Free probability for purely discrete eigenvalues of random matrices

By Benoit COLLINS, Takahiro HASEBE and Noriyoshi SAKUMA

Abstract. In this paper, we study random matrix models which are obtained as a non-commutative polynomial in random matrix variables of two kinds: (a) a first kind which have a discrete spectrum in the limit, (b) a second kind which have a joint limiting distribution in Voiculescu's sense and are globally rotationally invariant. We assume that each monomial constituting this polynomial contains at least one variable of type (a), and show that this random matrix model has a set of eigenvalues that almost surely converges to a deterministic set of numbers that is either finite or accumulating to only zero in the large dimension limit. For this purpose we define a framework (cyclic monotone independence) for analyzing discrete spectra and develop the moment method for the eigenvalues of compact (and in particular Schatten class) operators. We give several explicit calculations of discrete eigenvalues of our model.

1. Introduction.

1.1. Background.

Free probability (see e.g. [**25**]) is a branch of operator algebras that was invented by Voiculescu for the purpose of studying properties of free group factors. Later Voiculescu discovered in [**23**] that free probability has also an application to the behavior of eigenvalues of non-commutative polynomials in independent large random matrices. This is one of the most striking success of free probability.

Let $M_n(\mathbb{C})$ be the set of all $n \times n$ matrices whose entries are complex values. When applying free probability to random matrices, the standard assumption is that a family of Hermitian matrices $B_1(n), \ldots, B_k(n) \in M_n(\mathbb{C})$ has a (*joint*) *limiting distribution* as $n \to \infty$, meaning that for any non-commutative $*$-polynomial P in k variables, $\mathrm{tr}_n(P(B_1(n), \ldots, B_k(n)))$ admits a finite limit as $n \to \infty$ where tr_n is the normalized trace such that $\mathrm{tr}_n(I_n) = 1$. Then Voiculescu's result [**24**] (see also [**23**]) states that if $U(n)$ is a Haar unitary random matrix, then with probability one, the enlarged family $\{B_1(n), \ldots, B_k(n), U(n)\}$ also has a joint limiting distribution almost surely as $n \to \infty$, and in the limit, $U(n)$ becomes free from $\{B_1(n), \ldots, B_k(n)\}$.

Furthermore, Haagerup–Thorbjørnsen [**13**], Male [**14**] and Collins–Male [**10**] obtained versions of Voiculescu's results in the context of operator norm convergence. What they proved is that the family of matrices $\{B_1(n), \ldots, B_k(n), U(n)\}$ admits a *strong*

(*joint*) *limiting distribution* as $n \to \infty$, which means that it has a joint limiting distribution as stated before, and in addition, for any non-commutative self-adjoint $*$-polynomial P in $k+1$ non-commuting variables, the random matrix $P(B_1(n), \ldots, B_k(n), U(n))$ has no *outliers*, i.e. no eigenvalues outside the limiting support of the spectrum. Before the above results in free probability, several 'single random matrix models' were known to have strong limiting distributions; for example, this is the case for Wigner matrices under some assumptions, and in particular for Gaussian unitary ensembles (GUE) and Wishart matrices (see [**1**, Theorem 2.1.22 and Bibliographical notes]).

On the other hand, in the last 10 years, random matrix models that do not have strong limiting distributions have become fashionable. The literature is abundant. We refer for examples to [**3**], [**5**], [**8**], [**19**], [**20**], [**22**] and in particular to the pioneering work of Baik–Ben-Arous–Péché [**2**] and Péché [**18**] where outliers of finite rank deformations of Wishart matrices and of GUEs have been studied, respectively.

1.2. Our model.

The purpose of this paper is to investigate random matrix models similar to those studied in [**2**], [**3**], [**5**], [**18**], but our models admit *purely discrete spectra* when the dimension tends to infinity. In other words, our model has, as a limiting spectrum, the eigenvalues of a selfadjoint compact operator on a Hilbert space. The precise definition of our model is as follows. For simplicity, the dependence on n being dropped, let $\{A_1, \ldots, A_k\}$ be a family of $n \times n$ deterministic matrices which has a limiting joint distribution with respect to the *non-normalized trace* Tr_n, i.e. for any non-commutative $*$-polynomial P without a constant term, the following limit exists:

$$\lim_{n \to \infty} \mathrm{Tr}_n(P(A_1, \ldots, A_k)). \tag{1.1}$$

Let $\{B_1, \ldots, B_\ell\}$ be a family of $n \times n$ deterministic matrices which has a limiting joint distribution with respect to tr_n. Let U be an $n \times n$ Haar unitary (we can treat several independent Haar unitaries, but for now we restrict to a single Haar unitary). We investigate the limiting eigenvalues of

$$P(A_1, \ldots, A_k, UB_1U^*, \ldots, UB_\ell U^*), \tag{1.2}$$

where P is a $k+\ell$ variables selfadjoint non-commutative $*$-polynomial $P(\{x_i\}_{i=1}^k, \{y_j\}_{j=1}^\ell)$ such that $P(\{0\}_{i=1}^k, \{y_j\}_{j=1}^\ell) = 0$.

1.3. Main results.

The main results of this paper are as follows.

(i) We introduce and investigate *cyclic monotone independence* which is a universal computation rule for mixed moments with respect to weights (Section 3) as abstraction of the formula [**21**, Lemma 3.1].

(ii) The pair of tuples $(\{A_i\}_{i=1}^k, \{U^*B_jU\}_{j=1}^\ell)$ above is asymptotically cyclic monotone with respect to $(\mathrm{Tr}_n, \mathrm{tr}_n)$ almost surely (Theorem 4.3).

(iii) The eigenvalues of (1.2) converge to deterministic eigenvalues of a compact operator almost surely (Corollary 4.4). We then extend (by functional calculus) this result to

compact operators $A_1, \ldots, A_k$ for which the limit (1.1) may not exist (Theorem 4.7).

(iv) We compute in Theorem 5.1 the limiting eigenvalues of (1.2) explicitly when the polynomial P is of the forms

$$\sum_{i=1}^{k} x_i y_i x_i^*, \qquad \sum_{i=1}^{k} y_i x_i y_i^*, \qquad xy + yx, \qquad \mathrm{i}(xy - yx).$$

(v) We also discuss a generalization of the model (1.2) when several independent Haar unitaries appear. We show the almost sure convergence of eigenvalues (Corollary 4.10) and compute explicit eigenvalues for some polynomials of random matrices (Proposition 5.2).

Our model (1.2) is closely related to the recent work of Shlyakhtenko [21] where asymptotic infinitesimal freeness was proved for $\{A_1, \ldots, A_k\}$ and $\{UB_1U^*, \ldots, UB_\ell U^*\}$ when $A_1, \ldots, A_k$ are fixed matrices of finite size. We strengthen this result with a self-contained proof, and then show an almost sure asymptotic convergence result.

Shlyakhtenko gives a very interesting interpretation of his result in terms of locations of outliers (discrete spectra) and continuous spectra, assuming that the outliers exist. His arguments strongly suggest that infinitesimal freeness can be useful for outlier problems studied by Baik–Ben-Arous–Péché [2], Péché [18] and others. On the other hand, our research is devoted to purely discrete spectra. In our discrete spectrum model, we are able to show that the "outliers" indeed exist almost surely for completely general polynomials in general matrices $A_1, \ldots, A_k$ converging to compact operators and rotationally invariant random matrices.

On methodology, our model needs a new method outside the standard techniques in free probability, since our model (1.2) *converges to* 0 in distribution in the usual sense [21]:

$$\lim_{n \to \infty} \mathrm{tr}_n \left(P(A_1, \ldots, A_k, UB_1U^*, \ldots, UB_\ell U^*)^p \right) = 0, \qquad p \in \mathbb{N}. \qquad (1.3)$$

To analyze the discrete spectrum, we develop the moment method with respect to the *non-normalized trace* Tr_n. The most important point is that the convergence of moments with respect to Tr_n as $n \to \infty$ implies the pointwise convergence of eigenvalues. Together with this moment method, the *Weingarten calculus* developed in free probability [9], [11] enables us to compute moments with respect to Tr_n and prove the pointwise convergence of eigenvalues. When continuous and discrete spectra are mixed, it is not obvious if our method can somehow be extended.

1.4. Organization of this paper.

After this introduction, Section 2 gathers preliminary materials in order to handle eigenvalue distributions of non-commutative random variables that are compact operators. In Section 3 we introduce the notion, central to this paper, of cyclic monotone independence which is a special case of infinitesimal freeness, in the framework of a non-commutative probability space with a tracial weight. Section 4 shows the almost sure asymptotic cyclic monotone independence of $\{A_1, \ldots, A_k\}$ and $\{UB_1U^*, \ldots, UB_\ell U^*\}$.

Finally, Section 5 provides examples of eigenvalues of our model (1.2) in large n limit.

2. Convergence of eigenvalues.

In this section, we consider convergence of eigenvalues in a general setting. First, we introduce an order and the classification of eigenvalues to prove theorems in this section. After then, we obtain characterization of convergence of eigenvalues for Schatten class operators from a viewpoint of the non-normalized trace. The results play crucial roles to handle eigenvalues using non-commutative probability theory.

We use the following notations in this paper (in particular in this section).

(1) $C_{0,\mathrm{b}}(\mathbb{R})$: The set of real-valued bounded continuous functions on $\mathbb{R}$ that vanish in a neighborhood of 0.

(2) $\|\cdot\|_{[\alpha,\beta]}$: The supremum norm on $C[\alpha,\beta]$, $-\infty < \alpha < \beta < \infty$.

(3) $C_{0,\mathrm{b}}^{\infty}(\mathbb{R})$: The set of functions $f \in C_{0,\mathrm{b}}(\mathbb{R})$ that are infinitely differentiable.

(4) Tr_H: The trace on a separable Hilbert space H. When $H = M_n(\mathbb{C})$, Tr_H is denoted by Tr_n.

(5) tr_n: The normalized trace $(1/n)\,\mathrm{Tr}_n$ on $M_n(\mathbb{C})$.

(6) $\mathcal{S}^p(H)$: The set of p-Schatten class operators on a separable Hilbert space H.

(7) $\|\cdot\|_p$: The p-Schatten norm. If a is a selfadjoint compact operator with eigenvalues $\{\lambda_i\}_{i \geq 1}$ then $\|a\|_p = (\sum_{i=1}^{\infty} |\lambda_i|^p)^{1/p}$.

2.1. Order for eigenvalues.
It is useful to regard eigenvalues as a multiset.

DEFINITION 2.1.   For a selfadjoint compact operator a, we denote by $\mathrm{EV}(a)$ the multiset of its eigenvalues. The disjoint union of multisets counts the multiplicity, e.g. $\{3,2,1,1,0,0,\dots\} \sqcup \{2,1,1,0,0,\dots\} = \{3,2,2,1,1,1,1,0,0,\dots\}$.

Let us make remarks about this definition.

- We may also view $\mathrm{EV}(a)$ as a positive measure μ on $\mathbb{R}$, namely $\mu(\{\lambda\})$ is the dimension of the eigenspace of a with eigenvalue λ.

- Alternatively, we may view $\mathrm{EV}(a)$ as the collection of all real sequences (x_n) tending to zero, quotiented by the equivalence relation $(x_n) \sim (y_n)$ iff there exists a permutation σ of $\mathbb{N}$ such that $x_n = y_{\sigma(n)}$ for all n.

- Whenever needs be, we extend the notion of eigenvalues abstractly to selfadjoint elements of a $*$-algebra with a tracial weight even if the weight is not a trace on a separable Hilbert space. This is defined in Section 3 and is related to Section 5.

In order to discuss the convergence of eigenvalues, it is useful to order them in a nice way.

DEFINITION 2.2. We say that a sequence of real numbers $\{r_i\}_{i=1}^{\infty}$ converging to 0 is *properly arranged* if $|r_i| \geq |r_{i+1}|$ for all $i \in \mathbb{N}$. Note that the proper arrangement may not be unique.

Due to the non-uniqueness of proper arrangement, it is sometimes better to decompose a sequence into the nonnegative part and nonpositive part. The proper arrangement of the nonnegative part $\{r_i \mid i \in \mathbb{N}, r_i \geq 0\}$ is unique and is denoted by $\{r_i^+\}_{i=1}^{\infty}$, and similarly $\{r_i^-\}_{i=1}^{\infty}$ denotes the unique proper arrangement of $\{r_i \mid i \in \mathbb{N}, r_i \leq 0\}$.

From now on, we always assume that eigenvalues $\{\lambda_i\}_{i \geq 1}$ are properly arranged, namely

$$|\lambda_1| \geq |\lambda_2| \geq \cdots, \tag{2.1}$$

and their nonnegative and nonpositive parts are also properly arranged uniquely,

$$\lambda_1^+ \geq \lambda_2^+ \geq \cdots \geq 0 \geq \cdots \geq \lambda_2^- \geq \lambda_1^-. \tag{2.2}$$

The properly arranged eigenvalues of a selfadjoint compact operator a on a separable Hilbert space are denoted by $\{\lambda_i(a)\}_{i \geq 1}$. If the dimension of the Hilbert space is finite then we understand that the index i stops at the dimension. Instead of $\lambda_i(a)^{\pm}$ we use the notation $\lambda_i^{\pm}(a)$ for the properly arranged nonnegative and nonpositive parts of the eigenvalues.

2.2. Convergence of Eigenvalues.

DEFINITION 2.3. Let $a, a_k, k = 1, 2, 3, \ldots$ be selfadjoint compact operators on separable Hilbert spaces H, H_k, respectively. We say that a_k *converges to a in eigenvalues* if $\lim_{k \to \infty} \lambda_i^u(a_k) = \lambda_i^u(a)$ for any $i \in \mathbb{N}$ and $u \in \{+, -\}$. If a sequence stops at a finite i then infinitely many 0's are to be added in the end. Convergence in eigenvalues is denoted by

$$\lim_{k \to \infty} \mathrm{EV}(a_k) = \mathrm{EV}(a).$$

REMARK 2.4. It seems also natural to define the convergence $a_k \to a$ in eigenvalues by

$$\lim_{k \to \infty} \lambda_i(a_k) = \lambda_i(a) \qquad i \in \mathbb{N}, \tag{$*$}$$

but this is not good. For example if $\lambda_i(a_k) = \{-1, 1 - 1/k, 1/2, 1/3, 1/4, \ldots\}$ and $\lambda_i(a) = \{1, -1, 1/2, 1/3, 1/4, \ldots\}$, then $a_k \to a$ in eigenvalues, but the convergence does not hold in the sense of $(*)$.

Note that the set of eigenvalue sequences

$$\{\{\lambda_i\}_{i \geq 1} \subset \mathbb{R} : \lim_{i \to \infty} \lambda_i = 0\} \tag{2.3}$$

is metrizable by

$$d(\{\lambda_i\}_{i\geq 1}, \{\mu_i\}_{i\geq 1}) := \sum_{i=1}^{\infty} \frac{1}{2^i} \frac{|\lambda_i^+ - \mu_i^+|}{1 + |\lambda_i^+ - \mu_i^+|} + \sum_{i=1}^{\infty} \frac{1}{2^i} \frac{|\lambda_i^- - \mu_i^-|}{1 + |\lambda_i^- - \mu_i^-|}, \qquad (2.4)$$

where infinitely many 0's are to be added in the end if the sequence $\lambda_i^{\pm}$ or $\mu_i^{\pm}$ ends at a finite i. This metric is compatible with the convergence in eigenvalues.

PROPOSITION 2.5.   *Let $a, a_k, k = 1, 2, 3, \ldots$ be selfadjoint compact operators on separable Hilbert spaces $H, H_k, k = 1, 2, 3, \ldots$, respectively. Then the following are equivalent.*

(1) *a_k converges to a in eigenvalues (cf. Definition 2.3).*

(2) *$\lim_{k\to\infty} \mathrm{Tr}_{H_k}(f(a_k)) = \mathrm{Tr}_H(f(a))$ for any $f \in C_{0,\mathrm{b}}(\mathbb{R})$.*

(3) *$\lim_{k\to\infty} \mathrm{Tr}_{H_k}(f(a_k)) = \mathrm{Tr}_H(f(a))$ for any $f \in C_{0,\mathrm{b}}^{\infty}(\mathbb{R})$.*

PROOF.   Let $\lambda_i := \lambda_i(a)$ and $\lambda_i(k) := \lambda_i(a_k)$ for simplicity. Recall that we arrange the eigenvalues in the way $\lambda_1^+ \geq \lambda_2^+ \geq \cdots$ and $\lambda_1^- \leq \lambda_2^- \leq \cdots$, and similarly for $\{\lambda_i^{\pm}(k)\}_{i\geq 1}$.

(1)$\Rightarrow$(2). Take $f \in C_{0,\mathrm{b}}(\mathbb{R})$ then $f \equiv 0$ on $(-\delta, \delta)$ for some $\delta > 0$. Only finitely many eigenvalues of a are contained in $(-\delta, \delta)^c$, say $\lambda_i^+, i = 1, 2, \ldots, \ell$ and $\lambda_i^-, i = 1, 2, \ldots, m$. Convergence in eigenvalues implies that there exists k_0 such that $0 \leq \lambda_{\ell+1}^+(k) < \delta$ for all $k \geq k_0$. This implies that $0 \leq \lambda_i^+(k) < \delta$ for all $k \geq k_0$ and $i \geq \ell + 1$. Similar facts hold for negative eigenvalues. Therefore, for sufficiently large k we have

$$\mathrm{Tr}_{H_k}(f(a_k)) = \sum_{i=1}^{\ell} f(\lambda_i^+(k)) + \sum_{i=1}^{m} f(\lambda_i^-(k))$$

$$\xrightarrow{k\to\infty} \sum_{i=1}^{\ell} f(\lambda_i^+) + \sum_{i=1}^{m} f(\lambda_i^-) = \mathrm{Tr}_H(f(a)). \qquad (2.5)$$

(2)$\Rightarrow$(3). Obvious.

(3)$\Rightarrow$(1). If we take $f \in C_{0,\mathrm{b}}^{\infty}(\mathbb{R})$ such that $f \geq 0$ on $\mathbb{R}$, $f(x) = 0$ for $|x| \leq \|a\|$ and $f(x) = 1$ for $|x| > 2\|a\|$, then $\mathrm{Tr}_H(f(a)) = 0$. So for sufficiently large k we have $\mathrm{Tr}_H(f(a_k)) < 1$ and hence a_k has no eigenvalues in $\{x \in \mathbb{R} : |x| > 2\|a\|\}$. This implies that the eigenvalues of $\{a_k : k \in \mathbb{N}\}$ are uniformly bounded, i.e. they are contained in a common interval $[-\alpha, \alpha]$.

If $\lambda_1^+(k)$ does not converge to λ_1^+ then there exists a subsequence $(k_j)_{j\geq 1}$, a real number $\mu_1^+ \in [0, \alpha]$ such that $\mu_1^+ \neq \lambda_1^+$ and $\lambda_1^+(k_j) \to \mu_1^+$ as $j \to \infty$. We derive a contradiction below. Let $\varepsilon_1 := |\lambda_1^+ - \mu_1^+| > 0$.

- Case $0 \leq \lambda_1^+ < \mu_1^+$. We take a nonnegative function $f \in C_{0,\mathrm{b}}^{\infty}(\mathbb{R})$ such that $f(x) = 1$ for $x \geq \mu_1^+ - \varepsilon_1/4$ and $f(x) = 0$ for $x \leq \lambda_1^+ + \varepsilon_1/4$. Then all eigenvalues of a lie outside of the support of f. Therefore $\mathrm{Tr}_{H_{k_j}}(f(a_{k_j})) \geq f(\lambda_1^+(k_j)) = 1$ for large j, but $\mathrm{Tr}_H(f(a)) = 0$, a contradiction.

- Case $0 \leq \mu_1^+ < \lambda_1^+$. We take a nonnegative function $g \in C_{0,\mathrm{b}}^{\infty}(\mathbb{R})$ such that $g(x) = 1$

for $x \geq \lambda_1^+ - \varepsilon_1/4$ and $g(x) = 0$ for $x \leq \mu_1^+ + \varepsilon_1/4$. Then $\mathrm{Tr}_{H_{k_j}}(f(a_{k_j})) = 0$ for large j, but $\mathrm{Tr}_H(f(a)) \geq f(\lambda_1^+) = 1$, a contradiction.

Thus we conclude that $\lambda_1^+(k)$ converges to λ_1^+ as $k \to \infty$. Then we go to the induction: Suppose that $\lambda_i^+(k) \to \lambda_i^+$ as $k \to \infty$ for every $i = 1, \ldots, \ell - 1$ and suppose that $\lambda_\ell^+(k)$ does not converge to λ_ℓ^+. Then as before there exists a subsequence $(k_j)_{j \geq 1}$, a real number $\mu_\ell^+ \in [0, \lambda_{\ell-1}^+]$ such that $\mu_\ell^+ \neq \lambda_\ell^+$ and $\lambda_\ell^+(k_j) \to \mu_\ell^+$ as $j \to \infty$. Let $\varepsilon := |\lambda_\ell^+ - \mu_\ell^+| > 0$.

- Case $\lambda_\ell^+ < \mu_\ell^+$. We take a nonnegative function $f \in C_{0,\mathrm{b}}^\infty(\mathbb{R})$ such that $f(x) = 1$ for $x \geq \mu_\ell^+ - \varepsilon/4$ and $f(x) = 0$ for $x \leq \lambda_\ell^+ + \varepsilon/4$. Then $\mathrm{Tr}_{H_{k_j}}(f(a_{k_j})) \geq \ell$ for large j, but $\mathrm{Tr}_H(f(a)) = \sum_{i=1}^{\ell-1} f(\lambda_\ell^+) = \ell - 1$, a contradiction.

- Case $\mu_\ell^+ < \lambda_\ell^+$. We take a nonnegative function $g \in C_{0,\mathrm{b}}^\infty(\mathbb{R})$ such that $g(x) = 1$ for $x \geq \lambda_\ell^+ - \varepsilon/4$ and $g(x) = 0$ for $x \leq \mu_\ell^+ + \varepsilon/4$. Then $\mathrm{Tr}_{H_{k_j}}(f(a_{k_j})) = \ell - 1$ for large j, but $\mathrm{Tr}_H(f(a)) \geq \ell$, a contradiction.

Thus $\lambda_\ell^+(k)$ converges to λ_ℓ^+ as $k \to \infty$. By induction we conclude that $\lambda_i^+(k)$ converges to λ_i^+ as $k \to \infty$ for every $i \geq 1$. Similarly we can prove the convergence of $\lambda_i^-(k)$ to λ_i^-. $\qquad\square$

We define a notion of the distributional convergence of a tuple of compact operators, which is motivated from Definition 2.3 and Proposition 2.5.

DEFINITION 2.6. Given selfadjoint compact operators $a_i, a_i(n), i = 1, \ldots, k$ on separable Hilbert spaces H, H_n respectively, we say that $(a_1(n), \ldots, a_k(n))$ *converges in compact distribution* to $(a_1, \ldots, a_k)$ with respect to $\mathrm{Tr}_{H_n}, \mathrm{Tr}_H$ as $n \to \infty$ if for every function $f_i \in C_{0,\mathrm{b}}(\mathbb{R}), i = 1, \ldots, k$, $p \in \mathbb{N}$ and tuple $(i_1, \ldots, i_p) \in \{1, \ldots, k\}^p$ we have that

$$\lim_{n \to \infty} \mathrm{Tr}_{H_n}(f_{i_1}(a_{i_1}(n)) \cdots f_{i_p}(a_{i_p}(n))) = \mathrm{Tr}_H(f_{i_1}(a_{i_1}) \cdots f_{i_p}(a_{i_p})). \tag{2.6}$$

2.3. Moment method for convergence of Eigenvalues.

In noncommutative probability, the moment method is an important tool to prove weak convergence of probability measures. Here, we show a counterpart for eigenvalues.

LEMMA 2.7. *Let $\alpha > \delta > 0$ and let $p \in \mathbb{N}$. If $f \in C^p(\mathbb{R})$ and $f \equiv 0$ on $[-\delta, \delta]$, then for any $\varepsilon > 0$, there exists a polynomial P such that $0 = P(0) = P'(0) = \cdots = P^{(p-1)}(0)$, $\|f - P\|_{[-\alpha, \alpha]} < \varepsilon$ and $|P(x)| \leq \varepsilon |x|^p$ for $x \in [-\delta, \delta]$.*

PROOF. Without loss of generality we may assume that $\alpha = 1$. We only consider $p = 2$ for simplicity; the general case is similar. By Weierstrass' approximation we can find a polynomial $R_0(x)$ such that $\|f'' - R_0\|_{[-\alpha, \alpha]} < \varepsilon/8$. This implies that $|R_0(0)| < \varepsilon/8$, so the polynomial $R(x) := R_0(x) - R_0(0)$ satisfies that $\|f'' - R_0\|_{[-\alpha, \alpha]} < \varepsilon/4$. Then we define $Q(x) := \int_0^x R(y)\,dy$. Note that $Q(x)$ does not have a constant term. For $x \in [-1, 1]$ one has that $|f'(x) - Q(x)| = |\int_0^x f''(y)\,dy - \int_0^x R(y)\,dy| \leq \int_{-1}^1 |f''(y) - R(y)|\,dy \leq \varepsilon/2$. Then we define $P(x) := \int_0^x Q(y)\,dy$, which satisfies $P(0) = P'(0) = 0$. Then for any $x \in$

$[-1,1]$ one has that $|f(x)-P(x)| = |\int_0^x f'(y)\,\mathrm{d}y - \int_0^x Q(y)\,\mathrm{d}y| \le \int_{-1}^1 |f'(y)-Q(y)|\,\mathrm{d}y \le \varepsilon$. For $x \in [-\delta, \delta]$, we have

$$
\begin{aligned}
|P(x)| = \left|\int_0^x \mathrm{d}y \int_0^y R(z)\,\mathrm{d}z\right| &\le \int_{|y|\le|x|} \mathrm{d}y \int_{|z|\le|x|} |R(z)|\,\mathrm{d}z \\
&= 2|x| \int_{|z|\le|x|} |R(z)|\,\mathrm{d}z \le \varepsilon x^2.
\end{aligned}
\tag{2.7}
$$

On the last line we have used the fact that $f''(z) = 0$ for $z \in [-\delta, \delta]$ and $\|f''-R\|_{[-\alpha,\alpha]} \le \varepsilon/4$. $\qquad\square$

PROPOSITION 2.8. *Let $a, a_k, k = 1, 2, 3, \dots$ be selfadjoint operators on separable Hilbert spaces $H, H_k, k = 1, 2, 3, \dots$, respectively. Suppose that there exists $p \in \mathbb{N}$ such that $a \in \mathcal{S}^p(H)$ and $a_k \in \mathcal{S}^p(H_k)$ for all $k \in \mathbb{N}$. Suppose that $\mathrm{Tr}_{H_k}(a_k^n) \to \mathrm{Tr}_H(a^n)$ as $k \to \infty$ for any integer $n \ge p$. Then a_k converges to a in eigenvalues.*

PROOF. We may assume that p is an even integer since $\mathcal{S}^p \subset \mathcal{S}^{p+1}$. By Proposition 2.5, it suffices to show that for any bounded $C^p(\mathbb{R})$ function f such that $f \equiv 0$ on $[-\delta, \delta]$ for some $\delta > 0$, it holds that $\mathrm{Tr}_{H_k}(f(a_k)) \to \mathrm{Tr}_H(f(a))$ $(k \to \infty)$.

If the dimensions of some of H, H_k are finite, then the following proof is still available by adding infinitely many 0's to the eigenvalues, so let us assume that the dimensions of Hilbert spaces are all infinite. Let $\{\lambda_i\}_{i=1}^\infty$, $\{\lambda_i(k)\}_{i=1}^\infty$ be the properly arranged eigenvalues of a, a_k, respectively.

Let $\alpha := \sup_{k\ge 1} \|a_k\|_p < \infty$ and suppose that $\alpha > 0$; otherwise a, a_k are all zero elements. It is easy to see by Chebyshev's inequality that

$$
\#\{i \in \mathbb{N} : |\lambda_i(k)| > \delta\} = \#\{i \in \mathbb{N} : \lambda_i(k)^p > \delta^p\} \le \frac{\mathrm{Tr}(a_k^p)}{\delta^p} \le \frac{\alpha^p}{\delta^p},
\tag{2.8}
$$

and similarly, $\#\{i \in \mathbb{N} : |\lambda_i| > \delta\} \le \alpha^p/\delta^p$. Therefore if we define $i_0 := [\alpha^p/\delta^p] \in \mathbb{N}$ then $|\lambda_i(k)|, |\lambda_i| \le \delta$ for all $k \in \mathbb{N}$ and $i > i_0$.

Since $\|a\|_p^p, \|a_k\|_p^p \le \alpha^p$, the eigenvalues of a, a_k are all contained in the interval $[-\alpha, \alpha]$. Given $\varepsilon > 0$, by Lemma 2.7 we can find a polynomial P of the form $P(x) = a_p x^p + a_{p+1} x^{p+1} + \cdots + a_q x^q$ such that $\|f - P\|_{[-\alpha,\alpha]} < \varepsilon/i_0$ and $|P(x)| \le \varepsilon \alpha^{-p} x^p$ for $x \in [-\delta, \delta]$. Then for all $k \in \mathbb{N}$ we have that

$$
\begin{aligned}
\left|\sum_{i=1}^\infty f(\lambda_i(k)) - \sum_{i=1}^\infty f(\lambda_i)\right| &= \left|\sum_{i=1}^{i_0} f(\lambda_i(k)) - \sum_{i=1}^{i_0} f(\lambda_i)\right| \\
&\le \left|\sum_{i=1}^{i_0} f(\lambda_i(k)) - \sum_{i=1}^{i_0} P(\lambda_i(k))\right| + \left|\sum_{i=1}^{i_0} P(\lambda_i(k)) - \sum_{i=1}^{i_0} P(\lambda_i)\right| \\
&\quad + \left|\sum_{i=1}^{i_0} P(\lambda_i) - \sum_{i=1}^{i_0} f(\lambda_i)\right| \\
&\le 2\varepsilon + \left|\sum_{i=1}^{i_0} P(\lambda_i(k)) - \sum_{i=1}^{i_0} P(\lambda_i)\right|.
\end{aligned}
\tag{2.9}
$$

We note here that

$$\sum_{i=i_0+1}^{\infty} |P(\lambda_i(k))| \leq \varepsilon \alpha^{-p} \sum_{i=i_0+1}^{\infty} |\lambda_i(k)|^p \leq \varepsilon \alpha^{-p} \|a_k\|_p^p \leq \varepsilon, \qquad (2.10)$$

and similarly, $\sum_{i=i_0+1}^{\infty} |P(\lambda_i)| \leq \varepsilon$. Therefore,

$$\left| \sum_{i=1}^{i_0} P(\lambda_i(k)) - \sum_{i=1}^{i_0} P(\lambda_i) \right| \leq 2\varepsilon + \left| \sum_{i=1}^{\infty} P(\lambda_i(k)) - \sum_{i=1}^{\infty} P(\lambda_i) \right|. \qquad (2.11)$$

Since $\mathrm{Tr}_{H_k}(a_k^n) = \sum_{i=1}^{\infty} \lambda_i(k)^p \to \sum_{i=1}^{\infty} \lambda_i^p = \mathrm{Tr}_H(a^n)$ for all $n \geq p$ and the coefficients of $1, x, \dots, x^{p-1}$ of P are zero, there exists $k_0 \in \mathbb{N}$ such that

$$\left| \sum_{i=1}^{\infty} P(\lambda_i(k)) - \sum_{i=1}^{\infty} P(\lambda_i) \right| \leq \varepsilon \qquad (2.12)$$

for $k \geq k_0$. From (2.9), (2.11), (2.12) we have that

$$\left| \sum_{i=1}^{\infty} f(\lambda_i(k)) - \sum_{i=1}^{\infty} f(\lambda_i) \right| \leq 5\varepsilon, \qquad k \geq k_0, \qquad (2.13)$$

the conclusion. $\qquad \square$

COROLLARY 2.9. *Let a, b be selfadjoint operators on separable Hilbert spaces H, K, respectively, such that $a \in \mathcal{S}^p(H)$ and $b \in \mathcal{S}^p(K)$ for some $p \in \mathbb{N}$. If $\mathrm{Tr}_H(a^n) = \mathrm{Tr}_K(b^n)$ for every integer $n \geq p$ then $\mathrm{EV}(a) = \mathrm{EV}(b)$.*

Actually in Proposition 2.8, we need not a priori assume the existence of a limiting operator a.

PROPOSITION 2.10. *Let $p \in 2\mathbb{N}$ and $H_k, k \in \mathbb{N}$ be separable Hilbert spaces. Let $a_k \in \mathcal{S}^p(H_k), k = 1, 2, 3, \dots$ be selfadjoint. Suppose that the limit $\alpha_n := \lim_{k \to \infty} \mathrm{Tr}_{H_k}(a_k^n) \in \mathbb{R}$ exists at least for all integers $n \geq p$. Then there exist a separable Hilbert space H and a selfadjoint operator $a \in \mathcal{S}^p(H)$ such that a_k converges to a in eigenvalues, and moreover, $\alpha_n = \mathrm{Tr}_H(a^n)$ for at least $n \geq p + 1$.*

REMARK 2.11. (1) It is useless to assume that $p \in 2\mathbb{N} - 1$, because then the conclusion will only be $a \in \mathcal{S}^{p+1}(H)$.

(2) We cannot conclude that $\alpha_p = \mathrm{Tr}_H(a^p)$. For example we may take as eigenvalues of a_k

$$\lambda_i(k) = \begin{cases} \dfrac{1}{i}, & 1 \leq i \leq k, \\[2mm] \dfrac{1}{k}, & k+1 \leq i \leq k + k^2, \\[2mm] \dfrac{1}{i - k^2}, & i \geq k + k^2 + 1. \end{cases}$$

Then $\lambda_i := \lim_{k\to\infty} \lambda_i(k) = 1/i$ for all $i \in \mathbb{N}$, but $\sum_{i=1}^{\infty} \lambda_i(k)^2 = \zeta(2) + 1 \neq \sum_{i=1}^{\infty} \lambda_i^2$.

Proof.   Let $\{\lambda_i(k)\}_{i\geq 1}$ be the properly arranged eigenvalues of a_k, and let $\alpha := \sup_{k\in\mathbb{N}} \|a_k\|_p$. Since $|\lambda_i(k)|$ are uniformly bounded by α, there exists a subsequence $\{k(1,j)\}_{j\geq 1}$ of $\mathbb{N}$ such that $\lambda_1(k(1,j))$ converges to some λ_1. Then we can take a subsequence $\{k(2,j)\}_{j\geq 1}$ of $\{k(1,j)\}_{j\geq 1}$ such that $\lambda_2(k(2,j))$ converges to some λ_2, and continue this procedure. Let $k_m := k(m,m)$ and $\lambda_i(m) := \lambda_i(k_m)$. By construction $\{k_m\}_{m\geq i} \subset \{k(i,j)\}_{j\geq 1}$ for each $i \in \mathbb{N}$. Therefore $\lim_{m\to\infty} \lambda_i(m) = \lambda_i$ for every $i \in \mathbb{N}$. Then $\{\lambda_i\}_{i\geq 1}$ is properly arranged since so is $\{\lambda_i(k)\}_{i\geq 1}$. By Fatou's lemma

$$\sum_{i=1}^{\infty} \lambda_i^p \leq \varliminf_{m\to\infty} \sum_{i=1}^{\infty} \lambda_i(m)^p \leq \alpha^p. \tag{2.14}$$

Therefore, there exists a separable Hilbert space H and a selfadjoint $a \in \mathcal{S}^p(H)$ such that $\mathrm{EV}(a) = \{\lambda_i\}_{i\geq 1}$. Note that $\beta_n := \sum_{i=1}^{\infty} \lambda_i^n$ is absolutely convergent for $n \geq p$ since $\mathcal{S}^n(H)$ is increasing on $n \geq 1$.

We want to show that $\beta_n = \alpha_n$ for $n \geq p+1$. For any $\varepsilon > 0$, by Chebyshev's inequality (2.8) (now we use $\varepsilon > 0$ instead of $\delta > 0$), there exists $i_0 \in \mathbb{N}$ such that $|\lambda_{i_0+1}(k)| \leq \varepsilon$ for all $k \in \mathbb{N}$, and hence $|\lambda_{i_0+1}| \leq \varepsilon$. For $n \geq p+1$ we have

$$|\alpha_n - \beta_n| \leq |\alpha_n - \mathrm{Tr}_{H_k}(a_k^n)| + |\mathrm{Tr}_{H_k}(a_k^n) - \beta_n|$$

$$\leq |\alpha_n - \mathrm{Tr}_{H_k}(a_k^n)| + \left| \sum_{i=1}^{i_0} (\lambda_i(k)^n - \lambda_i^n) \right| + \sum_{i=i_0+1}^{\infty} (|\lambda_i(k)|^n + |\lambda_i|^n)$$

$$\leq |\alpha_n - \mathrm{Tr}_{H_k}(a_k^n)| + \left| \sum_{i=1}^{i_0} (\lambda_i(k)^n - \lambda_i^n) \right| + \varepsilon^{n-p} \left(\sum_{i=i_0+1}^{\infty} (\lambda_i(k)^p + \lambda_i^p) \right).$$
$$\tag{2.15}$$

The first two terms converge to 0 if we put $k = k_m$ and let m tend to infinity. The last term is bounded by $2\varepsilon^{n-p}\alpha^p$, so we get $|\alpha_n - \beta_n| \leq 2\alpha^p \varepsilon^{n-p}$ and hence $\alpha_n = \beta_n$ for $n \geq p+1$. By Corollary 2.9 (applied to now $p+1$), the limit eigenvalues $\{\lambda_i\}_{i\geq 1}$ do not depend on the choice of the subsequence $\{\lambda(k_m)\}_{m\geq 1}$. Since the space of eigenvalue distributions is metrizable (see (2.4)), a_k converges in eigenvalues to a.   □

3.   Cyclic monotone independence.

Our aim is to analyze asymptotic behavior of discrete eigenvalues of random matrices. For this we would like to abstract the notion of the non-normalized trace Tr_H on a Hilbert space H. So we define a non-commutative measure space that replaces a state in a non-commutative probability space with a *weight*.

3.1.   Algebraic non-commutative measure space.
A *non-commutative measure space* is a pair $(\mathcal{A}, \omega)$ where $\mathcal{A}$ is a (unital or non-unital) ∗-algebra over $\mathbb{C}$. Let ω be a tracial weight meaning that

- ω is defined on a (possibly non-unital) ∗-subalgebra $\mathrm{D}(\omega)$ of $\mathcal{A}$ and $\omega : \mathrm{D}(\omega) \to \mathbb{C}$

is linear,

- ω is positive, i.e. $\omega(a^*a) \geq 0$ for every $a \in \mathrm{D}(\omega)$,

- $\omega(a^*) = \overline{\omega(a)}$ for all $a \in \mathrm{D}(\omega)$,

- $\omega(ab) = \omega(ba)$ for all $a, b \in \mathrm{D}(\omega)$.

Moreover, if $\mathcal{A}$ is unital, $\mathrm{D}(\omega) = \mathcal{A}$ and $\omega(1_{\mathcal{A}}) = 1$ then we call $(\mathcal{A}, \omega)$ *a non-commutative probability space.*

Let $(\mathcal{A}, \omega)$ be a non-commutative measure space and let $a_1, \dots, a_k \in \mathrm{D}(\omega)$. The *distribution* of $(a_1, \dots, a_k)$ is the family of (*mixed*) *moments*

$$\{\omega(a_{i_1}^{\varepsilon_1} \dots a_{i_p}^{\varepsilon_p}) : p \geq 1, 1 \leq i_1, \dots, i_p \leq k, (\varepsilon_1, \dots, \varepsilon_p) \in \{1, *\}^p\}. \tag{3.1}$$

Given non-commutative measure spaces $(\mathcal{A}, \omega), (\mathcal{B}, \xi)$ and elements $a_1, \dots, a_k \in \mathrm{D}(\omega)$, $b_1, \dots, b_k \in \mathrm{D}(\xi)$, we say that $(a_1, \dots, a_k)$ has *the same distribution* as $(b_1, \dots, b_k)$ if

$$\omega(a_{i_1}^{\varepsilon_1} \cdots a_{i_p}^{\varepsilon_p}) = \xi(b_{i_1}^{\varepsilon_1} \cdots b_{i_p}^{\varepsilon_p}) \tag{3.2}$$

for any choice of $p \in \mathbb{N}$, $1 \leq i_1, \dots, i_p \leq k$ and $(\varepsilon_1, \dots, \varepsilon_p) \in \{1, *\}^p$.

The distribution of $(a_1, \dots, a_k)$ is a *trace class distribution* if there exist a separable Hilbert space H and $x_1, \dots, x_k \in (\mathcal{S}^1(H), \mathrm{Tr}_H)$ (trace class operators) such that the distribution of $(a_1, \dots, a_k)$ is the same as that of $(x_1, \dots, x_k)$. In this case we define the non-zero eigenvalues of a self-adjoint $*$-polynomial $P(a_1, \dots, a_k)$ to be the eigenvalues of $P(x_1, \dots, x_k)$.

Let $(\mathcal{A}, \omega), \{(\mathcal{A}_n, \omega_n)\}_{n \geq 1}$ be non-commutative measure spaces and $a_1, \dots, a_k \in \mathrm{D}(\omega), a_1(n), \dots, a_k(n) \in \mathrm{D}(\omega_n)$. We say that $(a_1(n), \dots, a_k(n))$ *converges in distribution* to $(a_1, \dots, a_k)$ if

$$\lim_{n \to \infty} \omega_n(a_{i_1}(n)^{\varepsilon_1} \cdots a_{i_p}(n)^{\varepsilon_p}) = \omega(a_{i_1}^{\varepsilon_1} \cdots a_{i_p}^{\varepsilon_p}) \tag{3.3}$$

for any choice of $p \in \mathbb{N}$, $1 \leq i_1, \dots, i_p \leq k$ and $(\varepsilon_1, \dots, \varepsilon_p) \in \{1, *\}^p$. If the distributions of selfadjoint elements $a_n \in \mathrm{D}(\omega_n)$ and $a \in \mathrm{D}(\omega)$ are trace class then we can define the concept of *convergence in eigenvalues* in a natural way (cf. Definition 2.3), i.e. $\lambda_i^u(a_n) \to \lambda_i^u(a)$ for every $i \in \mathbb{N}$ and $u \in \{+, -\}$.

PROPOSITION 3.1. *Let $\{(\mathcal{A}_n, \omega_n)\}_{n \geq 1}, (\mathcal{A}, \omega)$ be non-commutative measure spaces. Suppose that $a_i(n) \in \mathrm{D}(\omega_n), a_i \in \mathrm{D}(\omega), i = 1, 2, \dots, k$ have trace class distributions and that $(a_1(n), \dots, a_k(n))$ converges in distribution to $(a_1, \dots, a_k)$. Then for any selfadjoint non-commutative $*$-polynomial P without a constant term, $P(a_1(n), \dots, a_k(n))$ converges in eigenvalues to $P(a_1, \dots, a_k)$.*

PROOF. This follows from Proposition 2.8. □

3.2. Cyclic monotone independence.

We introduce a kind of independence in a non-commutative probability space with a tracial weight $(\mathcal{C}, \omega, \tau)$. For $*$-subalgebras $\mathcal{A}, \mathcal{B}$ of $\mathcal{C}$ such that $1_{\mathcal{C}} \in \mathcal{B}$, let $\mathrm{Ideal}_{\mathcal{B}}(\mathcal{A})$ be the ideal generated by $\mathcal{A}$ over $\mathcal{B}$. More precisely,

$$\mathrm{Ideal}_{\mathcal{B}}(\mathcal{A}) := \mathrm{span}\{b_0 a_1 b_1 \cdots a_n b_n : n \in \mathbb{N}, a_1, \ldots, a_n \in \mathcal{A}, b_0, \ldots, b_n \in \mathcal{B}\}, \qquad (3.4)$$

which is a $*$-subalgebra of $\mathcal{C}$ containing $\mathcal{A}$.

DEFINITION 3.2. Let $(\mathcal{C}, \omega, \tau)$ be a non-commutative probability space with a tracial weight ω.

(1) Let $\mathcal{A}, \mathcal{B}$ be $*$-subalgebras of $\mathcal{C}$ such that $1_{\mathcal{C}} \in \mathcal{B}$. We say that the pair $(\mathcal{A}, \mathcal{B})$ is *cyclically monotonically independent* or more simply *cyclically monotone* with respect to (ω, τ) if

- $\mathrm{Ideal}_{\mathcal{B}}(\mathcal{A}) \subset \mathrm{D}(\omega)$;
- for any $n \in \mathbb{N}, a_1, \ldots, a_n \in \mathcal{A}$ and any $b_1, \ldots, b_n \in \mathcal{B}$, we have that

$$\omega(a_1 b_1 a_2 b_2 \cdots a_n b_n) = \omega(a_1 a_2 \cdots a_n)\tau(b_1)\tau(b_2)\cdots\tau(b_n).$$

(2) Given $a_1, \ldots, a_k \in \mathrm{D}(\omega)$ and $b_1, \ldots, b_\ell \in \mathcal{C}$, the pair $(\{a_1, \ldots, a_k\}, \{b_1, \ldots, b_\ell\})$ is *cyclically monotone* if $(\mathrm{alg}\{a_1, \ldots, a_k\}, \mathrm{alg}\{1_{\mathcal{C}}, b_1, \ldots, b_\ell\})$ is cyclically monotone. Note that we *do not* assume that $\mathrm{alg}\{a_1, \ldots, a_k\}$ contains the unit of $\mathcal{C}$.

REMARK 3.3. This definition is similar to monotone independence of Muraki [15], but the difference is

$$\omega(b_0 a_1 b_1 a_2 b_2 \cdots a_n b_n) = \omega(a_1 a_2 \cdots a_n)\tau(b_1)\tau(b_2)\cdots\tau(b_{n-1})\tau(b_0 b_n).$$

The factor $\tau(b_0 b_n)$ is to be replaced by $\tau(b_0)\tau(b_n)$ in the monotone case.

DEFINITION 3.4 (Cyclic monotone product). Let $(\mathcal{A}, \omega)$ be a non-commutative measure space and let $(\mathcal{B}, \tau)$ be a non-commutative probability space. Let $\mathcal{A} * \mathcal{B}$ be the algebraic free product of $\mathcal{A}$ and $\mathcal{B}$, $1_{\mathcal{B}}$ being identified with the unit $\mathbb{C}1_{\mathcal{A}*\mathcal{B}}$,

$$\mathbb{C}1_{\mathcal{A}*\mathcal{B}} \oplus \mathcal{B}^0 \oplus \mathcal{A} \oplus (\mathcal{A} \otimes \mathcal{B}^0) \oplus (\mathcal{B}^0 \otimes \mathcal{A}) \oplus (\mathcal{A} \otimes \mathcal{B}^0 \otimes \mathcal{A}) \oplus \cdots,$$

where $\mathcal{B} = \mathbb{C}1_{\mathcal{B}} \oplus \mathcal{B}^0$ is a direct sum decomposition as a vector space. We define a linear functional $\omega \rhd \tau$ on an ideal of $\mathcal{A} * \mathcal{B}$ by

$$(\omega \rhd \tau)(b_0 a_1 b_1 a_2 b_2 \cdots a_n b_n) := \omega(a_1 a_2 \cdots a_n)\tau(b_1)\tau(b_2)\cdots\tau(b_{n-1})\tau(b_0 b_n), \qquad (3.5)$$

$$\mathrm{D}(\omega \rhd \tau) := \mathrm{D}(\omega) \oplus (\mathrm{D}(\omega) \otimes \mathcal{B}^0) \oplus (\mathcal{B}^0 \otimes \mathrm{D}(\omega)) \oplus (\mathrm{D}(\omega) \otimes \mathcal{B}^0 \otimes \mathrm{D}(\omega)) \oplus \cdots \qquad (3.6)$$

for $n \in \mathbb{N}, a_1, \ldots, a_n \in \mathrm{D}(\omega), b_0, b_1, \ldots, b_n \in \mathcal{B} \simeq \mathcal{B}^0 \oplus \mathbb{C}1_{\mathcal{A}*\mathcal{B}}$. We call $\omega \rhd \tau$ the *cyclic monotone product* of ω and τ.

PROPOSITION 3.5. *The cyclic monotone product $\omega \rhd \tau$ is a tracial positive linear functional on $\mathrm{D}(\omega \rhd \tau)$ and hence $(\mathcal{A} * \mathcal{B}, \omega \rhd \tau)$ becomes a non-commutative measure space.*

PROOF. Let $x = b_0 a_1 b_1 \cdots a_n b_n$ and $y = b_0' a_1' b_1' \cdots a_m' b_m'$ where $a_i, a_i' \in \mathrm{D}(\omega), b_i, b_i' \in \mathcal{B}^0 \cup \mathbb{C}1_{\mathcal{A}*\mathcal{B}}$. The traciality follows from

$$(\omega \trianglerighteq \tau)(xy) = \omega(a_1 a_2 \cdots a_n a_1' \cdots a_m')\tau(b_0 b_m')\tau(b_n b_0') \prod_{i=1}^{n-1} \tau(b_i) \prod_{i=1}^{m-1} \tau(b_i')$$

$$= \omega(a_1' \cdots a_m' a_1 a_2 \cdots a_n)\tau(b_m' b_0)\tau(b_0' b_n) \prod_{i=1}^{n-1} \tau(b_i) \prod_{i=1}^{m-1} \tau(b_i')$$

$$= (\omega \trianglerighteq \tau)(yx). \tag{3.7}$$

For the positivity it is enough to prove $(\omega \trianglerighteq \tau)(x^*x) \geq 0$ when

$$x = \sum_{i=1}^{n} \lambda_i b_{i,0} a_{i,1} b_{i,1} a_{i,2} b_{i,2} \cdots a_{i,m(i)} b_{i,m(i)}, \tag{3.8}$$

where $\lambda_i \in \mathbb{C}$, $a_{i,j} \in \mathrm{D}(\omega)$ and $b_{i,j} \in \mathcal{B}^0 \cup \mathbb{C}1_{\mathcal{A}*\mathcal{B}}$. By the positivity of ω we have that

$$\omega\left(\left(\sum_{i=1}^{n} \lambda_i a_{i,1} \cdots a_{i,m(i)}\right)^* \left(\sum_{i=1}^{n} \lambda_i a_{i,1} \cdots a_{i,m(i)}\right)\right) \geq 0. \tag{3.9}$$

Then we have

$$\sum_{i,j=1}^{n} \overline{\lambda_i}\lambda_j \omega(a_{i,m(i)}^* \cdots a_{i,1}^* a_{j,1} \cdots a_{j,m(j)}) \geq 0 \tag{3.10}$$

for any $(\lambda_1, \ldots, \lambda_n) \in \mathbb{C}^n$. Thus the $n \times n$ matrix

$$A = \left(\omega(a_{i,m(i)}^* \cdots a_{i,1}^* a_{j,1} \cdots a_{j,m(j)})\right)_{i,j=1}^{n} \tag{3.11}$$

is positive definite. We also show that the matrices

$$B' := \left(\tau(b_{i,0}^* b_{j,0})\right)_{i,j=1}^{n}, \tag{3.12}$$

$$B'' := \left(\tau(b_{i,1}^*)\tau(b_{j,1}) \cdots \tau(b_{i,m(i)-1}^*)\tau(b_{j,m(j)-1})\tau(b_{i,m(i)}^* b_{j,m(j)})\right)_{i,j=1}^{n} \tag{3.13}$$

are positive definite. It is easy to show that B' is positive definite. The matrix B'' is also positive definite since

$$\sum_{i,j=1}^{n} \overline{\lambda_i}\lambda_j \tau(b_{i,1}^*)\tau(b_{j,1}) \cdots \tau(b_{i,m(i)-1}^*)\tau(b_{j,m(j)-1})\tau(b_{i,m(i)}^* b_{j,m(j)})$$

$$= \tau\left(\left(\sum_{i=1}^{n} \lambda_i \tau(b_{i,1}) \cdots \tau(b_{i,m(i)-1})b_{i,m(i)}\right)^* \left(\sum_{i=1}^{n} \lambda_i \tau(b_{i,1}) \cdots \tau(b_{i,m(i)-1})b_{i,m(i)}\right)\right) \geq 0 \tag{3.14}$$

for any vector $(\lambda_1, \ldots, \lambda_n) \in \mathbb{C}^n$.

The Schur product $A \circ B' \circ B''$ is also positive definite. Now using the definition of cyclic monotone product, we have

$$0 \le (\lambda_1, \ldots, \lambda_n)^*((A \circ B' \circ B'')(\lambda_1, \ldots, \lambda_n))$$

$$= (\omega \trianglerighteq \tau)\left(\left(\sum_{i=1}^{n} \lambda_i b_{i,0} a_{i,1} b_{i,1} \cdots a_{i,m(i)} b_{i,m(i)}\right)^* \left(\sum_{i=1}^{n} \lambda_i b_{i,0} a_{i,1} b_{i,1} \cdots a_{i,m(i)} b_{i,m(i)}\right)\right).$$

Therefore we obtain $(\omega \trianglerighteq \tau)(x^*x) \ge 0$.   □

REMARK 3.6.   It turns out that this scalar product is not faithful at the algebraic level. For example, with the above notation, if a, b are selfadjoint non-zero, and $\tau(b) = 0$, then, calling $c = aba$, one sees that $(\omega \trianglerighteq \tau)(c^2) = \omega(a^4)\tau(b)^2 = 0$.

Let $\widetilde{\tau} \colon \mathcal{A} * \mathcal{B} \to \mathbb{C}$ be the free product map of the zero map on $\mathcal{A}$ and the trace τ on $\mathcal{B}$. This map $\widetilde{\tau}$ is a tracial state on $\mathcal{A} * \mathcal{B}$, and gives a triple $(\mathcal{A} * \mathcal{B}, \omega \trianglerighteq \tau, \widetilde{\tau})$. Then $(\mathrm{D}(\omega), \mathcal{B})$ is cyclically monotone with respect to $(\omega \trianglerighteq \tau, \widetilde{\tau})$. This construction is a universal one.

When applying to random matrices we need the asymptotic version of independence.

DEFINITION 3.7.   Let $(\mathcal{C}_n, \omega_n, \tau_n)$ be non-commutative probability spaces with tracial weights.   Let $a_1(n), \ldots, a_k(n) \in \mathrm{D}(\omega_n)$ and $b_1(n), \ldots, b_\ell(n) \in \mathcal{C}_n$.   The pair $(\{a_i(n)\}_{i=1}^{k}, \{b_i(n)\}_{i=1}^{\ell})$ is *asymptotically cyclically monotone* if there exist a non-commutative probability space $(\mathcal{C}, \omega, \tau)$ with a tracial weight, $a_1, \ldots, a_k \in \mathrm{D}(\omega)$ and $b_1, \ldots, b_\ell \in \mathcal{C}$ such that

(1) $(\{a_1, \ldots, a_k\}, \{b_1, \ldots, b_\ell\})$ is cyclically monotone;

(2) for any non-commutative $*$-polynomial $P(x_1, \ldots, x_k, y_1, \ldots, y_\ell)$ such that

$$P(0, \ldots, 0, y_1, \ldots, y_\ell) = 0,$$

we have $P(a_1(n), \ldots, a_k(n), b_1(n), \ldots, b_\ell(n)) \in \mathrm{D}(\omega_n)$ and

$$\lim_{n \to \infty} \omega_n(P(a_1(n), \ldots, a_k(n), b_1(n), \ldots, b_\ell(n))) = \omega(P(a_1, \ldots, a_k, b_1, \ldots, b_\ell)).$$

3.3.   Relation to infinitesimal freeness.

Biane, Goodman and Nica introduced a kind of freeness related to type B noncrossing partitions in [**7**] and then Belinschi and Shlyakhtenko formulated it as infinitesimal freeness, which is freeness with respect to a parametrized state τ_t $(t \in (-\varepsilon, \varepsilon))$ up to the order $o(t)$ [**4**] (see also Février and Nica's work [**12**]). More precisely, let $\{\tau_t\}_{t \in (-\varepsilon, \varepsilon)}$ be a family of traces on an unital $*$-algebra $\mathcal{C}$. Let $\tau := \tau_0$ and suppose that τ_t is differentiable in the sense that the limit

$$\tau'(x) := \lim_{t \to 0} \frac{\tau_t(x) - \tau(x)}{t} \tag{3.15}$$

exists for every $x \in \mathcal{C}$. Moreover, suppose that $*$-subalgebras $\mathcal{A}, \mathcal{B}$ are free with respect to τ_t for every $t \in (-\varepsilon, \varepsilon)$, then we have a computation formula such as

$$\tau_t(ab) = \tau_t(a)\tau_t(b), \tag{3.16}$$

$$\tau_t(aba') = \tau_t(aa')\tau_t(b), \tag{3.17}$$

$$\tau_t(aba'b') = \tau_t(aa')\tau_t(b)\tau_t(b') + \tau_t(a)\tau_t(a')\tau_t(bb') - \tau_t(a)\tau_t(a')\tau_t(b)\tau_t(b'), \qquad (3.18)$$

for $a, a' \in \mathcal{A}$ and $b, b' \in \mathcal{B}$. Putting $t = 0$ we get these formulas for τ, and moreover if we take the derivative regarding t at 0 we get the formulas

$$\tau'(ab) = \tau'(a)\tau(b) + \tau(a)\tau'(b), \qquad (3.19)$$

$$\tau'(aba') = \tau'(aa')\tau(b) + \tau(aa')\tau'(b), \qquad (3.20)$$

$$\begin{aligned}
\tau'(aba'b') = {} & \tau'(aa')\tau(b)\tau(b') + \tau(aa')\tau'(b)\tau(b') + \tau'(aa')\tau(b)\tau'(b') \\
& + \tau'(a)\tau(a')\tau(bb') + \tau(a)\tau'(a')\tau(bb') + \tau(a)\tau(a')\tau'(bb') \\
& - (4 \text{ terms}). \qquad (3.21)
\end{aligned}$$

These computation formulas give some kind of universal formulas for mixed moments.

Conversely, given τ, τ' we can define infinitesimal freeness. Let $\mathcal{C}$ be a unital $*$-algebra, let τ be a tracial state on $\mathcal{C}$ and let τ' be a tracial linear functional which satisfies $\tau'(1_{\mathcal{C}}) = 0$. The triple $(\mathcal{C}, \tau', \tau)$ is called an *infinitesimal non-commutative probability space* or non-commutative probability space of type B.

DEFINITION 3.8.   Let $(\mathcal{C}, \tau', \tau)$ be an infinitesimal non-commutative probability space. Let $\mathcal{A}$ and $\mathcal{B}$ be $*$-subalgebras of $\mathcal{C}$, which may not contain the unit of $\mathcal{C}$. Let $\tau_t := \tau + t\tau'$ for $t \in \mathbb{R}$. We say that $\mathcal{A}$ and $\mathcal{B}$ are infinitesimally free if for any $n \in \mathbb{N}$, $a_1, \ldots, a_n \in \mathcal{A}$ and $b_1, \ldots, b_n \in \mathcal{B}$

$$\tau_t((a_1 - \tau_t(a_1)1_{\mathcal{C}})(b_1 - \tau_t(b_1)1_{\mathcal{C}})(a_2 - \tau_t(a_2)1_{\mathcal{C}}) \cdots (b_n - \tau_t(b_n)1_{\mathcal{C}})) = o(t). \qquad (3.22)$$

REMARK 3.9.   In the above definition τ_t may not be positive, but it does not matter in defining infinitesimal freeness.

Our cyclic monotone independence is a special case of infinitesimal freeness, which was essentially proved in [**21**].

PROPOSITION 3.10.   *Let $(\mathcal{C}, \tau', \tau)$ be an infinitesimal non-commutative probability space and let $\mathcal{A}, \mathcal{B}$ be its $*$-subalgebras such that $1_{\mathcal{C}} \in \mathcal{B}$ and $\mathrm{Ideal}_{\mathcal{B}}(\mathcal{A}) \subset \ker(\tau)$. Then $\mathcal{A}, \mathcal{B}$ are infinitesimally free if and only if $(\mathcal{A}, \mathcal{B})$ is cyclically monotone with respect to (τ', τ).*

REMARK 3.11.   We assumed the positivity of the weight and the state in the definition of cyclic monotone independence, but we may drop it. The conclusion of Proposition 3.10 is to be understood in this generalized setup.

PROOF.   Suppose that $\mathcal{A}, \mathcal{B}$ are infinitesimally free. Let $\tau_t := \tau + t\tau'$. Since $\mathcal{A}, \mathcal{B}$ are free with respect to τ_t up to $o(t)$, it can be shown that

$$\tau_t(a_1 b_1 \cdots a_n b_n) = \tau_t(a_1 a_2 \cdots a_n)\tau_t(b_1)\tau_t(b_2) \cdots \tau_t(b_n) + R(t) + o(t), \qquad (3.23)$$

where $R(t)$ is the sum of monomials such that every monomial in $R(t)$ contains at least two factors of the form $\tau_t(a_{k_1} \cdots a_{k_p}), p < n$. For the proof apply the formula for products of free random variables [**17**, Theorem 14.4]. Our assumption implies that $\tau|_{\mathcal{A}} = 0$, and

so $R(t) = O(t^2)$. Comparing the coefficients of t in (3.23) we get the formula

$$\tau'(a_1 b_1 \cdots a_n b_n) = \tau'(a_1 a_2 \cdots a_n)\tau(b_1)\tau(b_2)\cdots\tau(b_n), \qquad (3.24)$$

implying the cyclic monotone independence.

For the converse direction, we check Definition 3.8. We have the decomposition

$$
\begin{aligned}
&\tau_t((a_1 - \tau_t(a_1)1_{\mathcal{C}})(b_1 - \tau_t(b_1)1_{\mathcal{C}})(a_2 - \tau_t(a_2)1_{\mathcal{C}})\cdots(b_n - \tau_t(b_n)1_{\mathcal{C}}))\\
&= \tau((a_1 - \tau_t(a_1)1_{\mathcal{C}})(b_1 - \tau_t(b_1)1_{\mathcal{C}})(a_2 - \tau_t(a_2)1_{\mathcal{C}})\cdots(b_n - \tau_t(b_n)1_{\mathcal{C}}))\\
&\quad + t\tau'((a_1 - \tau_t(a_1)1_{\mathcal{C}})(b_1 - \tau_t(b_1)1_{\mathcal{C}})(a_2 - \tau_t(a_2)1_{\mathcal{C}})\cdots(b_n - \tau_t(b_n)1_{\mathcal{C}}))\\
&=: J_1 + tJ_2. \qquad (3.25)
\end{aligned}
$$

By the assumption $\mathrm{Ideal}_{\mathcal{B}}(\mathcal{A}) \subset \ker(\tau)$, we can see that

$$
\begin{aligned}
J_1 &= \tau((-\tau_t(a_1)1_{\mathcal{C}})(b_1 - \tau_t(b_1)1_{\mathcal{C}})(-\tau_t(a_2)1_{\mathcal{C}})\cdots(b_n - \tau_t(b_n)1_{\mathcal{C}}))\\
&= (-1)^n t^n \tau'(a_1)\cdots\tau'(a_n)\tau((b_1 - \tau_t(b_1)1_{\mathcal{C}})(b_2 - \tau_t(b_2)1_{\mathcal{C}})\cdots(b_n - \tau_t(b_n)1_{\mathcal{C}}))\\
&= O(t^2). \qquad (3.26)
\end{aligned}
$$

Note that when $n = 1$, we can show $J_1 = O(t^2)$ since $\tau(b_1 - \tau_t(b_1)1_{\mathcal{C}}) = t\tau'(b_1)$. For J_2, since $\tau_t(a_i) = -t\tau'(a_i)$ we can see that

$$J_2 = \tau'(a_1(b_1 - \tau_t(b_1)1_{\mathcal{C}})a_2\cdots a_n(b_n - \tau_t(b_n)1_{\mathcal{C}})) + O(t). \qquad (3.27)$$

By cyclic monotonicity we obtain

$$
\begin{aligned}
&\tau'(a_1(b_1 - \tau_t(b_1)1_{\mathcal{C}})a_2\cdots a_n(b_n - \tau_t(b_n)1_{\mathcal{C}}))\\
&= \tau'(a_1 a_2 \cdots a_n)(\tau(b_1) - \tau_t(b_1))\cdots(\tau(b_n) - \tau_t(b_n))\\
&= \tau'(a_1 a_2 \cdots a_n)(-t)^n \tau'(b_1)\tau'(b_2)\cdots\tau'(b_n)\\
&= O(t). \qquad (3.28)
\end{aligned}
$$

Combining (3.27) and (3.28) we obtain $J_2 = o(1)$. Therefore, $J_1 + tJ_2 = o(t)$. $\qquad \square$

For later use we define a notion of the convergence of elements in non-commutative probability spaces to elements in an infinitesimal non-commutative probability space. This definition is inspired by [4], [21].

DEFINITION 3.12. Suppose that $(\mathcal{C}, \tau', \tau)$ is an infinitesimal non-commutative probability space, $(\mathcal{C}_n, \tau_n)$ are non-commutative probability spaces and $c_1, \ldots, c_\ell \in \mathcal{C}, c_1(n), \ldots, c_\ell(n) \in \mathcal{C}_n$. We say that the tuple $((c_1(n), \ldots, c_\ell(n)), \tau_n)$ *converges up to the first order* to $((c_1, \ldots, c_\ell), (\tau', \tau))$ if

(i) $((c_1(n), \ldots, c_\ell(n)), \tau_n)$ converges in distribution to $((c_1, \ldots, c_\ell), \tau)$,

(ii) for any non-commutative $*$-polynomial $P(y_1, \ldots, y_\ell)$,

$$\lim_{n\to\infty} \frac{\tau_n(P(c_1(n), \ldots, c_\ell(n))) - \tau(P(c_1, \ldots, c_\ell))}{1/n} = \tau'(P(c_1, \ldots, c_\ell)).$$

3.4. Eigenvalues of polynomials of cyclic monotone elements.

In this section we compute the eigenvalues of polynomials of cyclic monotone elements. We need a combinatorial lemma to compute the eigenvalues of the anti-commutator.

LEMMA 3.13. *For every integer $n \geq 0$ we have*

$$\sum_{\ell=0}^{n-2m} \binom{\ell+m-1}{m-1}\binom{n-\ell-m}{m}(-1)^\ell = \binom{[n/2]}{m}, \qquad m = 0, 1, \ldots, [n/2],$$

where $\binom{p}{-1} := \delta_{p,-1}$ for integers p.

PROOF. If $m = 0$ the identity is easy, so we assume that $m \geq 1$. The desired formula can be written in the form

$$\sum_{p=0}^{\infty} \binom{p}{m-1}\binom{r-p}{m}(-1)^p = (-1)^{m-1}\binom{[(r+1)/2]}{m}, \qquad m = 1, \ldots, [(r+1)/2]. \quad (3.29)$$

In order to prove this we use the identity

$$\frac{x^m}{(1-x)^{m+1}} = \sum_{p=0}^{\infty}\binom{p}{m}x^p. \quad (3.30)$$

We observe that

$$\frac{(-x)^{m-1}}{(1+x)^m} \cdot \frac{x^m}{(1-x)^{m+1}} = \left(\sum_{p=0}^{\infty}\binom{p}{m-1}(-x)^p\right)\left(\sum_{q=0}^{\infty}\binom{q}{m}x^p\right)$$

$$= \sum_{r=0}^{\infty}\left(\sum_{p=0}^{\infty}\binom{p}{m-1}\binom{r-p}{m}(-1)^p\right)x^r. \quad (3.31)$$

The left hand side can be computed as

$$\frac{(-x)^{m-1}}{(1+x)^m} \cdot \frac{x^m}{(1-x)^{m+1}} = \frac{(-1)^{m-1}(1+x)x^{2m-1}}{(1-x^2)^{m+1}}$$

$$= (-1)^{m-1}(1+x)\sum_{p=0}^{\infty}\binom{p}{m}x^{2p-1}$$

$$= (-1)^{m-1}\sum_{p=0}^{\infty}\binom{p}{m}(x^{2p-1}+x^{2p})$$

$$= (-1)^{m-1}\sum_{r=0}^{\infty}\binom{[(r+1)/2]}{m}x^r. \quad (3.32)$$

The comparison of (3.31) and (3.32) implies the conclusion. $\square$

THEOREM 3.14. *Let $(\mathcal{C}, \omega, \tau)$ be a non-commutative probability space with a tracial*

weight. Let $a, a_1, \ldots, a_k \in \mathrm{D}(\omega)$ and let $b, b_1, \ldots, b_k \in \mathcal{C}$. Suppose that $(a, a_1, \ldots, a_k)$ has a trace class distribution with respect to ω (see Section 3.1) and that the pair $(\{a, a_1, \ldots, a_k\}, \{b, b_1, \ldots, b_k\})$ is cyclically monotone in $(\mathcal{C}, \omega, \tau)$.

(1) *Suppose that $a_1, \ldots, a_k$ are selfadjoint. Let $B := (\tau(b_i^* b_j))_{i,j=1}^k \in M_k(\mathbb{C})$. Then*

$$\mathrm{EV}\left(\sum_{i=1}^k b_i a_i b_i^*\right) = \mathrm{EV}\left(\sqrt{B}\,\mathrm{diag}(a_1, \ldots, a_k)\sqrt{B}\right),$$

where $\sqrt{B}\,\mathrm{diag}(a_1, \ldots, a_k)\sqrt{B}$ is viewed as an element of $(M_k(\mathbb{C}) \otimes \mathcal{C}, \mathrm{Tr}_k \otimes \omega)$.

(2) *Suppose that $b_1, \ldots, b_k$ are selfadjoint. Then*

$$\mathrm{EV}\left(\sum_{i=1}^k a_i b_i a_i^*\right) = \mathrm{EV}\left(\sum_{i=1}^k \tau(b_i) a_i a_i^*\right).$$

(3) *Suppose that a, b are selfadjoint. Let $p = \sqrt{\tau(b^2)} + \tau(b), q = -(\sqrt{\tau(b^2)} - \tau(b))$. Then*

$$\mathrm{EV}(ab + ba) = (p\,\mathrm{EV}(a)) \sqcup (q\,\mathrm{EV}(a)).$$

(4) *Suppose that a, b are selfadjoint. Let $r := \sqrt{\tau(b^2) - \tau(b)^2}$. Then*

$$\mathrm{EV}(\mathrm{i}[a, b]) = (r\,\mathrm{EV}(a)) \sqcup (-r\,\mathrm{EV}(a)).$$

REMARK 3.15.   Our proofs are rather direct combinatorial arguments, but it may be possible to prove (3) and (4) by generalizing Nica and Speicher's computation of the distributions of commutators and anti-commutators of free random variables [16] to the infinitesimal free case.

PROOF.   (1) Let $x := \sum_{i=1}^k b_i a_i b_i^*$. Since B is symmetric nonnegative definite matrix, we can define $A = (\alpha_{ij})_{i,j=1}^k := \sqrt{B}$. Then $\tau(b_i^* b_j) = \sum_m \alpha_{im}\alpha_{mj}$. For $n \in \mathbb{N}$ we have

$$
\begin{aligned}
\omega(x^n) &= \sum_{i_1, \ldots, i_n} \omega(b_{i_1} a_{i_1} b_{i_1}^* b_{i_2} a_{i_2} b_{i_2}^* \cdots b_{i_n} a_{i_n} b_{i_n}^*) \\
&= \sum_{i_1, \ldots, i_n} \tau(b_{i_1}^* b_{i_2})\tau(b_{i_2}^* b_{i_3}) \cdots \tau(b_{i_n}^* b_{i_1})\omega(a_{i_1} \cdots a_{i_n}) \\
&= \sum_{i_1, \ldots, i_n, m_1, \ldots, m_n} \alpha_{i_1 m_1}\alpha_{m_1 i_2}\alpha_{i_2 m_2}\alpha_{m_2 i_3} \cdots \alpha_{i_n m_n}\alpha_{m_n i_1}\omega(a_{i_1} \cdots a_{i_n}) \\
&= \sum_{m_1, \ldots, m_n} \omega\left(\left(\sum_{i_1} \alpha_{m_n i_1} a_{i_1} \alpha_{i_1 m_1}\right)\left(\sum_{i_2} \alpha_{m_1 i_2} a_{i_2} \alpha_{i_2 m_2}\right)\right. \\
&\qquad\qquad\qquad\qquad\qquad \left. \cdots \left(\sum_{i_n} \alpha_{m_{n-1} i_n} a_{i_n} \alpha_{i_n m_n}\right)\right)
\end{aligned}
$$

$$\begin{aligned}
&= \sum_{m_1,\ldots,m_n} \omega(\gamma_{m_n m_1} \gamma_{m_1 m_2} \cdots \gamma_{m_{n-1} m_n}) \\
&= (\mathrm{Tr}_k \otimes \omega)(G^n),
\end{aligned} \tag{3.33}$$

where $G = (\gamma_{\ell m})_{\ell,m=1}^k := A \operatorname{diag}(a_1,\ldots,a_k) A \in (M_k(\mathbb{C}) \otimes \mathcal{C})_{\mathrm{sa}}$. Corollary 2.9 implies that $\mathrm{EV}(x) = \mathrm{EV}(G)$.

(2) The claim follows from

$$\begin{aligned}
\omega\left(\left(\sum_{i=1}^k a_i b_i a_i^*\right)^n\right) &= \sum_{i_1,\ldots,i_n} \omega(a_{i_1} b_{i_1} a_{i_1}^* a_{i_2} b_{i_2} a_{i_2}^* \cdots a_{i_n} b_{i_n} a_{i_n}^*) \\
&= \sum_{i_1,\ldots,i_n} \omega(a_{i_1} \tau(b_{i_1}) a_{i_1}^* a_{i_2} \tau(b_{i_2}) a_{i_2}^* \cdots a_{i_n} \tau(b_{i_n}) a_{i_n}^*) \\
&= \omega\left(\left(\sum_{i=1}^k \tau(b_i) a_i a_i^*\right)^n\right)
\end{aligned} \tag{3.34}$$

and from Corollary 2.9.

(3) We may assume that $\tau(b) \neq 0$ since the general case can be covered by approximation. Let $c_1 := ab, c_2 := ba$. Then

$$\omega((ab+ba)^n) = \sum_{(i_1,\ldots,i_n) \in \{1,2\}^n} \omega(c_{i_1} \cdots c_{i_n}). \tag{3.35}$$

Given $\mathbf{i} = (i_1,\ldots,i_n) \in \{1,2\}^n$ the set of *ascents* is defined by

$$\mathrm{asc}(\mathbf{i}) := \{1 \le m \le n \mid (i_m, i_{m+1}) = (1,2)\} \tag{3.36}$$

with the convention that $i_{n+1} = i_1$. Then

$$\omega(c_{i_1} \cdots c_{i_n}) = \omega(a^n) \tau(b)^n \alpha^{2\# \mathrm{asc}(\mathbf{i})}, \tag{3.37}$$

where $\alpha := \tau(b^2)^{1/2}/\tau(b)$. Let $\mathcal{NC}_{1,2;1}(n)$ be the set of noncrossing partitions of $\{1,\ldots,n\}$ such that every block has cardinality 1 (singleton) or 2 (pair block), each singleton has depth 0 or 1 and each pair block has depth 0. To each $\mathbf{i} = (i_1,\ldots,i_n) \in \{1,2\}^n$ such that $i_1 = 1$ we associate $\pi(\mathbf{i}) \in \mathcal{NC}_{1,2;1}(n)$ in the following procedure (with the convention that $i_{n+1} = i_1 \ (= 1)$):

- If $i_m = i_{m+1}$ and $1 \le m \le n$ then $\{m\}$ is a singleton of $\pi(\mathbf{i})$;

- If $(i_m, i_{m+1}) = (1,2)$ and $1 \le m \le n-1$ then there exists a unique minimal integer $m+1 \le \ell \le n$ such that $(i_\ell, i_{\ell+1}) = (2,1)$. Then $\{m, \ell\}$ is a pair block of $\pi(\mathbf{i})$.

The correspondence $\{(i_1,\ldots,i_n) \in \{1,2\}^n \mid i_1 = 1\} \to \mathcal{NC}_{1,2;1}(n)$ is bijective and $\# \mathrm{asc}(\mathbf{i}) = \# \mathrm{pair}(\pi(\mathbf{i}))$. Then

$$\sum_{\substack{(i_1,\ldots,i_n) \in \{1,2\}^n \\ i_1 = 1}} \omega(c_{i_1} \cdots c_{i_n}) = \omega(a^n) \tau(b)^n \sum_{\pi \in \mathcal{NC}_{1,2;1}(n)} \alpha^{2\# \mathrm{pair}(\pi)}$$
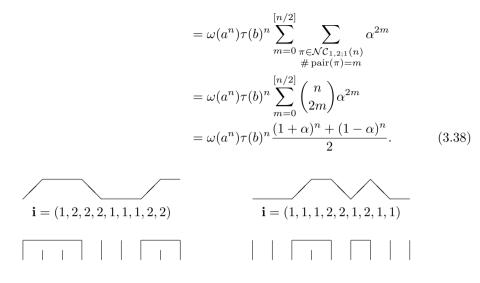
$$= \omega(a^n)\tau(b)^n \sum_{m=0}^{[n/2]} \sum_{\substack{\pi \in \mathcal{NC}_{1,2;1}(n) \\ \# \operatorname{pair}(\pi)=m}} \alpha^{2m}$$

$$= \omega(a^n)\tau(b)^n \sum_{m=0}^{[n/2]} \binom{n}{2m} \alpha^{2m}$$

$$= \omega(a^n)\tau(b)^n \frac{(1+\alpha)^n + (1-\alpha)^n}{2}. \tag{3.38}$$

$\mathbf{i} = (1, 2, 2, 2, 1, 1, 1, 2, 2)$ $\mathbf{i} = (1, 1, 1, 2, 2, 1, 2, 1, 1)$

Figure 1. $\pi(1, 2, 2, 2, 1, 1, 1, 2, 2)$.   Figure 2. $\pi(1, 1, 1, 2, 2, 1, 2, 1, 1)$.

We then compute the sum over $\mathbf{i}$ such that $i_1 = 2$. Let $i \mapsto i^\star$ be the flip of 1 and 2. Then the map $\mathbf{i} \mapsto \mathbf{i}^\star := (i_1^\star, \ldots, i_n^\star)$ defines an involution on $\{1,2\}^n$ whose restriction to $\{(i_1, \ldots, i_n) \in \{1,2\}^n \mid i_1 = 1\}$ is a bijection onto $\{(i_1, \ldots, i_n) \in \{1,2\}^n \mid i_1 = 2\}$. Moreover, it holds that

$$\omega(c_{i_1} \cdots c_{i_n}) = \omega(c_{i_1^\star} \cdots c_{i_n^\star}). \tag{3.39}$$

Thus

$$\sum_{\substack{(i_1,\ldots,i_n) \in \{1,2\}^n \\ i_1 = 2}} \omega(c_{i_1} \cdots c_{i_n}) = \sum_{\substack{(i_1,\ldots,i_n) \in \{1,2\}^n \\ i_1 = 1}} \omega(c_{i_1} \cdots c_{i_n}). \tag{3.40}$$

Therefore we conclude that

$$\omega((ab + ba)^n) = \omega(a^n)\tau(b)^n((1+\alpha)^n + (1-\alpha)^n)$$
$$= \omega(a^n)(p^n + q^n), \tag{3.41}$$

so Corollary 2.9 implies the conclusion.

(4) Let $d_1 := ab, d_2 := -ba$ as before. Then

$$\omega((\mathrm{i}[a,b])^n) = \sum_{(i_1,\ldots,i_n) \in \{1,2\}^n} \mathrm{i}^n \omega(d_{i_1} \cdots d_{i_n}). \tag{3.42}$$

We show that $\omega((\mathrm{i}[a,b])^n) = 0$ when n is odd. If we apply the involution $\mathbf{i} \mapsto \mathbf{i}^\star$ defined in the proof of (3), then

$$\omega(d_{i_1} \cdots d_{i_n}) = (-1)^n \omega(d_{i_1^\star} \cdots d_{i_n^\star}). \tag{3.43}$$

Therefore if n is odd then

$$\omega((\mathrm{i}[a,b])^n) = \sum_{(i_1,\dots,i_n)\in\{1,2\}^n} \mathrm{i}^n \omega(d_{i_1}\cdots d_{i_n})$$

$$= \sum_{\substack{(i_1,\dots,i_n)\in\{1,2\}^n \\ i_1=1}} \mathrm{i}^n \omega(d_{i_1}\cdots d_{i_n}) + \sum_{\substack{(i_1,\dots,i_n)\in\{1,2\}^n \\ i_1=1}} \mathrm{i}^n \omega(d_{i_1^\star}\cdots d_{i_n^\star})$$

$$= 0. \tag{3.44}$$

Hereafter we may assume that n is even. Let $\mathrm{insing}(\pi)$ be the set of inner singletons of $\pi \in \mathcal{NC}_{1,2;1}(n)$, that is, the set of singletons with depth 1. Given $\mathbf{i} = (i_1,\dots,i_n) \in \{1,2\}^n$, it holds that

$$\omega(d_{i_1}\cdots d_{i_n}) = (-1)^{\#\,\mathrm{insing}(\pi(\mathbf{i}))+\#\,\mathrm{pair}(\pi(\mathbf{i}))}\omega(a^n)\tau(b)^n\alpha^{2\#\,\mathrm{pair}(\pi(\mathbf{i}))}, \tag{3.45}$$

where α is the number defined in the proof of (3). For fixed integers $0 \le m \le n/2, 0 \le \ell \le n-2m$, we have that

$$\#\{\pi \in \mathcal{NC}_{1,2;1}(n) \mid \#\,\mathrm{pair}(\pi) = m, \#\,\mathrm{insing}(\pi) = \ell\} = \binom{\ell+m-1}{m-1}\binom{n-m-\ell}{m}, \tag{3.46}$$

which is to be understood as 1 when $m = \ell = 0$ and 0 when $m = 0, \ell > 0$. This is because there are $\binom{\ell+m-1}{m-1}$ ways to place ℓ unlabeled singletons inside m labeled pair blocks, and then there are $\binom{n-m-\ell}{m}$ ways to place the other $n-2m-\ell$ unlabeled singletons outside the m labeled pair blocks (the number of ways to distribute $n-2m-\ell$ unlabeled balls into $m+1$ labeled boxes). Thus

$$\sum_{\substack{(i_1,\dots,i_n)\in\{1,2\}^n \\ i_1=1}} \mathrm{i}^n \omega(d_{i_1}\cdots d_{i_n})$$

$$= \mathrm{i}^n \omega(a^n)\tau(b)^n \sum_{m=0}^{n/2}(-\alpha^2)^m \sum_{\ell=0}^{n-2m} \binom{\ell+m-1}{m-1}\binom{n-m-\ell}{m}(-1)^\ell, \tag{3.47}$$

and by lemma 3.13, we have that

$$\sum_{(i_1,\dots,i_n)\in\{1,2\}^n} \mathrm{i}^n \omega(d_{i_1}\cdots d_{i_n}) = \mathrm{i}^n \omega(a^n)\tau(b)^n \sum_{m=0}^{n/2}(-\alpha^2)^m \binom{n/2}{m}$$

$$= (-1)^{n/2}\omega(a^n)\tau(b)^n(1-\alpha^2)^{n/2}$$

$$= \omega(a^n)r^n. \tag{3.48}$$

Since we know (3.43), we get

$$\omega((\mathrm{i}[a,b])^n) = 2r^n\omega(a^n) \tag{3.49}$$

for even n. If $\lambda_1, \lambda_2,\dots$ are the eigenvalues of a then for every nonnegative integer n

$$\omega((\mathrm{i}[a,b])^n) = (r\lambda_1)^n + (r\lambda_2)^n + \cdots + (-r\lambda_1)^n + (-r\lambda_2)^n + \cdots, \tag{3.50}$$

showing the conclusion by Corollary 2.9. ◻

4. The link with random matrices.

In this section, we shall introduce and prove asymptotic cyclic monotone independence of Haar invariant random matrices which have limiting compact distributions and random matrices which have limiting distributions with respect to the normalized trace. Theorem 4.1 is essentially equivalent to [**21**, Lemma 3.1, Lemma 3.2] by Shlyakhtenko. We supply a proof of this theorem using the *Weingarten calculus*. Actually in our proof we can remove the assumption of some norm boundedness in Lemma 3.1 of Shlyakhtenko, and so we can unify the proofs of [**21**, Lemma 3.1, Lemma 3.2]. Moreover, we can prove the almost sure asymptotic independence and almost sure convergence of discrete eigenvalues, still without the norm boundedness in the trace class setting. We generalize the results to the compact setup, but then we need the norm boundedness.

First, we introduce the tool called the Weingarten calculus, summarizing results in [**11**] (see also the references [**9**] and [**17**, Lecture 23]).

4.1. The Weingarten calculus.

Let S_I be the symmetric group acting on a finite set I and in particular let S_k be the symmetric group $S_{\{1,2,\ldots,k\}}$. The identity of S_k is denoted by 1_k. Let $\mathbb{E}$ be an expectation in a probability space and let $U = (u_{ij})_{i,j=1}^n$ be a normalized Haar unitary random matrix, i.e. the law of U is the normalized Haar measure on the unitary group U_n. Let $\mathcal{E}$ be a linear map of $M_n(\mathbb{C})^{\otimes k}$ defined by

$$\mathcal{E}(A) = \mathbb{E}[U^{\otimes k} A (U^*)^{\otimes k}]. \tag{4.1}$$

Let $\{\delta_\sigma\}_{\sigma \in S_k}$ be the canonical basis of $\mathbb{C}[S_k]$ and let Φ be a linear map from $M_n(\mathbb{C})^{\otimes k}$ to $\mathbb{C}[S_k]$ defined by

$$\Phi(A) = \sum_{\sigma \in S_k} \mathrm{Tr}_{M_n(\mathbb{C})^{\otimes k}}(\rho(\sigma)^* A)\delta_\sigma, \tag{4.2}$$

where $\rho : S_k \to M_n(\mathbb{C})^{\otimes k}$ is the natural representation

$$(\rho(\sigma))(v_1 \otimes \cdots \otimes v_k) := v_{\sigma^{-1}(1)} \otimes \cdots \otimes v_{\sigma^{-1}(k)}, \qquad v_i \in \mathbb{C}^n. \tag{4.3}$$

For $\sigma \in S_k$ and $n \geq k$, we define the Weingarten function $\mathrm{Wg}(\sigma, n)$ as

$$\mathrm{Wg}(\sigma, n) = \mathbb{E}[u_{11} \cdots u_{kk} \overline{u}_{1\sigma(1)} \cdots \overline{u}_{k\sigma(k)}]. \tag{4.4}$$

It follows from [**11**, Equation (9)] that the function $\sigma \mapsto \mathrm{Wg}(\sigma, n), S_k \to \mathbb{C}$ is a linear combination of the characters of irreducible representations of S_k, and so $\mathrm{Wg}(\tau\sigma\tau^{-1}, n) = \mathrm{Wg}(\sigma, n)$. The conjugate classes of a symmetric group are determined by the structure of cycle decomposition. Since σ and σ^{-1} have the same structure of cycle decomposition, we have $\mathrm{Wg}(\sigma^{-1}, n) = \mathrm{Wg}(\sigma, n)$ for every $\sigma \in S_k$.

What is important is asymptotics of the Weingarten function for large n. Let Cat_p be the Catalan number

$$\mathrm{Cat}_p := \frac{(2p)!}{p!(p+1)!}, \qquad p \in \mathbb{N}. \tag{4.5}$$

For $\sigma \in S_k$ let $|\sigma|$ be the length function, that is, the minimal number of transpositions to express σ as the product of them. Let $\sigma = c_1 \cdots c_{\ell(\sigma)}$ be the cycle decomposition of $\sigma \in S_k$ and then let

$$\mathrm{M\ddot{o}b}(\sigma) := \prod_{i=1}^{\ell(\sigma)} (-1)^{|c_i|} \mathrm{Cat}_{|c_i|}. \tag{4.6}$$

Note that $|\sigma| = \sum_{i=1}^{\ell(\sigma)} |c_i|$ and $\ell(\sigma) = k - |\sigma|$. Then it is known that [**11**, Corollary 2.7]

$$\mathrm{Wg}(\sigma, n) = n^{-k-|\sigma|}(\mathrm{M\ddot{o}b}(\sigma) + O(n^{-2})). \tag{4.7}$$

The Weingarten function can obviously be regarded as the element of $\mathbb{C}[S_k]$

$$\mathrm{Wg} = \sum_{\sigma \in S_k} \mathrm{Wg}(\sigma, n) \delta_\sigma. \tag{4.8}$$

It was shown in [**11**] that for all $A, B \in M_n(\mathbb{C})^{\otimes k}$

$$\Phi(A\mathcal{E}(B)) = \Phi(A)\Phi(B)\mathrm{Wg}. \tag{4.9}$$

In this paper we consider $A = A_1 \otimes \cdots \otimes A_k$, $B = B_1 \otimes \cdots \otimes B_k$ where $A_i, B_i \in M_n(\mathbb{C})$, $i = 1, \ldots, k$. For a cycle $c = (i_1 i_2 \cdots i_m)$, let A_c be the product $A_{i_1} \cdots A_{i_m}$ and let Tr_σ be the product of traces Tr_n according to the cycle decomposition $\sigma = c_1 c_2 \cdots c_{\ell(\sigma)}$,

$$\mathrm{Tr}_\sigma(A_1, \ldots, A_k) = \prod_{i=1}^{\ell(\sigma)} \mathrm{Tr}_n(A_{c_i}). \tag{4.10}$$

For example if $\sigma = (1,3)(2)$ then $\mathrm{Tr}_\sigma(A_1, A_2, A_3) = \mathrm{Tr}_n(A_1 A_3) \mathrm{Tr}_n(A_2)$. Note that this notation is well-defined thanks to the cyclic property of the trace. We can then show that

$$\mathrm{Tr}_{M_n(\mathbb{C})^{\otimes k}}(\rho(\sigma)^* A) = \mathrm{Tr}_\sigma(A_1, \ldots, A_k). \tag{4.11}$$

Then (4.9) reads, for every $\sigma \in S_k$,

$$\begin{aligned} &\mathbb{E}[\mathrm{Tr}_\sigma(A_1 U B_1 U^*, \ldots, A_k U B_k U^*)] \\ &= \sum_{\substack{\sigma_1, \sigma_2, \sigma_3 \in S_k \\ \sigma_1 \sigma_2 \sigma_3 = \sigma}} \mathrm{Tr}_{\sigma_1}(A_1, \ldots, A_k) \mathrm{Tr}_{\sigma_2}(B_1, \ldots, B_k) \mathrm{Wg}(\sigma_3, n). \end{aligned} \tag{4.12}$$

This formula is the main tool of our analysis below.

4.2. Asymptotic cyclic monotone independence of random matrices.

We prove asymptotic cyclic monotone independence on average in the trace class setup. Note that we can replace $(\{A_1, \ldots, A_k\}, \{U B_1 U^*, \ldots, U B_\ell U^*\})$ with

$(\{UA_1U^*,\ldots,UA_kU^*\},\{B_1,\ldots,B_\ell\})$ in the following asymptotic results.

THEOREM 4.1. *Let $U = U(n)$ be an $n \times n$ Haar unitary and $A_i = A_i(n), B_j = B_j(n), i = 1,\ldots,k,\ j = 1,\ldots,\ell$ be $n \times n$ random matrices. Suppose that*

(1) *$((A_1,\ldots,A_k), \mathbb{E}\otimes\mathrm{Tr}_n)$ converges in distribution to a k-tuple of trace class operators as $n \to \infty$,*

(2) *$((B_1,\ldots,B_\ell), \mathbb{E}\otimes\mathrm{tr}_n)$ converges in distribution to an ℓ-tuple of elements in a non-commutative probability space as $n \to \infty$,*

(3) *$\{A_1,\ldots,A_k\},\{B_1,\ldots,B_\ell\}, U$ are independent.*

Then the pair $(\{A_i\}_{i=1}^k, \{UB_jU^*\}_{j=1}^\ell)$ is asymptotically cyclically monotone with respect to $(\mathbb{E}\otimes\mathrm{Tr}_n, \mathbb{E}\otimes\mathrm{tr}_n)$.

PROOF. We may assume that $\ell = k$ since otherwise we may add 0's to A_i's or the identity matrices to B_j's. For simplicity UB_iU^* is abbreviated to B_i. Note that it suffices to show that the expectation

$$\mathbb{E}[\mathrm{Tr}_n(A_1B_1\cdots A_kB_k)] \qquad (4.13)$$

factorizes following cyclic monotone independence, since a general monomial can be written in the form $B_0'A_1'\cdots A_m'B_{m+1}'$ where $A_i' \in \mathrm{alg}\{A_1,\ldots,A_k\}$ and $B_i' \in \mathrm{alg}\{1_n, B_1,\ldots,B_k\}$. By the traciality, the distribution of $B_0'A_1'B_1'\cdots A_m'B_{m+1}'$ with respect to the non-normalized trace is the same as that of $A_1'B_1'\cdots A_m'B_{m+1}'B_0'$, so every monomial reduces to (4.13). Also, a concrete construction of a limiting non-commutative probability space with a tracial weight can be given by the cyclic monotone product in Definition 3.2. Thus, it suffices to show the factorization of the expectation (4.13).

Let Z be the circular permutation, i.e. $Z = (1\cdots k)$. In this context, by the Weingarten formula (4.12) we have

$$\mathbb{E}[\mathrm{Tr}_n(A_1B_1\cdots A_kB_k)]$$
$$= \sum_{\substack{\sigma_1,\sigma_2,\sigma_3\in S_k\\ \sigma_1\sigma_2\sigma_3=Z}} \mathbb{E}[\mathrm{Tr}_{\sigma_1}(A_1,\ldots,A_k)]\mathbb{E}[\mathrm{Tr}_{\sigma_2}(B_1,\ldots,B_k)]\mathrm{Wg}(\sigma_3,n). \qquad (4.14)$$

Next we make a decay analysis. Since $(A_1,\ldots,A_k)$ has a limiting trace class distribution, then the leading behavior of $\mathrm{Tr}_{\sigma_1}(A_1,\ldots,A_k)$ is $O(n^0)$. Since $(B_1,\ldots,B_k)$ has a limiting distribution with respect to the normalized trace, the behavior of $\mathrm{Tr}_{\sigma_2}(B_1,\ldots,B_k)$ is $O(n^{k-|\sigma_2|})$. Finally, we know that $\mathrm{Wg}(\sigma_3,n)$ behaves as $O(n^{-k-|\sigma_3|})$. Therefore, for a triple $\sigma_1,\sigma_2,\sigma_3$ such that $\sigma_1\sigma_2\sigma_3 = Z$, the contribution of the summand is $O(n^{-|\sigma_2|-|\sigma_3|})$. Therefore, the asymptotics is driven by the summands for which $|\sigma_2| = |\sigma_3| = 0$, i.e. $\sigma_2 = \sigma_3 = 1_k$. This forces $\sigma_1 = Z$, and there is only one such summand. As a conclusion,

$$\mathbb{E}[\mathrm{Tr}_n(A_1B_1\cdots A_kB_k)] = \mathbb{E}[\mathrm{Tr}_n(A_1A_2\cdots A_k)]\prod_{i=1}^k \mathbb{E}[\mathrm{tr}_n(B_i)] + O(n^{-1}). \qquad (4.15)$$
$$\square$$

4.3. Almost sure convergence.

Some further analysis shows that

$$\mathrm{Cov}(\mathrm{Tr}(A_1 B_1 \cdots A_k B_k)) = O(n^{-1}).$$

Interestingly, this is quite different from classical random matrix models inspired from free probability theory [**9**], [**11**], where the covariance behaves rather like $O(n^{-2})$. As it is classically known, a behavior of $O(n^{-2})$, summable in n, allows to prove the almost sure convergence of the traces of random matrix models. Here, since we do not have the behavior $O(n^{-2})$, we need to investigate other functionals.

From now on, we assume that the matrices A_i, B_i are deterministic, but by conditioning A_i, B_i to be constant, we can generalize the results to the case when A_i, B_i are random matrices independent of the Haar unitary U such that $((A_1, \ldots, A_k), \mathrm{Tr}_n)$ and $((B_1, \ldots, B_k), \mathrm{tr}_n)$ almost surely converge in distributions to deterministic elements. For example this allows us to take $B_1 = \cdots = B_\ell$ to be a GUE G, and in this case we do not need to rotate G by a Haar unitary since G is rotationally invariant.

LEMMA 4.2.   *Let $A_i = A_i(n), B_i = B_i(n), i = 1, \ldots, k$, be $n \times n$ deterministic matrices and let $U = U(n)$ be a Haar unitary random matrix. Suppose that*

(1) *$((A_1, \ldots, A_k), \mathrm{Tr}_n)$ converges in distribution to a k-tuple of trace class operators as $n \to \infty$,*

(2) *$((B_1, \ldots, B_k), \mathrm{tr}_n)$ converges in distribution to a k-tuple of elements in a non-commutative probability space as $n \to \infty$.*

Then

$$\mathbb{E}\Big[\big| \mathrm{Tr}_n(A_1 U B_1 U^* \cdots A_k U B_k U^*) - \mathbb{E}[\mathrm{Tr}_n(A_1 U B_1 U^* \cdots A_k U B_k U^*)] \big|^4 \Big] = O(n^{-2}).$$

PROOF.   We prove a more general result, namely, instead of taking k matrices A_i and k matrices B_i, we take $4k$ matrices for each kind satisfying the same assumption. Moreover, for notational simplicity $U B_i U^*$ is abbreviated to B_i. Let

$$X_i = \mathrm{Tr}_n\big(A_{(i-1)k+1} B_{(i-1)k+1} \cdots A_{(i-1)k+k} B_{(i-1)k+k} \big), \qquad i = 1, 2, 3, 4, \qquad (4.16)$$

$$\mathring{X}_i = X_i - \mathbb{E}[X_i], \tag{4.17}$$

which are complex-valued random variables. We prove that

$$\mathbb{E}[\mathring{X}_1 \mathring{X}_2 \mathring{X}_3 \mathring{X}_4] = O(n^{-2}). \tag{4.18}$$

This implies the lemma if we define X_2, X_3, X_4 such that $X_3 = X_1, X_4 = X_2 = \overline{X_1}$. The last condition $X_2 = \overline{X_1}$ can be realized by taking $A_{k+1} = A_k^*, B_{k+1} = B_{k-1}^*, \ldots, A_{2k} = A_1^*, B_{2k} = B_k^*$, since

$$\overline{X_1} = \mathrm{Tr}_n((A_1 B_1 \cdots A_k B_k)^*) = \mathrm{Tr}_n(A_k^* B_{k-1}^* A_{k-1}^* \cdots A_2^* B_1^* A_1^* B_k^*). \tag{4.19}$$

We shall denote by I_i the interval $\{(i-1)k+1, (i-1)k+2, \ldots, (i-1)k+k\}$, by

Z_i the cyclic permutation $((i-1)k+1, (i-1)k+2, \ldots, (i-1)k+k)$ of S_{I_i} and by $Z^{\cup 4}$ the permutation $Z_1 Z_2 Z_3 Z_4$ of S_{4k}. Thanks to the Weingarten formula (4.12) we have formulas to compute the moments $\mathbb{E}[\mathring{X}_1 \mathring{X}_2 \mathring{X}_3 \mathring{X}_4]$. If we expand this mixed moment, we obtain 16 (products of) expectations. Our notation is

$$\mathbb{E}[\mathring{X}_1 \mathring{X}_2 \mathring{X}_3 \mathring{X}_4] = \sum_{A \subset \{1,2,3,4\}} \mathbb{E}_A, \tag{4.20}$$

$$\mathbb{E}_A = (-1)^{\# A} \mathbb{E}\left[\prod_{i \in A} X_i\right] \prod_{i \in A^c} \mathbb{E}[X_i]. \tag{4.21}$$

For example, our notation means $\mathbb{E}_{\{1,3,4\}} = -\mathbb{E}[X_1 \mathbb{E}[X_2] X_3 X_4]$. In this specific example, the Weingarten formula boils down to a sum over permutations in $S_{I_1 \cup I_3 \cup I_4}$ for the evaluation of $\mathbb{E}[X_1 X_3 X_4]$, and those in S_{I_2} for the evaluation of $\mathbb{E}[X_2]$.

It follows from the Weingarten formula that for each $A \subset \{1,2,3,4\}$ there exists a number $f_A(\sigma_1, \sigma_2, n)$ such that

$$\mathbb{E}_A = \sum_{\substack{\sigma_1, \sigma_2, \sigma_3 \in S_{4k} \\ \sigma_1 \sigma_2 \sigma_3 = Z^{\cup 4}}} \mathrm{Tr}_{\sigma_1}(A_1, \ldots, A_{4k}) \, \mathrm{Tr}_{\sigma_2}(B_1, \ldots, B_{4k}) f_A(\sigma_1, \sigma_2, n). \tag{4.22}$$

Note that σ_3 may be omitted but it is written for clearer understanding. The number f_A is either zero, or a product of (two, three or four) Weingarten functions. Again, instead of going through heavy notation that would perhaps not ease the understanding, let us describe $f_A(\sigma_1, \sigma_2, n)$ in the generic case $A = \{1,3,4\}$. Recall that $\mathbb{E}_A = -\mathbb{E}[X_2]\mathbb{E}[X_1 X_3 X_4]$. The product of Weingarten formulas (4.12) for $\sigma = Z_2$ and for $\sigma = Z_1 Z_3 Z_4$ only gives permutations of $S_{I_2} \times S_{I_1 \cup I_3 \cup I_4}$, and so, for $\sigma_1, \sigma_2, \sigma_3 \in S_{4k}$ such that $\sigma_1 \sigma_2 \sigma_3 = Z^{\cup 4}$,

$$f_A(\sigma_1, \sigma_2, n) = \begin{cases} -\mathrm{Wg}(\sigma_{3|I_2}, n) \, \mathrm{Wg}(\sigma_{3|I_1 \cup I_3 \cup I_4}, n), \\ \qquad \text{if } \sigma_1, \sigma_2 \text{ leave } I_2 \text{ and } I_1 \cup I_3 \cup I_4 \text{ invariant}, \\ 0, \qquad \text{otherwise}. \end{cases} \tag{4.23}$$

The general case is alike, and hence we can define a function $f_A(\sigma_1, \sigma_2, n)$ for every subset A. This gives us the expression

$$\mathbb{E}[\mathring{X}_1 \mathring{X}_2 \mathring{X}_3 \mathring{X}_4] = \sum_{\substack{\sigma_1, \sigma_2, \sigma_3 \in S_{4k} \\ \sigma_1 \sigma_2 \sigma_3 = Z^{\cup 4}}} \mathrm{Tr}_{\sigma_1}(A_1, \ldots, A_{4k}) \, \mathrm{Tr}_{\sigma_2}(B_1, \ldots, B_{4k}) f(\sigma_1, \sigma_2, n), \tag{4.24}$$

where

$$f(\sigma_1, \sigma_2, n) = \sum_{A \subset \{1,2,3,4\}} f_A(\sigma_1, \sigma_2, n). \tag{4.25}$$

As seen above, $f(\sigma_1, \sigma_2, n)$ is a signed sum of products of Wg functions on various permutation groups.

By assumption, $\mathrm{Tr}_{\sigma_1}(A_1, \ldots, A_{4k}) = O(1)$ as $n \to \infty$ and $\mathrm{Tr}_{\sigma_2}(B_1, \ldots, B_{4k}) =$

$O(n^{\ell(\sigma_2)}) = O(n^{4k-|\sigma_2|})$. From (4.7) it follows that

$$\text{Tr}_{\sigma_2}(B_1, \ldots, B_{4k}) f(\sigma_1, \sigma_2, n) = O(n^{-|\sigma_2|-|\sigma_3|}). \tag{4.26}$$

We show that actually

$$\text{Tr}_{\sigma_2}(B_1, \ldots, B_{4k}) f(\sigma_1, \sigma_2, n) = O(n^{-2}), \tag{4.27}$$

or equivalently

$$f(\sigma_1, \sigma_2, n) = O(n^{-4k+|\sigma_2|-2}) \tag{4.28}$$

for each fixed σ_1, σ_2 (and σ_3) $\in S_{4k}$ (such that $\sigma_1\sigma_2\sigma_3 = Z^{\cup 4}$).

Given $\sigma_1, \sigma_2, \sigma_3 \in S_{4k}$ such that $\sigma_1\sigma_2\sigma_3 = Z^{\cup 4}$, we define an equivalence relation on $\{1, \ldots, 4k\}$: $i \sim j$ if there exists $\tau \in \text{Grp}\langle \sigma_1, \sigma_2, \sigma_3 \rangle$ such that $\tau(i) = j$. Since this group contains $Z^{\cup 4}$, every interval I_i must be a subset of some equivalence class. Then, the permutations $\sigma_1, \sigma_2, \sigma_3$ associate a set partition $\pi(\sigma_1, \sigma_2, \sigma_3) = \{P_1, \ldots, P_m\}$ of $\{1, 2, 3, 4\}$ such that the subsets of $\{1, \ldots, 4k\}$,

$$\bigcup_{i \in P_1} I_i, \bigcup_{i \in P_2} I_i, \ldots, \tag{4.29}$$

are exactly the equivalence classes generated by the actions of $\sigma_1, \sigma_2, \sigma_3$. A subset A of $\{1, 2, 3, 4\}$ also associates the set partition $\pi(A) = \{A, \{b\} : b \in A^c\}$ of $\{1, 2, 3, 4\}$. Then only f_A for which $\pi(A)$ is coarser than or equal to $\pi(\sigma_1, \sigma_2, \sigma_3)$ contributes to f in the sum (4.25) and the other f_A's are zero. We distinguish several cases according to the set partition associated to $\sigma_1, \sigma_2, \sigma_3$.

(i) $\pi(\sigma_1, \sigma_2, \sigma_3) = \{\{1, 2, 3, 4\}\}$, or equivalently, the group $\text{Grp}\langle \sigma_1, \sigma_2, \sigma_3 \rangle$ acts on $\{1, 2, \ldots, 4k\}$ transitively. In this case, f_A is zero unless $A = \{1, 2, 3, 4\}$, and so $f = f_{\{1,2,3,4\}} = \text{Wg}(\sigma_3, n)$. We have to exclude $|\sigma_2| + |\sigma_3| = 0$, as the condition $\sigma_1\sigma_2\sigma_3 = Z^{\cup 4}$ contradicts transitivity. We can also exclude $|\sigma_2| + |\sigma_3| = 1$. Indeed this means that one is the identity, and the other one is a transposition, and again the condition $\sigma_1\sigma_2\sigma_3 = Z^{\cup 4}$ is incompatible with a transitive action. Therefore, $|\sigma_2| + |\sigma_3| \geq 2$, and so $\text{Tr}_{\sigma_2}(B_1, \ldots, B_{4k}) f(\sigma_1, \sigma_2, n) = O(n^{-2})$.

(ii) $\pi(\sigma_1, \sigma_2, \sigma_3)$ is a pair partition, e.g. $\{\{1, 2\}, \{3, 4\}\}$. In this case again we have $f = f_{\{1,2,3,4\}} = \text{Wg}(\sigma_3, n)$, and from a similar reasoning we must have $|\sigma_2| + |\sigma_3| \geq 2$ and hence $\text{Tr}_{\sigma_2}(B_1, \ldots, B_{4k}) f(\sigma_1, \sigma_2, n) = O(n^{-2})$.

In the other cases some nice cancellation occurs between Wg functions.

(iii) $\pi(\sigma_1, \sigma_2, \sigma_3)$ has two blocks with cardinality 1 and 3, say $\{\{1, 3, 4\}, \{2\}\}$. The equivalence classes are $I_1 \cup I_3 \cup I_4$ and I_2. The only indices A's for which f_A is non-zero are $A = \{1, 3, 4\}, \{1, 2, 3, 4\}$. By inspection we see that

$$f_{\{1,3,4\}} = -\text{Wg}(\sigma_{3|I_2}, n) \text{Wg}(\sigma_{3|I_1 \cup I_3 \cup I_4}, n),$$
$$f_{\{1,2,3,4\}} = \text{Wg}(\sigma_3, n).$$

By (4.7) and the multiplicativity (4.6) of Moebius functions, we obtain

$$
\begin{aligned}
f &= f_{\{1,3,4\}} + f_{\{1,2,3,4\}} \\
&= n^{-4k-|\sigma_3|}(-\operatorname{M\ddot{o}b}(\sigma_{3|I_2})\operatorname{M\ddot{o}b}(\sigma_{3|I_1\cup I_3\cup I_4}) + \operatorname{M\ddot{o}b}(\sigma_3) + O(n^{-2})) \\
&= O(n^{-4k-|\sigma_3|-2}).
\end{aligned}
$$

Thus $\operatorname{Tr}_{\sigma_2}(B_1,\ldots,B_{4k})f(\sigma_1,\sigma_2,n) = O(n^{-|\sigma_2|-|\sigma_3|-2})$.

(iv) $\pi(\sigma_1,\sigma_2,\sigma_3)$ has 3 blocks, say $\{\{1\},\{2\},\{3,4\}\}$. The indices A's for which f_A is non-zero are $A = \{3,4\},\{1,3,4\},\{2,3,4\},\{1,2,3,4\}$. We see that

$$
\begin{aligned}
f &= f_{\{3,4\}} + f_{\{1,3,4\}} + f_{\{2,3,4\}} + f_{\{1,2,3,4\}} \\
&= \operatorname{Wg}(\sigma_{3|I_1},n)\operatorname{Wg}(\sigma_{3|I_2},n)\operatorname{Wg}(\sigma_{3|I_3\cup I_4},n) - \operatorname{Wg}(\sigma_{3|I_2},n)\operatorname{Wg}(\sigma_{3|I_1\cup I_3\cup I_4},n) \\
&\quad - \operatorname{Wg}(\sigma_{3|I_1},n)\operatorname{Wg}(\sigma_{3|I_2\cup I_3\cup I_4},n) + \operatorname{Wg}(\sigma_3,n) \\
&= n^{-4k-|\sigma_3|}\big(\operatorname{M\ddot{o}b}(\sigma_{3|I_1})\operatorname{M\ddot{o}b}(\sigma_{3|I_2})\operatorname{M\ddot{o}b}(\sigma_{3|I_3\cup I_4}) - \operatorname{M\ddot{o}b}(\sigma_{3|I_2})\operatorname{M\ddot{o}b}(\sigma_{3|I_1\cup I_3\cup I_4}) \\
&\qquad - \operatorname{M\ddot{o}b}(\sigma_{3|I_1})\operatorname{M\ddot{o}b}(\sigma_{3|I_2\cup I_3\cup I_4}) + \operatorname{M\ddot{o}b}(\sigma_3) + O(n^{-2})\big) \\
&= O(n^{-4k-|\sigma_3|-2}).
\end{aligned}
$$

(v) $\pi(\sigma_1,\sigma_2,\sigma_3) = \{\{1\},\{2\},\{3\},\{4\}\}$, namely, every I_i is invariant under the actions of $\sigma_1,\sigma_2,\sigma_3$. In this case f_A contribute to f for all the 16 subsets $A \subset \{1,2,3,4\}$. By multiplicativity, the dominant contribution to f is the sum of 16 Möbius functions

$$
\pm \operatorname{M\ddot{o}b}(\sigma_{3|I_1})\operatorname{M\ddot{o}b}(\sigma_{3|I_2})\operatorname{M\ddot{o}b}(\sigma_{3|I_3})\operatorname{M\ddot{o}b}(\sigma_{3|I_4})
$$

multiplied by $n^{-4k-|\sigma_3|}$. Exactly half of them have minus signs, so they cancel. Thus $f = O(n^{-4k-|\sigma_3|-2})$.

This concludes the proof of (4.27). □

THEOREM 4.3.   *Let $A_i = A_i(n), B_j = B_j(n), i = 1,\ldots,k, j = 1,\ldots,\ell$ be $n \times n$ deterministic matrices and let $U = U(n)$ be a Haar unitary random matrix. Suppose that*

(1) *$((A_1,\ldots,A_k),\operatorname{Tr}_n)$ converges in distribution to a k-tuple of trace class operators as $n \to \infty$,*

(2) *$((B_1,\ldots,B_\ell),\operatorname{tr}_n)$ converges in distribution to an ℓ-tuple of elements in a non-commutative probability space as $n \to \infty$.*

Then the pair $(\{A_i\}_{i=1}^k, \{UB_jU^*\}_{j=1}^\ell)$ is asymptotically cyclically monotone almost surely with respect to $(\operatorname{Tr}_n,\operatorname{tr}_n)$.

PROOF.   For notational convenience we write UB_iU^* simply as B_i. From the arguments in the proofs of Theorem 4.1, it suffices to show that when $k = \ell$

$$
\lim_{n\to\infty}\operatorname{Tr}_n(A_1B_1\cdots A_kB_k) = \lim_{n\to\infty}\operatorname{Tr}_n(A_1\cdots A_k)\prod_{i=1}^k\lim_{n\to\infty}\operatorname{tr}_n(B_i) \quad \text{a.s.} \tag{4.30}
$$

One may use the standard Borel–Cantelli argument, but a simpler argument is possible. By Lemma 4.2 and by monotone convergence we have

$$\mathbb{E}\left[\sum_{n=1}^{\infty}\left|\operatorname{Tr}_n(A_1 B_1 \cdots A_k B_k) - \mathbb{E}\left[\operatorname{Tr}_n(A_1 B_1 \cdots A_k B_k)\right]\right|^4\right] < \infty, \qquad (4.31)$$

and so

$$\sum_{n=1}^{\infty}\left|\operatorname{Tr}_n(A_1 B_1 \cdots A_k B_k) - \mathbb{E}\left[\operatorname{Tr}_n(A_1 B_1 \cdots A_k B_k)\right]\right|^4 < \infty \quad \text{a.s.,} \qquad (4.32)$$

which implies that $\lim_{n\to\infty}|\operatorname{Tr}_n(A_1 B_1 \cdots A_k B_k) - \mathbb{E}[\operatorname{Tr}_n(A_1 B_1 \cdots A_k B_k)]| = 0$ a.s. By Theorem 4.1 we know that

$$\lim_{n\to\infty}\mathbb{E}[\operatorname{Tr}_n(A_1 B_1 \cdots A_k B_k)] = \lim_{n\to\infty}\operatorname{Tr}_n(A_1 \cdots A_k)\prod_{i=1}^{k}\lim_{n\to\infty}\operatorname{tr}_n(B_i), \qquad (4.33)$$

so we get the conclusion. □

Proposition 2.10 implies the following. Note that we can only conclude that the limiting operator is Hilbert Schmidt.

COROLLARY 4.4.   *Under the assumptions of Theorem 4.3, for any selfadjoint $*$-polynomial $P(x_1,\ldots,x_k,y_1,\ldots,y_\ell)$ such that $P(0,\ldots,0,y_1,\ldots,y_\ell) = 0$, the Hermitian random matrix $P(A_1,\ldots,A_k,UB_1U^*,\ldots,UB_\ell U^*)$ converges in eigenvalues to a selfadjoint Hilbert Schmidt operator almost surely. The limiting eigenvalues can be computed by using cyclic monotone independence.*

Examples of the limiting eigenvalues are computed in Section 5.

Our result implies the almost sure version of Shlyakhtenko's asymptotic infinitesimal freeness.

COROLLARY 4.5.   *Let $A_i = A_i(n), B_j = B_j(n), i = 1,\ldots,k, \; j = 1,\ldots,\ell$ be $n \times n$ deterministic matrices and let $U = U(n)$ be a Haar unitary random matrix. In addition to the assumption (1) of Theorem 4.3, we assume that $((B_1,\ldots,B_\ell),\operatorname{tr}_n)$ converges up to the first order to an ℓ-tuple of elements in an infinitesimal non-commutative probability space (see Definition 3.12). Then for any $*$-polynomial P in the unital noncommutative $*$-polynomial ring $\mathcal{C} := \mathbb{C}[x_1,\ldots,x_k,y_1,\ldots,y_\ell]$ the limits*

$$\tau(P) := \lim_{n\to\infty}\operatorname{tr}_n(P(A_1,\ldots,A_k,UB_1U^*,\ldots,UB_\ell U^*)), \qquad (4.34)$$

$$\tau'(P) := \lim_{n\to\infty}\frac{\operatorname{tr}_n(P(A_1,\ldots,A_k,UB_1U^*,\ldots,UB_\ell U^*)) - \tau(P)}{1/n} \qquad (4.35)$$

exist almost surely, and thus $(\mathcal{C},\tau',\tau)$ is an infinitesimal non-commutative probability space. Moreover, $\{x_i\}_{i=1}^k, \{y_j\}_{j=1}^\ell$ are infinitesimally free with respect to (τ',τ).

PROOF.   We decompose $P = Q + R$ where $Q = Q(\{x_i\}_{i=1}^k, \{y_j\}_{j=1}^\ell)$ and $R =$

$R(\{y_j\}_{j=1}^{\ell})$ such that $Q(0,\ldots,0,y_1,\ldots,y_\ell) = 0$. By Theorem 4.3 $\mathrm{Tr}_n(Q)$ converges almost surely to a finite real number, so $\tau(Q) = 0$. Hence $\tau'(Q) = \lim_{n\to\infty} \mathrm{Tr}_n(Q)$ converges almost surely. Since $((B_1,\ldots,B_\ell),\mathrm{tr}_n)$ converges up to the first order then $\tau(R)$ and $\tau'(R)$ converge too. Therefore the limits $\tau(P), \tau'(P)$ exist.

Let $\mathcal{A} := \mathbb{C}[x_1,\ldots,x_k]_0$ (not containing the unit) and let $\mathcal{B} := \mathbb{C}[y_1,\ldots,y_\ell]$ (containing the unit). Then $\mathrm{Ideal}_\mathcal{B}(\mathcal{A}) \subset \ker(\tau)$ since, as we saw, $\tau(Q) = 0$. Since $(\mathcal{A},\mathcal{B})$ is cyclically monotone by Theorem 4.3, $\mathcal{A},\mathcal{B}$ are infinitesimally free by Proposition 3.10. $\square$

Through the calculation in [**21**] Shlyakhtenko suggested that infinitesimal freeness is applicable to outliers, but a rigorous proof is not obtained yet.

PROBLEM 4.6.   Combining the calculation of Shlyakhtenko [**21**] and our almost sure convergence (and other ideas if needed), is it possible to rigorously prove the phase transition phenomena of outliers found by Baik et al. [**2**], Péché [**18**], and more generally by Benaych-Georges and Nadakuditi [**5**] and Belinschi et al. [**3**]?

4.4.   General compact case.

Actually, we do not need trace class distributions in order to obtain the almost sure convergence of eigenvalues, the compact setup is enough. In this section we denote by $\|\cdot\|$ the operator norm on $M_n(\mathbb{C})$.

THEOREM 4.7.   Let $A_i = A_i(n), B_j = B_j(n), i = 1,\ldots,k, j = 1,\ldots,\ell$ be deterministic $n \times n$ matrices and $U = U(n)$ be an $n \times n$ Haar unitary such that

(1) $A_1,\ldots,A_k$ are Hermitian,

(2) $((A_1,\ldots,A_k),\mathrm{Tr}_n)$ converges in compact distribution to a k-tuple of compact operators $((a_1,\ldots,a_k),\mathrm{Tr}_H)$ as $n \to \infty$ (see Definition 2.6),

(3) $((B_1,\ldots,B_\ell),\mathrm{tr}_n)$ converges in distribution to an ℓ-tuple of elements in a noncommutative probability space as $n \to \infty$,

(4) $\sup_{n\in\mathbb{N}} \|B_i(n)\| < \infty$ for every $i = 1,\ldots,\ell$.

Let $P(x_1,\ldots,x_k,y_1,\ldots,y_\ell)$ be a selfadjoint $*$-polynomial with selfadjoint variables $x_1,\ldots,x_k$ such that $P(0,\ldots,0,y_1,\ldots,y_\ell) = 0$. Then $P(A_1,\ldots,A_k,UB_1U^*,\ldots,UB_\ell U^*)$ converges in eigenvalues to a deterministic compact operator almost surely.

PROOF.   We may assume that $k = \ell$. For simplicity UB_pU^* is abbreviated to B_p. By assumption, every A_p converges in eigenvalues. We can then find some sequence $\{\varepsilon_j\}_{j\geq 1}$ such that $\varepsilon_j \downarrow 0$ and

$$\{\varepsilon_j : j \in \mathbb{N}\} \cap \{|\lambda_i^u(A_p(n))|, \lim_{N\to\infty} |\lambda_i^u(A_p(N))| : i, n \in \mathbb{N}, 1 \leq p \leq k, u \in \{+,-\}\} = \emptyset.$$

Let f_j be a continuous function on $\mathbb{R}$ such that f_j is non-decreasing and

$$f_j(x) = \begin{cases} 0, & |x| < \varepsilon_{j+1}, \\ x, & |x| > \varepsilon_j. \end{cases} \tag{4.36}$$

Let $A_p^{(j)}, a_p^{(j)}$ be the truncations $f_j(A_p), f_j(a_p)$ respectively, so that they are finite rank operators and in particular trace class operators. By the definition of convergence in compact distribution, $((A_1^{(j)}, \ldots, A_k^{(j)}), \mathrm{Tr}_n)$ converges in distribution to the trace class operators $((a_1^{(j)}, \ldots, a_k^{(j)}), \mathrm{Tr}_H)$ as $n \to \infty$. Under such circumstances, for each fixed $j \in \mathbb{N}$ we apply Corollary 4.4 to the pair $(\{A_1^{(j)}, \ldots, A_k^{(j)}\}, \{B_1, \ldots, B_k\})$. Thus, the random eigenvalues of the polynomial $P(A_1^{(j)}, \ldots, A_k^{(j)}, B_1, \ldots, B_k)$, denoted by $\{\lambda_i^{(j)}(n)\}_{i \geq 1}$, converge to some deterministic eigenvalues $\{\lambda_i^{(j)}\}_{i \geq 1} \in \ell^2(\mathbb{R})$ as $n \to \infty$:

$$\lim_{n \to \infty} (\lambda_i^{(j)})^{\pm}(n) = (\lambda_i^{(j)})^{\pm}, \qquad i, j \in \mathbb{N}. \tag{4.37}$$

It is follows by functional calculus that

$$\sup_{n \in \mathbb{N}, 1 \leq p \leq k} \left\| A_p^{(j)} - A_p \right\| \leq \varepsilon_j, \qquad j \in \mathbb{N}, \tag{4.38}$$

$$\sup_{n \in \mathbb{N}, 1 \leq p \leq k} \left\| A_p^{(j)} - A_p^{(j')} \right\| \leq \varepsilon_j, \qquad 1 \leq j \leq j'. \tag{4.39}$$

After the use of several triangular inequalities, we can show by (4.38), (4.39) and the assumption (4) that the random variables

$$\delta_j := \sup_{n \in \mathbb{N}} \left\| P(A_1, \ldots, A_k, B_1, \ldots, B_k) - P(A_1^{(j)}, \ldots, A_k^{(j)}, B_1, \ldots, B_k) \right\|, \qquad 1 \leq j, \tag{4.40}$$

$$\delta_{j,j'} := \sup_{n \in \mathbb{N}} \left\| P(A_1^{(j)}, \ldots, A_k^{(j)}, B_1, \ldots, B_k) - P(A_1^{(j')}, \ldots, A_k^{(j')}, B_1, \ldots, B_k) \right\|, \qquad j \leq j' \tag{4.41}$$

converge to 0 almost surely as $j \to \infty$.

Let $\{\lambda_i(n)\}_{i \geq 1}$ be the random eigenvalues of $P(A_1, \ldots, A_k, B_1, \ldots, B_k)$. By Weyl's inequality for eigenvalues [**6**, Corollary III.2.6.], we have

$$\left| \lambda_i^{\pm}(n) - (\lambda_i^{(j)})^{\pm}(n) \right| \leq \delta_j \text{ a.s.}, \qquad n \in \mathbb{N}, \tag{4.42}$$

$$\left| (\lambda_i^{(j)})^{\pm}(n) - (\lambda_i^{(j')})^{\pm}(n) \right| \leq \delta_{j,j'} \text{ a.s.}, \qquad n \in \mathbb{N}. \tag{4.43}$$

The second inequality (4.43) gives us $|(\lambda_i^{(j)})^{\pm} - (\lambda_i^{(j')})^{\pm}| \leq \delta_{j,j'}$ for $j \leq j'$, and so $\{(\lambda_i^{(j)})^{\pm}\}_{j \geq 1}$ is a Cauchy sequence and has a limit $\lambda_i^{\pm}$ as $j \to \infty$ for each fixed i. The first inequality (4.42) gives us

$$\varlimsup_{n \to \infty} \left| \lambda_i^{\pm}(n) - \lambda_i^{\pm} \right| \leq \varlimsup_{n \to \infty} \left| \lambda_i^{\pm}(n) - (\lambda_i^{(j)})^{\pm}(n) \right| + \varlimsup_{n \to \infty} \left| (\lambda_i^{(j)})^{\pm}(n) - (\lambda_i^{(j)})^{\pm} \right|$$
$$+ \left| (\lambda_i^{(j)})^{\pm} - \lambda_i^{\pm} \right|$$
$$\leq \delta_j + \left| (\lambda_i^{(j)})^{\pm} - \lambda_i^{\pm} \right| \text{ a.s.} \tag{4.44}$$

By letting $j \to \infty$ we get the almost sure convergence $\lambda_i^{\pm}(n) \to \lambda_i^{\pm}$ as $n \to \infty$.

Finally we prove that $\lambda_i^{\pm} \to 0$ as $i \to \infty$, so the limiting eigenvalues correspond to a compact operator. We denote by $s_1(X) \geq s_2(X) \geq \cdots \geq 0$ the singular values of a compact operator X. Note that in our notation of proper arrangement, for a self-adjoint operator X it holds that $s_i(X) = |\lambda_i(X)|$. Note also that singular values satisfy $s_i(XYZ) \leq \|X\|\|Z\|s_i(Y)$ and $s_{i+j-1}(X+Y) \leq s_i(X) + s_i(Y)$ which can be proved by the mini-max principle [**6**, Corollary III.1.2 and Problem III.6.2].

Suppose that the polynomial P is of the form

$$P(A_1, \ldots, A_k, B_1, \ldots, B_k) = \sum_{\ell=1}^{m} X_\ell, \tag{4.45}$$

where X_ℓ is a monomial containing some $A_{p(\ell)}$. Take $\varepsilon > 0$. By assumption (2), there exists some $i_0 \in \mathbb{N}$ such that

$$\sup_{n\in\mathbb{N},1\leq p\leq k} |\lambda_i(A_p)| \leq \frac{\varepsilon}{m}, \qquad i \geq i_0. \tag{4.46}$$

Let X_ℓ' be the monomial X_ℓ with the matrix $A_{p(\ell)}$ removed from it. Then we get

$$s_{mi-m+1}(P) \leq \sum_{\ell=1}^{m} \|X_\ell'\|s_i(A_{p(\ell)}) \leq M\varepsilon, \qquad i \geq i_0, \tag{4.47}$$

$$M := \sup_{n\in\mathbb{N},1\leq \ell\leq k} \|X_\ell'\|. \tag{4.48}$$

Since $\sup_{n\in\mathbb{N}} \|A_p(n)\| < \infty$ and $\sup_{n\in\mathbb{N}} \|B_p(n)\| < \infty$ then M is finite almost surely too. This shows that $\sup_{n\in\mathbb{N}} s_i(P) = \sup_{n\in\mathbb{N}} |\lambda_i(n)|$ converges to 0 as $i \to \infty$ almost surely, and so $\lim_{i\to\infty} \lambda_i = 0$. □

4.5. Several Haar unitaries case.

Theorems 4.1, 4.3, 4.7 and Corollaries 4.4, 4.5 can be generalized to the case when several independent Haar unitaries are involved. Proofs are just to combine our results of asymptotic cyclic monotonicity with asymptotic freeness between B_i's. For example Theorem 4.1 can be generalized as follows.

THEOREM 4.8. *Let* $A_i = A_i(n), B_{ij} = B_{ij}(n), i, j = 1, \ldots, k$ *be* $n \times n$ *random matrices and* $U_i = U_i(n), i = 1, \ldots, k$ *be independent Haar unitary random matrices. Suppose that*

(1) $((A_1, \ldots, A_k), \mathbb{E}\otimes\mathrm{Tr}_n)$ *converges in distribution to a k-tuple of trace class operators as $n \to \infty$,*

(2) *for each i,* $((B_{i1}, \ldots, B_{ik}), \mathbb{E}\otimes\mathrm{tr}_n)$ *converges in distribution to a k-tuple of elements in a non-commutative probability space as $n \to \infty$,*

(3) *the families* $\{A_i\}_{i=1}^{k}, \{B_{ij}\}_{i,j=1}^{k}, \{U_i\}_{i=1}^{k}$ *are independent.*

Then the pair $(\{A_i\}_{i=1}^{k}, \{U_i B_{ij} U_i^*\}_{i,j=1}^{k})$ *is asymptotically cyclically monotone with respect to* $(\mathbb{E} \otimes \mathrm{Tr}_n, \mathbb{E} \otimes \mathrm{tr}_n)$.

Proof. We take a Haar unitary U independent of all A_i, B_{ij} and U_i. Let $\tilde{B}_{ij} := U_i B_{ij} U_i^*$. By [**9**, Theorem 3.1] it follows that $\{\tilde{B}_{1j}\}_{j=1}^k, \ldots, \{\tilde{B}_{kj}\}_{j=1}^k$ are asymptotically free with respect to $\mathbb{E} \otimes \mathrm{tr}_n$, so $((\tilde{B}_{11}, \tilde{B}_{12}, \ldots, \tilde{B}_{kk}), \mathbb{E} \otimes \mathrm{tr}_n)$ converges in distribution to a k^2-tuple of elements in a non-commutative probability space as $n \to \infty$. Theorem 4.1 implies that the pair $(\{A_i\}_{i=1}^k, \{U\tilde{B}_{ij}U^*\}_{i,j=1}^k)$ is asymptotically cyclically monotone with respect to $(\mathbb{E} \otimes \mathrm{Tr}_n, \mathbb{E} \otimes \mathrm{tr}_n)$. Since $(UU_1, \ldots, UU_k)$ has the same distribution as $(U_1, \ldots, U_k)$, we conclude that the pair $(\{A_i\}_{i=1}^k, \{\tilde{B}_{ij}\}_{i,j=1}^k)$ is also asymptotically cyclically monotone with respect to $(\mathbb{E} \otimes \mathrm{Tr}_n, \mathbb{E} \otimes \mathrm{tr}_n)$. $\qquad\square$

The same technique allows us to generalize Lemma 4.2 to the several Haar unitaries case, and so we obtain the almost sure convergence, namely the generalization of Theorem 4.3.

THEOREM 4.9. *Let $A_i = A_i(n), B_{ij} = B_{ij}(n), i, j = 1, \ldots, k$ be $n \times n$ deterministic matrices and $U_i = U_i(n), i = 1, \ldots, k$ be independent Haar unitary random matrices. Suppose that*

(1) *$((A_1, \ldots, A_k), \mathrm{Tr}_n)$ converges in distribution to a k-tuple of trace class operators as $n \to \infty$,*

(2) *for each i, $((B_{i1}, \ldots, B_{ik}), \mathrm{tr}_n)$ converges in distribution to a k-tuple of elements in a non-commutative probability space as $n \to \infty$,*

Then the pair $(\{A_i\}_{i=1}^k, \{U_i B_{ij} U_i^*\}_{i,j=1}^k)$ is asymptotically cyclically monotone with respect to $(\mathrm{Tr}_n, \mathrm{tr}_n)$ almost surely.

Corollary 4.4 is generalized in the following form.

COROLLARY 4.10. *Under the assumptions of Theorem 4.9, for any selfadjoint $*$-polynomial $P(\{x_i\}_{i=1}^k, \{y_{ij}\}_{i,j=1}^k)$ such that $P(\{0\}_{i=1}^k, \{y_{ij}\}_{i,j=1}^k) = 0$, the Hermitian random matrix $P(\{A_i\}_{i=1}^k, \{U_i B_{ij} U_i^*\}_{i,j=1}^k)$ converges in eigenvalues to a selfadjoint Hilbert Schmidt operator almost surely. The limiting eigenvalues can be computed by asymptotic cyclic monotonicity of $(\{A_i\}_{i=1}^k, \{U_i B_{ij} U_i^*\}_{i,j=1}^k)$ and asymptotic freeness of $\{\{U_1 B_{1j} U_1^*\}_{j=1}^k, \ldots, \{U_k B_{kj} U_k^*\}_{j=1}^k\}$.*

Corollary 4.5 and Theorem 4.7 can be similarly generalized, the explicit statements of which are omitted. We also mention that we can obtain the above results if we take $B_{ij} = G_i$ where $G_1, \ldots, G_k$ are independent GUEs normalized so that each G_i converges in distribution. In this case we may remove the Haar unitaries U_i since the independent GUEs provide independent Haar unitaries when diagonalized.

In the case of a single Haar unitary, considering the pair $(\{UA_iU^*\}_{i=1}^k, \{B_j\}_{j=1}^\ell)$ is equivalent to considering the pair $(\{A_i\}_{i=1}^k, \{UB_jU^*\}_{j=1}^\ell)$. However, in the several Haar unitaries case, the pair $(\{U_i A_{ij} U_i^*\}_{i,j=1}^k, \{B_i\}_{i=1}^k)$ becomes rather trivial.

PROPOSITION 4.11. *Let $A_i = A_i(n), B_i = B_i(n), i = 1, \ldots, k$ be $n \times n$ deterministic matrices and $U_i = U_i(n), i = 1, \ldots, k$ be independent Haar unitaries. Suppose that*

(1) $((A_1, \ldots, A_k), \mathrm{Tr}_n)$ converges in distribution to a k-tuple of trace class operators as $n \to \infty$,

(2) $((B_1, \ldots, B_k), \mathrm{tr}_n)$ converges in distribution to a k-tuple of elements in a non-commutative probability space as $n \to \infty$,

(3) $\sup_{n \in \mathbb{N}} \|B_i(n)\| < \infty$ for every $i = 1, \ldots, k$.

For any tuple $(i_1, \ldots, i_k) \in \{1, \ldots, k\}^k$ such that at least two of them are distinct, we have the almost sure convergence

$$\lim_{n \to \infty} \mathrm{Tr}_n(U_{i_1} A_1 U_{i_1}^* B_1 \cdots U_{i_k} A_k U_{i_k}^* B_k) = 0.$$

PROOF.   We may suppose that $1, 2 \in \{i_1, \ldots, i_k\}$. By $\mathbb{E}^{U_1}$ we denote the expectation with respect to U_1, leaving $U_2, \ldots, U_k$ unchanged (that is, the conditional expectation onto the σ-field generated by $U_2, \ldots, U_k$). We show that almost surely

$$\mathbb{E}^{U_1}\left[\left|\mathrm{Tr}_n(U_{i_1} A_1 U_{i_1}^* B_1 \cdots U_{i_k} A_k U_{i_k}^* B_k)\right|^2\right] = O(n^{-2}). \tag{4.49}$$

By the cyclic property of the trace and by the obvious property $\overline{\mathrm{Tr}_n(X)} = \mathrm{Tr}_n(X^*)$, the LHS of (4.49) equals

$$\mathbb{E}^{U_1}[\mathrm{Tr}_n(C_1 U_1 D_1 U_1^* \cdots C_\ell U_1 D_\ell U_1^*) \, \mathrm{Tr}_n(C_{\ell+1} U_1 D_{\ell+1} U_1^* \cdots C_{2\ell} U_1 D_{2\ell} U_1^*)], \tag{4.50}$$

where C_i's are products of $B_p, B_p^*, U_q A_r U_q^*$ and $U_q A_r^* U_q^*$ with $q \geq 2$ and $D_i \in \{A_r^\varepsilon : \varepsilon \in \{1, *\}, 1 \leq r \leq k\}$. At least one of the matrices $C_1, \ldots, C_\ell$ must have a factor $U_2 A_r U_2^*$ for some r and similarly for $\{C_{\ell+1}, \ldots, C_{2\ell}\}$. Let $Z^{\cup 2}$ be the permutation $(1, \ldots, \ell)(\ell + 1, \ldots, 2\ell)$. By the Weingarten formula (4.12), the quantity (4.50) equals

$$\sum_{\substack{\sigma_1, \sigma_2, \sigma_3 \in S_{2\ell} \\ \sigma_1 \sigma_2 \sigma_3 = Z^{\cup 2}}} \mathrm{Tr}_{\sigma_1}(C_1, \ldots, C_{2\ell}) \, \mathrm{Tr}_{\sigma_2}(D_1, \ldots, D_{2\ell}) \, \mathrm{Wg}(\sigma_3, n). \tag{4.51}$$

By assumption $\mathrm{Tr}_{\sigma_2}(D_1, \ldots, D_{2\ell}) = O(1)$ and by inspection $\mathrm{Tr}_{\sigma_1}(C_1, \ldots, C_{2\ell}) = O(n^{2\ell-2})$. By (4.7) $\mathrm{Wg}(\sigma_3, n)$ behaves as $O(n^{-2\ell-|\sigma_3|})$. Thus, we obtain the behavior $O(n^{-2})$ of (4.49). By taking the sum $\sum_{n=1}^{\infty}$ and by using the conditional monotone convergence theorem we obtain the conclusion.   □

Therefore, a monomial of $\{U_i A_i U_i^*\}_{i=1}^k, \{B_i\}_{i=1}^k$ is trivial with respect to Tr_n if it contains at least two distinct factors from $\{U_i A_i U_i^*\}_{i=1}^k$. The nontrivial case is only when one factor from $\{U_i A_i U_i^*\}_{i=1}^k$ appears in each monomial; for instance

$$\sum_{i=1}^k B_i U_i A_i U_i^* B_i^*. \tag{4.52}$$

We compute the eigenvalues of this model in Proposition 5.2.

5. Some concrete computation of eigenvalues.

5.1. Eigenvalues of polynomials of random matrices.

We provide explicit discrete eigenvalues of some polynomials of random matrices converging to compact operators. Results in Section 4 show that the computation of the eigenvalues reduces to the eigenvalues of polynomials of cyclically monotone elements. Then the computations in Section 3.4 give the corresponding results on large random matrices converging to trace class operators. We can then show the results in the compact setup (i.e. the same assumptions as in Theorem 4.7) by approximation.

THEOREM 5.1.   *Let* $A_i = A_i(n), B_i = B_i(n), i = 1, \ldots, k$ *be deterministic* $n \times n$ *matrices and* $U = U(n)$ *be an* $n \times n$ *Haar unitary such that*

(1) $A_1, \ldots, A_k$ *are Hermitian,*

(2) $((A_1, \ldots, A_k), \mathrm{Tr}_n)$ *converges in compact distribution to a* k-*tuple of compact operators* $((a_1, \ldots, a_k), \mathrm{Tr}_H)$ *as* $n \to \infty$ *(see Definition 2.6),*

(3) $((B_1, \ldots, B_k), \mathrm{tr}_n)$ *converges in distribution to a* k-*tuple of elements in a noncommutative probability space as* $n \to \infty$,

(4) $\sup_{n \in \mathbb{N}} \|B_i(n)\| < \infty$ *for every* $i = 1, \ldots, k$.

In this case let $\beta_i := \lim_{n\to\infty} \mathrm{tr}_n(B_i)$, $\beta_{ij} := \lim_{n\to\infty} \mathrm{tr}_n(B_i^* B_j)$ *and* $B := (\beta_{ij})_{i,j=1}^k$. *The following statements hold true.*

(i) *We have*

$$\lim_{n\to\infty} \mathrm{EV}\left(\sum_{i=1}^k U B_i U^* A_i (U B_i U^*)^* \right) = \mathrm{EV}\left(\sqrt{B} \operatorname{diag}(a_1, \ldots, a_k) \sqrt{B} \right) \ a.s.,$$

where $\sqrt{B} \operatorname{diag}(a_1, \ldots, a_k) \sqrt{B}$ *is viewed as an element of* $(M_k(\mathbb{C}) \otimes \mathcal{S}^1(H), \mathrm{Tr}_k \otimes \mathrm{Tr}_H)$.

(ii) *Suppose that* $B_1, \ldots, B_k$ *are Hermitian. Then*

$$\lim_{n\to\infty} \mathrm{EV}\left(\sum_{i=1}^k A_i U B_i U^* A_i \right) = \mathrm{EV}\left(\sum_{i=1}^k \beta_i a_i^2 \right) \ a.s.$$

(iii) *Suppose that* $k = 1$ *and* B_1 *is Hermitian. Let* $p = \sqrt{\beta_{11}} + \beta_1, q = -(\sqrt{\beta_{11}} - \beta_1)$. *Then*

$$\lim_{n\to\infty} \mathrm{EV}(A_1 U B_1 U^* + U B_1 U^* A_1) = (p \, \mathrm{EV}(a_1)) \sqcup (q \, \mathrm{EV}(a_1)) \ a.s.$$

(iv) *Suppose that* $k = 1$ *and* B_1 *is Hermitian. Let* $r := \sqrt{\beta_{11} - \beta_1^2}$. *Then*

$$\lim_{n\to\infty} \mathrm{EV}(\mathrm{i}[A_1, U B_1 U^*]) = (r \, \mathrm{EV}(a_1)) \sqcup (-r \, \mathrm{EV}(a_1)) \ a.s.$$

PROOF.   We only show (i) since the other cases are similar. We reuse the notations and proof of Theorem 4.7. In particular UB_iU^* is abbreviated to B_i. Now the polynomial P is

$$P(x_1, \ldots, x_k, y_1, \ldots, y_k) = \sum_{i=1}^{k} y_i x_i y_i^*. \qquad (5.1)$$

Recall that

$$\{\lambda_i(n)\}_{i \geq 1} = \mathrm{EV}\left(P(A_1, \ldots, A_k, B_1, \ldots, B_k)\right), \qquad (5.2)$$

$$\{\lambda_i\}_{i \geq 1} = \lim_{n \to \infty} \mathrm{EV}\left(P(A_1, \ldots, A_k, B_1, \ldots, B_k)\right) \text{ a.s.}, \qquad (5.3)$$

$$\{\lambda_i^{(j)}(n)\}_{i \geq 1} = \mathrm{EV}\left(P(A_1^{(j)}, \ldots, A_k^{(j)}, B_1, \ldots, B_k)\right), \qquad (5.4)$$

$$\{\lambda_i^{(j)}\}_{i \geq 1} = \lim_{n \to \infty} \mathrm{EV}\left(P(A_1^{(j)}, \ldots, A_k^{(j)}, B_1, \ldots, B_k)\right) \text{ a.s.} \qquad (5.5)$$

From the result in the trace class setup in Theorem 3.14 and asymptotic cyclic monotonicity in Corollary 4.4, we have the identity

$$\{\lambda_i^{(j)}\}_{i \geq 1} = \mathrm{EV}\left(\sqrt{B}\,\mathrm{diag}(a_1^{(j)}, \ldots, a_k^{(j)})\sqrt{B}\right). \qquad (5.6)$$

Now we define another sequence of eigenvalues

$$\{\lambda_i'\}_{i \geq 1} := \mathrm{EV}\left(\sqrt{B}\,\mathrm{diag}(a_1, \ldots, a_k)\sqrt{B}\right). \qquad (5.7)$$

Our goal is to demonstrate that $\lambda_i = \lambda_i'$ for every $i \in \mathbb{N}$. For this we use the inequality

$$\left|\lambda_i^{\pm} - (\lambda_i')^{\pm}\right| \leq \left|\lambda_i^{\pm} - \lambda_i^{\pm}(n)\right| + \left|\lambda_i^{\pm}(n) - (\lambda_i^{(j)})^{\pm}(n)\right|$$
$$+ \left|(\lambda_i^{(j)})^{\pm}(n) - (\lambda_i^{(j)})^{\pm}\right| + \left|(\lambda_i^{(j)})^{\pm} - (\lambda_i')^{\pm}\right|. \qquad (5.8)$$

We proved in the proof of Theorem 4.7 that the first term on the RHS converges to 0 as $n \to \infty$ and proved in (4.42) that the second term is bounded by δ_j uniformly on n. The third term converges to 0 as $n \to \infty$ from the result in the trace class setup. Taking $\overline{\lim}_{j \to \infty} \overline{\lim}_{n \to \infty}$ we get

$$\left|\lambda_i^{\pm} - (\lambda_i')^{\pm}\right| \leq \overline{\lim_{j \to \infty}} \left|(\lambda_i^{(j)})^{\pm} - (\lambda_i')^{\pm}\right|. \qquad (5.9)$$

By (5.6) it now suffices to show that

$$\lim_{j \to \infty} \mathrm{EV}\left(\sqrt{B}\,\mathrm{diag}(a_1^{(j)}, \ldots, a_k^{(j)})\sqrt{B}\right) = \mathrm{EV}\left(\sqrt{B}\,\mathrm{diag}(a_1, \ldots, a_k)\sqrt{B}\right), \qquad (5.10)$$

which follows from the fact

$$\lim_{j \to \infty} \left\|\sqrt{B}\,\mathrm{diag}(a_1^{(j)}, \ldots, a_k^{(j)})\sqrt{B} - \sqrt{B}\,\mathrm{diag}(a_1, \ldots, a_k)\sqrt{B}\right\| = 0 \qquad (5.11)$$

and Weyl's inequality [**6**, Corollary III.2.6] which bounds the difference of eigenvalues by the operator norm of the difference of elements.   □

When several independent Haar unitaries are involved we can still compute the eigenvalues. For example we obtain the following.

PROPOSITION 5.2.   *Under the assumptions of Theorem 5.1, we take independent Haar unitaries $U_i = U_i(n), i = 1, \ldots, k$.*

(i) *Let $\beta_i := \lim_{n \to \infty} \mathrm{tr}_n(B_i)$. Then*

$$\lim_{n \to \infty} \mathrm{EV}\left(\sum_{i=1}^{k} U_i B_i U_i^* A_i (U_i B_i U_i^*)^* \right) = \mathrm{EV}\left(\sum_{i=1}^{k} |\beta_i|^2 a_i \right) \ a.s.$$

(ii) *Let $\gamma_i := \lim_{n \to \infty} \mathrm{tr}_n(B_i^* B_i)$. Then we have*

$$\lim_{n \to \infty} \mathrm{EV}\left(\sum_{i=1}^{k} B_i U_i A_i U_i^* B_i^* \right) = \mathrm{EV}(\gamma_1 a_1) \sqcup \cdots \sqcup \mathrm{EV}(\gamma_k a_k) \ a.s.$$

PROOF.   Suppose that $((A_1, \ldots, A_k), \mathrm{Tr}_n)$ converges in distribution to trace class operators; the general compact case is proved by approximation.

(i) By Corollary 4.10, the computation formula is obtained by the asymptotic cyclic monotonicity of $(\{A_i\}_{i=1}^{k}, \{U_i B_i U_i^*\}_{i=1}^{k})$ and then the asymptotic freeness of $U_1 B_1 U_1^*$, $\ldots, U_k B_k U_k^*$. This implies that we only need to replace the covariance matrix B in Theorem 5.1(i) with $B' = (\overline{\beta_i} \beta_j)_{i,j=1}^{k} \in M_k(\mathbb{C})$. The limiting eigenvalues are given by those of $\sqrt{B'} \, \mathrm{diag}(a_1, \ldots, a_k) \sqrt{B'}$. Since $\sqrt{B'} = (\sum_{i=1}^{k} |\beta_i|^2)^{-1/2} B'$, the conclusion follows easily.

(ii) By Proposition 4.11, for each $\ell \in \mathbb{N}$ we have

$$\mathrm{Tr}_n\left(\left(\sum_{i=1}^{k} B_i U_i A_i U_i^* B_i^* \right)^{\ell} \right) = \sum_{i=1}^{k} \mathrm{Tr}_n\left((B_i U_i A_i U_i^* B_i^*)^{\ell} \right) + o(1)$$

$$= \sum_{i=1}^{k} \mathrm{Tr}_n\left((A_i (U_i^* B_i^* B_i U_i))^{\ell} \right) + o(1). \qquad (5.12)$$

By the cyclic monotonicity of the pair $(\{A_i\}_{i=1}^{k}, \{U_i^* B_i^* B_i U_i\}_{i=1}^{k})$ (see Theorem 4.3), we have the almost sure convergence

$$\lim_{n \to \infty} \mathrm{Tr}_n\left(\left(\sum_{i=1}^{k} B_i U_i A_i U_i^* B_i^* \right)^{\ell} \right) = \sum_{i=1}^{k} \lim_{n \to \infty} \mathrm{Tr}_n(A_i^{\ell}) \lim_{n \to \infty} \mathrm{tr}_n(B_i^* B_i)^{\ell}$$

$$= \sum_{i=1}^{k} \mathrm{Tr}_H((\gamma_i a_i)^{\ell}), \qquad (5.13)$$

and the conclusion follows by Proposition 2.8.   □

5.2.   Numerical illustration.

In this subsection, we give numerical illustrations of main theorems of this paper.

Example 5.3.   Let $\{Z_i\}_i$ be a family of independent $n \times n$ non-selfadjoint Gaussian random matrices, that is, each Z_i has entries that are independent identically distributed with the standard complex normal distribution. We consider a sample covariance matrix $X_i = Z_i Z_i^* / (2n)$ and a diagonal matrix $D = \text{diag}(2^{-1}, 2^{-2}, 2^{-3}, \ldots, 2^{-n})$. From Theorem 4.1, the matrices D and X are asymptotically cyclically monotone. Table 1 shows numerical simulations of mixed moments and moments decomposed by cyclic monotone independence when $n = 500$.

Table 1.   Moments ($n = 500$).

$\text{Tr}(DX_1)$	1.65916	$\text{Tr}(D)\,\text{tr}(X_1)$	1.64044
$\text{Tr}(DX_1D)$	1.11507	$\text{Tr}(D^2)\,\text{tr}(X_1)$	1.08068
$\text{Tr}(DX_1DX_2)$	1.08634	$\text{Tr}(D^2)\,\text{tr}(X_1)\,\text{tr}(X_2)$	1.08034
$\text{Tr}(DX_1DX_2D)$	1.02348	$\text{Tr}(D^3)\,\text{tr}(X_1)\,\text{tr}(X_2)$	1.01548
$\text{Tr}(DX_1DX_2DX_3)$	1.00131	$\text{Tr}(D^3)\,\text{tr}(X_1)\,\text{tr}(X_2)\,\text{tr}(X_3)$	1.0176

Example 5.4.   We give numerical realization of Proposition 5.2(ii). Let D be the $n \times n$ diagonal matrix $\text{diag}(2^{-1}, 2^{-2}, \ldots, 2^{-n})$ and $U_i, i = 1, 2, \ldots$ be $n \times n$ independent Haar unitary random matrices. Then the eigenvalues of $D + \sum_{i=1}^{k} U_i D U_i^*$ can be computed by Proposition 5.2 and converge to

$$(\underbrace{2^{-1}, \ldots, 2^{-1}}_{k \text{ times}}, \underbrace{2^{-2}, \ldots, 2^{-2}}_{k \text{ times}}, \ldots). \tag{5.14}$$

In Figure 3 we show the first 14 eigenvalues for $n = 1000$.
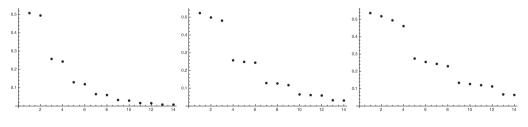

Figure 3.   Eigenvalues $\lambda_1, \ldots, \lambda_{14}$ of $\left\{ D + \sum_{i=1}^{k} U_i D U_i^* \right\}_{k=1,2,3}$ ($n = 1000$).

Example 5.5.   We compute random matrix models of Theorem 5.1 (iii) and (iv). Assume that $n = 300$. Let D be the $n \times n$ diagonal matrix $\text{diag}(2^{-1}, 2^{-2}, \ldots, 2^{-n})$ and U be an $n \times n$ Haar unitary. Let Z be an $n \times n$ non-selfadjoint Gaussian random matrix. We consider $A = UDU^*$ and $B = ZZ^*/(2n)$. First, we consider a random matrix model for Theorem 5.1 (iii). We compute $\text{EV}(AB + BA)$, $\text{EV}(pA) \sqcup \text{EV}(qA)$ and the theoretical

limiting eigenvalues up to 6th. In this case, we obtained $p = 2.4174$ and $q = -0.41351$. We also compute moments of $AB + BA$ with respect to the non-normalized trace Tr_n and factorized moments arising from cyclic monotone independence.

Second, we consider a random matrix model for Theorem 5.1 (iv). We compute $\mathrm{EV}(\mathrm{i}(AB - BA))$, $\mathrm{EV}(rA) \sqcup \mathrm{EV}(-rA)$ and the theoretical limiting eigenvalues up to 10th. In this case, we have $r = 0.98972$. We also compute the moments and factorized moments.
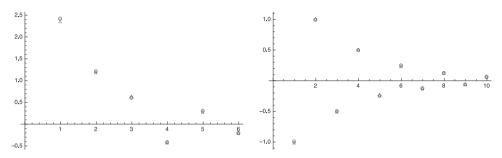

Figure 4.   Left: comparing $\mathrm{EV}(AB + BA)$ (triangle), $\mathrm{EV}(pA) \sqcup \mathrm{EV}(qA)$ (circle) and the theoretical limiting eigenvalues (square). Right: comparing $\mathrm{EV}(\mathrm{i}(AB - BA))$ (triangle), $\mathrm{EV}(rA) \sqcup \mathrm{EV}(-rA)$ (circle) and the theoretical limiting eigenvalues (square).

Table 2.   Moments ($n = 300$).

k	$\mathrm{Tr}((AB+BA)^k)$	$\mathrm{Tr}(A^k)(p^k+q^k)$	k	$\mathrm{Tr}((i(AB-BA))^{2k})$	$\mathrm{Tr}(2A^{2k})r^{2k}$
1	3.91122	4.00779	1	2.70837	2.61211
2	7.71417	8.01973	2	2.22769	2.04693
3	14.9663	16.0641	3	2.16952	1.90957

References

[1]   G. Anderson, A. Guionnet and O. Zeitouni, An introduction to random matrices, Cambridge Stud. Adv. Math., **118**, Cambridge University Press, Cambridge, 2010.

[2]   J. Baik, G. Ben Arous and S. Péché, Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices, Ann. Probab., **33** (2005), 1643–1697.

[3]   S. T. Belinschi, H. Bercovici, M. Capitaine and M. Février, Outliers in the spectrum of large deformed unitarily invariant models, Ann. Probab., **45** (2017), no. 6A, 3571–3625.

[4]   S. T. Belinschi and D. Shlyakhtenko, Free probability of type B: analytic interpretation and applications, Amer. J. Math., **134** (2012), 193–234.

[5]   F. Benaych-Georges and R. R. Nadakuditi, The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices, Adv. Math., **227** (2011), 494–521.

[6]   R. Bhatia, Matrix Analysis, Graduate Texts in Math., **169**, Springer-Verlag, New York, 1997, xii+347 pp.

[7]   P. Biane, F. Goodman and A. Nica, Non-crossing cumulants of type B, Trans. Amer. Math. Soc., **335** (2003), 2263–2303.

[8] M. Capitaine, C. Donati-Martin and D. Féral, The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations, Ann. Probab., **37** (2009), 1–47.

[9] B. Collins, Moments and cumulants of polynomial random variables on unitary groups, the Itzykson–Zuber integral, and free probability, Int. Math. Res. Not., **2003** (2003), 953–982.

[10] B. Collins and C. Male, The strong asymptotic freeness of Haar and deterministic matrices, Ann. Sci. Éc. Norm. Supér. (4), **47** (2014), 147–163.

[11] B. Collins and P. Śniady, Integration with respect to the Haar measure on unitary, orthogonal and symplectic group, Commun. Math. Phys., **264** (2006), 773–795.

[12] M. Février and A. Nica, Infinitesimal non-crossing cumulants and free probability of type B, J. Funct. Anal., **258** (2010), 2983–3023.

[13] U. Haagerup and S. Thorbjørnsen, A new application of random matrices: $\mathrm{Ext}(C^*_{red}(F_2))$ is not a group, Ann. of Math. (2), **162** (2005), 711–775.

[14] C. Male, The norm of polynomials in large random and deterministic matrices, with an appendix by Dimitri Shlyakhtenko, Probab. Theory Related Fields, **154** (2012), 477–532.

[15] N. Muraki, Monotonic independence, monotonic central limit theorem and monotonic law of small numbers, Infin. Dimens. Anal. Quantum Probab. Relat. Top., **4** (2001), 39–58.

[16] A. Nica and R. Speicher, Commutators of free random variables, Duke Math. J., **92** (1998), 553–592.

[17] A. Nica and R. Speicher, Lectures on the combinatorics of free probability, London Math. Soc. Lecture Note Series, **335**, Cambridge University Press, Cambridge, 2006, xvi+417 pp.

[18] S. Péché, The largest eigenvalue of small rank perturbations of Hermitian random matrices, Probab. Theory Related Fields, **134** (2006), 127–173.

[19] A. Pizzo, D. Renfrew and A. Soshnikov, On finite rank deformations of Wigner matrices, Ann. Inst. Henri Poincaré Probab. Stat., **49** (2013), 64–94.

[20] D. Renfrew and A. Soshnikov, On finite rank deformations of Wigner matrices II: Delocalized perturbations, Random Matrices Theory Appl., **2** (2013), 1250015, 36 pp.

[21] D. Shlyakhtenko, Free probability of type B and asymptotics of finite-rank perturbations of random matrices, arXiv:1509.08841v1.

[22] T. Tao, Outliers in the spectrum of iid matrices with bounded rank perturbations, Probab. Theory Related Fields, **155** (2013), 231–263.

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

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

[25] D. Voiculescu, K. Dykema and A. Nica, Free Random Variables, CRM Monograph Series, **1**, Amer. Math. Soc., 1992.

Benoit Collins
Department of Mathematics
Graduate School of Science
Kyoto University
Sakyo, Kyoto 606-8502, Japan
E-mail: collins@math.kyoto-u.ac.jp

Takahiro Hasebe
Department of Mathematics
Hokkaido University
Kita 10, Nishi 8, Kita-ku
Sapporo 060-0810, Japan
E-mail: thasebe@math.sci.hokudai.ac.jp

Noriyoshi Sakuma
Department of Mathematics
Aichi University of Education
Hirosawa 1 Igaya-cho
Kariya-shi 448-854, Japan
E-mail: sakuma@auecc.aichi-edu.ac.jp