Free Convolution and the Random
Sum of Matrices

By

Roland SPEICHER*

Abstract

We show that the spectral measure of the sum of two selfadjoint matrices is 'almost surely' given by the free convolution (in the sense of Voiculescu) of the spectral measures of the two matrices if their size tends to infinity.

§1. Introduction

Assume we know the spectral measures μ_A and μ_B of two $n \times n$-matrices A and B. What can we say about the spectral measure of the sum of A and B? One way to answer such a question is to average over all possibilities which are compatible with the given information.

If we knew that the two matrices commute, then we would average over all possibilities for choosing two diagonal matrices with the given spectral measures. The result would then be the convolution of the measures μ_A and μ_B.

But here we want to consider the case where we have no additional information (like commutativity) about the relation between A and B. Then the canonical procedure is to choose two fixed (e. g., diagonal) matrices A and B with the given spectral measures and average over all spectral measures of $A + UBU^*$ for all unitary matrices U, where the averaging measure for U is the Haar measure on the group $U(n)$. The question is, how the resulting average measure is connected with the starting measures μ_A and μ_B. For finite n the connection seems to be very complicated, but in the limit where n tends to ∞ it gets handable again. Indeed, it was shown by Voiculescu [Voi 3] that in this limit the averaging procedure gives a measure which is the free convolution of μ_A and μ_B. This free convolution (and the corresponding notion of 'free independence') has a nice algebraic characterization and was introduced

by Voiculescu [Voi 1] in a totally different context. He also recognized that this new kind of convolution can be treated very analogously to the usual convolution and he developed an analytic aparatus for free convolution comparable to the usual Fourier transformation machinery for the usual convolution ([Voi 2], see also [Maa, Spe]). To sum it up, free convolution can be defined and handled easily without any reference to averaging procedures for matrices.

The aim of this paper is to improve Voiculescu's result in the respect, that the above statement is not only true for an average over all possible sums, but also for almost all possibilities itself. Thus we shall prove (informally spoken): If A and B are two selfadjoint infinite matrices with spectral measures μ_A and μ_B, then the spectral measure of a typical sum $A+UBU^*$ is for almost all unitary U given by the free convolution of μ_A and μ_B. Of course, the notion of infinite matrices and 'almost all' has to be given a precise meaning by approximating A, B and U by finite-dimensional $n \times n$-matrices A_n, B_n and U_n.

The problem of a direct, more appealing, formulation in terms of operators on an infinite-dimensional Hilbert space $\mathcal{H}$ suffers from the absence of a canonical measure on the group of all unitary operators on $\mathcal{H}$.

For technical reasons, we shall only deal with the case where μ_A and μ_B have compact support, because we can then restrict to the consideration of moments. However, we believe that the result is also true in the non-compact case, see [Maa] for an elegant treatment of free convolution in this case.

We hope that our result supports our philosophy that it is not too bad to model two (or more) operators, the relation among which is obscured to us because of initial lack of knowledge or because of a complex dynamics, by operators which are freely independent.

The paper is organized as follows. In the next chapter we shall recall the definition and some basic facts about free independence and free convolution and fix our notations for later use. The third chapter contains the formulation and the proof of our main theorem.

§2. Some Facts about Free Independence and Fixing of Notations

The following definitions of 'free independence' and 'free convolution' are essentially taken from [Voi3].

Let C be a unital $*$-algebra and φ a state on C, i.e. a positive linear functional with $\varphi(1)=1$. A family of subalgebras $C_i \subset C$ $(i \in I)$ with $1 \in C_i$ is called a *free* family of subalgebras (or C_i are called *freely independent*) in (C, φ) if $\varphi(a_1 \cdots a_n)=0$ whenever we have $a_1, \cdots, a_n$ such that $a_j \in C_{i(j)}$ with $i(j) \neq i(j+1)$ for all $1 \leq j \leq n-1$ and $\varphi(a_k)=0$ for all $1 \leq k \leq n$.

Given pairs (C_k, φ_k) $(k \in J)$ of unital $*$-algebras C_k and states φ_k on C_k,

there exists a state φ on the free product algebra $C := \underset{k \in J}{\star} C_k$ (with identification of the units) such that $\varphi \mid C_k = \varphi_k$ and such that the subalgebras $(C_k)_{k \in J}$ form a free family of subalgebras in (C, φ). The state φ is uniquely determined by these conditions and is called the *free product* of the states φ_k.

Note that the above characterization of free independence allows to express all moments $\varphi(b_1 \cdots b_n)$ whenever $b_j \in C_{i(j)}$ as polynomials in moments of the form $\varphi(b_{k_1} \cdots b_{k_r})$ with $i(k_1) = i(k_2) = \cdots = i(k_r)$, i.e. $b_{k_1} \cdots b_{k_r} \in C_{i(k_1)}$. In particular, we have stochastic independence between the different subalgebras in the sense $\varphi(a_1 a_2) = \varphi(a_1) \varphi(a_2)$ whenever $a_1 \in C_{i(1)}$, $a_2 \in C_{i(2)}$ and $i(1) \neq i(2)$. But note also that the polynomial can be quite complicated for other moments, e. g. for $a \in C_1$, $b \in C_2$ we have

$$\varphi(abab) = \varphi(a)\varphi(a)\varphi(bb) + \varphi(aa)\varphi(b)\varphi(b) - \varphi(a)\varphi(a)\varphi(b)\varphi(b).$$

By $C\langle X_1, \cdots, X_s \rangle$ we denote the free algebra with the s generators $X_1, \cdots, X_s$, i.e. all polynomials in s non-commuting indeterminants. Let now two probability measures μ_1 and μ_2 on $\boldsymbol{R}$ (with compact support) be given. Then μ_i can be regarded as the distribution of the random variable $X_i = X_i^*$ on $(C\langle X_i \rangle, \varphi_i)$, i.e. φ_i is determined by

$$\varphi_i(X_i^k) = \int x^k d\mu_i(x) \qquad \text{for all } k \in \boldsymbol{N}_0.$$

Now consider the free product $\varphi_1 \star \varphi_2$ on $C\langle X_1 \rangle \star C\langle X_2 \rangle = C\langle X_1, X_2 \rangle$. The random variables X_1 and X_2 (i.e. the $*$-algebras generated by them) are freely independent in $(C\langle X_1, X_2 \rangle, \varphi_1 \star \varphi_2)$ and we denote by $\mu_1 \boxplus \mu_2$ the distribution of $X_1 + X_2$ with respect to $\varphi_1 \star \varphi_2$, i.e. $\mu_1 \boxplus \mu_2$ is determined by

$$\int x^k d(\mu_1 \boxplus \mu_2)(x) = \varphi_1 \star \varphi_2((X_1 + X_2)^k) \qquad \text{for all } k \in \boldsymbol{N}_0.$$

The probability measure $\mu_1 \boxplus \mu_2$ is called the *free convolution* of μ_1 and μ_2. By the above remark, the moments of $\mu_1 \boxplus \mu_2$ are polynomials in the moments of μ_1 and μ_2. It is clear that $\mu_1 \boxplus \mu_2$ also has compact support. Free convolution can also be defined without problems when the support of μ_1 or μ_2 is not compact, cf. [Maa].

Free independence and free convolution behave in some respects quite analogous to the usual independence and convolution (cf. [Voi1, 2, Spe]), but one should note that the free product is a completely non-commutative concept. An unusual feature of free convolution is the fact that $\boxplus$ is not distributive, i.e. in general $\mu \boxplus (\lambda \nu_1 + (1-\lambda)\nu_2) \neq \lambda \mu \boxplus \nu_1 + (1-\lambda)\mu \boxplus \nu_2$.

In the rest of this chapter we shall fix our notations for the following.

We shall use the following probability spaces: Let Ω_n be the probability space of all $n \times n$-matrices, where all entries are independent, complex-valued, Gaussian distributed random variables with mean zero and variance $1/n$. By

Ω_{gauss} we denote the product space of all Ω_n (equipped with the canonical product measure), i. e.

$$\Omega_{\mathrm{gauss}} := \bigotimes_{n=1}^{\infty} \Omega_n .$$

In a similar way, we denote by $\Omega_{\mathrm{unitary}}$ the product space of random sequences of unitary matrices

$$\Omega_{\mathrm{unitary}} := \bigotimes_{n=1}^{\infty} \Omega_n^{\mathrm{unitary}} ,$$

where $\Omega_n^{\mathrm{unitary}}$ is the group $U(n)$ of unitary $n \times n$-matrices equipped with the normalized Haar measure.

If μ_i ($i \in \mathbf{N}$) and μ are probability measures on $\mathbf{R}$, then we denote by $\mu_n \xrightarrow{w} \mu$ the weak convergence of μ_i to μ. We recall that in the case where μ has compact support it is determined by its moments and it is sufficient to check the convergence of all moments (cf. [Bil, Theorem 30.2]).

By M_n we denote the complex $*$-algebra of $n \times n$-matrices, equipped with the normalized trace tr_n given by

$$\mathrm{tr}_n(A) := \frac{1}{n} \sum_{i=1}^{n} a_{ii} \qquad \text{for } A = (a_{jk})_{j,k=1,\cdots,n} \in M_n .$$

The adjoint of a matrix A is denoted by A^*. By $|A| := (A^*A)^{1/2}$ we denote the absolute value of $A \in M_n$. The expression $\prod_{i=1}^{n} A_i$ means the product $A_1 \cdots A_n$ in the 'right' order.

For some selfadjoint $R \in M_n$ we denote by μ_R the spectral measure of R with respect to the trace tr_n, i. e. μ_R is determined by

$$\mathrm{tr}_n(R^k) = \int x^k d\mu_R(x) \qquad \text{for all } k \in \mathbf{N}_0 .$$

In connection with central limit theorems for the free independence it was seen [Voil, Spe] that the limiting states can be described as free Gaussian states on some $C\langle X_1, \cdots, X_s \rangle$ in the following way: Let $\mathcal{V} = \{V_1, \cdots, V_p\}$ be a partition of the set $\{1, \cdots, r\}$, i. e. the V_i are ordered and disjoint sets and $\{1, \cdots, r\} = \bigcup_{i=1}^{p} V_i$. Then $\mathcal{V}$ is called *non crossing* [Kre], if for all $i, j = 1, \cdots, p$ with $V_i = (v_1, \cdots, v_n)$ ($v_1 < \cdots < v_n$) and $V_j = (w_1, \cdots, w_m)$ ($w_1 < \cdots < w_m$) we have

$$w_k < v_1 < w_{k+1} \Longleftrightarrow w_k < v_n < w_{k+1} \qquad (k = 1, \cdots, m-1) .$$

(In [Spe] we used the term 'admissible' instead of 'non crossing'.) We will denote the set of all partitions of $\{1, \cdots, r\}$ by $\mathcal{P}(1, \cdots, r)$ and the set of all non crossing partitions by $\mathcal{P}_{NC}(1, \cdots, r)$. We can reformulate the definition of 'non crossing' in a recursive way: The partition $\mathcal{V} = \{V_1, \cdots, V_p\}$ is non crossing if at least one of the V_i is a segment of $(1, \cdots, r)$, i. e. it has the form $V_i = (k, k+1, k+2, \cdots, k+m)$ and $\{V_1, \cdots, V_{i-1}, V_{i+1}, \cdots, V_p\}$ is a non

crossing partition of $\{1, \cdots, r\} \setminus V_i$ (interpreted in a canonical way). We shall especially have to consider non crossing partitions $\mathcal{V} = \{V_1, \cdots, V_p\}$ where each V_i contains exactly two elements. We shall write $V_i = (e_i, z_i)$ with $e_i < z_i$. The set of all such partitions will be denoted by $\mathscr{P}_{NC}^2(1, \cdots, r)$, where of course r has to be of the form $r = 2m$.

A state φ on $C\langle X_1, \cdots, X_s \rangle$ is called a *free Gaussian state*, if we have for all $r \in N$ and $k(1), \cdots, k(r) \in \{1, \cdots, s\}$

$$\varphi(X_{k(1)} \cdots X_{k(r)})$$

$$= \begin{cases} 0, & r \text{ odd} \\ \displaystyle\sum_{\substack{V = \{(e_1, z_1), \cdots, (e_m, z_m)\} \\ \in \mathscr{P}_{NC}^2(1, \cdots, 2m)}} \varphi(X_{k(e_1)} X_{k(z_1)}) \cdots \varphi(X_{k(e_m)} X_{k(z_m)}), & r = 2m. \end{cases}$$

The matrix $(\varphi(X_i X_j))_{i,j=1,\cdots,s}$ is called the *covariance matrix* of φ.

§3. Statement and Proof of the Main Theorem

Theorem. *Let* $\mathcal{A} = (A_n)_{n \in N}$ *and* $\mathcal{B} = (B_n)_{n \in N}$ *be two sequences of selfadjoint* $n \times n$-*matrices* A_n *and* B_n *such that*

$$\mu_{A_n} \xrightarrow{w} \mu_{\mathcal{A}} \quad \text{and} \quad \mu_{B_n} \xrightarrow{w} \mu_{\mathcal{B}}$$

for some probability mesures $\mu_{\mathcal{A}}$ *and* $\mu_{\mathcal{B}}$ *on* $\mathbb{R}$. *If* $\mu_{\mathcal{A}}$ *and* $\mu_{\mathcal{B}}$ *have compact support, then*

$$\mu_{A_n + U_n B_n U_n^*} \xrightarrow{w} \mu_{\mathcal{A}} \boxplus \mu_{\mathcal{B}}$$

for almost all random sequences of unitary matrices $\mathcal{U} = (U_n)_{n \in N} \in \Omega_{\text{unitary}}$.

The rough idea of the proof is the same as in [Voi3], we only have to adapt some arguments to the 'almost sure' situation.

The idea is the following: One first shows that infinitely large random matrices $\mathcal{X} \in \Omega_{\text{gauss}}$ and non-random diagonal matrices are freely independent, which implies also the free independence between polynomials in $\mathcal{X}$ and diagonal matrices. Then one notices that the phase $\mathcal{U} := \mathcal{X}(\mathcal{X}^* \mathcal{X})^{-1/2}$ of a random matrix $\mathcal{X} \in \Omega_{\text{gauss}}$ gives an element $\mathcal{U} \in \Omega_{\text{unitary}}$ (under respectation of the measure) and tries to approximate $\mathcal{U}$ by polynomials in $\mathcal{X}$. In this way one also gets the free independence between random unitary matrices $\mathcal{U} \in \Omega_{\text{unitary}}$ and diagonal matrices. But free independence between $\{\mathcal{A}, \mathcal{B}\}$ and $\{\mathcal{U}, \mathcal{U}^*\}$ implies free independence between $\mathcal{A}$ and $\mathcal{U} \mathcal{B} \mathcal{U}^*$.

The crucial observation made by Voiculescu is that not only $n \times n$-matrices with independent entries become freely independent in the limit $n \to \infty$, but that one can also include non-random diagonal matrices in this argument. This

allows to start in our theorem with arbitrary spectral measures and not only with free Gaussian measures. To say it in another way, large matrices do not only give an interpretation of Wigner's semi-circle distribution, but much more generally for the concept of 'free independence'.

We shall now start with proving the free independence between random matrices in Ω_{gauss} and non-random diagonal matrices.

Lemma. *For almost all $\mathfrak{X}=(X_n)_{n\in N}\in\Omega_{\text{gauss}}$ we have*:
i) *Let P be a polynomial in two non-commuting indeterminants. Then*

$$\lim_{n\to\infty}\text{tr}_n(P(X_n, X_n^*))=\varphi(P(X, X^*)),$$

where φ is a free Gaussian state on $C\langle X, X^*\rangle$ with the covariance matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

ii) *Let $\mathscr{D}^1=(D_n^1)_{n\in N}, \cdots, \mathscr{D}^s=(D_n^s)_{n\in N}$ be s sequences of diagonal $n\times n$-matrices D_n^j such that*

$$\lim_{n\to\infty}\text{tr}_n(P(D_n^1, \cdots, D_n^s))=\rho(P(D^1, \cdots, D^s))$$

for all polynomials P in s indeterminants, where ρ is a state on $C\langle D^1, \cdots, D^s\rangle$. Then

$$\lim_{n\to\infty}\text{tr}_n(P(X_n, D_n^1, \cdots, D_n^s, X_n^*))=\rho \star \varphi(P(X, D^1, \cdots, D^s, X^*))$$

for all polynomials P in $(s+2)$ non-commuting indeterminants, where $\rho \star \varphi$ is the free product of ρ and φ on $C\langle X, D^1, \cdots, D^s, X^*\rangle$.

Remarks.
1) Clearly, i) is contained in ii). The main aim of the first part is to introduce the state φ.
2) One should note that φ restricted to $C\langle X^*X\rangle$ gives the 'square' of Wigner's semi-circle distribution (cf. [Voil, 3, Spe]), i.e.

$$\lim_{n\to\infty}\text{tr}_n((X_n^*X_n)^k)=\frac{1}{\pi}\int_{-2}^{+2}(x^2)^k \sqrt{1-(x/2)^2}dx.$$

Proof of Lemma. It is sufficient to consider the monomials $X_n^{i(1)} \cdots X_n^{i(r)}$ for all choices of $r\in N$ and $i(k)\in\{0, 1, \cdots, s, s+1\}$, where

$$X_n^0 := X_n$$

$$X_n^j := D_n^j \qquad (1\leqq j\leqq s)$$

$$X_n^{s+1} := X_n^*.$$

Let in the following r and $i(1), \cdots, i(r)$ be fixed and consider the random variable

$$S_n := \operatorname{tr}_n(X_n^{i(1)} \cdots X_n^{i(r)})$$

$$= \frac{1}{n} \sum_{k_1, \ldots, k_r=1}^{n} X_n^{i,1}(k_1, k_2) X_n^{i(2)}(k_2, k_3) \cdots X_n^{i(r)}(k_r, k_1),$$

where $X_n^{i(j)}(k, l)$ is the (k, l)-entry in the matrix $X_n^{i(j)}$.

Firstly, we shall show that the expectation value $E[S_n]$ converges to the asserted expression. Secondly, we shall give an argument which guarantees also the almost sure convergence.

The first part is already treated in [Voi3], we shall only add to this treatment an 'explicit' formula for the limit. That $\rho \star \varphi$ takes on the same value in this case can also be seen easily, but we omit the proof of this since it follows from [Voi3].

We shall show that the limit for S_n has the following form: Let us call a non crossing partition $\mathcal{V} = \{V_1, \cdots, V_p\}$ of $\{1, \cdots, r\}$ *valid* if one of the following two conditions is fulfilled:

(1) We have $p=1$ and $i(v) \in \{1, \cdots, s\}$ for all $v=1, \cdots, r$ — in this case we define $\rho \star \varphi(\mathcal{V}) := \rho(D^{i(1)} \cdots D^{i(r)})$.

(2) The partition $\mathcal{V}$ contains a segment V_i, i.e. $V_i = (v, v+1, \cdots, v+m)$, such that

$$i(v), i(v+m) \in \{0, s+1\}$$

$$i(v) \neq i(v+m)$$

$$i(v+1), \cdots, i(v+m-1) \in \{1, \cdots, s\}$$

and such that $\mathcal{V} \setminus \{V_i\}$ is a valid partition of $\{1, \cdots, r\} \setminus V_i$. In this second case we define

$$\rho \star \varphi(V_i) := \varphi(X^{i(v)} X^{i(v+m)}) \rho(D^{i(v+1)} \cdots D^{i(v+m-1)})$$

and recursively

$$\rho \star \varphi(\mathcal{V}) := \rho \star \varphi(V_i) \cdot \rho \star \varphi(\mathcal{V} \setminus \{V_i\}).$$

For example, for $\mathcal{V} = \{(1, 9), (2, 3, 7, 8), (4, 5, 6)\}$ (where $i(1), i(9) \in \{0, s+1\}$) we have

$$\rho \star \varphi(\mathcal{V}) = \varphi(X^{i(2)} X^{i(8)}) \rho(D^{i(3)} D^{i(7)}) \varphi(X^{i(4)} X^{i(6)}) \rho(D^{i(5)}) \rho(D^{i(1)} D^{i(9)}).$$

If we now denote the set of all valid partitions of $\{1, \cdots, r\}$ by $\mathcal{P}_v(1, \cdots, r)$, then we shall show that almost surely

$$\lim_{n \to \infty} S_n = \sum_{\mathcal{V} \in \mathcal{P}_v(1, \cdots, r)} \rho \star \varphi(\mathcal{V}).$$

Now let us first consider the expectation value of S_n. For getting a contribution of the tuple $(k_1, \cdots, k_r)$ to

$$E[S_n] = \frac{1}{n} \sum_{k_1, \ldots, k_r=1}^{n} E[X_n^{i(1)}(k_1, k_2) \cdots X_n^{i(r)}(k_r, k_1)]$$

the tuple must be such that each appearing matrix element of X_n occurs at least twice in the product (under consideration of $X_n^*(k, l) = \bar{X}_n(l, k)$), because different entries of the matrices X_n are independent and have expectation zero.

Denote by I the set of the positions of the X_n and X_n^*, i.e.

$$I := \{k \mid i(k) \in \{0, s+1\}\}.$$

Given an index-tuple $(k_1, \cdots, k_r)$, we put $k_{r+1} := k_1$ and call a pair (k_j, k_{j+1}) for $1 \le j \le r$ a *step* of $(k_1, \cdots, k_r)$. We shall refine this notation by calling the pair

- a *free step*, if $i(j) \in \{0, s+1\}$ and k_{j+1} has not appeared before, i.e. $k_{j+1} \ne k_l$ for all l with $1 \le l \le j$.
- a *null step*, if $i(j) \in \{1, \cdots, s\}$ and $k_j = k_{j+1}$.
- a *repetive step*, if $i(j) \in \{0, s+1\}$ and k_{j+1} has appeared before, i.e. there is a l with $1 \le l \le j$ such that $k_{j+1} = k_l$.

We say that two steps (k_j, k_{j+1}) and (k_l, k_{l+1}) *agree*, if $i(j), i(l) \in \{0, s+1\}$ and

$$\begin{cases} (k_j, k_{j+1}) = (k_l, k_{l+1}), & \text{if } i(j) = i(l) \\ (k_j, k_{j+1}) = (k_{l+1}, k_l), & \text{if } i(j) \ne i(l). \end{cases}$$

Now one sees as in [Wig] that we only have to consider index-tuples $(k_1, \cdots, k_r)$ which have the following properties: They contain only free, null or repetive steps and the number of the free steps is equal to the number of the repetive steps, namely they have no choice but to be $|I|/2$. This implies that in such a tuple each repetive step (k_j, k_{j+1}) has to agree with exactly one prior free step (k_l, k_{l+1}). Let us call such a tuple a *valid* one. Combinatorial arguments show that the contribution of all non valid tuples vanishes in the limit, their sum is at most of order $1/n$ (cf. [Wig]).

Now consider a valid tuple $(k_1, \cdots, k_r)$. It defines a valid partition $\mathcal{V}$ in the following way: Let m be the smallest number such that (k_m, k_{m+1}) is a repetive step. Then this step has to be equal to exactly one prior free step, say (k_l, k_{l+1}) with $l < m$. This is only possible if $(k_{l+1}, k_l) = (k_m, k_{m+1})$ (i.e. a X_n is paired with a X_n^* and vice versa) and if all steps between l and m are null steps. Then we define $V_1 := (l+1, \cdots, m)$ and take this part from k_{l+1} to k_m out of our tuple $(k_1, \cdots, k_r)$. Since $k_l = k_{m+1}$, we get again a valid tuple and we can repeat the above procedure. But in this way we exactly build up in a recursive way a valid partition and it is easy to see that the summation over all tuples resulting in the same valid partition $\mathcal{V}$ gives the contribution $\rho \star \varphi(\mathcal{V})$ to $\lim_{n \to \infty} \mathrm{E}[S_n]$. Since on the other hand all valid partitions can appear, the assertion is proved.

It remains to prove that S_n converges also almost surely. This can be done by showing a sufficient good estimate for the variance $\mathrm{Var}[S_n]$ of the form

$$(*) \qquad \mathrm{Var}\,[S_n] \leq \frac{c}{n^2},$$

where the constant c is independent of n. This is sufficient because one can conclude

$$\mathrm{E}\Big[\sum_{n=1}^{\infty}(S_n - \mathrm{E}[S_n])^2\Big] = \sum_{n=1}^{\infty}\mathrm{Var}\,[S_n] < \infty,$$

which implies

$$\sum_{n=1}^{\infty}(S_n - \mathrm{E}[S_n])^2 < \infty \qquad \text{almost surely,}$$

and therefore

$$\lim_{n\to\infty}(S_n - \mathrm{E}[S_n]) = 0 \qquad \text{almost surely.}$$

Thus it remains to prove $(*)$. We have

$$\mathrm{Var}\,[S_n] = \mathrm{E}[S_n^2] - \mathrm{E}[S_n]^2$$

$$= \frac{1}{n^2}\sum_{k_1,\,\cdots,k_r=1}^{\infty}\sum_{l_1,\,\cdots,l_r=1}^{\infty}$$

$$\{\mathrm{E}[X_n^{i(1)}(k_1,\,k_2)\cdots X_n^{i(r)}(k_r,\,k_1)X_n^{i(1)}(l_1,\,l_2)\cdots X_n^{i(r)}(l_r,\,l_1)] -$$

$$\mathrm{E}[X_n^{i(1)}(k_1,\,k_2)\cdots X_n^{i(r)}(k_r,\,k_1)]\cdot\mathrm{E}[X_n^{i(1)}(l_1,\,l_2)\cdots X_n^{i(r)}(l_r,\,l_1)]\}.$$

For all tuples $(k_1,\,\cdots,\,k_r)$, $(l_1,\,\cdots,\,l_r)$ where each step of $(k_1,\,\cdots,\,k_r)$ does not agree with any step of $(l_1,\,\cdots,\,l_r)$ the k-factors are independent of the l-factors, hence such tuples make no contribution. Now consider on the contrary those tuples where at least one step of $(k_1,\,\cdots,\,k_r)$ agrees with one step of $(l_1,\,\cdots,\,l_r)$. There are finitely many possibilities which k-step agrees with which l-step. We fix one of these possibilities, say we consider the set of all tuples where $(k_m,\,k_{m+1})$ agrees with $(l_j,\,l_{j+1})$ for fixed m and j with $1\leq m,\,j\leq r$. Of course, we can restrict to the case that both of these pairs belong to a X_n or a X_n^*, i. e. that $i(m),\,i(j)\in\{0,\,s+1\}$.

Consider furthermore first only such tuples with $i(m)=i(j)$, hence $(k_m,\,k_{m+1}) =(l_j,\,l_{j+1})$. Then we can rewrite

$$X_n^{i(1)}(k_1,\,k_2)\cdots X_n^{i(r)}(k_r,\,k_1)X_n^{i(1)}(l_1,\,l_2)\cdots X_n^{i(r)}(l_r,\,l_1)$$

as

$$X_n^{i(m)}(k_m,\,k_{m+1})\cdots X_n^{i(r)}(k_r,\,k_1)X_n^{i(1)}(k_1,\,k_2)\cdots X_n^{i(m-1)}(k_{m-1},\,k_m)$$

$$X_n^{i(j)}(l_j,\,l_{j+1})\cdots X_n^{i(r)}(l_r,\,l_1)X_n^{i(1)}(l_1,\,l_2)\cdots X_n^{i(j-1)}(l_{j-1},\,l_j)$$

and apply our results from the first part of this proof to the $2r$-tuple

$$(k_m,\,\cdots,\,k_r,\,k_1,\,\cdots,\,k_{m-1},\,k_m=l_j,\,l_{j+1},\,\cdots,\,l_r,\,l_1,\,\cdots,\,l_{j-1}).$$

They say that all our considered tuples are not valid (since $i(m)=i(j)$ means a

pairing of a X_n with a X_n or a pairing of a X_n^* with a X_n^*), which results in a sum of order $1/n^2$ (note that the factor $1/n$ in our considerations on $E[S_n]$ is now replaced by $1/n^2$).

Now consider the case $i(m) \neq i(j)$, which means $(k_m, k_{m+1}) = (l_{j+1}, l_j)$. In this case our problem is equivalent to the consideration of the $2r$-tuple

$$(k_{m+1}, \cdots, k_r, k_1, \cdots, k_m, k_{m+1} = l_j, l_{j+1}, \cdots, l_r, l_1, \cdots, l_{j-1}).$$

But again such tuples are not valid, because the steps (k_m, k_{m+1}) and (l_j, l_{j+1}) agree, but both are repetive steps (k_{m+1} has appeared before as first element of the sequence and l_{j+1} has appeared before as k_m), whereas in valid tuples always repetive steps are paired with free steps. Thus again the contribution of this case is of order $1/n^2$.

All in all, the sum of all tuples for which the k-factors and the l-factors are not independent can be estimated as less than c/n^2, where c is some combinatorial factor independent of n. This finishes the proof. $\qquad \diamond$

Now we can prove our theorem.

Proof of Theorem. First, we can assume without restriction of generality that all A_n and B_n are diagonal matrices: We can diagonalize A_n and B_n in the form $A_n = V_n A_n^d V_n^*$ and $B_n = W_n B_n^d W_n^*$. Then

$$\mu_{A_n + U_n B_n U_n^*} = \mu_{A_n^d + (V_n^* U_n W_n) B_n^d (V_n^* U_n W_n)^*}$$

and the assertion is equivalent to the assertion for $\mathcal{A}^d = (A_n^d)_{n \in N}$ and $\mathcal{B}^d = (B_n^d)_{n \in N}$, since the map $U_n \mapsto V_n^* U_n W_n$ preserves the Haar measure on $U(n)$.

Let

$$F : \Omega_{\text{gauss}} \longrightarrow \Omega_{\text{unitary}}$$

$$\mathcal{X} = (X_n) \longmapsto \mathcal{U} = (U_n)$$

be the measurable mapping defined by

$$U_n := F(X_n) := X_n (X_n^* X_n)^{-1/2},$$

which is defined almost everywhere. Since the measure on Ω_{unitary} is the image measure under this mapping of the measure on Ω_{gauss} (cf. [Voi3]), it is sufficient to prove

$$\mu_{A_n + F(X_n) B_n F(X_n)^*} \xrightarrow{\ w\ } \mu_{\mathcal{A}} \boxplus \mu_{\mathcal{B}}$$

for almost all $\mathcal{X} = (X_n)_{n \in N} \in \Omega_{\text{gauss}}$.

Choose now such an $\mathcal{X} = (X_n)_{n \in N} \in \Omega_{\text{gauss}}$ where we can assume that $U_n := X_n (X_n^* X_n)^{-1/2}$ is defined for all $n \in N$. Since $\mu_{\mathcal{A}}$ and $\mu_{\mathcal{B}}$, and hence $\mu_{\mathcal{A}} \boxplus \mu_{\mathcal{B}}$, have compact support, they are determined by their moments and it suffices to show the convergence of all moments of $\mu_{A_n + U_n B_n U_n^*}$ to the corresponding mo-

ments of $\mu_{\mathcal{A}} \boxplus \mu_{\mathcal{B}}$. This is guaranteed if we check the characterizing property of free independence, namely: Let $P_1, \cdots, P_{2r}$ be polynomials in one indeterminant such that for all $i=1, \cdots, r$

$$\lim_{n \to \infty} \operatorname{tr}_n(P_{2i-1}(A_n)) = 0$$

and

$$\lim_{n \to \infty} \operatorname{tr}_n(P_{2i}(U_n B_n U_n^*)) = \lim_{n \to \infty} \operatorname{tr}_n(P_{2i}(B_n)) = 0 .$$

Then we have to show

$$\lim_{n \to \infty} \operatorname{tr}_n\Big(\prod_{i=1}^{r} P_{2i-1}(A_n) P_{2i}(U_n B_n U_n^*) \Big)$$

$$= \lim_{n \to \infty} \operatorname{tr}_n\Big(\prod_{i=1}^{r} P_{2i-1}(A_n) U_n P_{2i}(B_n) U_n^* \Big) = 0 .$$

We shall show this by approximating U_n by polynomials in X_n and X_n^*. One knows that $\lim_{n\to\infty} \|X_n\| = 2$ almost surely (cf. e. g. [Sil, Gem]), i. e. we can assume that there exists an $M \in \mathbf{R}$, such that $\|X_n\| \leq M$ for all $n \in \mathbf{N}$. Choose now for $\varepsilon > 0$ a polynomial g_ε with real coefficients which is a good approximation of $1/\sqrt{x}$ in the sense that we have

$$\frac{1}{\pi} \int_{-2}^{+2} (1 - g_\varepsilon(x^2)x)^{2l} \sqrt{1-(x/2)^2}\, dx \leq \varepsilon \qquad \text{for some fixed } l \text{ with } 4r \leq 2^l$$

and

$$\sup_{x \in [-M^2, M^2]} |x g_\varepsilon(x)^2| \leq 1 .$$

Now put

$$Q_\varepsilon(X_n, X_n^*) := X_n g_\varepsilon(X_n^* X_n) .$$

Then we can write

$$U_n = Q_\varepsilon(X_n, X_n^*) + R_n^{(\varepsilon)} ,$$

where $Q_\varepsilon(X_n, X_n^*)$ is a polynomial in X_n and X_n^*, whose norm can be estimated as

$$\|Q_\varepsilon(X_n, X_n^*)\|^2 = \|g_\varepsilon(X_n^* X_n) X_n^* X_n g_\varepsilon(X_n^* X_n)\| \leq 1 ,$$

and $R_n^{(\varepsilon)}$ is an $n \times n$-matrix, which is small in the following sense:

$$\operatorname{tr}_n(|R_n^{(\varepsilon)}|^{2l}) = \operatorname{tr}_n((R_n^{(\varepsilon)*} R_n^{(\varepsilon)})^{2^{l-1}})$$

$$= \operatorname{tr}_n(\{[(X_n^* X_n)^{-1/2} - g_\varepsilon(X_n^* X_n)] X_n^* X_n [(X_n^* X_n)^{-1/2} - g_\varepsilon(X_n^* X_n)]\}^{2^{l-1}})$$

$$= \operatorname{tr}_n(\{[1 - g_\varepsilon(X_n^* X_n)(X_n^* X_n)^{1/2}][1 - (X_n^* X_n)^{1/2} g_\varepsilon(X_n^* X_n)]\}^{2^{l-1}}) .$$

By the second remark after our lemma and by the fact that φ on $\mathbf{C}\langle X^*X \rangle$ is determined by its moments we have for all continuous, bounded f on $\mathbf{R}$

$$\lim_{n\to\infty} \mathrm{tr}_n(f(X_n^* X_n)) = \frac{1}{\pi}\int_{-2}^{+2} f(x^2)\sqrt{1-(x/2)^2}\,dx\ .$$

Since we have $\|X_n^* X_n\| \le M^2$ for all $n \in \mathbf{N}$, we can even take any continuous function f on $\mathbf{R}$ and thus we get

$$\lim_{n\to\infty} \mathrm{tr}_n(|R_n^{(\varepsilon)}|^{2l}) = \frac{1}{\pi}\int_{-2}^{+2}(1-g_\varepsilon(x^2)x)^{2l}\sqrt{1-(x/2)^2}\,dx \le \varepsilon\ .$$

Since in all monomials in the polynomial $Q_\varepsilon(X_n, X_n^*)$ the number of X_n and the number of X_n^* is different, we can infer from our lemma that

$$\lim_{n\to\infty} \mathrm{tr}_n(Q_\varepsilon(X_n, X_n^*)) = 0\ .$$

Now we have

$$\lim_{n\to\infty} \mathrm{tr}_n\Big(\prod_{i=1}^{r} P_{2i-1}(A_n)U_n P_{2i}(B_n)U_n^*\Big)$$

$$= \lim_{n\to\infty} \mathrm{tr}_n\Big(\prod_{i=1}^{r} P_{2i-1}(A_n)Q_\varepsilon(X_n, X_n^*)P_{2i}(B_n)Q_\varepsilon(X_n, X_n^*)^*\Big)$$

$$+ \lim_{n\to\infty} \sum \mathrm{tr}_n\Big(\prod_{i=1}^{r} P_{2i-1}(A_n)S_n^{(2i-1)}P_{2i}(B_n)S_n^{(2i)}\Big),$$

where in the first term all U_n are replaced by $Q_\varepsilon(X_n, X_n^*)$ and the sum in the second term runs over all $S_n^{(1)}, \cdots, S_n^{(2r)} \in \{Q_\varepsilon(X_n, X_n^*), R_n^{(\varepsilon)}\}$, where at least one $S_n^{(j)}$ is equal to $R_n^{(\varepsilon)}$.

By our lemma, the first term vanishes in the limit $n\to\infty$. (One should note that one needs indeed a little stronger version of the lemma, because we can not be sure that the joint distribution of A_n and B_n exists in the limit $n\to\infty$. But since our $Q_\varepsilon(X_n, X_n^*)$ has the special property that it really seperates A_n from B_n in the sense that in each monomial of $Q_\varepsilon(X_n, X_n^*)$ the number of X_n is different from the number of X_n^*, the valid partitions in the calculation of the considered term never combine a A_n with a B_n, i.e. the joint distribution is not really needed. For the argument that the non-valid terms do not contribute it is sufficient to know that the possibly appearing joint moments of A_n and B_n remain bounded as n goes to ∞. But this is guaranteed by the assumption of a compact support for $\mu_{\mathcal{A}}$ and $\mu_{\mathcal{B}}$.)

A summand in the second term can be estimated as follows: From

$$|\mathrm{tr}_n(AB)|^2 \le \mathrm{tr}_n(|A|^2)\,\mathrm{tr}_n(|B|^2) \qquad (A, B \in M_n)$$

we can deduce by induction

$$|\mathrm{tr}_n(A_1 \cdots A_k)|^{2^l} \le \mathrm{tr}_n(|A_1|^{2^l}) \cdots \mathrm{tr}_n(|A_k|^{2^l})$$

for all $k \le 2^l$ and $A_1, \cdots, A_k \in M_n$. This implies for our fixed l with $4r \le 2^l$

$$\left|\mathrm{tr}_n\Big(\prod_{i=1}^{r} P_{2i-1}(A_n)S_n^{2i-1}P_{2i}(B_n)S_n^{(2i)^*}\Big)\right|^{2l}$$

$$\leqq \prod_{i=1}^{r} \mathrm{tr}_n(|P_{2i-1}(A_n)|^{2l})\,\mathrm{tr}_n(|P_{2i}(B_n)|^{2l}) \cdot \prod_{j=1}^{2r} \mathrm{tr}_n(|S_n^{(j)}|^{2l}).$$

(Note that $\mathrm{tr}_n(|S_n^{(j)^*}|^{2l})=\mathrm{tr}_n(|S_n^{(j)}|^{2l})$.) In the limit we can estimate these factors as follows:

$$\lim_{n\to\infty}\prod_{i=1}^{r}\mathrm{tr}_n(|P_{2i-1}(A_n)|^{2l})=\prod_{i=1}^{r}\int |P_{2i-1}(x)|^{2l}\,d\mu_{\mathcal{A}}(x)=:M_{\mathcal{A}}<\infty$$

$$\lim_{n\to\infty}\prod_{i=1}^{r}\mathrm{tr}_n(|P_{2i}(B_n)|^{2l})=\prod_{i=1}^{r}\int |P_{2i}(x)|^{2l}\,d\mu_{\mathcal{B}}(x)=:M_{\mathcal{B}}<\infty.$$

If $S_n^{(j)}=Q_\varepsilon(X_n, X_n^*)$ then we have

$$\mathrm{tr}_n(|S_n^{(j)}|^{2l})\leqq\|Q_\varepsilon(X_n, X_n^*)\|^{2l}\leqq 1,$$

and for $S_n^{(j)}=R_n^{(\varepsilon)}$ we get

$$\lim_{n\to\infty}\mathrm{tr}_n(|S_n^{(j)}|^{2l})=\lim_{n\to\infty}\mathrm{tr}_n(|R_n^{(\varepsilon)}|^{2l})\leqq\varepsilon.$$

Since at least one of the $S_n^{(j)}$ is equal to $R_n^{(\varepsilon)}$ we get for $\varepsilon\leqq 1$

$$\left|\lim_{n\to\infty}\mathrm{tr}_n\Big(\prod_{i=1}^{r} P_{2i-1}(A_n)U_n P_{2i}(B_n)U_n^*\Big)\right|\leqq(2^{2r}-1)^{2l}\sqrt[2l]{\varepsilon M_{\mathcal{A}}M_{\mathcal{B}}}.$$

Since this is valid for all ε with $0<\varepsilon\leqq 1$ and since $M_{\mathcal{A}}$ and $M_{\mathcal{B}}$ are independent of ε we arrive finally at

$$\lim_{n\to\infty}\mathrm{tr}_n\Big(\prod_{i=1}^{r} P_{2i-1}(A_n)U_n P_{2i}(B_n)U_n^*\Big)=0,$$

which we wanted to show. ◇

References

[Bil] Billingsley, P., *Probability and Measure*, John Wiley & Sons, New York, 1979.

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

[Kre] Kreweras, G., Sur les partitions non croisees d'un cycle, *Discr. Math.*, 1 (1972), 333–350.

[Maa] Maassen, H., Addition of Freely Independent Random Variables, *J. Funct. Anal.*, 106 (1992), 409–438.

[Sil] Silverstein, J. W., Eigenvalues and eigenvectors of large dimensional sample covariance matrices, *Contemporary Mathematics*, 50 (1986), 153–160.

[Spe] Speicher, R., A New Example of 'Independence' and 'White Noise', *Probab. Th. Rel. Fields*, 84 (1990), 141–159.

[Voi1] Voiculescu, D., Symmetries of some reduced free product C*-algebras, in *Operator*

Algebras and their Connection with Topology and Ergodic Theory, Lecture Notes in Mathematics 1132, Springer Verlag, Heidelberg 1985, 556–588.

[Voi2] Voiculescu, D., Addition of certain non-commuting random variables, *J. Funct. Anal.*, **66** (1986), 323–346.

[Voi3] ————, Limit laws for random matrices and free products, *Invent. math.*, **104** (1991), 201–220.

[Wig] Wigner, E. P., Characteristic vectors of bordered matrices with infinite dimensions, *Ann. of Math.*, **62** (1955), 548–564.