Strong asymptotic freeness for independent uniform variables on compact groups associated to non-trivial representations

Charles Bordenave, Benoit Collins

Strong asymptotic freeness for independent uniform variables on compact groups associated to non-trivial representations

Charles Bordenave and Benoît Collins

December 14, 2020

Abstract

Asymptotic freeness of independent Haar distributed unitary matrices was discovered by Voiculescu. Many refinements have been obtained, including strong asymptotic freeness of random unitaries and strong asymptotic freeness of random permutations acting on the orthogonal of the Perron-Frobenius eigenvector. In this paper, we consider a new matrix unitary model appearing naturally from representation theory of compact groups. We fix a non-trivial signature ρ, i.e. two finite sequences of non-increasing natural numbers, and for n large enough, consider the irreducible representation $V_{n,\rho}$ of $\mathbb{U}_n$ associated to the signature ρ. We consider the quotient $\mathbb{U}_{n,\rho}$ of $\mathbb{U}_n$ viewed as a matrix subgroup of $\mathbb{U}(V_{n,\rho})$, and show that strong asymptotic freeness holds in this generalized context when drawing independent copies of the Haar measure. We also obtain the orthogonal variant of this results. Thanks to classical results in representation theory, this result is closely related to strong asymptotic freeness for tensors, which we establish as a preliminary. In order to achieve this result, we need to develop four new tools, each of independent theoretical interest: (i) a centered Weingarten calculus and uniform estimates thereof, (ii) a systematic and uniform comparison of Gaussian moments and unitary moments of matrices, (iii) a generalized and simplified operator valued non-backtracking theory in a general C^*-algebra, and finally, (iv) combinatorics of tensor moment matrices.

1 Introduction

Asymptotic freeness in the large dimension limit of independent GUE random matrices is a very important phenomenon that was discovered by Voiculescu in the nineties [29]. It allows to understand the spectrum, in the limit of large dimension, and under very general assumptions, of non-commuting polynomials in independent random matrices. Subsequently, Voiculescu proved that independent Haar unitary random matrices are almost surely asymptotically free as the dimension goes to infinity [30].

While the above result addresses unitary groups, i.e. group of symmetries of an Euclidian space, it is natural to consider other group of symmetries, and in particular of a finite set, i.e. symmetric groups. Here, the counterpart of Voiculescu's results were solved by Nica in [25]. Asymptotic freeness for random permutations notably provides fine spectral information for operators acting on a random Schreier graph or on random coverings of a fixed graph. It has also become folklore that asymptotic freeness of some specific non-random permutations can be achieved easily through exhibiting explicitly a sequence of finite quotients of a free group whose kernels intersect to its trivial subgroup.

As soon as compact matrix groups are involved, the joint asymptotic behavior of independent random variables can legitimately be expected to depend on which representation of the group is being considered, and for example, although the above results are conclusive examples of asymptotic freeness irrespective of the representation in the case of the symmetric group, we are not aware of any complete result in this direction for unitary or orthogonal groups, and the initial results deal rather with the very particular case of the fundamental representation. In a recent development by the second author together with Gaudreau Lamarre and Male addressed part of the problem in [6], in the case when a signature is fixed.

However, although asymptotic freeness describes efficiently the macroscopic spectrum of non-commuting polynomials in the generators in large dimension, it is not not enough to analyze the existence of eigenvalues away from the limiting spectrum. The absence of such eigenvalues (also known as outliers) is an arduous problem whose positive answer corresponds to strong asymptotic freeness. We refer to Section 2 for a formal definition of strong asymptotic freeness.

The first breakthrough in this direction was achieved in [15] where the authors proved strong asymptotic freeness of independent GUE matrices, and it was followed by many improvements. A notable progress was strong asymptotic freeness of Haar unitaries in [7]. Later, both authors in [4] showed that strong asymptotic freeness holds for random independent permutations (i.e. Haar unitaries on the symmetric group) when viewed as $n \times n$ matrices.

To continue the parallel between asymptotic freeness and strong asymptotic freeness, let us point out here that all results on strong asymptotic freeness involving independent copies Haar measure on groups are basically only valid with respect to the fundamental representation – with the notable exception of [4], where we show that we can also extend the result to the tensor product of two fundamental representations. For very specific operators, though, partial results have also been obtained in the context of quantum expanders, see [27, 16, 17]. In other words, there is no hint at the fact that the same strong asymptotic freeness would hold true if one were considering all representations simultaneously – or equivalently, for our purposes, the left regular representation.

This seems however to be a very natural question, as there exists sequences of permutations

(such as those of the LPS Ramanujan graphs, [21]) for which a simple polynomial – the sum of generators – is known to behave well for all non-trivial irreducible representations. This prompts us to state a question: given an integer n, consider a non-trivial signature ρ, and the irreducible representation $V_{n,\rho}$ of $\mathbb{U}_n$ associated to the signature. We consider the quotient $\mathbb{U}_{n,\rho}$ of $\mathbb{U}_n$ viewed as a matrix subgroup of $\mathbb{U}(V_{n,\rho})$. In particular, the Haar measure on $\mathbb{U}_{n,\rho}$ has 'less randomness' than the Haar measure on the full unitary group $\mathbb{U}(V_{n,\rho})$, in the sense that the group is of much smaller dimension – say, as a manifold. Then the question is: *for which sequence of pairs* (n, ρ) *is it true that* d *independent Haar random variables of* $\mathbb{U}_{n,\rho}$ *are strongly asymptotically free?*

As a trivial observation, $\dim(V_{n,\rho})$ must tend to infinity because freeness does not occur in finite dimension, so either n or ρ must tend to infinity. Fixing n and letting ρ tend to infinity seems to be a fascinating problem to which we do not have any answer. Basically, our main theorem is an answer to this theorem when ρ is fixed and n tends to infinity. Our result is that strong asymptotic freeness holds as soon as ρ is non-trivial. Our result can be made uniform provided that ρ varies slowly as a function of n. The precise statement can be found in Corollary 3.

The above corollary is a consequence of our main theorem, Theorem 2. This result establishes strong asymptotic freeness for tensor products of the fundamental representation and of the contragredient representation on the orthogonal of fixed points. We believe the main theorem to be of independent interest in mathematical physics, but for the purpose of this introduction, let us note that, as explained in Section 2, thanks to elementary results in representation theory and a few classical facts of operator algebras, Corollary 3 and Theorem 2 are equivalent.

Our main result, Theorem 2, can be interpreted as a 0-1 law over natural sequences of irreducible representations: either it is trivial (if the sequence of representations is one-dimensional) or strongly asymptotically free (in all other cases). In other words, this quantifies very precisely the fact that the sole known obstructions to strong asymptotic freeness are the fixed points of a representation. Hence, there does not seem to be intermediate behavior between strong asymptotic freeness and triviality.

Let us digress a bit to hint at the fact that strong asymptotic freeness can naturally be expected to be much harder to achieve than plain asymptotic freeness. To be more specific, let us try to investigate whether the operator norm of a non-commutative linear function of independent copies of Haar-distributed variables could be bounded above uniformly (which must happen in the case of strong asymptotic freeness thanks to the Haagerup inequality). A naive attempt could consist in a combination of a net argument and a union bound – and this attempt would succeed in the case of the context of the fundamental representation. For this purpose, for $\varepsilon, \eta > 0$, let us observe that the size of an η-net of vectors in the unit ball of $V_{n,\rho}$ is of the order of $(C/\eta)^{\dim(V_{n,\rho})}$, whereas for any 1-Lipschitz function, the likelihood of being ε-away from the median is of order $\exp(-c\varepsilon^2 n)$,

this follows from Gromov' comparison theorem, see [1, Theorem 4.4.27]. The involved exponential speeds n and $\dim(V_{n,\rho})$ are comparable as n grows iff $V_{n,\rho}$ is the fundamental representation or its contragredient. In all other cases $\dim(V_{n,\rho}) \gg n$ and it is impossible, given η, to fix ε such that $(C/\eta)^{\dim(V_{n,\rho})} \cdot \exp(-c\varepsilon^2 n)$ remains small uniformly over the dimension. Hence, we might expect that 'soft' geometric techniques such as exposed in the monograph [2] are not of use here and make it less obvious that strong asymptotic must hold.

Likewise, analytic proofs extending [7] and [15] are not handy in this case because of the tensor structure of the objects – there is no prospect for an easy folding trick like in [7] or a Schwinger-Dyson equation like in [15] because of the tensor structure. In view of this lack of geometric or analytic methods, a natural approach turns out to be based on moments, which is the core technology of this paper. It is implemented with the help of an operator valued non-backtracking theory. Non-backtracking theory was observed by a few authors to be a very powerful to study outliers in the context of graph theory, see e.g. [11, 3]. A version of non-backtracking theory with non-commuting weights was developed further in [4] for the purpose of proving strong asymptotic freeness of random permutations. Here, to achieve our goal, we generalize this theory even further, beyond the case of permutation operators, to the case of all bounded operators. The results are gathered in Section 4.

However, moment methods require a good understanding of moments of the random objects under consideration. While moment estimates for random permutations boil down to combinatorial formulas, the situation is more involved for random unitary or orthogonal matrices. In these cases, formulas for moments require Weingarten calculus, as developed in [10, 5]. One of our key results relies in the comparison of moments of unitary random matrices and Gaussian random matrices, which extends the line of research of [19, 13] and others. Our main result in this direction is Theorem 12, which quantifies how well unitary random matrices can be estimated by Gaussian matrices. A key preliminary result to achieve this goal is to develop a systematic method to handle centering, for which we use a bracket notation. This is based on another notion of lattices and similar moment theoretic work was initiated in [9]. Our main result here is Theorem 9.

The paper is organized as follows. This introduction is followed by Section 2, that states the main results and describes the steps of the proof and its applications. Section 3 contains the necessary new developments on uniform centered Weingarten functions. Section 4 describes non-commutative non-backtracking theory in full generality. Finally, Section 5 is devoted to trace estimates based on the previous sections, and the completion of the proof of the main theorems.

Acknowledgements This paper was initiated while the first author was a visiting JSPS scholar at KU in 2018 and he acknowledges the hospitality of JSPS and of Kyoto University. BC was

supported by JSPS KAKENHI 17K18734 and 17H04823 and CB by ANR grant ANR-16-CE40-0024.

2 Strong asymptotic freeness for random unitary tensors

2.1 Main result

In the sequel, for an integer $k \geq 1$, we set $[\![k]\!] = \{1, \ldots, k\}$. For any compact group, there exists a unique probability measure that is left and right invariant under translation, it is called the normalized Haar measure. In this paper we are interested in the case of the unitary group $\mathbb{U}_n$ and the orthogonal group $\mathbb{O}_n$.

In order to keep the paper to a reasonable length, we choose to focus on the unitary case. The orthogonal case boils down to technical modifications of the unitary case, which we summarize in Subsection 3.5. Let $d \geq 2$ be an integer and let $U_1, \cdots, U_d$ be independent Haar distributed random unitary matrices in $\mathbb{U}_n$. For $d+1 \leq i \leq 2d$, we set $U_i = U_{i-d}^*$. We consider the involution on $[\![2d]\!]$, defined by $i^* = i + d$ for $1 \leq i \leq d$ and $i^* = i - d$ for $d+1 \leq i \leq 2d$. We thus have $U_{i^*} = U_i^*$ for all $i \in [\![2d]\!]$.

If $M \in M_n(C)$ and $\varepsilon \in \{\cdot, -\}$, we set $M^\varepsilon = M$ for $\varepsilon = \cdot$ and $M^\varepsilon = \bar{M}$ for $\varepsilon = -$. We fix a triple of integers $q = q_- + q_+ \geq 1$. For each $i \in [\![2d]\!]$, we introduce the unitary matrix in $\mathbb{U}_{n^q}$,

$$V_i = \bar{U}_i^{\otimes q_-} \otimes U_i^{\otimes q_+} = U_i^{\varepsilon_1} \otimes \cdots \otimes U_i^{\varepsilon_q}, \tag{1}$$

where $(\varepsilon_1, \ldots, \varepsilon_q) \in \{., -\}^q$ is the symbolic sequence: $\varepsilon_p = -$ for $p \leq q_-$ and $\varepsilon_p = \cdot$ otherwise. We recall the following fact:

Proposition 1. *The average matrix $\mathbf{E}V$ is an orthogonal projection onto the vector subspace H of elements invariant under (left) multiplication by $\bar{U}^{q_-} \otimes U^{q_+}$ for all $U \in \mathbb{U}_n$.*

Although this result is classical and its proof is elementary, it is not very well referenced in our specific context so we include an outline for the sake of self-containedness.

Outline of proof. The fact that $\mathbf{E}V$ is a projection follows from a direct application of the fact that the distribution of a Haar distributed element U is the same as $\tilde{U} \cdot \hat{U}$ where $\tilde{U}$ and $\hat{U}$ are two iid copies of U. The fact that it is selfadjoint follows from the fact that U and U^* have the same distribution. The fact that every invariant vector is invariant under the mean is trivial. Finally, let $x \notin H$. By definition, there exists V such that $Vx \neq x$. Since V is unitary and therefore preserves the Euclidean norm, it follows from the strict convexity of the Euclidean norm that $\|\mathbf{E}Vx\|_2 < \|x\|_2$ and therefore x is not in the image of $\mathbf{E}V$, which concludes the proof. □

We refer to Subsection 3.2 for a method to compute $\mathbf{E}V$. For example if the vector $(\varepsilon_1, \ldots, \varepsilon_q)$ is not balanced, that is $q_- \neq q_+$ then $\mathbf{E}V = 0$. We set

$$[V_i] = V_i - \mathbf{E}V_i = V_i - P_H P_H^*. \tag{2}$$

For a fixed integer $r \geq 1$, let $(a_0, a_1, \ldots, a_{2d})$ be matrices in $M_r(\mathbb{C})$. We introduce the following matrix on $\mathbb{C}^r \otimes \mathbb{C}^{n^q}$:

$$A = a_0 \otimes 1 + \sum_{i=1}^{2d} a_i \otimes V_i, \tag{3}$$

where 1 is the identity. From what precedes, the vector space $H_r = \mathbb{C}^r \otimes H$ is an invariant subspace of A and its adjoint A^*. We denote by $H_r^\perp$ the orthogonal of H_r and denote by $A_{|H_r^\perp}$ the restriction of A to this vector space. Our goal is to describe the spectrum of $A_{|H_r^\perp}$ as n goes to infinity.

To this end, the key observation is the following. Consider the unitary representation of the free group $\mathbb{F}_d$ on $\mathbb{C}^{n^q}$ defined by $\pi(g_i) = V_i$, where $(g_i)_{i \in [\![2d]\!]}$ are the free generators and their inverses: $g_{i^*} = g_i^{-1}$. Note that this representation π is random and depends implicitly on n and (q_-, q_+). The matrix A is the image by π of the following operator in $\ell^2(\mathbb{F}_d)$ in the (left)-group algebra on $\mathbb{F}_d$:

$$A_\star = a_0 \otimes 1 + \sum_{i=1}^{2d} a_i \otimes \lambda(g_i), \tag{4}$$

where $g \mapsto \lambda(g)$ is the left-regular representation (that is, left multiplication by group elements).

The main technical result of this paper is the following theorem. In the sequel $\|\cdot\|$ denotes the operator norm of an operator in an Hilbert space: $\|T\| = \sup_{x \neq 0} \|Tx\|_2 / \|x\|_2$ where $\|x\|_2$ is the Euclidean norm in the Hilbert space.

Theorem 2. *There exists a universal constant $c > 0$ such that the following holds. Let $r \geq 1$ be an integer and let $(a_0, a_1, \ldots, a_{2d})$ be matrices in $M_r(\mathbb{C})$. If $q = q(n)$ is a sequence such that $q \leq c \ln(n) / \ln(\ln(n))$ for all n large enough, then with probability one, we have*

$$\lim_{n \to \infty} \|A_{|H_r^\perp}\| = \|A_\star\|.$$

Thanks to the "linearization trick" of the theory of operator algebras (see Pisier [26] and the monograph by Mingo and Speicher [23, p256] for an accessible treatment), Theorem 2 has an apparently much more general corollary that we describe now.

Let us start with some basic notation of representation theory. For any n, let $\rho = \rho(n)$ be a signature (to lighten the notation we omit the dependence in n). More precisely, ρ is a pair of Young diagrams: two pairs of non-negative integer sequences $\lambda = (\lambda_1, \lambda_2, \ldots)$ and $\mu = (\mu_1, \mu_2, \ldots)$ satisfying the following properties: $\mu_i \geq \mu_{i+1}$, $\lambda_i \geq \lambda_{i+1}$ and $\sum_i \lambda_i + \mu_i < \infty$. We introduce $l(\lambda) = \max\{i, \lambda_i > 0\}$ and $|\lambda| = \sum_i \lambda_i$ (and likewise for μ). For notation, we refer for example

to [31]. If $n \geq l(\lambda) + l(\mu)$ we call $V_{n,\rho}$ the irreducible representation of $\mathbb{U}_n$ whose highest weights are $\lambda_1 \geq \cdots \geq \lambda_{l(\lambda)} \geq 0 \geq \cdots \geq 0 \geq -\mu_{l(\mu)} \geq \ldots \geq -\mu_1$. The signature formulation is just a reformulation of highest weight theory, but it is convenient as it allows to define simultaneously representations for all unitary groups $\mathbb{U}_n$ with $n \geq l(\lambda)+l(\mu)$. We call $\mathbb{U}_{n,\rho}$ the quotient of $\mathbb{U}_n$ under the above representation map. We view it as a matrix subgroup of $\mathbb{U}(V_{n,\rho})$. Letting U_i be a Haar distributed random matrix in $\mathbb{U}_n$, we call W_i its image under the irreducible representation map associated to ρ. Since the map is surjective, W_i is Haar distributed according to $\mathbb{U}_{n,\rho}$. Consider any operator on $\mathbb{C}^r \otimes \ell^2(\mathbb{F}_d)$ of the form

$$P_\star = \sum_{g \in \mathbb{F}_d} a_g \otimes \lambda(g),$$

where for all $g \in \mathbb{F}_d$, a_g is a matrix in $M_r(\mathbb{C})$. We assume that a_g is non-zero for a finite number group elements. In other words, $P_\star$ is a matrix-valued non-commutative polynomial. The image of $P_\star$ by the representation π is denoted by P:

$$P = \sum_{g \in \mathbb{F}_d} a_g \otimes W(g),$$

where $W(g) = W_{i_1} \ldots W_{i_k}$ if $g = g_{i_1} \ldots g_{i_k}$. Obviously, H_r and $H_r^\perp$ are again invariant subspaces of P and P^*. The operator $P_\star$ (and thus P) is self-adjoint if the following condition is met:

$$a_{g^{-1}} = a_g^* \quad \text{for all } g \in \mathbb{F}_d. \tag{5}$$

The following result is a corollary of Theorem 2.

Corollary 3. *Let $r \geq 1$ be an integer, P and $P_\star$ be as above and c be as in Theorem 2. If $q = q(n) = |\lambda| + |\mu|$ is a sequence such that $q \leq c \ln(n)/\ln(\ln(n))$ for all n large enough, then with probability one, we have*

$$\lim_{n \to \infty} \|P_{|H_r^\perp}\| = \|P_\star\|.$$

Moreover, if (5) holds then, with probability one, the Hausdorff distance between the spectrum of $P_{|H_r^\perp}$ and $P_\star$ goes to 0 as n goes to infinity.

We refer to [4, Section 6] for details on why Theorem 2 is sufficient to allow general polynomials in the above Corollary 3. As for the fact that we can replace tensors by irreducible representation, this follows from the following two facts: (i) contracting by an orthogonal projection does not increase the operator norm, and (ii) there exists an orthogonal projection Q that commutes with all V_i's such that on $\mathrm{Im}(Q)$, $W_i = QV_iQ$, when $q_+ = |\lambda|, q_- = |\mu|$. We refer for example to [14] for details on the tools involved such as Schur-Weyl duality.

Corollary 3 establishes the almost-sure strong asymptotic freeness of unitary representations of independent Haar-distributed unitary matrices provided that the unitary representation is a sub-representation of a tensor product representation of not too large dimension. We observe that this last point is critical. Indeed, the determinant is representation of dimension 1 (which is a sub-representation of the tensor product representation with $q = n$) and there is no freeness in non-trivial finite dimensional spaces.

2.2 Overview of the proof

We now explain the strategy behind the proof of Theorem 2. The strategy is the same than the one used in [4] for permutation matrices. The model in [4] is however very different and new technical achievements were necessary for tensor products of unitary matrices.

First of all, from [4], it is enough to prove that, for any $\varepsilon > 0$, we have

$$\|A_{|H_r^\perp}\| \leq \|A_\star\| + \varepsilon, \tag{6}$$

with high probability. To achieve this, it is possible to restrict ourselves to self-adjoint operators. Indeed, the operator norm of an operator M is also the square root of the right-most point in the spectrum of the non-negative operator

$$\begin{pmatrix} 0 & M \\ M^* & 0 \end{pmatrix} = E_{12} \otimes M + E_{21} \otimes M^*,$$

where $E_{ij} \in M_2(\mathbb{C})$ is the canonical matrix whose entry (i, j) is equal to 1 all other entries are equal to 0. In particular, if A is of the form (3), at the cost of changing r in $2r$, we may assume without loss of generality that

$$a_{i^*} = a_i^* \quad \text{for all } i \in [\![2d]\!]. \tag{7}$$

For such a given operator A, we will define a family of companion operators, denoted by B_μ and indexed by a real parameter μ, with the property that if, for all $\varepsilon > 0$, with high probability, for all μ, we have

$$\rho(B_\mu) \leq \rho((B_\star)_\mu) + \varepsilon \tag{8}$$

then (6) holds. In the above expression, $(B_\star)_\mu$ is the companion operator of $A_\star$ defined in (4) and $\rho(M) = \sup(|\lambda| : \lambda \in \sigma(M))$ is the spectral radius. These operators B_μ are called the non-backtracking operators. This will be explained in Section 4. This part is an extension of [4] in a more general setting.

It is not a priori obvious why the claim (8) is easier to prove than claim (6). A reason is that the powers of $(B_\star)_\mu$ are much simpler to compute than the powers of $A_\star$.

In Section 5, we prove that (8) holds by using the expected high trace method popularized by Füredi and Komlós [12] in random matrix theory. It can be summatized as follows, assume that we aim at an upper bound of the form $\rho(M) \leq (1+o(1))\theta$ for some $\theta > 0$ and $M \in M_n(\mathbb{C})$ random. We observe that for any ℓ integer,

$$\rho(M)^{2\ell} \leq \|M^\ell\|^2 = \|M^\ell (M^*)^\ell\| \leq \mathrm{tr}(M^\ell (M^*)^\ell).$$

Moreover, at the last step, we loose a factor at most n (and typically of this order). If we can prove that

$$\mathbf{E}\mathrm{tr}(M^\ell (M^*)^\ell) \leq n\theta^{2\ell},$$

then we will deduce from Markov inequality that the probability that $\rho(M) \leq n^{1/(2\ell)}(1+\delta)\theta$ is at least $1-(1+\delta)^{-\ell}$. In particular, if $n^{1/(2\ell)} \to 1$, that is $\ell \gg \ln(n)$, then this last upper bound is sharp enough for our purposes.

A usual strategy to evaluate $\mathbf{E}\mathrm{tr}(M^\ell (M^*)^\ell)$ is to expand the trace and the powers as the sum of product matrix entries and then use the linearity of the expectation. We thus need to combine two ingredients: (i) a sharp upper bound on the expectation of the product of matrix entries in terms of combinatorial properties of the entries and (ii) a counting machinery to estimate the number of entries in the trace which have the given combinatorial properties.

In our setting, the matrix M is the non-backtracking matrix B_μ and $\theta = \rho((B_\star)_\mu)$ for a fixed μ. Among the difficulties, the value of θ is unknown exactly and also a probabilistic control of an event where all μ are considered is needed in (8). These issues were already present in [4]. Here, in addition, the presence of the tensor products will create a significant complication in the counting arguments of ingredient (ii). An important step for ingredient (i) is performed in Section 3 where we prove a new estimate on the expectation of the product of a large number of entries of a Haar unitary matrix. This will be done by developing on recent results on Weingarten calculus.

3 High order Gaussian approximation for random unitary matrices

The goal of this section is to develop an efficient machinery to compare expectation of products of entries of a Haar distributed unitary matrix with the average of the same product when we replace the unitary matrix by a complex Gaussian matrix. The main results of these sections are Theorem 11 and Theorem 12 below.

3.1 Wick calculus

In this subsection, we recall the classical Wick formula. We then introduce a centered version which is new.

3.1.1 Wick formula

Let V be a real vector space of real centered Gaussian variables. It is called a real Gaussian space. Similarly, let W be a complex vector space of complex centered Gaussian variables. It is called a complex Gaussian space. In both cases, the addition is the regular addition of real (resp. complex) valued random variables and, and likewise for the scaling. W has a natural real structure, in the sense that the collection $\{\alpha \mathrm{Re}(x) + \beta \mathrm{Im}(x), x \in W, \alpha, \beta \in \mathbb{R}\}$ is a real Gaussian vector space. Conversely, given $V, \tilde{V}$ two independent copies of a real vector Gaussian space (together with a canonical isomorphism $x \to \tilde{x}$), then the collection $\lambda(x + i\tilde{x}), x \in V, \lambda \in \mathbb{C}$ is a collection of complex random variables that all have the law of the standard complex distribution up to a scalar. This collection has a natural structure of complex vector space.

Let $(x_i)_{i\in I}$ be iid real centered Gaussian variables indexed by a countable set I. Then $\mathrm{span}((x_i)_{i\in I})$ is a real vector space, and any real vector space can be realized in this way. Finally, a real Gaussian vector space comes with the scalar product $(x, y) \mapsto \mathbf{E}(xy)$ and similarly, a complex Gaussian vector space comes with the Hilbert product $(x, y) \mapsto \mathbf{E}(\overline{x}y)$.

Wick's Theorem asserts that the scalar product is enough to recover completely the structure of the Gaussian space, in other words, there is a one to one correspondence (in law) between Gaussian spaces and their Hilbert structure, see Janson [18, Chapter 3]. Everything relies on a moment formula (which, it the case of Gaussian variables, determines the distribution). In the real case,

$$\mathbf{E}(x_1 \dots x_k) = \sum_{p \in P_k} \mathbf{E}_p(x_1, \dots, x_k),$$

where P_k is the collection of pair partitions of $[\![k]\!]$, typically denoted by

$$p = \{p_1, \dots, p_{k/2}\} = \{\{i_1, j_1\}, \dots, \{i_{k/2}, j_{k/2}\}\} \tag{9}$$

with $i_l < j_l$ and $i_l < i_{l+1}$ (obviously P_k is empty when k is odd), and, under this notation,

$$\mathbf{E}_p(x_1, \dots, x_k) = \mathbf{E}(x_{i_1} x_{j_1}) \dots \mathbf{E}(x_{i_{k/2}} x_{j_{k/2}}).$$

In the complex case,

$$\mathbf{E}(g_1 \dots g_k \overline{h}_1 \dots \overline{h}_k) = \sum_{\sigma \in S_k} \mathbf{E}_\sigma(g_1, \dots, g_k, h_1, \dots, h_k). \tag{10}$$

where S_k is the permutation group on $[\![k]\!]$ and

$$\mathbf{E}_\sigma(g_1,\dots,g_k,h_1,\dots,h_k)=\prod_{l=1}^{k}\mathbf{E}(g_l\overline{h}_{\sigma(l)}).$$

Note that the complex case can be deduced directly from the real case thanks to the real structure of complex Gaussian spaces (S_k appears as a subset of P_{2k} by identifying a permutation $\sigma\in S_{2k}$ with the partition $p=\{p_1,\dots,p_k\}$ of P_{2k} with $p_l=\{l,k+\sigma(l)\}$). As for the real case, the formula is trivial in the case where the x_i's are all the same (this is the formula for the moments of a Gaussian). For the general case, after observing that both the right-hand side and the left-hand side are k-linear and symmetric, we conclude by a polarization identity.

3.1.2 Centered Wick formula

If X is a vector valued integrable random variable, we call $[X]$ the centered random variable

$$[X]=X-\mathbf{E}(X). \tag{11}$$

For integer $T\geq 1$, the Wick formula can be extended as follows: if $\pi=(\pi_t)_{t\in[\![T]\!]}$ is a partition of $[\![k]\!]$, we have, for real Gaussian variables,

$$\mathbf{E}\left(\prod_{t=1}^{T}[\prod_{i\in\pi_t}x_i]\right)=\sum_{p\in P_k(\pi)}\mathbf{E}_p(x_1,\dots,x_k),$$

where $P_k(\pi)$ is the subset of P_k of pair partitions $p=\{p_1,\dots,p_{k/2}\}$ with the following property: for each block π_t of the partition, there exists at least one j such that the pair p_j of p has one element in π_t and one element outside. Although this formula is not mainstream in probability theory, it follows directly from standard sieve formulas (see Lemma 13 below for the complex case). Obviously, a similar formula holds for a complex Gaussian space:

$$\mathbf{E}\left(\prod_{t=1}^{T}[\prod_{i\in\pi_t}g_i\bar{h}_i]\right)=\sum_{\sigma\in S_k(\pi)}\mathbf{E}_\sigma(g_1,\dots,g_k,h_1,\dots,h_k),$$

where $S_k(\pi)$ is the subset of S_k of permutations σ such that for each block π_t, there exists at least one $i\in\pi_t$ such that $\sigma(i)\notin\pi_t$. We will use these results in the proof of Theorem 12 when we compare Gaussian moments and unitary moments.

3.2 Weingarten calculus with brackets

3.2.1 Unitary Weingarten formula

In the sequel, if $x = (x_1, \ldots, x_k)$ and $y = (y_1, \ldots, y_k)$ are multi-indices in $[\![n]\!]^k$, for $\sigma \in S_k$ we set

$$\delta_\sigma(x, y) = \prod_{l=1}^{k} \delta_{x_l, y_{\sigma(l)}},$$

where $\delta_{i,j}$ is the usual Kronecker delta (1 if $i = j$ and 0 otherwise). Similarly, if k is even and $p \in P_k$ is a pair partition, $\delta_p(x)$ is a generalized Kronecker delta: namely, $\delta_p(x)$ takes the value 1 if for any block of p, the coefficients of x are the same, and it takes the value 0 in all other cases. This is a product of $k/2$ Kronecker deltas.

We start by recalling the analog of Wick calculus for the normalized Haar measure on the unitary group $\mathbb{U}_n$.

Theorem 4. *Let $k, n \geq 1$ be integers and x, y, x', y' in $[\![n]\!]^k$. There exists a function $\mathrm{Wg}(\cdot, \cdot, n) : S_k \times S_k \to \mathbb{R}$ such that if $U = (U_{ij})$ is Haar distributed on $\mathbb{U}_n$,*

$$\mathbf{E}\left(\prod_{l=1}^{k} U_{x_l y_l} \overline{U}_{x'_l y'_l} \right) = \sum_{p,q \in S_k} \delta_p(x, x') \delta_q(y, y') \mathrm{Wg}(p, q, n).$$

This function is uniquely defined iff $k \leq n$, see Theorem 5 for a formula.

The sequel of this subsection is devoted to providing a modern description of Wg. It follows from commutativity that $\mathrm{Wg}(p, q, n)$ actually only depends on the conjugacy class of pq^{-1} so in the unitary case we will write $\mathrm{Wg}(pq^{-1}, n) = \mathrm{Wg}(p, q, n)$, so Wg becomes a central function on S_k.

Let $\sigma \in S_k$. We denote by $|\sigma| = k - \ell(\sigma)$ where $\ell(\sigma)$ is the number of disjoint cycles in the cycle decomposition of σ. Classically, $|\sigma|$ is also the minimal number m such that σ can be written as a product of m transpositions. In particular $(-1)^{|\sigma|}$ is the signature of σ. For $l \geq 0$, we call $P(\sigma, l)$ the collection of solutions of

$$\sigma = (i_1, j_1) \ldots (i_{|\sigma|+l}, j_{|\sigma|+l}), \tag{12}$$

where $j_p \leq j_{p+1}$ and $i_p < j_p$. Note that $P(\sigma, l) = 0$ unless $l = 2g$ is even (since composing by a single transposition alternates the signature).

The following theorem is a combination of Theorem 2.7 and Lemma 2.8 in [8] (beware that the definition of the map $l \to P(\sigma, l)$ is shifted by $|\sigma|$ in this reference), see also [22]. In the statement below and in the sequel, if S is a finite set, we denote by $|S|$ its cardinal number (not to be confused with $|\sigma|$ for $\sigma \in S_k$).

Theorem 5. *For $\sigma \in S_k$, we have the expansion*

$$\mathrm{Wg}(\sigma, n) = (-1)^{|\sigma|} n^{-k-|\sigma|} \sum_{g \geq 0} |P(\sigma, 2g)| n^{-2g}. \tag{13}$$

This expansion is formal in the sense that $\mathrm{Wg}(\sigma, n)$ is a rational fraction in n and its power series expansion in the neighbourood of infinity is as above.

Since the poles of $n \to \mathrm{Wg}(\sigma, n)$ are known to be in the set $\{-k+1, \ldots, k-1\}$, it follows that the power series expansion is actually convergent as soon as $n \geq k$. The following result can be found in [8, Theorem 3.1], it is an estimate of the number of solutions of (12). It allows subsequently to give estimates on the Weingarten function.

Proposition 6. *Let k be a positive integer. For any permutation $\sigma \in S_k$ and integer $g \geq 0$, we have*

$$(k-1)^g |P(\sigma, 0)| \leq |P(\sigma, 2g)| \leq \left(6k^{7/2}\right)^g |P(\sigma, 0)|.$$

For our purposes we will just need the following corollary:

Corollary 7. *If $\sigma \in S_k$ and $12k^{7/2} \leq n^2$,*

$$|\mathrm{Wg}(\sigma, n)| \leq \left(1 + 24k^{7/2} n^{-2}\right) n^{-k-|\sigma|} 4^{|\sigma|}.$$

Proof. By [22, Corollary 2.11], we have

$$|P(\sigma, 0)| = \prod_{i=1}^{\ell(\sigma)} C_{\mu_i - 1},$$

where C_n is the n-th Catalan number and μ_i is the length of the i-th cycle of σ. Since $C_n \leq 4^n$, we get $|P(\sigma, 0)| \leq 4^{|\sigma|}$. From Proposition 6, we deduce that for any integer $g \geq 0$,

$$|P(\sigma, 2g)| \leq 4^{|\sigma|} \left(6k^{7/2}\right)^g. \tag{14}$$

From Theorem 5, we deduce that

$$|\mathrm{Wg}(\sigma, n)| \leq n^{-k-|\sigma|} 4^{|\sigma|} \sum_{g \geq 0} \left(\frac{6k^{7/2}}{n^2}\right)^g.$$

The conclusion follows by using that $(1-x)^{-1} \leq 1 + 4x$ for all $0 \leq x \leq 1/2$. □

3.2.2 The centered case

For a symbol $\varepsilon \in \{\cdot, -\}$ and $z \in \mathbb{C}$, we take the notation that $z^\varepsilon = z$ if $\varepsilon = \cdot$ and $z^\varepsilon = \overline{z}$ if $\varepsilon = -$. Our goal is to compute for $U = (U_{ij})$ Haar distributed on $\mathbb{U}_n$, expresssions of the form

$$\mathbf{E}\left(\prod_{t=1}^{T}[\prod_{l=1}^{k_t} U_{x_{tl}y_{tl}}^{\varepsilon_{tl}}]\right), \tag{15}$$

where we have use the bracket defined in (11) and then to estimate it in a useful way.

The polynomial to be integrated can be expanded into 2^T terms for which the Weingarten formula can be applied each time. However, such an approach does not yield good estimates because the sum is signed and the items to be summed do not have the right decay in large dimension. On the other hand, it turns out that it is possible to obtain a non-signed formula which (necessarily) yields the right asymptotics. This is the purpose of what follows.

Proposition 8. *Let k be even, $\pi = (\pi_t)_{t\in[\![T]\!]}$ be a partition of $[\![k]\!]$, let $\varepsilon \in \{\cdot, -\}^k$ be a balanced sequence (in the sense that it has as many $\cdot$ than $-$) and x, y in $[\![n]\!]^k$. There exists a generalized Weingarten function $\mathrm{Wg}[\pi](p, q, n)$ such that*

$$\mathbf{E}\left(\prod_{t=1}^{T}[\prod_{i\in\pi_t} U_{x_iy_i}^{\varepsilon_i}]\right) = \sum_{p,q\in P_k^\varepsilon} \delta_p(x)\delta_q(y)\mathrm{Wg}[\pi](p, q, n),$$

where $P_k^\varepsilon \subset P_k$ are the pair partitions that match an $\varepsilon_i = \cdot$ with an $\varepsilon_{i'} = -$ (seen as bijections from the set of i's such that $\varepsilon_i = \cdot$ to the set of i's such that $\varepsilon_i = -$).

Proof. Let $A \subset [\![T]\!]$ and $X_1, \ldots, X_T$ be random variables. We introduce the following temporary notation:

$$\mathbf{E}_A(X_1, \ldots, X_T) = \mathbf{E}\left(\prod_{t\in A} X_t\right)\prod_{t\notin A}\mathbf{E}(X_t).$$

It follows from this definition that

$$\mathbf{E}([X_1]\ldots[X_T]) = \sum_{A\subset[\![T]\!]} \mathbf{E}_A(X_1, \ldots, X_T)(-1)^{T-|A|}. \tag{16}$$

We apply this equation to

$$X_t := \prod_{i\in\pi_t} U_{x_iy_i}^{\varepsilon_i}.$$

We claim that for a given A, there exists a function $\mathrm{Wg}_A[\pi](p, q, n)$ such that

$$\mathbf{E}_A(X_1, \ldots, X_T) = \sum_{p,q\in P_k^\varepsilon} \delta_p(x)\delta_q(y)\mathrm{Wg}_A[\pi](p, q, n).$$

In order to define Wg_A precisely, we call π_A the partition of $[\![k]\!]$ whose blocks are $\pi_t, t \notin A$, and a last block that complements the previous ones (in other words, we simply merge all blocks in A and leave the other unchanged). From Theorem 4, this yields the following description of Wg_A: if either the permutation p or the permutation q fails to respect the partition π_A (that is does not leave all blocks invariant) then its value is zero. Or else, it is a product of Weingarten functions obtained by restricting the permutations over the block of π_A:

$$\mathrm{Wg}_A[\pi](p,q,n) = \prod_{b\in\pi_A} \mathrm{Wg}(p_b, q_b, n), \tag{17}$$

where the product is over all blocks b of π_A and p_b, q_b are the restrictions of p, q to the block b. The fact that the above holds true follows from an application of the Weingarten formula, Theorem 4, for each expectation factor appearing in the product.

In turn, the explicit formula we obtain for the generalized Weingarten formula becomes

$$\mathrm{Wg}[\pi](p,q,n) = \sum_{A\subset[\![T]\!]} (-1)^{T-|A|}\mathrm{Wg}_A[\pi](p,q,n), \tag{18}$$

and this concludes the proof. . □

Example. We take $T = 2$ and for $k_1 \in [\![k]\!]$, $\pi_1 = [\![k_1]\!]$, $\pi_2 = [\![k]\!]\backslash[\![k_1]\!]$. Then, if p or q do not leave invariant π, we have $\mathrm{Wg}[\pi](p,q,n) = \mathrm{Wg}(p,q,n)$. Otherwise, we get $\mathrm{Wg}[\pi](p,q,n) = \mathrm{Wg}(p,q,n) - \mathrm{Wg}(p_{\pi_1}, q_{\pi_1}, n)\mathrm{Wg}(p_{\pi_2}, q_{\pi_2}, n)$, where p_{π_t} is the restriction of p to the block π_t.

We now give an analog of Theorem 5. We consider the setting of Proposition 8. Let $\varepsilon \in \{\cdot, -\}^k$ be a balanced sequence and π be a partition of $[\![k]\!]$. Recall that we identify pair partitions $p, q \in P_k^\varepsilon$ with bijections from the i's such that $\varepsilon_i = \cdot$ to the set of i's such that $\varepsilon_i = -$. For $p, q \in P_k^\varepsilon$, we call $P[\pi](p,q,l)$ the collection of solutions of

$$p = (i_1, j_1)\dots(i_{|pq^{-1}|+l}, j_{|pq^{-1}|+l})q \tag{19}$$

where $j_p \leq j_{p+1}$, $i_p < j_p$ and $\varepsilon_{i_p} = \varepsilon_{j_p} = \cdot$, and that satisfies the following property: the solution can be restricted to *no single block* of π in the following sense: if there exists a block b of π such that p, q and each transposition (i_p, j_p) leave b invariant, then the solution is not in $P[\pi](p,q,l)$, otherwise, the solution is in $P[\pi](p,q,l)$.

These notations allow us to reformulate combinatorially the centered Weingarten function used in Proposition 8:

Theorem 9. *Let k be even, $\pi = (\pi_t)_{t\in[\![T]\!]}$ be a partition of $[\![k]\!]$, let $\varepsilon \in \{\cdot, -\}^k$ be a balanced sequence. For all $p, q \in P_k^\varepsilon$, we have the expansion*

$$\mathrm{Wg}[\pi](p,q,n) = (-1)^{|pq^{-1}|} n^{-k/2-|pq^{-1}|} \sum_{g\geq 0} |P[\pi](p,q,2g)| n^{-2g}. \tag{20}$$

Proof. We set $\sigma = pq^{-1}$, it is a bijection on the set $[\![\dot{k}]\!]$ of elements $[\![k]\!]$ such that $\varepsilon_i = \cdot$. For $A \subset [\![T]\!]$, we start with the formula for $\mathrm{Wg}_A[\pi]$ in (17), where we recall that π_A is the partition obtained for $\pi = (\pi_t)$ by merging all blocks in A. We apply Theorem 5 to each term $\mathrm{Wg}(p_b, q_b, n) = \mathrm{Wg}(p_b q_b^{-1}, n)$ in the product (17). One finds that

$$\mathrm{Wg}_A[\pi](p, q, n) = (-1)^{|\sigma|} n^{-k/2-|\sigma|} \sum_{g \geq 0} |P_A[\pi](p, q, 2g)| n^{-2g}, \tag{21}$$

where $P_A[\pi](p, q, 2g)$ is the empty set if either p or q do not respect the partition π_A (that is, do not leave all blocks of π_A invariant) and else, it is the collection of solutions of (19), $\sigma = (i_1, j_1) \dots (i_{|\sigma|+2g}, j_{|\sigma|+2g})$ where $j_p \leq j_{p+1}$ and $i_p < j_p$ in $[\![\dot{k}]\!]$ and every transposition (i_p, j_p) respects the partition π_A. This follows from the facts that the condition "$j_p \leq j_{p+1}$ and $i_p < j_p$" is a total order on transpositions and that all other quantities in the sum (13) are multiplicative over blocks of π_A.

We observe from the definitions that $P_{[\![T]\!]}[\pi](p, q, 2g) = P(pq^{-1}, 2g)$ and that

$$P[\pi](p, q, 2g) = P(pq^{-1}, 2g) \setminus \bigcup_{t=1}^{T} P_{[\![T]\!] \setminus \{t\}}[\pi](p, q, 2g).$$

Indeed, $P_{[\![T]\!] \setminus \{t\}}[\pi](p, q, 2g)$ is the set of solutions of (19) such that p, q and the transpositions leave π_t invariant. In addition, from the definition, we have for all sets A, B in $[\![T]\!]$,

$$P_A[\pi](p, q, 2g) \cap P_B[\pi](p, q, 2g) = P_{A \cap B}[\pi](p, q, 2g).$$

We recall that the sieve formula asserts that any sets $S_t \subset S$ we have

$$\left| S \setminus \bigcup_{t=1}^{T} S_t \right| = |S| + \sum_{\emptyset \neq B \subset [\![T]\!]} (-1)^{|B|} \left| \bigcap_{t \in B} S_t \right| = |S| + \sum_{A \subsetneq [\![T]\!]} (-1)^{T-|A|} \left| \bigcap_{t \notin A} S_t \right|, \tag{22}$$

We apply the sieve formula to the sets $S_t = P_{[\![T]\!] \setminus \{t\}}[\pi](p, q, 2g)$ and $S = P(pq^{-1}, 2g)$. From what precedes

$$\bigcap_{t \notin A} P_{[\![T]\!] \setminus \{t\}}[\pi](p, q, n) = P_{\bigcap_{t \notin A} [\![T]\!] \setminus \{t\}}[\pi](p, q, n) = P_A[\pi](p, q, n).$$

We thus have proved that

$$P[\pi](p, q, 2g) = \sum_{A \subset [\![T]\!]} (-1)^{T-|A|} |P_A[\pi](p, q, 2g)|.$$

Plugging this last identity in Equation (18) and Equation (21), this completes the proof. □

Out of this, we are able to propose the key estimate for the centered Weingarten function. In the statement below, if p and q are two partitions of $[\![k]\!]$ then $p \vee q$ is the finest partition that is coarser than both p and q.

Theorem 10. *Let k even with $2k^{7/2} \leq n^2$, $\pi = (\pi_t)_{t \in [\![T]\!]}$ be a partition of $[\![k]\!]$, let $\varepsilon \in \{\cdot, -\}^k$ be a balanced sequence. For all $p, q \in P_k^{\varepsilon}$, the following estimate holds true:*

$$|\mathrm{Wg}[\pi](p, q, n)| \leq (1 + 3k^{7/2}n^{-2})n^{-k/2-|pq^{-1}|}4^{|pq^{-1}|}(k^{7/4}n^{-1})^r,$$

where r is the number of blocks of π to which $p \vee q$ can be restricted.

Before we supply the proof, we give the main idea, which is quite simple: since in Theorem 9, we realize Weingarten functions as unsigned sums, is is enough to estimate each summand separately, In turn, our estimate is quite blunt, and relies solely on the inclusion $P[\pi](p, q, l) \subset P(p, q, l)$. In other words an estimate that is good enough for our purposes is achieved just thanks to the fact that a partial connectedness condition kills the first terms of a series.

Proof of Theorem 10. Set $\sigma = pq^{-1}$. From Theorem 9 we have

$$|\mathrm{Wg}[\pi](p, q, n)| = n^{-k/2-|\sigma|} \sum_{g \geq 0} |P[\pi](p, q, 2g)| n^{-2g}.$$

We observe that $P[\pi](p, q, 2g) \subset P(p, q, 2g)$. We also claim that $P[\pi](p, q, 2g) = \emptyset$ as soon as $2g < r$. Indeed, if b is a block of $p \vee q$, then any solution of (19) with $l = 0$ satisfy i_p, j_p is in b. Hence, if a block π_t is a union of blocks of $p \vee q$ then at least two extra transpositions from an element in π_t to another block have to be added to be an admissible solution of (19). These two extra transpositions could be shared between two such blocks of π. It follows that $P[\pi](p, q, 2g) \neq \emptyset$ implies $2g \geq r$ and

$$|\mathrm{Wg}[\pi](p, q, n)| \leq n^{-k/2-|\sigma|} \sum_{g \geq r/2} |P(p, q, 2g)| n^{-2g}.$$

The right-hand side can be estimated thanks to (14) (applied to $k/2$). We get if $c = 3/2^{5/2}$ and $ck^{7/2}n^{-2} < 1$,

$$|\mathrm{Wg}[\pi](p, q, n)| \leq n^{-k/2-|\sigma|}4^{|\sigma|} \frac{(ck^{7/2}n^{-2})^{r/2}}{1 - ck^{7/2}n^{-2}}.$$

If we assume further $2ck^{7/2}n^{-2} \leq 1$, we obtain

$$|\mathrm{Wg}[\pi](p, q, n)| \leq \left(1 + \frac{4ck^{7/2}}{n^2}\right) n^{-k/2-|\sigma|}4^{|\sigma|}(ck^{7/2}n^{-2})^{r/2},$$

as requested (since $c \simeq 0.53$). □

3.3 From Weingarten calculus to Wick calculus

3.3.1 Case without brackets

We start with x, y in $[\![n]\!]^k$ and a balanced sequence $\varepsilon \in \{\cdot, -\}^k$. If $U = (U_{ij})$ is Haar distributed on $\mathbb{U}_n$ We want to compare $|\mathbf{E}(U^{\varepsilon_1}_{x_1y_1} \dots U^{\varepsilon_k}_{x_ky_k})|$ with the matrix U replaced by $G_{ij}/\sqrt{n}$, where G_{ij} are independent complex standard Gaussian variables.

We need a new definition. Let $x, y \in [\![n]\!]^k$. If $u \in [n]$, we define the number of *left arms* of $u \in [n]$ in (x, y) as $\sum_i \mathbf{1}(x_i = u)$ and the number of *right arms* as $\sum_i \mathbf{1}(y_i = u)$. The pair (x, y) is called an *even sequence* if for any $u \in [n]$, the number of left and right arms are even.

Our result is as follows

Theorem 11. *Let k be even, x, y in $[\![n]\!]^k$ and let $\varepsilon \in \{\cdot, -\}^k$ be a balanced sequence. If $n \geq 4$ and $2k^{7/2} \leq n^2$ then*

$$n^{k/2} \left| \mathbf{E}\left(\prod_{i=1}^{k} U^{\varepsilon_i}_{x_iy_i} \right) \right| \leq \left(1 + 3k^{7/2}n^{-2}\right) \mathbf{E}\left(\prod_{i=1}^{k} \left(G^{\varepsilon_i}_{x_iy_i} + kn^{-1/4} \right) \right).$$

Moreover, the above expectation on the left-hand side is zero unless (x, y) is an even sequence.

Proof. From Theorem 4, we have

$$n^{k/2}\mathbf{E}\left(\prod_{i=1}^{k} U^{\varepsilon_i}_{x_iy_i} \right) = \sum_{p,q \in P_k^{\varepsilon}} \delta_p(x)\delta_q(y) n^{k/2} \mathrm{Wg}(p, q, n), \tag{23}$$

where we have identified a pair partition in P_k^{ε} with a bijection from the set $[\![\dot{k}]\!] \subset [\![k]\!]$ of i's such that $\varepsilon_i = \cdot$ to the set of i's such that $\varepsilon_i = -$. The final statement of the theorem follows directly.

Given a pair (p, q) involved in the right-hand side of (23), we introduce a subset A of $[\![k]\!]$ which is the maximal subset such that p_A, q_A, the restrictions of p and q to A, are well defined and describe the *same* pair partition. In other words, A is the union of all common pairs of p and q. Note that A could be empty or $[\![k]\!]$, but it has an even number of elements.

Our strategy is, to evaluate simultaneously all (p, q) that yield a same A. If we can find $\delta, \eta \geq 0$ such that for all $A \subset [\![k]\!]$, the sum in (23) restricted to pairs (p, q) that yield A is bounded above by

$$(1 + \delta)\mathbf{E}\left(\prod_{i \in A} G^{\varepsilon_i}_{x_iy_i} \right) \eta^{|A^c|} \tag{24}$$

then we would deduce that

$$n^{k/2} \left| \mathbf{E}\left(\prod_{i=1}^{k} U^{\varepsilon_i}_{x_iy_i} \right) \right| \leq (1 + \delta)\mathbf{E}\left(\prod_{i=1}^{k} \left(G^{\varepsilon_i}_{x_iy_i} + \eta \right) \right)$$

since we have the following expansion:

$$\mathbf{E}\left(\prod_{i=1}^{k}\left(G^{\varepsilon_i}_{x_i y_i}+\eta\right)\right)=\sum_{A\subset[\![k]\!]}\mathbf{E}\left(\prod_{i\in A}G^{\varepsilon_i}_{x_i y_i}\right)\prod_{i\in A^c}\eta.$$

To that end, we set $c=3/\sqrt{2}$ and $\sigma=pq^{-1}$ (seen as a bijection on $[\![\dot{k}]\!]$). According to Corollary 7 (applied to $k/2$), if $ck^{7/2}\leq n^2$ and $\delta=2ck^{7/2}/n^2$, we have

$$n^{k/2}\mathrm{Wg}(p,q,n)\leq(1+\delta)\left(\frac{4}{n}\right)^{|\sigma|}.$$

It is standard that if $\tau\in S_m$ then $|\tau|=m-\sum_l c_l$ where c_l is the number of cycles of length l. Since $\sum lc_l=m\geq 2(\sum_l c_l)-c_1$, we get $|\tau|\geq(m-c_1)/2$. By assumption, all cycles of σ in $A^c\cap[\![\dot{k}]\!]$ have length at least 2, so the number of cycles of length 1 of σ is at most $|A\cap[\![\dot{k}]\!]|=|A|/2$. We deduce that

$$|\sigma|\geq(k-|A|)/4=|A^c|/4.$$

We thus have proved that, if $n\geq 4$, and $\eta_0=(4/n)^{1/4}$,

$$n^{k/2}\mathrm{Wg}(p,q,n)\leq(1+\delta)\eta_0^{|A^c|}.$$

Therefore, the sum in (23) restricted to pair partitions p,q that yield A is upper bounded by

$$(1+\delta)\eta_0^{|A^c|}((|A^c|/2)!)^2\sum_{\tau\in P_A^\varepsilon}\delta_\tau(x)\delta_\tau(y),$$

where P_A^ε is the set of bijections on A from the set i's such that $\varepsilon_i=\cdot$ to the set of i's such that $\varepsilon_i=-$. Moreover, the term $((|A^c|/2)!)^2$ accounts for the choices of (p_{A^c},q_{A^c}). From Wick formula (10), we have

$$\mathbf{E}\left(\prod_{i\in A}G^{\varepsilon_i}_{x_i y_i}\right)=\sum_{\tau\in P_A^\varepsilon}\delta_\tau(x)\delta_\tau(y).$$

Finally, we recall that $m!\leq(me^{-1})^m$. We deduce that (24) holds with δ as above and $\eta=\eta_0 e^{-1}(k/2)$. It concludes the proof. □

3.3.2 Case with brackets

We now move to the case with brackets. If $U=(U_{ij})$ is Haar distributed on $\mathbb{U}_n$, we want to compare expectations as in Equation (15) with the matrix U replaced by $G_{ij}/\sqrt{n}$, where G_{ij} are independent complex standard Gaussian variables. The main result in this direction is

Theorem 12. *Let k even with $2k^{7/2} \le n^2$ and $n \ge 4$, $\pi = (\pi_t)_{t\in[\![T]\!]}$ be a partition of $[\![k]\!]$ such that each block has at most ℓ elements. Let $\varepsilon \in \{\cdot, -\}^k$ be a balanced sequence. For any x, y in $[\![n]\!]^k$, we have*

$$n^{k/2}\left|\mathbf{E}\left(\prod_{t=1}^{T}[\prod_{i\in\pi_t} U^{\varepsilon_i}_{x_i y_i}]\right)\right| \le (1+\delta)\mathbf{E}\left(\prod_{t=1}^{T}\left([\prod_{i\in\pi_t} G^{\varepsilon_i}_{x_i y_i}]+\eta\right)\right),$$

with $\delta = 3k^{7/2}n^{-2}$ and $\eta = 2k^{\ell}n^{-1/4}$. Moreover, if each block π_t contains an element with $\varepsilon_i = \cdot$ and another with $\varepsilon_i = -$, the same bound holds with $\eta = 2k^{\ell}n^{-1/2}$. Finally, the above expectation on the left-hand side is zero unless (x, y) is an even sequence.

We start by evaluating the average of products of brackets of shifted Gaussian variables.

Lemma 13. *Let k be even, let $\varepsilon \in \{\cdot, -\}^k$ be a balanced sequence and let $\pi = (\pi_t)_{t\in[\![T]\!]}$ be a partition of $[\![k]\!]$. If $(g_i)_{i\in[\![k]\!]}$ be a complex Gaussian vector then, for any complex number η, we have*

$$\mathbf{E}\left(\prod_{t=1}^{T}[\prod_{i\in\pi_t} g_i^{\varepsilon_i}]\right) = \sum_{p\in P^{\varepsilon}(\pi)}\prod_{(i,j)\in p}\mathbf{E}(g_i\bar{g}_j),$$

where $P^{\varepsilon}(\pi)$ is the set (possibly empty) of pair partitions on $[\![k]\!]$ such that all pairs (i, j) satisfy $\varepsilon_i = \cdot$, $\varepsilon_j = -$ and such that for each block of π there exists a pair (i, j) with one element in the block and the other outside.

Proof. Using (16), we have

$$\mathbf{E}\left(\prod_{t=1}^{T}[\prod_{i\in\pi_t} g_i^{\varepsilon_i}]\right) = \sum_{A\subset[\![T]\!]}(-1)^{T-|A|}\mathbf{E}\left(\prod_{t\in A}\prod_{i\in\pi_t} g_i^{\varepsilon_i}\right)\prod_{t\notin A}\mathbf{E}\left(\prod_{i\in\pi_t} g_i^{\varepsilon_i}\right).$$

For a given $A \subset [\![T]\!]$, let π_A be the partition obtained from π by merging all blocks in A into a single block. We apply (10) to each expectation, arguing as in the proof of Theorem 9, we get that

$$\mathbf{E}\left(\prod_{t\in A}\prod_{i\in\pi_t} g_i^{\varepsilon_i}\right)\prod_{t\notin A}\mathbf{E}\left(\prod_{i\in\pi_t} g_i^{\varepsilon_i}\right) = \sum_{p\in P_A^{\varepsilon}} E_p,$$

where P_A^{ε} is the set (possibly empty) of pair partitions on $[\![k]\!]$ such that all pairs (i, j) satisfy $\varepsilon_i = \cdot$, $\varepsilon_j = -$ and i and j are in the same block of π_A. Moreover, if $p \in P_k^{\varepsilon} = P_{[\![T]\!]}^{\varepsilon}$ we have set

$$E_p = \prod_{(i,j)\in p}\mathbf{E}(g_i\bar{g}_j).$$

We thus have checked the identity

$$\mathbf{E}\left(\prod_{t=1}^{T}[\prod_{i\in\pi_t} g_i^{\varepsilon_i}]\right) = \sum_{p\in P_k^{\varepsilon}} E_p \sum_{A\subset[\![T]\!]}(-1)^{T-|A|}\mathbf{1}(p\in P_A^{\varepsilon}).$$

By an application of the sieve formula (22) to $S = P_k^\varepsilon$ and $S_t = P^\varepsilon_{[\![T]\!]\setminus\{t\}}$, we find that

$$\sum_{A\subset[\![T]\!]} (-1)^{T-|A|}\mathbf{1}(p\in P_A^\varepsilon)$$

is 1 or 0 depending on $p \in P^\varepsilon(\pi)$ or not (we argue as in the proof of Theorem 9). □

Proof of Theorem 12. We follow the same strategy as in the proof of Theorem 11. By Proposition 8,

$$n^{k/2}\mathbf{E}\left(\prod_{t=1}^{T}[\prod_{i\in\pi_t} U^{\varepsilon_i}_{x_iy_i}]\right) = \sum_{p,q\in P_k^\varepsilon} \delta_p(x)\delta_q(y) n^{k/2}\mathrm{Wg}[\pi](p,q,n), \tag{25}$$

where as usual we have identified a pair partition in P_k^ε with a bijection from the set $[\![\dot{k}]\!] \subset [\![k]\!]$ of i's such that $\varepsilon_i = \cdot$ to the set of i's such that $\varepsilon_i = -$. The last statement of the theorem follows directly.

On the other hand, for $B \subset [\![T]\!]$, we set $[\![k]\!]_B = \cup_{t\in B}\pi_t$. We define P_B^ε as the set of pair partitions on $[\![k]\!]_B$ whose pairs (i,j) are such that $\varepsilon_i = \cdot$, $\varepsilon_j = -$ and we let $P_B^\varepsilon(\pi) \subset P_B^\varepsilon$ be the pair partitions such that for each block of π there exists a pair (i,j) with one element in the block and the other outside. By Lemma 13, we have, for any numbers η_1, η_2,

$$\begin{aligned}\mathbf{E}\left(\prod_{t\in T}\left([\prod_{i\in\pi_t}(G^{\varepsilon_i}_{x_iy_i})] + \eta_1 + \eta_2\right)\right) &= \sum_{A\subset[\![T]\!]} \eta_2^{|A|^c}\sum_{B\subset A}\eta_1^{|B^c|}\sum_{\tau\in P_B^\varepsilon(\pi)}\mathbf{E}\left(\prod_{t\in B}[\prod_{i\in\pi_t}(G^{\varepsilon_i}_{x_iy_i})]\right)\\ &= \sum_{A\subset[\![T]\!]} \eta_2^{|A|^c}\sum_{B\subset A}\eta_1^{|B^c|}\sum_{\tau\in P_B^\varepsilon(\pi)}\delta_\tau(x)\delta_\tau(y), \end{aligned}\tag{26}$$

where $B^c = A\setminus B$ is the complement of B in A.

Given a pair (p,q) involved in the right-hand side of (25), we introduce a subset A of $[\![T]\!]$ which is the maximal subset such that p_A, q_A, the restrictions of p and q to $[\![k]\!]_A$, are well defined and describe the *same* pair partition, say $\tau \in P_A^\varepsilon$. According to Theorem 10, if $\sigma = pq^{-1}$, we have

$$n^{k/2}|\mathrm{Wg}[\pi](p,q,n)| \le (1+\delta)\left(\frac{4}{n}\right)^{|\sigma|}\eta^r,$$

with $\delta = 3k^{7/2}n^{-2}$, $\eta = k^{7/4}n^{-1}$ and r is the number of blocks of π to which $p\vee q$ can be restricted. The number of cycles of σ of length 1 is at most $k/2 - |A^c|/2$ (remark: it is at most $k/2 - |A^c|$ if each block contains an element with $\varepsilon_i = \cdot$ and another with $\varepsilon_i = -$). As in the proof of Theorem 11, we deduce that $|\sigma| \ge |A|^c/4$ and, if $n \ge 4$,

$$n^{k/2}|\mathrm{Wg}[\pi](p,q,n)| \le (1+\delta)\eta_0^{|A^c|}\eta^r,$$

with $\eta_0 = (4/n)^{1/4}$. We note also that r is additive over the restrictions of π to $[\![k]\!]_A$ and $[\![k]\!]_{A^c}$. Since there are at most $((|[\![k]\!]_{A^c}|/2)!)^2$ choices for the restrictions of p and q to $[\![k]\!]_{A^c}$, the contribution in (25) of the sum over all pairs (p, q) which yields the same set A is thus upper bounded by

$$(1+\delta)\eta_0^{|A^c|}((|[\![k]\!]_{A^c}|/2)!)^2 \sum_{\tau \in P_A^\varepsilon} \eta^{r(\tau)} \delta_\tau(x)\delta_\tau(y), \tag{27}$$

where $r(\tau)$ is the number of blocks of π which is a union of pairs of τ.

By assumption $|[\![k]\!]_{A^c}| \leq \ell|A|^c$. Using $m! \leq (me^{-1})^m$, we deduce that the expression (27) is upper bounded by

$$(1+\delta)\eta_2^{|A^c|} \sum_{\tau \in P_A^\varepsilon} \eta^{r(\tau)} \delta_\tau(x)\delta_\tau(y), \tag{28}$$

with $\eta_2 = k^\ell n^{-1/4}$. We now give a closer look at the sum in (28). Let us call $B \subset A$ the complement in A of the blocks of π which are union of pairs of τ. We have $r(\tau) = |B^c|$ where $B^c = A \backslash B$. For a given set $B \subset [\![k]\!]_A$, let us call $P_{A,B}^\varepsilon$ the subset P_A^ε which yields the set B. By assumption, the restriction of τ in $P_{A,B}^\varepsilon$ to B is a pair partition in $P_B^\varepsilon(\pi)$. Conversely, recall that the block of π have at most ℓ elements. For a given pair partition τ' in $P_B^\varepsilon(\pi)$, there are at most $(\ell-1)^{|B^c|/2}$ partitions in $P_{A,B}^\varepsilon$ whose restriction to B is τ'. If $\eta_1 = \ell^{\ell/2}\eta$, we thus have proved that

$$\sum_{\tau \in P_A^\varepsilon} \eta^{r(\tau)} \delta_\tau(x)\delta_\tau(y) \leq \sum_{B \subset A} \eta_1^{|B^c|} \sum_{\tau \in P_B^\varepsilon(\pi)} \delta_\tau(x)\delta_\tau(y).$$

Note also that $\eta_1 \leq \eta_2$ for our choice of k. In view of Equation (26), this concludes the proof. □

Remark. Note by passing that the lower bound in Proposition 6 can be used to show that there is some symmetry in Theorem 11 and Theorem 12. Namely, it is possible in these results to swap the roles of the Gaussian entries and of the unitary entries (the values of the constants δ, η need to be adjusted). This just an esthetic comment about the sharpness as the comparison, as we just need the bound as stated in Theorem 12.

3.4 Moment bounds for product of unitaries

We conclude this section with a corollary of Theorem 12. Let x, y be two sequences in $[\![n]\!]^k$. The *multiplicity* of $e = (a, b) \in [\![n]\!]^2$ is defined as $\sum_i \mathbf{1}((x_i, y_i) = e)$. The set of pairs of multiplicity at least one is the set of visited pairs $\cup_i\{(x_i, y_i)\}$. Moreover if $\pi = (\pi_t)_{t \in [\![T]\!]}$ is a partition of $[\![k]\!]$, we say that π_t is an *isolated* block of (x, y) if for all $i \in \pi_t$, for all , $(x_i, y_i) \neq (x_j, y_j)$ for all $j \in [\![k]\!]\backslash\pi_t$ (in other words, (x_i, y_i) is of multiplicity 0 in the sequence $(x_j, y_j)_{j \in [\![k]\!]\backslash\pi_t}$).

Corollary 14. *Let $k = qT$ even with $k^{q+1} \le n^{1/4}$, $\pi = (\pi_t)_{t\in[\![T]\!]}$ be a partition of $[\![k]\!]$ such that each block contains q elements. Let $\varepsilon \in \{\cdot, -\}^k$ be a balanced sequence. For any x, y in $[\![n]\!]^k$, we have, for some universal constant $c > 0$,*

$$\left|\mathbf{E}\left(\prod_{t=1}^{T}[\prod_{i\in\pi_t} U_{x_iy_i}^{\varepsilon_i}]\right)\right| \le cn^{-\frac{k}{2}}\eta^{b+\frac{e_1}{q}}k^{\frac{m_4}{2}},$$

where $\eta = ck^{q/2}n^{-1/8}$, e_1 is the number of pairs of multiplicity 1, b is the number of isolated (x_t, y_t) and m_4 is the sum of multiplicities of pairs with multiplicity at least 4.

Proof. Let η, δ be as in Theorem 12 with $\ell = q$. From (26), we have

$$I = (1+\delta)^{-1}n^{k/2}\left|\mathbf{E}\left(\prod_{t=1}^{T}[\prod_{i\in\pi_t} U_{x_iy_i}^{\varepsilon_i}]\right)\right| \le \sum_{A\subset[\![T]\!]}\eta^{|A|^c}\sum_{\tau\in P_A^\varepsilon(\pi)}\delta_\tau(x)\delta_\tau(y). \tag{29}$$

From the definition of $P_A^\varepsilon(\pi)$, we have $\delta_\tau(x)\delta_\tau(y) = 0$ if A contains a pair of odd multiplicity or if A intersects an isolated block of (x, y). We set $[\![k]\!]_A = \cup_{t\in A}\pi_t$. If (G_{ij}) and Z are independent standard complex Gaussian variables,

$$\sum_{\tau\in P_A^\varepsilon(\pi)}\delta_\tau(x)\delta_\tau(y) \le \sum_{\tau\in P_A^\varepsilon}\delta_\tau(x)\delta_\tau(y) = \mathbf{E}\prod_{i\in[\![k]\!]_A}G_{x_iy_i}^{\varepsilon_i} = \prod_{(a,b)\in[\![n]\!]^2}\mathbf{E}[Z^{\dot m_{ab}(A)}Z^{\bar m_{ab}(A)}],$$

where $m_{ab}^\varepsilon(A)$ is the number of times that $(x_i, y_i, \varepsilon_i) = (a, b, \varepsilon)$ for $i \in [\![k]\!]_A$. If m_{ab} is the multiplicity of (a, b) in (x, y), we deduce that

$$\mathbf{E}\prod_{i\in[\![k]\!]_A}G_{x_iy_i}^{\varepsilon_i} \le \prod_{(a,b)\in[\![n]\!]^2}\mathbf{E}|Z|^{m_{ab}},$$

From Wick formula, for even m, $\mathbf{E}|Z|^m = (m/2)! \le m^{m/2}$ and $\mathbf{E}|Z|^2 = 1$. We deduce that

$$\mathbf{E}\prod_{i\in[\![k]\!]_A}G_{x_iy_i}^{\varepsilon_i} \le k^{m_4/2}.$$

Let T_0 be set of $t \in T$ such that π_t is isolated or contains a pair (x_i, y_i) of multiplicity one. We set $t_0 = |T_0|$. From (29), we find

$$\begin{aligned} I &\le k^{m_4/2}\sum_{A\subset[\![T]\!],A\cap T_0=\emptyset}\eta^{|A^c|} \\ &\le k^{m_4/2}\sum_{s=0}^{T-t_0}\binom{T-t_0}{s}\eta^{t_0+s} \\ &= k^{m_4/2}\eta^{t_0}(1+\eta)^{T-t_0}, \end{aligned}$$

where we have upper bounded all possibilities of sets A^c of size $t_0 + s$ in terms of its intersections with the set T_0 and its complement. We get,

$$I \leq k^{m_4/2}(1+\eta)^T \eta^{t_0}.$$

By assumption, $t_0 \geq \max(b, e_1/q) \geq (b + e_1/q)/2$. The statement of the corollary follows by using that, for $u, v \geq 0$, $(1+u)^v \leq e^{uv} \leq 1 + e^c uv$ if $uv \leq c$. □

Remark. Corollary 14 is tailored for our needs. This is not an optimal consequence of Theorem 12. Along the lines of the proof of Corollary 14, it is possible to obtain sharper bounds, for example by using that the odd moments of a Gaussian random variable vanish.

3.5 The orthogonal case

This paper focuses on random unitary matrices, however as we claim in the introduction, the obvious variant of our results works exactly the same way for sequences of orthogonal groups. In this subsection, we outline how to adapt the above statements on the unitary group to the orthogonal case

Firstly, there exists a counterpart of Proposition 8 which can be stated as follows:

Proposition 15. *Let k be even, $\pi = (\pi_t)_{t \in [\![T]\!]}$ be a partition of $[\![k]\!]$, and x, y in $[\![n]\!]^k$. There exists a generalized Weingarten function $\mathrm{Wg}_O[\pi](p,q,n)$ such that*

$$\mathbf{E}\left(\prod_{t=1}^{T} [\prod_{i \in \pi_t} O_{x_i y_i}^{\varepsilon_i}] \right) = \sum_{p,q \in P_k} \delta_p(x) \delta_q(y) \mathrm{Wg}_O[\pi](p,q,n).$$

Intriguingly, this formula seems to be simpler. Indeed, there is no need to consider conjugates and therefore specify matchings or balancing conditions since the entries O_{ij} of a random orthogonal matrix are all real. Its proof is based on the existence of the orthogonal Weingarten function and then reproduces the argument of Proposition 8.

The subsequent estimate is the orthogonal counterpart of Theorem 10.

Theorem 16. *Let k even with $4k^{7/2} \leq n^2$, $\pi = (\pi_t)_{t \in [\![T]\!]}$ be a partition of $[\![k]\!]$. For all $p, q \in P_k$, the following estimate holds true:*

$$|\mathrm{Wg}_O[\pi](p,q,n)| \leq (1 + 12k^{7/2}n^{-1}) n^{-k/2 - |pq^{-1}|} 4^{|pq^{-1}|} (2k^{7/4}n^{-1})^r,$$

where r is the number of blocks of π to which $p \vee q$ can be restricted.

Without entering into details, the proof is essentially the same than the proof of Theorem 10. The main input is an analog of Theorem 5 and Proposition 6 given in [8] by Theorem 4.6 and Theorem 4.9. We note that in [8] these two results are slightly more difficult to prove than their unitary counterpart. This yields an orthogonal version of Theorem 12 and Corollary 14.

4 Strong asymptotic freeness through non-backtracking operators

The aim of this section is to extend [4, Section 3] to general bounded operators in Hilbert spaces.

4.1 Spectral mapping formulas

We consider $(b_1, \dots, b_\ell)$ elements in $\mathcal{B}(\mathcal{H})$ where $\mathcal{H}$ is a Hilbert space. We assume that the set $[\![\ell]\!]$ is endowed with an involution $i \mapsto i^*$ (and $i^{**} = i$ for all i).

The non-backtracking operator associated to the ℓ-tuple of matrices $(b_1, \dots, b_\ell)$ is the operator on $\mathcal{B}(\mathcal{H} \otimes \mathbb{C}^\ell)$ defined by

$$B = \sum_{j \neq i^*} b_j \otimes E_{ij}, \tag{30}$$

where $E_{ij} \in M_d(\mathbb{R})$ are the canonical matrix elements. We also define the left non-backtracking operator as

$$\widetilde{B} = \sum_{j \neq i^*} b_i \otimes E_{ij}. \tag{31}$$

Note that if the b_i's are invertible, then B and $\widetilde{B}$ are conjugate

$$\widetilde{B} = DBD^{-1}$$

with $D = \sum_i b_i \otimes E_{ii}$. The non-backtracking operators are used in this paper to give an alternative description of the spectrum of an operator of the form (3). We start by a result in the reverse direction. In the sequel for shorter notation, the identity operator in $\mathcal{H}$ is denoted by 1 and if a is in $\mathcal{B}(\mathcal{H})$ and $\lambda \in \mathbb{C}$, we write $a - \lambda$ in place of $a - \lambda 1$.

Proposition 17. *Let B be as above and let $\lambda \in \mathbb{C}$ satisfy $\lambda^2 \notin \{\sigma(b_i b_{i^*}) : i \in [\![\ell]\!]\}$. Define the operator A_λ on $\mathcal{H}$ through*

$$A_\lambda = b_0(\lambda) + \sum_{i=1}^{\ell} b_i(\lambda)\,, \qquad b_i(\lambda) = \lambda b_i(\lambda^2 - b_{i^*} b_i)^{-1} \quad \text{and} \quad b_0(\lambda) = -1 - \sum_{i=1}^{\ell} b_i(\lambda^2 - b_{i^*} b_i)^{-1} b_{i^*}. \tag{32}$$

Then $\lambda \in \sigma(B)$ if and only if $0 \in \sigma(A_\lambda)$. Similarly, $\lambda \in \sigma(\widetilde{B})$ if and only if $0 \in \sigma(A_\lambda)$.

Proof. Let us assume that λ is in the point spectrum of B, i.e. there is a non-zero vector $v \in \mathcal{H} \otimes \mathbb{C}^\ell$ such that $Bv = \lambda v$. Let $E_i, i \in [\![\ell]\!]$ be the canonical basis of $\mathbb{C}^\ell$. We can write $v = \sum_i v_i \otimes E_i$. We define a vector $u \in \mathcal{H}$ by

$$u = \sum_j b_j v_j. \tag{33}$$

Let us show that u is a non-zero vector $u \in \mathcal{H}$ such that $A_\lambda u = 0$. Component-wise, the equation $Bv = \lambda v$ can be written

$$\lambda v_i = \sum_{j \neq i^*} b_j v_j = u - b_{i^*} v_{i^*}. \tag{34}$$

In the above equation, if we replace i by i^* we get

$$\lambda v_{i^*} = u - b_i v_i. \tag{35}$$

Multiplying Equation (34) by λ and substituting the last term of its right-hand side with Equation (35) multipled by b_{i^*}, we get

$$\lambda^2 v_i = \lambda u - b_{i^*} u + b_{i^*} b_i v_i.$$

This can be rewritten as

$$(\lambda^2 - b_{i^*} b_i) v_i = (\lambda - b_{i^*}) u.$$

Since $\lambda^2 - a_{i^*} a_i$ is invertible, we get, for all $i \in [\![\ell]\!]$,

$$v_i = (\lambda^2 - b_{i^*} b_i)^{-1} (\lambda - b_{i^*}) u. \tag{36}$$

If u was the zero vector, then all v_i would be zero, therefore v would be zero, which contradicts that v is an eigenvector. This proves that u is not zero.

Let us now prove that $A_\lambda u = 0$, i.e. its kernel is non trivial. According to the definition of A_λ, we need to prove that

$$u = \sum_i \big(-b_i (\lambda^2 - b_{i^*} b_i)^{-1} b_{i^*} + \lambda b_i (\lambda^2 - b_{i^*} b_i)^{-1} \big) u.$$

The right-hand side is equal to

$$\sum_i b_i (\lambda^2 - b_{i^*} b_i)^{-1} (\lambda - b_{i^*}) u.$$

Substituting with the help of Equation (36) for each i, this is equal to $\sum_i b_i v_i$, which completes the claim from the definition of u in (33).

Conversely, if 0 is in the point spectrum of A_λ with eigenvector u, we define v_i with Equation (36). However, the above computations imply that $Bv = \lambda v$ and $u = \sum_j b_j v_j$ as per the original definition of u, therefore u would be zero, which contradicts the assumption. We thus have proved so far that 0 is in the discrete spectrum of A_λ if and only if λ is in the point spectrum of B.

The same equivalence holds for $\widetilde{B}$. Indeed, if $\widetilde{B} v = \lambda v$ then we define $u = \sum_i v_i$. Repeating the above computation, we find $v_i = b_i (\lambda^2 - b_i b_{i^*})^{-1} (\lambda - b_i) u$ and $A_\lambda u = 0$. Conversely, if $A_\lambda u = 0$, then we define $v = \sum_i v_i \otimes E_i$ with $v_i = a_i (\lambda^2 - a_i a_{i^*})^{-1} (\lambda - a_i) u$. We get $Bv = \lambda v$. Arguing as above, the equivalence follows.

Next we handle the continuous spectrum. If λ is in the continuous spectrum of B, then there exists a sequence of unit vectors $v^{(n)} \in \mathcal{H} \otimes \mathbb{C}^\ell$ such that $\|\lambda v^{(n)} - Bv^{(n)}\|_2 \to 0$ (see for example Kowalski [20, p19]) . To each $v^{(n)}$ we associate a vector $u^{(n)} \in \mathcal{H}$ thanks to Equation (33). In the first part of the proof, all equations are continuous. Specifically, Equations (34)-(36) remain correct if the right-hand side is perturbed additively by a vector ε_n whose norm goes to 0 as $n \to \infty$. This implies that $\|A_\lambda u^{(n)}\|_2 \to 0$, i.e. 0 is in the spectrum of A_λ.

Conversely, if 0 is in the continuous spectrum of A_λ, let $u^{(n)} \in \mathcal{H} \otimes \mathbb{C}^\ell$ such that $\|A_\lambda u^{(n)}\|_2 \to 0$ as $n \to \infty$. We define $v^{(n)}$ through equation (36). As above, it is possible to prove that $\|\lambda v^{(n)} - Bv^{(n)}\|_2 \to 0$ as $n \to \infty$, so λ is in the spectrum of B,

Finally, we handle the residual spectrum. Let $v \in \mathcal{H} \otimes \mathbb{C}^\ell$ non-zero that is not an element in the closure of the image of $B - \lambda 1$, and without loss of generality let us assume that v is orthogonal to the image of $B - \lambda 1$. Then, for all $x \in \mathcal{H} \otimes \mathbb{C}^d$, $\langle Bx - \lambda x, v\rangle = 0$. This implies that $B^* v = \overline{\lambda} v$, that is $\overline{\lambda}$ is in the point spectrum of B^*. Now, observe that

$$B^* = \sum_{i \neq j^*} b_j^* \otimes E_{ji} = \sum_{i \neq j^*} b_i^* \otimes E_{ij} = \sum_{i \neq j^*} \tilde{b}_i \otimes E_{ij},$$

with $\tilde{b}_i = b_i^*$. In particular, B^* is a left non-backtracking operator as defined in (31) with underlying operators the $\tilde{b}_i$'s. From what precedes, $\overline{\lambda}$ is in the point sprectrum of B^* if and only if 0 is in the point spectrum of $\widetilde{A}_\lambda = \tilde{b}_0 + \sum_{i=1}^d \tilde{b}_i$ defined by, for $i \in [\![\ell]\!]$,

$$\tilde{b}_i = \overline{\lambda} b_i^* (\overline{\lambda}^2 - b_{i^*}^* b_i^*)^{-1} \quad \text{and} \quad \tilde{b}_0 = -1 - \sum_{i=1}^d b_i^* (\overline{\lambda}^2 - b_{i^*}^* b_i^*)^{-1} b_{i^*}^*.$$

It straightforward to check that $\tilde{b}_i = b_i(\lambda)^*$ for each $i \in [\![\ell]\!]$. Thus, we have $\widetilde{A}_\lambda = A_\lambda^*$. From what precedes, we get that $\overline{\lambda}$ is in the discrete sprectrum of B^* if and only if 0 is in the discrete spectrum of A_λ^*. Hence, we have proved that if λ is in the residual spectrum of B then 0 is in the spectrum of A_λ (since for any bounded operator T on a Hilbert space, $\sigma(T^*)$ is the complex conjugate of the set $\sigma(T)$, see Reed-Simon [28, Theorem VI.7]).

Conversely, if 0 is in the residual spectrum of A_λ, then 0 is in the discrete spectrum of A_λ^*. From what precedes, we get that $\overline{\lambda}$ is in discrete spectrum of B^*. In particular, λ is in the spectrum of B. The proposition is proved. □

We now consider the situation where the Hilbert space $\mathcal{H}$ is of the form $\mathbb{C}^r \otimes \mathcal{K}$, where $\mathcal{K}$ is a Hilbert space and

$$b_i = a_i \otimes V_i$$

with $a_i \in M_r(\mathbb{C})$ and V_i unitary operator on $\mathcal{K}$ such that for all $i \in [\![\ell]\!]$,

$$V_{i^*} = V_i^*.$$

Moreover, we assume that $\ell = 2d$ and the involution i^* is an Section 2. In this specific case, we have

$$B = \sum_{i \neq j^*} a_j \otimes V_j \otimes E_{ij} \quad \text{and} \quad \widetilde{B} = \sum_{i \neq j^*} a_i \otimes V_i \otimes E_{ij}. \tag{37}$$

The operator A_λ in (32) is given by

$$A_\lambda = a_0(\lambda) \otimes 1 + \sum_{i=1}^{2d} a_i(\lambda) \otimes V_i \tag{38}$$

with

$$a_i(\lambda) = \lambda a_i(\lambda^2 - a_{i^*} b_i)^{-1} \quad \text{and} \quad a_0(\lambda) = -1 - \sum_{i=1}^{2d} a_i(\lambda^2 - a_{i^*} a_i)^{-1} a_{i^*}.$$

If $\mathcal{K} = \ell^2(X)$ with X countable and V_i a permutation operator, then Proposition 17 recovers Proposition 9 of [4] up to the minor point that B is slightly modified into $B = \sum_{j \neq i^*} a_j \otimes S_i \otimes E_{ij}$, but it is easy to check that both B are conjugate to each other so they have the same spectrum. It follows that all results of [4, Section 3] can be extended to this more general setting.

We now review two keys results [4, Section 3] that will be used in the sequel. We start with a kind of converse of Proposition 17, in the sense that for a given self-adjoint operator A of the form

$$A = a_0 \otimes 1 + \sum_{i=1}^{2d} a_i \otimes V_i, \tag{39}$$

satisfying the symmetry condition (7) and a real number μ, we look for a non-backtracking operator B_μ which detects if $\mu \in \sigma(A)$. To perform this, we need to introduce the resolvent of the operator $A_\star$ defined in (4). For $\mu \notin \sigma(A_\star)$, we set

$$G(\mu) = (\mu - A_\star)^{-1}.$$

For $x, y \in \mathbb{F}_d$, we define $G_{xy}(\mu) \in M_r(\mathbb{C})$ as the matrix $P_x G(\mu) P_y^*$ where P_x is the projection onto the vector space $\mathbb{C}^r \otimes \delta_x$. We denote by o the neutral element of $\mathbb{F}_d$ for its group structure and g_i the i-th generator of $\mathbb{F}_d$. Finally, if D is a bounded set in $\mathbb{C}$, the convex hull of D is denoted by $\mathrm{hull}(D)$.

Proposition 18. *Let A be as in* (39) *satisfying* (7) *and $\mu \notin \mathrm{hull}(\sigma(A_\star))$. Define the matrices for $i \in [\![2d]\!]$,*

$$\hat{a}_i(\mu) = G_{oo}(\mu)^{-1} G_{og_i}(\mu).$$

Let $B_\mu = \sum_{i \neq j^*} \hat{a}_j(\mu) \otimes V_j \otimes E_{ij}$ be the corresponding non-backtracking operator. Then $\mu \notin \sigma(A)$ if and only if $1 \notin \sigma(B_\mu)$. The same statement holds for the corresponding left non-backtracking operator $\widetilde{B}_\mu$.

This statement is [4, Proposition 10] extended for a general Hilbert space and general unitaries V_i's. The same proof applies in this case thanks to Proposition 17.

4.2 Spectral radius of non-backtracking operators

We conclude this section with a sharp criterion to guarantee in terms of non-backtracking operators that the spectrum of an operator A is in a neighborhood of the spectrum of the operator $A_\star$. The following result is [4, Theorem 12]. Again, thanks to the improvement of Proposition 17, we can now state it a more general context.

Theorem 19. *Let A be as in* (39) *satisfying* (7) *and $A_\star$ the corresponding free operator defined by* (4)*. The following two results hold true:*

(i) For any $\mu \notin \mathrm{hull}(\sigma(A_\star))$, we have $\rho((B_\star)_\mu) < 1$, where $(B_\star)_\mu = \sum_{i \neq j^*} \hat{a}_j(\mu) \otimes \lambda(g_j) \otimes E_{ij}$ is the non-backtracking operator on $\mathbb{F}_d$ with weights as in Proposition 18.

(ii) For any $\varepsilon > 0$, there exists $\delta > 0$ such that if for all real μ at distance at least ε from $\mathrm{hull}(\sigma(A_\star))$*,*

$$\rho(B_\mu) < \rho((B_\star)_\mu) + \delta,$$

then $\mathrm{hull}(\sigma(A))$ *is in an ε-neighbourhood of* $\mathrm{hull}(\sigma(A_\star))$*.*

Moreover, the same holds for the left non-backtracking operators.

The theorem is a simple consequence of the following two claims:

$$\rho(B_\star) = \sup\{\lambda \geq 0 : \lambda \in \sigma(B_\star)\},$$

and

$$\text{the map from } M_r(\mathbb{C})^{2d} \text{ to } \mathbb{R}\text{: } (b_1, \ldots, b_{2d}) \mapsto \rho(B_\star) \text{ is continuous.} \tag{40}$$

We refer to [4] for details.

5 Expected high trace of non-backtracking matrices

In this section, we prove Theorem 2 by computing the spectral radius of the non-backtracking matrices associated to the random matrix A defined in (3).

5.1 Main technical statement

We now come back to the setting of Theorem 2. Let $U_1, \cdots, U_d$ be independent Haar distributed random unitary matrices in $\mathbb{U}_n$. We define V_i as in (1) and its centered version $[V_i]$ as in (2).

The left non-backtracking operator of associated to the weights $(a_1 \otimes [V_1], \ldots, a_{2d} \otimes [V_{2d}])$ is

$$B = \sum_{i \neq j^*} a_i \otimes [V_i] \otimes E_{ij}. \tag{41}$$

Note that we have omitted the tilde in (31) for shorter notation. The main technical result of this section is an upper bound on the spectral radius of B in terms on the spectral radius of the corresponding operator on the free group $\mathbb{F}_d$:

$$B_\star = \sum_{i \neq j^*} a_i \otimes \lambda(g_i) \otimes E_{ij}, \tag{42}$$

where $\lambda(g)$ is the left-regular representation of $g \in \mathbb{F}_d$. Let $\rho(B_\star)$ be the spectral radius of the $B_\star$. Recall that $\rho(B) \leq \|B^\ell\|^{1/\ell}$ for all integer $\ell \geq 1$ and the sequence $\|B^\ell\|^{1/\ell}$ is non-increasing in ℓ.

Theorem 20. *There exists a universal constant $c > 0$ such that if $q \leq c \ln n / \ln\ln n$ then the following holds. For any $\varepsilon > 0$, there exists a constant $C \geq 1$ (depending on ε, d and r) such that for any $(a_1, \dots, a_{2d}) \in M_r(\mathbb{C})$ with $\max_i \|a_i\| \leq 1$, if $\ell = \lfloor Cq \rfloor$, the event*

$$\|B^\ell\|^{1/\ell} \leq \rho(B_\star) + \varepsilon$$

holds with probability at least $1 - C\exp\left(-(\ln n)^2\right)$.

In the next subsection, we prove Theorem 20. In the remaining subsection, we prove Theorem 2 from Theorem 20.

5.2 Proof of Theorem 20

In the sequel, an entry of the matrix $[V_i]$ will be denoted by

$$[V_i]_{xy} = \prod_{p=1}^{q} [U_i^{\varepsilon_q}]_{x_p y_p},$$

with $x = (x_1, \dots, x_q) \in [\![n]\!]^q$ and $y = (y_1, \dots, y_q) \in [\![n]\!]^q$. We also set

$$\vec{E} = [\![n]\!]^q \times [\![2d]\!].$$

In analogy with usual non-bactracking matrices, an element $e = (x, i)$ of $\vec{E}$ is thought as the directed edge attached to x associated to the i-th unitary matrix V_i. If $e = (x, i), f = (y, j) \in \vec{E}$, the matrix-valued entry (e, f) of B defined in (41) is

$$B_{ef} = a_i [V_i]_{xy} \mathbf{1}_{i \neq j^*} \in M_r(\mathbb{C}). \tag{43}$$

We start with Weyl formula for the spectral radius. If o is the unit of F_d, we set

$$\rho_k = \left(\max \|B_\star^k \varphi \otimes \delta_e\| \right)^{\frac{1}{k}},$$

where the maximum is over all $e = (o, i), i \in [\![2d]\!]$ and $\varphi \in \mathbb{C}^r$ of unit norm. From Weyl formula, ρ_k converges to $\rho(B_\star)$ as k goes to infinity (see for example [24, Theorem 1.3.6]). In the remainder of the proof, we fix $\varepsilon > 0$, then there exists an integer k_0 large enough such that for all $k \geq k_0$,

$$\rho_k \leq \rho = \rho(B_\star) + \varepsilon. \tag{44}$$

Let ℓ be an integer. We fix an integer θ. In order to get a good probabilistic estimate, we will upper bound the expectation of $\left\|B^\ell\right\|^{2\theta}$ for θ large enough. We write

$$\left\|B^\ell\right\|^{2\theta} = \left\|B^\ell (B^\ell)^*\right\|^\theta = \left\|\left(B^\ell (B^\ell)^*\right)^\theta\right\|.$$

In particular, if $\mathrm{tr}_{\mathbb{C}^r}$ denotes the reduced trace on $\mathbb{C}^r$ for operators on $\mathbb{C}^r \otimes \mathbb{C}^{\vec{E}}$, we get

$$\left\|B^\ell\right\|^{2\theta} \leq \mathrm{tr}_{\mathbb{C}^r}\left\{\left(B^\ell (B^\ell)^*\right)^\theta\right\}.$$

We expand the trace in terms of the matrix-valued entries of B:

$$\mathrm{tr}_{\mathbb{C}^r}\left\{\left(B^\ell (B^\ell)^*\right)^\theta\right\} \;=\; \sum_{e_\alpha \in \vec{E}, \alpha \in [\![2\theta]\!]} \prod_{\alpha=1}^{\theta} (B^\ell)_{e_{2\alpha-1} e_{2\alpha}} ((B^\ell)^*)_{e_{2\alpha} e_{2\alpha+1}},$$

with $e_{2\theta+1} = e_1$. We expand further and use that $(B^*)_{ef} = (B_{fe})^*$, we obtain

$$\mathrm{tr}_{\mathbb{C}^r}\left\{\left(B^\ell (B^\ell)^*\right)^\theta\right\} \;=\; \sum_{\gamma} \prod_{\alpha=1}^{2\theta} \prod_{t=1}^{\ell} B^{\varepsilon_\alpha}_{\gamma_t^\alpha \gamma_{t+1}^\alpha},$$

with $B^{\varepsilon_\alpha}_{ef}$ is equal to B_{ef} for odd α and $(B_{fe})^*$ for even α and the sum is over all $\gamma = (\gamma^1, \ldots, \gamma^{2\theta})$, $\gamma^\alpha = (\gamma_1^\alpha, \cdots, \gamma_{\ell+1}^\alpha)$ in $\vec{E}^{\ell+1}$ with the boundary conditions, for all $\alpha \in [\![2\theta]\!]$,

$$\gamma_1^{2\alpha} = \gamma_1^{2\alpha+1} \quad \text{and} \quad \gamma_{\ell+1}^{2\alpha} = \gamma_{\ell+1}^{2\alpha-1}, \tag{45}$$

with the convention $\gamma^{2\theta+1} = \gamma^1$, see Figure 1.

Let us write $\gamma_t^\alpha = (x_t^\alpha, i_t^\alpha)$. From (43), we find

$$\mathrm{tr}_{\mathbb{C}^r}\left\{B^\ell (B^\ell)^*\right\} = \sum_{\gamma \in P_{\ell,\theta}} \prod_{\alpha}^{2\theta} \prod_{t=1}^{\ell} a_{i_t^\alpha}^{\varepsilon_\alpha} [V_{i_t^\alpha}]^{\varepsilon_\alpha}_{x_t^\alpha x_{t+1}^\alpha},$$

where $(a_i^{\varepsilon_\alpha}, [V_i]^{\varepsilon_\alpha}_{xy})$ is equal to $(a_i, [V_i]_{xy})$ or $(a_i^*, [\bar{V}_i]_{xy})$ depending on the parity of α and $P_{\ell,\theta}$ is the set of all $\gamma = (\gamma^1, \ldots, \gamma^{2\theta})$ as above which are also non-backtracking, that is, for all $t \in [\![\ell]\!]$ and $\alpha \in [\![2\theta]\!]$,

$$i_{t+1}^\alpha \neq i_t^{\alpha *}. \tag{46}$$

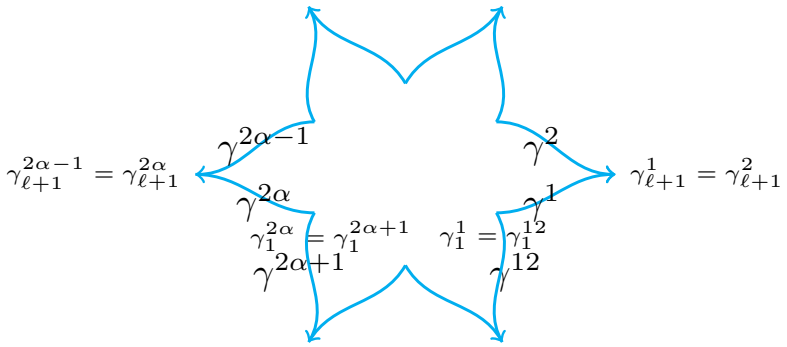

Figure 1: A path $\gamma = (\gamma^1, \dots, \gamma^{12})$ in $P_{\ell,6}$.

(Note that $[\bar{V}_i]_{xy} = [V_{i^*}]_{yx}$ by construction). We finally take the expectation and find:

$$\mathbf{E}\left\|B^\ell\right\|^{2\theta} \le r \sum_{\gamma \in P_{\ell,\theta}} a(\gamma)p(\gamma), \tag{47}$$

where we have introduced the algebraic and probabilist weights of an element $\gamma \in P_\ell$ defined as

$$a(\gamma) = \prod_{\alpha=1}^{2\theta} \left\| \prod_{t=1}^{\ell} a_{i_t^\alpha}^{\varepsilon_\alpha} \right\| \quad \text{and} \quad p(\gamma) = \left| \mathbf{E}\left\{ \prod_{\alpha=1}^{2\theta} \prod_{t=1}^{\ell} [V_{i_t^\alpha}]_{x_t^\alpha x_{t+1}^\alpha}^{\varepsilon_\alpha} \right\} \right|.$$

We organize the terms on the right-hand side of (47) by introducing the following equivalence class on elements of $P_{\ell,\theta}$. We say that a permutation σ on $\vec{E}$ is a *tensor permutation* if there exists a permutation τ on $[\![n]\!]$ and a family of permutations $(\beta_x)_{x\in[\![n]\!]^q}$ on $[\![2d]\!]$ such that for all $(x,i) \in [\![n]\!]^q \times [\![2d]\!]$, $\beta_x(i)^* = \beta_x(i^*)$ and $\sigma(x,i) = (\tau(x), \beta_x(i))$ with $\tau(x_1, \dots, x_q) = (\tau(x_1), \dots, \tau(x_q))$. We write $\gamma \sim \gamma'$ if there exists a tensor permutation σ such that $\gamma' = \sigma(\gamma) = (\sigma(\gamma_t^\alpha))_{\alpha\in[\![2\theta]\!], t\in[\![\ell+1]\!]}$. Note that if $\gamma \sim \gamma'$, then $p(\gamma) = p(\gamma')$. In more combinatorial language, an equivalence class of $P_{\ell,\theta}$ is an unlabeled path where the labels are the vertex entries in $[\![n]\!]$ and the edge colors in $[\![2d]\!]$. We will use colored graphs defined formally as follows.

Definition 21 (Colored edge and graph)**.** *Let X be a countable set and C a countable set with an involution $* : C \to C$. A* colored edge *$[x,i,y]$ is an equivalence class of $X \times C \times X$ equipped with the equivalence relation $(x,i,y) \sim (y,i^*,x)$. A* colored graph *$G = (V,E)$ is the pair formed by a vertex set $V \subset X$ and a set of colored egdes $[x,i,y]$ with $x,y \in V$. The* degree *of $v \in V$ is $\sum_{e=[x,i,y]\in E}(\mathbf{1}(v=x) + \mathbf{1}(v=y))$.*

We now define a colored graph which is naturally associated to an element $\gamma \in P_{\ell,\theta}$. We write $x_t^\alpha = (x_{t,1}^\alpha, \dots, x_{t,q}^\alpha) \in [\![n]\!]^q$ and define the colored graph $G_\gamma = (V_\gamma, E_\gamma)$ with color set $[\![2d]\!]$, $V_\gamma =$

$\{x^\alpha_{t,p} : \alpha \in [\![2\theta]\!], t \in [\![\ell+1]\!], p \in [\![q]\!]\} \subset [n]$ and $E_\gamma = \{[x^\alpha_{t,p}, i^\alpha_t, x^\alpha_{t+1,p}] : \alpha \in [\![2\theta]\!], t \in [\![\ell]\!], p \in [\![q]\!]\}$. The *multiplicity* of an edge $e \in E_\gamma$ is defined as $m(e) = \sum_{t\in[\![\ell]\!],\alpha\in[\![2\theta]\!],p\in[\![q]\!]} \mathbf{1}([x^\alpha_{t,p}, i^\alpha_t, x^\alpha_{t+1,p}] = e)$. If $e = |E_\gamma|$, e_1 is the number of edges of multiplicity one and $e_{\geq 2}$ is the number of edges of multiplicity at least two, we have

$$e = e_1 + e_{\geq 2} \quad \text{and} \quad e_1 + 2e_{\geq 2} \leq 2q\ell\theta.$$

We deduce that

$$e \leq q\ell\theta + e_1/2. \tag{48}$$

Moreover, from the boundary condition (45), the graph G_γ has at most q connected components. If $v = |V_\gamma|$, it follows that

$$e - v + q \geq 0. \tag{49}$$

In particular, combining the last two inequalities, for any $\gamma \in P_{\ell,\theta}$,

$$\chi = q(\ell\theta + 1) + e_1/2 - v \geq 0. \tag{50}$$

For integers v, e_1 such that (50) holds, we denote by $P_{\ell,\theta}(v, e_1)$ the set of elements in $P_{\ell,\theta}$ with v vertices and e_1 edges of multiplicity one. Note that if $\gamma \sim \gamma'$, the number of vertices and the number of edges with a given multiplicity are equal. It follows that we may define $\mathcal{P}_{\ell,\theta}(v, e_1)$ as the set of equivalence classes with v vertices and e_1 edges of multiplicity one. Our first lemma is a rough bound on $\mathcal{P}_{\ell,\theta}(v, e_1)$.

Lemma 22. *If $\chi = q(\ell\theta + 1) + e_1/2 - v$, we have,*

$$|\mathcal{P}_{\ell,\theta}(v, e_1)| \leq v^{q-1}(2q\ell\theta)^{6\chi}.$$

Proof. We order the set $T = [\![q]\!] \times [\![2\theta]\!] \times [\![\ell+1]\!]$ with the lexicographic order on the first two coordinates and for the last coordinate: $(p, \alpha, t) \succ (p, \alpha, t')$ if α is odd and $t > t'$ or if α is even and $t < t'$. In words, the natural order is reversed for the last coordinate depending on the parity of α.

We think of an element $s = (p, \alpha, t) \in T$ as a time. We also define the set $\vec{E}_1 = [\![n]\!] \times [\![2d]\!]$ ordered with the lexicographic order. We associate in an injective way to an element $\gamma \in P_{\ell,\theta}$ the sequence $(\gamma_s)_{s\in T} \in \vec{E}_1^T$ with $\gamma_s = (y_s, i_s) \in \vec{E}_1$. For concreteness, we may define a canonical element in each equivalence class as follows. We say that $\gamma \in P_{\ell,\theta}$ is *canonical* if $(\gamma_s)_{s\in T}$ is minimal for the lexicographic order in its equivalence class. For example, if $\gamma \in P_{\ell,\theta}(v, e_1)$ is canonical then $\gamma_{(1,1,1)} = (1, 1)$, $V_\gamma = [\![v]\!]$ and the vertices appear for the first time in the sequence (γ_s) in order. Our goal is to give an upper bound on the number of canonical elements in $P_{\ell,\theta}(v, e_1)$.

We define $T_0 \subset T$ as the set of (p, α, t) such that if α is odd $t \in [\![\ell]\!]$ and if α is even $t - 1 \in [\![\ell]\!]$. For $s \in T$, we write $s + 1$ for the successor of T in the total order and $t - 1$ for its predecessor,

with the convention that $(1,1,1)-1=0$ and $(q,2\theta,1)+1=\infty$. The set T_0 is the subset of s in T such that s and $s+1$ have the same first two coordinates. We explore iteratively the sequence (i_s, y_{s+1}), $s \in T_0$, and we also build a growing sequence of forests (F_s) (graphs without cycles). We define F_0 as the graph with no edge and vertex set $\{1\}$. At time $s \in T_0$, if the addition to F_{s-1} of the edge $[y_s, i_s, y_{s+1}]$ does not create a cycle, then F_s is the forest spanned by F_s and $[y_s, i_s, y_{s+1}]$, we then say that $[y_s, i_s, y_{s+1}]$ is a *tree edge*. Otherwise $F_s = F_{s-1}$. Let us say that a time $s \in T_0$ is a *first time* if the vertex y_{s+1} has not been seen before. Then, $[y_s, i_s, y_{s+1}]$ is a tree edge which is said to be associated to y_{s+1}. We call the vertices $y_{(p,1,1)} \in V_\gamma$ with $p \in [\![q]\!]$, the *seeds*. Due to the boundary condition (45), each vertex $y \in V_\gamma$ different from a seed has an associated tree edge. Hence, the number of tree edges associated to a vertex, say f, satisfies

$$f \geq v - q.$$

If s is a first time then the value of (i_s, y_{s+1}) is determined by the preceding values $(i_{s'}, y_{s'+1})$, $s' < s$ and the value of the seeds: we have $i_s = 1$ and $y_{s+1} = k+1$ where k is the number of distinct vertices which had been seen so far. We say that a time $s \in T_0$ is a *tree time* if $[y_s, i_s, y_{s+1}]$ is a tree edge which has already been visited. Finally, the other times are called *important times*. Due to the non-backtracking constraint (46), the sequence γ can be decomposed as the successive repetitions of: (i) a sequence of first times (possibly empty), (ii) an important time or a time in $T \backslash T_0$, (iii) a sequence of tree times (possibly empty).

We note that all edges are visited at least twice except e_1 of them. We deduce that the number of important times is at most

$$2q\ell\theta - 2f + e_1 \leq 2q(\ell\theta+1) - 2v + e_1 = 2\chi.$$

We mark the important times by the vector $(i_s, y_{s+1}, y_\tau, i_\tau)$ where $\tau > s$ is the next time which is not a tree time. Observe that there is a unique non-backtracking path between two vertices of a tree. It follows that the sequence γ is uniquely determined by the value $y_{(p,1,1)}$ of the seeds, the positions of the important times and their marks.

There are at most $2q\ell\theta$ possibilities for the position of an important time and $(2q\ell\theta)^2$ possibilities for its mark (there are at most $2q\ell\theta$ edges in G_γ). In particular, the number of distinct canonical paths in $P_{\ell,\theta}(v, e_1)$ is at most

$$v^{q-1}(2q\ell\theta)^{6\chi},$$

where v^{q-1} accounts for the possible values of the seeds (recall that $y_{(1,1,1)} = 1$). □

We now estimate the probabilistic weight $p(\gamma)$. To this end, we introduce a finer partition of the set $P_{\ell,\theta}$. Let $\gamma \in P_{\ell,\theta}$ with $\gamma_t^\alpha = (x_t^\alpha, i_t^\alpha), t \in [\![\ell+1]\!]$. We define a *bracket* as a colored edge

$[x,i,y]$ on $[\![n]\!]^q$ with color set $[\![2d]\!]$ (in the sense of Definition 21). We associate to γ the sequence of brackets $([x_t^\alpha, i_t^\alpha, x_{t+1}^\alpha])$, $t \in [\![\ell]\!], \alpha \in [\![2\theta]\!]$. We say that a bracket $[x_t^\alpha, i_t^\alpha, x_{t+1}^\alpha]$ of γ is *isolated* if for all $p \in [\![q]\!]$, the colored edge of G_γ, $e = [x_{t,p}^\alpha, i_t^\alpha, x_{t+1,p}^\alpha]$ is not visited at a different time: that is for all $(t', \alpha', p') \in [\![\ell]\!] \times [\![2\theta]\!] \times [\![q]\!]$ with $(t', \alpha') \neq (t, \alpha)$, we have $e \neq [x_{t',p'}^{\alpha'}, i_{t'}^{\alpha'}, x_{t'+1,p'}^{\alpha'}]$. For integer b, we denote by $P_{\ell,\theta}(v, e_1, b)$ the set of $\gamma \in P_{\ell,\theta}(v, e_1)$ with b isolated brackets.

Lemma 23. *Let $\gamma \in P_{\ell,\theta}(v, e_1, b)$ and $\chi = q(\ell\theta+1) + e_1/2 - v$. If $v > q(\ell\theta+1)$ then $p(\gamma) = 0$. Otherwise, there exists a universal constant $c > 0$, such that*

$$p(\gamma) \leq cn^{-q\ell\theta} \eta^{b+\frac{e_1}{q})} (cq\ell\theta)^{2\chi},$$

with $\eta = (cq\ell\theta)^{q/2} n^{-1/8}$.

Proof. Let T, T_0 endowed with a total order be as in the proof of Lemma 22. By Theorem 12, $p(\gamma) = 0$ unless the sequence $(x_{t,p}^\alpha, x_{t+1,p}^\alpha), (p, \alpha, t) \in T_0$ is an even sequence. In particular, this last condition implies that $2(v-q) \leq 2q\ell\theta$, since for all vertices x of V_γ different from the seeds $(x_{(1,1,p)}), p \in [\![q]\!]$, we may associate at least two elements $(t, \alpha, p) \in T_0$ such that $x = x_{t+1,p}^\alpha$ (for α odd) or $x = x_{t,p}^\alpha$ (for α even). It gives the first claim.

We now prove the second claim. By Corollary 14, it suffices to prove that

$$m_{\geq 4} \leq 4\chi, \tag{51}$$

where $m_{\geq 4} = \sum_e \mathbf{1}(m_e \geq 4) m_e$ is the sum of multiplicities of edges of G_γ with multiplicity at least 4. Indeed, let e_{23} be the number of edges of multiplicity 2 or 3. We have

$$e_1 + 2e_{23} + m_{\geq 4} \leq 2q\ell \quad \text{and} \quad e_1 + e_{23} + \frac{m_{\geq 4}}{4} \geq e \geq v - q,$$

where $e = |E_\gamma|$ is the number of edges and where we have used (49). We cancel e_{23} and we find at

$$-e_1 + \frac{m_{\geq 4}}{2} \leq 2q\ell\theta - 2v + 2q.$$

We obtain (51). □

Our final lemma is the estimation of $\sum a(\gamma)$ where the sum is over an equivalence class. This is our main combinatorial ingredient.

Lemma 24. *Let $\gamma \in P_{\ell,\theta}(v, e_1, b)$, $\chi = q(\ell\theta+1) + e_1/2 - v$ and let ρ be as in (44). There exists a constant $C > 1$ (depending on k_0, d and $\max_i \|a_i\|$), such that*

$$A(\gamma) = \sum_{\gamma' : \gamma' \sim \gamma} a(\gamma') \leq n^v c^{b+q\theta+\chi} \rho^{2\ell\theta}.$$

We start with a preliminary lemma.

Lemma 25. *Let ρ be as in (44). There exists a constant $c > 1$ (depending on k_0, d and $\max_i \|a_i\|$) such that for any integer $k \geq 1$,*

$$\sum \left\| \prod_{t=1}^{k} a_{i_t} \right\|^2 \leq c^2 \rho^{2k} \quad \text{and} \quad \sum \left\| \prod_{t=1}^{k} a_{i_t} \right\| \leq c^k \rho^k,$$

where the sums are over all $(i_1, \cdots, i_k)$ such that $i_{t+1} \neq i_t^*$. In particular,

$$\max \left\| \prod_{t=1}^{k} a_{i_t} \right\| \leq c\rho^k$$

where the maximum is over all $(i_1, \cdots, i_k)$ such that $i_{t+1} \neq i_t^*$.

Proof. Recall the definition of ρ_k was defined above (44). As above Proposition 18, for $e, f \in \mathbb{F}_d \times [\![2d]\!]$, $(B_\star)_{ef} \in M_r(\mathbb{C})$ is the matrix $P_e B_\star P_f$ with P_f the orthogonal projection onto $\mathbb{C}^r \otimes \delta_f$. If $e = (o, i)$ and $f = (g, j)$ are in $\mathbb{F}_d \otimes [\![2d]\!]$, we have

$$(B_\star^k)_{ef} = \prod_{t=1}^{k} a_{i_t},$$

if $g = g_{i_1} \cdots g_{i_k}$ with $i_1 = i$, $i_{t+1} \neq i_t^*$ and $j \neq i_k^*$ (note that g is written in reduced form). Otherwise, we have $(B_\star^k)_{ef} = 0$. We deduce that

$$\rho_k^{2k} = (2d-1) \cdot \max_{i \in [\![2d]\!]} \sum \left\| \prod_{t=1}^{k} a_{i_t} \right\|^2 = (2d-1) \cdot \max_{i \in [\![2d]\!]} \sum \left\| \prod_{t=1}^{k} a_{i_t^*} \right\|^2,$$

where the sum is over all $(i_1, \cdots, i_k)$ such that $i_1 = i$ and $i_{t+1} \neq i_t^*$. Let $C = \max_i \|a_i\|$, we find from (44) that for all $k \geq k_0$,

$$\sum \left\| \prod_{t=1}^{k} a_{i_t} \right\|^2 \leq \frac{2d}{2d-1} \rho^{2k}.$$

where the sum is over all $(i_1, \cdots, i_k)$ such that and $i_{t+1} \neq i_t^*$. Up to a large multiplicative constant on the right-hand side, the above inequality is true for all $k \geq 1$. This is the first claimed statement. For the second statement, we apply Cauchy-Schwarz inequality and get:

$$\sum \left\| \prod_{t=1}^{k} a_{i_t} \right\| \leq \left(2d(2d-1)^{k-1} \right)^{\frac{1}{2}} \left(\sum \left\| \prod_{t=1}^{k} a_{i_t} \right\|^2 \right)^{\frac{1}{2}}.$$

Modifying the constant c, we obtain the second statement. The last statement is an immediate consequence of the first statement. □

We may now prove Lemma 24.

Proof of Lemma 24. We start by introducing a new decomposition a path $\gamma \in P_\gamma$. As usual, we write $\gamma \in P_{\ell,\theta}$ with $\gamma_t^\alpha = (x_t^\alpha, i_t^\alpha) \in [\![n]\!]^q \times [\![2d]\!]$, $x_t^\gamma = (x_{t,p}^\gamma)_{p\in[\![q]\!]}$. Let $T = [\![2\theta]\!] \times [\![\ell]\!]$ and let $T_{\geq 3}$ be the set of $(\alpha, t) \in T$ such that for some $p \in [\![q]\!]$, $x_{t,p}^\alpha$ has degree at least 3 in G_γ (recall that the degree of a vertex was defined in Definition 21). We set $T_\star = T_{\geq 3} \cup ([\![2\theta]\!] \times \{1, \ell+1\})$ and write $l = |T_\star|$, $T_\star = \{(\alpha, t_j^\alpha) : j \in [\![l_\alpha]\!]\}$ with $\sum_\alpha l_\alpha = l$, $(t_j^\alpha)_j$ increasing, $t_1^\alpha = 1$, $t_{l_\alpha}^\alpha = \ell + 1$. Note that the set $T_\star$ is a function of the equivalence class of γ. We write $a(\gamma)$ defined below (47) as

$$a(\gamma) = \prod_{\alpha=1}^{2\theta} \left\| \prod_{j=1}^{l_\alpha - 1} \prod_{t=t_j^\alpha}^{t_{j+1}^\alpha - 1} a_{i_t^\alpha}^{\varepsilon_\alpha} \right\|. \tag{52}$$

We have $l \leq 4\theta + |T_{\geq 3}|$. We start by an upper bound on $|T_{\geq 3}|$. Let v_k and $v_{\geq k}$ be the number of vertices of V_γ with degree k and at least k. If $m_{\geq 3}$ is the sum of multiplicities at least 3 as in (51), we have the inequality

$$|T_{\geq 3}| \leq 2(v_{\geq 3} + m_{\geq 3}), \tag{53}$$

which follows from the simple observation that each edge neighboring a vertex of degree at least 3 is visited at most twice unless it is of multiplicity at least 3. We now estimate $v_{\geq 3}$. We have

$$v_1 + v_2 + v_{\geq 3} = v \quad \text{and} \quad v_1 + 2v_2 + 3v_{\geq 3} \leq \sum_k k v_k = 2e,$$

where $e = |E_\gamma|$ is the number of edges. Also arguing as in (48), we find $m_{\geq 3} \leq 6\chi$. We deduce that

$$v_{\geq 3} \leq 2(e - v) + v_1 \leq 2(q\ell\theta + e_1/2 - v) + 2q\theta = 2\chi + 2q(\theta - 1), \tag{54}$$

where the bound $v_1 \leq 2q\theta$ is a consequence of the assumption (46): only the vertices $\gamma_{1,p}^{2\alpha} = \gamma_{1,p}^{2\alpha+1}$ and $\gamma_{\ell+1,p}^{2\alpha} = \gamma_{\ell+1,p}^{2\alpha+1}$ with $p \in [\![q]\!]$ can possibly be of degree 1 in G_γ. Therefore, using (51), we deduce from (53) that

$$|T_{\geq 3}| \leq 16\chi + 4q(\theta - 1).$$

We thus obtain the upper bound:

$$l \leq 16\chi + 4q\theta. \tag{55}$$

We now proceed to the bound of the factors in (52) when we sum over all $\gamma' \sim \gamma$. Consider the color set $\{-1, 1\}$ equipped with the trivial involution $c^* = c$. We first define a colored graph, say Γ, on the vertex set $T = [\![2\theta]\!] \times [\![\ell]\!]$ on the color set C, by placing the edge $[(\alpha, t), 1, (\alpha', t')]$ between $(\alpha, t) \neq (\alpha', t')$ with color 1 if there exists p, p' in $[\![q]\!]$ such that $(x_{t,p}^\alpha, i_t^\alpha, x_{t+1,p}^\alpha) = (x_{t',p'}^{\alpha'}, i_{t'}^{\alpha'}, x_{t'+1,p'}^{\alpha'})$ and the edge $[(\alpha, t), -1, (\alpha', t')]$ with color -1 if $(x_{t,p}^\alpha, i_t^\alpha, x_{t+1,p}^\alpha) = (x_{t'+1,p'}^{\alpha'}, i_{t'}^{\alpha'*}, x_{t',p'}^{\alpha'})$. In words,

we place an edge between two elements of T if they share a common edge in G_γ, the color encoding the orientation.

Assume that $[(\alpha,t),1,(\alpha',t')]$ is a colored edge of Γ and that $t_j^\alpha \le t < t_{j+1}^\alpha$. Then, since all vertices $x_{t+\delta,p}^\alpha$ have degree 2 for all $\delta \in \mathbb{Z}$ such that $t_j^\alpha \le t+\delta < t_{j+1}^\alpha$, we have that $[(\alpha,t+\delta),1,(\alpha',t'+\delta)]$ is a colored edge for all $\delta \in \mathbb{Z}$ such that $t_j^\alpha \le t+\delta < t_{j+1}^\alpha$. Similarly, if instead $[(\alpha,t),-1,(\alpha',t')]$ is an edge of Γ, we have that $[(\alpha,t+\delta),-1,(\alpha',t'-\delta)]$ is an edge for all δ such that $t_j^\alpha \le t+\delta < t_{j+1}^\alpha$. We deduce easily that the degree of $(\alpha,t) \in T$ in Γ is constant for all $t \in [t_j^\alpha, t_{j+1}^\alpha)$. Moreover, by definition, $(\alpha,t) \in T$ has degree 0 in Γ iif $[x_t^\alpha, i_t^\alpha, x_{t+1}^\alpha]$ is an isolated bracket.

Summing over all possibilities for γ', we find from (52) that

$$I = \sum_{\gamma':\gamma'\sim\gamma} a(\gamma') \le (2d-1)^\theta n^v \sum \prod_{\alpha=1}^{2\theta} \left\| \prod_{j=1}^{l_\alpha - 1} \prod_{t=t_j^\alpha}^{t_{j+1}^\alpha - 1} a_{i_t^\alpha} \right\|, \tag{56}$$

where the sum is over all $(i_t^\alpha)_{(\alpha,t)\in T}$ in $[\![2d]\!]$ such that for all $(\alpha,t),(\alpha',t')$:

$$i_{t+1}^\alpha \neq i_t^{\alpha *},\ i_t^\alpha = i_{t'}^{\alpha'} \text{ if } [(\alpha,t),1,(\alpha',t')] \in \Gamma \text{ and } i_t^\alpha = (i_{t'}^{\alpha'})^* \text{ if } [(\alpha,t),-1,(\alpha',t')] \in \Gamma. \tag{57}$$

The factor $(2d-1)^\theta$ in (56) comes from the choices of $i_{\ell+1}^\alpha$.

For ease of notation, for $r=(\alpha,t)$ and $s=(\alpha,t')$ with $t<t'$ in $[\![\ell+1]\!]$ and $\alpha \in [\![2\theta]\!]$, we will denote by $[r,s)$ the sequence $((\alpha,t),(\alpha,t+1),\cdots,(\alpha,t'-1))$ and write $u \in [r,s)$ if $u=(\alpha,t'')$ with $t \le t'' < t'$. Such sequence $[r,s)$ will be called an *interval* of *length* $t'-t$. If $[r_1,s_1)$ and $[r_2,s_2)$ are two intervals, $r_i=(\alpha_i,t_i)$, $s_i=(\alpha_i,t_i')$ we say that $[r_1,s_1)$ and $[r_2',s_2')$ are 1-*paired* if they have the same length and, for all $0 \le \delta < t_i'-t_i$, either: $[(\alpha_1,t_1+\delta),1,(\alpha_2,t_2+\delta)]$ is an edge of Γ or $[(\alpha_1,t_1+\delta),-1,(\alpha_2,t_2'-\delta)]$ is an edge of Γ. We say that there are 2-*paired*, if there exists a third interval which is 1-paired to both of them. Iteratively, we define k-*paired* intervals. Finally, we say that two intervals are *paired* if they are k-paired for some $k \ge 1$. The key property is that if two intervals $[r_1,s_1)$ and $[r_2,s_2)$ are paired then, with the above notation, either

$$\text{for all } 0 \le \delta < t_i'-t_i\text{: } i_{t_1+\delta}^{\alpha_1} = i_{t_2+\delta}^{\alpha_2} \quad \text{or} \quad \text{for all } 0 \le \delta < t_i'-t_i\text{: } i_{t_1+\delta}^{\alpha_1} = (i_{t_2'-\delta}^{\alpha_2})^*. \tag{58}$$

Our goal is to find paired intervals $[t_j^\alpha, t_{j+1}^\alpha)$ of long length in (56) and then use Lemma 25 for each of them. More precisely, we claim that there exists a partition of the set T made of m intervals $[r_i,r_{i+1}), i \in [\![m]\!]$, with $m \le 8l$ and a classification of the intervals $[r_i,r_{i+1})$ into 3 types. An interval $[r_i,r_{i+1})$ is of type (1) iif all $u \in [r_i,r_{i+1})$ are of degree 0 in Γ. There is a matching of intervals of type (2) such that two matched intervals of type (2) are paired together (by 'matching', we mean a pair partition, we use the word matching to make a distinction between 'matched intervals' and

'paired intervals' defined above). Finally, any $(\alpha,t)\in T$ in an interval of type (3) is connected in Γ to an element in an interval of type (2).

Let us first assume the existence of such partition and conclude the proof of Lemma 24. For $c\in\{1,3\}$, we let m_c be the number of intervals of type (c) and $k_{c,j}$ be the length of the j-th interval. Let $2m_2$ be the number of intervals of type (2) and $k_{2,j}$ be the common length of the two j-th paired intervals. We get from (56) and (58)

$$\begin{aligned} I &\le (2d)^\theta n^v \prod_{j=1}^{m_1}\left\{\sum_{\substack{1\le t\le k_{1,j}\\ i_{t+1}\neq i_t^*}}\left\|\prod_{t=1}^{k_{1,j}}a_{i_t}\right\|\right\}\prod_{j=1}^{m_2}\left\{\sum_{\substack{1\le t\le k_{2,j}\\ i_{t+1}\neq i_t^*}}\left\|\prod_{t=1}^{k_{2,j}}a_{i_t}\right\|^2\right\}\prod_{j=1}^{m_3}\left\{\max_{\substack{1\le t\le k_{3,j}\\ i_{t+1}\neq i_t^*}}\left\|\prod_{t=1}^{k_{3,j}}a_{i_t}\right\|\right\}\\ &= (2d)^\theta n^v\prod_{c=1}^{3}\prod_{j=1}^{m_c} I_{c,j}. \end{aligned}$$

We note that to obtain the above expression we have used Cauchy-Schwarz inequality for some intervals of type (2): for example

$$\sum_{\substack{1\le t\le k\\ i_{t+1}\neq i_t^*}}\left\|\prod_{t=1}^{k}a_{i_t}\right\|\left\|\prod_{t=1}^{k}a_{i^*_{k-t+1}}\right\| \le \sqrt{\sum_{\substack{1\le t\le k\\ i_{t+1}\neq i_t^*}}\left\|\prod_{t=1}^{k}a_{i_t}\right\|^2}\sqrt{\sum_{\substack{1\le t\le k\\ i_{t+1}\neq i_t^*}}\left\|\prod_{t=1}^{k}a_{i^*_{k-t+1}}\right\|^2} = \sum_{\substack{1\le t\le k\\ i_{t+1}\neq i_t^*}}\left\|\prod_{t=1}^{k}a_{i_t}\right\|^2.$$

Then, by Lemma 25, we have $I_{1,j}\le c^{k_{1,j}}\rho^{k_{1,j}}$, $I_{2,j}\le c^2\rho^{2k_{2,j}}$ and $I_{3,j}\le c\rho^{k_{3,j}}$. So finally, we deduce that

$$I\le (2d)^\theta n^v c^b c^m \rho^{2\ell\theta}.$$

We are have used that $2m_2+m_3\le m$, $\sum_j k_{1,j}=b$ and $\sum k_{1,j}+2\sum k_{2,j}+\sum k_{3,j}=2\ell\theta$. Thus, from the assumption $m\le 8l$ and from (55), up to modifying the constant c, we deduce that the conclusion of Lemma 24 holds provided the existence of the partition $[r_j,r_{j+1}), j\in[\![m]\!]$, of T.

The rest of the proof is devoted to the construction of the partition. To that end, we define a natural greedy pairing algorithm. At a given step $i\ge 0$ of the algorithm, there will be a partition of T into intervals $[r_j^{(i)},r_{j+1}^{(i)})$. Some intervals of the current partition will be said to be *active*, the others are *frozen* (frozen intervals will remain unchanged in all subsequent partitions). The frozen intervals have been classified in types, the active intervals have not. The algorithm stops when all intervals are frozen. At step $i=0$, the initial partition is $[t_j^\alpha,t_{j+1}^\alpha)$, $\alpha\in[\![2\theta]\!], j\in[\![l_\alpha]\!]$, and all intervals are active. At each step of the algorithm, the number of active intervals decreases until there is none. Thus, the algorithm stops after at most l steps.

We have the following induction hypothesis: each active interval $[r_j^{(i)},r_{j+1}^{(i)})$ is contained in $[t_{j'}^\alpha,t_{j'+1}^\alpha)$ for some $\alpha\in[\![2m]\!]$ and $j'\in[\![l_\alpha]\!]$. Moreover, any (α,t) in a frozen interval of type (3) is connected in Γ to an element in a frozen interval of type (2). We now describe the induction step. We pick the longest active sequence, say $[r_{j_0}^{(i)},r_{j_0+1}^{(i)})$. There are several cases to consider:

(i) *All $u \in T$ with $u \in [r_{j_0}^{(i)}, r_{j_0+1}^{(i)})$ have common degree* 0. Then, we leave the partition unchanged, the interval $[r_{j_0}^{(i)}, r_{j_0+1}^{(i)})$ is classified as type (1) and becomes frozen.

(ii) *All $u \in T$ with $u \in [r_{j_0}^{(i)}, r_{j_0+1}^{(i)})$ have common degree at least* 1*, then we pick an interval $[s, s')$ such that $[s, s')$ and $[r_{j_0}^{(i)}, r_{j_0+1}^{(i)})$ are* 1*-paired (its existence is guaranteed by the induction hypothesis). The interval $[s, s')$ intersects a finite number, say α, of intervals, $\mathcal{I} = \{[r_{j_1+k-1}^{(i)}, r_{j_1+k}^{(i)}), k \in [\![\alpha]\!]\}$. We assume that the corresponding color is* 1. There are further subcases to consider:

(ii-a) *The intervals in $\mathcal{I}$ are all active and $\mathcal{I}$ does not contain $[r_{j_0}^{(i)}, r_{j_0+1}^{(i)})$.* Then, we merge the intervals in $\mathcal{I}$ into 3 intervals $[r_{j_1}^{(i)}, s)$, $[s, s')$ and $[s', r_{j_1+\alpha}^{(i)})$ (the two extreme intervals may be empty). The two extreme intervals are active. The intervals $[s, s')$ and $[r_{j_0}^{(i)}, r_{j_0+1}^{(i)})$ are frozen, they are classified as type (2) and they are matched together. The rest of the partition remains unchanged at step $i+1$. Note that if $\alpha = 1$ then the two extreme intervals are empty since we have picked the longest interval.

(ii-b) *The intervals in $\mathcal{I}$ are all active and $\mathcal{I}$ contains $[r_{j_0}^{(i)}, r_{j_0+1}^{(i)})$.* We then have that $\alpha \geq 2$ (the case $\alpha = 1$ is ruled out by the assumption that the corresponding color is 1) and $[r_{j_0}^{(i)}, r_{j_0+1}^{(i)})$ is one of the two extreme intervals of $\mathcal{I}$. Assume for example that $j_1 = j_0$. We denote by k and k_0 the lengths of $[r_{j_0}^{(i)}, s)$ and $[r_{j_0}^{(i)}, r_{j_0+1}^{(i)})$. Then, $1 \leq k \leq k_0$ and $h = k_0 + k$ is the total length of $[r_{j_1}^{(i)}, s')$. By construction if $[(\alpha, t), (\alpha, t'))$ and $[(\alpha, t+k), (\alpha, t'+k))$ are contained $[r_{j_1}^{(i)}, s')$ then they are paired. We let $k' = ak$ where $a \geq 1$ is the largest integer such that $ak \leq h/2$. We have $k' \geq h/4$ since otherwise, $k' + k \leq 2k' < h/2$. Assume for example that $h/4 \leq k' < h/3$. Let $r = h - 3k'$. We divide the sequence $(0, \ldots, h-1)$, into $(pk', \ldots, pk' + r - 1)$ for $p \in \{0, 1, 2, 3\}$ and $(qk' + r, \ldots, (q+1)k' - 1)$, for $q \in \{0, 1, 2\}$. Mapping the interval $[r_{j_1}^{(i)}, s')$ into the sequence $(0, \ldots, h-1)$, we have split $[r_{j_1}^{(i)}, r_{j_1+\alpha}^{(i)})$ into 7 frozen intervals at step $i+1$. The two intervals corresponding to $p \in \{0, 1\}$ are of type (2) and are matched together, similarly for the two intervals associated to $p \in \{2, 3\}$. Finally the three intervals corresponding to $q \in \{0, 1, 2\}$ are mutually paired, two of them are classified as type (2) and the remaining one as type (3). The interval $[s', r_{j_1+\alpha}^{(i)})$ becomes an active interval (if not empty). The rest of the partition remains unchanged at step $i+1$. The case $h/3 \leq k' \leq h/2$ is treated similarly with 5 intervals, 3 of length r and 2 of length $k' - r$ where $r = h - 2k'$ (if $r = 0$, 2 intervals are sufficient).

(ii-c) *The intervals in $\mathcal{I}$ are all frozen.* Note that by the induction hypothesis the intervals of $\mathcal{I}$ are of type (2) or (3) and all elements in $\mathcal{I}$ are connected in Γ to an element in a frozen interval of type (2). The interval $[r_{j_0}^{(i)}, r_{j_0+1}^{(i)})$ becomes a frozen interval of type (3). The rest of the

partition remains unchanged at step $i+1$.

(ii-d) *The set $\mathcal{I}$ contains both active and frozen intervals.* We split $\mathcal{I}$ into at most three set of contiguous intervals, say $\mathcal{I}_1, \dots, \mathcal{I}_k$ with $k \in \{2,3\}$, such that $\mathcal{I}_1$ contains only active intervals and the one or two others contain only frozen intervals. This is possible since at each step we have chosen the longest active interval: it follows that the union of frozen intervals is a union of disconnected intervals such each of these disconnected intervals has a length larger or equal than the longest active interval. We also split $[r_{j_0}^{(i)}, r_{j_0+1}^{(i)})$ into two or three intervals, $J_1, \dots, J_k$, such that for all $p \in [\![k]\!]$, J_p is paired with I_p which is contained in the union of the intervals in $\mathcal{I}_p$. For each p, we may apply one of the above steps.

(iii) *As in* (ii) *but the corresponding color is* -1. Up to minor modifications on the indices, the above argument for the color 1 extends in this case as well.

At each step, the number of active intervals decreases. It follows that the algorithm stops after at most l steps with a partition and a classification by types which has all desired properties. Since $[r_{j_0}^{(i)}, r_{j_0+1}^{(i)})$ is split into at most $7+1=8$ parts, the total number of intervals m is bounded by $8l$. This concludes the proof of Lemma 24. □

All ingredients are gathered to prove Theorem 20.

Proof of Theorem 20. We fix $\varepsilon > 0$. For some small $\delta > 0$ to be fixed later on, we set

$$\theta = \lceil n^{\delta/q} \rceil.$$

We start by assuming that (for $n \geq 3$),

$$q \leq \frac{1}{2^7} \frac{\ln n}{\ln \ln n} \quad \text{and} \quad C_0 q \leq \ell \leq 2C_0 q \tag{59}$$

for some constant C_0 to be also fixed. In this proof, we write c for a universal constant and C for a constant which depends on d, ε ans $\max_i \|a_i\|$.

From (47), with $A(\gamma)$ as in Lemma 24 and $\mathcal{P}_{\ell,\theta}(v, e_1)$ as in Lemma 22, we find

$$\mathbf{E}\Big(\|B^\ell\|^{2\theta} \Big) \leq \sum_{v=1}^{q(\ell\theta+1)} \sum_{e_1=0}^{\infty} |\mathcal{P}_{\ell,\theta}(v,e_1)| \max_{\gamma \in \mathcal{P}_{\ell,\theta}(v,e_1)} p(\gamma) A(\gamma),$$

where we have used, by Lemma 23 that $p(\gamma) = 0$ unless $v \leq q(\ell\theta+1)$. Set $\chi = q(\ell\theta+1) + e_1/2 - v$. By Lemma 23 and Lemma 24, we have,

$$p(\gamma)A(\gamma) \leq n^{-q\ell\theta} \Big((cq\ell\theta)^{q/2} n^{-1/8} \Big)^{b+\frac{e_1}{q}} (cq\ell\theta)^{2\chi}\, n^v C^{b+q\theta+\chi} \rho^{2\ell\theta}, \tag{60}$$

where b is the number of isolated brackets of γ. For our choice of parameters in (59), for any constant $c > 0$, we have for all n large enough: $(cq)^q \leq n^{1/2^7}$. It follows that, if δ is small enough, for all n large enough,

$$(cq\ell\theta)^{q/2} n^{-1/8} \leq n^{1/2^7+\delta/2-1/8} \leq n^{-1/10}.$$

We deduce that for all n large enough, the factor in b in (60) is bounded by 1. Therefore, for all n large enough, we have

$$\max_{\gamma \in \mathcal{P}_{\ell,\theta}(v,e_1)} p(\gamma) A(\gamma) \leq n^{-q\ell\theta} n^{-\frac{e_1}{10q}} (cq\ell\theta)^{2\chi} n^v C^{q\theta+\chi} \rho^{2\ell\theta}.$$

We set $r = q(\ell\theta + 1) - v$ and use Lemma 22. Since $v \leq 2q\ell\theta$, we obtain, for some new constant $C > 0$,

$$\mathbf{E}\Big(\|B^\ell\|^{2\theta}\Big) \leq \rho^{2\ell\theta} C^{q\theta} ((q\ell\theta)^9 n)^q \sum_{r=0}^{\infty} \left(\frac{(Cq\ell\theta)^8}{n}\right)^r \sum_{e_1=0}^{\infty} \left(\frac{(Cq\ell\theta)^4}{n^{\frac{1}{10q}}}\right)^{e_1}.$$

For our choices of parameters in (59), for all n large enough,

$$(Cq\ell\theta)^{4q} \leq (2CC_0 q^2\theta)^{4q} \leq n^{1/2^4+4\delta}.$$

This last expression is bounded by $n^{1/12}$ for δ small enough. In particular, the above geometric series in e_1 converges and

$$\sum_{e_1=0}^{\infty} \left(\frac{(Cq\ell\theta)^4}{n^{\frac{1}{10q}}}\right)^{e_1} \leq \sum_{e_1=0}^{\infty} n^{-\frac{e_1}{60q}} \leq \frac{1}{1 - n^{-\frac{1}{60q}}} \leq 2.$$

for all n large enough. Similarly, if δ is small enough, the series in r converges and is bounded by 2. We fix such choice of $\delta > 0$, we deduce that for all n large enough,

$$\mathbf{E}\Big(\|B^\ell\|^{2\theta}\Big) \leq 4\rho^{2\ell\theta} C^{q\theta} ((q\ell\theta)^9 n)^q.$$

We fix the constant C_0 in (59) such that $C^{q\theta/(2\ell\theta)} \leq C^{1/(2C_0)} \leq 1 + \varepsilon$. We note that

$$n^{q/(2\ell\theta)} \leq \exp\left(\frac{\ln n}{2C_0 n^{\delta/q}}\right).$$

If we further assume that $q \leq \max(\delta/2, 2^{-7}) \ln n/(\ln\ln n)$, we get that $n^\theta \geq n^{\delta/q} \geq (\ln n)^2$. It follows that for these choices of parameters, for all n large enough,

$$4((q\ell\theta)^9 n)^q \leq (1+\varepsilon)^{2\ell\theta}.$$

We thus have checked that

$$\mathbf{E}\Big(\|B^\ell\|^{2\theta}\Big) \leq \big(\rho(1+\varepsilon)^2\big)^{2\ell\theta}.$$

We have $\rho(B_\star) \le \|B_\star\| \le C$ for some $C > 0$. Also, from (44), we have $\rho \le \rho(B_\star) + \varepsilon$. We get

$$\mathbf{E}\Big(\|B^\ell\|^{2\theta}\Big) \le (\rho(B_\star) + \varepsilon')^{2\ell\theta},$$

where ε' can be taken arbitrarily small if ε is small enough. From Markov inequality, we obtain that

$$\mathbb{P}\Big(\|B^\ell\|^{1/\ell} \ge (1+\varepsilon')(\rho(B_\star) + \varepsilon')\Big) \le (1+\varepsilon')^{-2\ell\theta}.$$

Adjusting the value of ε and the constants, we obtain easily the required statement. □

5.3 Proof of Theorem 2

The orthogonal projection of A defined in (3) onto $H_r^\perp$ can be written as

$$[A] = a_0 \otimes 1 + \sum_{i=1}^{2d} a_i \otimes [V_i]. \tag{61}$$

We fix $\varepsilon > 0$ and let $\delta = \delta(\varepsilon)$ be as in Theorem 19(ii). In view of Theorem 19, the event

$$\Big\{\|A_{|H_r^\perp}\| \ge \|A_\star\| + \varepsilon\Big\}$$

is contained in the event

$$\mathcal{E}_\varepsilon = \bigcup_{a=(a_1,\dots a_{2d}) \in S_\varepsilon^{2d}} \mathcal{E}_\delta(a),$$

where

$$\mathcal{E}_\delta(a) = \{\rho(B) \ge \rho(B_\star) + \delta\},$$

with $B = B(a)$ is an in (41), $B_\star(a) = B_\star$ is as in (42) and

$$S_\varepsilon = \{b \in M_r(\mathbb{C}) : \|b\| \le \varepsilon^{-1}\}.$$

To prove Theorem 2 it is thus sufficient to check that for any $\varepsilon > 0$, for all n large enough,

$$\mathbb{P}(\mathcal{E}_\varepsilon) \le n^{-2}. \tag{62}$$

To prove (62), we need to use a net on S_ε^{2d}. A similar argument appears in [4]. Due to the lack of uniform continuity of spectral radii, we perform the net argument with operator norms. From (41), for $a \in M_r(\mathbb{C})^{2d}$, the map $a \mapsto B(a)$ is linear and $\|B(a)\| \le (2d-1)\|a\|$, where

$$\|a\| = \sum_{i=1}^{2d} \|a_i\|.$$

Let $\ell = \lfloor Cq \rfloor$ be as in Theorem 20. The map $a \mapsto B^\ell(a)$ satisfies a deviation inequality

$$\begin{aligned} \|B^\ell(a) - B^\ell(a')\| &\leq \ell \max(\|B(a)\|, \|B(a')\|)^{\ell-1} \|B(a-a')\| \\ &\leq \ell(2d-1)^\ell \max(\|a\|, \|a'\|)^{\ell-1} \|a-a'\|. \end{aligned} \tag{63}$$

For a given $\eta > 0$, the net N_η of S_ε^{2d} is built as follows. First, since all matrix norms are equivalent and $M_r(\mathbb{C}) \simeq \mathbb{R}^{2r^2}$, there exists a subset $N_\eta^1 \subset \{b \in M_r(\mathbb{C}) : \|b\| \leq \varepsilon^{-1}\}$ of cardinal number at most $(c/(\varepsilon\eta))^{2r^2}$ such that for any $b \in S_\varepsilon$, there exists $b' \in N_\eta$ with $\|b - b'\| \leq \eta$ (the constant c depends on r). We set $N_\eta = (N_\eta^1)^{2d}$. From (63), for some new constant $c > 0$ (depending on ε, r, d), for any $a \in S_\varepsilon$, there exists $a' \in N_\eta$ such that

$$\|B^\ell(a) - B^\ell(a')\| \leq \ell(2d-1)\left(\frac{2d-1}{\varepsilon}\right)^{\ell-1} \eta \leq c^\ell \eta.$$

Besides, from (40), for all $\eta \leq \eta_0$ small enough,

$$\left|\rho(B_\star(a)) - \rho(B_\star(a'))\right| \leq \frac{\delta}{3}, \tag{64}$$

If $\eta = \min(\eta_0, (\delta/3c)^\ell)$ and $\mathcal{E}_{\delta/3}(a')$ does not hold, we deduce that

$$\begin{aligned} \|B^\ell(a)\| &\leq \|B^\ell(a')\| + \|B^\ell(a) - B^\ell(a')\| \\ &< \left(\rho(B_\star(a')) + \frac{\delta}{3}\right)^\ell + \left(\frac{\delta}{3}\right)^\ell \\ &< \left(\rho(B_\star(a')) + \frac{2\delta}{3}\right)^\ell \\ &< (\rho(B_\star(a)) + \delta)^\ell, \end{aligned}$$

where we have used (64) at the last line. We find that, for our choice of η,

$$\mathcal{E}_\varepsilon = \bigcup_{a \in S_\varepsilon^{2d}} \mathcal{E}_\delta(a) \subset \bigcup_{a \in N_\eta} \mathcal{E}_{\delta/3}(a),$$

and, for some $c_1 > 0$ (depending on ε, r and d),

$$|N_\eta| \leq c_1^\ell.$$

We may now use the union bound to obtain an estimate of (62):

$$\begin{aligned} \mathbb{P}(\mathcal{E}_\varepsilon) &\leq \sum_{a \in N_\eta} \mathbb{P}\big(\mathcal{E}_{\delta/3}(a)\big) \\ &\leq |N_\eta| C \exp(-(\ln n)^2), \end{aligned}$$

where at the second line, we have used Theorem 20 (the constant C depends on ε, d, r). For our choice of ℓ, we have $|N_\eta| \leq c_1^\ell \leq n$ for all n large enough. The bound (62) follows. □

References

[1] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[2] G. Aubrun and S. a. J. Szarek. *Alice and Bob meet Banach*, volume 223 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2017. The interface of asymptotic geometric analysis and quantum information theory.

[3] C. Bordenave. A new proof of Friedman's second eigenvalue theorem and its extension to random lifts. *Ann. Sci. Éc. Norm. Supér.*, to appear.

[4] C. Bordenave and B. Collins. Eigenvalues of random lifts and polynomials of random permutation matrices. *Ann. of Math. (2)*, 190(3):811–875, 2019.

[5] B. Collins. Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability. *Int. Math. Res. Not.*, (17):953–982, 2003.

[6] B. Collins and P. Y. Gaudreau Lamarre. $*$-freeness in finite tensor products. *Adv. in Appl. Math.*, 83:47–80, 2017.

[7] B. Collins and C. Male. The strong asymptotic freeness of Haar and deterministic matrices. *Ann. Sci. Éc. Norm. Supér. (4)*, 47(1):147–163, 2014.

[8] B. Collins and S. Matsumoto. Weingarten calculus via orthogonality relations: new applications. *ALEA Lat. Am. J. Probab. Math. Stat.*, 14(1):631–656, 2017.

[9] B. Collins and I. Nechita. Gaussianization and eigenvalue statistics for random quantum channels (III). *Ann. Appl. Probab.*, 21(3):1136–1179, 2011.

[10] B. Collins and P. Śniady. Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Comm. Math. Phys.*, 264(3):773–795, 2006.

[11] J. Friedman, A. Joux, Y. Roichman, J. Stern, and J. P. Tillich. *The action of a few random permutations on r-tuples and an application to cryptography*, pages 375–386. Springer Berlin Heidelberg, Berlin, Heidelberg, 1996.

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

[13] C. E. González-Guillén, C. Palazuelos, and I. Villanueva. Euclidean distance between Haar orthogonal and Gaussian matrices. *J. Theoret. Probab.*, 31(1):93–118, 2018.

[14] R. Goodman and N. R. Wallach. *Symmetry, representations, and invariants*, volume 255 of *Graduate Texts in Mathematics*. Springer, Dordrecht, 2009.

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

[16] A. W. Harrow. Quantum expanders from any classical Cayley graph expander. *Quantum Inf. Comput.*, 8(8-9):715–721, 2008.

[17] M. B. Hastings. Random unitaries give quantum expanders. *Phys. Rev. A (3)*, 76(3):032315, 11, 2007.

[18] S. Janson. *Gaussian Hilbert spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997.

[19] T. Jiang. How many entries of a typical orthogonal matrix can be approximated by independent normals? *Ann. Probab.*, 34(4):1497–1529, 2006.

[20] E. Kowalski. Spectral theory in Hilbert spaces. available at https://people.math.ethz.ch/ kowalski/spectral-theory.pdf, 2009.

[21] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[22] S. Matsumoto and J. Novak. Jucys-Murphy elements and unitary matrix integrals. *Int. Math. Res. Not. IMRN*, (2):362–397, 2013.

[23] J. A. Mingo and R. Speicher. *Free probability and random matrices*, volume 35 of *Fields Institute Monographs*. Springer, New York; Fields Institute for Research in Mathematical Sciences, Toronto, ON, 2017.

[24] G. J. Murphy. *C^*-algebras and operator theory*. Academic Press, Inc., Boston, MA, 1990.

[25] A. Nica. Asymptotically free families of random unitaries in symmetric groups. *Pacific J. Math.*, 157(2):295–310, 1993.

[26] G. Pisier. A simple proof of a theorem of Kirchberg and related results on C^*-norms. *J. Operator Theory*, 35(2):317–335, 1996.

[27] G. Pisier. Quantum expanders and geometry of operator spaces. *J. Eur. Math. Soc. (JEMS)*, 16(6):1183–1219, 2014.

[28] M. Reed and B. Simon. *Methods of modern mathematical physics. IV. Analysis of operators.* Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1978.

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

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

[31] D. P. Želobenko. *Compact Lie groups and their representations.* American Mathematical Society, Providence, R.I., 1973. Translated from the Russian by Israel Program for Scientific Translations, Translations of Mathematical Monographs, Vol. 40.