



**HAL**  
open science

# Norm of polynomials in large random and deterministic matrices

Camille Male

► **To cite this version:**

Camille Male. Norm of polynomials in large random and deterministic matrices. Probability Theory and Related Fields, 2011, pp.41. 10.1007/s00440-011-0375-2 . hal-00494600v1

**HAL Id: hal-00494600**

**<https://hal.science/hal-00494600v1>**

Submitted on 23 Jun 2010 (v1), last revised 25 Sep 2012 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Norm of polynomials in large random and deterministic matrices

C. Male\*

ABSTRACT:

*In this article we show the convergence of the operator norm (the largest singular value) of any polynomial in some random matrices. If  $\mathbf{X}_N = (X_1^{(N)}, \dots, X_p^{(N)})$  denotes a family of  $N \times N$  independent random matrices from the Gaussian Unitary Ensemble (GUE), then we state sufficient conditions on (possibly random) matrices  $\mathbf{Y}_N = (Y_1^{(N)}, \dots, Y_q^{(N)})$ , independent with  $\mathbf{X}_N$ , for which  $\|P(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)\|$  converges for all polynomial  $P$ . Limits are described by operator norms of objects from free probability theory. We give examples of diagonal matrices  $\mathbf{Y}_N$  for which the convergence holds. Convergence of the operator norm is shown to hold for some block matrices, even with rectangular gaussian blocks, a situation including non-white Wishart matrices.*

## 1 Statement of results

Let  $\mathbf{X}_N = (X_1^{(N)}, \dots, X_p^{(N)})$  be independent  $N \times N$  matrices of the normalized Gaussian Unitary Ensemble (GUE), i.e. for each  $j = 1, \dots, p$ , the matrix  $X_j^{(N)}$  is selfadjoint with entries

$$X_j^{(N)} = (X_{n,m})_{1 \leq n, m \leq N}$$

the set of random variables  $(X_{n,n})_{1 \leq n \leq N}$ , and  $(\sqrt{2}\operatorname{Re}(X_{n,m}), \sqrt{2}\operatorname{Im}(X_{n,m}))_{1 \leq n < m \leq N}$  forms a centered Gaussian vector with covariance matrix  $\frac{1}{N}\mathbf{1}_{N^2}$ .

Let  $\mathbf{Y}_N = (Y_1^{(N)}, \dots, Y_q^{(N)})$  be  $N \times N$  random matrices. Some assumptions on the generalized moments of this family, namely the convergence of its non commutative law, will be made in Section 3 after a recall in free probability theory.

Voiculescu showed in [26] that when the size of the matrices goes to infinity, the generalized moments of the family  $(\mathbf{X}_N, \mathbf{Y}_N)$  can be described by a non negative linear form  $\tau$  on the set  $\mathbb{C}\langle \mathbf{x}, \mathbf{y}, \mathbf{y}^* \rangle$  of non commutative polynomials in the  $p + 2q$  non commutative indeterminates  $\mathbf{x} = (x_1, \dots, x_p)$ ,  $\mathbf{y} = (y_1, \dots, y_q)$ ,  $\mathbf{y}^* = (y_1^*, \dots, y_q^*)$ : for all  $P$  in  $\mathbb{C}\langle \mathbf{x}, \mathbf{y}, \mathbf{y}^* \rangle$ ,

$$\lim_{N \rightarrow \infty} \tau_N \left[ P(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) \right] = \tau \left[ P(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) \right], \quad (1.1)$$

almost surely and in expectation, where  $\tau_N = \frac{1}{N} \operatorname{Tr}_N$  is the normalized trace.

A different problem is the question of convergence with respect to the operator norm instead of convergence relative to the trace in (1.1); in the context of free probability, that is known as strong convergence. Let  $\|P(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)\|$  denote the operator norm of  $P(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)$ , i.e. the norm induced by the Euclidean norm in  $\mathbb{C}^N$  or, equivalently, the largest singular value of  $P(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)$  which is equal to the root square of the largest eigenvalue of  $P(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)^* P(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)$ . For the polynomial  $P(\mathbf{x}, \mathbf{y}, \mathbf{y}^*)$  in (1.1), the operator norm is defined by

$$\|P(\mathbf{x}, \mathbf{y}, \mathbf{y}^*)\|_\tau := \lim_{k \rightarrow \infty} \left( \tau \left[ (P(\mathbf{x}, \mathbf{y}, \mathbf{y}^*)^* P(\mathbf{x}, \mathbf{y}, \mathbf{y}^*))^k \right] \right)^{1/2k}. \quad (1.2)$$

---

\*UMPA, ENS de Lyon, 46 allée d'Italie, 69007 Lyon, France. camille.male@umpa.ens-lyon.fr

The problem we address in this paper is to give sufficient conditions on the matrices  $\mathbf{Y}_N$  so that

$$\lim_{N \rightarrow \infty} \|P(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)\| = \|P(\mathbf{x}, \mathbf{y}, \mathbf{y}^*)\|_\tau, \text{ almost surely.} \quad (1.3)$$

The asymptotic behavior of extremal singular values or eigenvalues of large random matrices has been a longstanding question. It originally appeared in multivariate analysis, to resolve technical difficulties for computation of statistics written as integrals of unbounded functions with respect to the empirical spectral distribution of a random matrix, via the Helly-Bray theorem, .

The first results on the convergence of the largest eigenvalue for a Wigner or a Wishart matrix when the size goes to the infinity appeared in the early 1980's in the works of Geman [11], Juhász [17], Füredi and Komlós [10], Jonsson [16] and Silverstein [23]. In 1988, in the case of a Wigner matrix, by Bai and Yin stated in [5] 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 result is due to Yin, Bai, and Krishnaiah in [28] and Bai, Silverstein, and Yin [3]. The case of a complex matrix has been investigated later by Bai [2]. In this series of papers, where the assumptions on the matrices were progressively relaxed up to the optimal ones, proofs were basically combinatorial, and based on the truncation of entries.

The method of Stieltjes transform may be applied to investigate the support of the limiting empirical spectral distribution of a random matrix. In 1998 Bai and Silverstein showed [4] for large sample covariance matrices the convergence of the support of the limiting distribution.

In 2005, Haagerup and Thorbjørnsen showed [15] in the case where the matrices  $\mathbf{Y}_N$  are zero that the convergence in (1.3) holds. Their method (which will be presented in Section 3) is based on a linearization trick and on free harmonic analysis of some Stieltjes transforms. It has been used by Schultz to obtain the same result for Gaussian random matrices with real or symplectic entries [22], and by Capitaine and Donati-Martin for non Gaussian Wigner matrices with symmetric distribution of the entries satisfying a Poincaré's inequality and for Wishart matrices [7]. Some details of the method have been improved by Haagerup, Schultz, and Thorbjørnsen in [14] to extend the class of random matrices that satisfy Bai and Silverstein's "no eigenvalues outside the limiting spectra" phenomena [4].

This article is mainly devoted to the following theorem.

**Theorem 1** (Strong asymptotic freeness). *In Section 3 we make precise assumptions on (1) the generalized moments of  $\mathbf{Y}_N$ , (2) the concentration of the entries of matrices  $\mathbf{Y}_N$ , and on (3) the good rate of convergence for some Stieltjes transforms, under which for all polynomial  $P$  in  $p + 2q$  non commutative variables we have almost surely*

$$\lim_{N \rightarrow \infty} \|P(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)\| = \|P(\mathbf{x}, \mathbf{y}, \mathbf{y}^*)\|_\tau. \quad (1.4)$$

The first and the simpler matrix model that may be investigated to play the role of matrices  $\mathbf{Y}_N$  in Theorem 1 consists in diagonal matrices with real entries. Let  $\mu_1, \dots, \mu_q$  be probability measures in  $\mathbb{R}$  with compact support. For  $j = 1, \dots, q$ , denote by  $F_j$  the cumulative distribution function of  $\mu_j$  defined by: for all  $t$  in  $\mathbb{R}$ ,  $F_j(t) = \mu_j(]-\infty, t])$ . Let  $F_j^{-1}$ ,  $j = 1, \dots, q$  be the generalized inverse of  $F_j$ : for all  $u$  in  $]0, 1]$ ,  $F_j^{-1}(u) = \inf \{t \in \mathbb{R} \mid F_j(t) \geq u\}$  and<sup>1</sup>  $F_j^{-1}(0) = \lim_{u \rightarrow 0^+} F_j^{-1}(u)$ . Let  $\mathbf{D}_N = (D_1^{(N)}, \dots, D_q^{(N)})$  be the family of  $N \times N$  deterministic diagonal matrices defined by: for  $j = 1, \dots, q$

$$D_j^{(N)} = \text{diag} \left( F_j^{-1} \left( \frac{0}{N} \right), \dots, F_j^{-1} \left( \frac{N-1}{N} \right) \right). \quad (1.5)$$

Then the empirical spectral distribution of  $D_j^{(N)}$  converges weakly to the measure  $\mu_j$ , and hence the convergence (1.1) holds by Voiculescu's theorem for a certain linear form  $\tau$  on non commutative polynomials in selfadjoint indeterminates  $(\mathbf{x}, \mathbf{d}) = (x_1, \dots, x_p, d_1, \dots, d_q)$ . We show in Section 5.1 the following:

**Corollary 1** (diagonal matrices). *If for  $j = 1 \dots q$  the support of  $\mu_j$  is a single interval then Theorem 1 holds for  $\mathbf{D}_N$ : for all non commutative polynomial  $P \in \mathbb{C}\langle \mathbf{x}, \mathbf{d} \rangle$ ,*

$$\lim_{N \rightarrow \infty} \|P(\mathbf{X}_N, \mathbf{D}_N)\| = \|P(\mathbf{x}, \mathbf{d})\|_\tau.$$

---

<sup>1</sup>This definition made  $F_j^{-1}$  continuous at 0

Theorem 1 may be used to deduce the same result for Wishart matrices as for the GUE matrices. Let  $r \geq s$  be integers and let  $\mathbf{W}_N = (W_1^{(N)}, \dots, W_p^{(N)})$  be independent  $rN \times rN$  matrices of the normalized Wishart ensemble with parameter  $r/s$ , i.e. for each  $j = 1, \dots, p$ ,  $W_j^{(N)} = M_j^{(N)} M_j^{(N)*}$  where  $M_j^{(N)}$  is a  $rN \times sN$  matrix whose entries are random variables,

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

and the random variables  $(\sqrt{2}\operatorname{Re}(M_{n,m}), \sqrt{2}\operatorname{Im}(M_{n,m}))_{1 \leq n \leq rN, 1 \leq m \leq sN}$  form a centered Gaussian vector with covariance matrix  $\frac{1}{rN} \mathbf{1}_{2rsN^2}$ . Consider  $rN \times rN$  random matrices  $\mathbf{Y}_N = (Y_1^{(N)}, \dots, Y_q^{(N)})$  satisfying the assumptions of Theorem 1 or of the diagonal form of Corollary 1.

**Corollary 2** (Wishart matrices). *For all polynomial  $P$  in  $p + 2q$  non commutative variables*

$$\lim_{N \rightarrow \infty} \|P(\mathbf{W}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)\| = \|P(\mathbf{w}, \mathbf{y}, \mathbf{y}^*)\|_\tau, \quad a.s. \quad (1.6)$$

for a non negative linear form  $\tau$ .

Let  $\Sigma_N = (\Sigma_1^{(N)}, \dots, \Sigma_p^{(N)})$  be a family of  $rN \times rN$  non negative definite Hermitian matrix and define  $\mathbf{Z}_N = (Z_1^{(N)}, \dots, Z_p^{(N)})$  by: for  $j = 1, \dots, p$

$$Z_j^{(N)} = (\Sigma_j^{(N)})^{1/2} W_j^{(N)} (\Sigma_j^{(N)})^{1/2}. \quad (1.7)$$

The matrices  $Z_1^{(N)}, \dots, Z_p^{(N)}$  are called non-white Wishart matrices. Then if the matrices  $((\Sigma_1^{(N)})^{1/2}, \dots, (\Sigma_p^{(N)})^{1/2})$  are diagonal as in Corollary 1 we get the convergence of  $\|P(Z_1^{(N)}, \dots, Z_p^{(N)})\|$  for any polynomial  $P$ . For Wishart matrices with rational parameter, this extends the result of Capitaine and Donati-Martin [7], where the convergence of the norm of polynomials in white Wishart matrices (where  $\Sigma_j^{(N)} = \mathbf{1}_{rN}$  and  $\mathbf{Y}_N = 0$ ) has been investigated.

It will be shown as a consequence of Theorem 1 that the convergence of norms (1.4) also holds for block matrices.

**Corollary 3** (Block matrices). *Let  $\mathbf{X}_N, \mathbf{Y}_N, \tau$  be as in Theorem 1. For all  $\ell$  in  $\mathbb{N}$ , let  $(P_{u,v})_{1 \leq u, v \leq \ell}$  be non commutative polynomials. Then the operator norm of the  $\ell N \times \ell N$  block matrix*

$$\begin{pmatrix} P_{1,1}(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) & \dots & P_{1,\ell}(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) \\ \vdots & & \vdots \\ P_{\ell,1}(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) & \dots & P_{\ell,\ell}(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) \end{pmatrix}, \quad (1.8)$$

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

$$\begin{pmatrix} P_{1,1}(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) & \dots & P_{1,\ell}(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) \\ \vdots & & \vdots \\ P_{\ell,1}(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) & \dots & P_{\ell,\ell}(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) \end{pmatrix}, \quad (1.9)$$

where the linear form  $(\tau_\ell \otimes \tau)$  is given by

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

In the context of communication, rectangular block random matrices are sometimes investigated for the study of wireless Multiple-input Multiple-Output (MIMO) systems [18, 25]. In the case of Intersymbol-

Interference the channel matrix  $H$  reflects the channel effect during a transmission and is of the form

$$H = \begin{pmatrix} A_1 & A_2 & \dots & A_L & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & A_1 & A_1 & \dots & A_L & \mathbf{0} & & \vdots \\ \vdots & \mathbf{0} & A_1 & A_2 & \dots & A_L & \mathbf{0} & \\ & & \ddots & \ddots & \ddots & & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & & \mathbf{0} \\ \mathbf{0} & \dots & \dots & \mathbf{0} & A_1 & A_2 & \dots & A_L \end{pmatrix}, \quad (1.10)$$

where  $L$  is the length of the impulse response of the channel,  $(A_l)_{1 \leq l \leq L}$  are  $n_R \times n_T$  matrices that are very often modeled by random matrices,  $n_T$  is the number of transmitter antenna and  $n_R$  is the number of receiver antenna.

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 provides a large class of matrices of the form (1.10) for which the norm convergence holds i.e. the extremal singular values converge to the extrema of the singular spectra. For instance we investigate in Section 5.4 the following case:

**Corollary 4** (Rectangular band matrices). *Let  $r, t \geq 1$  be integers. Let  $H$  be the matrix given by (1.10) where  $n_R = rN$ ,  $n_T = tN$ , the matrices  $A_1, \dots, A_L$  are independent and for  $l = 1 \dots L$  the entries of the matrix  $A_l$  are independent identically distributed according to the complex Gaussian measure with mean  $m_l \in \mathbb{C}$  and variance  $\sigma_l^2/N$  ( $\sigma \geq 0$ ). Then the norm of  $H$  converges.*

The convergence also holds when the matrices  $A_l$ , for  $l = 1 \dots L$  are multiplied on the left and on the right by  $rN \times rN$  and respectively  $tN \times tN$  diagonal matrices of the form of Corollary 1.

In Section 2, we recall basic facts in free probability theory. Section 3 is dedicated to the method of Haagerup and Thorbjørnsen [15], and we introduce some notations to give the precise assumptions on the non Gaussian matrices  $\mathbf{Y}_N$ . We also give a sketch of the proof of Theorem 1. Theorem 1 is proven in Section 4. Section 5 contains the proofs of Corollaries 1-4.

**Acknowledgments:** We would like to thank Alice Guionnet for dedicating much time over many discussions to the subjects of this paper and, along with Manjunath Krishnapur and Ofer Zeitouni, for the communication of Lemma 15.

## 2 Recall on free probability theory

Even if this paper concerns random matrices, it is relevant to describe the limits in the context of free probability. We remind for reader convenience the following basic facts from this theory (see [21] and [1] for details).

1. A  $*$ -probability space  $(\mathcal{A}, *, \tau)$  consists in 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 trace  $\tau$ . A trace  $\tau$  is a unital linear functional  $\tau : \mathcal{A} \mapsto \mathbb{C}$  satisfying

$$\tau[\mathbf{1}] = 1, \quad \tau[ab] = \tau[ba], \quad \forall a, b \in \mathcal{A}. \quad (2.1)$$

The elements of  $\mathcal{A}$  are called non commutative random variables, or in short non commutative variables. We will always suppose that  $\tau$  is a tracial state, that means

$$\tau[a^*a] \geq 0, \quad \forall a \in \mathcal{A}. \quad (2.2)$$

The tracial state  $\tau$  will always supposed to be faithful, i.e. satisfies  $\tau[a^*a] = 0$  only if  $a = 0$ . A tracial state  $\tau$  satisfies the Cauchy-Schwarz inequality

$$|\tau[b^*a]|^2 \leq \tau[a^*a]\tau[b^*b], \quad \forall a, b \in \mathcal{A}. \quad (2.3)$$

2. Given  $(\mathcal{A}, *, \tau)$  a  $*$ -probability space, we can define a norm  $\|\cdot\|_\tau$  on  $\mathcal{A}$  by

$$\|a\|_\tau = \lim_{k \rightarrow \infty} \left( \tau[(a^*a)^k] \right)^{\frac{1}{2k}}, \forall a \in \mathcal{A}. \quad (2.4)$$

Then  $\|\cdot\|_\tau$  satisfies

$$\|ab\|_\tau \leq \|a\|_\tau \|b\|_\tau, \forall a, b \in \mathcal{A}, \quad (2.5)$$

$$\|a^*a\|_\tau = \|a\|_\tau^2, \forall a \in \mathcal{A}. \quad (2.6)$$

3. We denote by  $\mathbb{C}\langle \mathbf{x}, \mathbf{x}^* \rangle$  the set of non commutative polynomials in non commutative indeterminates  $\mathbf{x} = (x_1, \dots, x_p)$ ,  $\mathbf{x}^* = (x_1^*, \dots, x_p^*)$ . We define the non commutative law of a family  $\mathbf{a} = (a_1, \dots, a_p)$  of elements of  $\mathcal{A}$  as the linear functional  $\tau_{\mathbf{a}} : \mathbb{C}\langle \mathbf{x}, \mathbf{x}^* \rangle \rightarrow \mathbb{C}$  determined by

$$\tau_{\mathbf{a}}[P] = \tau[P(\mathbf{a}, \mathbf{a}^*)]. \quad (2.7)$$

In the following we write  $\tau$  a shorthand for  $\tau_{\mathbf{a}}$  when there is no danger of confusion. For convenience we use the notation  $P(\mathbf{a}, \mathbf{a}^*) = P(a_1, \dots, a_p, a_1^*, \dots, a_p^*)$ .

4. The families of non commutative 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  $P_1, \dots, P_K$

$$\tau \left[ P_1(\mathbf{a}_{i_1}, \mathbf{a}_{i_1}^*) \dots P_K(\mathbf{a}_{i_K}, \mathbf{a}_{i_K}^*) \right] = 0 \quad (2.8)$$

as soon as  $i_1 \neq i_2 \neq \dots \neq i_K$  and  $\tau[P_k(\mathbf{a}_{i_k}, \mathbf{a}_{i_k}^*)] = 0$  for  $k = 1, \dots, K$ . For free families  $\mathbf{a}_1, \dots, \mathbf{a}_n$ , the non commutative laws  $\tau_{\mathbf{a}_1}, \dots, \tau_{\mathbf{a}_n}$  completely determines the joint law  $\tau_{(\mathbf{a}_1, \dots, \mathbf{a}_n)}$ .

5. A family  $\mathbf{x} = (x_1, \dots, x_p)$  in  $\mathcal{A}$  is called a semicircular family when the non commutative variables are free, selfadjoint ( $x_i = x_i^*$ ,  $i = 1, \dots, p$ ), and for all  $k$  in  $\mathbb{N}$  and  $i = 1, \dots, p$

$$\tau[x_i^k] = \int t^k d\sigma(t), \quad (2.9)$$

with  $d\sigma(t) = \frac{1}{2\pi} \sqrt{4-t^2} \mathbf{1}_{|t| \leq 2} dt$  the semicircle distribution.

6. For  $N$  in  $\mathbb{N} \cup \{\infty\}$ , let  $(\mathcal{A}_N, *, \tau_N)$  be a sequence of  $*$ -probability spaces and let  $\mathbf{a}_N = (a_1^{(N)}, \dots, a_n^{(N)})$  in  $\mathcal{A}_N$  be non commutative random variables. We say that the sequence  $(\mathbf{a}_N)_{N \in \mathbb{N}}$  converges in law (or in moment) to  $\mathbf{a}_\infty$  when for all  $P \in \mathbb{C}\langle \mathbf{x}, \mathbf{x}^* \rangle$

$$\lim_{N \rightarrow \infty} \tau_{\mathbf{a}_N}[P] = \tau_{\mathbf{a}_\infty}[P]. \quad (2.10)$$

We say that  $\mathbf{a}_\infty$  is the non commutative limit of  $(\mathbf{a}_N)_{N \in \mathbb{N}}$ . When the sequence of families  $(\mathbf{a}_1^{(N)}, \dots, \mathbf{a}_n^{(N)})$  converges in law to free families  $(\mathbf{a}_1, \dots, \mathbf{a}_n)$ , we say that the families are asymptotically free.

7. A  $C^*$ -probability space  $(\mathcal{B}, *, \tau, \|\cdot\|)$  consists in a  $*$ -probability space  $(\mathcal{B}, *, \tau)$  and a norm  $\|\cdot\|$  such that  $(\mathcal{B}, *, \|\cdot\|)$  is a  $C^*$ -algebra. By the Gelfand-Naimark-Segal construction, one can always realized  $\mathcal{B}$  as a norm-closed  $C^*$ -subalgebra of the algebra of bounded operators on a Hilbert space. The norm  $\|\cdot\|$  is unique and coincides with the norm  $\|\cdot\|_\tau$  defined in Equation (2.4). The framework of  $C^*$ -probability theory is a good one for a meaningful harmonic analysis (see section 4.2); for instance, if  $a$  is a non commutative variable in a  $C^*$ -probability space  $(\mathcal{B}, *, \tau, \|\cdot\|)$ , the Stieltjes (or Cauchy) transform of  $a$ , given by

$$\forall \lambda \in \mathbb{C}, \text{ such that } \text{Im } \lambda > 0, \quad g(\lambda) = \frac{1}{\lambda} + \sum_{n=1}^{\infty} \frac{\tau[a^n]}{\lambda^{n+1}} = \tau[(\lambda \mathbf{1} - a)^{-1}],$$

is an analytic function.

The set of  $N \times N$  matrices  $M_N(\mathbb{C})$  is a  $*$ -probability space when equipped with the conjugate transpose involution and tracial state  $\tau_N = \frac{1}{N} \text{Tr}_N$  the normalized trace;  $\text{Tr}_N$  denotes the usual trace in  $M_N(\mathbb{C})$ , sum of diagonal elements of a matrix. In this case, the norm  $\|\cdot\|_{\tau_N}$  defined by Equation (2.4) is the operator norm i.e. the norm induced by the Euclidean norm in  $\mathbb{C}^N$  and will be denoted in short  $\|\cdot\|$  in the rest of this article. Equivalently, for any matrices  $A$  in  $M_N(\mathbb{C})$  the number  $\|A\|$  is the largest singular value of  $A$  i.e. the root square of the largest eigenvalue of the matrix  $A^*A$  (consequence of the spectral theorem). To obtain an algebra in the case of random matrices, one has to suppose that the entries are random variables with finite moments of any order. Hence we have two possible frameworks, with  $\tau_N$  or  $\mathbb{E}[\tau_N \cdot]$  for the tracial state, where  $\mathbb{E}$  stands for the expectation.

### 3 Method, assumptions and sketch of the proof of Theorem 1

Let  $\mathbf{X}_N = (X_1^{(N)}, \dots, X_p^{(N)})$  be independent GUE matrices, and  $\mathbf{Y}_N = (Y_1^{(N)}, \dots, Y_q^{(N)})$  be random matrices, independent with  $\mathbf{X}_N$  that satisfy the two following assumptions (a third one will be assumed, and will be presented after some notations):

**Assumption 1** (Moments assumption). *The non commutative law of  $\mathbf{Y}_N$  in  $(M_N(\mathbb{C}), *, \tau_N)$  converges almost surely and in expectation to the law of non commutative variables  $\mathbf{y} = (y_1, \dots, y_q)$  in a  $C^*$ -probability space and there exists a  $D \geq 0$  such that for  $j = 1, \dots, q$  we have almost surely*

$$\limsup_{N \rightarrow \infty} \|Y_j^{(N)}\| < D \quad (3.1)$$

**Assumption 2** (Poincaré's inequality). *We suppose that there exists  $\sigma > 0$  such that for every  $N$ , the joint law of the entries of the matrices  $\mathbf{Y}_N$  satisfies a Poincaré's inequality with constant  $\sigma/N$  i.e. for all  $f: \mathbb{R}^{2qN^2} \rightarrow \mathbb{C}$ , function of the entries of  $q$  matrices, of class  $C^1$  and such that  $\mathbb{E} \left[ |f(\mathbf{Y}_N)|^2 \right] < \infty$  one has*

$$\text{Var} \left( f(\mathbf{Y}_N) \right) \leq \frac{\sigma}{N} \mathbb{E} \left[ \|\nabla f(\mathbf{Y}_N)\|^2 \right], \quad (3.2)$$

where  $\nabla f$  denotes the gradient of  $f$ ,  $\text{Var}$  denotes the variance,  $\text{Var}(x) = \mathbb{E} \left[ |x - \mathbb{E}[x]|^2 \right]$ .

Voiculescu's Theorem [26], Assumption 1 and the fact that  $\mathbf{X}_N$  and  $\mathbf{Y}_N$  are independent imply that the non commutative law of  $(\mathbf{X}_N, \mathbf{Y}_N)$  in  $(M_N(\mathbb{C}), *, \mathbb{E}[\tau_N \cdot])$  converges to the law of non commutative random variables  $(\mathbf{x}, \mathbf{y})$  in a  $C^*$ -probability space  $(\mathcal{B}, *, \tau, \|\cdot\|_{\tau})$ , where the trace  $\tau$  is defined by the facts that

- $\mathbf{x} = (x_1, \dots, x_p)$  is a semicircular system (Item 5 in Section 2),
- $\mathbf{y} = (y_1, \dots, y_q)$  is the non commutative limit of  $\mathbf{Y}_N$ ,
- $\mathbf{x}, \mathbf{y}$  are free (Item 4).

The convergence also holds almost surely when the random non commutative law of the matrices is viewed in  $(M_N(\mathbb{C}), *, \tau_N)$  [1, Theorem 5.4.5]: for all polynomial  $P$  in  $p + 2q$  non commutative variables,

$$\lim_{N \rightarrow \infty} \tau_N \left[ P(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) \right] = \tau \left[ P(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) \right], \quad \text{a. s. and in expectation,} \quad (3.3)$$

Hence it is easy to see that one has (see [15, Lemma 7.2])

$$\liminf_{N \rightarrow \infty} \|P(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)\| \geq \|P(\mathbf{x}, \mathbf{y}, \mathbf{y}^*)\|, \quad \text{a. s.} \quad (3.4)$$

It remains to show that the limsup is smaller than the right hand side in (3.4). The method of Haagerup and Thorbjørnsen described in [15] is carried out in three steps:

**Step 1. Linearization Trick:** Inequality (3.4) known, the question of almost sure convergence of the norm of any polynomial in random matrices can be reduced to the question of the convergence of the spectra of any matrix-valued selfadjoint degree one polynomials in these matrices (see [15, parts 2 and 7]). More precisely, in order to get Theorem 1, it is sufficient to show that given  $\varepsilon > 0$ ,  $k$  positive integer,  $L$  a self adjoint degree one polynomial in  $M_k(\mathbb{C}) \otimes \mathbb{C}\langle \mathbf{x}, \mathbf{y}, \mathbf{y}^* \rangle$ , then there exists  $N_0$  such that for all  $N \geq N_0$

$$\text{Sp} \left( L(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) \right) \subset \text{Sp} \left( L(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) \right) + (-\varepsilon, \varepsilon) \quad (3.5)$$

**Step 2. From Stieltjes transform to spectra:** here we use a slight modification of the method in [15] due to [1, subsection 5.5] and [13]. For convenience, we will recall its proof in Appendix B. Fix a selfadjoint degree one polynomial  $L \in M_k(\mathbb{C}) \otimes \mathbb{C}\langle \mathbf{x}, \mathbf{y}, \mathbf{y}^* \rangle$ , and denote by  $G_N^{(1)}$  and  $G^{(1)}$  the  $M_k(\mathbb{C})$ -valued Stieltjes transforms of  $L(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)$  and  $L(\mathbf{x}, \mathbf{y}, \mathbf{y}^*)$  respectively, given by

$$G_N^{(1)}(\Lambda) = \mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ (\Lambda \otimes \mathbf{1}_N - L(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*))^{-1} \right] \right] \in M_k(\mathbb{C}), \quad (3.6)$$

$$G^{(1)}(\Lambda) = \mathbb{E} \left[ (\text{id}_k \otimes \tau) \left[ (\Lambda \otimes \mathbf{1} - L(\mathbf{x}, \mathbf{y}, \mathbf{y}^*))^{-1} \right] \right] \in M_k(\mathbb{C}), \quad (3.7)$$

for all matrices  $\Lambda$  in  $M_k(\mathbb{C})$  such that  $\text{Im } \Lambda := \frac{1}{2i}(\Lambda - \Lambda^*)$  is definite positive (we will write in short  $\text{Im } \Lambda > 0$ ). Then in order to show (3.5), since the matrices  $(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)$  satisfy a Poincaré's inequality (by the tensor property), it is sufficient to show that: for every  $\varepsilon, \tilde{\varepsilon} > 0$ , there exist  $N_0, c, \eta, \gamma, \alpha > 0$  such that for all  $N \geq N_0$ , for all matrices  $\Lambda$  in  $M_k(\mathbb{C})$  of the form  $\Lambda = \lambda \mathbf{1}_k$  such that  $\varepsilon \leq (\text{Im } \lambda)^{-1} \leq N^\gamma$  and  $|\text{Re } \lambda| \leq \tilde{\varepsilon}$

$$\|G_N^{(1)}(\Lambda) - G^{(1)}(\Lambda)\| \leq \frac{c}{N^{1+\eta}} \lambda^{-\alpha}, \quad (3.8)$$

where  $\|\cdot\|$  stands as before for the operator norm, i.e. the largest singular value in  $M_k(\mathbb{C})$ .

**Step 3. Proof of Estimate (3.8):** This part is the main purpose of this paper. For this task we show that the Stieltjes transform for random matrices (3.6) approximatively satisfies an equation predicted by free probability theory for (3.7). In [15], [14] and [7] the authors have used a generalization, in the context of operator-valued non commutative variables, of the fact that the Stieltjes transform of a non commutative random variable is invertible and its inverse can be expressed in term of its  $\mathcal{R}$ -transform (see Section 4.2). In this paper, since we deal with two families of random matrices that are asymptotically free, the so-called subordination property (which encodes the fundamental property of  $\mathcal{R}$ -transforms, namely linearity under free convolution) may be applied.

We first come back to the notations in order to state the last assumption and stress the main ideas of the proof. Fix a self adjoint degree one non commutative polynomial  $L \in M_k(\mathbb{C}) \otimes \mathbb{C}\langle \mathbf{x}, \mathbf{y}, \mathbf{y}^* \rangle$ . We can write

$$L(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) = a_0 \otimes \mathbf{1}_N + S_N + T_N \in M_k(\mathbb{C}) \otimes M_N(\mathbb{C}),$$

$$L(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) = a_0 \otimes \mathbf{1} + s + t \in M_k(\mathbb{C}) \otimes \mathbb{C}\langle \mathbf{x}, \mathbf{y}, \mathbf{y}^* \rangle,$$

where

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

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

$a_0, \dots, a_p \in M_k(\mathbb{C})$  being selfadjoint,  $b_1, \dots, b_q \in M_k(\mathbb{C})$ .

Define the Stieltjes transforms of the linear component of  $L$  in the matrices of  $\mathbf{Y}_N$  and, respectively, in the non commutative variables  $\mathbf{y}$ : for all  $\Gamma$  in  $M_k(\mathbb{C})$  such that  $\text{Im } \Gamma > 0$ ,

$$G_N^{(2)}(\Gamma) = \mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ (\Gamma \otimes \mathbf{1}_N - T_N)^{-1} \right] \right] \in M_k(\mathbb{C}),$$

$$G^{(2)}(\Gamma) = (\text{id}_k \otimes \tau) \left[ (\Gamma \otimes \mathbf{1} - t)^{-1} \right] \in M_k(\mathbb{C}).$$

In the following we will denote, for  $\gamma, \tilde{\gamma} > 0$ , by  $\mathcal{O}_{\gamma, \tilde{\gamma}}$  the set

$$\mathcal{O}_{\gamma, \tilde{\gamma}} = \left\{ \Lambda \in M_k(\mathbb{C}) \mid \text{Im } \Lambda > 0, \|(\text{Im } \Lambda)^{-1}\| \leq N^\gamma, \|\Lambda\| \leq N^{\tilde{\gamma}} \right\}, \quad (3.9)$$

where  $\|\cdot\|$  stands as before for the operator norm. We can now describe the last assumption on  $\mathbf{Y}_N$ :



**Assumption 3** (Rate of convergence for  $M_k(\mathbb{C})$ -valued Stieltjes transforms). *We suppose that the following speed of convergence of  $G_N^{(2)}$  to  $G^{(2)}$  holds: there exist  $N_2, \gamma_2, \tilde{\gamma}_2, c_2, \eta_2, \alpha_2 > 0$  such that for all  $N \geq N_2$ , for all matrices  $\Gamma$  in  $\mathcal{O}_{\gamma_2, \tilde{\gamma}_2}$  we have*

$$\left\| G_N^{(2)}(\Gamma) - G^{(2)}(\Gamma) \right\| \leq \frac{c_2}{N^{1+\eta_2}} \|(\operatorname{Im} \Gamma)^{-1}\|^{\alpha_2}. \quad (3.10)$$

**Remark:** The order  $1/N^{1+\eta}$  is the optimal one we can expect to obtain the convergence of the extremal singular values. Modifying a single eigenvalue of a matrix produces a perturbation of order  $1/N$  in  $G_N^{(1)}$  which can modify the asymptotic behavior of the norm of the matrix.

We show that we can obtain a similar control on the speed of convergence of  $G_N^{(1)}$  to  $G^{(1)}$  and then deduce Estimate (3.8):

**Theorem 2** (Estimation of  $M_k(\mathbb{C})$ -valued Stieltjes transforms). *Under Assumptions 1, 2 and 3, there exist constants  $N_0, \gamma, \tilde{\gamma}$  such that for all  $N \geq N_0$ , for all  $\Lambda$  in  $\mathcal{O}_{\gamma, \tilde{\gamma}}$ , one has*

$$\|G^{(1)}(\Lambda) - G_N^{(1)}(\Lambda)\| \leq \left(1 + 2\|(\operatorname{Im} \Lambda)^{-1}\|^2 \sum_{j=1}^p \|a_j\|^2\right) \left(\frac{c_1}{N^2} \|(\operatorname{Im} \Lambda)^{-1}\|^5 + \frac{c_2}{N^{1+\eta_2}} \|(\operatorname{Im} \Lambda)^{-1}\|^{\alpha_2}\right), \quad (3.11)$$

with  $c_2, \eta_2, \alpha_2$  of Assumption 3 and  $c_1 = \frac{2k^{9/2}\sigma}{N^2} \sum_{j=1}^p \|a_j\| \left(\sum_{j=1}^p \|a_j\| + 2\sum_{j=1}^q \|b_j\|\right)^2$ .

Hence, for every  $\varepsilon, \tilde{\varepsilon} > 0$ , for all matrices  $\Lambda$  of the form  $\Lambda = \lambda \mathbf{1}_k$  with  $\operatorname{Im} \lambda > 0$  such that  $\varepsilon \leq (\operatorname{Im} \lambda)^{-1} \leq N^\gamma$  and  $|\operatorname{Re} \lambda| \leq \tilde{\varepsilon}$ , we have for  $N$  large enough

$$\|G^{(1)}(\Lambda) - G_N^{(1)}(\Lambda)\| \leq \frac{c}{N^{1+\eta}} (\operatorname{Im} \lambda)^{-\alpha} \quad (3.12)$$

where  $\eta = \min(\eta_2, 1/2)$ ,  $\alpha = \max(7, \alpha_2 + 2)$  and  $c$  a positive constant that does not depend on  $\lambda$  or  $N$ .

**Remark:** The central limit theorem (CLT) for certain linear eigenvalue statistics of random matrices was considered in many papers (see [20] and the references therein). Results are often stated for statistics re-centered by their mean: for a large class of random matrix  $M_N$  and functions  $f$ , it is shown that the random variable  $N(\tau_N[f(M_N)] - \mathbb{E}[\tau_N[f(M_N)]])$  converges in law to a Gaussian random variable. We obtain from Theorem 2 and classical manipulations of Stieltjes transforms that if the CLT holds for a  $kN \times kN$  matrix  $M_N = L(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)$  ( $L$  being in  $M_k(\mathbb{C}) \otimes \mathbb{C}\langle \mathbf{x}, \mathbf{y}, \mathbf{y}^* \rangle$  self adjoint of degree 1) and  $f: \mathbb{R} \rightarrow \mathbb{R}$  a smooth function, then we get the CLT re-centered by the limit:  $N(\tau_N[f(M_N)] - \tau[f(m)])$  converges in law to a Gaussian random variable, where  $m$  is the non commutative limit of  $M_N$ .

The key point of the proof of Estimate (3.11) is to show an asymptotic subordination property for random matrices: we show in Section 4.1 that  $\forall \Lambda \in M_k(\mathbb{C})$  with  $\operatorname{Im} \Lambda > 0$

$$G_N^{(1)}(\Lambda) = G_N^{(2)}\left(\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - \mathcal{R}(G_N^{(1)}(\Lambda))\right) + \Theta_N(\Lambda), \quad (3.13)$$

where:

- $\mathcal{R}$  is the  $M_k(\mathbb{C})$ -amalgamated  $\mathcal{R}$ -transform of  $s$  (see Section 4.2):

$$\begin{aligned} \mathcal{R}: M_k(\mathbb{C}) &\rightarrow M_k(\mathbb{C}) \\ G &\mapsto \sum_{j=1}^p a_j G a_j \end{aligned} \quad (3.14)$$

- $\Theta_N$  an error term well controlled by concentration inequalities.

The so-called subordination property for operator-valued non commutative random variables encodes the relation of freeness in term of Stieltjes and  $\mathcal{R}$ -transforms (see Section 4.2). In this case the freeness between  $\mathbf{x}$  and  $\mathbf{y}$  gives: for  $\operatorname{Im} \Lambda > 0$

$$G^{(1)}(\Lambda) = G^{(2)}\left(\Lambda - a_0 - \mathcal{R}(G^{(1)}(\Lambda))\right). \quad (3.15)$$

To relate precisely these two equations we use analytical changes of variables (purpose of Section 4.3.1): at the source for the exact subordination property and at the goal for the asymptotic one. More precisely we introduce perturbations  $\tilde{G}_N^{(1)}$  of the Stieltjes transform  $G_N^{(1)}$  and  $\Lambda_N$  of the parameter  $\Lambda$ , such that

$$\begin{cases} \tilde{G}_N^{(1)}(\Lambda) &= G^{(2)}\left(\Lambda_N - a_0 - \mathcal{R}(\tilde{G}_N^{(1)}(\Lambda))\right), \\ G^{(1)}(\Lambda_N) &= G^{(2)}\left(\Lambda_N - a_0 - \mathcal{R}(G^{(1)}(\Lambda_N))\right). \end{cases} \quad (3.16)$$

We show (Lemma 3) that the fixed point equation

$$G_\Lambda = G^{(2)}\left(\Lambda - a_0 - \mathcal{R}(G_\Lambda)\right) \quad (3.17)$$

admits a unique solution  $G_\Lambda$  in  $M_k(\mathbb{C})$  whose imaginary part is non positive semi definite, provided  $(\text{Im } \Lambda)^{-1}$  is small enough. By analyticity of the functions we will conclude that

$$\tilde{G}_N^{(1)}(\Lambda) = G^{(1)}(\Lambda_N), \quad \forall \text{Im } \Lambda > 0. \quad (3.18)$$

The convergence of  $G_N^{(1)}$  to  $G^{(1)}$  is deduced from some controls on  $\tilde{G}_N^{(1)}$  and  $\Lambda_N$ .

## 4 Details of the proof of Theorem 2

### 4.1 Asymptotic subordination property for random matrices

Let  $\mathbf{X}_N = (X_1^{(N)}, \dots, X_p^{(N)})$  be independent G.U.E. matrices, and  $\mathbf{Y}_N = (Y_1^{(N)}, \dots, Y_q^{(N)})$  random matrices. Consider a selfadjoint degree one polynomial  $L$  in  $M_k(\mathbb{C}) \otimes \mathbb{C}\langle \mathbf{x}, \mathbf{y}, \mathbf{y}^* \rangle$ , and write

$$L(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) = a_0 \otimes \mathbf{1}_N + S_N + T_N,$$

with

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

the matrices  $a_0, \dots, a_p$  being selfadjoint.

For  $\text{Im } \Lambda > 0$ ,  $\text{Im } \Gamma > 0$ , define the  $k \times k$  random matrices

$$\begin{aligned} H_N^{(1)}(\Lambda) &= (\text{id}_k \otimes \tau_N) \left[ (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N)^{-1} \right], \\ H_N^{(2)}(\Gamma) &= (\text{id}_k \otimes \tau_N) \left[ (\Gamma \otimes \mathbf{1}_N - T_N)^{-1} \right]. \end{aligned}$$

Since  $L(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)$  and  $T_N$  are selfadjoint and  $\text{Im } \Lambda > 0$ ,  $\text{Im } \Gamma > 0$ , these matrices are well defined (see [15, Lemma 3.1]). The  $M_k(\mathbb{C})$ -valued Stieltjes transforms of  $L(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)$  and of  $T_N$  are defined respectively by

$$\begin{aligned} G_N^{(1)}(\Lambda) &= \mathbb{E} \left[ H_N^{(1)}(\Lambda) \right] = \mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N)^{-1} \right] \right], \\ G_N^{(2)}(\Gamma) &= \mathbb{E} \left[ H_N^{(2)}(\Gamma) \right] = \mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ (\Gamma \otimes \mathbf{1}_N - T_N)^{-1} \right] \right]. \end{aligned}$$

The purpose of this section is to show the following theorem which shows that  $G_N^{(1)}$  approximately satisfies (3.17).

**Theorem 3** (Asymptotic subordination property). *Under Assumptions 1 and 2, for  $\text{Im } \Lambda > 0$  define*

$$\Gamma_N(\Lambda) = \Lambda - a_0 - \sum_{j=1}^p a_j G_N^{(1)}(\Lambda) a_j.$$

In the following, we denote for simplicity  $\Gamma_N = \Gamma_N(\Lambda)$ . Then one has  $\text{Im } \Gamma_N > 0$  and the Stieltjes transforms satisfy the equation

$$G_N^{(1)}(\Lambda) = G_N^{(2)}\left(\Lambda - a_0 - \sum_{j=1}^p a_j G_N^{(1)}(\Lambda) a_j\right) - \Theta_N(\Lambda, \Gamma_N), \quad (4.1)$$

where

$$\begin{aligned} \|\Theta_N(\Lambda, \Gamma)\| &\leq \frac{c_1}{N^2} \|(\text{Im } \Gamma)^{-1}\| \|(\text{Im } \Lambda)^{-1}\|^3 \left( \|(\text{Im } \Gamma)^{-1}\| + \|(\text{Im } \Lambda)^{-1}\| \right), \\ \|(\text{Im } \Gamma_N)^{-1}\| &\leq \|(\text{Im } \Lambda)^{-1}\|, \end{aligned}$$

with  $c_1 = \frac{k^{9/2}\sigma}{N^2} \sum_{j=1}^p \|a_j\| \left( \sum_{j=1}^p \|a_j\| + 2 \sum_{j=1}^q \|b_j\| \right)^2$ ,  $\|\cdot\|$  being the operator norm in  $M_k(\mathbb{C})$ .

We first obtain in Section 4.1.1 an equation relating  $H_N^{(1)}$  and  $H_N^{(2)}$  and some related quantities. In Section 4.1.2 we deduce by using concentration properties an equation relating  $G_N^{(1)}$  and  $G_N^{(2)}$ . Theorem 3 is then obtained in Section 4.1.3.

#### 4.1.1 Mean Schwinger-Dyson equation for random Stieltjes transforms

To condense our equation we define for all  $\Lambda, \Gamma$  in  $M_k(\mathbb{C})$  such that  $\text{Im } \Lambda > 0$ ,  $\text{Im } \Gamma > 0$  the elements of  $M_k(\mathbb{C}) \otimes M_N(\mathbb{C})$

$$\begin{aligned} h_N^{(1)}(\Lambda) &= (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N)^{-1}, \\ h_N^{(2)}(\Gamma) &= (\Gamma \otimes \mathbf{1}_N - T_N)^{-1}, \end{aligned}$$

and recall that  $H_N^{(1)}(\Lambda) = (\text{id}_k \otimes \tau_N) \left[ h_N^{(1)}(\Lambda) \right]$ ,  $H_N^{(2)}(\Lambda) = (\text{id}_k \otimes \tau_N) \left[ h_N^{(2)}(\Lambda) \right]$ .

**Proposition 1** (Mean Schwinger-Dyson equation for random Stieltjes transforms). *For all  $\Lambda, \Gamma$  in  $M_k(\mathbb{C})$  such that  $\text{Im } \Lambda > 0$ ,  $\text{Im } \Gamma > 0$ , we have*

$$\mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ h_N^{(2)}(\Gamma) \left( \sum_{j=1}^p a_j H_N^{(1)}(\Lambda) a_j - \Lambda + a_0 + \Gamma \right) \otimes \mathbf{1}_N h_N^{(1)}(\Lambda) \right] + H_N^{(2)}(\Gamma) - H_N^{(1)}(\Lambda) \right] = 0. \quad (4.2)$$

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  $\sigma^2$  and  $F : \mathbb{R}^N \rightarrow \mathbb{C}$  a differentiable map such that  $F$  and its partial derivatives are polynomially bounded one has for  $i = 1, \dots, N$

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

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. If  $P$  is a monomial in  $p$  non commutative indeterminates  $x_1, \dots, x_p$ , one has for  $i = 1, \dots, p$

$$\mathbb{E} \left[ \tau_N \left[ X_i^{(N)} P(X_1^{(N)}, \dots, X_p^{(N)}) \right] \right] = \sum_{P=Lx_iR} \mathbb{E} \left[ \tau_N \left[ L(X_1^{(N)}, \dots, X_p^{(N)}) \right] \tau_N \left[ R(X_1^{(N)}, \dots, X_p^{(N)}) \right] \right],$$

the sum over all decompositions  $P = Lx_iR$  being viewed as the partial derivative. Such a formula extends to the case of smooth maps instead of polynomials. The case of the function  $(\Lambda \otimes \mathbf{1}_N - S_N)^{-1}$  is investigated in details in [15, Formula (3.9)], our proof is obtained by minor modifications.

*Proof.* Recall we have defined  $S_N = \sum_{j=1}^p a_j \otimes X_j^{(N)}$ . Denote by  $(\epsilon_{m,n})_{m,n=1,\dots,N}$  the canonical basis of  $M_N(\mathbb{C})$ : using Gaussian integration by parts, summed in Formula (3.9) of [15], we get the following

$$\begin{aligned}
& \mathbb{E} \left[ (\Gamma \otimes \mathbf{1}_N - T_N)^{-1} S_N (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N)^{-1} \right] \\
&= \sum_{j=1}^p \mathbb{E} \left[ (\Gamma \otimes \mathbf{1}_N - T_N)^{-1} (a_j \otimes X_j^{(N)}) (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N)^{-1} \right] \\
&= \sum_{j=1}^p \mathbb{E} \left[ (\Gamma \otimes \mathbf{1}_N - T_N)^{-1} \frac{1}{N} \sum_{m,n=1}^N (a_j \otimes \epsilon_{m,n}) (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N)^{-1} \right. \\
&\quad \left. \times (a_j \otimes \epsilon_{n,m}) (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N)^{-1} \right] \\
&= \mathbb{E} \left[ (\Gamma \otimes \mathbf{1}_N - T_N)^{-1} \left( \sum_{j=1}^p a_j H_N^{(1)}(\Lambda) a_j \otimes \mathbf{1}_N \right) (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N)^{-1} \right] \\
&= \mathbb{E} \left[ h_N^{(2)}(\Gamma) \left( \sum_{j=1}^p a_j H_N^{(1)}(\Lambda) a_j \otimes \mathbf{1}_N \right) h_N^{(1)}(\Lambda) \right]. \tag{4.3}
\end{aligned}$$

We take the partial trace in Equation (4.3) to obtain:

$$\mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ h_N^{(2)}(\Gamma) \left( \sum_{j=1}^p a_j H_N^{(1)}(\Lambda) a_j \otimes \mathbf{1}_N \right) h_N^{(1)}(\Lambda) \right] \right] = \mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ h_N^{(2)}(\Gamma) S_N h_N^{(1)}(\Lambda) \right] \right]. \tag{4.4}$$

We now rewrite  $S_N$  as follow:

$$S_N = (\Lambda - a_0 - \Gamma) \otimes \mathbf{1}_N + (\Gamma \otimes \mathbf{1}_N - T_N) - (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N).$$

Re-injecting this expression in Equation (4.4), one gets Equation (4.2):

$$\begin{aligned}
& \mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ h_N^{(2)}(\Gamma) \left( \sum_{j=1}^p a_j H_N^{(1)}(\Lambda) a_j \otimes \mathbf{1}_N \right) h_N^{(1)}(\Lambda) \right] \right] \\
&= \mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ h_N^{(2)}(\Gamma) (\Lambda - a_0 - \Gamma) \otimes \mathbf{1}_N h_N^{(1)}(\Lambda) + h_N^{(1)}(\Lambda) - h_N^{(2)}(\Gamma) \right] \right] \\
&= \mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ h_N^{(2)}(\Gamma) \left( (\Lambda - a_0 - \Gamma) \otimes \mathbf{1}_N \right) h_N^{(1)}(\Lambda) \right] + H_N^{(1)}(\Lambda) - H_N^{(2)}(\Gamma) \right].
\end{aligned}$$

□

#### 4.1.2 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 (4.2) a relation between  $G_N^{(1)}$  and  $G_N^{(2)}$ . We define the centered version of  $H_N^{(1)}$ , for  $\text{Im } \Lambda > 0$ , by

$$K_N^{(1)}(\Lambda) = H_N^{(1)}(\Lambda) - G_N^{(1)}(\Lambda), \text{ in } M_k(\mathbb{C}). \tag{4.5}$$

Introduce the random linear map

$$\begin{aligned}
l_{N,\Lambda,\Gamma} : M_k(\mathbb{C}) \otimes M_N(\mathbb{C}) &\rightarrow M_k(\mathbb{C}) \otimes M_N(\mathbb{C}) \\
M &\mapsto h_N^{(2)}(\Gamma) M h_N^{(1)}(\Lambda)
\end{aligned} \tag{4.6}$$

and its mean

$$L_{N,\Lambda,\Gamma} : M \mapsto \mathbb{E} \left[ l_{N,\Lambda,\Gamma}(M) \right]. \tag{4.7}$$

**Proposition 2** (Schwinger-Dyson equation for mean Stieltjes transforms). *For all  $\Lambda, \Gamma$  in  $M_k(\mathbb{C})$  such that  $\text{Im } \Lambda > 0, \text{Im } \Gamma > 0,$*

$$(\text{id}_k \otimes \tau_N) \left[ L_{N,\Lambda,\Gamma} \left( \left( \sum_{j=1}^p a_j G_N^{(1)}(\Lambda) a_j - \Lambda + a_0 + \Gamma \right) \otimes \mathbf{1}_N \right) \right] + G_N^{(2)}(\Gamma) - G_N^{(1)}(\Lambda) = \Theta_N(\Lambda, \Gamma), \quad (4.8)$$

where

$$\Theta_N(\Lambda, \Gamma) = -\mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ (l_{N,\Lambda,\Gamma} - L_{N,\Lambda,\Gamma}) \left( \sum_{j=1}^p a_j K_N^{(1)}(\Lambda) a_j \otimes \mathbf{1}_N \right) \right] \right] \quad (4.9)$$

can be controlled in operator norm by the following estimate:

$$\|\Theta_N(\Lambda, \Gamma)\| \leq \frac{c_1}{N^2} \|(\text{Im } \Gamma)^{-1}\| \|(\text{Im } \Lambda)^{-1}\|^3 \left( \|(\text{Im } \Gamma)^{-1}\| + \|(\text{Im } \Lambda)^{-1}\| \right), \quad (4.10)$$

with  $c_1 = \frac{k^{9/2}\sigma}{N^2} \sum_{j=1}^p \|a_j\| \left( \sum_{j=1}^p \|a_j\| + 2 \sum_{j=1}^q \|b_j\| \right)^2$

We show the estimate of  $\Theta_N$  in Appendix A. Equation (4.8) follows immediately from the previous proposition.

*Proof.* We first develop  $\Theta_N(\Lambda, \Gamma)$ :

$$\begin{aligned} -\Theta_N(\Lambda, \Gamma) &:= \mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ (l_{N,\Lambda,\Gamma} - L_{N,\Lambda,\Gamma}) \left( \sum_{j=1}^p a_j (H_N^{(1)}(\Lambda) - G_N^{(1)}(\Lambda)) a_j \otimes \mathbf{1}_N \right) \right] \right] \\ &= \mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ l_{N,\Lambda,\Gamma} \left( \sum_{j=1}^p a_j H_N^{(1)}(\Lambda) a_j \otimes \mathbf{1}_N \right) \right] \right] \\ &\quad - \mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ L_{N,\Lambda,\Gamma} \left( \sum_{j=1}^p a_j G_N^{(1)}(\Lambda) a_j \otimes \mathbf{1}_N \right) \right] \right]. \end{aligned}$$

By Equation (4.2) we get the following:

$$\begin{aligned} &\mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ l_{N,\Lambda,\Gamma} \left( \sum_{j=1}^p a_j H_N^{(1)}(\Lambda) a_j \otimes \mathbf{1}_N \right) \right] \right] \\ &= \mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ l_{N,\Lambda,\Gamma} \left( (\Lambda - a_0 - \Gamma) \otimes \mathbf{1}_N \right) \right] - H_N^{(2)}(\Gamma) + H_N^{(1)}(\Lambda) \right] \\ &= (\text{id}_k \otimes \tau_N) \left[ L_{N,\Lambda,\Gamma} \left( (\Lambda - a_0 - \Gamma) \otimes \mathbf{1}_N \right) \right] - G_N^{(2)}(\Gamma) + G_N^{(1)}(\Lambda), \end{aligned}$$

which gives Equation (4.8).  $\square$

### 4.1.3 Proof of Theorem 3: the asymptotic subordination property

Since the map  $L_{N,\Lambda,\Gamma}$  is linear, if in the Schwinger-Dyson Equation (4.8) we formally take  $\Gamma$  equals to

$$\Gamma_N = \Lambda - a_0 - \sum_{j=1}^p a_j G_N^{(1)}(\Lambda) a_j, \quad (4.11)$$

then we obtain the expected equation

$$G_N^{(1)}(\Lambda) = G_N^{(2)}(\Gamma_N) - \Theta_N(\Lambda, \Gamma_N). \quad (4.12)$$

We show that  $\text{Im } \Gamma_N > 0,$  so that Theorem 3 follows.

**Lemma 1** (Control of  $\Gamma_N$ ). *There exists  $K > 0$  such that for  $\text{Im } \Lambda > 0$*

$$\text{Im } \Gamma_N \geq \text{Im } \Lambda + \|(\text{Im } \Lambda)^{-1}\|^{-1} (\|\Lambda\| + K)^{-2} \sum_{j=1}^p a_j^2. \quad (4.13)$$

*In particular  $\text{Im } \Gamma_N > 0$  as soon as  $\text{Im } \Lambda > 0$  and*

$$\|(\text{Im } \Gamma_N)^{-1}\| \leq \|(\text{Im } \Lambda)^{-1}\|. \quad (4.14)$$

*Proof.* The  $a_j$ 's being selfadjoint, we have

$$\text{Im } \Gamma_N = \text{Im} \left( \Lambda - a_0 - \sum_{j=1}^p a_j G_N^{(1)}(\Lambda) a_j \right) = \text{Im } \Lambda - \sum_{j=1}^p a_j \text{Im} \left( G_N^{(1)}(\Lambda) \right) a_j. \quad (4.15)$$

Recall the definition of  $G_N^{(1)}$ :

$$G_N^{(1)}(\Lambda) = \mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N)^{-1} \right] \right].$$

The imaginary part of the  $kN \times kN$  matrix in the partial trace is non positive definite:

$$\begin{aligned} & \text{Im} \left[ (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N)^{-1} \right] \\ &= \frac{1}{2i} \left[ (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N)^{-1} - (\Lambda^* \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N)^{-1} \right] \\ &= -(\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N)^{-1} (\text{Im } \Lambda \otimes \mathbf{1}_N) (\Lambda^* \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N)^{-1}. \end{aligned}$$

For a non negative definite matrix  $A$ , the number  $\|A^{-1}\|^{-1}$  is the smallest eigenvalue of  $A$ , hence

$$\begin{aligned} & -\text{Im} \left[ (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N)^{-1} \right] \\ & \geq \left\| \left( -\text{Im} \left[ (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N)^{-1} \right] \right)^{-1} \right\|^{-1} \mathbf{1}_k \otimes \mathbf{1}_N \\ & \geq \|\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N\|^{-2} \|(\text{Im } \Lambda)^{-1}\|^{-1} \mathbf{1}_k \otimes \mathbf{1}_N, \end{aligned}$$

where we have used the fact that for any matrices  $A$  such that  $\text{Im } A > 0$ ,  $\|A^{-1}\| \leq \|(\text{Im } A)^{-1}\|$ . By positivity of  $(\text{id}_k \otimes \tau_N)$  we get

$$\begin{aligned} -\text{Im } H_N^{(1)}(\Lambda) & \geq \|(\Lambda - a_0) \otimes \mathbf{1}_N - S_N - T_N\|^{-2} \|(\text{Im } \Lambda)^{-1}\|^{-1} \mathbf{1}_k \\ & \geq (\|\Lambda\| + \|a_0\| + \|S_N\| + \|T_N\|)^{-2} \|(\text{Im } \Lambda)^{-1}\|^{-1} \mathbf{1}_k. \end{aligned}$$

For a  $N \times N$  normalized G.U.E. matrix  $X_N$  one has always  $\mathbb{E}[\|X_N\|] \leq 4$  (see [15, Lemma 5.1]). By Assumption 1, for  $j = 1, \dots, q$  one has  $\limsup \|Y_j^{(N)}\| < D$  almost surely. Hence there exists a  $\tilde{D}$  such that  $\mathbb{E}[\|T_N\|] \leq \tilde{D}$  almost surely. Using Jensen's inequality we then obtain the following

$$\begin{aligned} -\text{Im } G_N^{(1)}(\Lambda) & \geq \left( \|\Lambda\| + \|a_0\| + \mathbb{E}[\|S_N\|] + \mathbb{E}[\|T_N\|] \right)^{-2} \|(\text{Im } \Lambda)^{-1}\|^{-1} \mathbf{1}_k \\ & \geq \left( \|\Lambda\| + \|a_0\| + 4 \sum_{j=1}^p \|a_j\| + \tilde{D} \right)^{-2} \|(\text{Im } \Lambda)^{-1}\|^{-1} \mathbf{1}_k. \end{aligned}$$

As a result we obtain (4.13) from (4.15) with  $K = \|a_0\| + 4 \sum_{j=1}^p \|a_j\| + \tilde{D}$ .  $\square$

## 4.2 The subordination property in free probability

In this section, Equation (4.1) is illuminated in the context of free probability. The reader is referred to Speicher's note in [24] for detailed definitions and proofs, especially Section 22.7 concerning operator-valued free probability. For the origins of the subordination property, see the works of Voiculescu in [27] and Biane in [6].

Let  $(\mathcal{B}, *, \tau, \|\cdot\|)$  be a  $C^*$ -probability space. For any non commutative variable  $a$ , one can define a map  $\mathcal{R}_a : \mathbb{C} \rightarrow \mathbb{C}$  which is a free analogue of the log Fourier transform in the sense that it characterizes the non commutative law of  $a$  and is linearized under free convolution: if  $a, b$  are free non commutative variables, then the  $\mathcal{R}$ -transform  $\mathcal{R}_{a+b}$  of the sum  $a + b$  is given by

$$\mathcal{R}_{a+b} = \mathcal{R}_a + \mathcal{R}_b. \quad (4.16)$$

Define for all  $\lambda$  in  $\mathbb{C}$  such that  $\text{Im } \lambda > 0$  the Stieltjes transform of a selfadjoint element  $a$  by

$$G_a(\lambda) = \tau [ (\lambda - a)^{-1} ]. \quad (4.17)$$

Then the following relations between  $\mathcal{R}_a$  and  $G_a$  hold: for  $\text{Im } \lambda > 0$

$$\frac{1}{G_a(\lambda)} + \mathcal{R}_a(G_a(\lambda)) = \lambda, \quad (4.18)$$

$$G_a\left(\mathcal{R}_a(\lambda) + \frac{1}{\lambda}\right) = \lambda. \quad (4.19)$$

With equations (4.16), (4.18) and (4.19) admitted it is easy to derive the subordination property for scalar-valued non commutative variables.

**Proposition 3.** *Let  $a, b$  in  $\mathcal{B}$  be free non commutative variables. Denote by  $G_a$  and  $G_{a+b}$ , respectively, the Stieltjes transforms of  $a$  and  $a + b$ ; denote by  $\mathcal{R}_b$  and  $\mathcal{R}_{a+b}$ , respectively, the  $\mathcal{R}$ -transforms of  $b$  and  $a + b$ . Then for all  $\lambda$  in  $\mathbb{C}$  such that  $\text{Im } \lambda > 0$*

$$G_{a+b}(\lambda) = G_a\left(\lambda - \mathcal{R}_b(G_{a+b}(\lambda))\right). \quad (4.20)$$

*Proof.* for  $\text{Im } \lambda > 0$  by (4.18)

$$\frac{1}{G_{a+b}(\lambda)} + \mathcal{R}_{a+b}(G_{a+b}(\lambda)) = \lambda.$$

By freeness of  $a$  and  $b$  we can use Equation (4.16):

$$\lambda - \mathcal{R}_b(G_{a+b}(\lambda)) = \frac{1}{G_{a+b}(\lambda)} + \mathcal{R}_a(G_{a+b}(\lambda)).$$

Compose by  $G_a$  and use (4.19)

$$G_a\left(\lambda - \mathcal{R}_b(G_{a+b}(\lambda))\right) = G_a\left(\mathcal{R}_a(G_{a+b}(\lambda)) + \frac{1}{G_{a+b}(\lambda)}\right) = G_{a+b}(\lambda).$$

□

Now let  $(\mathbf{x}, \mathbf{y}) = (x_1, \dots, x_p, y_1, \dots, y_q)$  be a family of non commutative variables in a  $C^*$ -probability space such that,

1.  $\mathbf{x} = (x_1, \dots, x_p)$  is a semicircular system (see Equation (2.9))
2.  $\mathbf{y}$  is the non commutative limit of  $\mathbf{Y}_N$
3. The families  $\mathbf{x}$  and  $\mathbf{y}$  are free (see Property (2.8)).

We introduce the matrix-valued Stieltjes transforms:

$$\begin{aligned} \forall \Lambda, \Gamma \in M_k(\mathbb{C}) \text{ with } \operatorname{Im} \Lambda > 0, \operatorname{Im} \Gamma > 0 \\ G^{(1)}(\Lambda) &= (id_k \otimes \tau) \left[ \left( \Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - \sum_{j=1}^p a_j \otimes x_j - \sum_{j=1}^q b_j \otimes y_j + b_j^* \otimes y_j^* \right)^{-1} \right], \\ G^{(2)}(\Gamma) &= (id_k \otimes \tau) \left[ \left( \Gamma \otimes \mathbf{1}_N - \sum_{j=1}^q b_j \otimes y_j + b_j^* \otimes y_j^* \right)^{-1} \right]. \end{aligned}$$

with the matrix coefficients  $(a_j)_{0 \leq j \leq p}$  and  $(b_j)_{1 \leq j \leq q}$  as in the previous Section. Let  $\mathcal{R}$  be the map  $M_k(\mathbb{C}) \rightarrow M_k(\mathbb{C})$  defined by

$$\forall M \in M_k(\mathbb{C}), \mathcal{R}(M) = \sum_{j=1}^p a_j M a_j$$

It has been shown by Lehner [19] that  $\mathcal{R}$  is the so-called  $M_k(\mathbb{C})$ -amalgamated  $\mathcal{R}$ -transform of  $s = \sum_{j=1}^p a_j \otimes x_j$ , generalization of the  $\mathcal{R}$ -transform for  $M_k(\mathbb{C})$ -valued non commutative variables. The linearization Property (4.16) of  $\mathcal{R}$ -transforms for the free convolution of non commutative random variables extends to operator-valued non commutative variables (see [24]) as Relations (4.18) and (4.18) between  $M_k(\mathbb{C})$ -amalgamated  $\mathcal{R}$ -transforms and  $M_k(\mathbb{C})$ -valued Stieltjes transforms. In this case one gets:

**Proposition 4** (The subordination property). *The Stieltjes transforms satisfy*

$$\begin{aligned} \forall \Lambda \in M_k(\mathbb{C}) \text{ such that } \operatorname{Im} \Lambda > 0, \\ G^{(1)}(\Lambda) &= G^{(2)}\left(\Lambda - a_0 - \mathcal{R}(G^{(1)}(\Lambda))\right) \end{aligned} \quad (4.21)$$

### 4.3 Manipulations of the Stieltjes transforms

#### 4.3.1 Strategy for the end of the proof

We now suppose that  $\mathbf{Y}_N$  satisfies assumptions 1, 2 and 3. We have shown in Section 4.1 that if we define for  $\operatorname{Im} \Lambda > 0$

$$\Gamma_N = \Lambda - a_0 - \mathcal{R}(G_N^{(1)}(\Lambda)), \quad (4.22)$$

then we have the (almost) closed relation between the Stieltjes transforms:

$$G_N^{(1)}(\Lambda) = G_N^{(2)}\left(\Lambda - a_0 - \mathcal{R}(G_N^{(1)}(\Lambda))\right) - \Theta_N(\Lambda, \Gamma_N). \quad (4.23)$$

We introduce a perturbation of  $G_N^{(1)}$  to compare the subordination property for random matrices (4.23) with the one for the non commutative random variables (4.21). The asymptotic subordination property for random matrices (4.23) suggests to define  $\forall \Lambda \in M_k(\mathbb{C})$  with  $\operatorname{Im} \Lambda > 0$ ,

$$\tilde{G}_N^{(1)}(\Lambda) = G^{(2)}(\Gamma_N) = G_N^{(1)}(\Lambda) + \left(G^{(2)}(\Gamma_N) - G_N^{(2)}(\Gamma_N)\right) - \Theta_N(\Lambda, \Gamma_N). \quad (4.24)$$

Note that Estimate (3.9) in Assumption 3 and the control (4.10) of the norm of  $\Theta_N$  allow us to conclude that  $\tilde{G}_N^{(1)}(\Lambda)$  is a perturbation of  $G_N^{(1)}$  provided we can say that  $\Gamma_N$  is in the domain where the control of  $\|G_N^{(2)}(\Gamma) - G^{(2)}(\Gamma)\|$  holds.

Recall that we have define for all  $\gamma, \tilde{\gamma} > 0$  the set

$$\mathcal{O}_{\gamma, \tilde{\gamma}} = \left\{ \Lambda \in M_k(\mathbb{C}) \mid \operatorname{Im} \Lambda > 0, \|(\operatorname{Im} \Lambda)^{-1}\| \leq N^\gamma, \|\Lambda\| \leq N^{\tilde{\gamma}} \right\}. \quad (4.25)$$

We will show in the following:

**Lemma 2** (Domain of  $\Gamma_N$ ). *With  $\gamma_2, \tilde{\gamma}_2$  the constants of Assumption 3, there exist  $\gamma, \tilde{\gamma} > 0$ ,  $N_0 \geq 1$ , such that for all  $N \geq N_0$*

$$\Lambda \in \mathcal{O}_{\gamma, \tilde{\gamma}} \Rightarrow \Gamma_N \in \mathcal{O}_{\gamma_2, \tilde{\gamma}_2}. \quad (4.26)$$



With Lemma 2 taken for granted momentarily, we can write the precise estimate for  $\|\tilde{G}_N^{(1)}(\Lambda) - G_N^{(1)}(\Lambda)\|$ . Use the triangle inequality in Equation (4.24): for  $\text{Im } \Lambda > 0$

$$\|\tilde{G}_N^{(1)}(\Lambda) - G_N^{(1)}(\Lambda)\| \leq \|G^{(2)}(\Gamma_N) - G_N^{(2)}(\Gamma_N)\| + \|\Theta_N(\Lambda, \Gamma_N)\|.$$

With the constants of Lemma 2, take  $N \geq N_0$  and  $\Lambda$  in  $\mathcal{O}_{\gamma, \tilde{\gamma}}$ . Then  $\Gamma_N$  is in  $\mathcal{O}_{\gamma_2, \tilde{\gamma}_2}$ , the domain where Assumption 3 is satisfied and hence we have (3.9):

$$\|G^{(2)}(\Gamma_N) - G_N^{(2)}(\Gamma_N)\| \leq \frac{c_2}{N^{1+\eta_2}} \|(\text{Im } \Gamma_N)^{-1}\|^{\alpha_2}.$$

Recall that in Lemma 1 we have shown that  $\|(\text{Im } \Gamma_N)^{-1}\| \leq \|(\text{Im } \Lambda)^{-1}\|$ , so that

$$\|G^{(2)}(\Gamma_N) - G_N^{(2)}(\Gamma_N)\| \leq \frac{c_2}{N^{1+\eta_2}} \|(\text{Im } \Lambda)^{-1}\|^{\alpha_2}. \quad (4.27)$$

On the other hand, by the Estimate (4.10) for the norm of  $\Theta_N$ , we get: for  $\text{Im } \Lambda > 0$

$$\begin{aligned} \|\Theta_N(\Lambda, \Gamma_N)\| &\leq \frac{c_1}{N^2} \|(\text{Im } \Lambda)^{-1}\|^3 \|(\text{Im } \Gamma_N)^{-1}\| \left( \|(\text{Im } \Lambda)^{-1}\| + \|(\text{Im } \Gamma_N)^{-1}\| \right), \\ &\leq \frac{2c_1}{N^2} \|(\text{Im } \Lambda)^{-1}\|^5, \end{aligned}$$

where we have used again Lemma 1. Hence we have the following control of  $\tilde{G}_N^{(1)}$ : for  $N \geq N_0$ , for  $\Lambda$  in  $\mathcal{O}_{\gamma, \tilde{\gamma}}$

$$\|\tilde{G}_N^{(1)}(\Lambda) - G_N^{(1)}(\Lambda)\| \leq \frac{c_2}{N^{1+\eta_2}} \|(\text{Im } \Lambda)^{-1}\|^{\alpha_2} + \frac{2c_1}{N^2} \|(\text{Im } \Lambda)^{-1}\|^5. \quad (4.28)$$

By the definition of  $\Gamma_N$ ,  $\tilde{G}_N^{(1)}$  can be written

$$\tilde{G}_N^{(1)}(\Lambda) = G^{(2)}\left(\Lambda - a_0 - \mathcal{R}(G_N^{(1)}(\Lambda))\right). \quad (4.29)$$

In order to improve our comparison of the equation of subordination (4.21) for non commutative random variable and the latter one, we introduce a change of matrix variable  $\Lambda_N = \Lambda_N(\Lambda)$ :

$$\begin{aligned} \forall \Lambda \in M_k(\mathbb{C}) \text{ with } \text{Im } \Lambda > 0, \\ \Lambda_N = \Gamma_N + a_0 + \mathcal{R}(\tilde{G}_N^{(1)}(\Lambda)). \end{aligned} \quad (4.30)$$

By (4.22) and (4.30) we get

$$\Lambda - a_0 - \mathcal{R}(G_N^{(1)}(\Lambda)) = \Lambda_N - a_0 - \mathcal{R}(\tilde{G}_N^{(1)}(\Lambda)) \quad (4.31)$$

Hence, by using (4.31) in Equation (4.29) and taking  $\Lambda_N$  instead of  $\Lambda$  in (4.21), we see that  $G^{(1)}(\Lambda_N)$  and  $\tilde{G}_N^{(1)}(\Lambda)$  appear to be two solutions of the same fixed point equation:

$$\begin{cases} \tilde{G}_N^{(1)}(\Lambda) &= G^{(2)}\left(\Lambda_N - a_0 - \mathcal{R}(\tilde{G}_N^{(1)}(\Lambda))\right), \\ G^{(1)}(\Lambda_N) &= G^{(2)}\left(\Lambda_N - a_0 - \mathcal{R}(G^{(1)}(\Lambda_N))\right). \end{cases} \quad (4.32)$$

**Lemma 3** (Uniqueness of the fixed point). *For all  $\Lambda \in M_k(\mathbb{C})$  with  $\text{Im } \Lambda > 0$  and such that*

$$\|(\text{Im } \Lambda)^{-1}\| < \sqrt{\sum_{j=1}^p \|a_j\|^2},$$

the equation

$$M_\Lambda = G^{(2)}\left(\Lambda - a_0 - \mathcal{R}(M_\Lambda)\right) \quad (4.33)$$

admits a unique solution  $M_\Lambda$  in  $M_k(\mathbb{C})$  whose imaginary part is non positive semi definite.

To use this lemma it will be necessary to have a lower bound for the smallest eigenvalue of  $\text{Im } \Lambda_N$ . We will show the following:

**Lemma 4** (Control of  $\Lambda_N$ ). *There exist  $\gamma, \tilde{\gamma} > 0$ ,  $N_0 \geq 1$ , such that for all  $N \geq N_0$ , for all  $\Lambda$  in  $\mathcal{O}_{\gamma, \tilde{\gamma}}$  one has*

$$\|(\text{Im } \Lambda_N)^{-1}\| \leq 2\|(\text{Im } \Lambda)^{-1}\|. \quad (4.34)$$

Admitting these lemmas, we obtain the following:

**Proposition 5.** *There exists a  $N_0$  such that for  $N \geq N_0$ , for all  $\Lambda \in M_k(\mathbb{C})$  with  $\text{Im } \Lambda > 0$*

$$\tilde{G}_N^{(1)}(\Lambda) = G^{(1)}(\Lambda_N). \quad (4.35)$$

*Proof of Proposition 5.* According to Lemma 4, there exist  $\gamma, \tilde{\gamma} > 0$ ,  $N_0 \geq 1$  such that for every  $N \geq N_0$ , for every  $\Lambda$  in  $\mathcal{O}_{\gamma, \tilde{\gamma}}$  satisfying

$$\|(\text{Im } \Lambda)^{-1}\| < \frac{1}{2} \sqrt{\sum_{j=1}^p \|a_j\|^2}, \quad (4.36)$$

we have

$$\|(\text{Im } \Lambda_N)^{-1}\| \leq 2\|(\text{Im } \Lambda)^{-1}\| < \sqrt{\sum_{j=1}^p \|a_j\|^2}.$$

Hence  $\Lambda_N$  satisfies the assumption of Lemma 3 and by uniqueness of the fixed point we obtain the equality

$$G^{(1)}(\Lambda_N) = \tilde{G}_N^{(1)}(\Lambda) \quad (4.37)$$

on the connected set (with non empty interior) of  $k \times k$  matrices  $\Lambda$  in  $\mathcal{O}_{\gamma, \tilde{\gamma}}$  satisfying (4.36). The matrix-valued Stieltjes transforms are analytical function form the set of  $k \times k$  matrices with complex entries and positive definite imaginary part to  $M_k(\mathbb{C})$ . Hence Equality (4.37) extends for all  $\Lambda$  such that  $\text{Im } \Lambda > 0$ .  $\square$

Assuming Lemmas 2-4, we can now obtain Estimate (3.11) for  $\|G^{(1)}(\Lambda) - G_N^{(1)}(\Lambda)\|$ . By Proposition 5 there exists  $N_0 \geq 1$  such that for all  $N \geq N_0$ , for all  $\text{Im } \Lambda > 0$

$$\begin{aligned} \|G^{(1)}(\Lambda) - G_N^{(1)}(\Lambda)\| &= \|G_N^{(1)}(\Lambda) - \tilde{G}_N^{(1)}(\Lambda) + G^{(1)}(\Lambda_N) - G^{(1)}(\Lambda)\| \\ &\leq \|G_N^{(1)}(\Lambda) - \tilde{G}_N^{(1)}(\Lambda)\| + \|G^{(1)}(\Lambda_N) - G^{(1)}(\Lambda)\|. \end{aligned}$$

The first term was already estimated in Inequality (4.28): there exists a constant we again denote  $N_0$ , there exist  $\gamma, \tilde{\gamma}$ , such that for all  $N \geq N_0$ , for  $\Lambda$  in  $\mathcal{O}_{\gamma, \tilde{\gamma}}$  we have

$$\|G_N^{(1)}(\Lambda) - \tilde{G}_N^{(1)}(\Lambda)\| \leq \frac{c_2}{N^{1+\eta_2}} \|(\text{Im } \Lambda)^{-1}\|^{\alpha_2} + \frac{2c_1}{N^2} \|(\text{Im } \Lambda)^{-1}\|^5 \quad (4.38)$$

For the second term we have the following

$$\begin{aligned} &\|G^{(1)}(\Lambda) - G^{(1)}(\Lambda_N)\| \\ &= \left\| \mathbb{E} \left[ (\text{id}_k \otimes \tau) \left[ (\Lambda \otimes \mathbf{1} - a_0 \otimes \mathbf{1} - s - t)^{-1} - (\Lambda_N \otimes \mathbf{1} - a_0 \otimes \mathbf{1} - s - t)^{-1} \right] \right] \right\| \\ &= \left\| \mathbb{E} \left[ (\text{id}_k \otimes \tau) \left[ (\Lambda \otimes \mathbf{1} - a_0 \otimes \mathbf{1} - s - t)^{-1} (\Lambda_N \otimes \mathbf{1} - \Lambda \otimes \mathbf{1}) (\Lambda_N \otimes \mathbf{1} - a_0 \otimes \mathbf{1} - s - t)^{-1} \right] \right] \right\| \\ &\leq \|(\Lambda \otimes \mathbf{1} - a_0 \otimes \mathbf{1} - s - t)^{-1}\| \|\Lambda_N - \Lambda\| \|(\Lambda_N \otimes \mathbf{1} - a_0 \otimes \mathbf{1} - s - t)^{-1}\| \\ &\leq \|\Lambda_N - \Lambda\| \|(\text{Im } \Lambda)^{-1}\| \|(\text{Im } \Lambda_N)^{-1}\|. \end{aligned}$$

By Lemma 4, there exist constants we again denote  $N_0, \gamma, \tilde{\gamma}$ , such that, for  $N \geq N_0$  and  $\Lambda$  in  $\mathcal{O}_{\gamma, \tilde{\gamma}}$  we have additionally

$$\|(\text{Im } \Lambda_N)^{-1}\| \leq 2\|(\text{Im } \Lambda)^{-1}\|,$$

and hence

$$\|G^{(1)}(\Lambda) - G^{(1)}(\Lambda_N)\| \leq 2\|\Lambda_N - \Lambda\| \|(\operatorname{Im} \Lambda)^{-1}\|^2. \quad (4.39)$$

The term  $\|\Lambda_N - \Lambda\|$  is easy to estimate since by definition  $\Lambda_N = \Gamma_N + a_0 + \mathcal{R}(\tilde{G}_N^{(1)}(\Lambda))$  and  $\Gamma_N = \Lambda - a_0 - \mathcal{R}(G_N^{(1)}(\Lambda))$ :

$$\Lambda_N = \Lambda + \sum_{j=1}^p a_j (\tilde{G}_N^{(1)}(\Lambda) - G_N^{(1)}(\Lambda)) a_j$$

so that (4.39) gives

$$\|G^{(1)}(\Lambda) - G^{(1)}(\Lambda_N)\| \leq 2 \sum_{j=1}^p \|a_j\|^2 \|(\operatorname{Im} \Lambda)^{-1}\|^2 \|\tilde{G}_N^{(1)}(\Lambda) - G_N^{(1)}(\Lambda)\|. \quad (4.40)$$

Combining (4.38) and (4.40) gives the main estimate of this paper: for  $N \geq N_0$ , for  $\Lambda$  in  $\mathcal{O}_{\gamma, \tilde{\gamma}}$

$$\|G^{(1)}(\Lambda) - G^{(1)}(\Lambda_N)\| \leq \left(1 + 2 \sum_{j=1}^p \|a_j\|^2 \|(\operatorname{Im} \Lambda)^{-1}\|^2\right) \left(\frac{c_2}{N^{1+\eta_2}} \|(\operatorname{Im} \Lambda)^{-1}\|^{\alpha_2} + \frac{2c_1}{N^2} \|(\operatorname{Im} \Lambda)^{-1}\|^5\right).$$

In a nutshell, it remains to show Lemma 2-4.

#### 4.3.2 Proof of Lemma 2: control of $\Gamma_N$

Let  $\gamma_2, \tilde{\gamma}_2$  be the constants of Assumption 3. We want to find constants  $\gamma, \tilde{\gamma} > 0$ ,  $N_0 \geq 1$ , such that for all  $N \geq N_0$ :  $\Lambda \in \mathcal{O}_{\gamma, \tilde{\gamma}} \Rightarrow \Gamma_N \in \mathcal{O}_{\gamma_2, \tilde{\gamma}_2}$ . By Lemma 1, we always have  $\|(\operatorname{Im} \Gamma_N)^{-1}\| \leq \|(\operatorname{Im} \Lambda)^{-1}\|$ . Hence we focus on the control of  $\|\Gamma_N\|$ . For  $\operatorname{Im} \Lambda > 0$ ,

$$\|\Gamma_N\| = \left\| \Lambda - a_0 - \sum_{j=1}^p a_j G_N^{(1)}(\Lambda) a_j \right\| \leq \|\Lambda\| + \|a_0\| + \sum_{j=1}^p \|a_j\|^2 \|G_N^{(1)}(\Lambda)\|.$$

Remark that for  $A$  in  $M_k(\mathbb{C}) \otimes M_N(\mathbb{C})$ ,  $\|(\operatorname{id}_k \otimes \tau_N)A\| \leq \|A\|$ , so that

$$\begin{aligned} \|G_N^{(1)}(\Lambda)\| &= \left\| (\operatorname{id}_k \otimes \tau_N) \left[ (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N)^{-1} \right] \right\| \\ &\leq \left\| (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - S_N - T_N)^{-1} \right\| \\ &\leq \|(\operatorname{Im} \Lambda)^{-1}\| \end{aligned}$$

Hence, with  $\gamma < \min(\gamma_2, \tilde{\gamma}_2)$ ,  $\tilde{\gamma} < \min(\tilde{\gamma}_2)$ , there exists a  $N_0$  such that for all  $N \geq N_0$  one has  $\|\Gamma_N\| \leq N^{\gamma_2}$ .

#### 4.3.3 Proof of Lemma 3: uniqueness of the fixed point

Fix  $\Lambda \in M_k(\mathbb{C})$  with  $\operatorname{Im} \Lambda > 0$  and

$$\|(\operatorname{Im} \Lambda)^{-1}\| < \sqrt{\sum_{j=1}^p \|a_j\|^2}. \quad (4.41)$$

Denote in the following  $\psi(M) = \Lambda - a_0 - \sum_{j=1}^p a_j M a_j$ . We show that the function

$$\Phi_\Lambda : M \rightarrow G^{(2)}(\psi(M))$$

is a contraction on the set of matrices of  $M_k(\mathbb{C})$  whose imaginary part is non positive semi definite. For all  $M, \tilde{M}$  in  $M_k(\mathbb{C})$  such matrices, it follows from the proof of Lemma 1 that  $\Phi_\Lambda(M)$  and  $\Phi_\Lambda(\tilde{M})$  are well

defined and that  $\Phi_\Lambda$  maps the subset of matrices whose imaginary part is non positive semi definite into itself. Moreover

$$\begin{aligned}
& \|\Phi_\Lambda(M) - \Phi_\Lambda(\tilde{M})\| \\
&= \left\| (id_k \otimes \tau) \left[ \left( \psi(M) \otimes \mathbf{1} - t \right)^{-1} - \left( \psi(\tilde{M}) \otimes \mathbf{1} - t \right)^{-1} \right] \right\| \\
&= \left\| (id_k \otimes \tau) \left[ \left( \psi(M) \otimes \mathbf{1} - t \right)^{-1} \left( \sum_{j=1}^p a_j (M - \tilde{M}) a_j \right) \otimes \mathbf{1}_N \left( \psi(\tilde{M}) \otimes \mathbf{1} - t \right)^{-1} \right] \right\| \\
&\leq \left\| \left( \text{Im} (\psi(M) \otimes \mathbf{1} - t) \right)^{-1} \right\| \left\| \left( \text{Im} (\psi(\tilde{M}) \otimes \mathbf{1} - t) \right)^{-1} \right\| \sum_{j=1}^p \|a_j\|^2 \|M - \tilde{M}\| \\
&\leq \left\| (\text{Im } \Lambda)^{-1} \right\|^2 \sum_{j=1}^p \|a_j\|^2 \|M - \tilde{M}\|
\end{aligned}$$

Hence the function  $\Phi_\Lambda$  is a contraction when (4.41) holds and by Picard's theorem the fixed point equation  $M = \Phi_\Lambda(M)$  admits an unique solution  $M_\Lambda$  on the closed set of  $k \times k$  matrices whose imaginary part is non positive semi definite.

#### 4.3.4 Proof of Lemma 4: control of $\Lambda_N$

For convenience, we denote in this proof  $A = \text{Im } \Lambda$ ,  $A_N = \text{Im } \Lambda_N$  and

$$\varepsilon_N(A) = - \sum_{j=1}^p a_j \text{Im} \left( \tilde{G}_N^{(1)}(\Lambda) - G_N^{(1)}(\Lambda) \right) a_j.$$

Then by definition of  $\Lambda_N$  and  $\Gamma_N$

$$\begin{aligned}
A_N &= \text{Im } \Lambda_N = \text{Im} \left( \Gamma_N + a_0 + \mathcal{R}(\tilde{G}_N^{(1)}(\Lambda)) \right) \\
&= \text{Im} \left( \Lambda - a_0 - \mathcal{R}(G_N^{(1)}(\Lambda)) + a_0 + \mathcal{R}(\tilde{G}_N^{(1)}(\Lambda)) \right) \\
&= A - \varepsilon_N(A)
\end{aligned}$$

By Estimate (4.28) there exist  $N_0, \gamma, \tilde{\gamma}$  such that for every  $N \geq N_0$ , for  $\Lambda$  in  $\mathcal{O}_{\gamma, \tilde{\gamma}}$  we have

$$\begin{aligned}
\|\varepsilon_N(A)\| &\leq \sum_{j=1}^p \|a_j\|^2 \|G_N^{(1)}(\Lambda) - \tilde{G}_N^{(1)}(\Lambda)\| \\
&\leq \sum_{j=1}^p \|a_j\|^2 \frac{c_2}{N^{1+\eta_2}} \|A^{-1}\|^{\alpha_2} + \sum_{j=1}^p \|a_j\|^2 \frac{2c_1}{N^2} \|A^{-1}\|^5 \\
&\leq \frac{c}{N^{1+\eta}} \|A^{-1}\|^{\alpha(A)}
\end{aligned}$$

where  $c = \sum_{j=1}^p \|a_j\|^2 \max(2c_1, c_2)$ ,  $\eta = \min(1, \eta_2)$ , and

$$\alpha(A) = \begin{cases} \min(5, \alpha_2) & \text{if } \|A^{-1}\| \leq 1 \\ \max(5, \alpha_2) & \text{if } \|A^{-1}\| > 1 \end{cases}$$

We write

$$A_N = A(\mathbf{1}_N - A^{-1}\varepsilon_N(A)).$$

If  $\|A^{-1}\| \leq N^\gamma$  then

$$\|A^{-1}\varepsilon_N(A)\| \leq \|A^{-1}\| \|\varepsilon_N(A)\| \leq \frac{c}{N^{1+\eta}} N^{\gamma(\alpha(A)+1)} \leq cN^{\gamma(\alpha(A)+1)-1-\eta}. \quad (4.42)$$

We chose  $\gamma$  such that  $\gamma < (1 + \eta)/(\max(2c_1, c_2) + 1) < (1 + \eta)/(\alpha(A) + 1)$ . Hence the last term in (4.42) tends to zero uniformly in  $A$ , and there exists  $N_1$  (independent with  $A$ ) such that for all  $N \geq N_1$  we have  $\|A^{-1}\varepsilon_N(A)\| < 1$ . For such an  $A$  and  $N$ , we can write

$$A_N^{-1} = \sum_{k \geq 0} \left( A^{-1}\varepsilon_N(A) \right)^k A^{-1}.$$

Notice that the term in the sum for  $k = 0$  is just  $A^{-1}$ , so that

$$\|A_N^{-1} - A^{-1}\| \leq \sum_{k \geq 1} \left\| A^{-1}\varepsilon_N(A) \right\|^k \leq \sum_{k \geq 1} c^k N^{k(\gamma(\alpha+1)-1-\eta)}.$$

The r.h.s. vanishes as  $N \rightarrow \infty$ : more precisely, there exist  $\tilde{\gamma}_1, N_1$  (independent with  $A$ ), such that for all  $N \geq N_1$  one has  $\|A_N^{-1} - A^{-1}\| \leq N^{-\tilde{\gamma}_1}$ . Hence we get  $\|A_N^{-1}\| \leq \|A^{-1}\| + N^{-\tilde{\gamma}_1}$ . But since  $\|A^{-1}\| \geq \|A\|^{-1} \geq N^{-\tilde{\gamma}}$ , with  $\tilde{\gamma}_1 < \tilde{\gamma}$ , there exists a constant still denoted  $N_0$ , such that for  $N \geq N_0$ ,  $N^{-\tilde{\gamma}_1} < N^{-\tilde{\gamma}} \leq \|A^{-1}\|$  and hence we obtain as desired  $\|A_N^{-1}\| \leq 2\|A^{-1}\|$ .

## 5 Proof of Corollaries 1-4

### 5.1 Proof of Corollary 1: diagonal matrices

We first focus on the case of the uniform measure on  $[0, 1]$ . The corresponding matrix is then

$$D_N = \text{diag} \left( \frac{0}{N}, \dots, \frac{N-1}{N} \right). \quad (5.1)$$

We cannot directly apply Theorem 1 to  $D_N$  since it does not satisfy Assumption 3. We introduce an auxiliary random diagonal matrix  $\tilde{D}_N$  that satisfies Assumption 3 and which is sufficiently closed to  $D_N$  in operator norm in order to conclude that Corollary 1 holds for  $D_N$ . The generalization to any family of measures  $\mu_1, \dots, \mu_q$  whose support is a single interval will be obtained by density of polynomials.

#### An auxiliary random matrix

Let  $\tilde{D}_N$  be the  $N \times N$  random diagonal matrix given by

$$\tilde{D}_N = \text{diag} \left( \frac{0 + u_0}{N}, \dots, \frac{N-1 + u_{N-1}}{N} \right), \quad (5.2)$$

the  $(u_i)_{0 \leq i \leq N-1}$  being random variables, independent identically distributed according to the uniform measure on  $[0, 1]$ . We show that the matrix  $D_N$  satisfies the three assumptions of Theorem 1.

**Lemma 5.** *Assumption 1 is satisfied by  $\tilde{D}_N$ : for all  $k \in \mathbb{N}$*

$$\tau_N[\tilde{D}_N^k] \xrightarrow{N \rightarrow \infty} \tau[d^k] := \int_0^1 u^k du, \quad (5.3)$$

*almost surely and in expectation.*

*Proof.* The non commutative law of  $\tilde{D}_N$  in expectation equals the expected limit:

$$\mathbb{E}[\tau_N[\tilde{D}_N^k]] = \frac{1}{N} \sum_{i=0}^{N-1} \int_0^1 \left( \frac{i+u}{N} \right)^k du = \int_0^1 u^k du.$$

For the almost sure limit, remark the following:

$$\tau_N[\tilde{D}_N^k] = \frac{1}{N} \sum_{i=0}^{N-1} \left( \frac{i+u_i}{N} \right)^k = \frac{1}{N} \sum_{i=0}^{N-1} \left( \frac{i}{N} \right)^k + R_N, \quad (5.4)$$

where

$$R_N = \frac{1}{N} \sum_{i=0}^{N-1} \left( \left( \frac{i}{N} \right)^k - \left( \frac{i+u_i}{N} \right)^k \right).$$

The first term in (5.4) is a Riemann sum and converges to  $\tau[d^k]$ . For all  $u \in [0, 1]$  and for  $i = 0, \dots, N-1$  we have  $\left| \left( \frac{i}{N} \right)^k - \left( \frac{i+u_i}{N} \right)^k \right| \leq \frac{k}{N}$ . Then  $|R_N| \leq \frac{k}{N}$  tends to zero almost surely when  $N$  goes to infinity and hence  $\tau_N[\tilde{D}_N^k] \xrightarrow[N \rightarrow \infty]{} \tau[d^k]$ .  $\square$

**Lemma 6.** *Assumption 2 is satisfied by  $\tilde{D}_N$ .*

*Proof.* The uniform measure on  $[0, 1]$  satisfies a Poincaré's inequality with constant  $\pi^{-2}$  (see [8, Section 6.3]). Hence the law of the entries of  $\tilde{D}_N$  satisfies a Poincaré's inequality with constant  $\pi^{-2}N^{-2}$ .  $\square$

**Lemma 7.** *Assumption 3 is satisfied by  $\tilde{D}_N$ .*

*Proof.* Let  $a_0, a_1$  in  $M_k(\mathbb{C})$  be selfadjoint matrices. For all  $\Lambda$  in  $M_k(\mathbb{C})$  such that  $\text{Im } \Lambda > 0$ , define

$$\begin{aligned} G_N(\Lambda) &= \mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - a_1 \otimes \tilde{D}_N)^{-1} \right] \right] \\ G(\Lambda) &= (\text{id}_k \otimes \tau) \left[ (\Lambda \otimes \mathbf{1} - a_0 \otimes \mathbf{1} - a_1 \otimes d)^{-1} \right], \end{aligned}$$

where  $\tau$  is given by (5.3). Then, the mean law of  $\tilde{D}_N$  being the one of  $d$ , the functions are equal:

$$\begin{aligned} G_N(\Lambda) &= \mathbb{E} \left[ \frac{1}{N} \sum_{i=0}^{N-1} \left( \Lambda - a_0 - \left( \frac{i+u_i}{N} \right) a_1 \right)^{-1} \right] = \frac{1}{N} \sum_{i=0}^{N-1} \int_0^1 \left( \Lambda - a_0 - \left( \frac{i+u}{N} \right) a_1 \right)^{-1} du \\ &= \int_0^1 (\Lambda - a_0 - ua_1)^{-1} du = G(\Lambda). \end{aligned}$$

$\square$

**The case of the uniform measure on  $[0, 1]$**

Let  $D_N$  be the  $N \times N$  diagonal matrix given by

$$D_N = \text{diag} \left( \frac{0}{N}, \dots, \frac{N-1}{N} \right). \quad (5.5)$$

Let  $\tilde{D}_N$  be the auxiliary matrix defined in (5.2). By Theorem 1, since  $\tilde{D}_N$  satisfies the three assumptions, if  $\mathbf{X}_N = (X_1^{(N)}, \dots, X_N^{(p)})$  is a family of independent normalized GUE matrices, then for all polynomials  $P$  in  $p+1$  non commutative variables

$$\|P(\mathbf{X}_N, \tilde{D}_N)\| \xrightarrow[N \rightarrow \infty]{} \|P(\mathbf{x}, d)\|_\tau, \quad (5.6)$$

where  $\tau$  is defined by the fact that  $\mathbf{x}$  is a semicircular system, the law of  $d$  is given by (5.3),  $\mathbf{x}$  and  $d$  being free. For any polynomial  $P$ , it is easy to see that there exists  $N_0$  such that for  $N \geq N_0$ , almost surely we have

$$\|P(\mathbf{X}_N, \tilde{D}_N) - P(\mathbf{X}_N, D_N)\| \leq c \|\tilde{D}_N - D_N\| \leq \frac{c}{N}, \quad (5.7)$$

where  $c$  depends only on  $P$  (we have used that  $\|D_N\| \leq 1$ ,  $\|\tilde{D}_N\| \leq 1$  and that for  $N_0$  large enough and  $N \geq N_0$ ,  $\|X_j^{(N)}\| \leq 3$  almost surely for  $j = 1 \dots p$ ). Then

$$\left| \|P(\mathbf{X}_N, D_N)\| - \|P(\mathbf{x}, d)\|_\tau \right| \leq \left| \|P(\mathbf{X}_N, \tilde{D}_N)\| - \|P(\mathbf{x}, d)\|_\tau \right| + \|P(\mathbf{X}_N, \tilde{D}_N) - P(\mathbf{X}_N, D_N)\| \xrightarrow[N \rightarrow \infty]{} 0$$

almost surely.

### The general case

Let  $\mu_1, \dots, \mu_q$  be probability measures whose support consists in a single interval. Recall that this condition is equivalent for the generalized inverse of their cumulative distribution functions  $F_j^{-1}$  to be continuous on  $[0, 1]$ . Let  $\mathbf{D}_N = (D_1^{(N)}, \dots, D_q^{(N)})$  be the matrices given by: for  $j = 1 \dots q$

$$D_j^{(N)} = \text{diag} \left( F_j^{-1} \left( \frac{0}{N} \right), \dots, F_j^{-1} \left( \frac{N-1}{N} \right) \right), \quad (5.8)$$

and let  $D_N$  be the matrix given by (5.1). The non commutative law of  $\mathbf{D}_N$ , respectively,  $D_N$ , converges to the non commutative law of variables  $\mathbf{d} = (d_1, \dots, d_q)$ , respectively,  $d$  in a same  $C^*$ -probability space  $(\mathcal{B}, *, \tau, \|\cdot\|_\tau)$ : for all polynomial  $P$

$$\tau[P(\mathbf{d})] = \int_0^1 P \left( F_1^{-1}(u), \dots, F_q^{-1}(u) \right) du, \quad \tau[P(d)] = \int_0^1 P(u) du.$$

Hence  $\mathbf{d} = (F_1^{-1}(d), \dots, F_q^{-1}(d))$ . By Stone Weierstrass's theorem, for all  $\varepsilon > 0$  there exist polynomials  $P_1, \dots, P_q$  such that: for  $j = 1 \dots q$ , for  $u$  in  $[0, 1]$

$$|F_j^{-1}(u) - P_j(u)| \leq \varepsilon. \quad (5.9)$$

Let  $\mathbf{X}_N = (X_1^{(N)}, \dots, X_N^{(p)})$  be a family of independent normalized GUE matrices. We have seen that for all polynomials  $P$  in  $p+q$  non commutative variables

$$\|P(\mathbf{X}_N, P_1(D_N), \dots, P_q(D_N))\| \xrightarrow{N \rightarrow \infty} \|P(\mathbf{x}, P_1(d), \dots, P_q(d))\|_\tau.$$

By Inequality (5.9) it is easy to see that there exists  $N_0$  such that for  $N \geq N_0$  we have almost surely

$$\begin{aligned} \|P(\mathbf{X}_N, D_1^{(N)}, \dots, D_q^{(N)}) - P(\mathbf{X}_N, P_1(D_N), \dots, P_q(D_N))\| &\leq c\varepsilon \\ \|P(\mathbf{x}, d_1, \dots, d_q) - P(\mathbf{x}, P_1(d), \dots, P_q(d))\|_\tau &\leq c\varepsilon \end{aligned}$$

where  $c$  only depends on  $P$ . Then we have

$$\begin{aligned} \left| \|P(\mathbf{X}_N, \mathbf{D}_N)\| - \|P(\mathbf{x}, \mathbf{d})\|_\tau \right| &\leq \left\| P(\mathbf{X}_N, \mathbf{D}_N) - P(\mathbf{X}_N, P_1(D_N), \dots, P_q(D_N)) \right\| \\ &\quad + \left| \|P(\mathbf{X}_N, P_1(D_N), \dots, P_q(D_N))\| - \|P(\mathbf{x}, P_1(d), \dots, P_q(d))\|_\tau \right| \\ &\quad + \left\| P(\mathbf{x}, \mathbf{d}) - P(\mathbf{x}, P_1(d), \dots, P_q(d)) \right\|_\tau \end{aligned}$$

which vanishes when  $N$  goes to the infinity and  $\varepsilon$  goes to zero.

## 5.2 Proof of Corollary 2: Wishart matrices

Let  $r \geq s$  and  $(\mathbf{W}_N, \mathbf{Y}_N)$  be as in Corollary 2. It is well known that under Assumption 1 the non commutative law of  $(\mathbf{W}_N, \mathbf{Y}_N)$  in  $(M_N(\mathbb{C}), *, \tau_N)$  converges almost surely and in expectation to the non commutative law of  $(\mathbf{w}, \mathbf{y})$  in a  $C^*$ -probability space  $(\mathcal{A}, *, \tau, \|\cdot\|_\tau)$  where

1.  $\mathbf{w} = (w_1, \dots, w_p)$  are free identically distributed selfadjoint non commutative random variables with common law the Marchenko-Pastur distribution:  $\forall k \in \mathbb{N}, \forall j = 1, \dots, p$

$$\tau[w_j^k] = \int t^k d\sigma(t),$$

$$\text{with } d\sigma(t) = \frac{1}{2\pi t} \sqrt{4\lambda - (t - (1 + \sqrt{\lambda}))^2} \mathbf{1}_{(1-\sqrt{\lambda})^2 \leq t \leq (1+\sqrt{\lambda})^2} dt, \quad \lambda = r/s,$$

2.  $\mathbf{y} = (y_1, \dots, y_q)$  is the non commutative limit of  $\mathbf{Y}_N$ ,
3.  $\mathbf{w}$  and  $\mathbf{y}$  being free.

We prove Corollary 2 in the case where the matrices  $\mathbf{Y}_N$  satisfy the three assumptions of Theorem 1. The case where the matrices  $\mathbf{Y}_N$  are diagonal of the form of Corollary 1 can be deduced with minor modifications. 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  $\tilde{\mathbf{X}}_N, \tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*$  and some elementary matrices, where  $\tilde{\mathbf{X}}_N$  is a family of independent GUE matrices and  $\tilde{\mathbf{Y}}_N$  is a modification of  $\mathbf{Y}_N$ . We will obtain the result as a consequence of Theorem 1.

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

$$W_j^{(N)} = M_j^{(N)} M_j^{(N)*}, \quad (5.10)$$

where  $M_j^{(N)}$  is an  $rN \times sN$  complex Gaussian matrix with independent identically distributed entries, centered and of variance  $1/rN$ . Define the  $(r+s) \times (r+s)$  matrices

$$\epsilon_1 = \text{diag}(\mathbf{1}_r, \mathbf{0}_s), \quad (5.11)$$

$$\epsilon_2 = \text{diag}(\mathbf{0}_r, \mathbf{1}_s), \quad (5.12)$$

and the  $(r+s)N \times (r+s)N$  matrices  $\mathbf{e}_N = (e_1^{(N)}, e_2^{(N)})$ :

$$e_1^{(N)} = \epsilon_1 \otimes \mathbf{1}_N = \text{diag}(\mathbf{1}_{rN}, \mathbf{0}_{sN}), \quad (5.13)$$

$$e_2^{(N)} = \epsilon_2 \otimes \mathbf{1}_N = \text{diag}(\mathbf{0}_{rN}, \mathbf{1}_{sN}). \quad (5.14)$$

Let  $\tilde{\mathbf{X}}_N = (\tilde{X}_1^{(N)}, \dots, \tilde{X}_p^{(N)})$  be a family of  $p$  independent  $(r+s)N \times (r+s)N$  normalized GUE matrices, independent with  $\mathbf{Y}_N$  and such that for  $j = 1, \dots, p$ , the  $rN \times sN$  right upper matrix of  $\sqrt{\frac{r+s}{r}} \tilde{X}_j^{(N)}$  is  $M_j^{(N)}$ . Then for  $j = 1, \dots, p$ ,

$$\tilde{M}_j^{(N)} := \sqrt{\frac{r+s}{r}} e_1^{(N)} \tilde{X}_j^{(N)} e_2^{(N)} = \begin{pmatrix} \mathbf{0}_{rN} & M_j^{(N)} \\ \mathbf{0}_{sN, rN} & \mathbf{0}_{sN} \end{pmatrix}. \quad (5.15)$$

and

$$\tilde{W}_j^{(N)} := e_1^{(N)} (\tilde{M}_j^{(N)} + \tilde{M}_j^{(N)*})^2 = \begin{pmatrix} W_j^{(N)} & \mathbf{0}_{rN, sN} \\ \mathbf{0}_{sN, rN} & \mathbf{0}_{sN} \end{pmatrix}. \quad (5.16)$$

Let  $\tilde{\mathbf{Y}}_N = (\tilde{Y}_1^{(N)}, \dots, \tilde{Y}_p^{(N)})$  be the family of  $(r+s)N \times (r+s)N$  matrices defined by: for  $j = 1, \dots, q$

$$\tilde{Y}_j^{(N)} = \epsilon_1 \otimes Y_j^{(N)} = \begin{pmatrix} Y_j^{(N)} & \mathbf{0}_{rN, sN} \\ \mathbf{0}_{sN, rN} & \mathbf{0}_{sN} \end{pmatrix}. \quad (5.17)$$

Corollary 2 follows directly from the following lemma:

**Lemma 8.** 1. *The family of  $(r+s)N \times (r+s)N$  matrices  $(\tilde{\mathbf{Y}}_N, \mathbf{e}_N)$  satisfies the three assumptions of Theorem 1. In particular, by Voiculescu theorem, the non commutative law of  $(\tilde{\mathbf{X}}_N, \tilde{\mathbf{Y}}_N, \mathbf{e}_N)$  in  $(M_{(r+s)N}(\mathbb{C}), *, \tau_{(r+s)N})$  converges almost surely and in expectation to the non commutative law of variables  $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \mathbf{e})$  in a  $C^*$ -probability space  $(\tilde{\mathcal{A}}, *, \tilde{\tau}, \|\cdot\|_{\tilde{\tau}})$ , where  $\tilde{\mathbf{x}}$  is a free semicircular system, free with  $(\tilde{\mathbf{y}}, \mathbf{e})$ .*

2. *For all  $P \in \mathbb{C}\langle \mathbf{w}, \mathbf{y}, \mathbf{y}^* \rangle$  there exists a polynomial  $\tilde{P}$  in  $\mathbb{C}\langle \tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e} \rangle$  such that*

$$\epsilon_1 \otimes P(\mathbf{W}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) = \tilde{P}(\tilde{\mathbf{X}}_N, \tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{e}_N). \quad (5.18)$$

*Moreover, defining the non commutative indeterminates  $\tilde{\mathbf{m}} = (\tilde{m}_1, \dots, \tilde{m}_q)$  and  $\tilde{\mathbf{w}} = (\tilde{w}_1, \dots, \tilde{w}_q)$  by: for  $j = 1, \dots, q$*

$$\tilde{m}_j = \sqrt{\frac{r+s}{r}} e_1 \tilde{x}_j e_2, \quad \tilde{w}_j = e_1 (\tilde{m}_j + \tilde{m}_j^*)^2, \quad (5.19)$$

*then  $\tilde{P}$  is given by  $\tilde{P}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e}) = e_1 P(\tilde{\mathbf{w}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*)$ .*



3. For any  $Q$  in  $\mathbb{C}\langle \tilde{\mathbf{w}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^* \rangle$  one has that  $e_1$  and  $Q(\tilde{\mathbf{w}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*)$  commute and

$$\tilde{\tau}[e_1 Q(\tilde{\mathbf{w}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*)] = \frac{r}{r+s} \tau[Q(\mathbf{w}, \mathbf{y}, \mathbf{y}^*)]. \quad (5.20)$$

In particular

$$\|e_1 Q(\tilde{\mathbf{w}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*)\|_{\tilde{\tau}} = \|Q(\mathbf{w}, \mathbf{y}, \mathbf{y}^*)\|_{\tau}. \quad (5.21)$$

Assuming Lemma 8, Corollary 2 follows easily from Theorem 1: it is obvious that  $\|P(\mathbf{W}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)\| = \|\epsilon_1 \otimes P(\mathbf{W}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)\|$ . But by Item 2 of Lemma 8 and Theorem 1:

$$\|P(\mathbf{W}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)\| = \|\tilde{P}(\tilde{\mathbf{X}}_N, \tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{e}_N)\| \xrightarrow{N \rightarrow \infty} \|\tilde{P}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e})\|_{\tilde{\tau}} = \|e_1 P(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e})\|_{\tilde{\tau}}$$

Hence, by Item 4:

$$\|P(\mathbf{W}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)\| \xrightarrow{N \rightarrow \infty} \|P(\mathbf{w}, \mathbf{y}, \mathbf{y}^*)\|_{\tau}$$

We start with the matrix manipulations, i.e. the proof of Item 2 of Lemma 8, which motivates the introduction of the family of matrices  $(\tilde{\mathbf{Y}}_N, \mathbf{e}_N)$ . Item 3 is a direct consequence of the construction. The proof of the assumptions for  $(\tilde{\mathbf{Y}}_N, \mathbf{e}_N)$  (except the Poincaré's inequality, which is obvious) are stated in Lemmas 9 and 10.

*Proof of Items 2 and 3.* Let  $P$  be a polynomial in  $p + 2q$  non commutative indeterminates  $(\mathbf{w}, \mathbf{y}, \mathbf{y}^*)$ . Then by Definitions (5.16) and (5.17):

$$\begin{aligned} \hat{P}(\mathbf{W}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) &= e_1^{(N)} P(\tilde{\mathbf{W}}_N, \tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*) = e_1^{(N)} P(e_1^{(N)} (\tilde{\mathbf{M}}_N + \tilde{\mathbf{M}}_N^*)^2, \tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*) \\ &= e_1^{(N)} P\left(e_1^{(N)} \frac{r+s}{r} (e_1^{(N)} \tilde{\mathbf{X}}_N e_2^{(N)} + e_2^{(N)} \tilde{\mathbf{X}}_N e_1^{(N)})^2, \tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*\right) \end{aligned}$$

Define  $\tilde{P}$  in  $\mathbb{C}\langle \tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e} \rangle$  by

$$\tilde{P}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e}) = e_1 P\left(e_1 \frac{r+s}{r} (e_1 \tilde{\mathbf{x}} e_2 + e_2 \tilde{\mathbf{x}} e_1)^2, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*\right). \quad (5.22)$$

Then  $\tilde{P}$  satisfies (5.18) and  $\tilde{P}(\tilde{\mathbf{w}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e}) = e_1 P(\tilde{\mathbf{w}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*)$ . □

**Lemma 9.** *Assumption 1 is satisfied by  $(\tilde{\mathbf{Y}}_N, \mathbf{e}_N)$*

*Proof.* Let  $P \in \mathbb{C}\langle \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e} \rangle$ :

$$\begin{aligned} &\tau_{(r+s)N}[P(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{e}_N)] \\ &= \tau_{(r+s)N} \left[ \begin{pmatrix} P(\mathbf{Y}_N, \mathbf{Y}_N^*, \mathbf{1}_{rN}, \mathbf{0}_{rN}) & \mathbf{0}_{rN, sN} \\ \mathbf{0}_{sN, rN} & P(\mathbf{0}_{sN}, \dots, \mathbf{0}_{sN}, \mathbf{0}_{sN}, \mathbf{1}_{sN}) \end{pmatrix} \right] \\ &= \frac{r}{r+s} \tau_{rN}[P(\mathbf{Y}_N, \mathbf{Y}_N^*, \mathbf{1}_{rN}, \mathbf{0}_{rN})] + \frac{s}{r+s} \tau_{sN}[P(\mathbf{0}_{sN}, \dots, \mathbf{0}_{sN}, \mathbf{0}_{sN}, \mathbf{1}_{sN})] \\ &\xrightarrow{N \rightarrow \infty} \frac{r}{r+s} \tau[P(\mathbf{y}, \mathbf{y}^*, \mathbf{1}, \mathbf{0})] + \frac{s}{r+s} \tau_s[P(\mathbf{0}_s, \dots, \mathbf{0}_s, \mathbf{0}_s, \mathbf{1}_s)] \end{aligned} \quad (5.23)$$

almost surely and in expectation since  $\mathbf{Y}_N$  is supposed to satisfy Assumption 1. □

**Lemma 10.** *Assumption 3 is satisfied by  $(\tilde{\mathbf{Y}}_N, \mathbf{e}_N)$*

*Proof.* Let

$$L(\tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e}) = a_0 \otimes \mathbf{1} + a_1 \otimes e_1 + a_2 \otimes e_2 + \sum_{j=1}^q b_j \otimes y_j + b_j^* \otimes y_j^*$$

where  $a_0, a_1, a_2, b_1, \dots, b_q$  are in  $M_k(\mathbb{C})$ ,  $a_0, a_1, a_2$  being selfadjoint. Then the Stieltjes transform  $G_N$  of  $L(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{e}_N)$  is given by:  $\forall \Gamma \in M_k(\mathbb{C})$  such that  $\text{Im } \Gamma > 0$

$$\begin{aligned} G_N(\Gamma) &:= \mathbb{E} \left[ \tau_{(r+s)N} \left[ (\Gamma \otimes \mathbf{1}_{(r+s)N} - L(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{e}_N))^{-1} \right] \right] \\ &= \mathbb{E} \left[ \frac{r}{r+s} \tau_{rN} \left[ (\Gamma \otimes \mathbf{1}_{rN} - (a_0 + a_1) \otimes \mathbf{1}_{rN} - T_N)^{-1} \right] \right] + \frac{s}{r+s} (\Gamma - a_0 - a_1)^{-1}, \end{aligned}$$

where  $T_N = \sum_{j=1}^q b_j \otimes Y_j^{(N)} + b_j^* \otimes Y_j^{(N)*}$ . Similarly, by (5.23), the Stieltjes transform  $G$  of the non commutative random variables  $L(\tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e})$  is:  $\forall \Gamma \in M_k(\mathbb{C})$  such that  $\text{Im } \Gamma > 0$

$$G(\Gamma) = \frac{r}{r+s} \tau \left[ (\Gamma \otimes \mathbf{1} - (a_0 + a_1) \otimes \mathbf{1} - \sum_{j=1}^q b_j \otimes y_j + b_j^* \otimes y_j^*)^{-1} \right] + \frac{s}{r+s} (\Gamma - a_0 - a_1)^{-1}$$

Hence

$$\|G_N(\Gamma) - G(\Gamma)\| = \frac{r}{r+s} \|\tilde{G}_N(\Gamma) - \tilde{G}(\Gamma)\|$$

where  $\tilde{G}_N$  (respectively  $\tilde{G}$ ) is the Stieltjes transform of  $(a_0 + a_1) \otimes \mathbf{1}_{rN} - \sum_{j=1}^q b_j \otimes Y_j^{(N)} + b_j^* \otimes Y_j^{(N)*}$  (respectively of  $(a_0 + a_1) \otimes \mathbf{1} - \sum_{j=1}^q b_j \otimes y_j + b_j^* \otimes y_j^*$ ). Hence (3.10) holds for  $\|G_N(\Gamma) - G(\Gamma)\|$  since  $\mathbf{Y}_N$  is supposed to satisfy Assumption 3.  $\square$

### 5.3 Proof of Corollary 3: Block matrices

The convergence in norm for block matrices stated in Corollary 3 can be shown in an algebraic way as a consequence of the strong convergence (1.4). We provide here a direct proof: it is shown to be a consequence of Theorem 1 using some elementary block matrices that are compatible with the three assumptions of the theorem.

Let  $(\mathbf{X}_N, \mathbf{Y}_N)$  as in Theorem 1. By Assumption 1 and Voiculescu's theorem, the non commutative law of  $(\mathbf{X}_N, \mathbf{Y}_N)$  in  $(M_N(\mathbb{C}), *, \tau_N)$  converges to the non commutative law of non commutative variables  $(\mathbf{x}, \mathbf{y})$  in a  $C^*$ -probability space  $(\mathcal{A}, *, \tau, \|\cdot\|_\tau)$ .

Fix in the following a positive integer  $l$  and denote by  $(\epsilon_{u,v})_{1 \leq u, v \leq l}$  the canonical basis of  $M_l(\mathbb{C})$ . Define the family of  $lN \times lN$  matrices  $\mathbf{e}_N = (e_1^{(N)}, \dots, e_l^{(N)})$ : for all  $j = 1, \dots, l$

$$e_j^{(N)} = \epsilon_{j,1} \otimes \mathbf{1}_N = \begin{pmatrix} \mathbf{0}_N & \dots & \dots & \dots & \mathbf{0}_N \\ \vdots & & & & \vdots \\ \mathbf{0}_N & & & & \vdots \\ \mathbf{1}_N & & & & \vdots \\ \mathbf{0}_N & & & & \vdots \\ \vdots & & & & \vdots \\ \mathbf{0}_N & \dots & \dots & \dots & \mathbf{0}_N \end{pmatrix}, \quad (5.24)$$

the identity block being at position  $(j, 1)$ . 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  $lN \times lN$ , independent with  $\mathbf{Y}_N$ , and such that for  $j = 1, \dots, p$ , the  $N \times N$  block of  $\sqrt{l}X_j^{(N)}$  in position  $(1, 1)$  is  $X_j^{(N)}$ . Then we get

$$\epsilon_{1,1} \otimes \mathbf{X}_N = e_{1,1}^{(N)} \sqrt{l} \tilde{\mathbf{X}}_N e_{1,1}^{(N)}. \quad (5.25)$$

Let  $\tilde{\mathbf{Y}}_N = (\tilde{Y}_1^{(N)}, \dots, \tilde{Y}_q^{(N)})$  be the family of  $lN \times lN$  matrices defined by: for  $j = 1, \dots, q$ ,

$$\tilde{Y}_j^{(N)} = \epsilon_{1,1} \otimes Y_j^{(N)} = \begin{pmatrix} Y_j^{(N)} & \mathbf{0}_N & \dots & \mathbf{0}_N \\ \mathbf{0}_N & \mathbf{0}_N & \dots & \mathbf{0}_N \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_N & \mathbf{0}_N & \dots & \mathbf{0}_N \end{pmatrix}. \quad (5.26)$$

**Lemma 11.** 1. *The family of  $lN \times lN$  matrices  $(\tilde{\mathbf{Y}}_N, \mathbf{e}_N)$  satisfies the three assumptions of Theorem 1. In particular, by Voiculescu theorem, the non commutative law of  $(\tilde{\mathbf{X}}_N, \tilde{\mathbf{Y}}_N, \mathbf{e}_N)$  in  $(M_{lN}(\mathbb{C}), *, \tau_{lN})$  converges almost surely and in expectation to the non commutative law of variables  $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \mathbf{e})$  in a  $C^*$ -probability space  $(\tilde{\mathcal{A}}, *, \tilde{\tau}, \|\cdot\|_{\tilde{\tau}})$ , where  $\tilde{\mathbf{x}}$  is a free semicircular system, free with  $(\tilde{\mathbf{y}}, \mathbf{e})$ .*

2. For all  $B \in M_l(\mathbb{C}) \otimes \mathbb{C}\langle \mathbf{x}, \mathbf{y}, \mathbf{y}^* \rangle$ , there exists a polynomial  $\tilde{B}$  in  $\mathbb{C}\langle \tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e}, \mathbf{e}^* \rangle$  such that

$$B(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) = \tilde{B}(\tilde{\mathbf{X}}_N, \tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{e}_N, \mathbf{e}_N^*). \quad (5.27)$$

Moreover,  $\tilde{B}$  is in fact a polynomial in  $(e_{11}\sqrt{l}\tilde{\mathbf{x}}e_{11}, \tilde{\mathbf{y}}, \mathbf{e})$ ;

3. The law of  $(\epsilon_{11} \otimes \mathbf{x}, \epsilon_{11} \otimes \mathbf{y}, (\epsilon_{j,1} \otimes \mathbf{1})_{j=1,\dots,l})$  in  $(M_l(\mathbb{C}) \otimes \mathcal{A}, \cdot, \tau_l \otimes \tau)$  is the law of  $(e_{11}\sqrt{l}\tilde{\mathbf{x}}e_{11}, \tilde{\mathbf{y}}, \mathbf{e})$  in  $(\tilde{\mathcal{A}}, \cdot, \tilde{\tau})$  and

$$\|B(\mathbf{x}, \mathbf{y}, \mathbf{y}^*)\|_{\tau_l \otimes \tau} = \|\tilde{B}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e}, \mathbf{e}^*)\|_{\tilde{\tau}}. \quad (5.28)$$

Assuming Lemma 11 we obtain Corollary 3: Theorem 1 ensures that, almost surely

$$\|B(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)\| = \|\tilde{B}(\tilde{\mathbf{X}}_N, \tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{e}_N, \mathbf{e}_N^*)\| \xrightarrow{N \rightarrow \infty} \|\tilde{B}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e}, \mathbf{e}^*)\|_{\tilde{\tau}} = \|B(\mathbf{x}, \mathbf{y}, \mathbf{y}^*)\|_{\tau_l \otimes \tau}.$$

We start with the proof of Item 2 of Lemma 11. Item 3 of Lemma 11 is a direct consequence of Item 1 and of the proof of Item 2. The proof of the assumptions for  $(\tilde{\mathbf{Y}}_N, \mathbf{e}_N)$  (except the Poincaré's inequality, which is obvious) are stated in Lemmas 12 and 13.

*Proof of Items 2 and 3.* Let  $B \in M_l(\mathbb{C}) \otimes \mathbb{C}\langle \mathbf{x}, \mathbf{y}, \mathbf{y}^* \rangle$  and write

$$B(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) = \sum_{u,v=1}^l \epsilon_{u,v} \otimes P_{u,v}(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) = \begin{pmatrix} P_{1,1}(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) & \dots & P_{1,l}(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) \\ \vdots & & \vdots \\ P_{l,1}(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) & \dots & P_{l,l}(\mathbf{x}, \mathbf{y}, \mathbf{y}^*) \end{pmatrix}, \quad (5.29)$$

$(P_{u,v})_{1 \leq u,v \leq l}$  being non commutative polynomials in  $p + 2q$  indeterminates. Then by Definition (5.24) of  $\mathbf{e}_N$ :

$$B(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) = \sum_{u,v=1}^l e_{u,1}^{(N)} \cdot \epsilon_{1,1} \otimes P_{u,v}(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) \cdot e_{v,1}^{(N)*}.$$

It is easy to see that

$$B(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) = \sum_{u,v=1}^l e_{u,1}^{(N)} P_{u,v}(\epsilon_{1,1} \otimes \mathbf{X}_N, \tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*) e_{v,1}^{(N)*}.$$

Define the polynomial  $\tilde{B} \in \mathbb{C}\langle \tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e}, \mathbf{e}^* \rangle$  by

$$\tilde{B} = \sum_{u,v=1}^l e_{u,1} P_{u,v}(e_{1,1}\sqrt{l}\tilde{\mathbf{x}}e_{1,1}, \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*) e_{v,1}^*. \quad (5.30)$$

Then by (5.25) we get as desired  $B(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*) = \tilde{B}(\tilde{\mathbf{X}}_N, \tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{e}_N, \mathbf{e}_N^*)$  and  $\tilde{B}$  is a polynomial in the non commutative indeterminates  $(e_{11}\sqrt{l}\tilde{\mathbf{x}}e_{11}, \tilde{\mathbf{y}}, \mathbf{e})$ .  $\square$

We now show the assumptions for  $(\tilde{\mathbf{Y}}_N, \mathbf{e}_N)$ . Recall that since  $\mathbf{Y}_N$  satisfies Assumption 1, there exists a tracial state  $\tau$  such that for all non commutative polynomial  $P \in \mathbb{C}\langle \mathbf{y}, \mathbf{y}^* \rangle$ ,

$$\tau_N [ P(\mathbf{Y}_N, \mathbf{Y}_N^*) ] \xrightarrow{N \rightarrow \infty} \tau [ P(\mathbf{y}, \mathbf{y}^*) ] \quad \text{a.s. and in expectation.} \quad (5.31)$$

**Lemma 12.** *Assumption 1 is satisfied by  $(\tilde{\mathbf{Y}}_N, \mathbf{e}_N)$ . More precisely, the non commutative law of  $(\tilde{\mathbf{Y}}_N, \mathbf{e}_N)$  in  $(M_l(\mathbb{C}), \cdot, \tau_l)$  converges almost surely and in expectation to the non commutative law given by the trace  $\tilde{\tau}$  given by the following. Denote by  $\tilde{\mathbf{y}} = (\tilde{y}_1, \dots, \tilde{y}_q)$ ,  $\mathbf{e} = (e_1, \dots, e_l)$  non commutative indeterminates. For all monic monomial  $Q$  in  $\mathbb{C}\langle \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e}, \mathbf{e}^* \rangle$*

$$\tau_{lN} [ Q(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{e}, \mathbf{e}^*) ] \xrightarrow{N \rightarrow \infty} \tilde{\tau} [ Q(\tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e}, \mathbf{e}^*) ] \quad \text{a.s. and in expectation}$$

where

$$\tilde{\tau} [ Q(\tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e}, \mathbf{e}^*) ] = \tau_l \left[ Q(\epsilon_{1,1}, \dots, \epsilon_{1,1}, (\epsilon_{j,1})_{j=1,\dots,l}, (\epsilon_{1,j})_{j=1,\dots,l}) \right] \tau \left[ Q(\tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, (\mathbf{1})_{j=1,\dots,l}, (\mathbf{1})_{j=1,\dots,l}) \right],$$

and  $\epsilon = (\epsilon_{u,v})_{1 \leq u,v \leq l}$  denote the canonical basis of  $M_l(\mathbb{C})$ . Hence  $\tilde{\tau} = (\tau_k \otimes \tau)$ .

*Proof.* Recall that with the tensor product notation we have written in (5.24) and (5.26)

$$\begin{aligned}\tilde{\mathbf{Y}}_j^{(N)} &= \epsilon_{1,1} \otimes \mathbf{Y}_j^{(N)}, \quad j = 1, \dots, q, \\ e_j^{(N)} &= \epsilon_{j,1} \otimes \mathbf{1}_N, \quad j = 1, \dots, l.\end{aligned}$$

Now let  $Q \in \mathbb{C}\langle \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e}, \mathbf{e}^* \rangle$  be a non commutative monic monomial; then by the product rule for tensor product  $a \otimes b \cdot c \otimes d = ac \otimes bd$  we have

$$\begin{aligned}Q(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{e}, \mathbf{e}^*) &= Q(\epsilon_{1,1} \otimes \mathbf{Y}_N, \epsilon_{1,1} \otimes \mathbf{Y}_N^*, (\epsilon_{j,1} \otimes \mathbf{1}_N)_{j=1, \dots, l}, (\epsilon_{1,j} \otimes \mathbf{1}_N)_{j=1, \dots, l}) \\ &= Q(\epsilon_{1,1}, \dots, \epsilon_{1,1}, (\epsilon_{j,1})_{j=1, \dots, l}, (\epsilon_{1,j})_{j=1, \dots, l}) \otimes Q(\mathbf{Y}_N, \mathbf{Y}_N^*, (\mathbf{1}_N)_{j=1, \dots, l}, (\mathbf{1}_N)_{j=1, \dots, l}).\end{aligned}$$

Taking the normalized trace and letting  $N$  go to the infinity gives the result

$$\begin{aligned}\tau_{1N} \left[ Q(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{e}, \mathbf{e}^*) \right] &= (\tau_l \otimes \tau_N) \left[ Q(\epsilon_{1,1}, \dots, \epsilon_{1,1}, (\epsilon_{j,1})_{j=1, \dots, l}, (\epsilon_{1,j})_{j=1, \dots, l}) \otimes Q(\mathbf{Y}_N, \mathbf{Y}_N^*, (\mathbf{1}_N)_{j=1, \dots, l}, (\mathbf{1}_N)_{j=1, \dots, l}) \right] \\ &= \tau_l \left[ Q(\epsilon_{1,1}, \dots, \epsilon_{1,1}, (\epsilon_{j,1})_{j=1, \dots, l}, (\epsilon_{1,j})_{j=1, \dots, l}) \right] \tau_N \left[ Q(\mathbf{Y}_N, \mathbf{Y}_N^*, (\mathbf{1}_N)_{j=1, \dots, l}, (\mathbf{1}_N)_{j=1, \dots, l}) \right] \\ &\xrightarrow{N \rightarrow \infty} \tau_l \left[ Q(\epsilon_{1,1}, \dots, \epsilon_{1,1}, (\epsilon_{j,1})_{j=1, \dots, l}, (\epsilon_{1,j})_{j=1, \dots, l}) \right] \tau \left[ Q(\tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, (\mathbf{1})_{j=1, \dots, l}, (\mathbf{1})_{j=1, \dots, l}) \right]\end{aligned}$$

almost surely and in expectation.  $\square$

It remains to show that the convergence of  $M_k(\mathbb{C})$ -valued Stieltjes transforms holds rapidly. Since  $\mathbf{Y}_N$  is supposed to satisfy Assumption 3 we have the following. For all non commutative selfadjoint degree one polynomial  $L$  in  $(\mathbf{y}, \mathbf{y}^*)$  with coefficient in  $M_k(\mathbb{C})$ , with  $G_N^L$  denoting the Stieltjes transform of  $L(\mathbf{Y}_N, \mathbf{Y}_N^*)$ ,  $G^L$  the one of  $L(\mathbf{y}, \mathbf{y}^*)$ , respectively defined by: for  $\Gamma$  in  $M_k(\mathbb{C})$  such that  $\text{Im } \Gamma > 0$

$$\begin{aligned}G_N^L(\Gamma) &= \mathbb{E} \left[ (\text{id}_k \otimes \tau_N) \left[ (\Gamma \otimes \mathbf{1}_N - L(\mathbf{Y}_N, \mathbf{Y}_N^*))^{-1} \right] \right] \\ G^L(\Gamma) &= (\text{id}_k \otimes \tau) \left[ (\Gamma \otimes \mathbf{1} - L(\mathbf{y}, \mathbf{y}^*))^{-1} \right]\end{aligned}$$

there exist  $N_0, \gamma, \tilde{\gamma}, c, \eta, \alpha > 0$  such that for all  $N \geq N_0$ , for all matrices

$$\Gamma \in \mathcal{O}_{\gamma, \tilde{\gamma}} = \left\{ \Lambda \in M_k(\mathbb{C}) \mid \text{Im } \Lambda > 0, \|(\text{Im } \Lambda)^{-1}\| \leq N^\gamma, \|\Lambda\| \leq N^{\tilde{\gamma}} \right\},$$

we have

$$\|G_N^L(\Gamma) - G^L(\Gamma)\| \leq \frac{c}{N^{1+\eta}} \|(\text{Im } \Gamma)^{-1}\|^\alpha. \quad (5.32)$$

**Lemma 13.** *Assumption 3 is satisfied by  $(\mathbf{Y}_N, \mathbf{e}_N)$ .*

*Proof.* Consider a non commutative selfadjoint degree one polynomial  $\tilde{L}$  with coefficient in  $M_k(\mathbb{C})$  in non commutative indeterminates  $(\tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e}, \mathbf{e}^*)$ :

$$\tilde{L}(\tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e}, \mathbf{e}^*) = \tilde{a}_0 \otimes \mathbf{1} + \sum_{j=1}^l \tilde{a}_j \otimes e_j + \tilde{a}_j^* \otimes e_j^* + \sum_{j=1}^q \tilde{b}_j \otimes \tilde{y}_j + \tilde{b}_j^* \otimes \tilde{y}_j^*,$$

with  $a_0$  selfadjoint. We can reduce the problem by considering the polynomial  $\tilde{L}(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{e}_N, \mathbf{e}_N^*)$  as a polynomial  $L$  in  $(\mathbf{Y}_N, \mathbf{Y}_N^*)$ , selfadjoint but with coefficient in  $M_{lk}(\mathbb{C})$ : by Definitions (5.26) and (5.24)

of  $\tilde{\mathbf{Y}}_N$  and of  $\mathbf{e}_N$

$$\begin{aligned}
& \tilde{L}(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{e}_N, \mathbf{e}_N^*) \\
&= \tilde{a}_0 \otimes \mathbf{1}_{lN} + \sum_{j=1}^l \tilde{a}_j \otimes e_j^{(N)} + \tilde{a}_j^* \otimes e_j^{(N)*} + \sum_{j=1}^q \tilde{b}_j \otimes \tilde{Y}_j^{(N)} + \tilde{b}_j^* \otimes \tilde{Y}_j^{(N)*} \\
&= \tilde{a}_0 \otimes \mathbf{1}_l \otimes \mathbf{1}_N + \sum_{j=1}^l \tilde{a}_j \otimes \epsilon_{j,1} \otimes \mathbf{1}_N + \tilde{a}_j^* \otimes \epsilon_{1,j} \otimes \mathbf{1}_N + \sum_{j=1}^q \tilde{b}_j \otimes \epsilon_{1,1} \otimes Y_j^{(N)} + \tilde{b}_j^* \otimes \epsilon_{1,1} \otimes Y_j^{(N)*} \\
&= a_0 \otimes \mathbf{1}_N + \sum_{j=1}^q b_j \otimes Y_j^{(N)} + b_j^* \otimes Y_j^{(N)*} =: L(\mathbf{Y}_N, \mathbf{Y}_N)
\end{aligned}$$

where  $a_0 = \tilde{a}_0 \otimes \mathbf{1}_l + \sum_{j=1}^l \tilde{a}_j \otimes \epsilon_j + \tilde{a}_j^* \otimes \epsilon_j^*$  is selfadjoint,  $b_j = \tilde{b}_j \otimes \epsilon_{1,1}$  for  $j = 1, \dots, q$ . The  $M_k(\mathbb{C})$ -valued Stieltjes transform  $G_N^{\tilde{L}}$  of  $\tilde{L}(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{e}_N, \mathbf{e}_N^*)$  can be written as the partial trace of the  $M_{lk}(\mathbb{C})$ -valued Stieltjes transform  $G_N^L$  of  $L(\mathbf{Y}_N, \mathbf{Y}_N^*)$ : for  $\tilde{\Gamma}$  in  $M_k(\mathbb{C})$  such that  $\text{Im } \tilde{\Gamma} > 0$

$$\begin{aligned}
G_N^{\tilde{L}}(\tilde{\Gamma}) &= \mathbb{E} \left[ (id_k \otimes \tau_N) \left[ (\tilde{\Gamma} \otimes \mathbf{1}_{lN} - \tilde{L}(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{e}_N, \mathbf{e}_N^*))^{-1} \right] \right] \\
&= (id_k \otimes \tau) \left[ \mathbb{E} \left[ (id_{lk} \otimes \tau_N) \left[ (\tilde{\Gamma} \otimes \mathbf{1}_{lN} - \tilde{L}(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*, \mathbf{e}_N, \mathbf{e}_N^*)) \right] \right] \right] \\
&= (id_k \otimes \tau) \left[ G_N^L(\Gamma) \right],
\end{aligned}$$

where  $\Gamma = \tilde{\Gamma} \otimes \mathbf{1}_l$  and  $G_N^L$  is the  $M_{lk}(\mathbb{C})$ -valued Stieltjes transform of  $L(\tilde{\mathbf{Y}}_N, \tilde{\mathbf{Y}}_N^*)$ . Remark that we have  $\text{Im } \Gamma > 0$ ,  $\|\Gamma\| = \|\tilde{\Gamma}\|$  and  $\|(\text{Im } \Gamma)^{-1}\| = \|(\text{Im } \tilde{\Gamma})^{-1}\|$ .

Similarly the  $M_k(\mathbb{C})$ -valued Stieltjes transform of  $L(\tilde{\mathbf{y}}, \tilde{\mathbf{y}}^*, \mathbf{e}, \mathbf{e}^*)$  can be re-written: for  $\tilde{\Gamma}$  in  $M_k(\mathbb{C})$  such that  $\text{Im } \tilde{\Gamma} > 0$

$$G^{\tilde{L}}(\tilde{\Gamma}) = (id_k \otimes \tau) \left[ G^L(\Gamma) \right]$$

Hence by Cauchy-Schwarz inequality and (5.32): for  $\tilde{\Gamma}$  in  $\mathcal{O}_{\gamma, \tilde{\gamma}}$  we have

$$\begin{aligned}
\|G_N^{\tilde{L}}(\tilde{\Gamma}) - G^{\tilde{L}}(\tilde{\Gamma})\| &= \left\| (id_k \otimes \tau) \left[ G_N^L(\Gamma) - G^L(\Gamma) \right] \right\| \leq \frac{1}{\sqrt{l}} \left\| G_N^L(\Gamma) - G^L(\Gamma) \right\|_e \\
&\leq \sqrt{k} \left\| G_N^L(\Gamma) - G^L(\Gamma) \right\| \leq l^{1+\eta} \sqrt{k} \frac{c}{(lN)^{1+\eta}} \|(\text{Im } \Gamma)^{-1}\|^\alpha
\end{aligned}$$

Hence the family  $(\tilde{\mathbf{Y}}_N, \mathbf{e}_N)$  satisfies Assumption 3 with a constant multiplied by  $l^{1+\eta} \sqrt{k}$ .  $\square$

#### 5.4 Proof of Corollary 4: Rectangular band matrices

We only sketch the proof of this example, details are obtained by minor modification of the proofs of Corollaries 2 and 3. Let  $H$  be as in Corollary 4:

$$H = \begin{pmatrix} A_1 & A_2 & \dots & A_L & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & A_1 & A_1 & \dots & A_L & \mathbf{0} & & \vdots \\ \vdots & \mathbf{0} & A_1 & A_2 & \dots & A_L & \mathbf{0} & \\ & & \ddots & \ddots & \ddots & & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \dots & \mathbf{0} & A_1 & A_2 & \dots & A_L \end{pmatrix}. \quad (5.33)$$

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 which blocks consist in sums of  $tN \times tN$  matrices of the form  $A_l^* A_m$ ,  $l, m = 1 \dots L$ . For any  $l = 1 \dots L$ ,  $A_l = m_l + \sigma_l M_l$  were  $M_l$  is  $rN \times tN$  complex Gaussian matrix with independent identically distributed entries, centered and of variance  $1/N$ .

By minor modifications of the proof of Corollary 2, we get that the norm of any polynomial in the matrices  $\mathbf{A}_N = (A_l^* A_m)_{l,m=1..L}$  and in matrices  $\mathbf{Y}_N$  that satisfy Assumptions 1-3 converges when  $N$  goes to the infinity. By the same reasoning as in the proof of Corollary 3, we get that the convergence holds for square block matrices and in particular for  $H^*H$ . Hence the result follows.

## A Proof of Proposition 2: estimate of $\Theta_N(\Lambda, \Gamma)$

We use the Poincaré inequality to control  $\Theta_N$ : if  $(g_1, \dots, g_N)$  are independent identically distributed centered real Gaussian variables with variance  $v^2$  and  $F$  is a differentiable map  $\mathbb{R}^N \rightarrow \mathbb{C}$  such that  $F$  and its partial derivatives are polynomially bounded, then (see [9, Theorem 2.1])

$$\text{Var}\left(F(g_1, \dots, g_N)\right) \leq v^2 \mathbb{E}\left[\|\nabla F(g_1, \dots, g_N)\|^2\right].$$

The Poincaré inequality is compatible with tensor product, then by Assumption (3.2) such a formula will be valid when  $F$  is a function of the matrices  $\mathbf{X}_N$  and  $\mathbf{Y}_N$  with  $v^2 = \frac{\sigma}{N}$  (without loss of generality, one may suppose  $\sigma \geq 1$ ).

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  $K$  integer, we denote the Euclidean norm of a  $K \times K$  matrix  $A = (a_{m,n})_{1 \leq m, n \leq K}$  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,

$$\|A\| \leq \|A\|_e \leq \sqrt{K} \|A\|, \quad (\text{A.1})$$

$$\|A\| \leq \sqrt{K} \|A\|_\infty \leq \sqrt{K} \|A\|_e, \quad (\text{A.2})$$

$$\|AB\| \leq \|A\|_e \|B\|. \quad (\text{A.3})$$

When  $A$  is in  $M_k(\mathbb{C}) \otimes M_N(\mathbb{C})$ , its euclidean norm is defined by considering  $A$  as a  $kN \times kN$  matrix. In the following we will write an element  $\omega$  of  $M_k(\mathbb{C}) \otimes M_N(\mathbb{C})$

$$\omega = \sum_{m,n=1}^N \sum_{u,v=1}^k w_{u,v}^{m,n} \epsilon_{u,v} \otimes \epsilon_{m,n} = \sum_{m,n=1}^N w^{(m,n)} \otimes \epsilon_{m,n} = \sum_{u,v=1}^k \epsilon_{u,v} \otimes w_{(u,v)}, \quad (\text{A.4})$$

where for  $m, n = 1, \dots, N$  and  $u, v = 1, \dots, k$ ,  $w_{u,v}^{m,n}$  is a complex,  $\omega^{(m,n)}$  is a  $k \times k$  matrix, and  $w_{(u,v)}$  is a  $N \times N$  matrix; we use the same notation for the canonical bases of  $M_k(\mathbb{C})$  and  $M_N(\mathbb{C})$ .

We denote for convenience  $M = \sum_{j=1}^p a_j K_N^{(1)}(\Lambda) a_j$  and denote by  $(\tilde{h}_N^{(1)}, \tilde{h}_N^{(2)})$  an independent copy of  $(h_N^{(1)}, h_N^{(2)})$ . Recall that by definitions (4.6) and (4.7):

$$\begin{aligned} l_{N,\Lambda,\Gamma} &: A \in M_k(\mathbb{C}) \mapsto h_N^{(2)}(\Gamma) A h_N^{(1)}(\Lambda) \in M_k(\mathbb{C}), \\ L_{N,\Lambda,\Gamma} &: A \in M_k(\mathbb{C}) \mapsto \mathbb{E}\left[l_{N,\Lambda,\Gamma}(A)\right] \in M_k(\mathbb{C}). \end{aligned}$$

Then with the notations of (A.4) we have

$$\begin{aligned} &(\text{id}_k \otimes \tau_N) \left[ (l_{N,\Lambda,\Gamma} - L_{N,\Lambda,\Gamma})(M \otimes \mathbf{1}_N) \right] \\ &= (\text{id}_k \otimes \tau_N) [h_N^{(2)}(\Gamma)(M \otimes \mathbf{1}_N) h_N^{(1)}(\Lambda)] - \mathbb{E} \left[ (\text{id}_k \otimes \tau_N) [\tilde{h}_N^{(2)}(\Gamma)(M \otimes \mathbf{1}_N) \tilde{h}_N^{(1)}(\Lambda)] \mid M \right] \\ &= \frac{1}{N} \sum_{m,n=1}^N \left[ \left( h_N^{(2)}(\Gamma) \right)^{(m,n)} M \left( h_N^{(1)}(\Lambda) \right)^{(n,m)} - \mathbb{E} \left[ \left( \tilde{h}_N^{(2)}(\Gamma) \right)^{(m,n)} M \left( \tilde{h}_N^{(1)}(\Lambda) \right)^{(n,m)} \mid M \right] \right]. \end{aligned}$$

To estimate the operator norm of  $\Theta_N$  we use the domination by the infinity norm (A.2) in order to split the contributions due to  $M$  and due to  $l_{N,\Lambda,\Gamma} - L_{N,\Lambda,\Gamma}$ : we get

$$\begin{aligned}
\|\Theta_N(\Lambda, \Gamma)\| &= \left\| \mathbb{E} \left[ (\text{id}_k \otimes \tau_N) [(l_{N,\Lambda,\Gamma} - L_{N,\Lambda,\Gamma}) (M \otimes \mathbf{1}_N)] \right] \right\| \\
&\leq \sqrt{k} \left\| \mathbb{E} \left[ \frac{1}{N} \sum_{m,n=1}^N \left( h_N^{(2)}(\Gamma) \right)^{(m,n)} M \left( h_N^{(1)}(\Lambda) \right)^{(n,m)} - \mathbb{E} \left[ \left( \tilde{h}_N^{(2)}(\Gamma) \right)^{(m,n)} M \left( \tilde{h}_N^{(1)}(\Lambda) \right)^{(n,m)} \mid M \right] \right] \right\|_\infty \\
&\leq k^{5/2} \max_{\substack{1 \leq u, v \leq k \\ 1 \leq u', v' \leq k}} \left| \mathbb{E} \left[ M_{u',v'} \times \frac{1}{N} \sum_{m,n=1}^N \left( h_N^{(2)}(\Gamma) \right)_{u,u'}^{m,n} \left( h_N^{(1)}(\Lambda) \right)_{v',v}^{n,m} - \mathbb{E} \left[ \left( h_N^{(2)}(\Gamma) \right)_{u,u'}^{m,n} \left( h_N^{(1)}(\Lambda) \right)_{v',v}^{n,m} \right] \right] \right| \\
&\leq k^{5/2} \max_{u,v,u',v'} \mathbb{E} \left[ |M_{u',v'}| \times \left| \tau_N [h_N^{(1,2)}(\Lambda, \Gamma)_{u,v}{}_{u',v'}] - \mathbb{E} [\tau_N [h_N^{(1,2)}(\Lambda, \Gamma)_{u,v}{}_{u',v'}]] \right| \right] \\
&\leq k^{5/2} \max_{u,v,u',v'} \mathbb{E} \left[ |M_{u',v'}| \times \left| \tau_N [k_N^{(1,2)}(\Lambda, \Gamma)_{u,v}{}_{u',v'}] \right| \right],
\end{aligned}$$

where we have denoted

$$\begin{aligned}
h_N^{(1,2)}(\Lambda, \Gamma)_{u,v}{}_{u',v'} &= \left( h_N^{(2)}(\Gamma) \right)_{(u,u')} \left( h_N^{(1)}(\Lambda) \right)_{(v',v)}, \\
k_N^{(1,2)}(\Lambda, \Gamma)_{u,v}{}_{u',v'} &= h_N^{(1,2)}(\Lambda, \Gamma)_{u,v}{}_{u',v'} - \mathbb{E} \left[ h_N^{(1,2)}(\Lambda, \Gamma)_{u,v}{}_{u',v'} \right].
\end{aligned}$$

Remark that by (A.3), for  $u', v' = 1, \dots, k$

$$|M_{u',v'}| = \left| \left( \sum_{j=1}^p a_j K_N^{(1)}(\Lambda) a_j \right)_{u',v'} \right| \leq \left\| \sum_{j=1}^p a_j K_N^{(1)}(\Lambda) a_j \right\|_e \leq \sum_{j=1}^p \|a_j\| \|K_N^{(1)}(\Lambda)\|_e.$$

Then by Cauchy-Schwarz inequality we get:

$$\begin{aligned}
\|\Theta_N(\Lambda, \Gamma)\| &\leq k^{5/2} \sum_{j=1}^p \|a_j\| \left( \mathbb{E} [\|K_N^{(1)}(\Lambda)\|_e^2] \max_{u,v,u',v'} \mathbb{E} \left[ \left| \tau_N [k_N^{(1,2)}(\Lambda, \Gamma)_{u,v}{}_{u',v'}] \right|^2 \right] \right)^{1/2} \\
&\leq k^{5/2} \sum_{j=1}^p \|a_j\| \left( \sum_{u,v=1}^k \text{Var}(H_N^{(1)}(\Lambda))_{u,v} \max_{u,v,u',v'} \text{Var} \left( \tau_N [h_N^{(1,2)}(\Lambda, \Gamma)_{u,v}{}_{u',v'}] \right) \right)^{1/2}. \quad (\text{A.5})
\end{aligned}$$

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

$$\begin{aligned}
\left( H_N^{(1)}(\Lambda) \right)_{u,v} &= F_{u,v}^{(1)}(X_1^{(N)}, \dots, X_p^{(N)}, Y_1^{(N)}, \dots, Y_q^{(N)}), \\
\tau_N \left[ h_N^{(1,2)}(\Lambda, \Gamma)_{u,v}{}_{u',v'} \right] &= F_{u,v,u',v'}^{(2)}(X_1^{(N)}, \dots, X_p^{(N)}, Y_1^{(N)}, \dots, Y_q^{(N)}),
\end{aligned}$$

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

$$\begin{aligned}
F_{u,v}^{(1)}(\mathbf{A}, \mathbf{B}) &= \left( (\text{id}_k \otimes \tau_N) [(\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1}] \right)_{u,v} \\
&= \frac{1}{N} (\text{Tr}_k \otimes \text{Tr}_N) [(\epsilon_{v,u} \otimes \mathbf{1}_N) (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1}],
\end{aligned}$$

$$\begin{aligned}
F_{u,v,u',v'}^{(2)}(\mathbf{A}, \mathbf{B}) &= \tau_N \left[ \left( (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1} \right)_{(u,u')} \left( (\Gamma \otimes \mathbf{1}_N - \tilde{T}_N)^{-1} \right)_{(v',v)} \right] \\
&= \frac{1}{N} (\text{Tr}_k \otimes \text{Tr}_N) \left[ (\epsilon_{v,u} \otimes \mathbf{1}_N) (\Gamma \otimes \mathbf{1}_N - \tilde{T}_N)^{-1} (\epsilon_{u',v'} \otimes \mathbf{1}_N) (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1} \right].
\end{aligned}$$

The functions and their partial derivatives are bounded (see [15, Lemma 4.6] with minor modifications), so that, by the Poincaré inequality with constant  $\frac{\sigma}{N}$  satisfied by the law of  $(\mathbf{X}_N, \mathbf{Y}_N)$

$$\begin{aligned} \text{Var} \left( H_N^{(1)}(\Lambda) \right)_{u,v} &\leq \frac{\sigma}{N} \mathbb{E} \left[ \left\| \nabla F_{u,v}^{(1)}(X_1^{(N)}, \dots, X_p^{(N)}, Y_1^{(N)}, \dots, Y_q^{(N)}) \right\|^2 \right], \\ \text{Var} \left( \tau_N [h_N^{(1,2)}(\Lambda, \Gamma)_{u,v}] \right)_{u',v'} &\leq \frac{\sigma}{N} \mathbb{E} \left[ \left\| \nabla F_{u,v,u',v'}^{(2)}(X_1^{(N)}, \dots, X_p^{(N)}, Y_1^{(N)}, \dots, Y_q^{(N)}) \right\|^2 \right]. \end{aligned}$$

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

$$\begin{aligned} \text{Var} \left( H_N^{(1)}(\Lambda) \right)_{u,v} &\leq \frac{\sigma}{N} \mathbb{E} \left[ \max_{(\mathbf{V}, \mathbf{W}) \in \mathcal{W}} \left| \frac{d}{dt} \Big|_{t=0} F_{u,v}^{(1)}(\mathbf{X}_N + t\mathbf{V}, \mathbf{Y}_N + t\mathbf{W}) \right|^2 \right], \\ \text{Var} \left( \tau_N [h_N^{(1,2)}(\Lambda, \Gamma)_{u,v}] \right)_{u',v'} &\leq \frac{\sigma}{N} \mathbb{E} \left[ \max_{(\mathbf{V}, \mathbf{W}) \in \mathcal{W}} \left| \frac{d}{dt} \Big|_{t=0} F_{u,v,u',v'}^{(2)}(\mathbf{X}_N + t\mathbf{V}, \mathbf{Y}_N + t\mathbf{W}) \right|^2 \right]. \end{aligned}$$

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

$$\begin{aligned} &\left| \frac{d}{dt} \Big|_{t=0} F_{u,v}^{(1)}(\mathbf{A} + t\mathbf{V}, \mathbf{B} + t\mathbf{W}) \right|^2 \\ &= \left| \frac{d}{dt} \Big|_{t=0} \frac{1}{N} (\text{Tr}_k \otimes \text{Tr}_N) \left[ (\epsilon_{v,u} \otimes \mathbf{1}_N) \left( \Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - \sum_{j=1}^p a_j \otimes (A_j + tV_j) \right. \right. \right. \\ &\quad \left. \left. \left. - \sum_{j=1}^q b_j \otimes (B_j + tW_j) + b_j \otimes (B_j^* + tW_j^*) \right)^{-1} \right] \right|^2 \\ &= \left| \frac{1}{N} (\text{Tr}_k \otimes \text{Tr}_N) \left[ (\epsilon_{v,u} \otimes \mathbf{1}_N) (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1} \right. \right. \\ &\quad \left. \left. \times \left( \sum_{j=1}^p a_j \otimes V_j + \sum_{j=1}^q b_j \otimes W_j + b_j^* \otimes W_j^* \right) (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1} \right] \right|^2. \end{aligned}$$

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

$$\begin{aligned} &\left| \frac{d}{dt} \Big|_{t=0} F_{u,v}^{(1)}(\mathbf{A} + t\mathbf{V}, \mathbf{B} + t\mathbf{W}) \right|^2 \\ &\leq \frac{1}{N^2} \left\| (\epsilon_{v,u} \otimes \mathbf{1}_N) (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1} \right\|_e^2 \\ &\quad \times \left\| \left( \sum_{j=1}^p a_j \otimes V_j + \sum_{j=1}^q b_j \otimes W_j + b_j^* \otimes W_j^* \right) (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1} \right\|_e^2 \end{aligned}$$

Using (A.3) to split Euclidean norms into the product of an operator norm and an Euclidean norm, we get:

$$\begin{aligned} &\left| \frac{d}{dt} \Big|_{t=0} F_{u,v}^{(1)}(\mathbf{A} + t\mathbf{V}, \mathbf{B} + t\mathbf{W}) \right|^2 \\ &\leq \frac{1}{N^2} \|\epsilon_{v,u} \otimes \mathbf{1}_N\|_e^2 \left\| (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1} \right\|^2 \left\| \sum_{j=1}^p a_j \otimes V_j + \sum_{j=1}^q b_j \otimes W_j + b_j^* \otimes W_j^* \right\|_e^2 \\ &\leq \frac{k}{N} \|(\text{Im } \Lambda)^{-1}\|^4 \left\| \sum_{j=1}^p a_j \otimes V_j + \sum_{j=1}^q b_j \otimes W_j + b_j^* \otimes W_j^* \right\|_e^2 \end{aligned}$$

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

$$\left\| \sum_{j=1}^p a_j \otimes V_j + \sum_{j=1}^q b_j \otimes W_j + b_j^* \otimes W_j^* \right\|_e \leq \sum_{j=1}^p \|a_j\|_e + 2 \sum_{j=1}^q \|b_j\|_e \leq \sqrt{k} \left( \sum_{j=1}^p \|a_j\| + \sum_{j=1}^q 2\|b_j\| \right).$$



Hence we finally obtain an estimate of  $\text{Var}(H_N^{(1)}(\Lambda))_{u,v}$

$$\text{Var}\left(H_N^{(1)}(\Lambda)\right)_{u,v} \leq \frac{k^2 \sigma}{N^2} \left( \sum_{j=1}^p \|a_j\| + 2 \sum_{j=1}^q \|b_j\| \right)^2 \|(\text{Im } \Lambda)^{-1}\|^4. \quad (\text{A.6})$$

We obtain a similar estimate for  $\text{Var}\left(\tau_N[h_N^{(1,2)}(\Lambda, \Gamma)_{u,v}]\right)$ . The partial derivative of  $F_{u,v,u',v'}^{(2)}$  gives in this case two terms:  $\forall(\mathbf{V}, \mathbf{W}) \in \mathcal{W}$ ,  $\forall(\mathbf{A}, \mathbf{B}) \in \text{M}_N(\mathbb{C})^{p+q}$

$$\begin{aligned} & \frac{d}{dt} \Big|_{t=0} F_{u,v,u',v'}^{(2)}(\mathbf{A} + t\mathbf{V}, \mathbf{B} + t\mathbf{W}) \\ &= \frac{1}{N} (\text{Tr}_k \otimes \text{Tr}_N) \left[ (\epsilon_{v,u} \otimes \mathbf{1}_N) (\Gamma \otimes \mathbf{1}_N - \tilde{T}_N)^{-1} \left( \sum_{j=1}^q b_j \otimes W_j + b_j^* \otimes W_j^* \right) \right. \\ & \quad \times (\Gamma \otimes \mathbf{1}_N - \tilde{T}_N)^{-1} (\epsilon_{u',v'} \otimes \mathbf{1}_N) (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1} \\ & \quad + (\epsilon_{v,u} \otimes \mathbf{1}_N) (\Gamma \otimes \mathbf{1}_N - \tilde{T}_N)^{-1} (\epsilon_{u',v'} \otimes \mathbf{1}_N) (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1} \\ & \quad \left. \times \left( \sum_{j=1}^p a_j \otimes V_j^{(N)} + \sum_{j=1}^q b_j \otimes W_j^{(N)} + b_j^* \otimes W_j^{(N)*} \right) (\Lambda \otimes \mathbf{1}_N - a_0 \otimes \mathbf{1}_N - \tilde{S}_N - \tilde{T}_N)^{-1} \right]. \end{aligned}$$

We then get the following

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

Hence we have

$$\begin{aligned} & \text{Var}\left(\tau_N[h_N^{(1,2)}(\Lambda, \Gamma)_{u,v}]\right) \\ & \leq \frac{k^2 \sigma}{N^2} \left( \sum_{j=1}^p \|a_j\| + 2 \sum_{j=1}^q \|b_j\| \right)^2 \|(\text{Im } \Gamma)^{-1}\|^2 \|(\text{Im } \Lambda)^{-1}\|^2 \left( \|(\text{Im } \Gamma)^{-1}\| + \|(\text{Im } \Lambda)^{-1}\| \right)^2 \quad (\text{A.7}) \end{aligned}$$

We then obtain as desired, by (A.5), (A.6) and (A.7):

$$\|\Theta_N(\Lambda, \Gamma)\| \leq c_1 \|(\text{Im } \Gamma)^{-1}\| \|(\text{Im } \Lambda)^{-1}\|^3 \left( \|(\text{Im } \Gamma)^{-1}\| + \|(\text{Im } \Lambda)^{-1}\| \right),$$

where  $c_1 = \frac{k^{9/2} \sigma}{N^2} \sum_{j=1}^p \|a_j\| \left( \sum_{j=1}^p \|a_j\| + 2 \sum_{j=1}^q \|b_j\| \right)^2$ .

## B Proof of Step 2: from Stieltjes transforms to spectra

For  $\lambda$  in  $\mathbb{C}$  with  $\text{Im } \lambda > 0$ , define

$$g_n(\lambda) = \mathbb{E} \left[ (\tau_k \otimes \tau_N) \left[ (\lambda \mathbf{1}_{kN} - L(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*))^{-1} \right] \right] = \tau_k \left[ G_N^{(1)}(\lambda \mathbf{1}_k) \right], \quad (\text{B.1})$$

$$g(\lambda) = (\tau_k \otimes \tau) \left[ (\lambda \mathbf{1}_k \otimes \mathbf{1} - L(\mathbf{x}, \mathbf{y}, \mathbf{y}^*))^{-1} \right] = \tau_k \left[ G^{(1)}(\lambda \mathbf{1}_k) \right]. \quad (\text{B.2})$$

Then, by (3.12), for every  $\varepsilon, \tilde{\varepsilon} > 0$ , for all  $\lambda$  in  $\mathbb{C}$  with  $\text{Im } \lambda > 0$  such that  $\varepsilon \leq (\text{Im } \lambda)^{-1} \leq N^\gamma$  and  $|\text{Re } \lambda| \leq \tilde{\varepsilon}$ , we have for  $N$  large enough

$$|g(\lambda) - g_N(\lambda)| \leq \frac{c}{N^{1+\eta}} \lambda^{-\alpha}, \quad (\text{B.3})$$

where  $c, \eta, \alpha$  are positive constants.

In the following, for any smooth function  $f : \mathbb{R} \mapsto \mathbb{R}$ , we will denote the real random variable  $D_N(f) = (\tau_k \otimes \tau_N) \left[ f(L(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)) \right]$ .

**Lemma 14.** For every smooth function  $f : \mathbb{R} \mapsto \mathbb{R}$  non negative, compactly supported and vanishing on a neighborhood of the spectrum of  $L(\mathbf{x}, \mathbf{y}, \mathbf{y}^*)$ , there exists a constant, still denoted  $c$ , such that for all  $N$

$$\left| \mathbb{E}[D_N(f)] \right| \leq \frac{c}{N^{1+\eta}}. \quad (\text{B.4})$$

*Proof.* see [1, Lemma 5.5.5] with minor modifications.  $\square$

**Lemma 15.** With  $f$  as above, there exists  $\kappa > 0$  such that

$$N^{1+\kappa} D_N(f) \xrightarrow{N \rightarrow \infty} 0, \text{ a.s.} \quad (\text{B.5})$$

*Proof.* The law of the random matrices satisfying a Poincaré's inequality with constant  $\frac{\sigma}{N}$  and  $L$  being a polynomial of degree one, for all Lipschitz function  $\Psi : M_{kN}(\mathbb{C}) \mapsto \mathbb{R}$  [12, Lemma 5.2]:

$$\mathbb{P}\left( \left| \Psi(L(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)) - \mathbb{E}\left[ \Psi(L(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)) \right] \right| \geq \delta \right) \leq K_1 e^{-K_2 \frac{\sqrt{N}\delta}{|\Psi|_{\mathcal{L}}}}, \quad (\text{B.6})$$

where  $K_1, K_2$  are positive constants and  $|\Psi|_{\mathcal{L}} = \sup_{A \neq B \in M_{kN}(\mathbb{C})} \frac{|\Psi(A) - \Psi(B)|}{\|A - B\|_2}$ .

For any  $f$  as in the lemma, we define  $\Phi_N^{(f)} : M_{kN}(\mathbb{C}) \mapsto \mathbb{R}$  by

$$\Phi_N^{(f)}(A) = (\tau_k \otimes \tau_N)[f(A)] \quad (\text{B.7})$$

and denote for  $N$  and  $0 < \kappa < \frac{\eta}{4}$

$$\mathcal{B}_{N,\kappa} = \left\{ A \in M_{kN}(\mathbb{C}) \mid \Phi_N^{(f'^2)}(A) \leq \frac{1}{N^{4\kappa}} \right\}. \quad (\text{B.8})$$

Define  $\Psi_N^{(f)} : M_{kN}(\mathbb{C}) \mapsto \mathbb{R}$  by:  $\forall A \in M_N(\mathbb{C})$

$$\Psi_N^{(f)}(A) = \max \left( 0, \sup_{B \in \mathcal{B}_{N,\kappa}} \left\{ \Phi_N^{(f)}(A) - |\Phi_N^{(f)}|_{\mathcal{L}} \|A - B\|_2 \right\} \right), \quad (\text{B.9})$$

and denote  $\tilde{D}_N(f) = \Psi_N^{(f)}(L(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*))$ . Then  $\Psi_N^{(f)}$  is Lipschitz, with  $|\Psi_N^{(f)}|_{\mathcal{L}} \leq |(\Phi_N^{(f)})|_{\mathcal{B}_{N,\kappa}}|_{\mathcal{L}}$ . But if  $f$  is a polynomial it is easy to see that  $\Phi_N^{(f)}$  is differentiable and

$$|(\Phi_N^{(f)})|_{\mathcal{B}_{N,\kappa}}|_{\mathcal{L}} \leq \frac{1}{\sqrt{kN}} \|(\Phi_N^{(f'^2)})|_{\mathcal{B}_{N,\kappa}}\|_{\infty}^{1/2}, \quad (\text{B.10})$$

and then it is the case when  $f$  is not a polynomial by density of polynomials. Hence  $|\Psi_N^{(f)}|_{\mathcal{L}} \leq \frac{1}{\sqrt{k}} N^{-1/2-2\kappa}$ . Moreover, by the Tchebychev inequality

$$\mathbb{P}(\mathcal{B}_{N,\kappa}^c) = \mathbb{P}\left(D_N(f'^2) \geq \frac{1}{N^{4\kappa}}\right) \leq N^{4\kappa} \mathbb{E}\left[D_N(f'^2)\right] \leq \frac{\sigma}{N^{1+\eta-4\kappa}}, \quad (\text{B.11})$$

where we used (B.4) and the fact that  $f'^2$  also vanishes in a neighborhood of the spectrum of  $L(\mathbf{x}, \mathbf{y}, \mathbf{y}^*)$ ; at last, since  $\Psi_N^{(j)}$  and  $\Phi_N^{(j)}$  are equals in  $\mathcal{B}_{N,\kappa}$  and  $\|\Psi_N^{(j)}\|_{\infty} \leq \|\Phi_N^{(j)}\|_{\infty}$ ,

$$\left| \mathbb{E}[\tilde{D}_N(f) - D_N(f)] \right| \leq \|\Phi_N^{(j)}\|_{\infty} \mathbb{P}(\mathcal{B}_{N,\kappa}^c) \leq \|\Phi_N^{(j)}\|_{\infty} \frac{\sigma}{N^{1+\eta-4\kappa}} \quad (\text{B.12})$$

Now, by (B.6) applied to  $\Psi_N^{(j)}$ :

$$\begin{aligned} & \mathbb{P}\left( \left| D_N(f) - \mathbb{E}[D_N(f)] \right| > \frac{\delta}{N^{1+\kappa}} \cap \mathcal{B}_{N,\kappa} \right) \\ & \leq \mathbb{P}\left( \left| \tilde{D}_N(f) - \mathbb{E}[\tilde{D}_N(f)] \right| > \frac{\delta}{N^{1+\kappa}} - \left| \mathbb{E}[\tilde{D}_N(f) - D_N(f)] \right| \right) \\ & \leq K_1 \exp\left( -\sqrt{k} K_2 N^{\kappa} (\delta - \left| \mathbb{E}[\tilde{D}_N(f) - D_N(f)] \right|) \right) \end{aligned}$$

By (B.11), (B.12), Lemma 14 and the Borel-Cantelli lemma,  $D_N(f)$  is almost surely of order  $N^{1+\kappa}$  at most.  $\square$

**Proposition 6.** For every  $\varepsilon > 0$  there exists  $N_0$  such that for  $N \geq N_0$

$$\mathrm{Sp}\left(L(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*)\right) \subset \mathrm{Sp}\left(L(\mathbf{x}, \mathbf{y}, \mathbf{y}^*)\right) + (-\varepsilon, \varepsilon) \quad (\text{B.13})$$

*Proof.* By (3.1) and [1, Exercise 2.1.27], there exists  $N_0 \in \mathbb{N}$  and  $D \geq 0$  such that the spectral radii of the matrices  $(\mathbf{X}_N, \mathbf{Y}_N)$  is bounded by  $D$  for all  $N \geq N_0$  almost surely, and then there exists a  $M \geq 0$  such that almost surely

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

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

$$(\tau_k \otimes \tau_N)\left[f(L(\mathbf{X}_N, \mathbf{Y}_N, \mathbf{Y}_N^*))\right] \geq N^{-1} \geq N^{-1-\kappa}$$

in contradiction with Lemma 15. □

## 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] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] P. Biane. Processes with free increments. *Math. Z.*, 227(1):143–174, 1998.
- [7] M. Capitaine and C. Donati-Martin. Strong asymptotic freeness for Wigner and Wishart matrices. *Indiana Univ. Math. J.*, 56(2):767–803, 2007.
- [8] D. Chafai. Inégalités de Poincaré et de Gross pour les mesures de Bernoulli, de Poisson, et de Gauss. <http://hal.archives-ouvertes.fr/ccsd-00012428>.
- [9] L. H. Y. Chen. An inequality for the multivariate normal distribution. *J. Multivariate Anal.*, 12(2):306–315, 1982.
- [10] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.
- [11] S. Geman. A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2):252–261, 1980.
- [12] 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.
- [13] A. Guionnet, M. Krishnapur, and O. Zeitouni. The single-ring theorem. arXiv:0909.2214v1, preprint, <http://arxiv4.library.cornell.edu/abs/0909.2214>.
- [14] U. Haagerup, H. Schultz, and S. Thorbjørnsen. A random matrix approach to the lack of projections in  $C_{\mathrm{red}}^*(\mathbb{F}_2)$ . *Adv. Math.*, 204(1):1–83, 2006.

- [15] U. Haagerup and S. Thorbjørnsen. A new application of random matrices:  $\text{Ext}(C_{\text{red}}^*(\mathbb{F}_2))$  is not a group. *Ann. of Math. (2)*, 162(2):711–775, 2005.
- [16] 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.
- [17] 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.
- [18] E. G. Larsson and P. Stoica. *Space-time block coding for wireless communications*. Cambridge University Press, Cambridge, 2003.
- [19] F. Lehner. Computing norms of free operators with matrix coefficients. *Amer. J. Math.*, 121(3):453–486, 1999.
- [20] A. Lytova and L. Pastur. Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.*, 37(5):1778–1840, 2009.
- [21] 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.
- [22] H. Schultz. Non-commutative polynomials of independent Gaussian random matrices. The real and symplectic cases. *Probab. Theory Related Fields*, 131(2):261–309, 2005.
- [23] 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.
- [24] R. Speicher. Free probability theory, contribution for handbook on random matrix theory. to be published by Oxford University Press, <http://arxiv.org/abs/0911.0087>.
- [25] A. M. Tulino and S. Verdú. *Random matrices and wireless communications*, volume vol. 1 of *Foundations and Trends in Communications and Information Theory*. Now Publishers Inc., 2004.
- [26] D. Voiculescu. Free probability theory: random matrices and von Neumann algebras. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)*, pages 227–241, Basel, 1995. Birkhäuser.
- [27] D. Voiculescu. The analogues of entropy and of Fisher’s information measure in free probability theory. VI. Liberation and mutual free information. *Adv. Math.*, 146(2):101–166, 1999.
- [28] 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.