A NEW APPROACH TO STRONG CONVERGENCE

CHI-FANG CHEN, JORGE GARZA-VARGAS, JOEL A. TROPP,
AND RAMON VAN HANDEL

ABSTRACT. A family of random matrices $\boldsymbol{X}^N = (X_1^N, \ldots, X_d^N)$ converges strongly to a family $\boldsymbol{x} = (x_1, \ldots, x_d)$ in a C^*-algebra if $\|P(\boldsymbol{X}^N)\| \to \|P(\boldsymbol{x})\|$ for every noncommutative polynomial P. This phenomenon plays a central role in several recent breakthrough results on random graphs, geometry, and operator algebras. However, strong convergence is notoriously difficult to prove and has generally required delicate problem-specific methods.

In this paper, we develop a new approach to strong convergence that uses only soft arguments. Our method exploits the fact that for many natural models, the expected trace of $P(\boldsymbol{X}^N)$ is a rational function of $\frac{1}{N}$ whose lowest order asymptotics are easily understood. We develop a general technique to deduce strong convergence directly from these inputs using the inequality of A. and V. Markov for univariate polynomials and elementary Fourier analysis. To illustrate the method, we develop the following applications.

1. We give a short proof of the result of Friedman that random regular graphs have a near-optimal spectral gap, and obtain a sharp understanding of the large deviations probabilities of the second eigenvalue.

2. We prove a strong quantitative form of the strong convergence property of random permutation matrices due to Bordenave and Collins.

3. We extend the above to any stable representation of the symmetric group, providing many new examples of the strong convergence phenomenon.

1. INTRODUCTION

Let $\boldsymbol{X}^N = (X_1^N, \ldots, X_d^N)$ be a sequence of d-tuples of random matrices of increasing dimension, and $\boldsymbol{x} = (x_1, \ldots, x_d)$ be a d-tuple of elements of a C^*-algebra. Then $\boldsymbol{X}^N$ is said to *converge strongly* to $\boldsymbol{x}$ if

$$\|P(X_1^N, \ldots, X_d^N)\| \xrightarrow{N \to \infty} \|P(x_1, \ldots, x_d)\| \quad \text{in probability}$$

for every noncommutative polynomial P.

The notion of strong convergence has its origin in the work of Haagerup and Thorbjørnsen [30] in the context of complex Gaussian (GUE) matrices. It has since proved to be a powerful tool in a broad range of applications. Notably, strong convergence plays a central role in a series of recent breakthroughs in the study of random lifts of graphs [7], random covering spaces [37, 44, 36, 46], random minimal surfaces [62], and operator algebras [30, 29, 33, 4, 34, 8]; we refer to the recent paper [8] for a more extensive discussion and bibliography.

In many cases, norm convergence can be highly nontrivial to establish even for the simplest polynomials P. For example, let $\bar{S}_1^N, \ldots, \bar{S}_d^N$ be i.i.d. random permutation matrices of dimension N. Then the linear polynomial

$$\bar{A}^N = \bar{S}_1^N + \bar{S}_1^{N*} + \cdots + \bar{S}_d^N + \bar{S}_d^{N*}$$

is the adjacency matrix of a random $2d$-regular graph. Every $2d$-regular graph has largest eigenvalue $2d$ with eigenvector 1_N, while the Alon–Boppana bound states that the second eigenvalue is at least $2\sqrt{2d-1} - o(1)$ [53]. It was conjectured by Alon, and proved in a breakthrough monograph of Friedman [27], that the random $2d$-regular graph satisfies $\|\bar{A}^N|_{\{1_N\}^\perp}\| = 2\sqrt{2d-1} + o(1)$, so that such graphs have near-optimal second eigenvalue. The much more general result that random permutation matrices converge strongly, due to Bordenave and Collins [7], is of major importance in many recent applications cited above.

Given the power of the strong convergence phenomenon, it may be unsurprising that strong convergence has been notoriously difficult to prove (see, e.g., the discussion in [9]). In those cases where strong convergence has been established, the proofs rely on problem-specific methods and often require delicate computations. These obstacles have hampered efforts to establish strong convergence in new situations and to obtain a sharper understanding of this phenomenon.

In this paper, we develop a new approach to strong convergence that uses only soft arguments, and which requires limited problem-specific inputs as compared to previous methods. We will illustrate the method in the context of random permutation matrices and other representations of the symmetric group, resulting in surprisingly short proofs of strong convergence and new quantitative and qualitative information on the strong convergence phenomenon. In particular:

- We obtain a new proof of Friedman's theorem on the second eigenvalue of random regular graphs [27] that enables a sharp characterization of the tail behavior of the second eigenvalue. As a byproduct, our results settle several open problems on the behavior of random regular graphs.

- We obtain a nonasymptotic form of the strong convergence of random permutation matrices due to Bordenave and Collins [7, 8], which significantly improves the best known convergence rate for arbitrary polynomials.

- We establish strong convergence of random matrices defined by any stable representation of the symmetric group (in the sense of [25]). This provides a large new family of examples of the strong convergence phenomenon, and illustrates that strong convergence arises in settings far beyond the standard representation.

Further applications to several other models will appear in forthcoming work.

Let us emphasize at the outset that the main difficulty in establishing strong convergence is to prove the upper bound

$$\|P(X_1^N, \ldots, X_d^N)\| \le \|P(x_1, \ldots, x_d)\| + o(1) \quad \text{with probability } 1 - o(1)$$

as $N \to \infty$. The corresponding lower bound then follows automatically under mild assumptions that certainly hold for all situations considered in this paper; see, e.g., [44, pp. 17-18]. We therefore focus on the upper bound throughout this paper.

Organization of this paper. This paper is organized as follows. In section 2, we first outline the core ingredients of our approach in a general setting, while section 3 presents the main results obtained in this paper.

The next two sections introduce some general tools on which our methods are based. Section 4 recalls properties of polynomials and distributions that form the basis for our approach, while section 5 recalls basic properties of words in random permutation matrices that form the input for our methods.

The rest of the paper is devoted to the proofs of our main results. Sections 6 and 7 prove master inequalities for polynomials and smooth functions, respectively. These inequalities are applied in section 8 to random regular graphs, and in section 9 to strong convergence of random permutation matrices. Finally, section 10 extends the above results to general stable representations of the symmetric group.

Notation. In the following, $a \lesssim b$ denotes that $a \leq Cb$ for a universal constant C.

We denote by $\mathcal{P}_q$ the space of real polynomials $h : \mathbb{R} \to \mathbb{R}$ of degree at most q, and by $\mathcal{P}$ the space of all real polynomials. We denote by $h^{(m)}$ the mth derivative of a univariate function, and by $\|h\|_{C^m[a,b]} := \sum_{j=0}^{m} \sup_{x \in [a,b]} |h^{(j)}(x)|$.

We will denote by $\mathrm{M}_N(\mathbb{C})$ the space of $N \times N$ matrices with complex entries. The trace of a matrix M is denoted as $\mathrm{Tr}\, M$, and the normalized trace is denoted as $\mathrm{tr}_N M := \frac{1}{N} \mathrm{Tr}\, M$. The identity matrix is denoted $\mathbf{1}_N \in \mathrm{M}_N(\mathbb{C})$, and $1_N \in \mathbb{C}^N$ is the vector all of whose entries equal one. The symmetric group on N letters is denoted as $\mathbf{S}_N$, and the free group with free generators $g_1, \ldots, g_d$ is denoted $\mathbf{F}_d$.

We will on occasion use standard C^*-algebra terminology, though it will not play a major role in this paper. We refer to [52] for an excellent introduction.

2. Outline

The main results of this paper will be presented in section 3, which can be read independently of the present section. However, as the new approach that underlies their proofs is much more broadly applicable, we begin in this section by introducing the core ingredients of the method in a general setting.

We aim to explain how to prove that $\|X^N\| \leq \|X_{\mathrm{F}}\| + o(1)$, where X^N is an N-dimensional self-adjoint random matrix and X_{F} is an element of a C^*-probability space $(\mathcal{A}, \tau)$. We assume these models are related by

$$\lim_{N \to \infty} \mathbf{E}[\mathrm{tr}_N(X^N)^p] = \tau(X_{\mathrm{F}}^p) \qquad \text{for all } p \in \mathbb{N}, \tag{2.1}$$

which implies *weak convergence* of the spectral distribution of X^N to that of X_{F}. The moments $\mathbf{E}[\mathrm{tr}_N(X^N)^p]$ generally admit a combinatorial description whose behavior as $N \to \infty$ is well understood, so that (2.1) is readily accessible.

To date, proofs of *strong* convergence were based on resolvent equations [30, 61], interpolation [20, 55, 56, 2], or the moment method [7, 9, 8]. The first two approaches rely on analytic tools, such as integration by parts formulae or free stochastic calculus, that are only available for special models. In contrast, the only known way to access the properties of models such as those based on random permutations or group representations is through moment computations. We therefore begin by recalling the challenges faced by the moment method.

2.1. **Prelude: the moment method.** The classical moment method aims to deduce norm convergence from moment convergence by noting that (2.1) implies

$$\left(\mathbf{E}\|X^N\|^{2p}\right)^{1/2p} \leq \left(\mathbf{E}[\mathrm{Tr}(X^N)^{2p}]\right)^{1/2p} \leq N^{1/2p}\left(\|X_{\mathrm{F}}\| + o(1)\right) \tag{2.2}$$

as $N \to \infty$. This does not in itself suffice to prove $\|X^N\| \leq \|X_{\mathrm{F}}\| + o(1)$, as the right-hand side diverges when p is fixed. The essence of the moment method is that the desired bound would follow if (2.2) remains valid for $p = p(N) \gg \log N$, as then $N^{1/2p} = 1 + o(1)$. We emphasize this is a much stronger property than (2.1). In practice, the implementation of this method faces two major obstacles.

- While moment asymptotics for fixed p are accessible in many situations, the case where p grows rapidly with N is often a difficult combinatorial problem. Despite the availability of tools to facilitate the analysis of large moments for strong convergence, such as nonbacktracking and linearization methods (see, e.g., [7, 8]), the core of the analysis remains dependent on delicate computations that do not readily carry over to new situations.

- A more fundamental problem is that the inequality (2.2) that forms the foundation for the moment method may fail to hold altogether for any $p \gg \log N$. To see why, suppose that $\mathbf{P}[\|X^N\| \geq \|X_{\mathrm{F}}\| + \varepsilon] \geq N^{-c}$ for some $\varepsilon, c > 0$. Then

$$\mathbf{E}[\mathrm{Tr}(X^N)^{2p}] \geq N^{-c}(\|X_{\mathrm{F}}\| + \varepsilon)^{2p},$$

which precludes the validity of (2.2) for $p \gg \log N$. It was observed by Friedman [27] that this situation arises already in the setting of random regular graphs due to the presence with probability N^{-c} of dense subgraphs called "tangles". The application of the moment method in this setting requires conditioning on the absence of tangles, which significantly complicates the analysis.

2.2. **A new approach.** The approach of this paper is also based on moment computations, which are however used in an entirely different manner. As we will explain below, our method deduces norm convergence from moments by first letting $N \to \infty$ and then $p \to \infty$, bypassing the challenges of the moment method.

Inputs. Our approach requires two basic inputs that we presently describe.

Let $h \in \mathcal{P}_q$ be a polynomial of degree at most q. Then $\mathbf{E}[\mathrm{tr}_N h(X^N)]$ is a linear combination of the moments $\mathbf{E}[\mathrm{tr}_N(X^N)^p]$ for $p \leq q$. We will exploit the fact that in many situations, the moments are rational functions of $\frac{1}{N}$:

$$\mathbf{E}[\mathrm{tr}_N h(X^N)] = \Phi_h(\tfrac{1}{N}) = \frac{f_h(\tfrac{1}{N})}{g_q(\tfrac{1}{N})}, \tag{2.3}$$

where f_h and g_q are polynomials that depend only on h and q, respectively. This phenomenon is very common; e.g., it arises in many settings from Weingarten calculus [21, 19] or from genus expansions [48, Chapter 1].

In general, the function Φ_h is extremely complicated. However, our methods will use only soft information on its structure: we require upper bounds on the degrees of f_h and g_q (which are proportional to q for the models considered in this paper), and we must show that g_q does not vanish near zero. Both properties can be read off almost immediately from a combinatorial expression for Φ_h.

The second input to our method is the asymptotic expansion as $N \to \infty$

$$\mathbf{E}[\mathrm{tr}_N h(X^N)] = \nu_0(h) + \frac{\nu_1(h)}{N} + O\left(\frac{1}{N^2}\right), \tag{2.4}$$

where ν_0 and ν_1 are linear functionals on the space $\mathcal{P}$ of real polynomials. Clearly

$$\nu_0(h) = \Phi_h(0) = \lim_{N \to \infty} \mathbf{E}[\mathrm{tr}_N h(X^N)] = \tau(h(X_{\mathrm{F}})),$$

$$\nu_1(h) = \Phi_h'(0) = \lim_{N \to \infty} N\big(\mathbf{E}[\mathrm{tr}_N h(X^N)] - \tau(h(X_{\mathrm{F}}))\big)$$

are defined by the lowest-order asymptotics of the moments. We will exploit that simple combinatorial expressions for ν_0 and ν_1 are readily accessible in practice.

Outline. Our basic aim is to achieve the following.

(a) We aim to show that the validity of (2.4) extends from polynomials to smooth functions h. In particular, we will show that ν_1 extends from a linear functional on polynomials to a compactly supported distribution.

(b) We aim to show that $\operatorname{supp}\nu_1 \subseteq [-\|X_{\mathrm{F}}\|, \|X_{\mathrm{F}}\|]$.

A bound on the norm then follows easily. Indeed, let $\chi : \mathbb{R} \to [0,1]$ be a smooth function so that $\chi = 0$ on $[-\|X_{\mathrm{F}}\|, \|X_{\mathrm{F}}\|]$ and $\chi = 1$ on $[-\|X_{\mathrm{F}}\| - \varepsilon, \|X_{\mathrm{F}}\| + \varepsilon]^c$. Then $\nu_0(\chi) = \nu_1(\chi) = 0$ by (b), and thus (a) yields

$$\mathbf{P}[\|X^N\| \geq \|X_{\mathrm{F}}\| + \varepsilon] \leq \mathbf{E}[\operatorname{Tr}\chi(X^N)] = O\left(\frac{1}{N}\right).$$

As $\varepsilon > 0$ was arbitrary, $\|X^N\| \leq \|X_{\mathrm{F}}\| + o(1)$ in probability.

We now explain the steps that will be used to prove (a) and (b). We highlight two key features of our method: it is entirely based on moment computations; and it yields strong quantitative bounds essentially for free.

Step 1: the polynomial method. At the heart of our approach lies a general method to obtain a quantitative form of (2.4): we will show that

$$\left|\mathbf{E}[\operatorname{tr}_N h(X^N)] - \nu_0(h) - \frac{\nu_1(h)}{N}\right| \lesssim \frac{q^8}{N^2}\|h\|_{C^0[-K,K]} \tag{2.5}$$

for any $h \in \mathcal{P}_q$, where $\|X^N\| \leq K$ a.s. for all N. While achieving such a bound by previous methods would require hard analysis, it arises here from a soft argument: we can "upgrade" an asymptotic expansion (2.4) to a strong nonasymptotic bound (2.5) merely by virtue of the fact that Φ_h is rational.

To this end, observe that the left-hand side of (2.5) is nothing other than the remainder in the first-order Taylor expansion of Φ_h at zero, so that

$$\left|\Phi_h(\tfrac{1}{N}) - \Phi_h(0) - \tfrac{1}{N}\Phi_h'(0)\right| \leq \tfrac{1}{2N^2}\|\Phi_h''\|_{C^0[0,\frac{1}{N}]}.$$

This bound appears useless at first sight, as it relies on the behavior of $\Phi_h(x)$ for $x \notin J := \{\frac{1}{N} : N \in \mathbb{N}\}$ where its spectral interpretation is lost. However, as Φ_h is rational, we can control its behavior by its values on J alone by means of classical inequalities for arbitrary univariate polynomials f of degree q:

- The inequality $\|f'\|_{C^0[-1,1]} \leq q^2\|f\|_{C^0[-1,1]}$ of A. and V. Markov (Lemma 4.1).

- A corollary of the Markov inequality that $\|f\|_{C^0[-1,1]} \lesssim \sup_{x \in I} |f(x)|$ for any set $I \subset [-1,1]$ with $O(\frac{1}{q^2})$ spacing between its points (Lemma 4.2).

Applying these inequalities to the numerator and denominator of Φ_h yields (2.5) with minimal effort. We emphasize that the only inputs used here are upper bounds on the degrees of f_h and g_q in (2.3), and that g_q does not vanish near zero.

Step 2: the extension problem. We now aim to extend (2.5) to $h \in C^\infty$. This is not merely a technical issue: while $\mathbf{E}[\operatorname{tr}_N h(X^N)]$ and $\nu_0(h) = \tau(h(X_{\mathrm{F}}))$ are defined for $h \in C^\infty$ by functional calculus, it is unclear that $\nu_1(h)$ can even be meaningfully defined when h is not a polynomial.

These issues can be addressed using Fourier analysis. Consider first a polynomial $h \in \mathcal{P}_q$. Rather than applying (2.5) directly to h, we expand

$$h(x) = \sum_{j=0}^{q} a_j T_j(K^{-1}x)$$

in terms of Chebyshev polynomials T_j, and apply (2.5) to each T_j individually. This replaces $q^8 \|h\|_{C^0[-K,K]}$ by $\sum_{j=1}^{q} j^8 |a_j| \lesssim \|h\|_{C^9[-K,K]}$ in (2.5), where the latter inequality follows by classical Fourier analysis. Thus

$$\left| \mathbf{E}[\operatorname{tr}_N h(X^N)] - \nu_0(h) - \frac{\nu_1(h)}{N} \right| \lesssim \frac{1}{N^2} \|h\|_{C^9[-K,K]} \tag{2.6}$$

for every polynomial $h \in \mathcal{P}$. (We will in fact use a stronger inequality of Zygmund, cf. Lemma 4.4, that achieves better rates in our main results.)

The inequality (2.6) simultaneously serves two purposes. First, it ensures that its left-hand side extends uniquely to a compactly supported distribution, so that $\nu_1(h)$ can be uniquely defined for any $h \in C^\infty$. Consequently, (2.6) remains valid for any smooth function h as polynomials are dense in C^∞.

Step 3: the asymptotic moment method. It remains to bound the support of ν_1. To this end, we make fundamental use of a key property of compactly supported distributions: their support is bounded by the exponential growth rate of their moments (Lemma 4.9). In particular, if we define

$$\rho = \limsup_{p \to \infty} |\nu_1(x^p)|^{1/p} = \limsup_{p \to \infty} \lim_{N \to \infty} \left| N \left(\mathbf{E}[\operatorname{tr}_N (X^N)^p] - \tau(X_F^p) \right) \right|^{1/p},$$

then $\operatorname{supp} \nu_1 \subseteq [-\rho, \rho]$. Thus our method ultimately reduces to a form of the moment method, but where we first let $N \to \infty$ and only then $p \to \infty$.

In contrast to the moments of the random matrix X^N, the moments of ν_1 turn out to be easy to analyze for the models considered in this paper using a simple idea that is inspired by earlier work of Friedman [26, Lemma 2.4]: even though ν_1 does not have a clear spectral interpretation, its moments can be expressed as a sum of products of matrix elements of powers of X_F. As the sum only has polynomially many terms, the desired bound $\rho \leq \|X_F\|$ follows readily.

2.3. **Related work.** The observation that norm bounds follow from the validity of (2.4) for *smooth* functions h and a bound on first-order support $\operatorname{supp} \nu_1$ has been widely used since the works of Haagerup and Thorbjørnsen [30] and Schultz [61] (see also the work of Parraud [56] where higher-order expansions are considered). However, these works rely heavily on analytic and operator-algebraic tools that are not available for the kind of models considered in this paper. What is fundamentally new here is that our method achieves this aim using only simple moment computations, which are much more broadly applicable.

The polynomial method that lies at the heart of our approach is inspired by work in complexity theory [1]. We are aware of little precedent for the use of this method in a random matrix context, beside a distantly related idea that appears in Bourgain and Tzafriri [11, Theorem 2.3]. The first author used a variant of this idea in concurrent work to establish weak convergence of certain random unitary matrices that arise in quantum information theory [15, 14].

We emphasize that the appearance of Chebyshev polynomials in Step 2 above is unrelated to their connection with nonbacktracking walks that is used, for example, in [27, Lemma 2.3]. Indeed, our Chebyshev polynomials are normalized by the a.s. bound K on $\|X^N\|$ (e.g., $2d$ in Friedman's theorem) as opposed to the support of the limiting spectrum $\|X_F\|$ ($2\sqrt{2d-1}$ in Friedman's theorem).

Finally, we mention the interesting works [6, 23] that develop a method to bound the spectral radius of certain non-Hermitian matrices using $N \to \infty$ moment asymptotics with fixed p, similarly avoiding the challenges of the moment method in that setting. However, the methods developed in these works are unrelated to our approach and are not applicable to the problems considered in this paper.

3. Main results

In this section, we formulate and discuss the main results of this paper. As our primary aim is to achieve strong convergence, we will focus the presentation on upper bounds on the norm as explained in section 1. Let us note, however, that a byproduct of the analysis will also yield a quantitative form of weak convergence, which is of independent interest; see Corollary 7.2 below.

3.1. **Preliminaries.** Before we turn to our main results, we must briefly recall some basic facts about random permutation matrices and their limiting model. The following definitions will remain in force throughout this paper.

Definition 3.1. Let $\bar{S}_1^N, \dots \bar{S}_d^N$ be i.i.d. random permutation matrices of dimension N, and denote by $S_i^N := \bar{S}_i^N|_{\{1_N\}^\perp}$ their restriction to the invariant subspace $\{1_N\}^\perp \subset \mathbb{C}^N$. We will often write $\boldsymbol{S}^N = (S_1^N, \dots, S_d^N)$.

Definition 3.2. Let $\boldsymbol{s} = (s_1, \dots, s_d)$ be defined by $s_i := \lambda(g_i)$, where $g_1, \dots, g_d$ are the free generators of $\mathbf{F}_d$ and $\lambda : \mathbf{F}_d \to B(l^2(\mathbf{F}_d))$ is the left-regular representation defined by $\lambda(g)\delta_h = \delta_{gh}$. Define the vector state $\tau(a) := \langle \delta_e, a\delta_e \rangle$ on $B(l^2(\mathbf{F}_d))$.

The basic weak convergence property of random permutation matrices, due to Nica [50] (see Corollary 5.3 for a short proof), states that

$$\lim_{N \to \infty} \mathbf{E}\big[\operatorname{tr}_N P(\boldsymbol{S}^N, \boldsymbol{S}^{N*})\big] = \tau\big(P(\boldsymbol{s}, \boldsymbol{s}^*)\big)$$

for every noncommutative polynomial P. This property plays the role of (2.1) in section 2. The aim of the strong convergence problem is to prove that this convergence holds not only for the trace but also for the norm.

The basic inputs to the methods of this paper, as described in section 2.2, are well known in the present setting. They will be reviewed in section 5 below.

Remark 3.3. Even though S_i^N are $(N-1)$-dimensional matrices defined on $\{1_N\}^\perp$, we will normalize the trace tr_N by N rather than by $N-1$ as this leads to cleaner expressions for the rational functions that arise in the proof. This makes no difference to our results, and is mainly done for notational convenience.

3.2. **Random regular graphs.** Let $\bar{A}^N$ be the adjacency matrix of the random $2d$-regular graph with N vertices defined in section 1. Then $A^N := \bar{A}^N|_{\{1_N\}^\perp}$ is defined by the linear polynomial of random permutation matrices

$$A^N := S_1^N + S_1^{N*} + \cdots + S_d^N + S_d^{N*},$$

and the associated limiting model is

$$A_{\mathrm{F}} := s_1 + s_1^* + \cdots + s_d + s_d^*.$$

Note that A_{F} is nothing other than the adjacency matrix of the Cayley graph of $\mathbf{F}_d$ generated by $g_1, \ldots, g_d$ and their inverses, that is, of the $2d$-regular tree.

It is a classical fact due to Kesten [41] that $\|A_{\mathrm{F}}\| = 2\sqrt{2d-1}$. That

$$\|A^N\| \le 2\sqrt{2d-1} + o(1) \quad \text{with probability } 1 - o(1)$$

is due to Friedman [27]. Friedman's proof was simplified by Bordenave [5], and a new proof with an improved convergence rate was given by Huang and Yau [40]. Very recently, the latter approach was further developed in the impressive work of Huang, McKenzie and Yau [39] to achieve a nearly optimal convergence rate.

3.2.1. *An effective Friedman theorem.* As a first illustration of the methods of this paper, we will give a short new proof of Friedman's theorem in section 8.1. A direct implementation of the approach described in section 2 yields the following.

Theorem 3.4. *For every $d \ge 2$, $N \ge 1$, and $\varepsilon < 2d - 2\sqrt{2d-1}$, we have*

$$\mathbf{P}\big[\|A^N\| \ge 2\sqrt{2d-1} + \varepsilon\big] \lesssim \frac{1}{N}\left(\frac{d\log d}{\varepsilon}\right)^8 \log\left(\frac{2ed}{\varepsilon}\right).$$

Theorem 3.4 implies that when d is fixed as $N \to \infty$, we have[1]

$$\|A^N\| \le 2\sqrt{2d-1} + O_{\mathrm{P}}\big(\big(\tfrac{\log N}{N}\big)^{1/8}\big).$$

This rate falls short of the near-optimal $N^{-2/3+o(1)}$ rate in Friedman's theorem that was very recently established in [39]. However, the methods of [40, 39] rely heavily on the structure of random regular graphs, and it is unclear at present whether they could be applied to the study of strong convergence. In contrast, the methods of this paper will achieve the same rate for arbitrary polynomials of random permutation matrices, see section 3.3 below.

The quantitative nature of Theorem 3.4 also enables us to consider what happens when $d, N \to \infty$ simultaneously. It is an old question of Vu [63, §5] whether $\|A^N\| = (1+o(1))2\sqrt{2d-1}$ with probability $1-o(1)$ remains valid when $d, N \to \infty$ in an arbitrary manner. We can now settle this question for the permutation model of random regular graphs that is considered here (see Remark 3.8 below for a brief discussion of other models of random regular graphs).

Corollary 3.5. $\|A^N\| = (1 + o(1))2\sqrt{2d-1}$ *with probability $1 - o(1)$ whenever $N \to \infty$ and $d = d(N)$ depends on N in an arbitrary manner.*

Proof. That the conclusion holds for $d \ge (\log N)^5$ was proved in [12, §3.1.2]. In the complementary regime $d \le (\log N)^5$, Theorem 3.4 readily yields the upper bound $\|A^N\| \le (1+o(1))2\sqrt{2d-1}$ with probability $1-o(1)$, while the corresponding lower bound follows from the Alon–Boppana theorem [53]. $\qquad\square$

Let us emphasize that the presence of tangles, which form a fundamental obstruction to the moment method (cf. section 2.1), is completely ignored by the proof of Theorem 3.4. This seems unexpected, as the method is ultimately based on an estimate (2.5) for traces of high degree polynomials. However, the effect of tangles

---

[1]The notation $Z_N = O_{\mathrm{P}}(z_N)$ denotes that $\{Z_N/z_N\}_{N \ge 1}$ is bounded in probability.
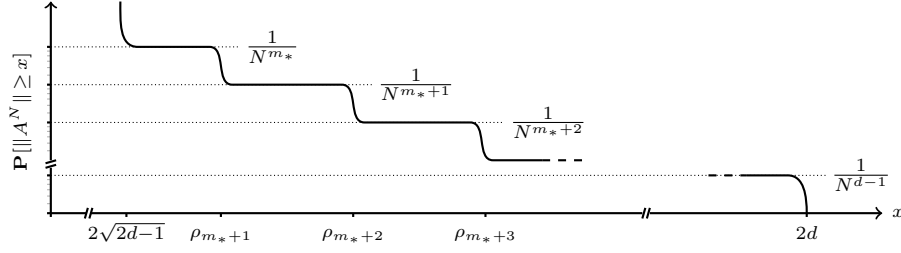
FIGURE 3.1. Staircase pattern of the tail probabilities of $\|A^N\|$.

on individual moments cancels out when they are combined to form bounded test functions h. One of the key features of the polynomial method is that it can capture these cancellations. (Note that (2.5) yields no useful information on moments of order $p \gg \log N$, as $\|h\|_{C^0[-K,K]} = K^p$ is exponential in p for $h(x) = x^p$.)

3.2.2. *The staircase theorem.* As was explained in section 2, the approach of this paper only requires an understanding of the first-order term ν_1 in the $\frac{1}{N}$-expansion of the moments. However, in the setting of random regular graphs, a detailed understanding of the higher-order terms was achieved by Puder [58] using methods of combinatorial group theory. When such additional information is available, the approach of this paper is readily adapted to achieve stronger results.

The following theorem will be proved in section 8.2 by taking full advantage of the results of [58]. We emphasize that this is the only part of this paper where we will use asymptotics of expected traces beyond the lowest order.

Theorem 3.6. *Define*

$$\rho_m := \begin{cases} 2\sqrt{2d-1} & \text{for } 2m-1 \le \sqrt{2d-1}, \\ 2m-1+\frac{2d-1}{2m-1} & \text{for } 2m-1 > \sqrt{2d-1}, \end{cases}$$

and let $m_* := \lfloor \frac{1}{2}(\sqrt{2d-1}+1)\rfloor$ be the largest integer m so that $2m-1 \le \sqrt{2d-1}$. Then for every $d \ge 2$, $m_* \le m \le d-1$, and $0 < \varepsilon < \rho_{m+1} - \rho_m$, we have

$$\mathbf{P}\big[\|A^N\| \ge \rho_m + \varepsilon\big] \le \frac{C_d}{N^m}\frac{1}{\varepsilon^{4(m+1)}}\log\left(\frac{2e}{\varepsilon}\right)$$

for all $N \ge 1$, and

$$\mathbf{P}\big[\|A^N\| \ge \rho_m + \varepsilon\big] \ge \frac{1-o(1)}{N^m}$$

as $N \to \infty$. Here C_d is a constant that depends on d only.

Theorem 3.6 reveals an unusual staircase pattern of the large deviations of $\|A^N\|$, which is illustrated in Figure 3.1. The lower bound, which follows from an elementary argument of Friedman [27, Theorem 2.11], arises due to the presence of tangles: Friedman shows that the presence of a vertex with $m > m_*$ self-loops, which happens with probability $\approx N^{1-m}$, gives rise to an outlier in the spectrum at location $\approx \rho_m$. The fact that our upper bound precisely matches this behavior shows that the tail probabilities of $\|A^N\|$ are completely dominated by these events.

Remark 3.7. Let us highlight two further consequences of Theorem 3.6.

1. Theorem 3.6 shows that $\|A^N\| \leq 2\sqrt{2d-1} + \varepsilon$ with probability at least $1 - \frac{c}{N^{m_*}}$ for any $\varepsilon > 0$. This closes the gap between the upper and lower bounds in the main result of [27], settling an open problem of Friedman [27, p. 2].

2. The lower bound of Theorem 3.6 shows that the $O(\frac{1}{N})$ probability bound of Theorem 3.4 cannot be improved in general for fixed $\varepsilon > 0$, as this bound is sharp when $m_* = 1$ (that is, when $2 \leq d \leq 4$).

Remark 3.8. The random regular graph model considered in this paper is known as the *permutation model*. Several other models of random regular graphs are also considered in the literature. All these models are known to be contiguous to the permutation model as $N \to \infty$ for fixed degree $2d$, so that any statement that holds with probability $1 - o(1)$ for the permutation model remains valid with probability $1 - o(1)$ for the other models; cf. [27, pp. 2–3] and the references therein.

It should be emphasized, however, that low probability events are not preserved by contiguity. In particular, the detailed quantitative picture in Theorem 3.6 is specific to the permutation model. The corresponding picture for other models of random regular graphs must be investigated on a case by case basis.

Similarly, the different models of random regular graphs are no longer contiguous when the degree diverges $d \to \infty$, so that high degree results as in Corollary 3.5 must be investigated separately for each model. Partial results on this problem for the uniform model of random regular graphs may be found in [3, 60, 35].

3.3. **Strong convergence of random permutation matrices.** The adjacency matrix of a random regular graph is one very special example of a polynomial of random permutation matrices. The much more general fact that the norm of *every* noncommutative polynomial of random permutation matrices converges to that of its limiting model is an important result due to Bordenave and Collins [7, 8]. Here we give a short proof that yields an effective form of this result.

A noncommutative polynomial $P \in \mathbb{C}\langle \boldsymbol{s}, \boldsymbol{s}^* \rangle$ of the free unitaries $\boldsymbol{s} = (s_1, \ldots, s_d)$ is naturally identified with an element of the group algebra $\mathbb{C}\mathbf{F}_d$ of the free group. We let $\|P\|_{C^*(\mathbf{F}_d)}$ be the norm of P in the full C^*-algebra of the free group; thus

$$\|P\|_{C^*(\mathbf{F}_d)} = \sup_{\boldsymbol{U}} \|P(\boldsymbol{U}, \boldsymbol{U}^*)\|,$$

where the supremum is taken over all d-tuples of unitary matrices [17, Theorem 7]. In practice, $\|P\|_{C^*(\mathbf{F}_d)}$ may be bounded trivially by the sum of the moduli of the coefficients of P, but we retain the general definition for clarity.

We can now formulate our main result on strong convergence of random permutation matrices. We state the result for polynomials $P \in \mathrm{M}_D(\mathbb{C}) \otimes \mathbb{C}\langle \boldsymbol{s}, \boldsymbol{s}^* \rangle$ with matrix (rather than scalar) coefficients, as the latter arise often in applications; their norm $\|P\|_{\mathrm{M}_D(\mathbb{C}) \otimes C^*(\mathbf{F}_d)}$ is defined as for scalar polynomials.

Theorem 3.9. *Let $d \geq 2$, and let $P \in \mathrm{M}_D(\mathbb{C}) \otimes \mathbb{C}\langle \boldsymbol{s}, \boldsymbol{s}^* \rangle$ be any self-adjoint noncommutative polynomial of degree q_0. Then we have*

$$\mathbf{P}\left[\|P(\boldsymbol{S}^N, \boldsymbol{S}^{N*})\| \geq \|P(\boldsymbol{s}, \boldsymbol{s}^*)\| + \varepsilon \right] \lesssim \frac{D}{N} \left(\frac{K q_0 \log d}{\varepsilon} \right)^8 \log \left(\frac{eK}{\varepsilon} \right)$$

for all $\varepsilon < K - \|P(\boldsymbol{s}, \boldsymbol{s}^*)\|$, where $K = \|P\|_{\mathrm{M}_D(\mathbb{C}) \otimes C^*(\mathbf{F}_d)}$.

This theorem will be proved in section 9. Note that the limiting norm $\|P(\boldsymbol{s}, \boldsymbol{s}^*)\|$ can in principle be computed explicitly using the results of [42].

Remark 3.10. The assumption that P is self-adjoint is made largely for convenience. To obtain strong convergence of a non-self-adjoint polynomial P of degree q_0, we may simply apply this result to the self-adjoint polynomial P^*P of degree $2q_0$.

Theorem 3.9 shows that $\|P(\boldsymbol{S}^N, \boldsymbol{S}^{N*})\| \leq \|P(\boldsymbol{s}, \boldsymbol{s}^*)\| + O_{\mathrm{P}}\left(\left(\frac{\log N}{N}\right)^{1/8}\right)$ for any polynomial P. This significantly improves the best known bound to date, due to Bordenave and Collins [8, Theorem 1.4], which yields fluctuations of order $\frac{\log \log N}{\log N}$. Our bound can be directly substituted in applications, such as to random lifts of graphs [7, §1.5], to improve the best known convergence rates.

Let us note that for fixed $\varepsilon > 0$, the tail probability of order $\frac{1}{N}$ in Theorem 3.9 cannot be improved in general, as is illustrated by Theorem 3.6 with $d = 2$.

Remark 3.11. While Theorem 3.9 yields much stronger quantitative bounds for fixed D then prior results, it can only be applied when $D = o(N)$. In contrast, it was recently shown in [8, Corollary 1.5] that strong convergence remains valid in the present setting even for D as large as $N^{(\log N)^{1/2}}$. Such bounds could be achieved using our methods if the supports of the higher-order terms in the $\frac{1}{N}$-expansion of the moments are still bounded by $\|P(\boldsymbol{s}, \boldsymbol{s}^*)\|$. While this is the case for continuous models such as GUE [56], it is simply false in the present setting: the proof of Theorem 3.6 shows that tangles can already arise in the second-order term. It is an interesting question whether the approach of this paper can be combined with conditioning on the absence of tangles to achieve improved bounds.

3.4. **Stable representations of the symmetric group.** Let $\pi_N : \mathbf{S}_N \to V_N$ be a finite-dimensional unitary representation of the symmetric group $\mathbf{S}_N$. Then we can define random matrices $\Pi_1^N, \ldots, \Pi_d^N$ of dimension $\dim V_N$ as

$$\Pi_i^N := \pi_N(\sigma_i), \tag{3.1}$$

where $\sigma_1, \ldots, \sigma_d$ are i.i.d. uniformly distributed elements of $\mathbf{S}_N$. When π_N is the standard representation, we recover the random permutation matrices $\Pi_i^N = S_i^N$ as a special case. Other representations, however, capture a much larger family of random matrix models. We will prove strong convergence for random matrices defined by any *stable* representation of $\mathbf{S}_N$ (see section 3.4.1), which yields a far-reaching generalization of Theorem 3.9. This result is of interest for several reasons:

- It provides many new examples of the strong convergence phenomenon.

- It shows that strong convergence can be achieved with much less randomness than in the standard representation: Π_i^N uses the same number of random bits as S_i^N, but has much larger dimension ($\dim V_N \asymp N^\alpha$ with α arbitrarily large). See [9] for analogous questions in the context of the unitary group.

- It supports long-standing questions on the expansion of random Cayley graphs of the symmetric group, for which extensive numerical evidence and conjectures are available; see [59] and the references therein.

- Random matrices defined by representations other than the standard representation arise in various applications [28, 32].

Remark 3.12. The question whether strong asymptotic freeness can be derandomized has been investigated in the theoretical computer science literature [49, 54] by means of pseudorandom permutation matrices. The results of the present section

suggest that high-dimensional representations of $\mathbf{S}_N$ can provide a different per-
spective on such questions. We omit a detailed discussion of the number of random
bits needed by our bounds as significantly better quantitative results are expected
to arise from additional group-theoretic input, see Remark 3.16 below.

3.4.1. *Stable representations.* The approach of this paper requires that the moments
of the random matrices of interest are rational functions of $\frac{1}{N}$. For this to be the
case, we cannot choose an arbitrary representation π_N of $\mathbf{S}_N$ for each N. Instead,
we will work with stable representations [25, 18] that are defined consistently for
different N. We briefly introduce the relevant notions here.

Following [31], denote by $\xi_i(\sigma)$ the number of fixed points of σ^i for $\sigma \in \mathbf{S}_N$. The
sequence $\xi_1(\sigma), \xi_2(\sigma), \ldots$ determines the conjugacy class of σ. If $\pi_N : \mathbf{S}_N \to V_N$ is
any finite-dimensional unitary representation, its character $\sigma \mapsto \operatorname{Tr} \pi_N(\sigma)$ is a class
function and can therefore be expressed as a polynomial of $\xi_1(\sigma), \xi_2(\sigma), \ldots$

Definition 3.13. A finite-dimensional unitary representation $\pi_N : \mathbf{S}_N \to V_N$ of
$\mathbf{S}_N$, defined for each $N \geq N_0$, is *stable* if there exists $r \in \mathbb{N}$ and a polynomial
$\varphi \in \mathbb{C}[x_1, \ldots, x_r]$ so that $\operatorname{Tr} \pi_N(\sigma) = \varphi(\xi_1(\sigma), \ldots, \xi_r(\sigma))$ for all N and $\sigma \in \mathbf{S}_N$.

Thus stable representations are representations of $\mathbf{S}_N$ whose characters are de-
fined by the same polynomial φ for all $N \geq N_0$. For example, the standard repre-
sentation is stable as it satisfies $\operatorname{Tr} \pi_N = \xi_1 - 1$ for all N.

The irreducible stable representations of $\mathbf{S}_N$ can be constructed explicitly as
follows. Fix a base partition $\lambda = (\lambda_1 \geq \cdots \geq \lambda_\ell > 0) \vdash |\lambda|$ of $|\lambda| = \sum_{i=1}^\ell \lambda_i$. Then
for every $N \geq |\lambda| + \lambda_1$, the irreducible representation of $\mathbf{S}_N$ defined by

$$\lambda[N] = (N - |\lambda| \geq \lambda_1 \geq \cdots \geq \lambda_\ell) \vdash N \tag{3.2}$$

is stable. Moreover, it follows from [31, Proposition B.2] that every stable repre-
sentation in the sense of Definition 3.13 is a direct sum of such irreducible repre-
sentations defined by fixed base partitions $\lambda^{(1)}, \ldots, \lambda^{(s)}$.

3.4.2. *Strong convergence.* Fix a stable representation $\pi_N : \mathbf{S}_N \to V_N$ defined for
$N \geq N_0$ by a character polynomial $\varphi \in \mathbb{C}[x_1, \ldots, x_r]$. We aim to prove strong
convergence of the random matrices $\Pi_1^N, \ldots, \Pi_d^N$ defined by (3.1).

We will not require that π_N is irreducible, but we assume it does not contain
the trivial representation. The dimension of Π_i^N is given by

$$D_N := \dim(V_N) = \operatorname{Tr} \pi_N(e) = \varphi(N, N, \ldots, N).$$

Thus D_N is a polynomial in N; we denote its degree by α, so that $D_N \asymp N^\alpha$.

We now formulate our main result on strong convergence of $\mathbf{\Pi}^N = (\Pi_1^N, \ldots, \Pi_d^N)$,
whose proof is contained in section 10 below.

Theorem 3.14. *Let $d \geq 2$, and let $P \in \mathrm{M}_D(\mathbb{C}) \otimes \mathbb{C}\langle \boldsymbol{s}, \boldsymbol{s}^* \rangle$ be any self-adjoint
noncommutative polynomial of degree q_0. Then we have*

$$\mathbf{P}\big[\|P(\mathbf{\Pi}^N, \mathbf{\Pi}^{N*})\| \geq \|P(\boldsymbol{s}, \boldsymbol{s}^*)\| + \varepsilon\big] \leq \frac{CD}{N}\left(\frac{Kq_0 \log d}{\varepsilon}\right)^{4(\alpha+1)} \log\left(\frac{eK}{\varepsilon}\right)$$

*for all $\varepsilon < K - \|P(\boldsymbol{s}, \boldsymbol{s}^*)\|$ and $N \geq N_0$. Here we define $K = \|P\|_{\mathrm{M}_D(\mathbb{C}) \otimes C^*(\mathbf{F}_d)}$,
and C is a constant that depends on the choice of stable representation.*

Remark 3.15. It is certainly possible to obtain an explicit expression for C from the proof; we have suppressed the dependence on the choice of stable representation for simplicity of presentation, and we did not optimize this dependence in the proof.

Remark 3.16. The bound $\|P(\mathbf{\Pi}^N, \mathbf{\Pi}^{N*})\| \le \|P(\boldsymbol{s}, \boldsymbol{s}^*)\| + O_{\mathrm{P}}\big(\big(\frac{\log N}{N}\big)^{1/4(\alpha+1)}\big)$ that follows from Theorem 3.14 becomes weaker when we consider stable representations of increasingly large dimension $D_N \asymp N^\alpha$. This is not a restriction of our method, however, but rather reflects a gap in the understanding of stable representations. In particular, [31, Conjecture 1.8] would be expected to yield an improved probability bound where $\frac{1}{N^{\varepsilon^4(\alpha+1)}}$ is replaced by $\frac{1}{N^{\alpha}\varepsilon^{8\alpha}}$, resulting in a convergence rate that is independent of α. Such an improved probability bound would also enable us to establish strong convergence in situations where α is allowed to diverge rapidly with N. New developments in this direction will appear in [13].

Theorem 3.14 can be readily applied to concrete situations. For example, it implies that random $2d$-regular Schreier graphs defined by the action of $\mathbf{S}_N$ on k-tuples of distinct elements of $\{1, \ldots, N\}$ have second eigenvalue $2\sqrt{2d-1} + o(1)$ with probability $1 - o(1)$, settling a question discussed in [31, §1.4]. Indeed, it is not difficult to see (cf. [31, §8]) that the restricted adjacency matrix $A^N = \bar{A}^N|_{\{1^\perp\}}$ of such a random graph can be represented as

$$A^N = \Pi_1^N + \Pi_1^{N*} + \cdots \Pi_d^N + \Pi_d^{N*}$$

for some stable representation of $\mathbf{S}_N$ (which depends on k) that does not contain the trivial representation, so that the conclusion follows immediately as a special case of Theorem 3.14. Applications of this model may be found in [28, 32]. (Let us note for completeness that the case $k = 1$ reduces to Friedman's theorem, while the case $k = 2$ was previously studied by Bordenave and Collins [7, §1.6].)

4. Basic tools

The aim of this section is to introduce the general facts on polynomials and compactly supported distributions that form the basis for the methods of this paper. While most of the tools in this section follow readily from known results, it is useful to state them in the particular form in which they will be needed.

4.1. **Markov inequalities.**
One of the key tools that will be used in our analysis is the following classical inequality of A. Markov and V. Markov. Here we recall that $(2k-1)!! := (2k-1)(2k-3)\cdots 5 \cdot 3 \cdot 1 = \frac{(2k)!}{2^k k!}$.

Lemma 4.1 (Markov inequality)**.** *Let $h \in \mathcal{P}_q$ and $a > 0$, $m \in \mathbb{N}$. Then we have*

$$\|h^{(m)}\|_{C^0[0,a]} \le \frac{1}{(2m-1)!!}\left(\frac{2q^2}{a}\right)^m \|h\|_{C^0[0,a]}.$$

Proof. Apply [10, p. 256(d)] to $P(x) = h(\frac{a}{2}x + \frac{a}{2})$. $\qquad\square$

Two basic issues will arise in our applications of this inequality. First, we will not be able to control the relevant functions on the entire interval $[0, a]$, but only on a discrete subset thereof. This issue will be addressed using a standard interpolation inequality that is itself a direct consequence of the Markov inequality.

Lemma 4.2 (Interpolation)**.** *Let $h \in \mathcal{P}_q$, and fix a subset $I \subseteq [0,a]$ such that $\delta := \sup_{x \in [0,a]} \inf_{y \in I} |x - y|$ satisfies $4q^2\delta \leq a$. Then*

$$\|h\|_{C^0[0,a]} \leq 2 \sup_{x \in I} |h(x)|.$$

Proof. Apply [16, Lemma 3(i), p. 91] to $P(x) = h(\frac{a}{2}x + \frac{a}{2})$. $\qquad\square$

The second issue is that we will aim to apply these inequalities to rational functions rather than to polynomials. However, in the cases of interest to us, the denominator of the rational function will be nearly constant. The following lemma extends the Markov inequality to this setting.

Lemma 4.3 (Rational Markov)**.** *Let $r = \frac{f}{g}$ be a rational function with $f, g \in \mathcal{P}_q$, and let $a > 0$ and $m \in \mathbb{N}$. Suppose that $c := \sup_{x,y \in [0,a]} \left| \frac{g(x)}{g(y)} \right| < \infty$. Then*

$$\|r^{(m)}\|_{C^0[0,a]} \leq m! \left(\frac{5cq^2}{a} \right)^m \|r\|_{C^0[0,a]}.$$

Proof. Applying the product rule to $f = rg$ yields

$$r^{(m)}g = f^{(m)} - \sum_{k=1}^{m} \binom{m}{k} r^{(m-k)} g^{(k)}.$$

As $\frac{1}{c} \| \frac{h}{g} \|_{C^0[0,a]} \leq \frac{\|h\|_{C^0[0,a]}}{\|g\|_{C^0[0,a]}} \leq \| \frac{h}{g} \|_{C^0[0,a]}$ for every function h, we can estimate

$$\|r^{(m)}\|_{C^0[0,a]} \leq \left\| \frac{f^{(m)}}{g} \right\|_{C^0[0,a]} + \sum_{k=1}^{m} \binom{m}{k} \|r^{(m-k)}\|_{C^0[0,a]} \left\| \frac{g^{(k)}}{g} \right\|_{C^0[0,a]}$$

$$\leq 2c \sum_{k=1}^{m} \binom{m}{k} \frac{1}{(2k-1)!!} \left(\frac{2q^2}{a} \right)^k \|r^{(m-k)}\|_{C^0[0,a]}$$

by applying Lemma 4.1 to $f^{(m)}$ and $g^{(k)}$. Here the factor 2 in the second line is due to the fact that the $f^{(m)}$ and $k = m$ terms in the first line yield the same bound.

We now reason by induction. Clearly the conclusion holds $m = 0$. Now suppose the conclusion holds up to $m - 1$. Then the above inequality yields

$$\|r^{(m)}\|_{C^0[0,a]} \leq \left(\frac{5cq^2}{a} \right)^m \|r\|_{C^0[0,a]} \cdot 2c \sum_{k=1}^{m} \binom{m}{k} \frac{(m-k)!}{(2k-1)!!} \left(\frac{2}{5c} \right)^k$$

$$\leq m! \left(\frac{5cq^2}{a} \right)^m \|r\|_{C^0[0,a]} \cdot 2c \left(\cosh\left(\sqrt{\frac{4}{5c}} \right) - 1 \right)$$

as $\binom{m}{k} \frac{(m-k)!}{(2k-1)!!} = \frac{2^k m!}{(2k)!}$. The result follows as $\cosh(\sqrt{x}) - 1 \leq \frac{5x}{8}$ for $x \in [0,1]$. $\quad\square$

The above bounds cannot be essentially improved without further assumptions. Lemma 4.1 attains equality for Chebyshev polynomials [10, p. 256(d)]. The optimality of Lemma 4.2 is discussed in [22], while the optimality of Lemma 4.3 is illustrated by considering $r(x) = \frac{1}{u-x}$ on $x \in [0,1]$ for $u > 1$.

4.2. Chebyshev polynomials. Let $h \in \mathcal{P}_q$ and fix $K > 0$. In this section, we will consider the behavior of h on the interval $[-K, K]$.

Denote by T_j the Chebyshev polynomial of degree j, defined by the relation $T_j(\cos\theta) = \cos(j\theta)$. Then h can be uniquely expressed as

$$h(x) = \sum_{j=0}^{q} a_j T_j(K^{-1}x) \tag{4.1}$$

for some real coefficients $a_0, \ldots, a_q$. Note that these coefficients are precisely the Fourier coefficients of the cosine series $h(K\cos\theta)$. We can therefore apply a classical result of Zygmund on absolute convergence of trigonometric series.

Lemma 4.4 (Zygmund). *Let h be as in* (4.1) *and define $f(\theta) := h(K\cos\theta)$. Then*

$$|a_0| \le \|h\|_{C^0[-K,K]},$$

and

$$\sum_{j=1}^{q} j^m |a_j| \lesssim \beta_* \|f^{(m+1)}\|_{L^\beta[0,2\pi]}$$

for every $m \in \mathbb{Z}_+$ and $\beta > 1$, where we defined $1/\beta_* := 1 - 1/\beta$.

Proof. That $|a_0| \le \|h\|_{C^0[-K,K]}$ follows immediately from $a_0 = \frac{1}{2\pi}\int_0^{2\pi} f(\theta)\, d\theta$. To obtain the second estimate, we note that

$$f^{(m)}(\theta) = \begin{cases} \sum_{j=0}^{q}(-1)^{m/2} j^m a_j \cos(j\theta) & \text{for even } m, \\ \sum_{j=0}^{q}(-1)^{(m+1)/2} j^m a_j \sin(j\theta) & \text{for odd } m. \end{cases}$$

The conclusion follows from [64, p. 242]. $\qquad\square$

A direct consequence is the following.

Corollary 4.5. *Let h be as in* (4.1) *and let $m \in \mathbb{Z}_+$. Then*

$$\sum_{j=0}^{q} j^m |a_j| \le c_{m,K}\, \|h\|_{C^{m+1}[-K,K]},$$

where the constant $c_{m,K}$ depends on m, K only.

Proof. That $f(\theta) = h(K\cos\theta)$ satisfies $\|f^{(m+1)}\|_{L^\infty[0,2\pi]} \le c_{m,K}\|h\|_{C^{m+1}[-K,K]}$ follows by the chain rule. It remains to apply Lemma 4.4 with $\beta = \infty$. $\qquad\square$

4.3. **Compactly supported distributions.** In this paper we will encounter only distributions on $\mathbb{R}$. We therefore adopt the following definition [38, §2.3].

Definition 4.6. A linear functional ν on $C^\infty(\mathbb{R})$ such that

$$|\nu(f)| \le c\|f\|_{C^m[-K,K]} \quad \text{for all } f \in C^\infty(\mathbb{R})$$

holds for some $c, K > 0$ and $m \in \mathbb{Z}_+$, is called a *compactly supported distribution*.

The linear functionals that will arise in this paper will not be defined *a priori* on $C^\infty(\mathbb{R})$, but rather only on the space $\mathcal{P}$ of univariate polynomials. It will be therefore essential to understand when linear functionals on $\mathcal{P}$ can be extended to compactly supported distributions. As we have not located the following result in the literature, we prove it here for completeness.

Lemma 4.7 (Hausdorff moment problem for compactly supported distributions). *Let ν be a linear functional on $\mathcal{P}$. Then the following are equivalent.*

1. *There exist $c, m, K \geq 0$ so that $|\nu(h)| \leq cq^m \|h\|_{C^0[-K,K]}$ for all $h \in \mathcal{P}_q$, $q \in \mathbb{N}$.*
2. *There exist $c, m, K \geq 0$ so that $|\nu(T_j(K^{-1}\cdot))| \leq cj^m$ for all $j \in \mathbb{N}$.*
3. *ν extends uniquely to a compactly supported distribution.*

Only the implication $1 \Rightarrow 3$ will be used in this paper; the equivalence of $2 \Leftrightarrow 3$ appears implicitly in [24, p. 38].

Proof. That $1 \Rightarrow 2$ is immediate as $\|T_j(K^{-1}\cdot)\|_{C^0[-K,K]} = 1$ by the definition of the Chebyshev polynomials T_j.

To prove that $2 \Rightarrow 3$, fix any $h \in \mathcal{P}$ and express it in the form (4.1). Then

$$|\nu(h)| \leq \sum_{j=0}^{q} |\nu(T_j(K^{-1}\cdot))| \, |a_j| \leq c \sum_{j=0}^{q} j^m |a_j| \leq c_{m,K} \|h\|_{C^{m+1}[-K,K]}$$

by Corollary 4.5. Thus the condition of Definition 4.6 is satisfied for all $h \in \mathcal{P}$. As $\mathcal{P}$ is dense in $C^\infty(\mathbb{R})$ with respect to the norm $\| \cdot \|_{C^{m+1}[-K,K]}$, it is clear that ν extends uniquely to a compactly supported distribution.

To prove that $3 \Rightarrow 1$, it suffices by Definition 4.6 to note that $\|h\|_{C^m[-K,K]} \leq c_{m,K} q^{2m} \|h\|_{C^0[-K,K]}$ for every $h \in \mathcal{P}_q$ and $q \in \mathbb{N}$ by Lemma 4.1. $\square$

Next, we recall the definition of support [38, §2.2].

Definition 4.8. The *support* $\operatorname{supp} \nu$ of a compactly supported distribution ν is the smallest closed set $A \subset \mathbb{R}$ so that $\nu(f) = 0$ for all $f \in C^\infty(\mathbb{R})$ with $f = 0$ on A.

It is clear that when the condition of Definition 4.6 is satisfied, we must have $\operatorname{supp} \nu \subseteq [-K, K]$. However, the support may in fact be much smaller. A key property of compactly supported distributions for our purposes is that their support can be bounded by the growth rate of their moments. While the analogous property of measures is straightforward, its validity for distributions requires justification.

Lemma 4.9. *Let ν be a compactly supported distribution. Then*

$$\operatorname{supp} \nu \subseteq [-\rho, \rho] \qquad with \qquad \rho = \limsup_{p \to \infty} |\nu(x^p)|^{1/p}.$$

Proof. It is clear from the definition of a compactly supported distribution that we must have $\rho < \infty$. Thus the function F defined by

$$F(z) := \sum_{p=0}^{\infty} \frac{\nu(x^p)}{z^{p+1}}$$

is analytic for $z \in \mathbb{C}$ with $|z|$ sufficiently large. It follows from [24, Lemma 1] (or from [61, Theorem 5.4]) that F can be analytically continued to $\mathbb{C} \backslash \operatorname{supp} \nu$ and that $\operatorname{supp} \nu$ is precisely the set of singular points of F. As the above expansion of F is absolutely convergent in a neighborhood of $z = \pm(\rho + \varepsilon)$ for all $\varepsilon > 0$, it follows that $\pm(\rho + \varepsilon) \notin \operatorname{supp} \nu$ for all $\varepsilon > 0$, which yields the conclusion. $\square$

4.4. **Test functions.** We finally recall a standard construction of smooth approximations of the indicator function $1_{[-\rho,\rho]^c}$ for which the bound of Lemma 4.4 is well controlled; its sharpness is discussed in [38, pp. 19–22]. Recall that $1/\beta_* = 1 - 1/\beta$.

Lemma 4.10 (Test functions)**.** *Fix $m \in \mathbb{Z}_+$ and $K, \rho, \varepsilon > 0$ so that $\rho + \varepsilon < K$. Then there exists a function $\chi \in C^{m+1}[-K, K]$ with the following properties.*

1. $\chi(x) \in [0,1]$ *for all* x, $\chi(x) = 0$ *for* $|x| \le \rho$, *and* $\chi(x) = 1$ *for* $|x| \ge \rho + \varepsilon$.

2. $\|f^{(m+1)}\|_{L^\beta[0,2\pi]} \le 4^{m+2} m^m (\frac{K}{\varepsilon})^{m+1/\beta_*}$ *for all* $\beta > 1$, *where* $f(\theta) := \chi(K\cos\theta)$.

Proof. Fix $\delta, \varphi > 0$ to be chosen below, and define the functions $H_a := \frac{1}{a} 1_{[0,a]}$ and

$$h(x) = \int_{-\infty}^x (H_{\delta/2} * \underbrace{H_{\delta/2m} * \cdots * H_{\delta/2m}}_{m \text{ times}})(y) \, dy.$$

The integrand is nonnegative, has $L^1(\mathbb{R})$-norm one, and is supported on $[0, \delta]$. Thus h is nondecreasing, $h(x) = 0$ for $x \le 0$, and $h(x) = 1$ for $x \ge \delta$. We define

$$\chi(x) = h\big(\arcsin\big(\tfrac{x}{K}\big) - \varphi\big) + h\big(-\arcsin\big(\tfrac{x}{K}\big) - \varphi\big)$$

so that f takes the simple form $f(\theta) = h\big(\frac{\pi}{2} - \theta - \varphi\big) + h\big(\theta - \frac{\pi}{2} - \varphi\big)$ for $\theta \in [0, \pi]$. Then we can choose the parameters $\varphi = \arcsin(\frac{\rho}{K})$ and $\delta = \arcsin(\frac{\rho+\varepsilon}{K}) - \arcsin(\frac{\rho}{K})$ so that 1. holds. Moreover, $h \in C^{m+1}(\mathbb{R})$ is constant near $\pm\frac{\pi}{2} - \varphi$ as $\rho + \varepsilon < K$, so that $\chi \in C^{m+1}[-K, K]$ by the chain rule. Finally

$$\|f^{(m+1)}\|_{L^\beta[0,2\pi]} \le 4\|H_{\delta/2} * H'_{\delta/2m} * \cdots * H'_{\delta/2m}\|_{L^\beta(\mathbb{R})}$$
$$\le 4\|H_{\delta/2}\|_{L^\beta(\mathbb{R})} \|H'_{\delta/2m}\|_{L^1(\mathbb{R})}^m \le 4^{m+2} m^m \big(\tfrac{1}{\delta}\big)^{m+1/\beta_*},$$

where the factor 4 in the first line arises by the triangle inequality and by splitting the L^β norm over the intervals $[0, \pi]$ and $[\pi, 2\pi]$, and the second line uses Young's inequality and $H'_a = \frac{1}{a}(\delta_0 - \delta_a)$. Then 2. follows by noting that $\delta \ge \frac{\varepsilon}{K}$. $\qquad\square$

5. Words in random permutation matrices

The aim of this section is to recall basic facts about random permutations that will form the input to the methods of this paper. These results are implicit in Nica [51], but are derived in simpler and more explicit form by Linial and Puder [43] whose presentation we follow. We emphasize that all the results of this section are essentially elementary in nature, so that analogous results are likely accessible in many other situations (see [21, 19] for a general perspective).

We will work with random permutation matrices $\boldsymbol{S}^N = (S_1^N, \ldots, S_d^N)$ and their limiting model $\boldsymbol{s} = (s_1, \ldots, s_d)$ as defined in section 3.1. In particular, we recall that $s_i := \lambda(g_i)$, where $g_1, \ldots, g_d$ are the free generators of $\mathbf{F}_d$.

It will be convenient to extend these definitions by setting $g_0 := e$, $g_{d+i} := g_i^{-1}$, and analogously $S_0^N := \mathbf{1}$, $S_{d+i}^N := (S_i^N)^{-1} = S_i^{N*}$ and $s_0 := \mathbf{1}$, $s_{d+i} := s_i^{-1} = s_i^*$. This convention will be in force in the remainder of the paper.

We aim to understand the expected trace of words in random permutation matrices and their inverses. To formalize this notion, denote by $\mathbf{W}_d$ the set of all finite words in the letters $g_0, \ldots, g_{2d}$. We implicitly identify $w \in \mathbf{W}_d$ with the word map $w : \mathrm{G}^d \to \mathrm{G}$ it induces on any group G. Thus $w(g_1, \ldots, g_d) \in \mathbf{F}_d$ is the reduction of the word w, while $w(S_1^N, \ldots, S_d^N)$ is the random matrix obtained by substituting $g_i \leftarrow S_i^N$ and multiplying these matrices in the order they appear in w.

5.1. **Rationality.** The first property we will need is that the expected trace of any word is a rational function of $\frac{1}{N}$. We emphasize that only very limited information on this function will be needed, which is collected in the following lemma.

Lemma 5.1. *Let $w \in \mathbf{W}_d$ be any word of length at most q and $N \geq q$. Then*

$$\mathbf{E}[\mathrm{tr}_N w(\boldsymbol{S}^N)] = \frac{f_w(\frac{1}{N})}{g_q(\frac{1}{N})},$$

where

$$g_q(x) := (1-x)^{d_1}(1-2x)^{d_2}\cdots(1-(q-1)x)^{d_{q-1}}$$

with $d_j := \min\left(d, \lfloor \frac{q}{j+1} \rfloor\right)$, and f_w, g_q are polynomials of degree at most $q(1+\log d)$.

Proof. Observe that $\mathbf{E}[\mathrm{Tr}\, w(S_1^N, \ldots, S_d^N)]$ is the expected number of fixed points of $w(\bar{S}_1^N, \ldots, \bar{S}_d^N)$ minus one (as we restrict to $\{1_N\}^\perp$). Thus [43, eq. (7)] yields

$$\mathbf{E}[\mathrm{tr}_N w(\boldsymbol{S}^N)] = -\frac{1}{N} + \sum_\Gamma \frac{(\frac{1}{N})^{e_\Gamma - v_\Gamma + 1} \prod_{l=1}^{v_\Gamma - 1}(1 - \frac{l}{N})}{\prod_{j=1}^d \prod_{l=1}^{e_\Gamma^j - 1}(1 - \frac{l}{N})}$$

where the sum is over a certain collection of connected graphs Γ with $v_\Gamma \leq q$ vertices and $e_\Gamma \leq q$ edges, and $e_\Gamma^j \geq 0$ are integers so that $e_\Gamma^1 + \cdots + e_\Gamma^d = e_\Gamma$ [43, p. 105]. Note that $e_\Gamma - v_\Gamma + 1 \geq 0$ as Γ is connected.

The denominator inside the sum can be written equivalently as

$$\prod_{j=1}^d \prod_{l=1}^{e_\Gamma^j - 1}(1 - \tfrac{l}{N}) = (1 - \tfrac{1}{N})^{d_1^\Gamma}(1 - \tfrac{2}{N})^{d_2^\Gamma}\cdots(1 - \tfrac{q-1}{N})^{d_{q-1}^\Gamma},$$

with $d_i^\Gamma := |\{1 \leq j \leq d : e_\Gamma^j \geq i+1\}| \leq d_i$ as $e_\Gamma^1 + \cdots + e_\Gamma^d \leq q$. Thus

$$\mathbf{E}[\mathrm{tr}_N w(\boldsymbol{S}^N)] = \frac{-\frac{1}{N}g_q(\frac{1}{N}) + \sum_\Gamma (\frac{1}{N})^{e_\Gamma - v_\Gamma + 1} \prod_{l=1}^{v_\Gamma - 1}(1 - \frac{l}{N}) \prod_{i=1}^{q-1}(1 - \frac{i}{N})^{d_i - d_i^\Gamma}}{g_q(\frac{1}{N})}.$$

To conclude, note that $\sum_{i=1}^{q-1} d_i \leq \int_1^q \min(d, \frac{q}{x})dx \leq q(1 + \log d) - \min(d, q)$ and $e_\Gamma - \sum_{i=1}^{q-1} d_i^\Gamma = |\{1 \leq j \leq d : e_\Gamma^j \geq 1\}| \leq \min(d, q)$. Thus g_q has degree at most $q(1 + \log d) - 1$ and each term in the sum has degree at most $q(1 + \log d)$. $\square$

5.2. **First-order asymptotics.** We now recall the first-order asymptotic behavior of expected traces of words in random permutation matrices. The following is a special case of a result of Nica [51], see [43, p. 124] for a simple proof.

An element of the free group $v \in \mathbf{F}_d$, $v \neq e$ is called a *proper power* if it can be written as $v = w^k$ for some $w \in \mathbf{F}_d$ and $k \geq 2$, and is called a *non-power* otherwise. We denote by $\mathbf{F}_d^{\mathrm{np}}$ the set of non-powers. Every $v \in \mathbf{F}_d$, $v \neq e$ can be written uniquely as $v = w^k$ for some $w \in \mathbf{F}_d^{\mathrm{np}}$ and $k \geq 1$, cf. [45, Proposition I.2.17].

Lemma 5.2. *Fix a word $w \in \mathbf{W}_d$ that does not reduce to the identity, and express its reduction as $w(g_1, \ldots, g_d) = v^k$ with $v \in \mathbf{F}_d^{\mathrm{np}}$ and $k \geq 1$. Then*

$$\lim_{N\to\infty} N\,\mathbf{E}[\mathrm{tr}_N w(\boldsymbol{S}^N)] = \omega(k) - 1,$$

where $\omega(k)$ denotes the number of divisors of k.

Proof. This follows from [51, Corollary 1.3] by noting as in the proof of Lemma 5.1 that $\mathbf{E}[\mathrm{Tr}\, w(\boldsymbol{S}^N)]$ is the expected number of fixed points of $w(\bar{\boldsymbol{S}}^N)$ minus one. $\square$

In particular, Lemma 5.2 implies the following.

Corollary 5.3 (Weak convergence). *For every word $w \in \mathbf{W}_d$, we have*

$$\lim_{N\to\infty} \mathbf{E}[\mathrm{tr}_N w(\boldsymbol{S}^N)] = \tau\big(w(\boldsymbol{s})\big).$$

Proof. Lemma 5.2 implies that $\mathbf{E}[\operatorname{tr}_N w(\boldsymbol{S}^N)] = o(1)$ for any word w that does not reduce to the identity. On the other hand, if w reduces to the identity, then $w(\boldsymbol{S}^N)$ is the identity matrix on $\{1_N\}^\perp$ and thus $\mathbf{E}[\operatorname{tr}_N w(\boldsymbol{S}^N)] = 1 - \frac{1}{N}$. The conclusion follows as clearly $\tau(s_{i_1} \cdots s_{i_k}) = \langle \delta_e, \lambda(g_{i_1}) \cdots \lambda(g_{i_k}) \delta_e \rangle = 1_{g_{i_1} \cdots g_{i_k} = e}$. $\qquad\square$

6. Master inequalities for random permutations I

6.1. **Master inequalities.** In this section, we develop a core ingredient of our method: inequalities for the normalized trace of polynomials of $\boldsymbol{S}^N$.

Theorem 6.1 (Master inequalities)**.** *Fix a self-adjoint noncommutative polynomial $P \in \mathrm{M}_D(\mathbb{C}) \otimes \mathbb{C}\langle \boldsymbol{s}, \boldsymbol{s}^* \rangle$ of degree q_0, and let $K = \|P\|_{\mathrm{M}_D(\mathbb{C}) \otimes C^*(\mathbf{F}_d)}$. Then there exists a linear functional ν_i on $\mathcal{P}$ for every $i \in \mathbb{Z}_+$ so that*

$$\left| \mathbf{E}[\operatorname{tr}_{DN} h(P(\boldsymbol{S}^N, \boldsymbol{S}^{N*}))] - \sum_{i=0}^{m-1} \frac{\nu_i(h)}{N^i} \right| \leq \frac{(4qq_0(1 + \log d))^{4m}}{N^m} \|h\|_{C^0[-K,K]}$$

for every $N, m, q \in \mathbb{N}$ and $h \in \mathcal{P}_q$.

An immediate consequence of Theorem 6.1 is the following corollary that we spell out separately for future reference.

Corollary 6.2. *In the setting of Theorem 6.1, we have*

$$|\nu_m(h)| \leq (4qq_0(1 + \log d))^{4m} \|h\|_{C^0[-K,K]}$$

for every $m \in \mathbb{Z}_+$, $q \in \mathbb{N}$, and $h \in \mathcal{P}_q$.

While we will need Theorem 6.1 in its full generality in some applications, many strong convergence results will already arise from the case $m = 2$ as was explained in section 2. The case $m = 1$ is of independent interest as it provides a quantitative form of Corollary 5.3; see, for example, Corollary 7.2 below.

6.2. **Proof of Theorem 6.1 and Corollary 6.2.** Fix a polynomial P as in Theorem 6.1 and $h \in \mathcal{P}_q$. Then $(\operatorname{tr}_D \otimes \mathrm{id})(h(P(\boldsymbol{S}^N, \boldsymbol{S}^{N*})))$ is a linear combination of words $w(\boldsymbol{S}^N)$ of length at most qq_0. Thus Lemma 5.1 yields

$$\mathbf{E}[\operatorname{tr}_{DN} h(P(\boldsymbol{S}^N, \boldsymbol{S}^{N*}))] = \Psi_h(\tfrac{1}{N}) = \frac{f_h(\tfrac{1}{N})}{g_{qq_0}(\tfrac{1}{N})}$$

for $N \geq qq_0$, where f_h is a polynomial of degree at most $qq_0(1 + \log d)$.

As Ψ_h has no poles near 0, we can define for each $m \in \mathbb{Z}_+$

$$\nu_m(h) := \frac{\Psi_h^{(m)}(0)}{m!}.$$

It is clear from the definition of Ψ_h that ν_m defines a linear functional on $\mathcal{P}$. Moreover, it follows immediately from Taylor's theorem that

$$\left| \mathbf{E}[\operatorname{tr}_{DN} h(P(\boldsymbol{S}^N, \boldsymbol{S}^{N*}))] - \sum_{i=0}^{m-1} \frac{\nu_i(h)}{N^i} \right| \leq \frac{\|\Psi_h^{(m)}\|_{C^0[0, \frac{1}{N}]}}{m!} \frac{1}{N^m} \tag{6.1}$$

provided N is large enough that Ψ_h has no poles on $[0, \frac{1}{N}]$. The main idea of the proof is that we will use Lemma 4.3 to bound the remainder term in the Taylor expansion. To this end we must show that the function g_{qq_0} is nearly constant.

Lemma 6.3. $e^{-1} \le g_q(x) \le 1$ *for every* $x \in [0, \frac{1}{q^2(1+\log d)}]$ *and* $q \in \mathbb{N}$.

Proof. The definition of g_q in Lemma 5.1 implies $(1 - qx)^{q(1+\log d)-1} \le g_q(x) \le 1$ for $x \in [0, \frac{1}{q}]$, where we used that $d_1 + \cdots + d_{q-1} \le q(1 + \log d) - 1$ as shown in the proof of Lemma 5.1. The conclusion follows using $(1 - \frac{1}{a})^{a-1} \ge e^{-1}$ for $a > 1$. $\square$

Next, we must obtain an upper bound on $\|\Psi_h\|_{C^0[0, \frac{1}{N}]}$.

Lemma 6.4. $\|\Psi_h\|_{C^0[0, \frac{1}{M}]} \le 2e\|h\|_{C^0[-K,K]}$ *for* $M = 2(qq_0(1 + \log d))^2$.

Proof. As $\|P(\boldsymbol{S}^N, \boldsymbol{S}^{N*})\| \le \|P\|_{\mathrm{M}_D(\mathbb{C}) \otimes C^*(\mathbf{F}_d)}$ almost surely, we have

$$|\Psi_h(\tfrac{1}{N})| = |\mathbf{E}[\mathrm{tr}_{DN}\, h(P(\boldsymbol{S}^N, \boldsymbol{S}^{N*}))]| \le \|h\|_{C^0[-K,K]}$$

for all $N \in \mathbb{N}$. We will extend this bound from the set $I = \{\frac{1}{N} : N \ge M\}$ to the interval $[0, \frac{1}{M}]$ using polynomial interpolation. First, note that

$$|f_h(\tfrac{1}{N})| \le |\Psi_h(\tfrac{1}{N})| \le \|h\|_{C^0[-K,K]}$$

as $g_{qq_0}(\frac{1}{N}) \le 1$ for $N \ge q$. The assumption of Lemma 4.2 is satisfied for $h \leftarrow f_h$, $q \leftarrow qq_0(1 + \log d)$, $a = \frac{1}{M}$ as soon as $M \ge 2(qq_0(1 + \log d))^2$, which yields

$$\|f_h\|_{C^0[0, \frac{1}{M}]} \le 2\|h\|_{C^0[-K,K]}.$$

It remains to apply Lemma 6.3 to estimate $\|\Psi_h\|_{C^0[0, \frac{1}{M}]} \le e\|f_h\|_{C^0[0, \frac{1}{M}]}$. $\square$

We can now conclude the proof.

Proof of Theorem 6.1 and Corollary 6.2. Let $M = 2(qq_0(1 + \log d))^2$. Then

$$\frac{\|\Psi_h^{(m)}\|_{C^0[0, \frac{1}{M}]}}{m!} \le (4qq_0(1 + \log d))^{4m}\|h\|_{C^0[-K,K]}$$

by Lemmas 4.3, 6.3, and 6.4, where we used that $(10e)^m 2e \le 4^{4m}$ for $m \ge 1$. Thus the proof of Theorem 6.1 for $N \ge M$ follows directly from (6.1).

The proof of Corollary 6.2 for $m \ge 1$ now follows by multiplying the bound of Theorem 6.1 by N^m and taking $N \to \infty$. As $\nu_0(h) = \lim_N \mathbf{E}[\mathrm{tr}_{DN}\, h(P(\boldsymbol{S}^N, \boldsymbol{S}^{N*}))]$, the conclusion of Corollary 6.2 evidently also holds for $m = 0$.

It remains to prove Theorem 6.1 for $N < M$. To this end, we can trivially bound the left-hand side of the inequality in Theorem 6.1 by

$$\left|\mathbf{E}[\mathrm{tr}_{DN}\, h(P(\boldsymbol{S}^N, \boldsymbol{S}^{N*}))] - \sum_{i=0}^{m-1} \frac{\nu_i(h)}{N^i}\right| \le \left(1 + \sum_{i=0}^{m-1} \frac{(4qq_0(1 + \log d))^{4i}}{N^i}\right)\|h\|_{C^0[-K,K]}$$

using the triangle inequality and Corollary 6.2. But note that $1 < \frac{2^k(qq_0(1+\log d))^{4k}}{N^k}$ for every $k \ge 1$ as we assumed $N < M$. We can therefore estimate

$$1 + \sum_{i=0}^{m-1} \frac{(4qq_0(1 + \log d))^{4i}}{N^i} \le \left(2^m + \sum_{i=0}^{m-1} 4^{4i}2^{m-i}\right) \frac{(qq_0(1 + \log d))^{4m}}{N^m},$$

and we conclude using $2^m + \sum_{i=0}^{m-1} 4^{4i}2^{m-i} \le 4^{4m}$. $\square$

7. MASTER INEQUALITIES FOR RANDOM PERMUTATIONS II

The aim of this short section is to extend the master inequalities of Theorem 6.1 from polynomials to general smooth test functions. This extension will play a key role in the proofs of strong convergence.

Theorem 7.1 (Smooth master inequalities). *Fix a self-adjoint noncommutative polynomial $P \in \mathrm{M}_D(\mathbb{C}) \otimes \mathbb{C}\langle \boldsymbol{s}, \boldsymbol{s}^* \rangle$ of degree q_0, and let $K = \|P\|_{\mathrm{M}_D(\mathbb{C}) \otimes C^*(\mathbf{F}_d)}$. Then there exists a compactly supported distribution ν_i for every $i \in \mathbb{Z}_+$ so that*

$$\left| \mathbf{E}[\mathrm{tr}_{DN}\, h(P(\boldsymbol{S}^N, \boldsymbol{S}^{N*}))] - \sum_{i=0}^{m-1} \frac{\nu_i(h)}{N^i} \right| \lesssim \frac{(4q_0(1+\log d))^{4m}}{N^m} \beta_* \|f^{(4m+1)}\|_{L^\beta[0,2\pi]}$$

for all $N, m \in \mathbb{N}$, $\beta > 1$, $h \in C^\infty(\mathbb{R})$, where $f(\theta) := h(K \cos \theta)$ and $1/\beta_* := 1 - 1/\beta$.

Proof. We first note that Corollary 6.2 ensures, by Lemma 4.7, that the linear functionals ν_i in Theorem 6.1 extend uniquely to compactly supported distributions.

Let $h \in \mathcal{P}_q$ and express it in the form (4.1). Rather than applying Theorem 6.1 to h directly, we apply it to the Chebyshev polynomials $T_j(K^{-1}\cdot)$ to obtain

$$\left| \mathbf{E}[\mathrm{tr}_{DN}\, h(P(\boldsymbol{S}^N, \boldsymbol{S}^{N*}))] - \sum_{i=0}^{m-1} \frac{\nu_i(h)}{N^i} \right| \leq \frac{(4q_0(1+\log d))^{4m}}{N^m} \sum_{j=1}^q j^{4m} |a_j|.$$

Here we used that the left-hand side of Theorem 6.1 vanishes when h is the constant function (as then $\nu_0(h) = h$, $\nu_i(h) = 0$ for $i \geq 1$), so the constant term in the Chebyshev expansion cancels. The proof is completed for $h \in \mathcal{P}_q$ by applying Lemma 4.4. The conclusion extends to $h \in C^\infty$ as $\mathcal{P}_q$ is dense in $C^\infty[-K, K]$. $\square$

The special case $m = 1$ is of independent interest, as it yields a quantitative formulation of weak convergence (cf. Corollary 5.3).

Corollary 7.2 (Quantitative weak convergence). *In the setting of Theorem 7.1,*

$$\left| \mathbf{E}[\mathrm{tr}_{DN}\, h(P(\boldsymbol{S}^N, \boldsymbol{S}^{N*}))] - (\mathrm{tr}_D \otimes \tau)(h(P(\boldsymbol{s}, \boldsymbol{s}^*))) \right| \lesssim \frac{(q_0(1+\log d))^4}{N} \beta_* \|f^{(5)}\|_{L^\beta[0,2\pi]}$$

for all $N \in \mathbb{N}$, $\beta > 1$, $h \in C^\infty(\mathbb{R})$, where $f(\theta) := h(K \cos \theta)$ and $1/\beta_* := 1 - 1/\beta$.

Proof. This is simply a restatement of the special case of Theorem 7.1 with $m = 1$, where we note that $\nu_0(h) = (\mathrm{tr}_D \otimes \tau)(h(P(\boldsymbol{s}, \boldsymbol{s}^*)))$ by Corollary 5.3. $\square$

Remark 7.3. When h is chosen to be a smooth approximation of the indicator function of an interval, Corollary 7.2 shows that the spectral distribution of $P(\boldsymbol{S}^N, \boldsymbol{S}^{N*})$ is well approximated by that of $P(\boldsymbol{s}, \boldsymbol{s}^*)$ at a mesoscopic scale. While far sharper results at the local scale are proved in [40, 39] for random regular graphs, no result of this kind appears to be known to date for arbitrary polynomials P.

In the remainder of this paper, we adopt the notation for the distributions ν_i as in Theorem 7.1. It should be emphasized, however, that the definition of ν_i depends both on the model and on the noncommutative polynomial P under consideration. As each of the following sections will be concerned with a specific model and choice of polynomial P, the meaning of ν_i will always be clear from context.

8. Random regular graphs

The aim of this section is to prove our main results on random regular graphs, the effective Friedman Theorem 3.4 and the staircase Theorem 3.6.

8.1. **Proof of Theorem 3.4.** The proof of Theorem 3.4 is based on Theorem 7.1 with $m = 2$. Let us spell out what this says in the present setting.

Lemma 8.1. *There exists a compactly supported distribution ν_1 such that for every $N \in \mathbb{N}$, $\beta > 1$, and $h \in C^\infty(\mathbb{R})$, we have*

$$\left| \mathbf{E}[\mathrm{tr}_N \, h(A^N)] - \tau(h(A_{\mathrm{F}})) - \tfrac{1}{N}\nu_1(h) \right| \lesssim \frac{(1 + \log d)^8}{N^2} \, \beta_* \|f^{(9)}\|_{L^\beta[0,2\pi]},$$

where $f(\theta) := h(2d\cos\theta)$ and $1/\beta_* = 1 - 1/\beta$.

Proof. Apply Theorem 7.1 with $m = 2$ and $P(\boldsymbol{s}, \boldsymbol{s}^*) = s_1 + s_1^* + \cdots + s_d + s_d^*$, using $\|P\|_{C^*(\mathbf{F}_d)} = 2d$ and $\nu_0(h) = \lim_N \mathbf{E}[\mathrm{tr}_N \, h(A^N)] = \tau(h(A_{\mathrm{F}}))$ by Corollary 5.3. □

The remaining ingredient of the proof is to show that $\mathrm{supp}\,\nu_1 \subseteq [-\|A_{\mathrm{F}}\|, \|A_{\mathrm{F}}\|]$. To this end, is suffices by Lemma 4.9 to understand the exponential growth rate of the moments of ν_1. We begin by computing a formula for the moments of ν_1.[2] Recall that $\mathbf{F}_d^{\mathrm{np}}$ denotes the non-power elements of $\mathbf{F}_d$ (see Lemma 5.2).

Lemma 8.2. *For all $p \in \mathbb{N}$, we have*

$$\nu_1(x^p) = -\tau(A_{\mathrm{F}}^p) + \sum_{k=2}^{p}(\omega(k) - 1) \sum_{v \in \mathbf{F}_d^{\mathrm{np}}} \sum_{i_1,\ldots,i_p=1}^{2d} 1_{g_{i_1}\cdots g_{i_p}=v^k}.$$

Proof. Note that

$$\nu_1(x^p) + \tau(A_{\mathrm{F}}^p) = \lim_{N\to\infty} N\left(\mathbf{E}[\mathrm{tr}_N(A^N)^p] - (1 - \tfrac{1}{N})\,\tau(A_{\mathrm{F}}^p)\right)$$

$$= \lim_{N\to\infty} \sum_{i_1,\ldots,i_p=1}^{2d} N\left(\mathbf{E}[\mathrm{tr}_N \, S_{i_1}^N \cdots S_{i_p}^N] - (1 - \tfrac{1}{N})\,\tau(\lambda(g_{i_1})\cdots\lambda(g_{i_p}))\right).$$

As $\tau(\lambda(g_{i_1})\cdots\lambda(g_{i_p})) = 1_{g_{i_1}\cdots g_{i_p}=e}$, the terms with $g_{i_1}\cdots g_{i_p} = e$ cancel in the sum as in the proof of Corollary 5.3. The conclusion now follows from Lemma 5.2 (note that the $k = 1$ term does not appear in the sum as $\omega(1) - 1 = 0$). □

In the present setting, a simple argument due to Friedman [26, Lemma 2.4] suffices to bound the growth rate of the moments. As a variant of this argument will be needed later in this paper, we recall the proof here.

Lemma 8.3. *For every $k \geq 2$ and $p \geq 1$, we have*

$$\sum_{v \in \mathbf{F}_d^{\mathrm{np}}} \sum_{i_1,\ldots,i_p=1}^{2d} 1_{g_{i_1}\cdots g_{i_p}=v^k} \leq (p+1)^4 \|A_{\mathrm{F}}\|^p.$$

---

[2]In the special setting of random regular graphs, a nonrigorous derivation of an explicit formula for ν_1 (that is, an explicit solution to the Hausdorff moment problem associated to Lemma 8.2) appears in the physics literature [47]. However, an explicit solution to the moment problem, which is not available in more general situations, is not needed to apply the methods of this paper.

Proof. We would like to argue that if $g_{i_1} \cdots g_{i_p} = v^k$, there must exist $a < b$ so that $g_{i_1} \cdots g_{i_a} = v$, $g_{i_{a+1}} \cdots g_{i_b} = v$, and $g_{i_{b+1}} \cdots g_{i_p} = v^{k-2}$. But this need not be true: if v is not cyclically reduced, the last letters of v may cancel the first letters of the next repetition of v, and then the cancelled letters need not appear in $g_{i_1} \cdots g_{i_p}$.

To eliminate this ambiguity, we note that any $v \in \mathbf{F}_d^{\mathrm{np}}$ can be uniquely expressed as $v = gwg^{-1}$ with $w \in \mathbf{F}_d^{\mathrm{np}}$, $g \in \mathbf{F}_d$ such that w is cyclically reduced and such that gwg^{-1} is reduced if $g \neq e$. Thus we may write for any $k \geq 2$

$$\sum_{v \in \mathbf{F}_d^{\mathrm{np}}} \sum_{i_1,\ldots,i_p=1}^{2d} 1_{g_{i_1} \cdots g_{i_p} = v^k} \leq \sum_{(w,g)} \sum_{0 \leq t_1 \leq t_2 \leq t_3 \leq t_4 \leq p} \sum_{i_1,\ldots,i_p=1}^{2d} \left(1_{g_{i_1} \cdots g_{i_{t_1}} = g} \times \right.$$
$$\left. 1_{g_{i_{t_1+1}} \cdots g_{i_{t_2}} = w} 1_{g_{i_{t_2+1}} \cdots g_{i_{t_3}} = w} 1_{g_{i_{t_3+1}} \cdots g_{i_{t_4}} = w^{k-2}} 1_{g_{i_{t_4+1}} \cdots g_{i_p} = g^{-1}} \right),$$

where the sum is over pairs (w, g) with the above properties.

The idea is now to express the indicators as $1_{g_{i_1} \cdots g_{i_t} = v} = \langle \delta_v, \lambda(g_{i_1}) \cdots \lambda(g_{i_t}) \delta_e \rangle$. Substituting this identity into the right-hand side of the above inequality yields

$$\sum_{v \in \mathbf{F}_d^{\mathrm{np}}} \sum_{i_1,\ldots,i_p=1}^{2d} 1_{g_{i_1} \cdots g_{i_p} = v^k} \leq \sum_{(w,g)} \sum_{0 \leq t_1 \leq t_2 \leq t_3 \leq t_4 \leq p} \left(\langle \delta_g, A_{\mathrm{F}}^{t_1} \delta_e \rangle \langle \delta_w, A_{\mathrm{F}}^{t_2-t_1} \delta_e \rangle \times \right.$$
$$\left. \langle \delta_w, A_{\mathrm{F}}^{t_3-t_2} \delta_e \rangle \langle \delta_{w^{k-2}}, A_{\mathrm{F}}^{t_4-t_3} \delta_e \rangle \langle \delta_{g^{-1}}, A_{\mathrm{F}}^{p-t_4} \delta_e \rangle \right)$$
$$\leq \sum_{0 \leq t_1 \leq t_2 \leq t_3 \leq t_4 \leq p} \| A_{\mathrm{F}} \|^{t_4-t_3} \left(\sum_{g \in \mathbf{F}_d} \langle \delta_g, A_{\mathrm{F}}^{t_1} \delta_e \rangle \langle \delta_{g^{-1}}, A_{\mathrm{F}}^{p-t_4} \delta_e \rangle \right) \times$$
$$\left(\sum_{w \in \mathbf{F}_d} \langle \delta_w, A_{\mathrm{F}}^{t_2-t_1} \delta_e \rangle \langle \delta_w, A_{\mathrm{F}}^{t_3-t_2} \delta_e \rangle \right)$$

where we used $|\langle \delta_{w^{k-2}}, A_{\mathrm{F}}^{t_4-t_3} \delta_e \rangle| \leq \| A_{\mathrm{F}} \|^{t_4-t_3}$. The conclusion now follows readily by applying the Cauchy–Schwarz inequality to the sums over g and w, as

$$\sum_{v \in \mathbf{F}_d} |\langle \delta_v, A_{\mathrm{F}}^t \delta_e \rangle|^2 = \| A_{\mathrm{F}}^t \delta_e \|^2 \leq \| A_{\mathrm{F}} \|^{2t}$$

for any $t \geq 0$. $\qquad\square$

Corollary 8.4. $\operatorname{supp} \nu_1 \subseteq [-\| A_{\mathrm{F}} \|, \| A_{\mathrm{F}} \|]$.

Proof. Lemmas 8.2 and 8.3 imply that $|\nu_1(x^p)| \leq (1 + p^2(p+1)^4) \| A_{\mathrm{F}} \|^p$ for all $p \geq 1$, so that the conclusion follows from Lemma 4.9. $\qquad\square$

We can now complete the proof of Theorem 3.4.

Proof of Theorem 3.4. Let χ be the test function provided by Lemma 4.10 with $m = 8$, $K = 2d$, and $\rho = \| A_{\mathrm{F}} \| = 2\sqrt{2d-1}$. As $\chi(x) = 0$ for $|x| \leq \| A_{\mathrm{F}} \|$, we have $\tau(\chi(A_{\mathrm{F}})) = 0$ and $\nu_1(\chi) = 0$ by Corollary 8.4. Thus Lemmas 8.1 and 4.10 yield

$$\mathbf{E}[\operatorname{tr}_N \chi(A^N)] \lesssim \frac{(\log d)^8}{N^2} \beta_* \left(\frac{d}{\varepsilon} \right)^{8 + 1/\beta_*}$$

for all $\varepsilon < 2d - 2\sqrt{2d-1}$ and $\beta_* \geq 1$.

To conclude, we note that as $\chi(x) \geq 1$ for $|x| \geq \|A_{\mathrm{F}}\| + \varepsilon$, we have

$$\mathbf{P}[\|A^N\| \geq \|A_{\mathrm{F}}\| + \varepsilon] \leq \mathbf{P}[\mathrm{Tr}\,\chi(A^N) \geq 1] \lesssim \frac{1}{N}\left(\frac{d\log d}{\varepsilon}\right)^8 \log\left(\frac{2ed}{\varepsilon}\right)$$

for $\varepsilon < 2d - 2\sqrt{2d-1}$, where we chose $\beta_* = 1 + \log(\frac{2d}{\varepsilon})$. $\qquad\square$

8.2. **Proof of Theorem 3.6.** The proof is based on results of Puder [58]. Let us begin by synthesizing the parts of [58] that we need here.

Lemma 8.5. *Define ρ_m as in Theorem 3.6. Then for any $2 \leq m \leq d$, we have*

$$\limsup_{p\to\infty} |\nu_m(x^p)|^{1/p} \leq \rho_m.$$

Proof. Fix a word $w \in \mathbf{W}_d$ of length q that does not reduce to the identity. By Lemmas 5.1 and 5.2, we can write for all sufficiently large N

$$N\,\mathbf{E}[\mathrm{tr}_N\,w(\boldsymbol{S}^N)] = \sum_{s=0}^{\infty} \frac{b_s(w)}{N^s},$$

as a rational function is analytic away from its poles. It is shown in [58, §5.5] that $b_s(w) = 0$ for $s \leq \pi(w) - 2$, that $b_{\pi(w)-1}(w) = |\mathrm{Crit}(w)|$, and that $|b_s(w)| \leq q^{2(s+1)}$ for $s \geq \pi(w)$, where $\pi(w) \in \{1, \ldots, d\} \cup \{\infty\}$ denotes the primitivity rank of w. We refer to [58, §2.2] for precise definitions of these notions.

Theorem 6.1 with $m = d + 1$ and $P(\boldsymbol{s}, \boldsymbol{s}^*) = s_1 + s_1^* + \cdots + s_d + s_d^*$ yields

$$\sum_{i_1,\ldots,i_p=1}^{2d} N\,\mathbf{E}[\mathrm{tr}_N\,S_{i_1}^N \cdots S_{i_p}^N]\,1_{g_{i_1}\cdots g_{i_p} \neq e} = N\big(\mathbf{E}[\mathrm{tr}_N(A^N)^p] - (1 - \tfrac{1}{N})\nu_0(x^p)\big)$$

$$= \nu_0(x^p) + \sum_{i=0}^{d-1} \frac{\nu_{i+1}(x^p)}{N^i} + O\left(\frac{1}{N^d}\right)$$

for $N \to \infty$ with p, d fixed (cf. the proof of Lemma 8.2). It follows that

$$\nu_m(x^p) = \sum_{i_1,\ldots,i_p=1}^{2d} b_{m-1}(g_{i_1}\cdots g_{i_p})\,1_{g_{i_1}\cdots g_{i_p} \neq e} = \sum_{r=1}^{m}\sum_{w\in\mathcal{W}_p^r} b_{m-1}(w)$$

for $m \geq 2$, where $\mathcal{W}_p^r = \{w = g_{i_1}\cdots g_{i_p} : 1 \leq i_1, \ldots, i_p \leq 2d, \pi(w) = r\}$. Thus

$$|\nu_m(x^p)| \leq mp^{2m} \max_{r=1,\ldots,m}\sum_{w\in\mathcal{W}_p^r} |\mathrm{Crit}(w)|,$$

where we used that $|\mathrm{Crit}(w)| \geq 1$ for $\pi(w) \neq \infty$. The conclusion now follows from the second (unnumbered) theorem in [58, §2.4]. $\qquad\square$

Corollary 8.6. *Let $d \geq 2$. Then $\mathrm{supp}\,\nu_m \subseteq [-\rho_m, \rho_m]$ for all $0 \leq m \leq d$.*

Proof. For $m = 0, 1$ the follows from $\|A_F\| = 2\sqrt{2d-1}$ and Corollary 8.4, while for $m = 2, \ldots, d$ this follows from Lemmas 8.5 and 4.9. $\qquad\square$

We can now complete the proof of Theorem 3.6.

Proof of Theorem 3.6. Fix $m_* \leq m \leq d-1$. Let χ be the test function provided by Lemma 4.10 with $m \leftarrow 4(m+1)$, $K \leftarrow 2d$, $\rho \leftarrow \rho_m$. Then $\nu_i(\chi) = 0$ for $i \leq m$ by

Corollary 8.6. Thus Theorem 7.1 with $h \leftarrow \chi$, $m \leftarrow m + 1$ and Lemma 4.10 yield

$$\mathbf{P}[\|A^N\| \geq \rho_m + \varepsilon] \leq \mathbf{P}[\mathrm{Tr}\,\chi(A^N) \geq 1] \lesssim \frac{C_d}{N^m} \frac{1}{\varepsilon^{4(m+1)}} \log\left(\frac{2e}{\varepsilon}\right),$$

where we chose $\beta_* = 1 + \log(\frac{2}{\varepsilon})$ (note that $\beta_* > 1$ as $\varepsilon < \rho_{m+1} - \rho_m \leq 2$).

For the lower bound, it is shown in the proof of [27, Theorem 2.11] that

$$\mathbf{P}[\|A^N\| \geq \alpha'] \geq (1 - o(1))N^{1-m}$$

for all $m > m_*$ and $\alpha' < \rho_m$. Choosing $\alpha' = \rho_{m-1} + \varepsilon$ completes the proof. $\qquad\square$

9. Strong convergence of random permutation matrices

The aim of this section is to prove Theorem 3.9. Most of the proof is identical to that of Theorem 3.4. The only difficulty in the present setting is to adapt the argument of Lemma 8.3 for bounding $\mathrm{supp}\,\nu_1$. This is not entirely straightforward, because the proof of Lemma 8.3 relied on an overcounting argument which is not applicable to general polynomials. Nonetheless, we will show that a more careful implementation of the idea behind the proof can avoid this issue.

Throughout the proof, we fix a polynomial P as in Theorem 3.9, and define

$$X^N := P(\boldsymbol{S}^N, \boldsymbol{S}^{N*}), \qquad X_{\mathrm{F}} := P(\boldsymbol{s}, \boldsymbol{s}^*).$$

To simplify the proof, we begin by applying a linearization trick of Pisier [57] in order to factor X_{F} into polynomials of degree one.

Lemma 9.1. *For any $\varepsilon > 0$, there exist $q \in \mathbb{N}$ and operators*

$$X_j := \sum_{i=0}^{2d} A_{j,i} \otimes \lambda(g_i), \qquad j = 1, \ldots, q$$

with matrix coefficients $A_{j,i}$ of dimensions $D_j \times D_j'$ with $D_1 = D_q' = D$, $D_j' = D_{j+1}$ for $j = 1, \ldots, q-1$, so that $X_{\mathrm{F}} = X_1 \cdots X_q$ and $\|X_j\| \leq (\|X_{\mathrm{F}}\| + \varepsilon)^{1/q}$ for all j.

Proof. This is an immediate consequence of [57, Theorem 1]. $\qquad\square$

In the following, we fix $\varepsilon > 0$ and a factorization of X_{F} as in Lemma 9.1. We will also define X_j^N by replacing $\lambda(g_i) \leftarrow S_i^N$ in the definition of X_j. Note that as $g_1, \ldots, g_d$ are free, $X_{\mathrm{F}} = X_1 \cdots X_q$ implies that $X^N = X_1^N \cdots X_q^N$.

It will be convenient to extend the indexing of the above operators cyclically: we will define $X_j := X_{((j-1) \bmod q)+1}$ for all $j \in \mathbb{N}$. This implies, for example, that $X_{\mathrm{F}}^p = X_1 X_2 \cdots X_{pq}$. We similarly define $A_{j,i} := A_{((j-1) \bmod q)+1,i}$ for $j \in \mathbb{N}$.

9.1. **The first-order support.** In the present setting, the moments of the linear functionals ν_0 and ν_1 in Theorem 6.1 satisfy

$$\nu_0(x^p) = \lim_{N \to \infty} \mathbf{E}[\mathrm{tr}_{DN}(X^N)^p] = \tau(X_F^p),$$
$$\nu_1(x^p) = \lim_{N \to \infty} N\big(\mathbf{E}[\mathrm{tr}_{DN}(X^N)^p] - \tau(X_F^p)\big).$$

We begin by writing an expression fo $\nu_1(x^p)$.

Lemma 9.2. *For every $p \in \mathbb{N}$, we have*

$$\nu_1(x^p) = -\tau(X_{\mathrm{F}}^p) + \sum_{k=2}^{pq}(\omega(k)-1) \sum_{v \in \mathbf{F}_d^{\mathrm{np}}} \sum_{i_1,\dots,i_{pq}=0}^{2d} a_{i_1,\dots,i_{pq}} 1_{g_{i_1}\cdots g_{i_{pq}}=v^k},$$

where we define

$$a_{i_1,\dots,i_{pq}} := \mathrm{tr}_D\left(A_{1,i_1}A_{2,i_2}\cdots A_{pq,i_{pq}}\right).$$

Proof. The proof is identical to that of Lemma 8.2. $\qquad\square$

We would like to repeat the proof of Lemma 8.3 in the present setting. Recall that the idea of the proof is to write $v^k = gwww^{k-2}g^{-1}$ where w is cyclically reduced, and then to reason that any word $g_{i_1}\cdots g_{i_t}$ that reduces to v^k is a concatenation of words $g_{i_1}\cdots g_{i_{t_1}} = g$, $g_{i_{t_1+1}}\cdots g_{i_{t_2}} = w$, $g_{i_{t_2+1}}\cdots g_{i_{t_3}} = w$, etc. We can therefore sum over all such concatenations in the expression for ν_1.

The problem with this argument is that the above decomposition is not unique: for example, $g_1 g_2 g_2^{-1} g_1 = g_1^2$ can be decomposed in two different ways $(g_1, g_2 g_2^{-1} g_1)$ or $(g_1 g_2 g_2^{-1}, g_1)$. Thus when we sum over all possible ways of generating such concatenations, we are overcounting the number of words that reduce to v^k. Unlike in Lemma 8.3, the coefficients $a_{i_1,\dots,i_{pq}}$ can have both positive and negative signs and thus we cannot upper bound the moments by overcounting.

We presently show how a more careful analysis can avoid overcounting. We begin by introducing the following basic notion.

Definition 9.3. For any $v \in \mathbf{F}_d$ and $0 \le i_1,\dots,i_k \le 2d$, we write $g_{i_1}\cdots g_{i_k} \triangleq v$ if $g_{i_1}\cdots g_{i_k} = v$ and $g_{i_l}\cdots g_{i_k} \ne v$ for $1 < l \le k$.

To interpret this notion, one should think of any word $g_{i_1}\cdots g_{i_k}$ as defining a walk in the Cayley graph of $\mathbf{F}_d$ when read from right to left, starting at the identity. Then $g_{i_1}\cdots g_{i_k} \triangleq v$ states that the walk defined by $g_{i_1}\cdots g_{i_k}$ reaches v *for the first time* at its endpoint. Just as we can write

$$1_{g_{i_1}\cdots g_{i_k}=v} = \langle \delta_v, \lambda(g_{i_1})\cdots\lambda(g_{i_k})\delta_e\rangle,$$

the equivalence notion of Definition 9.3 may also be expressed in terms of matrix elements of the left-regular representation.

Lemma 9.4. *For any $0 \le i_1,\dots,i_k \le 2d$ and $v \in \mathbf{F}_d$, we have*

$$1_{g_{i_1}\cdots g_{i_k} \triangleq v} = \langle \delta_v, \lambda(g_{i_1})\cdots\lambda(g_{i_k})\delta_e\rangle$$
$$- \sum_{s=1}^{k-1}\langle \delta_e, \lambda(g_{i_1})Q\lambda(g_{i_2})\cdots Q\lambda(g_{i_s})\delta_e\rangle\langle \delta_v, \lambda(g_{i_{s+1}})\cdots\lambda(g_{i_k})\delta_e\rangle,$$

where we defined $Q := \mathbf{1} - \delta_e\delta_e^*$.

Proof. One may simply note that term s in the sum is the indicator of the event that $g_{i_1}\cdots g_{i_k} = v$, $g_{i_{s+1}}\cdots g_{i_k} = v$, and $g_{i_t}\cdots g_{i_k} \ne v$ for $1 < t < s+1$. $\qquad\square$

We now use this identity to express the kind of sum that appears in Lemma 9.2 exactly (without overcounting) in terms of matrix elements of the operators X_j.

Lemma 9.5. *Fix $v_1, \ldots, v_\ell \neq e \in \mathbf{F}_d$ $(\ell \geq 2)$ so that $v_1 \cdots v_\ell$ is reduced. Define*

$$X_{(t,s)} := X_t \cdots X_s,$$
$$X_{[t,s]} := -X_t(\mathbf{1}_{D_{t+1}} \otimes Q)X_{t+1} \cdots (\mathbf{1}_{D_s} \otimes Q)X_s$$

for $t \leq s$ and $X_{(t,s)} = X_{[t,s]} := \mathbf{1}$ for $t > s$. Then

$$\sum_{i_1,\ldots,i_{pq}=0}^{2d} a_{i_1,\ldots,i_{pq}} 1_{g_{i_1} \cdots g_{i_{pq}}=v_1\cdots v_\ell} =$$

$$\sum_{1<t_2\leq s_2<\cdots<t_\ell\leq s_\ell\leq pq} \langle \delta_{v_1}, X_{(1,t_2-1)}\delta_e \rangle \prod_{r=2}^{\ell} \langle \delta_e, X_{[t_r,s_r-1]}\delta_e \rangle \langle \delta_{v_r}, X_{(s_r,t_{r+1}-1)}\delta_e \rangle$$

where we let $t_{\ell+1} := pq + 1$.

Proof. Suppose that $g_{i_1} \cdots g_{i_{pq}} = v_1 \cdots v_\ell$. As we assumed that $v_1 \cdots v_\ell$ is reduced, there are unique indices $1 < t_2 < \cdots < t_\ell \leq pq$ so that

$$g_{i_1} \cdots g_{i_{t_2-1}} = v_1, \qquad g_{i_{t_r}} \cdots g_{i_{t_{r+1}-1}} \triangleq v_r \quad \text{for } 2 \leq r \leq \ell.$$

Thus we can write

$$1_{g_{i_1} \cdots g_{i_{pq}}=v_1\cdots v_\ell} = \sum_{1<t_2<\cdots<t_\ell\leq pq} 1_{g_{i_1}\cdots g_{i_{t_2-1}}=v_1} \prod_{r=2}^{\ell} 1_{g_{i_{t_r}}\cdots g_{i_{t_{r+1}-1}}\triangleq v_r}.$$

The conclusion follows readily by applying Lemma 9.4 to the indicators in this identity and summing over the indices $i_1, \ldots, i_{pq}$. $\qquad\square$

With this identity in hand, we can proceed exactly as in Lemma 8.3. In the following lemma, recall that we chose an arbitrary $\varepsilon > 0$ in Lemma 9.1.

Lemma 9.6. *For every $k \geq 2$, we have*

$$\left| \sum_{v\in\mathbf{F}_d^{\mathrm{np}}} \sum_{i_1,\ldots,i_{pq}=0}^{2d} a_{i_1,\ldots,i_{pq}} 1_{g_{i_1}\cdots g_{i_{pq}}=v^k} \right| \leq 2(pq)^8(\|X_\mathrm{F}\| + \varepsilon)^p.$$

Proof. Suppose first that $k \geq 3$. Let $\mathcal{F}$ be the set of $v \in \mathbf{F}_d^{\mathrm{np}}$ that are cyclically reduced, and let $\mathcal{F}' := \mathbf{F}_d^{\mathrm{np}} \backslash \mathcal{F}$. Then every $v \in \mathcal{F}'$ can be uniquely expressed as $v = gwg^{-1}$ with $w \in \mathcal{F}$, $g \in \mathbf{F}_d$ so that gwg^{-1} is reduced. Applying Lemma 9.5 with $\ell = 5$ and $v_1 \leftarrow g$, $v_2 \leftarrow w$, $v_3 \leftarrow w$, $v_4 \leftarrow w^{k-2}$, $v_5 \leftarrow g^{-1}$ yields

$$\left| \sum_{v\in\mathcal{F}'} \sum_{i_1,\ldots,i_{pq}=0}^{2d} a_{i_1,\ldots,i_{pq}} 1_{g_{i_1}\cdots g_{i_{pq}}=v^k} \right| \leq$$

$$\sum_{1<t_2\leq s_2<\cdots<t_5\leq s_5\leq pq} \|X_{[t_2,s_2-1]}\|\|X_{[t_3,s_3-1]}\|\|X_{[t_4,s_4-1]}\|\|X_{[t_5,s_5-1]}\|\|X_{(s_4,t_5-1)}\| \times$$

$$\sum_{w\in\mathbf{F}_d} |\langle \delta_w, X_{(s_2,t_3-1)}\delta_e \rangle \langle \delta_w, X_{(s_3,t_4-1)}\delta_e \rangle| \sum_{g\in\mathbf{F}_d} |\langle \delta_g, X_{(1,t_2-1)}\delta_e \rangle \langle \delta_{g^{-1}}, X_{(s_5,pq)}\delta_e \rangle|.$$

Using that $\|X_{(t,s-1)}\| \leq (\|X_\mathrm{F}\| + \varepsilon)^{(s-t)/q}$ and $\|X_{[t,s-1]}\| \leq (\|X_\mathrm{F}\| + \varepsilon)^{(s-t)/q}$ by Lemma 9.1 and applying Cauchy–Schwarz as in the proof of Lemma 8.3, we can upper bound the right-hand side by $(pq)^8(\|X_\mathrm{F}\| + \varepsilon)^p$.

If we sum instead over $v \in \mathcal{F}$ on the left-hand side, we can bound in exactly the same manner by applying Lemma 9.5 with $\ell = 3$ and $v_1 \leftarrow w$, $v_2 \leftarrow w$, $v_3 \leftarrow w^{k-2}$. Thus the proof is complete for the case $k \geq 3$.

The proof for $k = 2$ is identical, except that we now omit the w^{k-2} terms. $\qquad\square$

Corollary 9.7. $\operatorname{supp} \nu_1 \subseteq [-\|X_{\mathrm{F}}\|, \|X_{\mathrm{F}}\|]$.

Proof. Lemmas 9.2 and 9.6 imply that $|\nu_1(x^p)| \leq \|X_{\mathrm{F}}\|^p + 2(pq)^{10}(\|X_{\mathrm{F}}\| + \varepsilon)^p$. The conclusion follows by first applying Lemma 4.9, and then letting $\varepsilon \to 0$. $\qquad\square$

9.2. **Proof of Theorem 3.9.** The rest of the proof contains no new ideas.

Proof of Theorem 3.9. Let χ be the test function provided by Lemma 4.10 with $m = 8$, $K = \|P\|_{\mathrm{M}_d(\mathbb{C}) \otimes C^*(\mathbf{F}_d)}$, and $\rho = \|X_{\mathrm{F}}\|$. Then $\nu_0(\chi) = \nu_1(\chi) = 0$ by Corollary 9.7. Thus Theorem 7.1 with $h \leftarrow \chi$, $m \leftarrow 2$ and Lemma 4.10 yield

$$\mathbf{P}[\|X^N\| \geq \|X_{\mathrm{F}}\| + \varepsilon] \leq \mathbf{P}[\operatorname{Tr} \chi(X^N) \geq 1] \lesssim \frac{D}{N}\left(\frac{Kq_0 \log d}{\varepsilon}\right)^8 \log\left(\frac{eK}{\varepsilon}\right)$$

for all $\varepsilon < K - \|X_{\mathrm{F}}\|$, where we chose $\beta_* = 1 + \log(\frac{K}{\varepsilon})$. $\qquad\square$

10. Stable representations of the symmetric group

The aim of this section is to prove Theorem 3.14. The main issue here is to extend the basic properties of random permutations matrices of section 5 to general stable representations. With analogues of these properties in place, the remainder of the proof is nearly the same as that of Theorem 3.9.

Throughout this section, we fix a stable representation as in section 3.4.2 and adopt the notations introduced there. In addition, we will denote by β the degree of the polynomial $\varphi(x_1, x_2^2, \ldots, x_r^r)$ and by $\kappa := \sup_{N \geq N_0} \frac{D_N}{N^\alpha}$.

10.1. **Basic properties.** We will deduce the requisite properties from the much more precise results of [31]. (In fact, the only facts that will be needed here are relatively elementary, as is explained in [43, Remark 31] and [31, §1.5.1].)

Let $w \in \mathbf{W}_d$ be a word. Then clearly

$$w(\mathbf{\Pi}^N) = \pi_N(w(\boldsymbol{\sigma}))$$

where $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_d)$ are i.i.d. uniformly distributed elements of $\mathbf{S}_N$. The distribution of $w(\boldsymbol{\sigma})$ defines a probability measure $\mathbf{P}_{w,N}$ on $\mathbf{S}_N$, called the word measure.

Lemma 10.1. *Let $w \in \mathbf{W}_d$ be any word of length at most q that does not reduce to the identity, and let $N \geq \max\{\beta q, N_0\}$. Then we have*

$$\mathbf{E}[\operatorname{Tr} w(\mathbf{\Pi}^N)] = \frac{f_w(\frac{1}{N})}{g_q(\frac{1}{N})},$$

where

$$g_q(x) := (1 - x)^{d_1}(1 - 2x)^{d_2} \cdots (1 - (\beta q - 1)x)^{d_{\beta q - 1}}$$

with $d_j := \min\left(d, \lfloor \frac{\beta q}{j+1} \rfloor\right)$, and f_w, g_q are polynomials of degree at most $\beta q(1 + \log d)$.

Proof. We may assume without loss of generality that w is cyclically reduced with length $1 \leq \ell \leq q$, as $\operatorname{Tr} w(\mathbf{\Pi}^N)$ is invariant under cyclic reduction.

The definition of the stable representation implies that

$$\mathbf{E}[\operatorname{Tr} w(\mathbf{\Pi}^N)] = \mathbf{E}_{w,N}[\varphi(\xi_1, \ldots, \xi_r)]. \tag{10.1}$$

Let $\xi_1^{\alpha_1} \cdots \xi_r^{\alpha_r}$ be a monomial of φ of degree at least one. Then for N sufficiently large, [31, Example 3.6, Definition 6.6, and Proposition 6.8] imply that

$$\mathbf{E}_{w,N}[\xi_1^{\alpha_1} \cdots \xi_r^{\alpha_r}] = \sum_\Gamma \frac{(\frac{1}{N})^{e_\Gamma - v_\Gamma} \prod_{i=1}^{v_\Gamma - 1}(1 - \frac{i}{N})}{\prod_{j=1}^d \prod_{i=1}^{e_\Gamma^j - 1}(1 - \frac{i}{N})}, \tag{10.2}$$

where the sum is over a certain collection of quotients Γ of the graph consisting of α_1 disjoint cycles of length ℓ, α_2 disjoint cycles of length 2ℓ, etc., whose edges are colored by $\{1, \ldots, d\}$. Here we denote by v_Γ the number of vertices of Γ, by e_Γ^j the number of edges of Γ colored by j, and by $e_\Gamma = e_\Gamma^1 + \cdots + e_\Gamma^d$.

It is immediate from the above construction that $v_\Gamma \leq \ell \sum_{i=1}^r i\alpha_i \leq \beta q$, and analogously $e_\Gamma \leq \beta q$. The validity of (10.2) for all $N \geq \beta q \geq \max_j e_\Gamma^j$ can be read off from the proof of [31, Proposition 6.8]. Moreover, we must have $e_\Gamma - v_\Gamma \geq 0$ for all Γ that appear in the sum, as $\mathbf{E}_{w,N}[\xi_1^{\alpha_1} \cdots \xi_r^{\alpha_r}] = O(1)$ by [31, Theorem 1.3] and the following discussion. Thus (10.2) is a rational function of $\frac{1}{N}$. The remainder of the proof proceeds exactly as in the proof of Lemma 5.1. $\qquad\square$

We now describe the lowest order asymptotics.

Lemma 10.2. *Fix a word $w \in \mathbf{W}_d$ that does not reduce to the identity, and express its reduction as $w(g_1, \ldots, g_d) = v^k$ with $v \in \mathbf{F}_d^{\mathrm{np}}$ and $k \geq 1$. Then*

$$\lim_{N \to \infty} \mathbf{E}[\operatorname{Tr} w(\mathbf{\Pi}^N)] = \varpi(k)$$

with $\varpi(1) = 0$ and $|\varpi(k)| \leq ck^\beta$ for $k \geq 2$. Here c is a constant that depends on φ.

Proof. Recall that φ is a sum of character polynomials of irreducible representations, and that we assumed that π_N does not contain the trivial representation. The case $k = 1$ therefore follows from Lemma 5.2 for the standard representation and from [31, Corollary 1.7] for all other irreducible components of φ.

For $k \geq 2$, let $\xi_1^{\alpha_1} \cdots \xi_r^{\alpha_r}$ be a monomial of φ. Applying [31, Theorem 1.3] and the following discussion as well as [31, Remark 7.3] yields

$$\lim_{N \to \infty} \mathbf{E}_{w,N}[\xi_1^{\alpha_1} \cdots \xi_r^{\alpha_r}] = \lim_{N \to \infty} \mathbf{E}_{v,N}[\xi_k^{\alpha_1} \cdots \xi_{rk}^{\alpha_r}] = \sum_{\mathcal{P} \in \operatorname{Partitions}(S)} \prod_{A \in \mathcal{P}} \sum_{j | \gcd(A)} j^{|A|-1},$$

where S is the multiset that contains jk with multiplicity α_j for $j = 1, \ldots, r$. By crudely estimating $j \leq \gcd(A) \leq rk$, the right-hand side is bounded by $B_{|S|}(rk)^{|S|}$ where $B_n \leq n!$ denotes the number of partitions of a set with n elements. The conclusion follows as $|S| = \sum_{j=1}^r \alpha_j \leq \beta$ and using (10.1). $\qquad\square$

Note that an immediate consequence of Lemma 10.2 is that the weak convergence as in Corollary 5.3 remains valid in the present setting.

10.2. **Proof of Theorem 3.14.** Fix a self-adjoint noncommutative polynomial $P \in \mathrm{M}_D(\mathbb{C}) \otimes \mathbb{C}\langle \boldsymbol{s}, \boldsymbol{s}^* \rangle$ of degree q_0, and let $h \in \mathcal{P}_q$. Then $(\operatorname{Tr} \otimes \operatorname{id})(h(P(\mathbf{\Pi}^N, \mathbf{\Pi}^{N*})))$ is a linear combination of words $w(\mathbf{\Pi}^N)$ of length at most qq_0. Summing separately

over words that do and do not reduce to the identity yields

$$\mathbf{E}[\operatorname{Tr} h(P(\mathbf{\Pi}^N, \mathbf{\Pi}^{N*}))] = \frac{\tilde{f}_h(\frac{1}{N})}{g_{qq_0}(\frac{1}{N})} + D_N\,(\operatorname{Tr}\otimes\tau)(h(P(\boldsymbol{s},\boldsymbol{s}^*)))$$

by Lemma 10.1, where $\tilde{f}_h$ is a polynomial of degree at most $\beta q q_0(1 + \log d)$. Thus

$$\mathbf{E}[\operatorname{tr}_{N^\alpha} h(P(\mathbf{\Pi}^N, \mathbf{\Pi}^{N*}))] = \Psi_h(\tfrac{1}{N}) = \frac{f_h(\frac{1}{N})}{g_{qq_0}(\frac{1}{N})}$$

where f_h is the polynomial of degree at most $\beta q q_0(1+\log d) + \alpha \le 2\beta q q_0(1 + \log d)$. Here we used that $\frac{D_N}{N^\alpha}$ is a polynomial of $\frac{1}{N}$ of degree α.

We can now repeat the proofs in sections 6–7 nearly verbatim with the replacements $q \leftarrow 2\beta N_0^{1/2} q$ (here we multiply by $N_0^{1/2}$ to ensure $M \ge N_0$ in Lemma 6.4) and $|\Psi_h(\frac{1}{N})| \le \frac{DD_N}{N^\alpha}\|h\|_{C^0[-K,K]} \le \kappa D\|h\|_{C^0[-K,K]}$ to conclude the following.

Lemma 10.3. *Let $d \ge 2$. Fix P as in Theorem 3.14 and $K = \|P\|_{\mathrm{M}_D(\mathbb{C})\otimes C^*(\mathbf{F}_d)}$. Then there exist compactly supported distributions ν_i such that*

$$\left| \mathbf{E}[\operatorname{tr}_{N^\alpha} h(P(\mathbf{\Pi}^N, \mathbf{\Pi}^{N*}))] - \sum_{i=0}^{\alpha} \frac{\nu_i(h)}{N^i} \right| \le \frac{CD\,(q_0 \log d)^{4(\alpha+1)}}{N^{\alpha+1}}\,\beta_*\|f^{(4(\alpha+1)+1)}\|_{L^\beta[0,2\pi]}$$

for all $N \ge N_0$, $\beta > 1$, and $h \in C^\infty(\mathbb{R})$. Here $f(\theta) := h(K\cos\theta)$, $1/\beta_* := 1 - 1/\beta$, and C is a constant that depends on the choice of stable representation.

Next, we must characterize the supports of the distributions ν_i.

Lemma 10.4. $\operatorname{supp}\nu_i \subseteq [-\|P(\boldsymbol{s},\boldsymbol{s}^*)\|, \|P(\boldsymbol{s},\boldsymbol{s}^*)\|]$ *for $i = 0, \dots, \alpha$.*

Proof. Suppose first that $0 \le i < \alpha$. As Lemma 10.2 implies that

$$\mathbf{E}[\operatorname{tr}_{N^\alpha} h(P(\mathbf{\Pi}^N, \mathbf{\Pi}^{N*}))] = \frac{D_N}{N^\alpha}\,(\operatorname{Tr}\otimes\tau)(h(P(\boldsymbol{s},\boldsymbol{s}^*))) + O\!\left(\frac{1}{N^\alpha}\right)$$

as $N \to \infty$ for every $h \in \mathcal{P}$, it follows that there is a constant c_i for every $i < \alpha$ so that $\nu_i(h) = c_i\,(\operatorname{Tr}\otimes\tau)(h(P(\boldsymbol{s},\boldsymbol{s}^*)))$. The claim follows directly.

We now consider $i = \alpha$. Then we can repeat the proof of Lemma 9.2, but using Lemma 10.2 instead of Lemma 5.2, to obtain the representation

$$\nu_\alpha(x^p) = D_0\tau(P(\boldsymbol{s},\boldsymbol{s}^*)^p) + \sum_{k=2}^{pq} \varpi(k) \sum_{v \in \mathbf{F}_d^{\mathrm{np}}} \sum_{i_1,\dots,i_{pq}=0}^{2d} a_{i_1,\dots,i_{pq}}\,1_{g_{i_1}\cdots g_{i_{pq}}=v^k}$$

where $D_0 := \varphi(0,\dots,0)$ is the constant term in the polynomial $N \mapsto D_N$. We therefore obtain $|\nu_\alpha(x^p)| \lesssim (pq)^{9+\beta}(\|P(\boldsymbol{s},\boldsymbol{s}^*)\| + \varepsilon)^p$ by Lemmas 9.6 and 10.2, and the conclusion follows by Lemma 4.9 and as $\varepsilon > 0$ is arbitrary. $\qquad\square$

The rest of the proof follows in the usual manner.

Proof of Theorem 3.14. Let χ be the test function provided by Lemma 4.10 with $m \leftarrow 4(\alpha+1)$ and $\rho \leftarrow \|P(\boldsymbol{s},\boldsymbol{s}^*)\|$. Then $\nu_i(\chi) = 0$ for all $i \le \alpha$ by Lemma 10.4. Thus Lemma 10.3 with $h \leftarrow \chi$, and Lemma 4.10 yield

$$\mathbf{P}[\|P(\mathbf{\Pi}^N, \mathbf{\Pi}^{N*})\| \ge \|P(\boldsymbol{s},\boldsymbol{s}^*)\| + \varepsilon] \le \frac{CD}{N}\left(\frac{Kq_0\log d}{\varepsilon}\right)^{4(\alpha+1)} \log\left(\frac{eK}{\varepsilon}\right),$$

where we chose $\beta_* = 1 + \log(\frac{K}{\varepsilon})$. $\qquad\square$

References

[1] S. Aaronson. The polynomial method in quantum and classical computing. In *Proc. 49th IEEE Symposium on Foundations of Computer Science*, page 3. IEEE, 2008.

[2] A. S. Bandeira, M. T. Boedihardjo, and R. van Handel. Matrix concentration inequalities and free probability. *Invent. Math.*, 234(1):419–487, 2023.

[3] R. Bauerschmidt, J. Huang, A. Knowles, and H.-T. Yau. Edge rigidity and universality of random regular graphs of intermediate degree. *Geom. Funct. Anal.*, 30(3):693–769, 2020.

[4] S. Belinschi and M. Capitaine. Strong convergence of tensor products of independent GUE matrices, 2022. Preprint arXiv:2205.07695.

[5] C. Bordenave. A new proof of Friedman's second eigenvalue theorem and its extension to random lifts. *Ann. Sci. Éc. Norm. Supér. (4)*, 53(6):1393–1439, 2020.

[6] C. Bordenave, D. Chafaï, and D. García-Zelada. Convergence of the spectral radius of a random matrix through its characteristic polynomial. *Probab. Theory Related Fields*, 182(3-4):1163–1181, 2022.

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

[8] C. Bordenave and B. Collins. Norm of matrix-valued polynomials in random unitaries and permutations, 2024. Preprint arxiv:2304.05714v2.

[9] C. Bordenave and B. Collins. Strong asymptotic freeness for independent uniform variables on compact groups associated to nontrivial representations. *Invent. Math.*, 237(1):221–273, 2024.

[10] P. Borwein and T. Erdélyi. *Polynomials and polynomial inequalities*, volume 161 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.

[11] J. Bourgain and L. Tzafriri. On a problem of Kadison and Singer. *J. Reine Angew. Math.*, 420:1–43, 1991.

[12] T. Brailovskaya and R. van Handel. Universality and sharp matrix concentration inequalities, 2023. Preprint arxiv:2201.05142v2.

[13] E. Cassidy, 2024. In preparation.

[14] C.-F. Chen, A. Bouland, F. G. S. L. Brandão, J. Docter, P. Hayden, and M. Xu. Efficient unitary designs and pseudorandom unitaries from permutations, 2024. Preprint arxiv:2404.16751.

[15] C.-F. Chen, J. Docter, M. Xu, A. Bouland, and P. Hayden. Efficient unitary t-designs from random sums, 2024. Preprint arxiv:2402.09335.

[16] E. W. Cheney. *Introduction to approximation theory*. AMS, Providence, RI, 1998.

[17] M. D. Choi. The full C^*-algebra of the free group on two generators. *Pacific J. Math.*, 87(1):41–48, 1980.

[18] T. Church, J. S. Ellenberg, and B. Farb. FI-modules and stability for representations of symmetric groups. *Duke Math. J.*, 164(9):1833–1910, 2015.

[19] B. Collins. Moment methods on compact groups: Weingarten calculus and its applications. In *ICM Vol. IV. Sections 5–8*, pages 3142–3164. EMS Press, Berlin, 2023.

[20] B. Collins, A. Guionnet, and F. Parraud. On the operator norm of non-commutative polynomials in deterministic matrices and iid GUE matrices. *Camb. J. Math.*, 10(1):195–260, 2022.

[21] B. Collins, S. Matsumoto, and J. Novak. The Weingarten calculus. *Notices Amer. Math. Soc.*, 69(5):734–745, 2022.

[22] D. Coppersmith and T. J. Rivlin. The growth of polynomials bounded at equally spaced points. *SIAM J. Math. Anal.*, 23(4):970–983, 1992.

[23] S. Coste, G. Lambert, and Y. Zhu. The characteristic polynomial of sums of random permutations and regular digraphs. *Int. Math. Res. Not. IMRN*, (3):2461–2510, 2024.

[24] R. Estrada and R. P. Kanwal. Moment sequences for a certain class of distributions. *Complex Variables Theory Appl.*, 9(1):31–39, 1987.

[25] B. Farb. Representation stability. In *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. II*, pages 1173–1196. Kyung Moon Sa, Seoul, 2014.

[26] J. Friedman. Relative expanders or weakly relatively Ramanujan graphs. *Duke Math. J.*, 118(1):19–35, 2003.

[27] J. Friedman. A proof of Alon's second eigenvalue conjecture and related problems. *Mem. Amer. Math. Soc.*, 195(910):viii+100, 2008.

[28] 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. In *STACS 96 (Grenoble, 1996)*, volume 1046 of *Lecture Notes in Comput. Sci.*, pages 375–386. Springer, Berlin, 1996.

[29] U. Haagerup, H. Schultz, and S. Thorbjørnsen. A random matrix approach to the lack of projections in $C_{\mathrm{red}}^*(\mathbb{F}_2)$. *Adv. Math.*, 204(1):1–83, 2006.

[30] 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.

[31] L. Hanany and D. Puder. Word measures on symmetric groups. *Int. Math. Res. Not. IMRN*, (11):9221–9297, 2023.

[32] M. B. Hastings and A. W. Harrow. Classical and quantum tensor product expanders. *Quantum Inf. Comput.*, 9(3-4):336–360, 2009.

[33] B. Hayes. A random matrix approach to the Peterson-Thom conjecture. *Indiana Univ. Math. J.*, 71(3):1243–1297, 2022.

[34] B. Hayes, D. Jekel, and S. K. Elayavalli. Consequences of the random matrix solution to the Peterson-Thom conjecture, 2023. Preprint arXiv:2308.14109.

[35] Y. He. Spectral gap and edge universality of dense random regular graphs, 2024. Preprint arXiv:2203.07317v4.

[36] W. Hide. Effective lower bounds for spectra of random covers and random unitary bundles, 2023. Preprint arxiv:2305.04584.

[37] W. Hide and M. Magee. Near optimal spectral gaps for hyperbolic surfaces. *Ann. of Math. (2)*, 198(2):791–824, 2023.

[38] L. Hörmander. *The analysis of linear partial differential operators. I*. Classics in Mathematics. Springer-Verlag, Berlin, 2003. Distribution theory and Fourier analysis.

[39] J. Huang, T. McKenzie, and H.-T. Yau. Optimal eigenvalue rigidity of random regular graphs, 2024. Preprint arxiv:2405.12161.

[40] J. Huang and H.-T. Yau. Spectrum of random d-regular graphs up to the edge. *Comm. Pure Appl. Math.*, 77(3):1635–1723, 2024.

[41] H. Kesten. Symmetric random walks on groups. *Trans. Amer. Math. Soc.*, 92:336–354, 1959.

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

[43] N. Linial and D. Puder. Word maps and spectra of random graph lifts. *Random Structures Algorithms*, 37(1):100–135, 2010.

[44] L. Louder and M. Magee. Strongly convergent unitary representations of limit groups, 2023. With an appendix by W. Hide and M. Magee. Preprint arxiv:2210.08953.

[45] R. C. Lyndon and P. E. Schupp. *Combinatorial group theory*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1977 edition.

[46] M. Magee and J. Thomas. Strongly convergent unitary representations of right-angled Artin groups, 2023. Preprint arxiv:2308.00863.

[47] F. L. Metz, G. Parisi, and L. Leuzzi. Finite-size corrections to the spectrum of regular random graphs: An analytical solution. *Phys. Rev. E*, 90:052109, Nov 2014.

[48] J. A. Mingo and R. Speicher. *Free probability and random matrices*, volume 35 of *Fields Institute Monographs*. Springer, New York, 2017.

[49] S. Mohanty, R. O'Donnell, and P. Paredes. Explicit near-Ramanujan graphs of every degree, 2019. Preprint arXiv:1909.06988v3.

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

[51] A. Nica. On the number of cycles of given length of a free word in several random permutations. *Random Structures Algorithms*, 5(5):703–730, 1994.

[52] A. Nica and R. Speicher. *Lectures on the combinatorics of free probability*. Cambridge, 2006.

[53] A. Nilli. On the second eigenvalue of a graph. *Discrete Math.*, 91(2):207–210, 1991.

[54] R. O'Donnell and X. Wu. Explicit near-fully X-Ramanujan graphs, 2020. Preprint arXiv:2009.02595.

[55] F. Parraud. On the operator norm of non-commutative polynomials in deterministic matrices and iid Haar unitary matrices. *Probab. Theory Related Fields*, 182(3-4):751–806, 2022.

[56] F. Parraud. Asymptotic expansion of smooth functions in polynomials in deterministic matrices and iid GUE matrices. *Comm. Math. Phys.*, 399(1):249–294, 2023.

[57] G. Pisier. On a linearization trick. *Enseign. Math.*, 64(3-4):315–326, 2018.

[58] D. Puder. Expansion of random graphs: new proofs, new results. *Invent. Math.*, 201(3):845–908, 2015.

[59] I. Rivin and N. T. Sardari. Quantum chaos on random Cayley graphs of $SL_2[\mathbb{Z}/p\mathbb{Z}]$. *Exp. Math.*, 28(3):328–341, 2019.

[60] A. Sarid. The spectral gap of random regular graphs. *Random Structures Algorithms*, 63(2):557–587, 2023.

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

[62] A. Song. Random minimal surfaces in spheres, 2024. Preprint arxiv:2402.10287.

[63] V. Vu. Random discrete matrices. In *Horizons of combinatorics*, volume 17 of *Bolyai Soc. Math. Stud.*, pages 257–280. Springer, Berlin, 2008.

[64] A. Zygmund. *Trigonometric series. Vol. I, II*. Cambridge, third edition, 2002.

Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, CA, USA
Email address: `chifang@caltech.edu`

Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA
Email address: `jgarzav@caltech.edu`

Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA
Email address: `jtropp@caltech.edu`

PACM, Princeton University, Princeton, NJ 08544, USA
Email address: `rvan@math.princeton.edu`