu_*^{\gamma,\sigma}) = c_n(\nu_\boxplus^{\gamma,\sigma}) = \begin{cases} \gamma + \int t \mathrm{d}\sigma(t) & \text{if } n = 1, \\[2mm] \int t^{n-2}(1 + t^2) \mathrm{d}\sigma(t) & \text{if } n \geq 2. \end{cases}$$

Hence the Bercovici-Pata bijection can be seen as (the continuous extension of) the map which transforms classical cumulants into free ones.

12.3. **Combinatorial definitions.** In this section, definitions and results can be found in the classical book [15] of Berge.

12.3.1. *Graphs.* A *graph* $G = (V, E)$ is characterized by a set V of vertices and a family $E = (e_i, i \in I)$ of edges:

(1) If the edges are elements of $V \times V$, G is a *directed* graph;
(2) If the edges are singletons or pairs of V, G is a *non-directed* graph.

Note that **multiple edges** are possible with this definition, since $E = (e_i, i \in I)$ is a *family* and not a *set*, which means that the e_i's are not pairwise distinct.

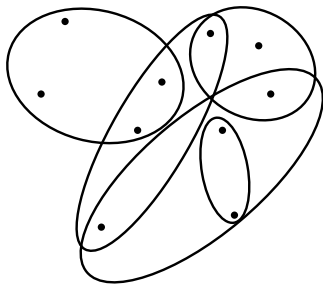In the following three definitions, we suppose G **non-directed**.

(1) A *connected component* of G is a block of the smallest partition of V which is above E w.r.t. the refinement order.
(2) A *path* in G is a sequence $(v_0, e_{i_1}, v_1, e_{i_2}, \ldots, e_{i_k}, v_k)$ with $k \geq 1$, $v_0, \ldots, v_k \in V$, $e_{i_j} = \{v_{j-1}, v_j\}$. If, moreover, the $v_0 = v_k$, then the path is said to be a *closed*. In this case, if $v_0, \ldots, v_{k-1}$ are pairwise distinct and $i_1, \ldots, i_k$ are pairwise distinct, the path is said to be a *cycle*.
(3) The *cyclomatic number* of G is equal to $|I| - |V| + n$, with n the number of connected components of G.

The following proposition can easily be proved by induction on the number of vertices of G.

Proposition 12.4. *For every non directed graph G, its cyclomatic number is non-negative. It vanishes if and only if G has no cycle.*

The previous definitions extend easily to the framework of a **directed** graph $G = (V, E = (e_i, i \in I))$. For example, a *path* in G is a sequence $(v_0, e_{i_1}, v_1, e_{i_2}, \ldots, e_{i_k}, v_k)$ with $\ell \geq 1$, $v_0, \ldots, v_k \in V$, $e_{i_j} = (v_{j-1}, v_j)$. In a directed graph, two cycles c, c' are said to be *disjoint* if they do not have any edge in common, *i.e.* if there is no $i \in I$ such that e_i appears both in c and in c' (whereas they can have some vertices in common). We say that the directed graph G is a *circuit* if a closed path of G visits each of its edges exactly once (vertices can be passed-by more than once). One can easily prove (by induction) that any circuit is a union of disjoint cycles.

12.3.2. *Hypergraphs.* A *hypergraph* is a pair $H = (V, E)$, where V is a set (the *vertices*) and $E = (E_i, i \in I)$ is a family (the *edges*), of non-empty subsets of V such that $\cup_{i \in I} E_i = V$. An example is given at Figure 10(a). If the edges have only one or two elements, then H reduces to a non directed graph.
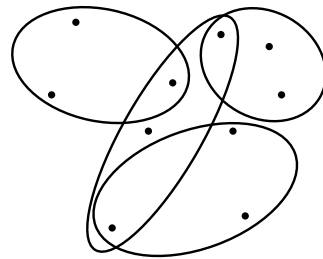
(a) A hypergraph   (b) A hypergraph with no cycle

FIGURE 10. Two examples of hypergraphs

The refinement order for hypergraphs defined on a same set of vertices is defined as the refinement order on their sets of edges.

A *connected component* of H is a block of the smallest partition of V which is above E w.r.t. the refinement order. For instance, there is only one connected component in the hypergraph of Figure 10(a), hence it is said to be *connected*.

Cycles are defined in hypergraphs as in non-directed graphs, except that cycles with length one are not accepted (in fact, they do not make sense) : a *cycle* of H is a sequence $(v_0, E_{i_1}, v_1, E_{i_2}, \ldots, v_{\ell-1}, E_{i_\ell}, v_\ell)$ with $\ell \geq 2$, $v_0, v_1, \cdots, v_{\ell-1}$ pairwise distincts, $v_0 = v_\ell$, $i_1, \ldots, i_\ell$ pairwise distinct and $v_{j-1}, v_j \in E_{i_j}$ for all j.

An example of hypergraph with no cycle is given in Figure 10(b). Notice that a hypergraph with no cycle is linear: two different edges have at most one vertex in common. Notice also that a hypergraph with only one vertex or only one edge has no cycle.

Proposition 12.5. *Let n be the number of connected components of H. Then:*

- $\sum_{i \in I} |E_i| - |I| - |V| + n \geq 0$;
- $\sum_{i \in I} |E_i| - |I| - |V| + n = 0$ *if and only if H has no cycle.*

Proof. This is a standard result (see for instance [15, Ch. 7, Prop. 4]), which is a consequence of Proposition 12.4. For the easyness of the reader, let us recall the short proof. We associate to H a non directed graph $G(H)$ with vertices $V \cup I$, and with an edge between $x \in V$ and $i \in I$ if and only if $x \in E_i$. The graph $G(H)$ is simple, i.e. there cannot be more than one edge between two of its vertices. Also, it has no loop, i.e. no edge from one vertex to itself. The number of edges is equal to $\sum_{i \in I} |E_i|$, and the number of connected components is n, so the cyclomatic number of $G(H)$ is equal to $\sum_{i \in I} |E_i| - (|V| + |I|) + n$. It is non-negative, and vanishes if and only $G(H)$ has no cycle, which means exactly that H has no cycle. $\square$

REFERENCES

[1] R. Adamczak *On the Marchenko-Pastur and circular laws for some classes of random matrices with dependent entries*, Electron. J. Probab., Vol. 16 (2011) 1065–1095.

[2] G. Anderson, A. Guionnet, O. Zeitouni *An Introduction to Random Matrices*. Cambridge studies in advanced mathematics, 118 (2009).

[3] G. Aubrun *Random points in the unit ball of ℓ_p^n*, Positivity 10, (2006) 755–759.

[4] S.T. Belinschi, A. Dembo, A. Guionnet *Spectral measure of heavy tailed band and covariance random matrices*. Comm. Math. Phys. 289 (2009), no. 3, 1023–10.

[5] G. Ben Arous, A. Guionnet *The spectrum of heavy tailed random matrices*. Comm. Math. Phys. 278 (2008), no. 3, 715–751.

[6] F. Benaych-Georges *Classical and free infinitely divisible distributions and random matrices*. Ann. Probab. Vol. 33, no. 3 (2005) 1134–1170.

[7] F. Benaych-Georges *Infinitely divisible distributions for rectangular free convolution: classification and matricial interpretation* Probab. Theory Related Fields Vol. 139, no. 1-2 (2007) 143–189.

[8] F. Benaych-Georges *Taylor expansions of R-transforms, application to supports and moments.* Indiana Univ. Math. J., Vol. 55, no. 2 (2006) 465–483.

[9] F. Benaych-Georges *Finite dimensional projections of the Brownian motion on large unitary groups.* Bull. Soc. Math. France, Vol. 139, no. 4 (2011) 593–610.

[10] F. Benaych-Georges, T. Cabanal-Duvillard *A matrix interpolation between classical and free max operations: I. The univariate case.* J. Theoret. Probab. Vol. 23, no. 2 (2010) 447–465.

[11] F. Benaych-Georges, T. Lévy *A continuous semigroup of notions of independence between the classical and the free one.* Ann. Probab., Vol. 39, no. 3 (2011), 904–938.

[12] H. Bercovici, V. Pata, with an appendix by P. Biane *Stable laws and domains of attraction in free probability theory.* Annals of Mathematics, 149. (1999) 1023–1060.

[13] H. Bercovici, D. Voiculescu *Lévy-Hinchin type theorems for multiplicative and additive free convolution* Pacific J. Math. 153 (1992), no. 2, 217–248.

[14] H. Bercovici, D. Voiculescu *Free convolution of measures with unbounded supports.* Indiana Univ. Math. J. 42 (1993) 733–773.

[15] C. Berge *Graphes et hypergraphes.* Deuxième édition, Collection Dunod Université, Série Violette, No. 604, Dunod, Paris, 1973.

[16] T. Cabanal-Duvillard *A matrix representation of the Bercovici-Pata bijection.* Electron. J. Probab. 10 (2005), no. 18, 632–661.

[17] G.P. Chistyakov, F. Götze *Limit theorems in free probability theory. I.* Ann. Probab. 36 (2008), 54–90.

[18] J. A. Domínguez-Molina, V. Pérez-Abreu, A. Rocha-Arteaga *Covariation representations for random matrix ensembles of free innitely divisible distributions and Hermitian Levy processes.* arXiv:1207.3831.

[19] B. V. Gnedenko, A. N. Kolmogorov *Limit distributions for sums of independent random variables.* Translated and annotated by K. L. Chung. With an Appendix by J. L. Doob. Addison-Wesley Publishing Company, Inc., Cambridge, Mass., 1954.

[20] F. Götze, A. Tikhomirov *Limit theorems for spectra of positive random matrices under dependence.* Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI), 311(Veroyatn. i Stat. 7):92–123, 299, 2004.

[21] F. Götze, A. N. Tikhomirov *Limit theorems for spectra of random matrices with martingale structure.* In Stein's method and applications, volume 5 of Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap., pages 181–193. Singapore Univ. Press, Singapore, 2005.

[22] C. Male *Distributions of traffics and their free product: an asymptotic freeness theorem for random matrices and a central limit theorem.* arXiv:1111.4662.

[23] V.A. Marchenko, L.A. Pastur *Distribution of eigenvalues for some sets of random matrices.* Mat. Sb. (N.S.), 72(114):4, 507–536.

[24] A. Nica, R. Speicher *Lectures on the combinatorics of free probability.* London Mathematical Society Lecture Note Series, 335. Cambridge University Press, Cambridge, 2006.

[25] A. Pajor, L.A. Pastur *On the Limiting Empirical Measure of the sum of rank one matrices with log-concave distribution.* Studia Math. 195 (2009), 11–29.

[26] V. Pérez-Abreu, N. Sakuma *Free generalized gamma convolutions.* Electron. Commun. Probab. 13 (2008), 526–539.

[27] V.V. Petrov *Limit theorems of probability theory* Oxford Studies in Probability, 4, 1995.

[28] K. Sato *Lévy processes and infinitely divisible distributions* Cambridge Studies in Advanced Mathematics, 1999.

[29] M. Schürmann, M. Skeide, S. Volkwardt *Transformations of Lévy Processes on Hopf algebras*, Commun. Stoch. Anal., Volume 4 (2010):4, 553–577.

[30] T. Tao *Topics in random matrix theory*, Graduate Studies in Mathematics, AMS, 2012.

[31] Y. Q. Yin, P. R. Krishnaiah *Limit theorem for the eigenvalues of the sample covariance matrix when the underlying distribution is isotropic.* Teor. Veroyatnost. i Primenen., 30(4):810–816, 1985.

[32] I. Zakharevich *A generalization of Wigner's law.* Commun. Math. Phys. 268(2), 403–414 (2006).

FLORENT BENAYCH-GEORGES: MAP 5, UMR CNRS 8145 - UNIVERSITÉ PARIS DESCARTES, 45 RUE DES SAINTS-PÈRES 75270 PARIS CEDEX 6, FRANCE AND CMAP, ÉCOLE POLYTECHNIQUE, ROUTE DE SACLAY, 91128 PALAISEAU CEDEX, FRANCE.

E-mail address: florent.benaych-georges@parisdescartes.fr

THIERRY CABANAL-DUVILLARD: MAP 5, UMR CNRS 8145 - UNIVERSITÉ PARIS DESCARTES, 45 RUE DES SAINTS-PÈRES 75270 PARIS CEDEX 6, FRANCE

E-mail address: Thierry.Cabanal-Duvillard@mi.parisdescartes.fr