The norm of polynomials in large random and deterministic matrices

Camille Male

The norm of polynomials in large random and deterministic matrices

Camille Male*

with an appendix by

Dimitri Shlyakhtenko

ABSTRACT:

Let $\mathbf{X}_N = (X_1^{(N)}, \dots, X_p^{(N)})$ be a family of $N \times N$ independent, normalized random matrices from the Gaussian Unitary Ensemble. We state sufficient conditions on matrices $\mathbf{Y}_N = (Y_1^{(N)}, \dots, Y_q^{(N)})$, possibly random but independent of $\mathbf{X}_N$, for which the operator norm of $P(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)$ converges almost surely for all polynomials P. Limits are described by operator norms of objects from free probability theory. Taking advantage of the choice of the matrices $\mathbf{Y}_N$ and of the polynomials P, we get for a large class of matrices the "no eigenvalues outside a neighborhood of the limiting spectrum" phenomena. We give examples of diagonal matrices $\mathbf{Y}_N$ for which the convergence holds. Convergence of the operator norm is shown to hold for block matrices, even with rectangular Gaussian blocks, a situation including non-white Wishart matrices and some matrices encountered in MIMO systems.

1 Introduction and statement of result

For a Hermitian $N \times N$ matrix H_N, let $\mathcal{L}_{H_N}$ denote its empirical eigenvalue distribution, namely

$$\mathcal{L}_{H_N} = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i},$$

*Ecole Normale Supérieure de Lyon, Unité de Mathématiques pures et appliquées, UMR 5669, 46 allée d'Italie, 69364 Lyon Cedex 07, France. camille.male@umpa.ens-lyon.fr

where δ_λ is the Dirac mass in λ and $\lambda_1, \dots, \lambda_N$ are the eigenvalues of H_N. The empirical eigenvalue distribution of large dimensional random matrices has been studied with much interest for a long time. One pioneering result is Wigner's theorem [**?**], from 1958. Let W_N be an $N \times N$ Wigner matrix. Then the theorem states that, under appropriate assumptions, the n-th moment of $\mathcal{L}_{W_N}$ converges in expectation to the n-th moment of the semicircular law as N goes to infinity for any integer n. This result has been generalized in many directions, notably by Arnold [2] for the almost sure convergence of the moments. The convergence of the empirical eigenvalue distribution for covariance matrices was first shown by Marčenko and Pastur [27] in 1967, and has been generalized in the late 1970's and the early 1980's by many people, including Grenander and Silverstein [16], Wachter [40], Jonsson [22], Yin and Krishnaiah [44], Bai, Yin and Krishnaiah [7] and Yin [42].

In 1991, Voiculescu [37] discovered a connection between large random matrices and free probability theory. He showed the so-called asymptotic freeness theorem, which has been generalized for instance in [21, 35, 39], which implies the almost sure weak convergence of the empirical eigenvalue distribution for Hermitian matrices H_N of the form

$$H_N = P(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*),$$

where

- P is a fixed polynomial in $2p+q$ non commutative indeterminates,
- $\mathbf{X}_N = (X_1^{(N)}, \dots, X_p^{(N)})$ is a family of independent $N \times N$ matrices of the normalized Gaussian Unitary Ensemble (GUE),
- $\mathbf{Y}_N = (Y_1^{(N)}, \dots, Y_q^{(N)})$ are $N \times N$ matrices with appropriate assumptions (see Theorem 1.3 below).

The limiting empirical eigenvalue distribution of H_N can be computed by using the notion of freeness. Recall that an $N \times N$ random matrix $X^{(N)}$ is said to be a normalized GUE matrix if it is Hermitian with entries $(X^{(N)}_{n,m})_{1 \leqslant n,m \leqslant N}$, such that the set of random variables $(X^{(N)}_{n,n})_{1 \leqslant n \leqslant N}$, and $(\sqrt{2}\mathrm{Re}\ (X^{(N)}_{n,m}), \sqrt{2}\mathrm{Im}\ (X^{(N)}_{n,m})\ )_{1 \leqslant n < m \leqslant N}$ forms a centered Gaussian vector with covariance matrix $\frac{1}{N}\mathbf{1}_{N^2}$. Moreover, the result of Voiculescu holds even for independent Wigner or Wishart matrices instead of GUE matrices, as it has been proved by Dykema [13] and Capitaine and Casalis [9] respectively.

Currently, it is known for some random matrices, as for example Wigner and Wishart matrices, that, almost surely, the eigenvalues of the matrix belong to a small neighborhood of the limiting eigenvalue distribution for N large enough. More formally, if H_N is a Hermitian matrix whose empirical eigenvalue distribution converges weakly to a probability measure μ it is observed in many situations [6, 43, 4, 5, 29] that : for all $\varepsilon > 0$, almost surely there exists $N_0 \geqslant 1$ such that for all $N \geqslant N_0$ one has

$$\mathrm{Sp}(\ H_N\ ) \subset \mathrm{Supp}\ (\ \mu\ ) + (-\varepsilon, \varepsilon), \tag{1.1}$$

where " Sp " means the spectrum and " Supp " means the support.

The convergence of the extremal eigenvalues to the edges of the spectrum of a single Wigner or Wishart matrix has been shown in the early 1980's by Geman [15], Juhász [24], Füredi and Komlós [14], Jonsson [23] and Silverstein [34, 33]. In 1988, in the case of a real Wigner matrix, Bai and Yin stated in [6] necessary and sufficient conditions for the convergence in terms of the first four moments of the entries of these matrices. In the case of a Wishart matrix, the similar result is due to Yin, Bai, and Krishnaiah [43] and Bai, Silverstein, and Yin [4]. The case of a complex matrix has been investigated later by Bai [3]. The phenomenon "no eigenvalues outside (a small neighborhood of) the support of the limiting distribution" has been shown in 1998 by Bai and Silverstein [5] for large sample covariance matrices and in 2008 by Paul and Silverstein [29] for large separable covariance matrices.

In 2005, Haagerup and Thorbjørnsen [19] have shown (1.1) using operator algebra techniques for matrices $H_N = P(X_1^{(N)}, \dots, X_p^{(N)})$, where P is a polynomial in p non commutative indeterminates and $X_1^{(N)}, \dots, X_p^{(N)}$ are independent, normalized $N \times N$ GUE matrices. This constitutes a real breakthrough in the context of free probability. Their method has been used by Schultz [31] to obtain the same result for Gaussian random matrices with real or symplectic entries, and by Capitaine and Donati-Martin [10] for Wigner matrices with symmetric distribution of the entries satisfying a Poincaré inequality and for Wishart matrices.

A consequence of the main result of the present article is that the phenomenon (1.1) holds in the setting considered by Voiculescu, i.e. for certain Hermitian matrices H_N of the form $H_N = P(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)$.

Theorem 1.1 (The spectrum of large Hermitian random matrices)**.** *Let* $\mathbf{X}_N = (X_1^{(N)}, \dots, X_p^{(N)})$ *be a family of independent, normalized GUE matrices and* $\mathbf{Y}_N = (Y_1^{(N)}, \dots, Y_q^{(N)})$ *be a family of* $N \times N$ *matrices, possibly random but independent of* $\mathbf{X}_N$*. Assume that for every Hermitian matrix*

H_N of the form

$$H_N = P(\mathbf{Y}_N, \mathbf{Y}_N^*),$$

where P is a polynomial in $2q$ non commutative indeterminates, we have with probability one that:

1. **Convergence of the empirical eigenvalue distribution:** *there exists a compactly supported measure μ on the real line such that the empirical eigenvalue distribution of H_N converges weakly to μ as N goes to infinity.*

2. **Convergence of the spectrum:** *for any $\varepsilon > 0$, almost surely there exists N_0 such that for all $N \geqslant N_0$,*

$$\mathrm{Sp}\big(\ H_N\ \big) \subset \mathrm{Supp}\ \big(\ \mu\ \big) + (-\varepsilon, \varepsilon). \tag{1.2}$$

Then almost surely the convergences of the empirical eigenvalue distribution and of the spectrum also hold for all Hermitian matrices $H_N = P(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)$, where P is a polynomial in $p + 2q$ non commutative indeterminates.

Theorem 1.1 is a straightforward consequence of Theorem 1.6 below, where the language of free probability is used. Moreover, Theorem 1.6 specifies Theorem 1.1 by giving a description of the limit of the empirical eigenvalue distribution. For readers convenience, we recall some definitions (see [28] and [1] for details).

Definition 1.2. *1. A *-probability space $(\mathcal{A}, .^*, \tau)$ consists of a unital $\mathbb{C}$-algebra $\mathcal{A}$ endowed with an antilinear involution $.^*$ such that $(ab)^* = b^*a^*$ for all a, b in $\mathcal{A}$, and a state τ. A state τ is a linear functional $\tau : \mathcal{A} \mapsto \mathbb{C}$ satisfying*

$$\tau[\mathbf{1}] = 1, \quad \tau[a^*a] \geqslant 0\ \forall a \in \mathcal{A}. \tag{1.3}$$

The elements of $\mathcal{A}$ are called non commutative random variables. We will always assume that τ is a trace, i.e. that it satisfies $\tau[ab] = \tau[ba]$ for every $a, b \in \mathcal{A}$. The trace τ is said to be faithful when it satisfies $\tau[a^*a] = 0$ only if $a = 0$.

2. The non commutative law of a family $\mathbf{a} = (a_1, \dots, a_p)$ of non commutative random variables is defined as the linear functional $P \mapsto \tau\big[P(\mathbf{a}, \mathbf{a}^*)\ \big]$, defined on the set of polynomials in $2p$ non commutative indeterminates. The convergence in law is the pointwise convergence relative to this functional.

3. *The families of non commutative random variables* $\mathbf{a}_1, \dots, \mathbf{a}_n$ *are said to be free if for all* K *in* $\mathbb{N}$*, for all non commutative polynomials* $P_1, \dots, P_K$
$$\tau\Big[P_1(\mathbf{a}_{i_1}, \mathbf{a}_{i_1}^*) \dots P_K(\mathbf{a}_{i_K}, \mathbf{a}_{i_K}^*) \ \Big] = 0 \qquad (1.4)$$
as soon as $i_1 \neq i_2 \neq \dots \neq i_K$ *and* $\tau\big[P_k(\mathbf{a}_{i_k}, \mathbf{a}_{i_k}^*) \ \big] = 0$ *for* $k = 1, \dots, K$*.*

4. *A family of non commutative random variables* $\mathbf{x} = (x_1, \dots, x_p)$ *is called a free semicircular system when the non commutative random variables are free, selfadjoint* ($x_i = x_i^*$, $i = 1, \dots, p$)*, and for all* k *in* $\mathbb{N}$ *and* $i = 1, \dots, p$*, one has*
$$\tau[x_i^k] = \int t^k d\sigma(t), \qquad (1.5)$$
with $d\sigma(t) = \frac{1}{2\pi}\sqrt{4-t^2}\ \mathbf{1}_{|t|\leqslant 2}\ dt$ *the semicircle distribution.*

Recall first the statement of Voiculescu's asymptotic freeness theorem.

Theorem 1.3 ([21, 35, 38, 39] The asymptotic freeness of $X_1^{(N)}, \dots, X_p^{(N)}$ and $\mathbf{Y}_N$). *Let* $\mathbf{X}_N = (X_1^{(N)}, \dots, X_p^{(N)})$ *be a family of independent, normalized GUE matrices and* $\mathbf{Y}_N = (Y_1^{(N)}, \dots, Y_q^{(N)})$ *be a family of* $N \times N$ *matrices, possibly random but independent of* $\mathbf{X}_N$*. Let* $\mathbf{x} = (x_1, \dots, x_p)$ *be a free semicircular system in a* **-probability space* $(\mathcal{A}, .^*, \tau)$ *and* $\mathbf{y} = (y_1, \dots, y_q)$ *in* $\mathcal{A}^q$ *be a family of non commutative random variables free from* $\mathbf{x}$*. Assume the following.*

1. **Convergence of $\mathbf{Y}_N$:** *Almost surely, the non commutative law of* $\mathbf{Y}_N$ *in* $(\mathrm{M}_N(\mathbb{C}), .^*, \tau_N)$ *converges to the non commutative law of* $\mathbf{y}$*, which means that for all polynomial* P *in* $2q$ *non commutative indeterminates, one has*
$$\tau_N\big[P(\mathbf{Y}_N, \mathbf{Y}_N^*)\big] \underset{N\rightarrow\infty}{\longrightarrow} \tau\big[P(\mathbf{y}, \mathbf{y}^*)\big], \qquad (1.6)$$
where τ_N *denotes the normalized trace of* $N \times N$ *matrices.*

2. **Boundedness of the spectrum:** *Almost surely, for* $j = 1, \dots, q$ *one has*
$$\limsup_{N\rightarrow\infty} \|Y_j^{(N)}\| < \infty, \qquad (1.7)$$
where $\|\cdot\|$ *denotes the operator norm.*

Then the non commutative law of $(\mathbf{X}_N, \mathbf{Y}_N)$ *in* $(\mathrm{M}_N(\mathbb{C}), .^*, \tau_N)$ *converges to the non commutative law of* $(\mathbf{x}, \mathbf{y})$*, i.e. for all polynomial* P *in* $p + 2q$ *non commutative indeterminates, one has*
$$\tau_N\big[P(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)\big] \underset{N\rightarrow\infty}{\longrightarrow} \tau\big[P(\mathbf{x}, \mathbf{y}, \mathbf{y}^*)\big]. \qquad (1.8)$$

In [19] Haagerup and Thorbjørnsen strengthened the connection between random matrices and free probability. Limits of random matrices have now to be seen in more elaborated structure, called $\mathcal{C}^*$-probability space, which is endowed with a norm.

Definition 1.4. *A $\mathcal{C}^*$-probability space $(\mathcal{A}, .^*, \tau, \| \cdot \|)$ consists of a *-probability space $(\mathcal{A}, .^*, \tau)$ and a norm $\| \cdot \|$ such that $(\mathcal{A}, .^*, \| \cdot \|)$ is a $\mathcal{C}^*$-algebra.*

By the Gelfand-Naimark-Segal construction, one can always realize $\mathcal{A}$ as a norm-closed $\mathcal{C}^*$-subalgebra of the algebra of bounded operators on a Hilbert space. Hence we can use functional calculus on $\mathcal{A}$. Moreover, if τ is a faithful trace, then the norm $\| \cdot \|$ is uniquely determined by the following formula (see [28, Proposition 3.17]):

$$\|a\| = \lim_{k\to\infty} \Big(\tau\big[\ (a^*a)^k \ \big] \Big)^{\frac{1}{2k}}, \forall a \in \mathcal{A}. \tag{1.9}$$

The main result of [19] is the following.

Theorem 1.5 ([19] The strong asymptotic freeness of independent GUE matrices)**.** *Let $X_1^{(N)}, \dots, X_p^{(N)}$ be independent, normalized $N \times N$ GUE matrices and let $x_1, \dots, x_p$ be a free semicircular system in a $\mathcal{C}^*$-probability space $(\mathcal{A}, .^*, \tau, \| \cdot \|)$ with a faithful trace. Then almost surely, one has: for all polynomials P in p non commutative indeterminates, one has*

$$\big\| P(X_1^{(N)}, \dots, X_p^{(N)}) \big\| \underset{N\to\infty}{\longrightarrow} \|P(x_1, \dots, x_p)\|. \tag{1.10}$$

This article is mainly devoted to the following theorem which is a generalization of Theorem 1.5 in the setting of Theorem 1.3.

Theorem 1.6 (The strong asymptotic freeness of $X_1^{(N)}, \dots, X_p^{(N)}, \mathbf{Y}_N$)**.** *Let $\mathbf{X}_N = (X_1^{(N)}, \dots, X_p^{(N)})$ be a family of independent, normalized GUE matrices and $\mathbf{Y}_N = (Y_1^{(N)}, \dots, Y_q^{(N)})$ be a family of $N \times N$ matrices, possibly random but independent of $\mathbf{X}_N$. Let $\mathbf{x} = (x_1, \dots, x_p)$ and $\mathbf{y} = (y_1, \dots, y_q)$ be a family of non commutative random variables in a $\mathcal{C}^*$-probability space $(\mathcal{A}, .^*, \tau, \| \cdot \|)$ with a faithful trace, such that $\mathbf{x}$ is a free semicircular system free from $\mathbf{y}$. Assume the following.*

Strong convergence of $\mathbf{Y}_N$: *Almost surely, for all polynomials P in $2q$ non commutative indeterminates, one has*

$$\tau_N\big[P(\mathbf{Y}_N, \mathbf{Y}_N^*)\big] \underset{N\to\infty}{\longrightarrow} \tau[P(\mathbf{y}, \mathbf{y}^*)], \tag{1.11}$$

$$\big\|P(\mathbf{Y}_N, \mathbf{Y}_N^*)\big\| \underset{N\to\infty}{\longrightarrow} \|P(\mathbf{y}, \mathbf{y}^*)\|. \tag{1.12}$$

Then, almost surely, for all polynomials P in $p+2q$ non commutative indeterminates, one has

$$\tau_N\big[P(\mathbf{X}_N,\mathbf{Y}_N,\mathbf{Y}_N^*)\big] \underset{N\to\infty}{\longrightarrow} \tau[P(\mathbf{x},\mathbf{y},\mathbf{y}^*)], \tag{1.13}$$

$$\big\|P(\mathbf{X}_N,\mathbf{Y}_N,\mathbf{Y}_N^*)\big\| \underset{N\to\infty}{\longrightarrow} \|P(\mathbf{x},\mathbf{y},\mathbf{y}^*)\|. \tag{1.14}$$

The convergence of the normalized traces stated in (1.13) is the content of Voiculescu's asymptotic freeness theorem and is recalled in order to give a coherent and complete statement. Theorem 1.1 is easily deduced from Theorem 1.6 by applying Hamburger's theorem [20] for the convergence of the measure and functional calculus for the convergence of the spectrum.

Organization of the paper: In Section 2 we give applications of Theorem 1.6 which are proved in Section 9. Sections 3 to 8 are dedicated to the proof of Theorem 1.6.

Acknowledgments: The author would like to thank Alice Guionnet for dedicating much time for many discussions to the subjects of this paper and, along with Manjunath Krishnapur and Ofer Zeitouni, for the communication of Lemma 8.2. He is very much obliged to Dimitri Shlyakhtenko for his contribution to this paper. He would like to thank Benoit Collins for pointing out an error in a previous version of Corollary 2.1 and giving the idea to fix it. He also likes to thank Mikael de la Salle for useful discussions.

2 Applications

2.1 Diagonal matrices

The first and the simpler matrix model that may be investigated to play the role of matrices $\mathbf{Y}_N$ in Theorem 1.6 consists of deterministic diagonal matrices with real entries and prescribed asymptotic spectral measure.

Corollary 2.1 (diagonal matrices)**.** *Let $\mathbf{X}_N=(X_1^{(N)},\dots,X_p^{(N)})$ be a family of independent, normalized GUE matrices and let $\mathbf{D}_N=(D_1^{(N)},\dots,D_q^{(N)})$ be $N\times N$ deterministic real diagonal matrices, such that for any $j=1,\dots,q$,*

1. *the empirical spectral distribution of $D_j^{(N)}$ converges weakly to a compactly supported probability measure μ_j,*

2. the diagonal entries of $D_j^{(N)}$ are non decreasing:

$$D_j^{(N)} = \mathrm{diag}\left(\lambda_1^{(N)}(j), \dots, \lambda_N^{(N)}(j)\right), \text{ with } \lambda_1^{(N)}(j) \leqslant \dots \leqslant \lambda_N^{(N)}(j),$$

3. for all $\varepsilon > 0$, there exists N_0 such that for all $N \geqslant N_0$, for all $j = 1 \dots q$,

$$\mathrm{Sp}(\ D_j^{(N)}\ ) \subset \ \mathrm{Supp}(\ \mu_j\ ) + (-\varepsilon, \varepsilon).$$

Let $v = (v_1, \dots, v_q)$ in $[0,1]^q$. We set $\mathbf{D}_N^v = \big(D_1^{(N)}(v_1), \dots, D_q^{(N)}(v_q)\big)$, where for any $j = 1, \dots, q$,

$$D_j^{(N)}(v_j) = \mathrm{diag}\left(\lambda_{1+\lfloor v_j N \rfloor}^{(N)}(j), \dots, \lambda_{N+\lfloor v_j N \rfloor}^{(N)}(j)\right), \text{ with indices modulo } N.$$

Let $\mathbf{x} = (x_1, \dots, x_p)$ and $\mathbf{d}^v = \big(d_1(v), \dots, d_q(v)\big)$ be non commutative random variables in a $\mathcal{C}^*$-probability space $(\mathcal{A}, .^*, \tau, \| \cdot \|)$ with a faithful trace, such that

1. $\mathbf{x}$ is a free semicircular system, free from $\mathbf{d}^v$,

2. The variables $d_1(v), \dots, d_q(v)$ commute, are selfadjoint and for all polynomials P in q indeterminates, one has

$$\tau[P(\mathbf{d}^v)\ ] = \int_0^1 P\Big(F_1^{-1}(u+v_1), \dots, F_q^{-1}(u+v_q)\ \Big)du. \tag{2.1}$$

For any $j = 1 \dots q$, the application F_j^{-1} is the (periodized) generalized inverse of the cumulative distribution function $F_j : t \mapsto \mu_j\big(\ ]-\infty, t]\big)$ of μ_j defined by: F_j^{-1} is 1-periodic and for all u in $]0,1]$, $F_j^{-1}(u) = \inf\big\{t \in \mathbb{R}\ \big|\ F_j(t) \geqslant u\big\}$.

Then, with probability one, for all polynomials P in $p+q$ non commutative indeterminates, one has

$$\tau_N\big[P(\mathbf{X}_N, \mathbf{D}_N^v)\big] \underset{N\to\infty}{\longrightarrow} \tau[P(\mathbf{x}, \mathbf{d}^v)] \tag{2.2}$$

$$\big\|P(\mathbf{X}_N, \mathbf{D}_N^v)\big\| \underset{N\to\infty}{\longrightarrow} \|P(\mathbf{x}, \mathbf{d}^v)\|, \tag{2.3}$$

for any v in $[0,1]^q$ except in a countable set.

Remark that the non commutative random variables $d_1, \dots, d_q$ can be realized as classical random variables, d_j being μ_j-distributed for $j = 1, \dots, q$. The dependence between the random variables is trivial since Formula (2.1) exhibits a deterministic coupling. The convergence of the normalized trace (2.2) actually holds for any v. In general, the convergence (2.3) of the norm can fail: the family of matrices $\mathbf{D}_N = (D_1^{(N)}, D_2^{(N)})$ where

$$D_1^{(N)} = \text{ diag } (\mathbf{0}_{\lfloor N/2 \rfloor}, \mathbf{1}_{N-\lfloor N/2 \rfloor}),\ D_1^{(N)} = \text{ diag } (\mathbf{0}_{\lfloor N/2 \rfloor+1}, \mathbf{1}_{N-\lfloor N/2 \rfloor-1})$$

gives a counterexample (consider their difference). Furthermore, let mention that it is clear that we always can take one of the v_i to be zero.

2.2 Non-white Wishart matrices

Theorem 1.6 may be used to deduce the same result for some Wishart matrices as for the GUE matrices. Let $r, s_1, \ldots, s_p \geqslant 1$ be integers. Let $\mathbf{Z}_N = (Z_1^{(N)}, \ldots, Z_p^{(N)})$ be a family of independent positive definite Hermitian random matrices such that for $j = 1, \ldots, p$ the matrix $Z_j^{(N)}$ is of size $s_jN \times s_jN$. Let $\mathbf{W}_N = \mathbf{W}_N(\mathbf{Z}) = (W_1^{(N)}, \ldots, W_p^{(N)})$ be the family of $rN \times rN$ matrices defined by: for each $j = 1, \ldots, p$, $W_j^{(N)} = M_j^{(N)}\ Z_j^{(N)}\ M_j^{(N)*}$, where $M_j^{(N)}$ is a $rN \times s_jN$ matrix whose entries are random variables,

$$M_j^{(N)} = (M_{n,m})_{\substack{1 \leqslant n \leqslant rN \\ 1 \leqslant m \leqslant s_j N}},$$

and the random variables $(\sqrt{2}\text{Re } (M_{n,m}), \sqrt{2}\text{Im } (M_{n,m})\ )_{1 \leqslant n \leqslant rN, 1 \leqslant m \leqslant s_j N}$ form a centered Gaussian vector with covariance matrix $\frac{1}{rN}\mathbf{1}_{2rs_jN^2}$. We assume that $M_1^{(N)}, \ldots, M_p^{(N)}, \mathbf{Z}_N$ are independent. The matrices $W_1^{(N)}, \ldots, W_p^{(N)}$ are called non-white Wishart matrices, the white case occurring when the matrices $Z_j^{(N)}$ are the identity matrices.

Corollary 2.2 (Wishart matrices)**.** *Let $\mathbf{Y}_N = (Y_1^{(N)}, \ldots, Y_q^{(N)})$ be a family of $rN \times rN$ random matrices, independent of $\mathbf{Z}_N$ and $\mathbf{W}_N$. Assume that the families of matrices $(Z_1^{(N)}), \ldots, (Z_q^{(N)}), \mathbf{Y}_N$ satisfy separately the assumptions of Theorem 1.6. Then, almost surely, for all polynomials P in $p + 2q$ non commutative indeterminates, one has*

$$\big\|P(\mathbf{W}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)\big\| \underset{N\to\infty}{\longrightarrow} \|P(\mathbf{w}, \mathbf{y}, \mathbf{y}^*)\|, \tag{2.4}$$

where $\| \cdot \|$ is given by Formula (1.9) with τ a faithful trace for which the non commutative random variables $\mathbf{w} = (w_1, \ldots, w_p)$ and $\mathbf{y} = (y_1, \ldots, y_q)$ are free.

In [29], motivated by applications in statistics and wireless communications, the authors study the global limiting behavior of the spectrum of the following matrix, referred as separable covariance matrix:

$$C_n = \frac{1}{n} A_n^{1/2} X_n B_n X_n^* A_n^{1/2},$$

where X_n is a $n \times m$ random matrix, $A_n^{1/2}$ is a nonnegative definite square root of the nonnegative definite $n \times n$ Hermitian matrix A_n and B_n is a $m \times m$ diagonal matrix with nonnegative diagonal entries. It is shown in [29] that, for n large enough, almost surely the eigenvalues of C_n belong in a small neighborhood of the limiting distribution under the following assumptions:

1. $m = m(n)$ with $c_n := n/m \underset{n\to\infty}{\longrightarrow} c > 0$.

2. The entries of X_n are independent, identically distributed, standardized complex and with a finite fourth moment.

3. The empirical eigenvalue distribution $\mathcal{L}_{A_n}$ (respectively $\mathcal{L}_{B_n}$) of A_n (respectively B_n) converges weakly to a compactly supported probability measure ν_a (respectively ν_b) and the operator norms of A_n and B_n are uniformly bounded.

4. By assumptions 1,2 and 3, it is known that almost surely $\mathcal{L}_{C_n}$ converges weakly to a probability measure $\mu^{(c)}_{\nu_a,\nu_b}$. This define a map $\Phi : (x, \nu_1, \nu_2) \mapsto \mu^{(x)}_{\nu_1,\nu_2}$ (the input x is a positive real number, the inputs ν_1 and ν_2 are probability measures on $\mathbb{R}^+$). Assume that for every $\varepsilon > 0$, there exists $n_0 \geqslant 1$ such that, for all $n \geqslant n_0$, one has
$$\operatorname{Supp}\left(\mu^{(c_n)}_{\mathcal{L}_{A_n},\mathcal{L}_{B_N}}\right) \subset \operatorname{Supp}\left(\mu^{(c)}_{\nu_a,\nu_b}\right) + (-\varepsilon, \varepsilon).$$

Now consider the following situation, where Corollary 2.2 may be applied

1' $n = n(N) = rN$, $m = m(N) = sN$ for fixed positive integers r and s,

2' the entries of X_n are independent, identically distributed, standardized complex Gaussian,

3' the empirical eigenvalue distribution of A_n (respectively B_n) converges weakly to a compactly supported probability measure,

4' for N large enough, the eigenvalues of A_n (respectively B_n) belong in a small neighborhood of its limiting distribution.

Then we obtain by Corollary 2.2 that for N large enough, almost surely the eigenvalues of C_n belong in a small neighborhood of the limiting distribution. The advantage of our version is the replacement of assumption 4 by assumption 4'. Replacing assumptions 1' and 2' by assumptions 1 and 2 could be an interesting question.

2.3 Block matrices

It will be shown as a consequence of Theorem 1.6 that the convergence of norms (1.14) also holds for block matrices.

Corollary 2.3 (Block matrices). *Let $\mathbf{X}_N, \mathbf{Y}_N, \mathbf{x}, \mathbf{y}$ and τ be as in Theorem 1.6. Almost surely, for all positive integer ℓ and for all non commutative polynomials $(P_{u,v})_{1\leqslant u,v\leqslant \ell}$, the operator norm of the $\ell N \times \ell N$ block matrix*

$$\begin{pmatrix} P_{1,1}(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) & \dots & P_{1,\ell}(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) \\ \vdots & & \vdots \\ P_{\ell,1}(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) & \dots & P_{\ell,\ell}(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) \end{pmatrix} \tag{2.5}$$

converges to the norm $\| \cdot \|_{\tau_\ell \otimes \tau}$ of

$$\begin{pmatrix} P_{1,1}(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) & \dots & P_{1,\ell}(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) \\ \vdots & & \vdots \\ P_{\ell,1}(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) & \dots & P_{\ell,\ell}(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) \end{pmatrix}, \tag{2.6}$$

where $\| \cdot \|_{\tau_\ell \otimes \tau}$ is given by the faithful trace $\tau_\ell \otimes \tau$ defined by

$$(\tau_\ell \otimes \tau) \left[\begin{pmatrix} P_{1,1}(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) & \dots & P_{1,\ell}(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) \\ \vdots & & \vdots \\ P_{\ell,1}(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) & \dots & P_{\ell,\ell}(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) \end{pmatrix} \right] = \tau \Big[\frac{1}{\ell} \sum_{i=1}^{\ell} P_{i,i}(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) \Big].$$

2.4 Channel matrices

We give a potential application of Theorem 1.6 in the context of communication, where rectangular block random matrices are sometimes investigated for the study of wireless Multiple-input Multiple-Output (MIMO) systems [25, 36]. In the case of Intersymbol-Interference, the channel matrix H reflects the channel effect during a transmission and is of the form

$$H = \begin{pmatrix} A_1 & A_2 & \dots & A_L & \mathbf{0} & \dots & & \dots & \mathbf{0} \\ \mathbf{0} & A_1 & A_2 & \dots & A_L & \mathbf{0} & & & \vdots \\ \vdots & \mathbf{0} & A_1 & A_2 & \dots & A_L & \mathbf{0} & & \\ & & \ddots & \ddots & \ddots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & & \dots & \mathbf{0} & A_1 & A_2 & \dots & A_L \end{pmatrix}, \tag{2.7}$$

$(A_l)_{1\leqslant \ell\leqslant L}$ are $n_R \times n_T$ matrices that are very often modeled by random matrices e.g. $A_1, \dots, A_L$ are independent and for $\ell = 1, \dots, L$ the entries of the matrix A_ℓ are independent identically distributed with finite variance. The number of matrices L is the length of the impulse response

of the channel, n_T is the number of transmitter antennas and n_R is the number of receiver antennas.
In order to calculate the capacity of such a channel, one must know the singular value distribution of H, which is predicted by free probability theory. Theorem 1.6 may be used to obtain the convergence of the singular spectrum for a large class of such matrices. For instance we investigate in Section 9.3 the following case:

Corollary 2.4 (Rectangular band matrices)**.** *Let r and t be integers. Consider a matrix H of the form (2.7) such that for any $\ell = 1, \dots, L$ one has $A_\ell = C_\ell M_\ell D_\ell$ where*

1. *$\mathbf{M} = (M_1, \dots, M_L)$ is a family of independent $rN \times tN$ random matrices such that for $\ell = 1, \dots, L$ the entries of M_ℓ are independent, Gaussian and centered with variance σ_ℓ^2/N,*

2. *the family of $rN \times rN$ matrices $\mathbf{C} = (C_1, \dots, C_L)$ and the family of $tN \times tN$ matrices $\mathbf{D} = (D_1, \dots, D_L)$ satisfy separately the assumptions of Theorem 1.6,*

3. *the families of matrices $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{D}$ are independent.*

Then, almost surely, the empirical eigenvalue distribution of HH^* converges weakly to a measure μ. Moreover, for any $\varepsilon > 0$, almost surely there exists N_0 such that the singular values of H belong to $\mathrm{Supp}(\mu) + (-\varepsilon, \varepsilon)$.

3 The strategy of proof

Let $\mathbf{X}_N = (X_1^{(N)}, \dots, X_p^{(N)})$ and $\mathbf{Y}_N = (Y_1^{(N)}, \dots, Y_q^{(N)})$ be as in Theorem 1.6. We start with some remarks in order to simplify the proof.

1. We can suppose that the matrices of $\mathbf{Y}_N$ are Hermitian. Indeed for any $j = 1, \dots, q$, one has $Y_j^{(N)} = \mathrm{Re}\ Y_j^{(N)} + i\ \mathrm{Im}\ Y_j^{(N)}$, where

$$\mathrm{Re}\ Y_j^{(N)} := \frac{1}{2}\big(Y_j^{(N)} + Y_j^{(N)*}\big), \quad \mathrm{Im}\ Y_j^{(N)} := \frac{1}{2i}\big(Y_j^{(N)} - Y_j^{(N)*}\big)$$

are Hermitian matrices. A polynomial in $\mathbf{Y}_N, \mathbf{Y}_N^*$ is obviously a polynomial in the matrices $\mathrm{Re}\ Y_1^{(N)}, \dots, \mathrm{Re}\ Y_q^{(N)}$, and $\mathrm{Im}\ Y_1^{(N)}, \dots, \mathrm{Im}\ Y_q^{(N)}$ and so the latter satisfies the assumptions of Theorem 1.6 as soon as $\mathbf{Y}_N$ does.

2. It is sufficient to prove the theorem for deterministic matrices $\mathbf{Y}_N$. Indeed, the matrices $\mathbf{X}_N$ and $\mathbf{Y}_N$ are independent. Then we can choose the underlying probability space to be of the form $\Omega = \Omega_1 \times \Omega_2$, with $\mathbf{X}_N$ (respectively $\mathbf{Y}_N$) a measurable function on Ω_1 (respectively Ω_2). The event "for all polynomials P the convergences (1.13) and (1.14) hold" is a measurable set $\tilde{\Omega} \subset \Omega$. Assume that the theorem holds for deterministic matrices. Then for almost all $\omega_2 \in \Omega_2$, there exists a set $\tilde{\Omega}_1(\omega_2)$ for which for all $\omega_1 \in \tilde{\Omega}_1$, (1.13) and (1.14) hold for $(\mathbf{X}_N(\omega_1), \mathbf{Y}_N(\omega_2))$. The set of such couples (ω_1, ω_2) is of outer measure one and is contained in $\tilde{\Omega}$, hence by Fubini's theorem $\tilde{\Omega}$ is of measure one.

3. It is sufficient to prove that for any polynomial the convergence of the norm in (1.14) holds almost surely (instead of almost surely the convergence holds for all polynomials). Indeed we can switch the words "for all polynomials with rational coefficients" and "almost surely" and both the left and the right hand side in (1.14) are continuous in P.

In the following, when we say that $\mathbf{Y}_N = (Y_1^{(N)}, \dots, Y_q^{(N)})$ is as in Section 3, we mean that $\mathbf{Y}_N$ is a family of deterministic Hermitian matrices satisfying (1.11) and (1.12).

Remark that by (1.12), almost surely the supremum over N of $\|Y_j^{(N)}\|$ is finite for all $j = 1, \dots, q$. Hence by Theorem 1.3, with probability one the non commutative law of $(\mathbf{X}_N, \mathbf{Y}_N)$ in $(\mathrm{M}_N(\mathbb{C}), .^*, \tau_N)$ converges to the law of non commutative random variables $(\mathbf{x}, \mathbf{y})$ in a *-probability space $(\mathcal{A}, .^*, \tau,)$: almost surely, for all polynomials P in $p+q$ non commutative indeterminates, one has

$$\tau_N\big[P(\mathbf{X}_N, \mathbf{Y}_N)\ \big] \underset{N\to\infty}{\longrightarrow} \tau[P(\mathbf{x}, \mathbf{y})], \tag{3.1}$$

where the trace τ is completely defined by:

- $\mathbf{x} = (x_1, \dots, x_p)$ is a free semicircular system,
- $\mathbf{y} = (y_1, \dots, y_q)$ is the limit in law of $\mathbf{Y}_N$,
- $\mathbf{x}, \mathbf{y}$ are free.

Since τ is faithful on the *-algebra spanned by $\mathbf{x}$ and $\mathbf{y}$, we can always assume that τ is a faithful trace on $\mathcal{A}$. Moreover, the matrices $\mathbf{Y}_N$ are uniformly bounded in operator norm. If we define $\|\cdot\|$ in $\mathcal{A}$ by Formula (1.9), then $\|y_j\|$ is finite for every $j = 1, \dots, q$. Hence, we can assume that $\mathcal{A}$ is a $\mathcal{C}^*$-probability space endowed with the norm $\|\cdot\|$.

Haagerup and Thorbjørnsen describe in [19] a method to show that for all non commutative polynomials P, almost surely one has

$$\big\|P(\mathbf{X}_N)\big\| \underset{N\to\infty}{\longrightarrow} \|P(\mathbf{x})\|. \tag{3.2}$$

We present in this section this method with some modification to fit our situation. First, it is easy to see the following.

Proposition 3.1. *For all non commutative polynomials P, almost surely one has*

$$\liminf_{N\to\infty} \big\|P(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)\big\| \geqslant \|P(\mathbf{x}, \mathbf{y}, \mathbf{y}^*)\|. \tag{3.3}$$

Proof. In a $\mathcal{C}^*$-algebra $(\mathcal{A}, .^*, \|\cdot\|)$, one has $\forall a \in \mathcal{A}$, $\|a\|^2 = \|a^*a\|$. Hence, without loss of generality, we can suppose that $H_N := P(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)$ is non negative Hermitian and $h := P(\mathbf{x}, \mathbf{y}, \mathbf{y}^*)$ is selfadjoint. Let $\mathcal{L}_N$ denote the empirical spectral distribution of H_N:

$$\mathcal{L}_N = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i},$$

where $\lambda_1, \dots, \lambda_N$ denote the eigenvalues of H_N and δ_λ the Dirac measure in $\lambda \in \mathbb{R}$. By (3.1) and Hamburger's theorem [20], almost surely $\mathcal{L}_N$ converges weakly to the compactly supported probability measure μ on $\mathbb{R}$ given by: for all polynomial P,

$$\int P \mathrm{d}\mu = \tau[P(h)].$$

Since τ is faithful, the extrema of the support of μ is $\|h\|$ ([28, proposition 3.15]). In particular, if $f : \mathbb{R} \to \mathbb{R}$ is a non negative continuous function whose support is the closure of a neighborhood of $\|h\|$ (f not indentically zero), then almost surely there exists a $N_0 \geqslant 0$ such that for all $N \geqslant N_0$ one has $\mathcal{L}_N(f) > 0$. Hence for $N \geqslant N_0$ some eigenvalues of H_N belong to the considered neighborhood of $\|h\|$ and so $\|H_N\| \geqslant \|h\|$. $\square$

It remains to show that the limsup is smaller than the right hand side in (3.3). The method is carried out in many steps.

Step 1. A linearization trick: *With inequality* (3.3) *established, the question of almost sure convergence of the norm of any polynomial in the considered random matrices can be reduced to the question of the convergence of the spectrum of any matrix-valued selfadjoint degree one polynomials in these matrices. More precisely, in order to get* (3.2)*, it is sufficient to show that for all* $\varepsilon > 0$*, k positive integer, L selfadjoint degree one polynomial with coefficients in* $\mathrm{M}_k(\mathbb{C})$*, almost surely there exists* N_0 *such that for all* $N \geqslant N_0$*,*

$$\mathrm{Sp}\big(\ L(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)\ \big) \ \subset \mathrm{Sp}\big(\ L(\mathbf{x}, \mathbf{y}, \mathbf{y}^*)\ \big) + (-\varepsilon, \varepsilon). \tag{3.4}$$

We refer the readers to [19, Parts 2 and 7] for the proof of this step, which is based on $\mathcal{C}^*$-algebra and operator space techniques. We only recall here the main ingredients. By an argument of ultraproduct it is sufficient to show the following: Let $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ be elements of a $\mathcal{C}^*$-algebra. Assume that for all selfadjoint degree one polynomials L with coefficients in $\mathrm{M}_k(\mathbb{C})$, one has

$$\mathrm{Sp}\big(\ L(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*)\ \big) \ \subset \mathrm{Sp}\big(\ L(\mathbf{x}, \mathbf{y}, \mathbf{y}^*)\ \big). \tag{3.5}$$

Then for all polynomials P one has $\|P(\mathbf{x}, \mathbf{y}, \mathbf{y}^*)\| \geqslant \|P(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*)\|$. The linearization trick used to prove that fact arises from matrix manipulations and Arveson's theorem: with a dilation argument, one deduces from (3.5) that there exists ϕ a unital $*$-homomorphism between the $\mathcal{C}^*$-algebra spanned by $(\mathbf{x}, \mathbf{y})$ and the one spanned by $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ such that one has $\phi(x_i) = \tilde{x}_i$ for $i = 1, \dots, p$, and $\phi(y_i) = \tilde{y}_i$ for $i = 1, \dots, q$. A $*$-homomorphism being always contractive, one gets the result.

We fix a selfadjoint degree one polynomial L with coefficients in $\mathrm{M}_k(\mathbb{C})$. To prove (3.4) we apply the method of Stieltjes transforms. We use an idea from Bai and Silverstein in [5]: we do not compare the Stieltjes transform of $L(\mathbf{X}_N, \mathbf{Y}_N)$ with the one of $L(\mathbf{x}, \mathbf{y})$, but with an intermediate quantity, where in some sense we have taken partially the limit N goes to infinity, only for the GUE matrices. To make it precise, we realize the non commutative random variables $\big(\mathbf{x}, \mathbf{y}, (\mathbf{Y}_N)_{N\geqslant 1}\big)$ in a same $\mathcal{C}^*$-probability space $(\mathcal{A}, .^*, \tau, \|\cdot\|)$ with faithful trace, where

- the families $\mathbf{x}$, $\mathbf{y}$, $\mathbf{Y}_1$, $\mathbf{Y}_2, \dots, \mathbf{Y}_N, \dots$ are free,
- for any polynomials P in q non commutative indeterminates $\tau[P(\mathbf{Y}_N)] := \tau_N[P(\mathbf{Y}_N)]$.

The intermediate object $L(\mathbf{x}, \mathbf{Y}_N)$ is therefore well defined as an element of $\mathcal{A}$. We use a theorem about norm convergence, due to D. Shlyakhtenko and stated in Appendix A, to relate the spectrum of $L(\mathbf{x}, \mathbf{Y}_N)$ with the spectrum of $L(\mathbf{x}, \mathbf{y})$.

Step 2. An intermediate inclusion of spectrum: *for all $\varepsilon > 0$ there exists N_0 such that for all $N \geqslant N_0$, one has*

$$\mathrm{Sp}\big(\ L(\mathbf{x}, \mathbf{Y}_N) \ \big) \ \subset \mathrm{Sp}\big(\ L(\mathbf{x}, \mathbf{y}) \ \big) + (-\varepsilon, \varepsilon). \tag{3.6}$$

We define the Stieltjes transforms g_{L_N} and g_{ℓ_N} of $L_N = L(\mathbf{X}_N, \mathbf{Y}_N)$ and respectively $\ell_N = L(\mathbf{x}, \mathbf{Y}_N)$ by the formulas

$$g_{L_N}(\lambda) = \mathbb{E}\Big[(\tau_k \otimes \tau_N)\Big[\big(\lambda \mathbf{1}_k \otimes \mathbf{1}_N - L(\mathbf{X}_N, \mathbf{Y}_N) \ \big)^{-1} \ \Big]\Big], \tag{3.7}$$

$$g_{\ell_N}(\lambda) = (\tau_k \otimes \tau)\Big[\big(\lambda \mathbf{1}_k \otimes \mathbf{1} - L(\mathbf{x}, \mathbf{Y}_N) \ \big)^{-1} \ \Big], \tag{3.8}$$

for all complex numbers λ such that $\mathrm{Im}\ \lambda > 0$.

Step 3. From Stieltjes transform to spectra: *In order to show (3.5) with (3.6) granted, it is sufficient to show the following: for every $\varepsilon > 0$, there exist $N_0, \gamma, c, \alpha > 0$ such that for all $N \geqslant N_0$, for all λ in $\mathbb{C}$ such that $\varepsilon \leqslant (\mathrm{Im}\ \lambda)^{-1} \leqslant N^{\gamma}$, one has*

$$|g_{L_N}(\lambda) - g_{\ell_N}(\lambda)| \leqslant \frac{c}{N^2} (\mathrm{Im}\ \lambda)^{-\alpha}. \tag{3.9}$$

The proof of Estimate (3.9) represents the main work of this paper. For this task we consider a generalization of the Stieltjes transform. We define the $\mathrm{M}_k(\mathbb{C})$-valued Stieltjes transforms G_{L_N} and G_{ℓ_N} of $L_N = L(\mathbf{X}_N, \mathbf{Y}_N)$ and respectively $\ell_N = L(\mathbf{x}, \mathbf{Y}_N)$ by the formulas

$$G_{L_N}(\Lambda) = \mathbb{E}\Big[(\mathrm{id}_k \otimes \tau_N)\Big[\big(\Lambda \otimes \mathbf{1}_N - L(\mathbf{X}_N, \mathbf{Y}_N) \ \big)^{-1} \ \Big]\Big], \tag{3.10}$$

$$G_{\ell_N}(\Lambda) = (\mathrm{id}_k \otimes \tau)\Big[\big(\Lambda \otimes \mathbf{1} - L(\mathbf{x}, \mathbf{Y}_N) \ \big)^{-1} \ \Big], \tag{3.11}$$

for all $k \times k$ matrices Λ such that the Hermitian matrix $\mathrm{Im}\ \Lambda := (\Lambda - \Lambda^*)/(2i)$ is positive definite. Since $g_{L_N}(\lambda) = \tau_k[G_{L_N}(\lambda \mathbf{1}_k)]$ and $g_{\ell_N}(\lambda) = \tau_k[G_{\ell_N}(\lambda \mathbf{1}_k)]$, a uniform control of $\|G_{L_N}(\Lambda) - G_{\ell_N}(\Lambda)\|$ will be sufficient to show (3.9). Here $\| \cdot \|$ denotes the operator norm.

Due to the block structure of the matrices under consideration, these quantities are more relevant than the classical Stieltjes transforms. The polynomial L is selfadjoint and of degree one, so we can write $L_N = a_0 \otimes \mathbf{1}_N + S_N + T_N$, $\ell_N = a_0 \otimes \mathbf{1} + s + T_N$, where

$$S_N = \sum_{j=1}^{p} a_j \otimes X_j^{(N)}, \ s = \sum_{j=1}^{p} a_j \otimes x_j, \ T_N = \sum_{j=1}^{q} b_j \otimes Y_j^{(N)},$$

and $a_0, \ldots, a_p, b_1, \ldots, b_q$ are Hermitian matrices in $\mathrm{M}_k(\mathbb{C})$. We also need to introduce the $\mathrm{M}_k(\mathbb{C})$-valued Stieltjes transforms G_{T_N} of T_N:

$$G_{T_N}(\Lambda) = (\mathrm{id}_k \otimes \tau_N)\Big[\big(\Lambda \otimes \mathbf{1} - T_N\ \big)^{-1}\Big], \tag{3.12}$$

for all Λ in $\mathrm{M}_k(\mathbb{C})$ such that Im Λ is positive definite.

The families $\mathbf{x}$ and $\mathbf{Y}_N$ being free in $\mathcal{A}$ and $\mathbf{x}$ being a free semicircular system, the theory of matrix-valued non commutative random variables gives us the following equation relating G_{ℓ_N} and G_{T_N}. It encodes the fundamental property of $\mathcal{R}$-transforms, namely the linearity under free convolution.

Step 4. The subordination property for $\mathrm{M}_k(\mathbb{C})$-valued non commutative random variables: *For all Λ in $\mathrm{M}_k(\mathbb{C})$ such that* Im Λ *is positive definite, one has*

$$G_{\ell_N}(\Lambda) = G_{T_N}\Big(\Lambda - a_0 - \mathcal{R}_s\big(G_{\ell_N}(\Lambda)\ \big)\ \Big), \tag{3.13}$$

where

$$\mathcal{R}_s : M \mapsto \sum_{j=1}^{p} a_j M a_j.$$

We show that the fixed point equation implicitly given by (3.13) is, in a certain sense, stable under perturbations. On the other hand, by the asymptotic freeness of $\mathbf{X}_N$ and $\mathbf{Y}_N$, it is expected that Equation (3.13) is asymptotically satisfied when G_{ℓ_N} is replace by G_{L_N}. Since, in order to apply Step 3, we want an uniform control, we make this connection precise by showing the following:

Step 5. The asymptotic subordination property for random matrices: *For all Λ in $\mathrm{M}_k(\mathbb{C})$ such that* Im Λ *is positive definite, one has*

$$G_{L_N}(\Lambda) = G_{T_N}\Big(\Lambda - a_0 - \mathcal{R}_s\big(G_{L_N}(\Lambda)\ \big)\ \Big) + \Theta_N(\Lambda), \tag{3.14}$$

where $\Theta_N(\Lambda)$ satisfies

$$\|\Theta_N(\Lambda)\| \ \leqslant\ \frac{c}{N^2}\,\big\|(\mathrm{Im}\ \Lambda)^{-1}\big\|^5$$

for a constant c and with $\|\cdot\|$ denoting the operator norm.

Organization of the proof

We tackle the different points of the proof described above in the following order:

- **Proof of Step 4.** The precise statement of the subordination property for $\mathrm{M}_k(\mathbb{C})$-valued non commutative random variables is contained in Proposition 4.2 and Proposition 4.3. We highlight in this section the relevance of matrix-valued Stieltjes transforms in a quite general framework.

- **Proof of Step 5.** The asymptotic subordination property for random matrices is stated in Theorem 5.1 in a more general situation. The matrices $\mathbf{Y}_N$ can be random, independent of $\mathbf{X}_N$, satisfying a Poincaré inequality, without assumption on their asymptotic properties. This result is based on the Schwinger-Dyson equation and on the Poincaré inequality satisfied by the law of $\mathbf{X}_N$.

- **Proof of Estimate (3.9).** The estimate will follow easily from the two previous items.

- **Proof of Step 2.** This part is based on $\mathcal{C}^*$-algebra techniques. Step 2 is a consequence of a result due to D. Shlyakhtenko which is stated Theorem A.1 of Appendix A. In a previous version of this article, when we did not know this result, we used the subordination property with $L(\mathbf{x}, \mathbf{Y}_N)$ replaced by $L(\mathbf{x}, \mathbf{y})$ and T_N replaced by its limit in law $t = \sum_{j=1}^q b_j \otimes y_j$. Hence we obtained Theorem 1.6 with additional assumptions on $\mathbf{Y}_N$, notably a uniform rate of convergence of G_{T_N} to the $\mathrm{M}_k(\mathbb{C})$-valued Stieltjes transform of t.

- **Proof of Step 3.** The method is quite standard once Steps 2 and 4 are established. We use a version due to [18] which is based on the use of local concentration inequalities.

4 Proof of Step 4: the subordination property for matrix-valued non commutative random variables

In random matrix theory, a classical method lies in the study of empirical eigenvalue distribution by the analysis of its Stieltjes transform. In many situation, it is shown that this functional satisfies a fixed point equation and a lot of properties of the considered random matrices are deduced from this fact. The purpose of this section is to emphasize that this method can be generalized in the case where the matrices have a macroscopic block structure.

Let $(\mathcal{A}, .^*, \tau, \| \cdot \|)$ be a $\mathcal{C}^*$-probability space with a faithful trace and

$k \geqslant 1$ an integer. The algebra $\mathrm{M}_k(\mathbb{C}) \otimes \mathcal{A}$, formed by the $k \times k$ matrices with coefficients in $\mathcal{A}$, inherits the structure of $\mathcal{C}^*$-probability space with trace $(\tau_k \otimes \tau)$ and norm $\| \cdot \|_{\tau_k \otimes \tau}$ defined by (1.9) with $\tau_k \otimes \tau$ instead of τ. We also shall consider the linear functional $(\mathrm{id}_k \otimes \tau)$, called the partial trace.

For any matrix Λ in $\mathrm{M}_k(\mathbb{C})$ we denote $\mathrm{Im}\ \Lambda$ the Hermitian matrix $\frac{1}{2i}(\Lambda - \Lambda^*)$. We write $\mathrm{Im}\ \Lambda > 0$ whenever the matrix $\mathrm{Im}\ \Lambda$ is positive definite and we denote

$$\mathrm{M}_k(\mathbb{C})^+ = \big\{ \Lambda \in \mathrm{M}_k(\mathbb{C}) \ \big| \ \mathrm{Im}\ \Lambda > 0 \big\}.$$

This lemma will be used throughout this paper. See [19, Lemma 3.1] for a proof.

Lemma 4.1. *Let z in $\mathrm{M}_k(\mathbb{C}) \otimes \mathcal{A}$ be selfadjoint. Then for any $\Lambda \in \mathrm{M}_k(\mathbb{C})^+$, the element $(\Lambda \otimes \mathbf{1} - z)$ is invertible and*

$$\big\| (\Lambda \otimes \mathbf{1} - z)^{-1} \big\|_{\tau_k \otimes \tau} \leqslant \| (\mathrm{Im}\ \Lambda)^{-1} \|. \tag{4.1}$$

On the right hand side, $\| \cdot \|$ denotes the operator norm in $\mathrm{M}_k(\mathbb{C})$.
For a selfadjoint non commutative random variable z in $\mathrm{M}_k(\mathbb{C}) \otimes \mathcal{A}$, its $\mathrm{M}_k(\mathbb{C})$-valued Stieltjes transform is defined by

$$\begin{array}{rccc} G_z : & \mathrm{M}_k(\mathbb{C})^+ & \to & \mathrm{M}_k(\mathbb{C}) \\ & \Lambda & \mapsto & (\mathrm{id}_k \otimes \tau)\Big[\big(\Lambda \otimes \mathbf{1} - z \big)^{-1} \Big]. \end{array}$$

The functional G_z is well defined by Lemma 4.1 and satifies

$$\forall \Lambda \in \mathrm{M}_k(\mathbb{C})^+, \ \| G_z(\Lambda) \| \leqslant \| (\mathrm{Im}\ \Lambda)^{-1} \|.$$

It maps $\mathrm{M}_k(\mathbb{C})^+$ to $\mathrm{M}_k(\mathbb{C})^- = \big\{ \Lambda \in \mathrm{M}_k(\mathbb{C}) \ \big| \ -\Lambda \in \mathrm{M}_k(\mathbb{C})^+ \big\}$ and is analytic (in k^2 complex variables on the open set $\mathrm{M}_k(\mathbb{C})^+ \subset \mathbb{C}^{k^2}$). Moreover, it can be shown (see [38]) that G_z is univalent on a set of the form $U_\delta = \big\{ \Lambda \in \mathrm{M}_k(\mathbb{C})^+ \ \big| \ \| \Lambda^{-1} \| < \delta \ \big\}$ for some $\delta > 0$, and its inverse $G_z^{(-1)}$ in U_δ is analytic on a set of the form $V_\gamma = \big\{ \Lambda \in \mathrm{M}_k(\mathbb{C})^- \ \big| \ \| \Lambda \| < \gamma \big\}$ for some $\gamma > 0$.

The amalgamated $\mathcal{R}$-transform over $\mathrm{M}_k(\mathbb{C})$ of $z \in \mathrm{M}_k(\mathbb{C}) \otimes \mathcal{A}$ is the function $\mathcal{R}_z : G_z(U_\delta) \to \mathrm{M}_k(\mathbb{C})$ given by

$$\mathcal{R}_z(\Lambda) = G_z^{(-1)}(\Lambda) - \Lambda^{-1}, \quad \forall \Lambda \in G_z(U_\delta).$$

The following proposition states the fundamental property of the amalgamated $\mathcal{R}$-transform, namely the subordination property, which is the keystone of our proof of Theorem 1.6.

Proposition 4.2. *Let* $\mathbf{x} = (x_1, \dots, x_p)$ *and* $\mathbf{y} = (y_1, \dots, y_q)$ *be selfadjoint elements of* $\mathcal{A}$ *and let* $\mathbf{a} = (a_1, \dots, a_p)$ *and* $\mathbf{b} = (b_1, \dots, b_q)$ *be* $k \times k$ *Hermitian matrices. Define the elements of* $\mathrm{M}_k(\mathbb{C}) \otimes \mathcal{A}$

$$s = \sum_{j=1}^{p} a_j \otimes x_j, \quad t = \sum_{j=1}^{q} b_j \otimes y_j.$$

Suppose that the families $\mathbf{x}$ *and* $\mathbf{y}$ *are free. Then one has*

1. **Linearity property:** *There is a* γ *such that, in the domain* V_γ*, one has*
$$\mathcal{R}_{s+t} = \mathcal{R}_s + \mathcal{R}_t. \tag{4.2}$$

2. **Subordination property:** *There is* δ *such that, for every* Λ *in* U_δ*, one has*
$$G_{s+t}(\Lambda) = G_t\Big(\Lambda - \mathcal{R}_s\big(\ G_{s+t}(\Lambda)\ \big)\ \Big). \tag{4.3}$$

3. **Semicircular case:** *If* $(x_1, \dots, x_p)$ *is a free semicircular system, then we get*
$$\mathcal{R}_s : \Lambda \mapsto \sum_{j=1}^{p} a_j \Lambda a_j. \tag{4.4}$$

Proof. The linearity property has been shown by Voiculescu in [38] and the $\mathcal{R}$-transform of s has been computed by Lehner in [26]. We deduce easily the subordination property since by Equation (4.2): there exists $\gamma > 0$ such that for all $\Lambda \in V_\gamma$,

$$G_t^{(-1)}(\Lambda) = G_{s+t}^{(-1)}(\Lambda) - \mathcal{R}_s(\Lambda).$$

Then there exists a $\delta > 0$ such that, with $G_{s+t}(\Lambda)$ instead of Λ in the previous equality,

$$G_t^{(-1)}\big(G_{s+t}(\Lambda)\ \big) = \Lambda - \mathcal{R}_s\big(G_{s+t}(\Lambda)\ \big).$$

We compose by $G_t^{(-1)}$ to obtain the result. □

The subordination property plays a key role in our problem: it describes G_{s+t} as a fixed point of a simple function involving s and t separately. Such a fixed point is unique and stable under some perturbation, as it is stated in Proposition 4.3 below. Remark first that, for $\mathcal{R}_s$ given by (4.4), for any Λ in $\mathrm{M}_k(\mathbb{C})^+$ and M in $\mathrm{M}_k(\mathbb{C})^-$,

$$\mathrm{Im}\ \big(\Lambda - \mathcal{R}_s(M)\ \big) = \mathrm{Im}\ \Lambda - \sum_{j=1}^{p} a_j\ \mathrm{Im}\ M\ a_j > 0 \tag{4.5}$$

and
$$\Big\|\Big(\mathrm{Im}\ \big(\Lambda - \mathcal{R}_s(M)\ \big)\ \Big)^{-1}\Big\| \leqslant \|\ (\mathrm{Im}\ \Lambda)^{-1}\|. \tag{4.6}$$
In particular, by analytic continuation, the subordination property holds actually for any $\Lambda \in \mathrm{M}_k(\mathbb{C})^+$ when $\mathbf{x}$ is a free semicircular system.

Proposition 4.3. *Let s and t be as in Proposition 4.2, with $\mathbf{x}$ a free semicircular system.*

1. **Uniqueness of the fixed point:** *For all $\Lambda \in \mathrm{M}_k(\mathbb{C})^+$ such that*
$$\big\|(\mathrm{Im}\ \Lambda)^{-1}\big\| < \sqrt{\sum_{j=1}^p \|a_j\|^2},$$
the following equation in $G_\Lambda \in \mathrm{M}_k(\mathbb{C})^-$,
$$G_\Lambda = G_t\Big(\Lambda - \mathcal{R}_s(\ G_\Lambda\ )\Big), \tag{4.7}$$
admits a unique solution G_Λ in $M_k(\mathbb{C})^-$ given by $G_\Lambda = G_{s+t}(\Lambda)$.

2. **Stability under analytic perturbations:** *Let $G : \Omega \to \mathrm{M}_k(\mathbb{C})^-$ be an analytic function on a simply connected open subset $\Omega \subset \mathrm{M}_k(\mathbb{C})^+$ containing matrices Λ such that $\|(\mathrm{Im}\ \Lambda)^{-1}\|$ is arbitrary small. Suppose that G satisfies: for all $\Lambda \in \Omega$,*
$$G(\Lambda) = G_t\Big(\Lambda - \mathcal{R}_s\big(G(\Lambda)\ \big)\ \Big) + \Theta(\Lambda), \tag{4.8}$$
where the function $\Theta : \Omega \to \mathrm{M}_k(\mathbb{C})$ is analytic and satisfies: there exists $\varepsilon > 0$ such that for all Λ in Ω,
$$\kappa(\Lambda) := \|\Theta(\Lambda)\|\ \|(\mathrm{Im}\ \Lambda)^{-1}\|\ \sum_{j=1}^p \|a_j\|^2 < 1 - \varepsilon.$$
Then one has: $\forall \Lambda \in \Omega$
$$\|G(\Lambda) - G_{s+t}(\Lambda)\| \leqslant \big(1 + c\ \|(\mathrm{Im}\ \Lambda)^{-1}\|^2\ \big)\ \|\Theta(\Lambda)\|, \tag{4.9}$$
where $c = \frac{1}{\varepsilon}\sum_{j=1}^p \|a_j\|^2$.

Proof. 1. Uniqueness of the fixed point:
Fix $\Lambda \in \mathrm{M}_k(\mathbb{C})^+$ such that
$$\big\|(\mathrm{Im}\ \Lambda)^{-1}\big\| < \sqrt{\sum_{j=1}^p \|a_j\|^2}. \tag{4.10}$$

Denote for any M in $\mathrm{M}_k(\mathbb{C})^-$ the matrix $\psi(M) = \Lambda - \mathcal{R}_s(M)$, which is in $\mathrm{M}_k(\mathbb{C})^+$ by (4.5). We show that the function

$$\Phi_\Lambda : M \to G_t\big(\psi(M)\ \big)$$

is a contraction on $\mathrm{M}_k(\mathbb{C})^-$. Remark that Φ_Λ maps $\mathrm{M}_k(\mathbb{C})^-$ into $\mathrm{M}_k(\mathbb{C})^-$. Moreover for all $M, \tilde{M}$ in $\mathrm{M}_k(\mathbb{C})^-$,

$$\begin{aligned}
&\|\Phi_\Lambda(M) - \Phi_\Lambda(\tilde{M})\| \\
&\quad = \left\|(id_k \otimes \tau)\left[\Big(\psi(M)\otimes \mathbf{1} - t\Big)^{-1} - \Big(\psi(\tilde{M})\otimes \mathbf{1} - t\Big)^{-1}\right]\right\| \\
&\quad = \left\|(id_k \otimes \tau)\left[\Big(\psi(M)\otimes \mathbf{1} - t\Big)^{-1}\Big(\sum_{j=1}^p a_j(M-\tilde{M})a_j\Big)\otimes \mathbf{1}_N \right.\right. \\
&\qquad\qquad\qquad \left.\left. \times\Big(\psi(\tilde{M})\otimes \mathbf{1} - t\Big)^{-1}\right]\right\| \\
&\quad \leqslant \left\|\Big(\mathrm{Im}\ \big(\psi(M)\otimes \mathbf{1} - t\big)\Big)^{-1}\right\| \left\|\Big(\mathrm{Im}\ \big(\psi(\tilde{M})\otimes \mathbf{1} - t\big)\Big)^{-1}\right\| \\
&\qquad\qquad\qquad \times \sum_{j=1}^p \|a_j\|^2\ \left\|M - \tilde{M}\right\| \\
&\quad \leqslant \left\|(\mathrm{Im}\ \Lambda)^{-1}\right\|^2 \sum_{j=1}^p \|a_j\|^2\ \|M - \tilde{M}\|.
\end{aligned}$$

Hence the function Φ_Λ is a contraction and by Picard's theorem the fixed point equation $M = \Phi_\Lambda(M)$ admits a unique solution M_Λ on the closed set of $k \times k$ matrices whose imaginary part is non positive semi-definite, which is necessarily G_{s+t} by the subordination property.

2. Stability under analytic perturbations:
We set $\tilde{G} : \Omega \to \mathrm{M}_k(\mathbb{C})^-$ given by: for all $\Lambda \in \Omega$,

$$\tilde{G}(\Lambda) = G(\Lambda) - \Theta(\Lambda) = G_t\Big(\Lambda - \mathcal{R}_s\big(G(\Lambda)\ \big)\ \Big).$$

We set $\tilde{\Lambda} : \Omega \to \mathrm{M}_k(\mathbb{C})$ given by: for all $\Lambda \in \Omega$

$$\tilde{\Lambda}(\Lambda) = \Lambda - \mathcal{R}_s(\Theta(\Lambda)) = \Lambda - \mathcal{R}_s\big(G(\Lambda)\ \big) + \mathcal{R}_s\big(\tilde{G}(\Lambda)\ \big).$$

In the following, we use $\tilde{\Lambda}$ as a shortcut for $\tilde{\Lambda}(\Lambda)$. One has $\tilde{\Lambda} - \mathcal{R}_s\big(\tilde{G}(\Lambda)\ \big) = \Lambda - \mathcal{R}_s\big(G(\Lambda)\ \big)$ which is in $\mathrm{M}_k(\mathbb{C})^+$ by (4.5). Hence we have: for all $\Lambda \in \Omega$,

$$\tilde{G}(\Lambda) \quad = \quad G_t\Big(\tilde{\Lambda} - \mathcal{R}_s\big(\tilde{G}(\Lambda)\ \big)\ \Big). \tag{4.11}$$

We want to estimate $\|(\mathrm{Im}\ \tilde{\Lambda})^{-1}\|$ in terms of $\|(\mathrm{Im}\ \Lambda)^{-1}\|$. For all Λ in Ω, we use the definition of $\tilde{\Lambda}$ and we write:

$$\mathrm{Im}\ \tilde{\Lambda} = \mathrm{Im}\ \Lambda\ \Big(\mathbf{1}_k - (\mathrm{Im}\ \Lambda)^{-1}\mathcal{R}_s\big(\Theta(\Lambda)\ \big)\Big).$$

Remark that

$$\|(\mathrm{Im}\ \Lambda)^{-1}\mathcal{R}_s\big(\Theta(\Lambda)\ \big)\| \leqslant \kappa(\Lambda) = \|\Theta(\Lambda)\|\ \|(\mathrm{Im}\ \Lambda)^{-1}\|\ \sum_{j=1}^{p}\|a_j\|^2 < 1-\varepsilon$$

by assumption. Then $\mathrm{Im}\ \tilde{\Lambda}$ is invertible and one has

$$(\mathrm{Im}\ \tilde{\Lambda})^{-1} = \sum_{\ell\geqslant 0}\Big(\ (\mathrm{Im}\ \Lambda)^{-1}\mathcal{R}_s\big(\Theta(\Lambda)\ \big)\ \Big)^{\ell}\ (\mathrm{Im}\ \Lambda)^{-1}.$$

We then obtain the following estimate

$$\begin{aligned}\|(\mathrm{Im}\ \tilde{\Lambda})^{-1}\| &\leqslant \Big\|\sum_{\ell\geqslant 0}\Big(\ (\mathrm{Im}\ \Lambda)^{-1}\mathcal{R}_s\big(\Theta(\Lambda)\ \big)\ \Big)^{\ell}\ (\mathrm{Im}\ \Lambda)^{-1}\Big\| \\ &\leqslant \frac{1}{1-\kappa(\Lambda)}\|(\mathrm{Im}\ \Lambda)^{-1}\| < \frac{1}{\varepsilon}\|(\mathrm{Im}\ \Lambda)^{-1}\|.\end{aligned}$$

By uniqueness of the fixed point and by (4.11), for all $\Lambda\in\Omega$ such that $\|(\mathrm{Im}\ \Lambda)^{-1}\| < \varepsilon\sqrt{\sum_{j=1}^{p}\|a_j\|^2}$, one has $\tilde{G}(\Lambda) = G_{s+t}(\tilde{\Lambda})$ (such matrices Λ exist by assumption on Ω). But the functions are analytic (in k^2 complex variables) so that the equality extends to Ω. Then for all $\Lambda\in\Omega$,

$$\|G(\Lambda) - G_{s+t}(\Lambda)\| \ \leqslant\ \|G(\Lambda)-\tilde{G}(\Lambda)\| + \|G_{s+t}(\tilde{\Lambda}) - G_{s+t}(\Lambda)\|.$$

For the first term we have by definition of $\tilde{G}$ that $\|G(\Lambda)-\tilde{G}(\Lambda)\| \leqslant \|\Theta(\Lambda)\|$. On the other hand, one has

$$\begin{aligned}&\|G_{s+t}(\Lambda) - G_{s+t}(\tilde{\Lambda})\| \\ &\quad= \Big\|(\mathrm{id}_k\otimes\tau)\big[(\Lambda\otimes\mathbf{1}-s-t)^{-1} - (\tilde{\Lambda}\otimes\mathbf{1}-s-t)^{-1}\ \big]\Big\| \\ &\quad= \Big\|(\mathrm{id}_k\otimes\tau)\big[(\Lambda\otimes\mathbf{1}-s-t)^{-1}(\tilde{\Lambda}\otimes\mathbf{1}-\Lambda\otimes\mathbf{1})(\tilde{\Lambda}\otimes\mathbf{1}-s-t)^{-1}\ \big]\Big\| \\ &\quad\leqslant \|(\Lambda\otimes\mathbf{1}-s-t)^{-1}\|\ \|\tilde{\Lambda}-\Lambda\|\ \|(\tilde{\Lambda}\otimes\mathbf{1}-s-t)^{-1}\| \\ &\quad\leqslant \frac{1}{\varepsilon}\ \big\|\mathcal{R}_s\big(\tilde{G}(\Lambda)\ \big) - \mathcal{R}_s\big(G(\Lambda)\ \big)\ \big\|\ \|(\mathrm{Im}\ \Lambda)^{-1}\|^2 \\ &\quad\leqslant \frac{1}{\varepsilon}\ \sum_{j=1}^{p}\|a_j\|^2\ \|(\mathrm{Im}\ \Lambda)^{-1}\|^2\ \|\Theta(\Lambda)\|.\end{aligned}$$

We then obtain as expected

$$\|G(\Lambda)-G_{s+t}(\Lambda)\| \leqslant \Big(1+\frac{1}{\varepsilon}\ \sum_{j=1}^{p}\|a_j\|^2\ \|(\mathrm{Im}\ \Lambda)^{-1}\|^2\Big)\ \|\Theta(\Lambda)\|.$$

□

5 Proof of Step 5: the asymptotic subordination property for random matrices

The purpose of this section is to prove Theorem 5.1 below, where it is stated that, for N fixed, the matrix-valued Stieltjes transforms of certain random matrices satisfy an asymptotic subordination property i.e. an equation as in (4.8). This result is independent with the previous part and does not involve the language of free probability.

Let $\mathbf{X}_N = (X_1^{(N)}, \dots, X_p^{(N)})$ be a family of independent, normalized $N \times N$ matrices of the GUE and $\mathbf{Y}_N = (Y_1^{(N)}, \dots, Y_q^{(N)})$ be a family of $N \times N$ random Hermitian matrices, independent of $\mathbf{X}_N$. We fix an integer $k \geqslant 1$ and Hermitian matrices $a_0, \dots, a_p, b_1, \dots, b_q \in \mathrm{M}_k(\mathbb{C})$. We set S_N and T_N the $kN \times kN$ block matrices

$$S_N = \sum_{j=1}^{p} a_j \otimes X_j^{(N)}, \qquad T_N = \sum_{j=1}^{q} b_j \otimes Y_j^{(N)}.$$

Define the $\mathrm{M}_k(\mathbb{C})$-valued Stieltjes transforms of $S_N + T_N$ and T_N: for all $\Lambda \in \mathrm{M}_k(\mathbb{C})^+ = \big\{ \Lambda \in \mathrm{M}_k(\mathbb{C}) \ \big| \ \mathrm{Im}\, \Lambda > 0 \big\}$,

$$\begin{aligned} G_{S_N+T_N}(\Lambda) &= \mathbb{E}\Big[(\mathrm{id}_k \otimes \tau_N)\Big[\big(\Lambda \otimes \mathbf{1}_N - S_N - T_N\big)^{-1} \Big] \Big], \\ G_{T_N}(\Lambda) &= \mathbb{E}\Big[(\mathrm{id}_k \otimes \tau_N)\Big[\big(\Lambda \otimes \mathbf{1}_N - T_N\big)^{-1} \Big] \Big]. \end{aligned}$$

We denote by $\mathcal{R}_s$ the functional

$$\begin{aligned} \mathcal{R}_s : \ \mathrm{M}_k(\mathbb{C}) &\to \ \mathrm{M}_k(\mathbb{C}) \\ M &\mapsto \ \sum_{j=1}^{p} a_j \ M \ a_j. \end{aligned}$$

Theorem 5.1 (Asymptotic subordination property)**.** *Assume that there exists $\sigma \geqslant 1$ such that the joint law of the entries of the matrices $\mathbf{Y}_N$ satisfies a Poincaré inequality with constant σ/N, i.e. for any $f : \mathbb{R}^{2qN^2} \to \mathbb{C}$ function of the entries of q matrices, of class $\mathcal{C}^1$ and such that $\mathbb{E}\Big[\ |f(\mathbf{Y}_N)|^2 \ \Big] < \infty$, one has*

$$\mathbb{V}ar\Big(f(\mathbf{Y}_N) \ \Big) \leqslant \frac{\sigma}{N} \ \mathbb{E}\Big[\|\nabla f(\mathbf{Y}_N)\|^2 \Big], \tag{5.1}$$

where ∇f denotes the gradient of f, $\mathbb{V}ar$ denotes the variance, $\mathbb{V}ar(\ x \ ) = \mathbb{E}\big[\ \big| \ x - \mathbb{E}[\ x \ ] \ \big|^2 \big]$.

Then for any $\Lambda \in \mathrm{M}_k(\mathbb{C})^+$*, the Stieltjes transforms* $G_{S_N+T_N}$ *and* G_{T_N} *satisfy*

$$G_{S_N+T_N}(\Lambda) = G_{T_N}\Big(\Lambda - \mathcal{R}_s\big(G_{S_N+T_N}(\Lambda)\ \big)\ \Big) + \Theta_N(\Lambda), \tag{5.2}$$

where Θ *is analytic* $\mathrm{M}_k(\mathbb{C})^+ \to \mathrm{M}_k(\mathbb{C})$ *and satisfies*

$$\|\Theta_N(\Lambda)\| \ \leqslant\ \frac{c}{N^2}\left\|(\mathrm{Im}\ \Lambda)^{-1}\right\|^5,$$

with $c = 2k^{9/2}\sigma\sum_{j=1}^p \|a_j\|^2\Big(\sum_{j=1}^p \|a_j\| + \sum_{j=1}^q \|b_j\|\Big)^2$*,* $\|\cdot\|$ *denoting the operator norm in* $M_k(\mathbb{C})$*.*

The proof of Theorem 5.1 is carried out in two steps.

- In Section 5.1 we state a mean Schwinger-Dyson equation for random Stieltjes transforms (Proposition 5.2).
- In Section 5.2 we deduce from Proposition 5.2 a Schwinger-Dyson equation for mean Stieltjes transforms (Proposition 5.3).

Theorem 5.1 is a direct consequence of Proposition 5.3 as it is shown in Section 5.3.

5.1 Mean Schwinger-Dyson equation for random Stieltjes transforms

For Λ, Γ in $\mathrm{M}_k(\mathbb{C})^+$, define the elements of $\mathrm{M}_k(\mathbb{C}) \otimes \mathrm{M}_N(\mathbb{C})$

$$\begin{aligned} h_{S_N+T_N}(\Lambda) &= (\Lambda \otimes \mathbf{1}_N - S_N - T_N)^{-1}, \\ h_{T_N}(\Gamma) &= (\Gamma \otimes \mathbf{1}_N - T_N)^{-1}, \end{aligned}$$

and $H_{S_N+T_N}(\Lambda) = (\mathrm{id}_k \otimes \tau_N)\Big[h_{S_N+T_N}(\Lambda)\Big]$, $H_{T_N}(\Lambda) = (\mathrm{id}_k \otimes \tau_N)\Big[h_{T_N}(\Lambda)\Big]$.

Proposition 5.2 (Mean Schwinger-Dyson equation for random Stieltjes transforms)**.** *For all* $\Lambda, \Gamma \in \mathrm{M}_k(\mathbb{C})^+$ *we have*

$$\begin{aligned} 0 = \mathbb{E}\Big[& H_{S_N+T_N}(\Lambda) - H_{T_N}(\Gamma) \\ & -(\mathrm{id}_k \otimes \tau_N)\Big[h_{T_N}(\Gamma)\Big(\mathcal{R}_s\big(H_{S_N+T_N}(\Lambda)\ \big) - \Lambda + \Gamma\Big) \otimes \mathbf{1}_N\ h_{S_N+T_N}(\Lambda)\Big]\Big]. \end{aligned} \tag{5.3}$$

The result is a consequence of integration by parts for Gaussian densities and of the formula for the differentiation of the inverse of a matrix. If $(g_1, \dots, g_N)$ are independent identically distributed centered real Gaussian variables with variance σ^2 and $F : \mathbb{R}^N \to \mathbb{C}$ a differentiable map such that F and its partial derivatives are polynomially bounded, one has for $i = 1, \dots, N$

$$\mathbb{E}\Big[\ g_i\ F(g_1, \dots, g_N)\ \Big] = \sigma^2 \mathbb{E}\Big[\frac{\partial F}{\partial x_i}(g_1, \dots, g_N)\ \Big].$$

This induces an analogue formula for independent matrices of the GUE, called the Schwinger-Dyson equation, where the Hermitian symmetry of the matrices plays a key role. For instance, if P is a monomial in p non commutative indeterminates, one has for $i = 1, \dots, p$,

$$\mathbb{E}\Big[\tau_N\Big[\ X_i^{(N)}\ P(\mathbf{X}_N)\ \Big]\ \Big] = \sum_{P = Lx_iR} \mathbb{E}\Big[\tau_N\Big[L(\mathbf{X}_N)\ \Big]\ \tau_N\Big[R(\mathbf{X}_N)\ \Big]\Big],$$

the sum over all decompositions $P = Lx_iR$ for L and R monomials being viewed as the partial derivative.

This formula has an analogue for analytical maps instead of polynomials. The case of the function $\mathbf{X}_N \mapsto (\Lambda \otimes \mathbf{1}_N - S_N)^{-1}$ is investigated in details in [19, Formula (3.9)], our proof is obtained by minor modifications.

Proof. Denote by $(\epsilon_{m,n})_{m,n=1,\dots,N}$ the canonical basis of $\mathrm{M}_N(\mathbb{C})$. By [19, Formula (3.9)] with minor modification, we get the following: for all Λ, Γ in $\mathrm{M}_k(\mathbb{C})^+$ and $j = 1, \dots, p$,

$$\begin{aligned}
&\mathbb{E}\Big[(\mathbf{1}_k \otimes X_j^{(N)})(\Lambda \otimes \mathbf{1}_N - S_N - T_N)^{-1}\ \Big|\ T_N\Big] \\
&\quad = \mathbb{E}\Big[\frac{1}{N}\sum_{m,n=1}^{N} (\mathbf{1}_k \otimes \epsilon_{m,n})(\Lambda \otimes \mathbf{1}_N - S_N - T_N)^{-1} \\
&\qquad\qquad \times (a_j \otimes \epsilon_{n,m})(\Lambda \otimes \mathbf{1}_N - S_N - T_N)^{-1}\ \Big|\ T_N\Big].
\end{aligned}$$

In these equations, $\mathbb{E}[\cdot|T_N]$ stands for the conditional expectation with respect to T_N. Furthermore, for any M in $\mathrm{M}_k(\mathbb{C}) \otimes \mathrm{M}_N(\mathbb{C})$, one has

$$\frac{1}{N}\sum_{m,n=1}^{N} (\mathbf{1}_k \otimes \epsilon_{m,n})\ M\ (\mathbf{1}_k \otimes \epsilon_{n,m}) = (\mathrm{id}_k \otimes \tau_N)[\ M\ ] \otimes \mathbf{1}_N.$$

Indeed the formula is clear if M is of the form $M = \tilde{M} \otimes \epsilon_{u,v}$ and extends by linearity. In particular, with $M = (\Lambda \otimes \mathbf{1}_N - S_N - T_N)^{-1}(a_j \otimes \mathbf{1}_N)$, we obtain that: for all Λ, Γ in $\mathrm{M}_k(\mathbb{C})^+$ and $j = 1, \ldots, p$,

$$\begin{aligned}
&\mathbb{E}\Big[(a_j \otimes X_j^{(N)})(\Lambda \otimes \mathbf{1}_N - S_N - T_N)^{-1} \ \Big| \ T_N\Big] \\
&\quad = \mathbb{E}\Big[(a_j \otimes \mathbf{1}_N)\Big((\mathrm{id}_k \otimes \tau_N)\Big[(\Lambda \otimes \mathbf{1}_N - S_N - T_N)^{-1}\Big] a_j \otimes \mathbf{1}_N\Big) \\
&\qquad\quad \times (\Lambda \otimes \mathbf{1}_N - S_N - T_N)^{-1} \ \Big| \ T_N\Big] \\
&\quad = \mathbb{E}\Big[\big(a_j H_{S_N+T_N} a_j \otimes \mathbf{1}_N\big) \ h_{S_N+T_N} \ \Big| \ T_N\Big].
\end{aligned}$$

Recall that $S_N = \sum_{j=1}^p a_j \otimes X_j^{(N)}$ and $\mathcal{R}_s : M \mapsto \sum_{j=1}^p a_j M a_j$, so that for all Λ, Γ in $\mathrm{M}_k(\mathbb{C})^+$, one has

$$\begin{aligned}
&\mathbb{E}\Big[(\Gamma \otimes \mathbf{1}_N - T_N)^{-1} \ S_N \ (\Lambda \otimes \mathbf{1}_N - S_N - T_N)^{-1}\Big] \\
&\quad = \mathbb{E}\Big[\ (\Gamma \otimes \mathbf{1}_N - T_N)^{-1} \ \sum_{j=1}^p \mathbb{E}\Big[(a_j \otimes X_j^{(N)}) \qquad (5.4)\\
&\qquad\quad \times (\Lambda \otimes \mathbf{1}_N - S_N - T_N)^{-1} \ \Big| \ T_N\Big] \ \Big] \\
&\quad = \mathbb{E}\Big[\ h_{T_N}(\Gamma) \ \mathbb{E}\Big[\big(\sum_{j=1}^p a_j H_{S_N+T_N}(\Lambda) a_j \otimes \mathbf{1}_N\big) \ h_{S_N+T_N}(\Lambda) \ \Big| \ T_N\Big] \\
&\quad = \mathbb{E}\Big[h_{T_N}(\Gamma) \ \Big(\mathcal{R}_s\big(H_{S_N+T_N}(\Lambda) \ \big) \otimes \mathbf{1}_N\Big) \ h_{S_N+T_N}(\Lambda)\Big]. \qquad (5.5)
\end{aligned}$$

We take the partial trace in Equation (5.5) to obtain:

$$\begin{aligned}
&\mathbb{E}\Big[(\mathrm{id}_k \otimes \tau_N)\Big[h_{T_N}(\Gamma) \ S_N \ h_{S_N+T_N}(\Lambda)\Big]\Big] \qquad (5.6)\\
&\quad = \ \mathbb{E}\Big[(\mathrm{id}_k \otimes \tau_N)\Big[h_{T_N}(\Gamma) \ \Big(\mathcal{R}_s\big(H_{S_N+T_N}(\Lambda) \ \big) \otimes \mathbf{1}_N\Big) \ h_{S_N+T_N}(\Lambda)\Big]\Big].
\end{aligned}$$

We now rewrite S_N as follow:

$$S_N = (\Lambda - \Gamma) \otimes \mathbf{1}_N + (\Gamma \otimes \mathbf{1}_N - T_N) - (\Lambda \otimes \mathbf{1}_N - S_N - T_N).$$

Re-injecting this expression in the left hand side of Equation (5.6), one gets Equation (5.3):

$$
\begin{aligned}
&\mathbb{E}\Big[(\mathrm{id}_k \otimes \tau_N)\Big[h_{T_N}(\Gamma)\ \Big(\mathcal{R}_s\big(H_{S_N+T_N}(\Lambda)\ \big)\otimes \mathbf{1}_N\Big)\ h_{S_N+T_N}(\Lambda)\Big]\Big] \\
&\quad = \ \mathbb{E}\Big[(\mathrm{id}_k \otimes \tau_N)\Big[h_{T_N}(\Gamma)\ (\Lambda-\Gamma)\otimes \mathbf{1}_N h_{S_N+T_N}(\Lambda) \\
&\qquad\qquad + h_{S_N+T_N}(\Lambda) - h_{T_N}(\Gamma)\Big]\Big] \\
&\quad = \ \mathbb{E}\Big[(\mathrm{id}_k \otimes \tau_N)\Big[\ h_{T_N}(\Gamma)\ \Big((\Lambda-\Gamma)\otimes \mathbf{1}_N\Big)\ h_{S_N+T_N}(\Lambda)\Big] \\
&\qquad\qquad + H_{S_N+T_N}(\Lambda)\ -\ H_{T_N}(\Gamma)\Big].
\end{aligned}
$$

□

5.2 Schwinger-Dyson equation for mean Stieltjes transforms

We use the concentration properties of the law of $(\mathbf{X}_N, \mathbf{Y}_N)$ to get from Equation (5.3) a relation between $G_{S_N+T_N}$ and G_{T_N}. We define the centered version of $H_{S_N+T_N}$ by: for all Λ in $\mathrm{M}_k(\mathbb{C})^+$,

$$K_{S_N+T_N}(\Lambda) \ = \ H_{S_N+T_N}(\Lambda) - G_{S_N+T_N}(\Lambda), \text{ in } \mathrm{M}_k(\mathbb{C}). \tag{5.7}$$

We introduce the random linear map

$$
\begin{array}{rccc}
l_{N,\Lambda,\Gamma}: & \mathrm{M}_k(\mathbb{C})\otimes \mathrm{M}_N(\mathbb{C}) & \to & \mathrm{M}_k(\mathbb{C})\otimes \mathrm{M}_N(\mathbb{C}) \\
& M & \mapsto & h_{T_N}(\Gamma)\ M\ h_{S_N+T_N}(\Lambda)
\end{array} \tag{5.8}
$$

and its mean

$$L_{N,\Lambda,\Gamma} : M \mapsto \mathbb{E}\Big[l_{N,\Lambda,\Gamma}(M)\ \Big]. \tag{5.9}$$

Remark that if M is a random matrix, then

$$L_{N,\Lambda,\Gamma}(M) = \mathbb{E}\big[h_{\tilde{T}_N}(\Gamma)\ M\ h_{\tilde{S}_N+\tilde{T}_N}(\Lambda)\big| M\big],$$

where $(\tilde{S}_N + \tilde{T}_N)$ is an independent copy of $(S_N + T_N)$ independent of M.

Proposition 5.3 (Schwinger-Dyson equation for mean Stieltjes transforms). *For all Λ, Γ in $\mathrm{M}_k(\mathbb{C})^+$, one has*

$$
\begin{aligned}
&G_{S_N+T_N}(\Lambda) - G_{T_N}(\Gamma) \\
&-(\mathrm{id}_k\otimes\tau_N)\Big[L_{N,\Lambda,\Gamma}\Big(\ \Big(\mathcal{R}_s\big(G_{S_N+T_N}(\Lambda)\ \big) - \Lambda + \Gamma\Big)\otimes \mathbf{1}_N\ \Big)\ \Big] = \Theta_N(\Lambda,\Gamma),
\end{aligned} \tag{5.10}
$$

where

$$\Theta_N(\Lambda,\Gamma)=\mathbb{E}\Big[(\mathrm{id}_k\otimes\tau_N)\Big[\ (l_{N,\Lambda,\Gamma}-L_{N,\Lambda,\Gamma})\Big(\mathcal{R}_s\big(K_{S_N+T_N}(\Lambda)\ \big)\otimes\mathbf{1}_N\Big)\ \Big]\Big] \tag{5.11}$$

is controlled in operator norm by the following estimate:

$$\|\Theta_N(\Lambda,\Gamma)\|\leqslant\frac{c}{N^2}\left\|(\mathrm{Im}\ \Gamma)^{-1}\right\|\left\|(\mathrm{Im}\ \Lambda)^{-1}\right\|^3\ \Big(\|(\mathrm{Im}\ \Gamma)^{-1}\|+\|(\mathrm{Im}\ \Lambda)^{-1}\|\Big), \tag{5.12}$$

with $c=k^{9/2}\sigma\sum_{j=1}^p\|a_j\|^2\Big(\sum_{j=1}^p\|a_j\|+\sum_{j=1}^q\|b_j\|\Big)^2$.

Proof of Proposition 5.3. We first expand $\Theta_N(\Lambda,\Gamma)$: for all Λ,Γ in $\mathrm{M}_k(\mathbb{C})^+$, we have

$$\begin{aligned}\Theta_N(\Lambda,\Gamma)\ &:=\ \mathbb{E}\Big[(\mathrm{id}_k\otimes\tau_N)\Big[\ (l_{N,\Lambda,\Gamma}-L_{N,\Lambda,\Gamma})\\ &\qquad\times\Big(\mathcal{R}_s\big(H_{S_N+T_N}(\Lambda)-G_{S_N+T_N}(\Lambda)\ \big)\otimes\mathbf{1}_N\Big)\ \Big]\Big]\\ &=\ \mathbb{E}\Big[(\mathrm{id}_k\otimes\tau_N)\Big[l_{N,\Lambda,\Gamma}\Big(\mathcal{R}_s\big(H_{S_N+T_N}(\Lambda)\ \big)\otimes\mathbf{1}_N\Big)\ \Big]\Big]\\ &\qquad-(\mathrm{id}_k\otimes\tau_N)\Big[L_{N,\Lambda,\Gamma}\Big(\mathcal{R}_s\big(G_{S_N+T_N}(\Lambda)\ \big)\otimes\mathbf{1}_N\Big)\ \Big].\end{aligned}$$

By Equation (5.3), we get the following:

$$\begin{aligned}&\mathbb{E}\Big[(\mathrm{id}_k\otimes\tau_N)\Big[l_{N,\Lambda,\Gamma}\Big(\mathcal{R}_s\big(H_{S_N+T_N}(\Lambda)\ \big)\otimes\mathbf{1}_N\Big)\Big]\Big]\\ &=\ \mathbb{E}\Big[(\mathrm{id}_k\otimes\tau_N)\Big[l_{N,\Lambda,\Gamma}\Big(\ (\Lambda-\Gamma)\otimes\mathbf{1}_N\Big)\ \Big]-H_{T_N}(\Gamma)+H_{S_N+T_N}(\Lambda)\Big]\\ &=\ (\mathrm{id}_k\otimes\tau_N)\Big[L_{N,\Lambda,\Gamma}\Big(\ (\Lambda-\Gamma)\otimes\mathbf{1}_N\Big)\ \Big]-G_{T_N}(\Gamma)+G_{S_N+T_N}(\Lambda),\end{aligned}$$

which gives Equation (5.10).

We use the Poincaré inequality to control the operator norm of Θ_N: if $(g_1,\dots,g_K)$ are independent identically distributed centered real Gaussian variables with variance v^2 and F is a differentiable map $\mathbb{R}^K\to\mathbb{C}$ such that F and its partial derivatives are polynomially bounded, then (see [11, Theorem 2.1])

$$\mathbb{V}\mathrm{ar}\Big(F(g_1,\dots,g_K)\ \Big)\leqslant v^2\mathbb{E}\Big[\ \|\nabla F(g_1,\dots,g_K)\ \|^2\ \Big].$$

The Poincaré inequality is compatible with tensor product and then such a formula is still valid when F is a function of the matrices $\mathbf{X_N}$ and $\mathbf{Y_N}$ with $v^2 = \frac{\sigma}{N}$.

We will often deal with matrices of size $k \times k$. Since the integer k is fixed, we can use intensively the equivalence of norms, the constants appearing will not modify the order of convergence. For any integer K, we denote the Euclidean norm of a $K \times K$ matrix $A = (a_{m,n})_{1 \leqslant m,n \leqslant K}$ by

$$\|A\|_e = \sqrt{\sum_{m,n=1}^{K} |a_{m,n}|^2},$$

and its infinity norm by

$$\|A\|_\infty = \max_{m,n=1,\dots,K} |a_{m,n}|.$$

Recall that if A, B are $K \times K$ matrices we have the following inequalities

$$\|A\| \leqslant \|A\|_e \leqslant \sqrt{K}\|A\|, \tag{5.13}$$
$$\|A\| \leqslant \sqrt{K}\|A\|_\infty \leqslant \sqrt{K}\|A\|_e, \tag{5.14}$$
$$\|AB\| \leqslant \|A\|_e \; \|B\|. \tag{5.15}$$

When A is in $\mathrm{M}_k(\mathbb{C}) \otimes \mathrm{M}_N(\mathbb{C})$, its Euclidean norm is defined by considering A as a $kN \times kN$ matrix. In the following we will write an element Z of $\mathrm{M}_k(\mathbb{C}) \otimes \mathrm{M}_N(\mathbb{C})$

$$\begin{aligned} Z &= \sum_{m,n=1}^{N} \sum_{u,v=1}^{k} Z_{u,v}^{m,n} \; \epsilon_{u,v} \otimes \epsilon_{m,n} = \sum_{m,n=1}^{N} Z^{(m,n)} \otimes \epsilon_{m,n} \qquad (5.16) \\ &= \sum_{u,v=1}^{k} \epsilon_{u,v} \otimes Z_{(u,v)}, \end{aligned}$$

where for $m, n = 1, \dots, N$ and $u, v = 1, \dots, k$, $Z_{u,v}^{m,n}$ is a complex number, $Z^{(m,n)}$ is a $k \times k$ matrix, and $Z_{(u,v)}$ is a $N \times N$ matrix; we use the same notation for the canonical bases of $\mathrm{M}_k(\mathbb{C})$ and $\mathrm{M}_N(\mathbb{C})$.
We fix Λ, Γ in $\mathrm{M}_k(\mathbb{C})^+$ until the end of this proof and we use for convenience the following notations:

$$\begin{aligned} M_N &= \mathcal{R}_s\big(K_{S_N+T_N}(\Lambda)\ \big) \\ h_N^{(1)} &= h_{S_N+T_N}(\Lambda) \\ h_N^{(2)} &= h_{T_N}(\Gamma) \\ l_N &= l_{N,\Lambda,\Gamma} \\ L_N &= L_{N,\Lambda,\Gamma}. \end{aligned}$$

We consider $(\tilde{h}_N^{(1)}, \tilde{h}_N^{(2)})$ an independent copy of $(h_N^{(1)}, h_N^{(2)})$, independent of $\mathbf{X}_N$ and $\mathbf{Y}_N$ (and hence of all the random variables considered). Recall that by definitions (5.8) and (5.9): for all Λ, Γ in $\mathrm{M}_k(\mathbb{C})^+$, we have

$$\begin{array}{rcl} l_N & : A \in \mathrm{M}_k(\mathbb{C}) \mapsto & h_N^{(2)} \; A \; h_N^{(1)} \in \mathrm{M}_k(\mathbb{C}), \\ L_N & : A \in \mathrm{M}_k(\mathbb{C}) \mapsto & \mathbb{E}\Big[l_N(A) \Big] \in \mathrm{M}_k(\mathbb{C}). \end{array}$$

With the notations of (5.16) we have

$$\begin{aligned} &(\mathrm{id}_k \otimes \tau_N)\Big[(l_N - L_N) \; (M_N \otimes \mathbf{1}_N)\Big] \\ &= \; (\mathrm{id}_k \otimes \tau_N)\big[h_N^{(2)} \; (M_N \otimes \mathbf{1}_N) \; h_N^{(1)}\big] \\ &\quad - \mathbb{E}\Big[(\mathrm{id}_k \otimes \tau_N)\big[\tilde{h}_N^{(2)} \; (M_N \otimes \mathbf{1}_N) \; \tilde{h}_N^{(1)}\big] \; \Big| \; M_N\Big] \\ &= \; \frac{1}{N} \sum_{m,n=1}^{N} \Bigg[\Big(h_N^{(2)}\Big)^{(m,n)} M_N \; \Big(h_N^{(1)}\Big)^{(n,m)} \\ &\quad - \mathbb{E}\Big[\Big(\tilde{h}_N^{(2)}\Big)^{(m,n)} M_N \; \Big(\tilde{h}_N^{(1)}\Big)^{(n,m)} \; \Big| \; M_N\Big] \Bigg]. \end{aligned}$$

To estimate the operator norm of Θ_N we use the domination by the infinity norm (5.14) in order to split the contributions due to M_N and due to $l_N - L_N$: we get

$$\begin{aligned} \|\Theta_N(\Lambda, \Gamma)\| &= \Big\| \mathbb{E}\Big[(\mathrm{id}_k \otimes \tau_N) \left[(l_N - L_N) \; (M_N \otimes \mathbf{1}_N)\right]\Big] \Big\| \\ &\leqslant \; \sqrt{k} \Bigg\| \mathbb{E}\Bigg[\frac{1}{N} \sum_{m,n=1}^{N} \Big(h_N^{(2)}\Big)^{(m,n)} M_N \; \Big(h_N^{(1)}\Big)^{(n,m)} \\ &\qquad\qquad - \mathbb{E}\Big[\Big(\tilde{h}_N^{(2)}\Big)^{(m,n)} M_N \; \Big(\tilde{h}_N^{(1)}\Big)^{(n,m)} \; \Big| \; M_N\Big] \Bigg] \Bigg\|_\infty \\ &\leqslant \; k^{5/2} \max_{\substack{1 \leqslant u,v \leqslant k \\ 1 \leqslant u',v' \leqslant k}} \Bigg| \mathbb{E}\Bigg[(M_N)_{u',v'} \times \frac{1}{N} \sum_{m,n=1}^{N} \Big(h_N^{(2)}\Big)^{m,n}_{u,u'} \Big(h_N^{(1)}\Big)^{n,m}_{v',v} \\ &\qquad\qquad - \mathbb{E}\Big[\Big(h_N^{(2)}\Big)^{m,n}_{u,u'} \Big(h_N^{(1)}\Big)^{n,m}_{v',v} \Big] \Bigg] \Bigg| \\ &\leqslant \; k^{5/2} \max_{u,v,u',v'} \mathbb{E}\Bigg[|(M_N)_{u',v'}| \times \Big| \tau_N\Big[\Big(h_N^{(1,2)}\Big)_{\substack{u,v \\ u',v'}} \Big] - \mathbb{E}\Big[\tau_N\Big[\Big(h_N^{(1,2)}\Big)_{\substack{u,v \\ u',v'}} \Big] \Big] \Big| \Bigg] \\ &\leqslant \; k^{5/2} \max_{u,v,u',v'} \mathbb{E}\Bigg[|(M_N)_{u',v'}| \times \Big| \tau_N\Big[\Big(k_N^{(1,2)}\Big)_{\substack{u,v \\ u',v'}} \Big] \Big| \Bigg], \end{aligned}$$

where we have denoted the $N \times N$ matrices

$$\begin{aligned}
\big(h_N^{(1,2)}\big)_{\substack{u,v\\u',v'}} &= \big(h_N^{(2)}\big)_{(u,u')}\big(h_N^{(1)}\big)_{(v',v)},\\
\big(k_N^{(1,2)}\big)_{\substack{u,v\\u',v'}} &= \big(h_N^{(1,2)}\big)_{\substack{u,v\\u',v'}} - \mathbb{E}\Big[\big(h_N^{(1,2)}\big)_{\substack{u,v\\u',v'}}\Big].
\end{aligned}$$

Remark that by (5.15), for $u', v' = 1, \dots, k$,

$$\begin{aligned}
|(M_N)_{u',v'}| &= \Big|\Big(\sum_{j=1}^{p} a_j K_{S_N+T_N}(\Lambda) a_j\Big)_{u',v'}\Big|\\
&\leqslant \Big\|\sum_{j=1}^{p} a_j K_{S_N+T_N}(\Lambda) a_j\Big\|_e\\
&\leqslant \sum_{j=1}^{p} \|a_j\|^2\ \|K_{S_N+T_N}(\Lambda)\|_e.
\end{aligned}$$

Then by Cauchy-Schwarz inequality we get:

$$\begin{aligned}
\|\Theta_N(\Lambda,\Gamma)\| &\leqslant k^{5/2}\sum_{j=1}^{p}\|a_j\|^2\Big(\mathbb{E}\big[\|K_{S_N+T_N}(\Lambda)\|_e^2\big]\\
&\qquad\qquad \times \max_{u,v,u',v'}\mathbb{E}\Big[\Big|\tau_N\big[\big(k_N^{(1,2)}\big)_{\substack{u,v\\u',v'}}\big]\ \Big|^2\ \Big]\ \Big)^{1/2}\\
&\leqslant k^{5/2}\sum_{j=1}^{p}\|a_j\|^2\Big(\sum_{u,v=1}^{k}\mathbb{V}\mathrm{ar}\big(H_{S_N+T_N}(\Lambda)\big)_{u,v} \qquad (5.17)\\
&\qquad\qquad \times \max_{u,v,u',v'}\mathbb{V}\mathrm{ar}\Big(\tau_N\big[\big(h_N^{(1,2)}\big)_{\substack{u,v\\u',v'}}\big]\ \Big)\ \Big)^{1/2}.
\end{aligned}$$

One is reduced to the study of variances of random variables. To use the Poincaré inequality, we write for $u, v, u', v' = 1, \dots, k$,

$$\begin{aligned}
\big(H_{S_N+T_N}(\Lambda)\ \big)_{u,v} &= F_{u,v}^{(1)}\big(\mathbf{X}_N, \mathbf{Y}_N\big),\\
\tau_N\Big[\big(h_N^{(1,2)}\big)_{\substack{u,v\\u',v'}}\Big] &= F_{u,v,u',v'}^{(2)}\big(\mathbf{X}_N, \mathbf{Y}_N\big),
\end{aligned}$$

where for all selfadjoint matrices $\mathbf{A} = (A_1, \dots, A_p)$ in $\mathrm{M}_N(\mathbb{C})$, for all $\mathbf{B} = (B_1, \dots, B_q)$ in $\mathrm{M}_N(\mathbb{C})$ and with $\tilde{S}_N = \sum_{j=1}^{p} a_j \otimes A_j$, $\tilde{T}_N = \sum_{j=1}^{q} b_j \otimes B_j$, we have set

$$\begin{aligned}
F_{u,v}^{(1)}(\mathbf{A},\mathbf{B}) &= \Big((\mathrm{id}_k \otimes \tau_N)\big[(\Lambda \otimes \mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1}\big]\ \Big)_{u,v}\\
&= \frac{1}{N}(\mathrm{Tr}_k \otimes \mathrm{Tr}_N)\big[(\epsilon_{v,u} \otimes \mathbf{1}_N)(\Lambda \otimes \mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1}\big],
\end{aligned}$$

$$
\begin{aligned}
&F^{(2)}_{u,v,u',v'}(\mathbf{A},\mathbf{B})\\
&= \tau_N\Big[\Big((\Lambda\otimes \mathbf{1}_N - \tilde S_N - \tilde T_N)^{-1}\Big)_{(u,u')}\Big((\Gamma\otimes \mathbf{1}_N - \tilde T_N)^{-1}\Big)_{(v',v)}\Big]\\
&= \frac{1}{N}(\mathrm{Tr}_k\otimes \mathrm{Tr}_N)\Big[(\epsilon_{v,u}\otimes \mathbf{1}_N)(\Gamma\otimes \mathbf{1}_N - \tilde T_N)^{-1}\\
&\qquad \times(\epsilon_{u',v'}\otimes \mathbf{1}_N)\,(\Lambda\otimes \mathbf{1}_N - \tilde S_N - \tilde T_N)^{-1}\Big].
\end{aligned}
$$

The functions and their partial derivatives are bounded (see [19, Lemma 4.6] with minor modifications), so that, since the law of $(\mathbf{X}_N,\mathbf{Y}_N)$ satisfies a Poincaré inequality with constant $\frac{\sigma}{N}$, one has

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}\Big(H_{S_N+T_N}(\Lambda)\Big)_{u,v} &\leqslant \frac{\sigma}{N}\mathbb{E}\Big[\big\|\nabla\, F^{(1)}_{u,v}(\mathbf{X}_N,\mathbf{Y}_N)\big\|^2\Big],\\
\mathbb{V}\mathrm{ar}\Big(\tau_N\big[\big(h_N^{(1,2)}\big)_{\substack{u,v\\u',v'}}\big]\Big) &\leqslant \frac{\sigma}{N}\mathbb{E}\Big[\big\|\nabla\, F^{(2)}_{u,v,u',v'}(\mathbf{X}_N,\mathbf{Y}_N)\big\|^2\Big].
\end{aligned}
$$

We define the set $\mathcal{W}$ of families $(\mathbf{V},\mathbf{W})$ of $N\times N$ Hermitian matrices, with $\mathbf{V}=(V_1,\ldots,V_p)$, $\mathbf{W}=(W_1,\ldots,W_q)$, of unit Euclidean norm in $\mathbf{R}^{(p+q)N^2}$. Then we have

$$
\begin{aligned}
&\mathbb{V}\mathrm{ar}\Big(H_{S_N+T_N}(\Lambda)\Big)_{u,v}\\
&\quad\leqslant \frac{\sigma}{N}\mathbb{E}\Big[\max_{(\mathbf{V},\mathbf{W})\in\mathcal{W}}\Big|\frac{d}{dt}_{|t=0}\, F^{(1)}_{u,v}(\mathbf{X}_N+t\mathbf{V},\mathbf{Y}_N+t\mathbf{W})\Big|^2\Big],
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{V}\mathrm{ar}\Big(\tau_N\big[\big(h_N^{(1,2)}\big)_{\substack{u,v\\u',v'}}\big]\Big)\\
&\quad\leqslant \frac{\sigma}{N}\mathbb{E}\Big[\max_{(\mathbf{V},\mathbf{W})\in\mathcal{W}}\Big|\frac{d}{dt}_{|t=0}\, F^{(2)}_{u,v,u',v'}(\mathbf{X}_N+t\mathbf{V},\mathbf{Y}_N+t\mathbf{W})\Big|^2\Big].
\end{aligned}
$$

For all $(\mathbf{V},\mathbf{W})$ in $\mathcal{W}$, for all selfadjoint $N\times N$ matrices $\mathbf{A}=(A_1,\ldots,A_1)$, $\mathbf{B}=(B_1,\ldots,B_1)$:

$$
\begin{aligned}
&\Big|\frac{d}{dt}_{|t=0}\, F^{(1)}_{u,v}(\mathbf{A}+t\mathbf{V},\mathbf{B}+t\mathbf{W})\Big|^2\\
&= \Big|\frac{d}{dt}_{|t=0}\,\frac{1}{N}(\mathrm{Tr}_k\otimes \mathrm{Tr}_N)\Big[(\epsilon_{v,u}\otimes \mathbf{1}_N)\\
&\qquad\times\Big(\Lambda\otimes \mathbf{1}_N-\sum_{j=1}^p a_j\otimes(A_j+tV_j)-\sum_{j=1}^q b_j\otimes(B_j+tW_j)\Big)^{-1}\Big]\Big|^2\\
&= \Big|\frac{1}{N}(\mathrm{Tr}_k\otimes \mathrm{Tr}_N)\Big[(\epsilon_{v,u}\otimes \mathbf{1}_N)(\Lambda\otimes \mathbf{1}_N-\tilde S_N-\tilde T_N)^{-1}\\
&\qquad\times\Big(\sum_{j=1}^p a_j\otimes V_j+\sum_{j=1}^q b_j\otimes W_j\Big)(\Lambda\otimes \mathbf{1}_N-\tilde S_N-\tilde T_N)^{-1}\Big]\Big|^2.
\end{aligned}
$$

The Cauchy-Schwarz inequality for $\mathrm{Tr}_k \otimes \mathrm{Tr}_N$ (i.e. for Tr_{kN}) gives

$$\begin{aligned}
&\left| \frac{d}{dt}_{|t=0} \; F^{(1)}_{u,v}(\mathbf{A}+t\mathbf{V},\mathbf{B}+t\mathbf{W}) \right|^2 \\
&\leqslant \; \frac{1}{N^2} \left\| (\epsilon_{v,u}\otimes \mathbf{1}_N)(\Lambda\otimes\mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1} \right\|_e^2 \\
&\qquad \times \; \left\| \Big(\sum_{j=1}^p a_j \otimes V_j + \sum_{j=1}^q b_j \otimes W_j \Big) (\Lambda\otimes\mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1} \right\|_e^2 .
\end{aligned}$$

Using (5.15) to split Euclidean norms into the product of an operator norm and an Euclidean norm, we get:

$$\begin{aligned}
&\left| \frac{d}{dt}_{|t=0} \; F^{(1)}_{u,v}(\mathbf{A}+t\mathbf{V},\mathbf{B}+t\mathbf{W}) \right|^2 \\
&\leqslant \; \frac{1}{N^2} \|\epsilon_{v,u}\otimes\mathbf{1}_N\|_e^2 \; \|(\Lambda\otimes\mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1}\|^2 \\
&\qquad \times \left\| \sum_{j=1}^p a_j\otimes V_j + \sum_{j=1}^q b_j\otimes W_j \right\|_e^2 \\
&\leqslant \; \frac{k}{N} \; \|(\mathrm{Im}\;\Lambda)^{-1}\|^4 \; \left\| \sum_{j=1}^p a_j\otimes V_j + \sum_{j=1}^q b_j\otimes W_j \right\|_e^2 .
\end{aligned}$$

Remark that, since $(\mathbf{V},\mathbf{W}) \in \mathcal{W}$, the norm of the matrices V_j and W_j is bounded by one. Then we have the following:

$$\begin{aligned}
&\left\| \sum_{j=1}^p a_j\otimes V_j + \sum_{j=1}^q b_j\otimes W_j + b_j^*\otimes W_j^* \right\|_e \\
&\leqslant \; \sum_{j=1}^p \|a_j\|_e + 2\sum_{j=1}^q \|b_j\|_e \leqslant \sqrt{k}\Big(\sum_{j=1}^p \|a_j\| + \sum_{j=1}^q \|b_j\| \Big).
\end{aligned}$$

Hence we finally obtain an estimate of $\mathbb{V}\mathrm{ar}(H_{S_N+T_N}(\Lambda)\;)_{u,v})$:

$$\mathbb{V}\mathrm{ar}\Big(H_{S_N+T_N}(\Lambda)\;\Big)_{u,v} \leqslant \frac{k^2\sigma}{N^2}\Big(\sum_{j=1}^p \|a_j\| + \sum_{j=1}^q \|b_j\| \Big)^2 \; \|(\mathrm{Im}\;\Lambda)^{-1}\|^4. \quad (5.18)$$

We obtain a similar estimate for $\mathbb{V}\mathrm{ar}\Big(\tau_N\big[(h_N^{(1,2)})_{\substack{u,v\\u',v'}} \big] \;\Big)$. The partial derivative of $F^{(2)}_{u,v,u',v'}$ gives two terms: $\forall(\mathbf{V},\mathbf{W}) \in \mathcal{W}$, $\forall(\mathbf{A},\mathbf{B}) \in \mathrm{M}_N(\mathbb{C})^{p+q}$

$$\begin{aligned}
&\frac{d}{dt}_{|t=0} \; F^{(2)}_{u,v,u',v'}(\mathbf{A}+t\mathbf{V},\mathbf{B}+t\mathbf{W}) \\
&= \frac{1}{N}(\mathrm{Tr}_k \otimes \mathrm{Tr}_N)\Big[(\epsilon_{v,u}\otimes \mathbf{1}_N)(\Gamma\otimes \mathbf{1}_N - \tilde{T}_N)^{-1}\Big(\sum_{j=1}^{q} b_j \otimes W_j\Big) \\
&\qquad \times (\Gamma\otimes \mathbf{1}_N - \tilde{T}_N)^{-1}(\epsilon_{u',v'}\otimes \mathbf{1}_N)(\Lambda\otimes \mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1} \\
&\qquad + (\epsilon_{v,u}\otimes \mathbf{1}_N)(\Gamma\otimes \mathbf{1}_N - \tilde{T}_N)^{-1}(\epsilon_{u',v'}\otimes \mathbf{1}_N)(\Lambda\otimes \mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1} \\
&\qquad \times \Big(\sum_{j=1}^{p} a_j\otimes V_j^{(N)} + \sum_{j=1}^{q} b_j \otimes W_j^{(N)}\Big)(\Lambda\otimes \mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1}\Big].
\end{aligned}$$

We then get the following:

$$\begin{aligned}
&\left|\frac{d}{dt}_{|t=0} \; F^{(2)}_{u,v,u',v'}(\mathbf{A}+t\mathbf{V},\mathbf{B}+t\mathbf{W})\right|^2 \\
&\leqslant \frac{k^2}{N}\Big(\sum_{j=1}^{p}\|a_j\| + \sum_{j=1}^{q}\|b_j\|\Big)^2 \|(\mathrm{Im}\;\Gamma)^{-1}\|^2 \\
&\qquad \times\|(\mathrm{Im}\;\Lambda)^{-1}\|^2\Big(\|(\mathrm{Im}\;\Lambda)^{-1}\| + \|(\mathrm{Im}\;\Gamma)^{-1}\|\Big)^2.
\end{aligned}$$

Hence we have

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}\Big(\tau_N\big[\big(h_N^{(1,2)}\big)_{\substack{u,v\\u',v'}}\big]\Big) \\
&\leqslant \frac{k^2\sigma}{N^2}\Big(\sum_{j=1}^{p}\|a_j\| + \sum_{j=1}^{q}\|b_j\|\Big)^2 \|(\mathrm{Im}\;\Gamma)^{-1}\|^2 \\
&\qquad \times\|(\mathrm{Im}\;\Lambda)^{-1}\|^2\;\Big(\|(\mathrm{Im}\;\Gamma)^{-1}\| + \|(\mathrm{Im}\;\Lambda)^{-1}\|\Big)^2. \qquad (5.19)
\end{aligned}$$

We then obtain as desired, by (5.17), (5.18) and (5.19):

$$\begin{aligned}
\|\Theta_N(\Lambda,\Gamma)\| &\leqslant k^{5/2}\sum_{j=1}^{p}\|a_j\|^2\Big(\sum_{u,v=1}^{k}\mathbb{V}\mathrm{ar}\big(H_{S_N+T_N}(\Lambda)\big)_{u,v} \\
&\qquad \times \max_{u,v,u',v'}\mathbb{V}\mathrm{ar}\Big(\tau_N\big[\big(h_N^{(1,2)}\big)_{\substack{u,v\\u',v'}}\big]\Big)\Big)^{1/2} \\
&\leqslant \frac{c}{N^2}\,\|(\mathrm{Im}\;\Gamma)^{-1}\|\,\|(\mathrm{Im}\;\Lambda)^{-1}\|^3 \\
&\qquad \times\Big(\|(\mathrm{Im}\;\Gamma)^{-1}\| + \|(\mathrm{Im}\;\Lambda)^{-1}\|\Big),
\end{aligned}$$

where $c = k^{9/2}\sigma\sum_{j=1}^{p}\|a_j\|^2\Big(\sum_{j=1}^{p}\|a_j\| + \sum_{j=1}^{q}\|b_j\|\Big)^2$.

□

5.3 Proof of Theorem 5.1

By (4.5), for all Λ in $\mathrm{M}_k(\mathbb{C})^+$, the matrix $\Lambda - \mathcal{R}_s\big(G_{S_N+T_N}(\Lambda)\ \big)$ is in $\mathrm{M}_k(\mathbb{C})^+$ and then it makes sense to choose $\Gamma = \Lambda - \mathcal{R}_s\big(G_{S_N+T_N}(\Lambda)\ \big)$ in Equation (5.10). We obtain for all Λ in $\mathrm{M}_k(\mathbb{C})^+$,

$$G_{S_N+T_N}(\Lambda) = G_{T_N}\Big(\Lambda - \mathcal{R}_s\big(G_{S_N+T_N}(\Lambda)\ \big)\ \Big) + \Theta_N(\Lambda),$$

where $\Theta_N(\Lambda) = \Theta_N\Big(\Lambda, \Lambda - \mathcal{R}_s\big(G_{S_N+T_N}(\Lambda)\ \big)\ \Big)$ is analytic in k^2 complex variables. Recall that by (4.6), we have $\big\|\big(\Lambda - \mathcal{R}_s\big(G_{S_N+T_N}(\Lambda)\ \big)\ \big)^{-1}\big\| \leqslant \|(\Lambda)^{-1}\|$, which gives (when replacing c in (5.12) by $c/2$) the expected estimate of $\Theta_N(\Lambda)$.

6 Proof of Estimate (3.9)

Let $(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{x}, \mathbf{y})$ be as in Section 3. We assume that $\big(\mathbf{x}, \mathbf{y}, (\mathbf{Y}_N)_{N\geqslant 1}\big)$ are realized in a same $\mathcal{C}^*$-probability space $(\mathcal{A}, .^*, \tau, \|\cdot\|)$ with faithful trace, where

- the families $\mathbf{x}$, $\mathbf{y}$, $\mathbf{Y}_1$, $\mathbf{Y}_2, \dots, \mathbf{Y}_N, \dots$ are free,
- for any polynomials P in q non commutative indeterminates $\tau[P(\mathbf{Y}_N)] := \tau_N[P(\mathbf{Y}_N)]$.

Consider L a degree one selfadjoint polynomial with coefficients in $\mathrm{M}_k(\mathbb{C})$. Define the Stieltjes transform of $L_N = L(\mathbf{X}_N, \mathbf{Y}_N)$ and $\ell_N = L(\mathbf{x}, \mathbf{Y}_N)$: for all $\lambda \in \mathbb{C}^+ = \big\{z \in \mathbb{C}\big|\ \mathrm{Im}\ z > 0\big\}$,

$$g_{L_N}(\lambda) = \mathbb{E}\Big[(\tau_k \otimes \tau_N)\Big[\big(\lambda \mathbf{1}_k \otimes \mathbf{1}_N - L_N\ \big)^{-1}\ \Big]\Big], \tag{6.1}$$

$$g_{\ell_N}(\lambda) = (\tau_k \otimes \tau)\Big[\big(\lambda \mathbf{1}_k \otimes \mathbf{1} - \ell_N\ \big)^{-1}\ \Big]. \tag{6.2}$$

One can always write $L_N = a_0 \otimes \mathbf{1}_N + S_N + T_N$, $\ell_N = a_0 \otimes \mathbf{1} + s + T_N$, where

$$S_N = \sum_{j=1}^p a_j \otimes X_j^{(N)},\ s = \sum_{j=1}^p a_j \otimes x_j,\ T_N = \sum_{j=1}^q b_j \otimes Y_j^{(N)},$$

and $a_0, \dots, a_p, b_1, \dots, b_q$ are Hermitian matrices in $\mathrm{M}_k(\mathbb{C})$. Define the $\mathrm{M}_k(\mathbb{C})$-valued Stieltjes transforms of $S_N + T_N$ and $s + T_N$: for all $\Lambda \in \mathrm{M}_k(\mathbb{C})^+ = \big\{\Lambda \in \mathrm{M}_k(\mathbb{C})\ \big|\ \mathrm{Im}\ \Lambda > 0\big\}$,

$$G_{S_N+T_N}(\Lambda) = \mathbb{E}\Big[(\mathrm{id}_k \otimes \tau_N)\Big[\big(\Lambda \otimes \mathbf{1}_N - S_N - T_N\big)^{-1}\Big]\Big],$$

$$G_{s+T_N}(\Lambda) = (\mathrm{id}_k \otimes \tau)\Big[\big(\Lambda \otimes \mathbf{1} - s - T_N\big)^{-1}\Big].$$

Then one has: for all λ in $\mathbb{C}^+$

$$g_{L_N}(\lambda) = \tau_k\Big[\ G_{S_N+T_N}(\lambda \mathbf{1}_k - a_0)\ \Big],\ g_{\ell_N}(\lambda) = \tau_k\Big[\ G_{s+T_N}(\lambda \mathbf{1}_k - a_0)\ \Big].$$

By Proposition 4.2, for any $\Lambda \in \mathrm{M}_k(\mathbb{C})^+$, one has

$$G_{s+T_N}(\Lambda) = G_{T_N}\Big(\Lambda - \mathcal{R}_s\big(G_{s+T_N}(\Lambda)\ \big)\ \Big).$$

On the other hand, since the matrices of $\mathbf{Y}_N$ are deterministic, we can apply Theorem 5.1 with $\sigma = 1$

$$G_{S_N+T_N}(\Lambda) = G_{T_N}\Big(\Lambda - \mathcal{R}_s\big(G_{S_N+T_N}(\Lambda)\ \big)\ \Big) + \Theta_N(\Lambda),$$

where $\|\Theta_N(\Lambda)\| \leqslant \frac{c}{N^2}\,\|(\mathrm{Im}\ \Lambda)^{-1}\|^5$ for a constant $c > 0$. Define

$$\Omega_\eta^{(N)} = \Big\{\Lambda \in \mathrm{M}_k(\mathbb{C})^+ \Big|\ \|(\mathrm{Im}\ \Lambda)^{-1}\| < N^\eta\Big\}.$$

Then for $\eta < 1/3$, there exists N_0 such that for all $N \geqslant N_0$ and for any Λ in $\Omega_\eta^{(N)}$, one has

$$\kappa(\Lambda) := \|\Theta_N(\Lambda)\|\ \|(\mathrm{Im}\ \Lambda)^{-1}\|\ \sum_{j=1}^p \|a_j\|^2 \leqslant \frac{c}{N^2}\|(\mathrm{Im}\ \Lambda)^{-1}\|^6 \leqslant cN^{6\eta-2} \leqslant \frac{1}{2}.$$

Then by Proposition 4.3 with $(t, G, \Theta, \Omega, \varepsilon) = (T_N, G_{S_N+T_N}, \Theta_N, \Omega_\eta^{(N)}, 1/2)$, one has

$$\begin{aligned}
&\|G_{s+T_N}(\Lambda) - G_{S_N+T_N}(\Lambda)\| \\
&\quad\leqslant\ \Big(1 + 2\sum_{j=1}^p \|a_j\|^2\ \|(\mathrm{Im}\Lambda)^{-1}\|^2\Big)\ \|\Theta(\Lambda)\| \\
&\quad\leqslant\ c\Big(1 + 2\sum_{j=1}^p \|a_j\|^2\ \|(\mathrm{Im}\ \Lambda)^{-1}\|^2\Big)\ \frac{\|(\mathrm{Im}\ \Lambda)^{-1}\|^5}{N^2}.
\end{aligned}$$

Hence for every $\varepsilon > 0$, there exist N_0 and γ such that for all $N \geqslant N_0$, for all λ in $\mathbb{C}$ such that $\varepsilon \leqslant (\mathrm{Im}\ \lambda)^{-1} \leqslant N^\gamma$, one has

$$|g_{L_N}(\lambda) - g_{\ell_N}(\lambda)| \leqslant \|G_{s+T_N}(\lambda\mathbf{1}_k - a_0) - G_{S_N+T_N}(\lambda\mathbf{1}_k - a_0)\| \leqslant \frac{c}{N^2}(\mathrm{Im}\ \lambda)^{-7}, \tag{6.3}$$

where c denotes now the constant $c = k^{9/2}\sum_{j=1}^p \|a_j\|\Big(\sum_{j=1}^p \|a_j\| + \sum_{j=1}^q \|b_j\|\Big)^2\Big(\varepsilon^{-2} + 2\sum_{j=1}^p \|a_j\|^2\Big)$.

7 Proof of Step 2: An intermediate inclusion of spectrum

For a review on the theory of $\mathcal{C}^*$-algebras, we refer the readers to [12] and [8]. Notably, Appendix A of the second reference contains facts about ultrafilters and ultraproducts that are used in this section.

Let $\big(\mathbf{x}, \mathbf{y}, (\mathbf{Y}_N)_{N\geqslant 1}\big)$ be as in Section 3. We assume that these non commutative random variables are realized in the same $\mathcal{C}^*$-probability space $(\mathcal{A}, .^*, \tau, \|\cdot\|)$ with faithful trace, where

- the families $\mathbf{x}$, $\mathbf{y}$, $\mathbf{Y}_1$, $\mathbf{Y}_2, \dots, \mathbf{Y}_N, \dots$ are free,
- for any polynomials P in q non commutative indeterminates $\tau[P(\mathbf{Y}_N)] := \tau_N[P(\mathbf{Y}_N)]$.

A consequence of Voiculescu's theorem and of Shlyakhtenko's Theorem A.1 in Appendix A is that for all polynomials P in $p+q$ non commutative indeterminates,

$$\tau[P(\mathbf{x}, \mathbf{Y}_N)] \underset{N\to\infty}{\longrightarrow} \tau[P(\mathbf{x}, \mathbf{y})], \tag{7.1}$$

$$\|P(\mathbf{x}, \mathbf{Y}_N)\| \underset{N\to\infty}{\longrightarrow} \|P(\mathbf{x}, \mathbf{y})\|. \tag{7.2}$$

In order to prove Step 2, it remains to show that (7.2) still holds when the polynomials P are $\mathrm{M}_k(\mathbb{C})$-valued. This fact is a folklore result in $\mathcal{C}^*$-algebra theory, we give a proof for readers convenience. We need first the two following lemmas.

Lemma 7.1. *Let $\mathcal{A}$ and $\mathcal{B}$ be unital $\mathcal{C}^*$-algebra. Let $\pi : \mathcal{A} \to \mathcal{B}$ be a morphism of unital *-algebra. Then π is contractive.*

Proof. It is easy to see that for any a in $\mathcal{A}$, the spectrum of $\pi(a)$ is included in the spectrum of a (since $\lambda \mathbf{1}_{\mathcal{A}} - a$ invertible implies that $\lambda \mathbf{1}_{\mathcal{A}} - \pi(a)$ is also invertible). Hence we get that for all a in $\mathcal{A}$

$$\|\pi(a)\|^2 = \|\pi(a^*a)\| \leqslant \|a^*a\| = \|a\|^2.$$

□

Lemma 7.2. *Let $\mathcal{A}$ be a unital $\mathcal{C}^*$-algebra. Then for any integer $k \geqslant 1$, there exists a unique $\mathcal{C}^*$-algebra structure on $\mathrm{M}_k(\mathbb{C}) \otimes \mathcal{A}$ compatible with the structure on $\mathcal{A}$. In particular, if $\mathcal{A}$ is a $\mathcal{C}^*$-probability space equipped with a faithful tracial state τ, then $\mathrm{M}_k(\mathbb{C}) \otimes \mathcal{A}$ is a $\mathcal{C}^*$-probability space with trace $(\tau_k \otimes \tau)$ and norm $\|\cdot\|_{\tau_k\otimes\tau}$, where τ_k is the normalized trace on $\mathrm{M}_k(\mathbb{C})$ and $\|\cdot\|_{\tau_k\otimes\tau}$ is given by Formula* (1.9)*.*

Sketch of the proof. For the existence we consider the norm given by the spectral radius. The uniqueness follows from Lemma 7.1. □

Proposition 7.3. *Let $k \geqslant 1$ be an integer. For all $N \geqslant 1$, let $\mathbf{z}_N = (z_1^{(N)}, \dots, z_p^{(N)})$, respectively $\mathbf{z} = (z_1, \dots, z_p)$, be self-adjoint non commutative random variables in a $\mathcal{C}^*$- probability space $(\mathcal{A}_N, .^*, \tau_N, \| \cdot \|_{\tau_N})$, respectively $(\mathcal{A}, .^*, \tau, \| \cdot \|_\tau)$. Assume that the traces τ_N and τ are faithful (hence the notation for the norms) and that for any polynomial P in p non commutative indeterminates,*

$$\tau_N[P(\mathbf{z}_N)] \underset{N\to\infty}{\longrightarrow} \tau[P(\mathbf{z})], \tag{7.3}$$

$$\|P(\mathbf{z}_N)\|_{\tau_N} \underset{N\to\infty}{\longrightarrow} \|P(\mathbf{z})\|_\tau. \tag{7.4}$$

Then for any polynomial P in p non commutative indeterminates with coefficients in $\mathrm{M}_k(\mathbb{C})$,

$$\|P(\mathbf{z}_N)\|_{\tau_k\otimes\tau_N} \underset{N\to\infty}{\longrightarrow} \|P(\mathbf{z})\|_{\tau_k\otimes\tau}. \tag{7.5}$$

We abuse notation and write with the same symbol the traces in $\mathrm{M}_k(\mathbb{C})$ and $\mathcal{A}_N$ when $N = k$. There is no danger of confusion.

Proof. For any positive integer k and any ultrafilter $\mathcal{U}$ on $\mathbb{N}$, we define the ultraproduct

$$\mathfrak{A}^{(k)} = \prod^{\mathcal{U}} \mathrm{M}_k(\mathbb{C}) \otimes \mathcal{A}_N,$$

which is the quotient of

$$\left\{ (a_N)_{N\geqslant 1} \,\middle|\, \forall N \geqslant 1,\ a_N \in \mathrm{M}_k(\mathbb{C}) \otimes \mathcal{A}_N \text{ and } \sup_{N\geqslant 1} \|a_N\| < \infty \right\},$$

by

$$\left\{ (a_N)_{N\geqslant 1} \,\middle|\, \forall N \geqslant 1,\ a_N \in \mathrm{M}_k(\mathbb{C}) \otimes \mathcal{A}_N \text{ and } \lim_{N\to\mathcal{U}} \|a_N\| = 0 \right\}.$$

The algebra $\mathfrak{A}^{(k)}$ is a $\mathcal{C}^*$-algebra whose norm $\| \cdot \|_{\mathfrak{A}^{(k)}}$ is given by: for all a in $\mathfrak{A}^{(k)}$, equivalence class of $(a_N)_{N\geqslant 1}$

$$\|a\|_{\mathfrak{A}^{(k)}} = \lim_{N\to\mathcal{U}} \|a_N\|_{\tau_k\otimes\tau_N}.$$

Furthermore $\mathfrak{A}^{(k)}$ is a $\mathcal{C}^*$-probability space which can be identified with $\mathrm{M}_k(\mathbb{C})\otimes\mathfrak{A}^{(1)}$. The trace $\tilde{\tau}$ on $\mathfrak{A}^{(1)}$ is given by: for all a in $\mathfrak{A}^{(1)}$, equivalence class of $(A_N)_{N\geqslant 1}$, one has

$$\tilde{\tau}[a] = \lim_{N\to\mathcal{U}} \tau[A_N].$$

If the classical limit as N goes to infinity exists, then the trace of a does not depends on the ultrafilter $\mathcal{U}$ and is given by the limit. The trace on $\mathfrak{A}^{(k)}$ is $(\tau_k \otimes \tilde{\tau})$. Notice that $(\tau_k \otimes \tilde{\tau})$ on $\mathfrak{A}^{(k)}$ is not faithful in general, which implies that the norm $\| \cdot \|_{\mathfrak{A}^{(k)}}$ and the norm $\| \cdot \|_{\tau_k \otimes \tilde{\tau}}$ given by $(\tau_k \otimes \tilde{\tau})$ with Formula (1.9) are not equal on the whole $\mathcal{C}^*$-algebra.
At last, we can equip $\mathfrak{A}^{(k)}$ with a structure of operator-valued $\mathcal{C}^*$-probability space. Define the unital sub-algebra $\mathcal{B}$ of $\mathfrak{A}^{(k)}$ as the set

$$\Big\{ \ b \otimes \mathbf{1}_{\mathfrak{A}^{(1)}} \ \Big| \ b \in \mathrm{M}_k(\mathbb{C}) \ \Big\} \subset \mathfrak{A}^{(k)}.$$

The conditional expectation in $\mathfrak{A}^{(k)}$ is given by $(\mathrm{id}_k \otimes \tilde{\tau}) : \mathfrak{A}^{(k)} \to \mathcal{B}$.

For $j = 1, \dots, p$, we denote by $\tilde{z}_j$ in $\mathfrak{A}^{(1)}$ the equivalence class of the sequence $(z_j^{(N)})_{N \geqslant 1}$. We have by definition of $\mathfrak{A}^{(k)}$: for all polynomial P in $p + 2q$ non commutative indeterminates with coefficients in $\mathrm{M}_k(\mathbb{C})$,

$$\| P(\mathbf{z}_N) \|_{\tau_N} \underset{N \to \mathcal{U}}{\longrightarrow} \| P(\tilde{\mathbf{z}}) \|_{\mathfrak{A}^{(k)}}$$

Let $\mathcal{C}^*(\tilde{\mathbf{z}})$ be the sub-algebra spanned by $\tilde{\mathbf{z}} = (\tilde{z}_1, \dots, \tilde{z}_p)$ in $\mathfrak{A}^{(1)}$ and let $\mathcal{C}^*(\mathbf{z})$ be the sub-algebra spanned by $\mathbf{z}$ in $\mathcal{A}$. Then by (7.4), the $\mathcal{C}^*$-algebras $\mathcal{C}^*(\tilde{\mathbf{z}})$ and $\mathcal{C}^*(\mathbf{z})$ are isomorphic. Hence we get an isomorphism of the *-algebras $\mathrm{M}_k(\mathbb{C}) \otimes \mathcal{C}^*(\tilde{\mathbf{z}})$ and $\mathrm{M}_k(\mathbb{C}) \otimes \mathcal{C}^*(\mathbf{z})$, and so an isomorphism of the $\mathcal{C}^*$-algebras by Lemma 7.1. Hence, for all polynomial P in $p + 2q$ non commutative indeterminates with coefficients in $\mathrm{M}_k(\mathbb{C})$,

$$\| P(\tilde{\mathbf{z}}) \|_{\mathfrak{A}^{(k)}} = \| P(\mathbf{z}) \|_{\tau_k \otimes \tilde{\tau}}$$

Hence we get

$$\| P(\mathbf{z}_N) \|_{\tau_k \otimes \tau_N} \underset{N \to \mathcal{U}}{\longrightarrow} \| P(\mathbf{z}) \|_{\tau_k \otimes \tilde{\tau}}$$

for all ultrafilter $\mathcal{U}$. Then the convergence holds when N goes to infinity. □

Proof of Step 2. Let L be a selfadjoint degree one polynomial in $p + q$ non commutative indeterminates with coefficients in $\mathrm{M}_k(\mathbb{C})$. Define $\ell_N = L(\mathbf{x}, \mathbf{Y}_N)$ and $\ell = L(\mathbf{x}, \mathbf{y})$. Then by Proposition 7.3, for all commutative polynomials P, one has

$$\| P(\ell_N) \|_{\tau_k \otimes \tau} \underset{N \to \infty}{\longrightarrow} \| P(\ell) \|_{\tau_k \otimes \tau}.$$

The convergence extends to continuous function on the real line and then, with an appropriate choice of test functions, Step 2 follows. □

8 Proof of Step 3: from Stieltjes transforms to spectra

Let $\mathbf{X}_N, \mathbf{Y}_N, \mathbf{x}$ and $\mathbf{y}$ be as in Section 3. As before $\mathbf{x}, \mathbf{y}$, and $\mathbf{Y}_N$ are assumed to be realized in a same $\mathcal{C}^*$-probability space $(\mathcal{A}, .^*, \tau, \|\cdot\|)$ with faithful trace. Let L be a selfadjoint degree one polynomial with coefficients in $\mathrm{M}_k(\mathbb{C})$.

For any function $f : \mathbb{R} \to \mathbb{R}$ and any Hermitian matrix A with spectral decomposition $A = U\mathrm{diag}\ (\lambda_1, \dots, \lambda_K)U^*$, with U unitary, we set the Hermitian matrix $f(A) = U\mathrm{diag}\ (f(\lambda_1), \dots, f(\lambda_K))U^*$. For any function $f : \mathbb{R} \mapsto \mathbb{R}$, we set

$$D_N(f) = (\tau_k \otimes \tau_N)\big[f(L(\mathbf{X}_N, \mathbf{Y}_N))\big].$$

By Step 2, for all $\varepsilon > 0$, there exists $N_0 \geqslant 1$ such that for all $N \geqslant N_0$, one has

$$\mathrm{Sp}\Big(\ L(\mathbf{x}, \mathbf{Y}_N)\ \Big) \subset \mathrm{Sp}\Big(\ L(\mathbf{x}, \mathbf{y})\ \Big) + (-\varepsilon, \varepsilon).$$

Hence, for any function f vanishing on a neighborhood of the spectrum of $L(\mathbf{x}, \mathbf{y})$, there exists $N_0 \geqslant 1$ such that for all $N \geqslant N_0$, the function f actually vanishes on a neighborhood of the spectrum of $L(\mathbf{x}, \mathbf{Y}_N)$. In particular, with μ_N (respectively ν_N) denoting the empirical eigenvalue distribution of $L_N = L(\mathbf{X}_N, \mathbf{Y}_N)$ (respectively $\ell_N = L(\mathbf{x}, \mathbf{Y}_N)$), one has

$$\mathbb{E}\Big[D_N(f)\Big] = \mathbb{E}\Big[\ \int f \ \mathrm{d}\mu_N\ \Big] = \mathbb{E}\Big[\ \int f \ \mathrm{d}\mu_N\ \Big] - \int f \ \mathrm{d}\nu_N. \qquad (8.1)$$

Furthermore, by Estimate (3.9), with the Stieltjes transforms of L_N and of ℓ_N defined by: for all λ in $\mathbb{C}^+$

$$\begin{aligned} g_{L_N}(\lambda) &= \mathbb{E}\Big[(\tau_k \otimes \tau_N)\Big[\ \big(\lambda \mathbf{1}_k \otimes \mathbf{1}_N - L_N\ \big)^{-1}\ \Big]\ \Big] = \mathbb{E}\Big[\ \int \frac{1}{\lambda - t}\mathrm{d}\mu_N(t)\Big] \\ g_{\ell_N}(\lambda) &= (\tau_k \otimes \tau)\Big[\ \big(\lambda \mathbf{1}_k \otimes \mathbf{1} - \ell_N\ \big)^{-1}\ \Big] = \int \frac{1}{\lambda - t}\mathrm{d}\nu_N(t), \end{aligned}$$

we have shown that: for any $\varepsilon > 0$ and $A > 0$, there exist $N_0, c, \eta, \gamma, \alpha > 0$ such that for all $N \geqslant N_0$, for all λ in $\mathbb{C}$ such that $\varepsilon \leqslant (\mathrm{Im}\ \lambda)^{-1} \leqslant N^\gamma$ and $|\mathrm{Re}\ \lambda| \leqslant A$

$$|g_{L_N}(\lambda) - g_{\ell_N}(\lambda)| \leqslant \frac{c}{N^2}(\mathrm{Im}\ \lambda)^{-\alpha}. \qquad (8.2)$$

With (8.1) and (8.2) established, it is easy to show with minor modifications of [1, Lemma 5.5.5] the following result.

Lemma 8.1. *For every smooth function $f : \mathbb{R} \to \mathbb{R}$ non negative, compactly supported and vanishing on a neighborhood of the spectrum of $L(\mathbf{x}, \mathbf{y})$, there exists a constant such that for all N large enough*

$$\Big|\mathbb{E}\big[D_N(f)\ \big]\ \Big| \leqslant \frac{c}{N^2}. \tag{8.3}$$

To get an almost sure control of $D_N(f)$, we use the fact that the entries of the matrices $\mathbf{X}_N$ satisfy a concentration inequality.

Lemma 8.2. *With f as in Lemma 8.1, there exists $\kappa > 0$ such that, almost surely*

$$N^{1+\kappa} D_N(f) \underset{N\to\infty}{\longrightarrow} 0. \tag{8.4}$$

Proof. The law of the random matrices satisfying a Poincaré inequality with constant $\frac{1}{N}$ and L being a polynomial of degree one, for all Lipschitz function $\Psi : \mathrm{M}_{kN}(\mathbb{C}) \mapsto \mathbb{R}$, by [17, Lemma 5.2] one has:

$$\mathbb{P}\Big(\ \big|\Psi(L_N) - \mathbb{E}\big[\Psi(L_N)\ \big]\ \big| \geqslant \delta\Big) \leqslant K_1 e^{-K_2 \frac{\sqrt{N}\delta}{|\Psi|_{\mathcal{L}}}}, \tag{8.5}$$

where K_1, K_2 are positive constants and $|\Psi|_{\mathcal{L}} = \sup_{A\neq B\in \mathrm{M}_{kN}(\mathbb{C})} \frac{|\Psi(A)-\Psi(B)|}{\|A-B\|_e}$. Recall that the Euclidean norm $\|\cdot\|_e$ of a matrix $A = (a_{i,j})_{i,j=1}^{kN}$ is given by

$$\|A\|_e = \sqrt{\sum_{i,j=1}^{kN} |a_{i,j}|^2}.$$

For any Hermitian matrices A in $\mathrm{M}_{kN}(\mathbb{C})$ and any function $f : \mathbb{R} \to \mathbb{R}$, we set

$$\Phi_N^{(f)}(A) = (\tau_k \otimes \tau_N)\big[f(A)\ \big]. \tag{8.6}$$

For all smooth function $f : \mathbb{R} \to \mathbb{R}$, $N \geqslant 1$ and $0 < \kappa < \frac{1}{2}$, we define

$$\mathcal{B}_{N,\kappa}^{(f)} = \left\{ A \in \mathrm{M}_{kN}(\mathbb{C}) \ \middle|\ A \text{ is Hermitian and } \left|\Phi_N^{(f'^2)}(A)\right| \leqslant \frac{1}{N^{4\kappa}} \right\}, \tag{8.7}$$

and denote $\rho_{N,\kappa}^{(f)} = |(\Phi_N^{(f)})_{|\mathcal{B}_{N,\kappa}}|_{\mathcal{L}}$. Define $\Psi_N^{(f)} : \mathrm{M}_{kN}(\mathbb{C}) \mapsto \mathbb{R}$ by: $\forall A \in \mathrm{M}_N(\mathbb{C})$

$$\Psi_N^{(f)}(A) = \sup_{B\in\mathcal{B}_{N,\kappa}^{(f)}} \Big\{ \Phi_N^{(f)}(B) - \rho_{N,\kappa}^{(f)}\ \|A-B\|_2\ \Big\}, \tag{8.8}$$

and denote $\tilde{D}_N(f) = \Psi_N^{(f)}(L_N)$. By [17, Proof of Lemma 5.9], $\Psi_N^{(f)}$ coincides with $\Phi_N^{(f)}$ on $\mathcal{B}_{N,\kappa}^{(f)}$ and is Lipschitz with constant $|\Psi_N^{(f)}|_{\mathcal{L}} \leqslant \rho_{N,\kappa}^{(f)}$.

For all Hermitian matrices A in $\mathrm{M}_{kN}(\mathbb{C})$, M in $\mathrm{M}_{kN}(\mathbb{C})$ and $n \geqslant 1$, one has $\frac{d}{dt}_{|t=0}(A+tM)^n = \sum_{m=0}^n A^m M A^{n-m-1}$ and then $\frac{d}{dt}_{|t=0}(\tau_k \otimes \tau_N)[(A+tM)^n] = (\tau_k \otimes \tau_N)[nA^{n-1}M]$. So for all polynomials P, one has $\mathrm{D}_A \Phi_N^{(P)}(M) = (\tau_k \otimes \tau_N)[P'(A)M]$. Hence, by density of polynomials, for any smooth function $f : \mathbb{R} \to \mathbb{R}$ one has $\mathrm{D}_A \Phi_N^{(f)}(M) = (\tau_k \otimes \tau_N)[f'(A)M]$. By the Cauchy-Schwarz inequality, we get

$$\begin{aligned}
\big|\mathrm{D}_A \Phi_N^{(f)}(M)\big|^2 &= |(\tau_k \otimes \tau_N)[f'(A)M]|^2 \\
&\leqslant (\tau_k \otimes \tau_N)[f'(A)^2] \times (\tau_k \otimes \tau_N)[M^*M] \\
&= \Phi_N^{(f'^2)}(A) \times \frac{\|M\|_e}{kN}.
\end{aligned}$$

Then, for any smooth function f, one has

$$\rho_{N,\kappa}^{(f)} \leqslant \frac{1}{\sqrt{kN}} \| \ (\Phi_N^{(f'^2)})_{|\mathcal{B}_{N,\kappa}^{(f)}} \ \|_\infty^{1/2}, \tag{8.9}$$

where $\|\cdot\|_\infty$ denotes the supremum of the considered function on the set of $kN \times kN$ Hermitian matrices. Hence we get that $|\Psi_N^{(f)}|_{\mathcal{L}} \leqslant \rho_{N,\kappa}^{(f)} \leqslant \frac{1}{\sqrt{k}} N^{-1/2-2\kappa}$.

We fix f a smooth function, non negative, compactly supported and vanishing on a neighborhood of the spectrum of $L(\mathbf{x},\mathbf{y})$. By the Tchebychev inequality

$$\mathbb{P}(L_N \notin \mathcal{B}_{N,\kappa}^{(f)}) \ = \ \mathbb{P}\Big(D_N(f'^2) \geqslant \frac{1}{N^{4\kappa}}\Big) \leqslant N^{4\kappa}\mathbb{E}\Big[D_N(f'^2)\Big] \leqslant \frac{C}{N^{2-4\kappa}}, \tag{8.10}$$

where we have used Lemma 8.1 (f'^2 also vanishes in a neighborhood of the spectrum of $L(\mathbf{x},\mathbf{y})$). Moreover, since $\Psi_N^{(f)}$ and $\Phi_N^{(f)}$ are equals in $\mathcal{B}_{N,\kappa}^{(f)}$ and $\|\Psi_N^{(f)}\|_\infty \leqslant \|\Phi_N^{(f)}\|_\infty$,

$$\Big|\mathbb{E}\big[\tilde{D}_N(f) - D_N(f)\big] \ \Big| \leqslant \|\Phi_N^{(f)}\|_\infty \mathbb{P}(L_N \notin \mathcal{B}_{N,\kappa}^{(f)}) \leqslant \|\Phi_N^{(f)}\|_\infty \frac{C}{N^{2-4\kappa}} \tag{8.11}$$

Now, by (8.5) applied to $\Psi_N^{(f)}$: for all $\delta > 0$

$$\begin{aligned}
&\mathbb{P}\bigg(\Big|D_N(f) - \mathbb{E}\big[D_N(f)\ \big]\ \Big| > \frac{\delta}{N^{1+\kappa}} \ \text{ and } \ L_N \in \mathcal{B}_{N,\kappa}^{(f)}\bigg) \\
&\leqslant \ P\bigg(\Big|\tilde{D}_N(f) - \mathbb{E}\big[\tilde{D}_N(f)\ \big]\ \Big| > \frac{\delta}{N^{1+\kappa}} \ - \ \Big|\mathbb{E}\big[\tilde{D}_N(f) - D_N(f)\big]\ \Big|\ \bigg) \\
&\leqslant \ K_1 \exp\bigg(- \sqrt{k} K_2 N^\kappa (\delta - \ \Big|\mathbb{E}\big[\tilde{D}_N(f) - D_N(f)\big]\ \Big|)\bigg)
\end{aligned}$$

By (8.10), (8.11), Lemma 8.1 and the Borel-Cantelli lemma, $D_N(f)$ is almost surely of order $N^{1+\kappa}$ at most. □

Proposition 8.3. *For every $\varepsilon > 0$, there exists N_0 such that for $N \geqslant N_0$*

$$\mathrm{Sp}\Big(\ L(\mathbf{X}_N, \mathbf{Y}_N) \ \Big) \ \subset \mathrm{Sp}\Big(\ L(\mathbf{x}, \mathbf{y}) \ \Big) + (-\varepsilon, \varepsilon) \qquad (8.12)$$

Proof. By (1.11) and [1, Exercise 2.1.27], almost surely there exists $N_0 \in \mathbb{N}$ and $D \geqslant 0$ such that the spectral radii of the matrices $(\mathbf{X}_N, \mathbf{Y}_N)$ is bounded by D for all $N \geqslant N_0$. Hence, there exists $M \geqslant 0$ such that almost surely one has

$$\mathrm{Sp}\Big(\ L(\mathbf{X}_N, \mathbf{Y}_N) \ \Big) \subset [-M, M].$$

Let $f : \mathbb{R} \mapsto \mathbb{R}$ non negative, compactly supported, vanishing on $\mathrm{Sp}(\ L(\mathbf{x}, \mathbf{y}) \ ) + (-\varepsilon/2, \varepsilon/2)$ and equal to one on $[-M, M] \smallsetminus \big(\mathrm{Sp}(\ L(\mathbf{x}, \mathbf{y})) + (-\varepsilon, \varepsilon) \ \big)$. Then almost surely for N large enough, no eigenvalue of $L(\mathbf{X}_N, \mathbf{Y}_N)$ belongs to the complementary of $\mathrm{Sp}(\ L(\mathbf{x}, \mathbf{y}) \ ) + (-\varepsilon, \varepsilon)$, since otherwise

$$(\tau_k \otimes \tau_N)\Big[f\big(L(\mathbf{X}_N, \mathbf{Y}_N) \ \big) \Big] \geqslant N^{-1} \geqslant N^{-1-\kappa}$$

in contradiction with Lemma 8.2. □

9 Proof of Corollaries 2.1, 2.2 and 2.4

9.1 Proof of Corollary 2.1: diagonal matrices

Let $\mathbf{D}_N = (D_1^{(N)}, \dots, D_q^{(N)})$ be as in Corollary 2.1. For any $j = 1, \dots, p$, the number of jump of F_j^{-1} is countable. We show that the convergence of the norm (2.3) holds when we chose $v = (v_1, \dots, v_q)$ in $[0,1]^q$ such that for any $k \neq \ell$ in $\{1, \dots, q\}$, the sets of jump points of $u \mapsto F_k^{-1}(u + v_k)$ and $u \mapsto F_\ell^{-1}(u + v_\ell)$ are disjoint. We show that for such a v, the family $\mathbf{D}_N^v$ satisfies the assumptions of Theorem 1.6. In all this section, we always denote $\lambda_i(j)$ instead of $\lambda_i^{(N)}(j)$ for any $i = 1, \dots, N$ and any $j = 1, \dots, q$.

The convergence of traces, case $v = (0, \dots, 0)$: Since the matrices commute, we only consider commutative polynomials. We start by showing that for all polynomials P,

$$\tau_N\Big[P(\mathbf{D}_N) \ \Big] \underset{N\to\infty}{\longrightarrow} \int_0^1 P\Big(F_1^{-1}(u), \dots, F_q^{-1}(u) \ \Big) du. \qquad (9.1)$$

Denote by μ the probability distribution of the random variable $\big(F_1^{-1}(U), \dots, F_q^{-1}(U) \ \big) \in \mathbb{R}^q$, where U is distributed according to the uniform distribution on $[0,1]$. In order to get (9.1), we show that the sequence of measure in $\mathbb{R}^q$

$$\Big(\frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i(1)}, \ldots, \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i(q)}\Big)$$

converges weakly to μ. This sequence is tight, since there exists a $B>0$ such that for all $j=1\ldots q$, for all $i=1\ldots N$, one has $\lambda_i(j)\in[-B,B]$. Hence it is sufficient to show the following: for all real numbers $a_1,\ldots,a_q$, for all $\varepsilon>0$, there exists $\eta>0$ such that

$$\limsup_{N\to\infty}\Big|\frac{1}{N}\sum_{i=1}^N \mathbf{1}_{]-\infty,a_1+\eta]}\big(\lambda_i(1)\big)\times\cdots\times\mathbf{1}_{]-\infty,a_q+\eta]}\big(\lambda_i(q)\big) \\ -\mu\big(\,]-\infty,a_1]\times\cdots\times]-\infty,a_q]\,\big)\Big|\leqslant\varepsilon. \tag{9.2}$$

Fix $(a_1,\ldots,a_q)$ in $\mathbb{R}^q$ and $\varepsilon>0$. Remark that one has

$$\mu\big(\,]-\infty,a_1]\times\cdots\times]-\infty,a_q]\,\big)=\min_{j=1\ldots q}F_j(a_j).$$

Let j_0 be an integer such that $F_{j_0}(a_{j_0})=\mu\big(\,]-\infty,a_1]\times\cdots\times]-\infty,a_q]\,\big)$. For any $j=1,\ldots,q$, the empirical spectral distribution of $D_j^{(N)}$ converges to μ_j. Then for all a in $\mathbb{R}$ point of continuity for F_j, one has

$$\frac{1}{N}\sum_{i=1}^N \mathbf{1}_{]-\infty,a]}\big(\lambda_i(j)\big)\underset{N\to\infty}{\longrightarrow}\mu_j\big(\,]-\infty,a]\,\big). \tag{9.3}$$

Let $\eta>0$ such that

- $\mu_{j_0}\big(\,]a_{j_0},a_{j_0}+\eta]\,\big)<\varepsilon/2$.
- for all $j=1,\ldots,q$, the real numbers $a_j+\eta$ and $a_{j_0}+\eta$ are points of continuity for F_j.

By (9.3) with $a=a_j+\eta$, there exists $N_0\geqslant1$ such that for all $N\geqslant N_0$ and $j=1,\ldots,q$, one has

$$F_j(a_j+\eta)-\varepsilon\leqslant\frac{1}{N}\mathrm{Card}\,\Big\{i=1\ldots N\;\Big|\;\lambda_i(j)\leqslant a_j+\eta\Big\}.$$

But $F_j(a_j+\eta)\geqslant F_j(a_j)\geqslant F_{j_0}(a_{j_0})$. Then we have

$$N\Big(F_{j_0}(a_{j_0})-\varepsilon\Big)\leqslant\mathrm{Card}\,\Big\{i=1\ldots N\;\Big|\;\lambda_i(j)\leqslant a_j+\eta\Big\}.$$

The $\lambda_i(j)$ are non decreasing, so we get

$$\forall j = 1 \dots q,\ \forall i \leqslant N\Big(F_{j_0}(a_{j_0}) - \varepsilon\Big),\ \lambda_i(j) \leqslant a_j + \eta. \tag{9.4}$$

On the other hand, by (9.3) with $j = j_0$ and $a = a_{j_0} + \eta$, there exists $N_0 \geqslant 1$ such that, for all $N \geqslant N_0$, one has

$$\frac{1}{N}\mathrm{Card}\Big\{i = 1 \dots N \;\Big|\; \lambda_i(j_0) \leqslant a_{j_0} + \eta\Big\} \leqslant F_{j_0}(a_{j_0} + \eta) + \varepsilon/2.$$

But $F_{j_0}(a_{j_0} + \eta) \leqslant F_{j_0}(a_{j_0}) + \varepsilon/2$, so that

$$\mathrm{Card}\Big\{i = 1 \dots N \;\Big|\; \lambda_i(j_0) \leqslant a_{j_0} + \eta\Big\} \leqslant N\Big(F_{j_0}(a_{j_0}) + \varepsilon\Big).$$

The $\lambda_i(j_0)$ are non decreasing, then we get

$$\forall i \geqslant N\Big(F_{j_0}(a_{j_0}) + \varepsilon\Big),\ \lambda_i(j_0) \geqslant a_{j_0} + \eta. \tag{9.5}$$

By (9.4) and (9.5) we obtain: for all $N \geqslant N_0$

$$\Big|\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{]-\infty, a_1+\eta]}\big(\lambda_i(1)\big) \times \cdots \times \mathbf{1}_{]-\infty, a_q+\eta]}\big(\lambda_i(q)\big) - F_{j_0}(a_{j_0} + \eta)\Big| \leqslant \varepsilon,$$

and then (9.2) is satisfied. So the convergence (9.1) holds when v is zero.

The convergence of traces, case v in $[0,1]^q$: To deduce the general case we shall need the following lemmas.

Lemma 9.1 (Quantiles of real diagonal matrices with sorted entries)**.** *Let $D_N =$ diag $(\lambda_1, \dots, \lambda_N)$ be an $N \times N$ real diagonal matrix with non decreasing entries along its diagonal. Assume that the empirical eigenvalue distribution of D_N converges weakly to a compactly supported probability measure μ. Let F denote the cumulative distribution function of μ and F^{-1} its generalized inverse. Let v in $(0,1)$ a point of continuity for F^{-1} and $(i_N)_{N\geqslant 1}$ a sequence of integers, with i_N in $\{1, \dots, N\}$, such that i_N/N tends to v. Then, one has*

$$\lambda_{i_N} \underset{N\to\infty}{\longrightarrow} F^{-1}(v).$$

In particular, we have the convergence of the quantile of order v:

$$\lambda_{1+\lfloor vN\rfloor} \underset{N\to\infty}{\longrightarrow} F^{-1}(v).$$

Proof. Denote $w = F^{-1}(v)$. We can always find $\eta \geqslant 0$, arbitrary small, such that $w-\eta$ and $w+\eta$ and points of continuity for F. Then, one has

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{]-\infty,w-\eta]}(\lambda_i) \underset{N\to\infty}{\longrightarrow} \mu\big(\,]-\infty, w-\eta] \, \big) = F(w-\eta).$$

Then, the λ_i being non decreasing, for any $\varepsilon > 0$ there exists N_0 such that for any $N \geqslant N_0$, one has

$$\forall i \geqslant \Big(F(w-\eta)+\varepsilon\Big)N, \quad \lambda_i \geqslant w-\eta. \tag{9.6}$$

Since v is a point of continuity for F^{-1}, we get that $F(w-\eta) < v$. We chose $\varepsilon < v - F(w-\eta)$. Then, we get $F(w-\eta)+\varepsilon < v$. Hence, there exists N_0 such that, for any $N \geqslant N_0$, one has $i_N \geqslant \big(F(w-\eta)+\varepsilon\big)N$ and so, by (9.6): for any $\eta > 0$, there exists N_0 such that for all $N \geqslant N_0$, one has $w-\eta \leqslant \lambda_{i_N}$. Hence, we get for all $\eta > 0$,

$$w-\eta \leqslant \liminf_{N\to\infty} \lambda_{i_N}.$$

With the same reasoning, we get that

$$\limsup_{N\to\infty} \lambda_{i_N} \geqslant w+\eta,$$

and hence, letting η go to zero, we obtain the expected result. □

Lemma 9.2 (Truncation of real diagonal matrices with sorted entries)**.** *Let $D_N =$ diag $(\lambda_1,\ldots,\lambda_N)$ an $N\times N$ real diagonal matrix with non decreasing entries along its diagonal. Assume that the empirical eigenvalue distribution of D_N converges weakly to a compactly supported probability measure μ. For any $v_1 < v_2$ in $[0,1]$, we set*

$$D_N^{(v_1,v_2)} = \text{diag } (\lambda_{1+\lfloor v_1 N\rfloor},\ldots,\lambda_{\lfloor v_2 N\rfloor}).$$

Let F denote the cumulative distribution function of μ and F^{-1} its generalized inverse. We set $w_1 = F^{-1}(v_1)$, $w_2 = F^{-1}(v_2)$, $a_1 = F(w_1) - v_1$ and $a_2 = v_2 - F(w_2^-)$. Then, the empirical eigenvalue distribution of $D_N^{(v_1,v_2)}$ converges weakly the probability measure proportional to

$$a_1\delta_{w_1} + \mu\Big(\ \cdot \cap \,]w_1, w_2[\ \Big) + a_2\delta_{w_2}.$$

Proof. We only show the lemma for $v_2 = 0$, the general case can be deduce by adapting the reasoning. We then use, for conciseness, the symbols v, w and a instead of v_1, w_1 and a_1 respectively.

If F is not continuous in w (i.e. if $\mu(w) \neq 0$) and $v \neq F(w)$, then for any α in $]0,(F(w)-v)/2[$, the map F^{-1} is continuous at the points $v+\alpha$ and $F(w)-\alpha$. By Lemma 9.1, we get that

$$\lim_{N\to\infty} \lambda_{1+\lfloor (v+\alpha)N\rfloor} = \lim_{N\to\infty} \lambda_{1+\lfloor (F(w)-\alpha)N\rfloor} = w. \tag{9.7}$$

Hence, for any continuous function f, we get

$$\frac{1}{N} \sum_{i=1+\lfloor (v+\alpha)N\rfloor}^{1+\lfloor (F(w)-\alpha)N\rfloor} f(\lambda_i) \underset{N\to\infty}{\longrightarrow} (a-2\alpha) f(w). \tag{9.8}$$

If F is continuous in w, we take $\alpha = 0$ in the following.

We can always find $\beta > 0$, arbitrary small, such that $F(w)+\beta$ is a point of continuity for F^{-1}. Remark that we then have

$$w = F^{-1}\big(F(w)\big) < F^{-1}\big(F(w)+\beta\big).$$

By Lemma 9.1, we get

$$\lambda_{1+\lfloor (F(w)+\beta)N\rfloor} \underset{N\to\infty}{\longrightarrow} F^{-1}\big(F(w)+\beta\big). \tag{9.9}$$

Moreover, we can always find γ in $]0, F^{-1}\big(F(w)+\beta\big)-w[$, arbitrary small, such that $w+\gamma$ is a point of continuity for F and $F(w+\gamma) < F(w)+\beta$. Then, by (9.9), we get that, for N large enough

$$\mathrm{Card}\left\{ i \geqslant 1+\lfloor (F(w)-\alpha)N\rfloor \,\middle|\, \lambda_i \leqslant w+\gamma \right\} \leqslant \lfloor (F(w)+\beta)N\rfloor - \lfloor (F(w)-\alpha)N\rfloor.$$

Hence, for any continuous function f, we get that for N large enough

$$\begin{aligned}
&\left| \frac{1}{N} \sum_{i=1+\lfloor (F(w)-\alpha)N\rfloor}^{N} f(\lambda_i) - \int_{]\omega,+\infty]} f(x)\mathrm{d}\mu(x) \right| \\
&\quad\leqslant \left| \frac{1}{N} \sum_{i=1}^{N} f(\lambda_i)\mathbf{1}_{]w+\gamma,+\infty]}(\lambda_i) - \int_{]\omega,+\infty]} f(x)\mathrm{d}\mu(x) \right| \\
&\qquad + \|f\|_\infty \frac{\lfloor (F(w)+\beta)N\rfloor - \lfloor (F(w)-\alpha)N\rfloor}{N}.
\end{aligned} \tag{9.10}$$

By (9.8) and (9.10), we obtain

$$\begin{aligned}
&\limsup_{N\to\infty} \left| \frac{1}{N} \sum_{i=1+\lfloor vN\rfloor}^{N} f(\lambda_i) - a f(w) - \int_{]\omega,+\infty]} f(x)\mathrm{d}\mu(x) \right| \\
&\quad\leqslant \|f\|_\infty \Big(4\alpha + \beta + \mu\big(]w, w+\gamma]\big)\Big).
\end{aligned}$$

Letting α, β, γ go to zero, we get the result. □

Let v in $[0,1]^q$. We now show that, for any polynomial P, one has

$$\tau_N\Big[P(\mathbf{D}_N^v)\ \Big] \underset{N\to\infty}{\longrightarrow} \int_0^1 P\Big(F_1^{-1}(u+v_1),\dots,F_q^{-1}(u+v_q)\ \Big)du. \quad (9.11)$$

At the possible price of relabeling the matrices, we assume $v_1 \geqslant \dots \geqslant v_q$ and set

$$\begin{aligned} N_1 &= N - \lfloor v_1 N\rfloor, \\ N_j &= \lfloor v_{j-1}N\rfloor - \lfloor v_j N\rfloor,\ \forall j=1,\dots,q. \end{aligned}$$

For any $j=1,\dots,q$, we decompose the matrices $D_j^{(N)}(v_j)$ into

$$D_j^{(N)}(v_j) = \text{ diag }(D_{j,1}^{(N)},\dots,D_{j,q}^{(N)}),$$

where for any $i=1,\dots,q$, the matrix $D_{j,i}^{(N)}$ is $N_i\times N_i$. We set for any $i=1,\dots,q$, the family $\mathbf{D}_N(i) = (D_{1,i}^{(N)},\dots,D_{q,i}^{(N)})$. For any $i,j=1,\dots,q$, we denote by $F_{i,j}$ the cumulative distribution function of the measure obtained in Lemma 9.2 with (D_N,μ,v_1,v_2) replaced by $(D_j^{(N)},\mu_j,v_{i-1},v_i)$. Then, for any polynomial P, one as

$$\tau_N[P(\mathbf{D}_N^v)] = \sum_{i=1}^q \frac{N_i}{N}\tau_{N_i}[P(\mathbf{D}_N(i))].$$

By Lemma 9.2 and by the case $v=(0,\dots,0)$, we deduce that

$$\tau_{N_i}[P(\mathbf{D}_N(i))] \underset{N\to\infty}{\longrightarrow} \frac{1}{v_{q-1}-v_q}\int_{v_q}^{v_{q-1}} P\Big(F_{i,1}^{-1}(u+v_1),\dots,F_{i,q}^{-1}(u+v_q)\ \Big)du,$$

with the convention $v_0=1$. The merge of the different terms for $i=1,\dots,q$ gives as expected

$$\tau_N\Big[P(\mathbf{D}_N^v)\ \Big] \underset{N\to\infty}{\longrightarrow} \int_0^1 P\Big(F_1^{-1}(u+v_1),\dots,F_q^{-1}(u+v_q)\ \Big)du. \quad (9.12)$$

The convergence of norms: Let $v=(v_1,\dots,v_q)$ in $[0,1]^q$ such that for any $k\neq\ell$ in $\{1,\dots,q\}$, the sets of jump points of $u\mapsto F_k^{-1}(u+v_k)$ and $u\mapsto F_\ell^{-1}(u+v_\ell)$ are disjoint. We now show that, for all polynomials P, one has

$$\|P(\mathbf{D}_N^v)\| \underset{N\to\infty}{\longrightarrow} \operatorname*{Sup}_{\text{Supp }\mu^v} \big|P\big|,$$

where μ^v is the probability distribution of the random variable $\big(F_1^{-1}(U+v_1),\dots,F_q^{-1}(U+v_q)\ \big)\in\mathbb{R}^q$, where U is distributed according to the uniform distribution on $[0,1]$. In view of the above, we have

$$\liminf\ \|P(\mathbf{D}_N^v)\| \geqslant \operatorname*{Sup}_{\text{Supp }\mu^v}\big|P\big|.$$

It is sufficient then to show that, for any $\eta > 0$, there exists $N_0 \geqslant N$ such that for all $i = 1, \dots, N$, one has

$$\Big(\lambda_{i+\lfloor v_1 N\rfloor}(1), \dots, \lambda_{i+\lfloor v_q N\rfloor}(q)\ \Big) \in \text{Supp } \mu^v + (-\eta, \eta)^q. \tag{9.13}$$

Indeed, by uniform continuity, for any polynomial P and $\varepsilon > 0$, there exists $\eta \geqslant 0$ such that, for all $(x_1, \dots, x_q)$ in Supp $\mu^v + [-1, 1]^q$ and $(y_1, \dots, y_q)$ in $\mathbb{R}^q$, one has

$$|y_j - x_j| < \eta \ \Rightarrow\ \Big|P(x_1, \dots, x_q) - P(y_1, \dots, y_q)\Big| < \varepsilon$$

and hence: for all $\varepsilon > 0$, there exist $\eta \geqslant 0$ and $N_0 \geqslant 1$ such that for all $N \geqslant N_0$, for all $i = 1, \dots, N$

$$\max_{i=1\dots N}\Big|P\big(\lambda_{i+\lfloor v_1 N\rfloor}(1), \dots, \lambda_{i+\lfloor v_q N\rfloor}(q)\ \big)\ \Big| \leqslant \max_{\text{Supp }\ \mu^v+(-\eta,\eta)^q} \big|P\big| \leqslant \max_{\text{Supp }\ \mu^v} |P|+\varepsilon.$$

Suppose that (9.13) is not true: there exist $\eta > 0$ and $(N_k)_{k\geqslant 1}$ an increasing sequence of positive integer such that for all $k \geqslant 1$, there exists i_k such that

$$\Big(\lambda^{(N_k)}_{i_k+\lfloor v_1 N_k\rfloor}(1), \dots, \lambda^{(N_k)}_{i_k+\lfloor v_q N_k\rfloor}(q)\ \Big) \notin \text{Supp } \mu^v + (-\eta, \eta)^q.$$

By compactness, one can always assume that i_k/N_k converges to u_0 in $[0, 1]$. For all j in $\{1, \dots, q\}$ except a possible j_0, we have that $u_0 + v_j$ is a point of continuity for F_j^{-1} and so, by Lemma 9.1, $\lambda^{(N_k)}_{i_k+\lfloor v_j N_k\rfloor}(j)$ converges to $F_j^{-1}(u_0 + v_j)$. Recall that

$$\text{Supp } \mu^v = \Big\{\big(F_1^{-1}(u + v_1), \dots, F_q^{-1}(u + v_q)\ \big)\ \Big|\ u \in [0, 1]\ \Big\}.$$

Then we have, for N large enough and for all u in $[0, 1]$, that $\big|\lambda^{(N_k)}_{i_k+\lfloor v_{j_0} N_k\rfloor}(j_0) - F_{j_0}^{-1}(u + v_{j_0})\big| > \eta$ i.e.

$$\text{dist }\big(\lambda^{(N_k)}_{i_k+\lfloor v_{j_0} N_k\rfloor}(j_0), \text{Supp } \mu_{j_0}\big) > \eta,$$

which is in contradiction with the fact that for N large enough the eigenvalues of $D^{(N)}_{j_0}$ belong to a small neighborhood of the support of μ_{j_0}.

9.2 Proof of Corollary 2.2: Wishart matrices

Let $r, s_1, \dots, s_p \geqslant 1$ and $(\mathbf{W}_N, \mathbf{Y}_N)$ be as in Corollary 2.2 and denote $s = s_1 + \dots + s_p$. We use matrix manipulations in order to see the norm of a polynomial in the $rN \times rN$ matrices $\mathbf{W}_N, \mathbf{Y}_N, \mathbf{Y}_N^*$ as the norm of a polynomial in $(r+s)N \times (r+s)N$ matrices $\tilde{\mathbf{X}}_N, \tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \tilde{\mathbf{Z}}_N$ and some elementary matrices, where $\tilde{\mathbf{X}}_N$ is a family of independent GUE matrices and $\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Z}}_N$ are modifications of $\mathbf{Y}_N, \mathbf{Z}_N$. We will obtain the result as a consequence of Theorem 1.6.

Define the $(r+s)N \times (r+s)N$ matrices $\mathbf{e}_N = (e_0^{(N)}, e_1^{(N)}, \dots, e_p^{(N)})$:

$$e_0^{(N)} = \begin{pmatrix} \mathbf{1}_{rN} & \mathbf{0}_{rN,sN} \\ \mathbf{0}_{sN,rN} & \mathbf{0}_{sN} \end{pmatrix}, \tag{9.14}$$

$$e_j^{(N)} = \begin{pmatrix} \mathbf{0}_{rN} & & & \\ & \mathbf{0}_{(s_1+\cdots+s_{j-1})N} & & \\ & & \mathbf{1}_{s_jN} & \\ & & & \mathbf{0}_{(s_{j+1}+\cdots+s_p)N} \end{pmatrix}, \quad j = 1, \dots, p. \tag{9.15}$$

Recall that by definition of the Wishart matrix model for $j = 1, \dots, p$

$$W_j^{(N)} = M_j^{(N)} Z_j^{(N)} M_j^{(N)*}, \tag{9.16}$$

where $M_j^{(N)}$ is an $rN \times s_jN$ complex Gaussian matrix with independent identically distributed entries, centered and of variance $1/rN$. Let $\tilde{\mathbf{X}}_N = (\tilde{X}_1^{(N)}, \dots, \tilde{X}_p^{(N)})$ be a family of p independent, normalized GUE matrices of size $(r+s)N \times (r+s)N$, independent of $\mathbf{Y}_N$ and $\mathbf{Z}_N$ and such that for $j = 1, \dots, p$, the $rN \times s_jN$ matrix $M_j^{(N)}$ appears as a sub-matrix of $\sqrt{\frac{r+s}{r}}\tilde{X}_j^{(N)}$ in the following way: if we denote $\tilde{M}_j^{(N)} = \sqrt{\frac{r+s}{r}} e_0^{(N)} \tilde{X}_j^{(N)} e_j^{(N)}$ then

$$\tilde{M}_j^{(N)} = \begin{pmatrix} \mathbf{0}_{rN} & & M_j^{(N)} & \\ & \mathbf{0}_{(s_1+\cdots+s_{j-1})N} & & \\ & & \mathbf{0}_{s_jN} & \\ & & & \mathbf{0}_{(s_{j+1}+\cdots+s_p)N} \end{pmatrix}. \tag{9.17}$$

Let $\tilde{\mathbf{Y}}_N = (\tilde{Y}_1^{(N)}, \dots, \tilde{Y}_q^{(N)})$ and $\tilde{\mathbf{Z}}_N = (\tilde{Z}_1^{(N)}, \dots, \tilde{Z}_p^{(N)})$ be the families of $(r+s)N \times (r+s)N$ matrices defined by:

$$\tilde{Y}_j^{(N)} = \begin{pmatrix} Y_j^{(N)} & \mathbf{0}_{rN,sN} \\ \mathbf{0}_{sN,rN} & \mathbf{0}_{sN} \end{pmatrix}, \quad j = 1, \dots, q, \tag{9.18}$$

$$\tilde{Z}_j^{(N)} = \begin{pmatrix} \mathbf{0}_{rN} & & & \\ & \mathbf{0}_{(s_1+\cdots+s_{j-1})N} & & \\ & & Z_j^{(N)} & \\ & & & \mathbf{0}_{(s_{j+1}+\cdots+s_p)N} \end{pmatrix}, \quad j = 1, \dots, p. \tag{9.19}$$

By assumption, with probability one the non commutative law of $\mathbf{Y}_N$ converges to the law of non commutative random variables $\mathbf{y} = (y_1, \dots, y_q)$ in a $\mathcal{C}^*$-probability space $(\mathcal{A}_0, .^*, \tau, \|\cdot\|)$ and for $j = 1 \dots p$ the non commutative law of Z_j converges to the law of a non commutative random variable z_j in a $\mathcal{C}^*$-probability space $(\mathcal{A}_j, .^*, \tau, \|\cdot\|)$ (we use the same notations for the functionals in the different spaces). All the traces under consideration are faithful. Let $\mathcal{B}$ denotes the product algebra $\mathcal{B}_0 \times \mathcal{B}_1 \times \cdots \times \mathcal{B}_p$. We equip $\mathcal{B}$ with the involution $.^*$ and the trace $\tilde\tau$ defined by: for all $(b_0, \dots, b_p)$ in $\mathcal{B}$

$$(b_0, \dots, b_p)^* = (b_0^*, \dots, b_p^*),$$
$$\tilde\tau\big[\ (b_0, \dots, b_p)\ \big] = \frac{r}{r+s}\tau(b_0) + \frac{s_1}{r+s}\tau(b_1) + \cdots + \frac{s_p}{r+s}\tau(b_p).$$

The trace $\tilde\tau$ is a faithful tracial state on $\mathcal{B}$. Equipped with $.^*$, $\tilde\tau$ and with the norm $\|\cdot\|$ defined by (1.9), the algebra $\mathcal{B}$ is a $\mathcal{C}^*$-probability space. Define $\tilde{\mathbf{y}} = (\tilde y_1, \dots, \tilde y_q)$, $\tilde{\mathbf{z}} = (\tilde z_1, \dots, \tilde z_q)$ and $\mathbf{e} = (e_0, \dots, e_p)$ by

$$\tilde y_j = (y_j, \mathbf{0}_{\mathcal{B}_1}, \dots, \mathbf{0}_{\mathcal{B}_p}),\ j = 1, \dots, q,$$
$$\tilde z_j = (\mathbf{0}_{\mathcal{B}_0}, \dots, \mathbf{0}_{\mathcal{B}_{j-1}}, z_j, \mathbf{0}_{\mathcal{B}_{j+1}}, \dots, \mathbf{0}_{\mathcal{B}_p}),\ j = 1, \dots, p,$$
$$e_j = (\mathbf{0}_{\mathcal{B}_0}, \dots, \mathbf{0}_{\mathcal{B}_{j-1}}, \mathbf{1}_{\mathcal{B}_j}, \mathbf{0}_{\mathcal{B}_{j+1}}, \dots, \mathbf{0}_{\mathcal{B}_p}),\ j = 0, \dots, q.$$

Lemma 9.3. *With probability one, the non commutative law of* $(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Z}}_N, \mathbf{e}_N)$ *in* $(\mathrm{M}_{(r+s)N}(\mathbb{C}), .^*, \tau_{(r+s)N})$ *converges to the law of* $(\tilde{\mathbf{y}}, \tilde{\mathbf{z}}, \mathbf{e})$ *in* $(\mathcal{B}, .^*, \tilde\tau)$.

Proof. Let P be a polynomial in $2p + 2q + 1$ non commutative indeterminates:

$$\begin{aligned}
&\tau_{(r+s)N}\big[P(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{Z}_N, \mathbf{e}_N)\ \big] \\
&= \quad \frac{r}{r+s}\tau_{rN}\big[P(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \underbrace{\mathbf{0}_{rN}, \dots, \mathbf{0}_{rN}}_{p}, \mathbf{1}_{rN}, \underbrace{\mathbf{0}_{rN}, \dots, \mathbf{0}_{rN}}_{p})\ \big] \\
&+ \quad \sum_{j=1}^p \frac{s_j}{s+r}\tau_{s_j}\big[P(\underbrace{\mathbf{0}_{s_jN}, \dots, \mathbf{0}_{s_jN}}_{2q+j-1}, Z_j^{(N)}, \underbrace{\mathbf{0}_{s_jN}, \dots, \mathbf{0}_{s_jN}}_{p}, \mathbf{1}_{s_jN}, \underbrace{\mathbf{0}_{s_jN}, \dots, \mathbf{0}_{s_jN}}_{p-j})\ \big] \\
&\underset{N\to\infty}{\longrightarrow} \quad \frac{r}{r+s}\tau\big[P(\mathbf{y}, \mathbf{y}^*, \underbrace{\mathbf{0}, \dots, \mathbf{0}}_{p}, \mathbf{1}, \underbrace{\mathbf{0}, \dots, \mathbf{0}}_{p})\ \big] \\
&\qquad + \sum_{j=1}^p \frac{s_j}{s+r}\tau\big[P(\underbrace{\mathbf{0}, \dots, \mathbf{0}}_{2q+j-1}, z_j, \underbrace{\mathbf{0}, \dots, \mathbf{0}}_{p}, \mathbf{1}, \underbrace{\mathbf{0}, \dots, \mathbf{0}}_{p-1})\ \big] && (9.20)\\
&= \quad \tilde\tau[P(\tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \tilde{\mathbf{z}}, \mathbf{e})\ ], && (9.21)
\end{aligned}$$

where the convergence holds almost surely since each term of the sum converges almost surely. □

Lemma 9.4. *For all polynomials P in $2p+2q+1$ non commutative indeterminates, almost surely*

$$\left\|P(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{Z}_N, \mathbf{e}_N)\ \right\| \underset{N\to\infty}{\longrightarrow} \|P(\tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \tilde{\mathbf{z}}, \mathbf{e})\ \|.$$

Proof. Lemma 9.4 follows easily since for any polynomial P in $2p+2q+1$ non commutative indeterminates, $\left\|P(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{Z}_N, \mathbf{e}_N)\ \right\|$ is the maximum of the $p+1$ real numbers

- $\|P(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \underbrace{\mathbf{0}_{rN}, \ldots, \mathbf{0}_{rN}}_{p}, \mathbf{1}_{rN}, \underbrace{\mathbf{0}_{rN}, \ldots, \mathbf{0}_{rN}}_{p})\|$,
- $\|P(\underbrace{\mathbf{0}_{s_jN}, \ldots, \mathbf{0}_{s_jN}}_{2q+j-1}, Z_j^{(N)}, \underbrace{\mathbf{0}_{s_jN}, \ldots, \mathbf{0}_{s_jN}}_{p}, \mathbf{1}_{s_jN}, \underbrace{\mathbf{0}_{s_jN}, \ldots, \mathbf{0}_{s_jN}}_{p-j})\|$, $j = 1, \ldots, p$,

and $\|P(\tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \tilde{\mathbf{z}}, \mathbf{e})\|_{\tilde{\tau}}$ is the maximum of the $p+1$ real numbers

- $\|P(\mathbf{y}, \mathbf{y}^*, \underbrace{\mathbf{0}, \ldots, \mathbf{0}}_{p}, \mathbf{1}, \underbrace{\mathbf{0}, \ldots, \mathbf{0}}_{p})\|$,
- $\|P(\underbrace{\mathbf{0}, \ldots, \mathbf{0}}_{2q+j-1}, z_j, \underbrace{\mathbf{0}, \ldots, \mathbf{0}}_{p}, \mathbf{1}, \underbrace{\mathbf{0}, \ldots, \mathbf{0}}_{p-1})\|$, $j = 1, \ldots, p$.

□

Let $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_p)$ be a free semicircular system in $\mathcal{C}^*$-probability space. Let $\tilde{\mathcal{A}}$ be the reduced free product $\mathcal{C}^*$-algebra of $\mathcal{B}$ and the $\mathcal{C}^*$-algebra spanned by $\tilde{\mathbf{x}}$. We still denotes by $\tilde{\tau}$ the trace on $\tilde{\mathcal{A}}$ and the norm considered $\|\cdot\|$ is given by (1.9) since the trace is faithful. By Voiculescu's theorem and by the independence of $\tilde{\mathbf{X}}_N$ and $(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Z}}_N)$, with probability one the non commutative law of $(\tilde{\mathbf{X}}_N, \tilde{\mathbf{Y}}_N, \tilde{\mathbf{Z}}_N, \mathbf{e}_N)$ in $(\mathrm{M}_{(r+s)N}(\mathbb{C}), .^*, \tau_{(r+s)N})$ converges to the non commutative law of $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}}, \mathbf{e})$ in $(\tilde{\mathcal{A}}, .^*, \tilde{\tau})$. Define the non commutative random variables $\tilde{\mathbf{m}} = (\tilde{m}_1, \ldots, \tilde{m}_q)$ and $\tilde{\mathbf{w}} = (\tilde{w}_1, \ldots, \tilde{w}_q)$ in $\tilde{\mathcal{A}}$ by: for $j = 1, \ldots, q$,

$$\tilde{m}_j = \sqrt{\frac{r+s}{r}} e_0 \tilde{x}_j e_j, \quad \tilde{w}_j = e_0(\tilde{m}_j\ \tilde{z}_j + \tilde{m}_j^*)^2. \tag{9.22}$$

Lemma 9.5. *For any polynomial P in $p+2q$ non commutative indeterminates, there exists a polynomial $\tilde{P}$ in $3p+2q+1$ non commutative indeterminates, such that one has*

$$\begin{pmatrix} P(\mathbf{W}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) & \mathbf{0}_{rN,sN} \\ \mathbf{0}_{sN,rN} & \mathbf{0}_{sN} \end{pmatrix} = \tilde{P}(\tilde{\mathbf{X}}_N, \tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \tilde{\mathbf{Z}}_N, \mathbf{e}_N), \tag{9.23}$$

$$e_0 P(\tilde{\mathbf{w}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*) = \tilde{P}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \tilde{\mathbf{z}}, \mathbf{e}).$$

Proof. We set $\tilde{\mathbf{W}}_N = (W_1^{(N)}, \dots, W_p^{(N)})$ given by: for $j = 1, \dots, p$,

$$\tilde{W}_j^{(N)} := e_0^{(N)}(\tilde{M}_j^{(N)} \tilde{Z}_j^{(N)} + \tilde{M}_j^{(N)*})^2 = \begin{pmatrix} W_j^{(N)} & \mathbf{0}_{rN,sN} \\ \mathbf{0}_{sN,rN} & \mathbf{0}_{sN} \end{pmatrix}. \quad (9.24)$$

Let P be a polynomial in $p + 2q$ non commutative indeterminates. By the block decomposition of $\tilde{\mathbf{W}}_N$ and $\tilde{\mathbf{Y}}_N$, one has

$$\begin{pmatrix} P(\mathbf{W}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) & \mathbf{0}_{rN,sN} \\ \mathbf{0}_{sN,rN} & \mathbf{0}_{sN} \end{pmatrix} = e_0^{(N)} \ P(\tilde{\mathbf{W}}_N, \tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*).$$

Furthermore, By definitions of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{W}}$: for $j = 1, \dots, p$

$$\begin{aligned} \tilde{W}_j^{(N)} &= e_0^{(N)}(\tilde{M}_j^{(N)} \tilde{Z}_j^{(N)} + \tilde{M}_j^{(N)*})^2 \\ &= e_0^{(N)} \frac{r+s}{r} (e_0^{(N)} \tilde{X}_j^{(N)} e_j^{(N)} \tilde{Z}_j^{(N)} + e_j^{(N)} \tilde{X}_j^{(N)} e_0^{(N)})^2. \end{aligned}$$

Define for $j = 1, \dots, p$ the non commutative polynomial P_j deduced by the formula

$$P_j(\tilde{x}_j, \tilde{z}_j, \mathbf{e}) = e_0 \frac{r+s}{r} (e_0 \tilde{x}_j e_j \tilde{z}_j + e_j \tilde{x}_j e_0)^2, \quad (9.25)$$

and define $\tilde{P}$ deduced by

$$\tilde{P}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \tilde{\mathbf{z}}, \mathbf{e}) = e_0 \ P\Big(P_1(\tilde{x}_1, \tilde{z}_1, \mathbf{e}), \dots, P_p(\tilde{x}_p, \tilde{z}_p, \mathbf{e}), \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*\Big). \quad (9.26)$$

The polynomials are defined without ambiguity if $\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \tilde{\mathbf{z}}, \mathbf{e}$ are seen as families of non commutative indeterminates (without any algebraic relation) instead of non commutative random variables. Remark that, by definition, for all $j = 1, \dots, p$ the non commutative random variable w_j equals $P_j(\tilde{x}_j, \tilde{z}_j, \mathbf{e})$. Hence it follows as expected that

$$\begin{aligned} \begin{pmatrix} P(\mathbf{W}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) & \mathbf{0}_{rN,sN} \\ \mathbf{0}_{sN,rN} & \mathbf{0}_{sN} \end{pmatrix} &= \tilde{P}(\tilde{\mathbf{X}}_N, \tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \tilde{\mathbf{Z}}_N, \mathbf{e}_N), \\ e_0 P(\tilde{\mathbf{w}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*) &= \tilde{P}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \tilde{\mathbf{z}}, \mathbf{e}). \end{aligned}$$

□

It is well known as a generalization of Voiculescu's theorem that, under Assumption 1 separately for $Z_1^{(N)}, , \dots, Z_p^{(N)}, \mathbf{Y}_N$ and by independence of the families, with probability one the non commutative law of $(\mathbf{W}_N, \mathbf{Y}_N)$ in $(\mathrm{M}_N(\mathbb{C}), .^*, \tau_N)$ converges to the non commutative law of $(\mathbf{w}, \mathbf{y})$ in a $\mathcal{C}^*$-probability space $(\mathcal{A}, .^*, \tau, \| \cdot \|)$ with faithful trace, where

1. $\mathbf{w} = (w_1, \dots, w_p)$ are free selfadjoint non commutative random variables,

2. $\mathbf{y} = (y_1, \ldots, y_q)$ is the limit in law of $\mathbf{Y}_N$,
3. $\mathbf{w}$ and $\mathbf{y}$ are free.

For any polynomial P in $p + 2q$ non commutative indeterminates

$$\begin{aligned}
\tau[P(\mathbf{w}, \mathbf{y}, \mathbf{y}^*)] &= \lim_{N\to\infty} \tau_{rN}\big[P(\mathbf{W}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)\ \big] \\
&= \lim_{N\to\infty} \frac{r+s}{r} \tau_{(r+s)N}\left[\begin{pmatrix} P(\mathbf{W}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) & \mathbf{0}_{rN,sN} \\ \mathbf{0}_{sN,rN} & \mathbf{0}_{sN} \end{pmatrix}\right] \\
&= \lim_{N\to\infty} \frac{r+s}{r} \tau_{(r+s)N}\big[\tilde{P}(\tilde{\mathbf{X}}_N, \tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \tilde{\mathbf{Z}}_N, \mathbf{e}_N)\ \big] \\
&= \frac{r+s}{r} \tilde{\tau}\big[\tilde{P}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \tilde{\mathbf{z}}, \mathbf{e})\ \big] \\
&= \frac{r+s}{r} \tilde{\tau}\big[e_0 P(\tilde{\mathbf{w}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*)\ \big],
\end{aligned}$$

where the limits are almost sure. In particular we obtain that, for all polynomials P in $p + 2q$ non commutative indeterminates, one has

$$\|e_0 P(\tilde{\mathbf{w}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*)\| = \|P(\mathbf{w}, \mathbf{y}, \mathbf{y}^*)\|. \tag{9.27}$$

By Lemmas 9.3 and 9.4, the family of $(r + s)N \times (r + s)N$ matrices $(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Z}}_N, \mathbf{e}_N)$ satisfies the assumptions of Theorem 1.6, hence for all polynomials P in $3p + 2q + 1$ non commutative indeterminates, with $\tilde{P}$ as in Lemma 9.5, almost surely one has

$$\|\tilde{P}(\tilde{\mathbf{X}}_N, \tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \tilde{\mathbf{Z}}_N, \mathbf{e}_N)\| \underset{N\to\infty}{\longrightarrow} \|\tilde{P}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \tilde{\mathbf{z}}, \mathbf{e})\|. \tag{9.28}$$

Remark that

$$\begin{aligned}
\|P(\mathbf{W}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)\| &= \left\| \begin{pmatrix} P(\mathbf{W}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) & \mathbf{0}_{rN,sN} \\ \mathbf{0}_{sN,rN} & \mathbf{0}_{sN} \end{pmatrix} \right\| \\
&= \|\tilde{P}(\tilde{\mathbf{X}}_N, \tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \tilde{\mathbf{Z}}_N, \mathbf{e}_N)\|,
\end{aligned}$$

$$\|\tilde{P}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \tilde{\mathbf{z}}, \mathbf{e})\| = \|e_0 P(\tilde{\mathbf{w}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*)\| = \|P(\mathbf{w}, \mathbf{y}, \mathbf{y}^*)\|.$$

Together with (9.28), this gives the expected result.

9.3 Proof of Corollary 2.4: Rectangular band matrices

We only give a sketch of the proof. Details are obtained by minor modification of the proofs of Corollaries 2.2 and 2.3. Let H be as in Corollary

2.4:

$$H = \begin{pmatrix} A_1 & A_2 & \dots & A_L & \mathbf{0} & \dots & & \dots & \mathbf{0} \\ \mathbf{0} & A_1 & A_1 & \dots & A_L & \mathbf{0} & & & \vdots \\ \vdots & \mathbf{0} & A_1 & A_2 & \dots & A_L & \mathbf{0} & & \\ & & \ddots & \ddots & \ddots & & \ddots & \vdots & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & & \dots & \mathbf{0} & A_1 & A_2 & \dots & A_L \end{pmatrix}. \tag{9.29}$$

We start with the following observation: the operator norm of H is the square root of the operator norm of H^*H, which is a square block matrix. Its blocks consist of sums of $tN \times tN$ matrices of the form $A_l^* A_m$, $l, m = 1 \dots L$. By minor modifications of the proof of Corollary 2.2, we get the almost sure convergence of the normalized trace and of the norm for any polynomial in the matrices $\mathbf{A}_N = (A_l^* A_m)_{l,m=1..L}$ as N goes to the infinity. By Proposition 7.3, we get that the convergences hold for square block matrices and in particular for any polynomial in H^*H. Hence the result follows by functional calculus.

A A theorem about norm convergence, by D. Shlyakhtenko[1]

Lemma Let (A, τ) be a C^*-algebra with a faithful trace τ, and consider B to be the universal C^*-algebra generated by A and elements $L^{(1)}, \dots, L^{(n)}$ satisfying $L^{(i)*} x L^{(j)} = \delta_{i=j} \tau(x)$ for all $x \in A$. Moreover, consider the linear functional ψ determined on $* - \mathrm{Alg}(A, \{L^{(j)}\}_j)$ by:

$\psi|_A = \tau$,

$\psi(x_0 L^{(i_1)} x_1 \cdots x_{k-1} L^{(i_k)} x_k y_0 L^{(j_1)*} y_1 \cdots y_{l-1} L^{(j_l)*} y_l) = 0$ whenever $x_1, \dots, x_k, y_0, \dots, y_l \in A$ and at least one of k and l is nonzero.

Then ψ extends to a state on B having a faithful GNS representation. Moreover, $(B, \psi) \cong (A, \tau) * (\mathcal{E}, \phi)$ where $(\mathcal{E}, \phi)$ is the C^*-algebra generated by n free creation operators $\ell_1, \dots, \ell_n$ on the full Fock space $\mathcal{F}(\mathbb{C}^n)$ and ϕ is the vacuum expectation.

Sketch of proof. Consider the A, A-Hilbert bimodule $\mathcal{H} = L^2(A, \tau) \otimes A$ with the inner product

$$\langle \xi \otimes a, \xi' \otimes a' \rangle_A = \langle \xi, \xi' \rangle_{L^2(\tau)} a^* a'$$

and the left and right A actions given by

$$x \cdot (\xi \otimes a) \cdot y = x\xi \otimes ay.$$

[1]Research supported by NSF grant DMS-0900776

Let B be the extended Cuntz-Pimsner algebra associated to $\mathcal{H}^{\oplus n}$ (see [30]), i.e. the universal C^*-algebra generated by A and operators L_h : $h \in \mathcal{H}$ satisfying the relations

$$\begin{aligned} L_h^* L_g &= \langle h, g\rangle_A, \qquad h, g \in \mathcal{H}^{\oplus n} \\ a L_h b &= L_{ahb}, \qquad h \in \mathcal{H}^{\oplus n},\ a, b \in A. \end{aligned}$$

It follows from the results of [32] that if we denote by $(\hat{B}, \hat{\psi})$ the free product $(A, \tau) * (\mathcal{E}, \phi)$, then:

$$\begin{gathered} \ell_i^* x \ell_j = \delta_{i=j}\tau(x), \qquad \forall x \in A, \\ \hat{\psi}(x_0 \ell_{i_1} x_1 \cdots x_{k-1} \ell_{i_k} x_k y_0 \ell_{j_1}^* y_1 \cdots y_{l-1} \ell_{j_l}^* y_l) = 0, \\ \forall x_1, \ldots, x_k, y_1, \ldots, y_l \in A,\ k + l > 0 \end{gathered}$$

If $h = (\sum_i \xi_i^{(k)} \otimes a_i^{(k)})_{k=1}^n \in (A \otimes A)^{\oplus n} \subset \mathcal{H}^{\oplus n}$ is a finite tensor, write

$$\ell_h = \sum_{k,i} \xi_i^{(k)} \ell_k a_i^{(k)}.$$

It then follows that

$$\begin{aligned} \ell_h^* \ell_g &= \langle h, g\rangle_A, \qquad h, g \in \mathcal{H}^{\oplus n} \\ a \ell_h b &= \ell_{ahb}, \qquad a, b \in A,\ h \in \mathcal{H}^{\oplus n} \end{aligned}$$

which in particular means that $\|\ell_h\|_2^2 = \|\ell_h^* \ell_h\| = \|h\|^2$ so that the mapping $h \mapsto \ell_h$ is an isometry. We then extend ℓ to a map from $\mathcal{H}^{\oplus n}$ into $\hat{B}$. Note that the extension of ℓ still satisfies $a\ell_h b = \ell_{ahb}$ whenever $a, b \in A$ and $h \in \mathcal{H}^{\oplus n}$.

From this we see that (by the universal property of B) there exists a $*$-homomorphism $\pi : B \to \hat{B}$, so that $\psi = \hat{\psi} \circ \pi$. Thus all we need to prove is that π is injective. But by [30, Prop. 3.3], it follows that B is isomorphic to the Toeplitz algebra $\mathcal{T}$ (since in this case obviously $\langle \mathcal{H}^{\oplus n}, \mathcal{H}^{\oplus n}\rangle_A = A$) acting on the Fock space $\mathcal{F} = \bigoplus_{k \geqslant 0} (\mathcal{H}^{\oplus n})^{\otimes_A k}$. If we denote by E the canonical conditional expectation from $\mathcal{T}$ onto A and consider the state $\theta = \tau \circ E$, then the resulting Hilbert space is the closure of $\mathcal{F}$ in the (faithful) norm $\|\xi\| = \tau(\langle \xi, \xi\rangle_A)^{1/2}$; from this we see that the GNS representation of B associated to the state θ on B is faithful. Since $\hat{B}$ is exactly this GNS representation, it follows that π is injective. $\square$

If A_N is a sequence of C^*-algebras and $\omega \in \beta\mathbb{N} \setminus \mathbb{N}$ is a free ultrafilter, we shall denote by

$$\mathfrak{A} = \prod^{\omega} A_N$$

the quotient

$$\prod^{\omega} A_N = \left(\prod_{N=1}^{\infty} A_N\right) / \left\{ (a_j)_{N=1}^{\infty} : \lim_{N\to\omega} \|a_N\| = 0 \right\}.$$

Then $\mathfrak{A}$ is a C^*-algebra.

Let now $X_N^{(j)}$, $j = 1, \ldots, n$, $N = 1, 2, \ldots$ be self-adjoint random variables and assume that $X^{(j)}$, $j = 1, \ldots, n$ are such that for any non-commutative polynomial P,

$$\begin{aligned} \tau_N(P(X_N^{(1)}, \ldots, X_N^{(n)})) &\to \tau(P(X^{(1)}, \ldots, X^{(n)})) \\ \|P(X_N^{(1)}, \ldots, X_N^{(n)})\| &\to \|P(X^{(1)}, \ldots, X^{(n)})\|. \end{aligned}$$

Let $L^{(j)}$, $j = 1, \ldots, n$ be a family of free creation operators, free from each other and from $\{X_N^{(j)}\}_{N,j} \cup \{X^{(j)}\}_j$. In other words, they satisfy:

$$L^{(j)*} x L^{(j)} = \tau(x), \qquad \forall x \in C^*(\{X_N^{(j)}\}_{N,j} \cup \{X^{(j)}\}_j)$$

We use the notations

$$\begin{aligned} A_N &= C^*(X_N^{(1)}, \ldots, X_N^{(n)}), \qquad B_N = C^*(X_N^{(1)}, \ldots, X_N^{(n)}, L^{(1)}, \ldots, L^{(n)}) \\ A &= C^*(X^{(1)}, \ldots, X^{(n)}), \qquad B = C^*(X^{(1)}, \ldots, X^{(n)}, L^{(1)}, \ldots, L^{(n)}) \end{aligned}$$

and we denote by τ_N and ψ_N the respective states on A_N and B_N ($\cong (A_N, \tau_N) * (\mathcal{E}, \phi)$). We denote by τ and ψ the respective states on A and B ($\cong (A, \tau) * (\mathcal{E}, \phi)$).

Consider now the ultrapowers

$$\mathfrak{A} = \prod^{\omega} A_N \subset \mathfrak{B} = \prod^{\omega} B_N.$$

The formula

$$\psi : (x_N)_{N=1}^{\infty} \mapsto \lim_{N\to\omega} \psi_N(x_N)$$

defines a state on $\mathfrak{B}$.

We shall denote by $\hat{X}^{(j)} \in \mathfrak{A}$ the sequence $(X_N^{(j)})_{j=1}^{N}$. Then by assumption, we have that the map α taking $X^{(j)}$ to $\hat{X}^{(j)}$ extends to a state-preserving isomorphism from (A, τ) into $\mathcal{B}$ with range $\hat{A} = C^*(\hat{X}^{(1)}, \ldots, \hat{X}^{(n)})$.

We shall also denote by $\hat{L}^{(j)}$ the constant sequence $(L^{(j)})_{N=1}^{\infty} \in \mathfrak{B}$. Then for any element of $\hat{A}$ represented by the sequence $x = (x_N)_{N=1}^{\infty}$ we have:

$$\hat{L}^{(j)*} x \hat{L}^{(i)} = \delta_{i=j} (\tau_N(x_N))_{N=1}^{\infty}$$

which (since the L^2 and operator norms coincide on multiples of identity) is equal to $\tau(x)1\delta_{i=j} \in \mathfrak{A}$. It follows from the universality property that

$$\hat{B} \stackrel{\text{def}}{=} C^*(\hat{X}^{(1)}, \ldots, \hat{X}^{(n)}, \hat{L}^{(1)}, \ldots, \hat{L}^{(n)})$$

is a quotient of $(A, \tau) * (\mathcal{E}, \phi)$, the quotient map β determined by the fact that it is α on A and takes ℓ_j to $\hat{L}^{(j)}$. On the other hand, if we consider the GNS-representation π of $\hat{B}$ with respect to the restriction of ψ, we easily get (by freeness from $\hat{A}$ and $\{\hat{L}^{(j)}\}_j$) that the image is isomorphic to $(A, \tau) * (\mathcal{E}, \phi)$. Thus $\pi \circ \beta = \mathrm{id}$ so that actually

$$\beta : (A, \tau) * (\mathcal{E}, \phi) \to \hat{B} = C^*(\hat{X}^{(1)}, \ldots, \hat{X}^{(n)}, \hat{L}^{(1)}, \ldots, \hat{L}^{(n)})$$

is an isomorphism.

Consider now a non-commutative $*$-polynomial P. Then

$$\begin{aligned}
&\|P(X^{(1)}, \ldots, X^{(n)}, \ell^{(1)}, \ldots \ell^{(n)})\|_{(A,\tau)*(\mathcal{E},\phi)} \\
&= \|P(\hat{X}^{(1)}, \ldots, \hat{X}^{(n)}, \hat{L}^{(1)}, \ldots, \hat{L}^{(n)})\|_{\mathfrak{B}} \\
&= \lim_{N \to \omega} \|P(X_N^{(1)}, \ldots, X_N^{(n)}, L^{(1)}, \ldots, L^{(n)})\|_{B_N}.
\end{aligned}$$

Since the left hand side does not depend on ω, we have proved:

Theorem A.1. *Let $X_N^{(j)} \in (A_N, \tau_N)$, $j = 1, \ldots, n$, $N = 1, 2, \ldots$ be self-adjoint random variables and assume that $X^{(j)} \in (A, \tau)$, $j = 1, \ldots, n$ are such that for any non-commutative polynomial P,*

$$\begin{aligned}
\tau(P(X_N^{(1)}, \ldots, X_N^{(n)})) &\to \tau(P(X^{(1)}, \ldots, X^{(n)})) \\
\|P(X_N^{(1)}, \ldots, X_N^{(n)})\|_{A_N} &\to \|P(X^{(1)}, \ldots, X^{(n)})\|_A.
\end{aligned}$$

Let $(\ell_1, \ldots, \ell_n) \in \mathcal{E}$ be free creation operators, and let $B_N = (\mathcal{E}, \phi) * (A_N, \tau_N)$, $B = (\mathcal{E}, \phi) * (A, \tau)$. Assume that the traces τ_j are faithful. Then for any non-commutative $*$-polynomial Q,

$$\|Q(X_N^{(1)}, \ldots, X_N^{(n)}, \ell_1, \ldots, \ell_n)\|_{B_N} \to \|Q(X^{(1)}, \ldots, X^{(n)}, \ell_1, \ldots, \ell_n)\|_B.$$

It should be noted that if $S_1, \ldots, S_n$ are free semicircular variables, free from $\{X_N^{(j)}\}_{N,j} \cup \{X^{(j)}\}_j$, then $C_N = C^*(X_N^{(1)}, \ldots, X_N^{(n)}, S_1, \ldots, S_n)$ is isometrically contained in B_N, while $C = C^*(X^{(1)}, \ldots, X^{(n)}, S_1, \ldots, S_n)$ is isometrically contained in B. Thus the analog of Theorem A with ℓ_j's replaced by a free semicircular family also holds.

References

[1] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*, volume 118 of *Cambridge studies in advanced mathematics*. Cambridge University Press, 2010.

[2] L. Arnold. On the asymptotic distribution of the eigenvalues of random matrices. *J. Math. Anal. Appl.*, 20:262–268, 1967.

[3] Z. D. Bai. Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica*, 9(3):611–677, 1999. With comments by G. J. Rodgers and Jack W. Silverstein; and a rejoinder by the author.

[4] Z. D. Bai, J. W. Silverstein, and Y. Q. Yin. A note on the largest eigenvalue of a large-dimensional sample covariance matrix. *J. Multivariate Anal.*, 26(2):166–168, 1988.

[5] Z. D. Bai and Jack W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.*, 26(1):316–345, 1998.

[6] Z. D. Bai and Y. Q. Yin. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *Ann. Probab.*, 16(4):1729–1741, 1988.

[7] Z. D. Bai, Y. Q. Yin, and P. R. Krishnaiah. On limiting spectral distribution of product of two random matrices when the underlying distribution is isotropic. *J. Multivariate Anal.*, 19(1):189–200, 1986.

[8] N. P. Brown and N. Ozawa. *C^*-algebras and finite-dimensional approximations*, volume 88 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2008.

[9] M. Capitaine and M. Casalis. Asymptotic freeness by generalized moments for Gaussian and Wishart matrices. Application to beta random matrices. *Indiana Univ. Math. J.*, 53(2):397–431, 2004.

[10] M. Capitaine and C. Donati-Martin. Strong asymptotic freeness for Wigner and Wishart matrices. *Indiana Univ. Math. J.*, 56(2):767–803, 2007.

[11] L. H. Y. Chen. An inequality for the multivariate normal distribution. *J. Multivariate Anal.*, 12(2):306–315, 1982.

[12] J. B. Conway. *A course in operator theory*, volume 21 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2000.

[13] K. Dykema. On certain free product factors via an extended matrix model. *J. Funct. Anal.*, 112(1):31–60, 1993.

[14] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.

[15] S. Geman. A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2):252–261, 1980.

[16] U. Grenander and J. W. Silverstein. Spectral analysis of networks with random topologies. *SIAM J. Appl. Math.*, 32(2):499–519, 1977.

[17] A. Guionnet. *Large random matrices: lectures on macroscopic asymptotics*, volume 1957 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2009. Lectures from the 36th Probability Summer School held in Saint-Flour, 2006.

[18] A. Guionnet, M. Krishnapur, and O. Zeitouni. The single-ring theorem. arXiv:0909.2214v1, preprint, http://arxiv4.library.cornell.edu/abs/0909.2214.

[19] U. Haagerup and S. Thorbjørnsen. A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(\mathbb{F}_2))$ is not a group. *Ann. of Math. (2)*, 162(2):711–775, 2005.

[20] H. Hamburger. Über eine Erweiterung des Stieltjesschen Momentenproblems. *Math. Ann.*, 82(3-4):168–187, 1921.

[21] F. Hiai and D. Petz. Asymptotic freeness almost everywhere for random matrices. *Acta Sci. Math. (Szeged)*, 66(3-4):809–834, 2000.

[22] D. Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.*, 12(1):1–38, 1982.

[23] D. Jonsson. On the largest eigenvalue of a sample covariance matrix. In *Multivariate analysis VI (Pittsburgh, Pa., 1983)*, pages 327–333. North-Holland, Amsterdam, 1985.

[24] F. Juhász. On the spectrum of a random graph. In *Algebraic methods in graph theory, Vol. I, II (Szeged, 1978)*, volume 25 of *Colloq. Math. Soc. János Bolyai*, pages 313–316. North-Holland, Amsterdam, 1981.

[25] E. G. Larsson and P. Stoica. *Space-time block coding for wireless communications*. Cambridge University Press, Cambridge, 2003.

[26] F. Lehner. Computing norms of free operators with matrix coefficients. *Amer. J. Math.*, 121(3):453–486, 1999.

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

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

[29] D. Paul and J. W. Silverstein. No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix. *J. Multivariate Anal.*, 100(1):37–57, 2009.

[30] M. V. Pimsner. A class of C^*-algebras generalizing both Cuntz-Krieger algebras and crossed products by $\mathbf{Z}$. In *Free probability theory (Waterloo, ON, 1995)*, volume 12 of *Fields Inst. Commun.*, pages 189–212. Amer. Math. Soc., Providence, RI, 1997.

[31] H. Schultz. Non-commutative polynomials of independent Gaussian random matrices. The real and symplectic cases. *Probab. Theory Related Fields*, 131(2):261–309, 2005.

[32] D. Shlyakhtenko. Some applications of freeness with amalgamation. *J. Reine Angew. Math.*, 500:191–212, 1998.

[33] J. W. Silverstein. The smallest eigenvalue of a large-dimensional Wishart matrix. *Ann. Probab.*, 13(4):1364–1368, 1985.

[34] J. W. Silverstein. On the weak limit of the largest eigenvalue of a large-dimensional sample covariance matrix. *J. Multivariate Anal.*, 30(2):307–311, 1989.

[35] S. Thorbjørnsen. Mixed moments of Voiculescu's Gaussian random matrices. *J. Funct. Anal.*, 176(2):213–246, 2000.

[36] A. M. Tulino and S. Verdú. *Random matrices and wireless communications*, volume vol. 1 of *Fundations and Trends in Communications and Information Theory*. Now Publishers Inc., 2004.

[37] D. Voiculescu. Limit laws for random matrices and free products. *Invent. Math.*, 104(1):201–220, 1991.

[38] D. Voiculescu. Operations on certain non-commutative operator-valued random variables. *Astérisque*, (232):243–275, 1995. Recent advances in operator algebras (Orléans, 1992).

[39] D. Voiculescu. A strengthened asymptotic freeness result for random matrices with applications to free entropy. *Internat. Math. Res. Notices*, (1):41–63, 1998.

[40] K. W. Wachter. The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probability*, 6(1):1–18, 1978.

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

[42] Y. Q. Yin. Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.*, 20(1):50–68, 1986.

[43] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah. On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields*, 78(4):509–521, 1988.

[44] Y. Q. Yin and P. R. Krishnaiah. A limit theorem for the eigenvalues of product of two random matrices. *J. Multivariate Anal.*, 13(4):489–507, 1983.