Strong convergence of uniformly random permutation representations of surface groups

Michael Magee, Doron Puder, and Ramon van Handel

Abstract

Let Γ be the fundamental group of a closed orientable surface of genus at least two. Consider the composition of a uniformly random element of $\mathrm{Hom}(\Gamma, S_n)$ with the $(n-1)$-dimensional irreducible representation of S_n. We prove the strong convergence in probability as $n \to \infty$ of this sequence of random representations to the regular representation of Γ.

As a consequence, for any closed hyperbolic surface X, with probability tending to one as $n \to \infty$, a uniformly random degree-n covering space of X has near optimal relative spectral gap — ignoring the eigenvalues that arise from the base surface X.

To do so, we show that the polynomial method of proving strong convergence can be extended beyond rational settings.

To meet the requirements of this extension we prove two new kinds of results. First, we show there are effective polynomial approximations of expected values of traces of elements of Γ under random homomorphisms to S_n. Secondly, we estimate the growth rates of probabilities that a finitely supported random walk on Γ is a proper power after a given number of steps.

Contents

1 Introduction

In the following $\Gamma = \Gamma_g$ will be the fundamental group of a closed orientable surface of genus $g \geqslant 2$. We view g and Γ as fixed in the paper. We denote by S_n the group of permutations of $[n] \stackrel{\text{def}}{=} \{1, \ldots, n\}$ and $\mathcal{U}(n)$ the group of $n \times n$ complex unitary matrices. The n-dimensional linear representation of S_n by permutation matrices is not irreducible but has an $(n-1)$-dimensional irreducible subrepresentation

$$\mathrm{std} : S_n \to \mathcal{U}(n-1)$$

obtained by removing the non-zero invariant vectors. The group Γ also has an important unitary representation called the *regular representation*:

$$\lambda : \Gamma \to \mathcal{U}(\ell^2(\Gamma)), \quad \lambda(\gamma)[f](x) \stackrel{\text{def}}{=} f(\gamma^{-1}x).$$

For $n \in \mathbf{N}$ the set of homomorphisms

$$\mathbb{X}_{g,n} \stackrel{\text{def}}{=} \mathrm{Hom}(\Gamma, S_n)$$

is finite and we equip it with the uniform probability measure. Let $\mathbf{C}[\Gamma]$ denote the group algebra of Γ. Every linear representation of Γ extends linearly to one of $\mathbf{C}[\Gamma]$.

Theorem 1.1. *For all $x \in \mathbf{C}[\Gamma]$, for $\phi_n \in \mathrm{Hom}(\Gamma, S_n)$ uniformly random, as $n \to \infty$*

$$\|\mathrm{std} \circ \phi_n(x)\| \to \|\lambda(x)\|$$

in probability. The norms on each side are operator norms.

In the language of strong convergence of unitary representations, Theorem 1.1 says precisely that the random representations $\mathrm{std} \circ \phi_n$ converge strongly in probability to λ. See [Mag25] for a survey on strong convergence. Theorem 1.1 is the analog, for surface groups, of the main theorem of Bordenave–Collins [BC19] for free groups. In §1.4 we cover the history of strong convergence results prior to Theorem 1.1.

Random hyperbolic surfaces. One of the primary motivations for Theorem 1.1 is a known important consequence on the spectral theory of random hyperbolic surfaces. A hyperbolic surface X is a complete Riemannian surface of constant curvature -1. Such a surface has an associated Laplacian operator Δ_X. If X is closed, the spectrum of Δ_X as an unbounded operator on $L^2(X)$ is discrete consisting of real eigenvalues that accumulate only at infinity. If X is connected, the bottom of the spectrum of Δ_X is a simple eigenvalue at 0 of multiplicity one.

The first non-zero eigenvalue, $\lambda_1(X)$, is of fundamental significance to e.g. the dynamics of the geodesic flow on X, the statistics of lengths of closed geodesics, and notions of connectedness of X. In cases where X is a congruence hyperbolic surface, Selberg recognized [Sel65] that the condition $\lambda_1(X) \geqslant \frac{1}{4}$ — the Selberg Eigenvalue conjecture — is the archimedean analog of the Ramanujan–Petersson conjectures from number theory, and from the point of view of the Selberg zeta function is also the analog of the Riemann hypothesis for these surfaces.

If $\{X_n\}_{n=1}^{\infty}$ are a sequence of closed hyperbolic surfaces with genera tending to ∞, then $\limsup_{n\to\infty} \lambda_1(X_n) \leqslant \frac{1}{4}$ [Hub74] and hence $\frac{1}{4}$ is the asymptotically optimal value of $\lambda_1(X_n)$ for such surfaces. The number $\frac{1}{4}$ arises here as the bottom of the spectrum of the Laplacian on hyperbolic space of two dimensions.

Theorem 1.2. *Let X be a closed connected hyperbolic surface. Let X_n denote a uniformly random degree n covering space of X.*[1] *With probability tending to one as $n \to \infty$*

$$\mathrm{spec}(\Delta_{X_n}) \cap \left[0, \frac{1}{4} - o(1)\right) = \mathrm{spec}(\Delta_X) \cap \left[0, \frac{1}{4} - o(1)\right),$$

with the same multiplicities on both sides.
In particular, if $\lambda_1(X) \geqslant \frac{1}{4}$, then $\lambda_1(X_n) \to \frac{1}{4}$ in probability as $n \to \infty$.

We omit the proof that Theorem 1.2 follows from Theorem 1.1, which is the same as in [LM25, Appendix A].[2]

The existence of covering spaces of hyperbolic surfaces that have optimal spectral gaps was proved in [HM23, LM25]. What Theorem 1.2 shows is that not only do such surfaces exist, but that this is in fact the *typical* behavior of covering spaces of closed hyperbolic surfaces. In §1.4 we describe prior results to Theorem 1.2.

1.1 Technical innovation I: The generalized polynomial method

The polynomial method, which was introduced in the recent work of Chen, Garza-Vargas, Tropp, and the third named author [CGVTvH24, CGVvH24] and further developed by de la Salle and the first named author [MdlS24b], provides a powerful approach for establishing strong convergence in situations that are outside the reach of previous methods. To date, however, all applications of the polynomial method have relied fundamentally on the property that the spectral statistics of the random matrices that arise in these applications are rational functions of $\frac{1}{n}$. This restrictive property

[1] If $\pi_1(X) \cong \Gamma$, then by definition, a uniformly random degree-n cover of X is given by a uniformly random element of $\mathbb{X}_{g,n}$ via the bijection between $\mathbb{X}_{g,n}$ and covers with a labeled fiber; see [MP23, §1].

[2] This passage as presented is non-effective. For an effective version one can see [Hid25].

fails manifestly for the model considered in this paper.[3]

The basis for the present paper is a refinement of the polynomial method, which makes it applicable to much more general situations where the dimension dependence of the spectral statistics need not even be analytic. This greatly expands the potential range of applications of this method. More precisely, this method will give rise to the following general criterion for strong convergence.

Let Λ be a finitely generated group with a finite generating set S. For $\gamma \in \Lambda$, denote by $|\gamma|$ the word length of γ with respect to S. Denote by λ the left regular representation of Λ, and for each $n \in \mathbf{N}$ let

$$\pi_n : \Lambda \to \mathcal{U}(n)$$

be a random unitary representation. We denote by $\operatorname{Tr} M$ the trace of a matrix M and by $\operatorname{tr} M$ its normalized trace. We let τ be the canonical trace on the reduced group C^*-algebra $C^*_{\mathrm{red}}(\Lambda)$. (The definitions of these notions are recalled in §1.6.)

To establish that $\pi_n \to \lambda$ strongly in probability, we will need two assumptions. The first is an effective $\frac{1}{n}$-expansion for the expected spectral statistics.

Assumption 1.3. *For every $\gamma \in \Lambda$, there exist $u_k(\gamma)$ with $u_0(\gamma) = \tau(\lambda(\gamma))$ and*

$$\left| \mathbb{E}\big[\operatorname{tr} \pi_n(\gamma)\big] - \sum_{k=0}^{q-1} \frac{u_k(\gamma)}{n^k} \right| \leqslant \frac{(Cq)^{Cq}}{n^q}$$

for all $q \geqslant |\gamma|$ and $n \geqslant Cq^C$, where $C \geqslant 1$ is a constant.

In the following, we will extend u_k linearly to the group algebra $\mathbf{C}[G]$. The second assumption is that u_1 is "tempered" in the sense of [CGVTvH24, MdlS24b].

Assumption 1.4. *For every self-adjoint $x \in \mathbf{C}[\Lambda]$, we have*

$$\limsup_{p\to\infty} |u_1(x^p)|^{1/p} \leqslant \|\lambda(x)\|.$$

The following theorem is proved in §3.

Theorem 1.5. *Suppose that Assumptions 1.3 and 1.4 hold. Then*

$$\|\pi_n(x)\| \leqslant \|\lambda(x)\| + o(1) \quad \textit{with probability} \quad 1 - o(1)$$

as $n \to \infty$ for every $x \in \mathbf{C}[\Lambda]$.

Theorem 1.5 gives a strong convergence upper bound. If $C^*_{\mathrm{red}}(\Lambda)$ has a unique trace, this automatically implies the lower bound [MdlS24b, §5.3]. We state the result here in

[3]Indeed, in the setting of this paper, the spectral statistics are not even analytic as a function of $\frac{1}{n}$. This can be verified by checking, by an explicit computation, that the coefficients $a_i(\gamma)$ in Theorem 1.6 grow faster than exponentially in i when γ is one of the standard generators of Γ. We omit the details.

a qualitative form for simplicity, but Theorem 3.3 below in fact provides a quantitative form of this result with a polynomial convergence rate.

Let us highlight two important features of the above result.

- Previous applications of the polynomial method required that $\mathbb{E}[\operatorname{tr} \pi_n(\gamma)]$ is a rational function of $\frac{1}{n}$. Assumption 1.3 is a drastic relaxation of this assumption: it requires only that the spectral statistics are in a Gevrey class (so they need not even be analytic, let alone rational) as a function of $\frac{1}{n}$.
- To verify Assumption 1.3, it is only necessary to consider individual group elements $\gamma \in \Lambda$ rather than elements of the group algebra $x = \sum_\gamma \alpha_\gamma \gamma \in \mathbf{C}[\Lambda]$.

The latter is crucial for applications. In principle, one may expect that a method that applies to polynomial (or rational) spectral statistics could be adapted to spectral statistics that are well approximated by polynomials. This is what Assumption 1.3 ensures. However, a key feature of the polynomial method is that it captures cancellations between the terms of $x \in \mathbf{C}[\Lambda]$ in the $\frac{1}{n}$-expansion of $\mathbb{E}[\operatorname{tr} \pi_n(x)]$ [CGVTvH24, §2.3]. It is not clear that such cancellations are preserved when the spectral statistics are approximated by polynomials. Since we do not know how to establish *a priori* bounds that capture cancellations, it is essential that this is not required by Assumption 1.3. We elaborate further on this point in the remarks at the end of §3.

1.2 Technical innovation II: Effective polynomial approximation

We aim to apply the criterion for strong convergence from Theorem 1.5 to the setting that $\Lambda = \Gamma = \Gamma_g$ is a surface group with $g \geqslant 2$, and

$$\pi_n = \operatorname{std} \circ \phi_n,$$

where ϕ_n is a uniformly distributed random element of $\operatorname{Hom}(\Gamma_g, S_n)$. Thus we need to establish Assumptions 1.3 and 1.4 for this setting. We begin with Assumption 1.3.

For $\sigma \in S_n$ let $\mathsf{fix}(\sigma)$ denote the number of fixed points of σ. For a random variable f on $\mathbb{X}_{g,n}$ we denote by $\mathbb{E}_{g,n}[f]$ the expectation of f with respect to the uniform measure. Natural random variables on $\mathbb{X}_{g,n}$ come from elements of γ: given $\gamma \in \Gamma$ we define

$$\mathsf{fix}_\gamma : \mathbb{X}_{g,n} \to \mathbf{Z}$$
$$\mathsf{fix}_\gamma : \phi \mapsto \mathsf{fix}(\phi(\gamma)).$$

Then we have

$$\mathbb{E}[\operatorname{Tr} \pi_n(\gamma)] = \mathbb{E}_{g,n}[\mathsf{fix}_\gamma] - 1. \tag{1.1}$$

An asymptotic expansion for $\mathbb{E}_{g,n}[\mathsf{fix}_\gamma]$ was obtained in the previous work [MP22, MP23] of the first two authors.

Theorem 1.6 ([MP23, Thm. 1.1])**.** *For each $\gamma \in \Gamma$ there exists a (unique) sequence $\{a_i(\gamma)\}_{i=-1}^{\infty}$ such that for any $q \in \mathbf{N}$, as $n \to \infty$*

$$\mathbb{E}_{g,n}[\mathsf{fix}_\gamma] = a_{-1}(\gamma)n + a_0(\gamma) + a_1(\gamma)n^{-1} + \cdots + a_{q-1}(\gamma)n^{-(q-1)} + O_{q,\gamma}\left(n^{-q}\right),$$

where the implied constant in the big O depends on q and γ.

In §4, we prove the following theorem that strengthens Theorem 1.6. Here and in the sequel, we fix a standard set of generators of Γ of size $2g$ coming from the presentation

$$\Gamma = \langle a_1, b_1, \ldots, a_g, b_g \mid [a_1, b_1] \cdots [a_g, b_g] \rangle, \tag{1.2}$$

and for $\gamma \in \Gamma$ write $|\gamma|$ for the word length of γ with respect to these generators.

Theorem 1.7. *There exists $C = C(g) > 1$ such that for all $q, n \in \mathbf{N}$ with $n \geqslant Cq^C$, and for all $\gamma \in \Gamma$ with $|\gamma| \leqslant q$,*

$$\left|\mathbb{E}_{g,n}[\mathsf{fix}_\gamma] - a_{-1}(\gamma)n - a_0(\gamma) - a_1(\gamma)n^{-1} - \cdots - a_{q-1}(\gamma)n^{-(q-1)}\right| \leqslant (Cq)^{Cq} n^{-q}.$$

The proof of this effective asymptotic expansion proceeds in two parts. The first part of the analysis relies on an exact expansion of $\mathbb{E}_{g,n}[\mathsf{fix}_\gamma]$ in terms of all irreducible representations of S_n that is developed in [MP23]. A careful analysis shows that we may truncate this expansion only to those representations defined by Young diagrams $\lambda \vdash n$ with $O(q)$ boxes outside the first row and column, with an effective error bound as in Assumption 1.3. This is Proposition 4.4.

It is not clear, however, how to analyze the truncated expansion using the methods of [MP23]. Instead, we introduce a new method for analyzing the truncated expansion by means of a combinatorial integration formula for the symmetric group. This enables us to show that the truncated expansion is in fact a rational function of $\frac{1}{n}$. An effective error bound for the truncated expansion can be then achieved using the analytic theory of polynomials, in a manner closely analogous to the polynomial method for random matrix models with rational spectral statistics [CGVTvH24, CGVvH24, MdlS24b].

1.3 Technical innovation III: Geometry of proper powers

It remains to establish Assumption 1.4 for surface groups.

For every element $1 \neq \gamma \in \Gamma$, there is a unique $\gamma_0 \in \Gamma$ and $k \in \mathbf{N}$ with $\gamma = \gamma_0^k$ such that γ_0 is not a proper power.[4] We denote by Γ_{np} the set of non-powers in Γ. Denote by $\omega(k)$ the number of positive divisors of k, and by abuse of notation, we define

$$\omega(\gamma) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } \gamma = 1, \\ \omega(k) & \text{if } \gamma = \gamma_0^k \text{ with } \gamma_0 \in \Gamma_{\mathrm{np}}. \end{cases}$$

[4]This follows immediately from the fact the non-trivial elements of Γ admit cyclic centralizer. The latter fact can be found, e.g., in [FM11, p. 23].

In the notation of Theorem 1.7, it is shown in [MP23, Thm. 1.2] that $a_0(\gamma) = \omega(\gamma)$. Thus u_1 in Assumption 1.4 can be expressed for any $x = \sum_{\gamma\in\Gamma} \alpha_\gamma \gamma \in \mathbf{C}[\Gamma]$ as

$$u_1(x) = \sum_{\gamma} \alpha_\gamma \left[\omega(\gamma) - 1\right] \tag{1.3}$$

using (1.1), see §6. Note that u_1 is determined only by the contributions of $\gamma \in \Gamma$ that are either the identity or a proper power, since $\omega(\gamma) - 1 = 0$ for $\gamma \in \Gamma_{\mathrm{np}}$.

We prove the following theorem in §5.

Theorem 1.8. *For any self adjoint $x \in \mathbf{C}[\Gamma]$,*

$$\limsup_{p\to\infty} |u_1(x^p)|^{1/p} \leqslant \|\lambda(x)\|\,.$$

To prove Theorem 1.8 it suffices, by [MdlS24b, Prop. 6.3] and as surface groups satisfy the rapid decay property [Jol90, Thm. 3.2.1], to control the exponential growth rate of the probability that a random walk in Γ with an arbitrary finite set of generators lands on a proper power.

In the free group setting, the latter is readily accomplished by a simple spectral argument that dates back to [Fri03, Lem. 2.4]. This argument relies on the following elementary property of free groups: if a word $w = a_{i_1} \cdots a_{i_q}$ in the free generators $a_1, \ldots, a_r$ reduces to a proper power γ_0^k and we write $\gamma_0 = bhb^{-1}$ where h is the cyclic reduction of γ_0, then w must be a concatenation of five consecutive sub-words that reduce to b, h, h, h^{k-2}, and b^{-1}, respectively.

The difficulty in the setting of surface groups is that such a property is no longer true. However, we will show in Lemma 5.6 that an approximate form of this property remains valid: if a word w in the standard generators of Γ is equivalent in Γ to a proper power γ_0^k, then there must exist $b, h \in \Gamma$ so that w is a concatenation of five consecutive sub-words which are equivalent in Γ to elements that are $c \log |w|$-close to b, h, h, h^{k-2}, and b^{-1} (here u is ℓ-close to v if u is equivalent in Γ to $\gamma v \gamma'$ with $|\gamma|, |\gamma'| \leqslant \ell$). The proof of this fact makes crucial use of the hyperbolic geometry of $\mathrm{Cay}(\Gamma)$. This property suffices to conclude the proof of Assumption 1.4 for surface groups.

1.4 Prior results

Strong convergence

The first result in the spirit of Theorem 1.1 was the breakthrough result of Haagerup and Thorbjørnsen [HT05], who proved that i.i.d. GUE random matrices strongly converge to free semicircular random variables. Let $\mathbf{F}_r$ denote a free group of rank r. Collins and Male proved in [CM14] that Haar-distributed elements of $\mathrm{Hom}(\mathbf{F}_r, \mathcal{U}(n)) \cong \mathcal{U}(n)^r$ strongly converge to the regular representation of $\mathbf{F}_r$. In another breakthrough [BC19], Bordenave and Collins proved that uniformly random elements of $\mathrm{Hom}(\mathbf{F}_r, S_n)$ composed with std strongly converge to the regular representation of $\mathbf{F}_r$.

Subsequent developments include results on higher dimensional representations of $\mathcal{U}(n)$ or S_n [BC24b, CGVTvH24, MdlS24b, Cas25], results related to Hayes' work [Hay22] on the Peterson-Thom conjecture [BC22, BC24a, Par24, CGVvH24], and work of Austin on annealed almost periodic entropy [Aus24].

Regarding the existence of finite dimensional unitary representations that strongly converge to the regular representation, results have been obtained for various classes of groups [LM25, MdlS24a, MT23] other than free groups. Most relevant to the current paper is the result of the first named author and Louder [LM25, Cor. 1.2] that for surface groups Γ as in this paper, there exists a sequence of $\{\phi_n \in \mathrm{Hom}(\Gamma, S_n)\}_{n=1}^{\infty}$ such that $\mathrm{std} \circ \phi_n$ strongly converge to λ. Theorem 1.1 shows not only that such homomorphisms exist, but that this is in fact the *typical* behavior.

The recent works [CGVTvH24, CGVvH24] introduced a new approach to strong convergence, called the *polynomial method*, that is based almost entirely on soft arguments, in contrast to previous works that were based on problem-specific analytic methods or delicate combinatorial estimates. This method has made it possible to establish strong convergence in previously inaccessible situations and to achieve improved quantitative results. These works form the basis for the generalized polynomial method that is developed in the present paper (see §1.1 and §3).

Hyperbolic surfaces

Whether there exists a sequence of closed hyperbolic surfaces $\{X_n\}_{n=1}^{\infty}$ with genera tending to ∞ such that $\lambda_1(X_n) \to \frac{1}{4}$ was an old question of Buser [Bus84]. In view of analogous results on optimal spectral gaps of random regular graphs [Alo86, Fri08, Pud15, Bor20] (for the state-of-the-art see the recent breakthrough [HMY25]), it is natural to approach this question through the study of random hyperbolic surfaces. We presently discuss previous results on three models of hyperbolic surfaces:

1. The Brooks–Makover (BM) model and variations, obtained by constructing a random cusped hyperbolic surface out of gluing m ideal triangles and then compactifying the cusps.

2. The Weil–Petersson (WP) random model that comes from the Weil–Petersson Kähler form on the moduli space of genus g hyperbolic surfaces.

3. Models of random degree n covering spaces.

We say events hold w.h.p. (with high probability) if they hold with probability tending to one as the relevant parameter $(m/g/n)$ tends to infinity. In random cover models we understand that λ_1 is measured ignoring the eigenvalues of the base surface.

The first result that random closed hyperbolic surfaces in the BM model have a uniform positive lower bound on λ_1 was due to Brooks–Makover [BM04]. The same result for WP random surfaces was obtained by Mirzakhani in [Mir13]. The explicit bound $\lambda_1 \geqslant \frac{3}{16} - o(1)$ w.h.p. for closed surfaces was first obtained in [MNP22] for the uniform random covering model. Following this, the same result was obtained in the WP model by Wu–Xue [WX22] and Lipnowski–Wright [LW24]. The latter was improved to $\lambda_1 \geqslant \frac{2}{9} - o(1)$ w.h.p. by Anantharaman–Monk [AM23].

The optimal bound $\lambda_1 \geqslant \frac{1}{4} - o(1)$ w.h.p. was finally achieved by Hide and the first named author [HM23] in a random cover model for *non-closed* hyperbolic surfaces. This suffices, by means of a compactification procedure, to settle Buser's conjecture in the affirmative. The same proof applies verbatim to a model of gluing hyperbolic triangles[5] and as a result, one obtains $\lambda_1 \geqslant \frac{1}{4} - o(1)$ w.h.p. in close cousins of the BM model for closed surfaces. In [LM25], the existence of covers of closed hyperbolic surfaces such that $\lambda_1 \geqslant \frac{1}{4} - o(1)$ was proved without a compactification procedure.

Very recently, it was shown in the impressive work of Anantharaman–Monk [AM25] that $\lambda_1 \geqslant \frac{1}{4} - o(1)$ w.h.p. in the WP model of closed surfaces. This shows that the optimal spectral gap property is in fact the typical behavior, in the sense of the Weil-Petersson measure, of closed hyperbolic surfaces as $g \to \infty$.

Despite the above advances, the question whether optimal spectral gaps $\lambda_1 \geqslant \frac{1}{4} - o(1)$ are also typical for random covering spaces has remained open. The reason is that strong convergence of uniform random homomorphisms of surface groups to S_n was outside the reach of previous methods. Thus [HM23, LM25] had to rely on reductions to the case of free groups: either by working with non-compact surfaces, or by embedding surface groups into an iterated sequence of extensions of centralizers of a free group which results in a complicated randomization that is far from uniform. The reason we can resolve this question here (Theorem 1.2) is that the new methods introduced in this paper resolve the corresponding strong convergence problem (Theorem 1.1).[6]

1.5 Paper organization

The remainder of this paper is organized as follows. In §2, we state some basic inequalities for polynomials and rational functions that will be used in the sequel. The proofs of Theorems 1.5, 1.7, and 1.8 are subsequently developed in §3, §4, and §5, respectively. Finally, we assemble the above ingredients in §6 to complete the proof of Theorem 1.1.

1.6 Notation

We write $\mathbf{N} = \{1, 2, 3, \ldots\}$, $\mathbf{Z}_+ = \{0, 1, 2, \ldots\}$, and $\mathbf{Z} = \{\ldots, -1, 0, 1, \ldots\}$. For $n \in \mathbf{N}$ we write $[n] = \{1, \ldots, n\}$. The falling Pochhammer symbol and double factorial are

$$(n)_m \stackrel{\text{def}}{=} n(n-1)\cdots(n-m+1), \qquad (2n-1)!! \stackrel{\text{def}}{=} (2n-1)(2n-3)\cdots 3 \cdot 1.$$

A set partition of $[n]$ is a collection of disjoint subsets whose union is $[n]$. We write $\mathrm{Part}([n])$ for the set partitions of $[n]$.

[5] Although it was not pointed out at the time, the proof of [HM23] works just as well for random covers of the quotient of $\mathbb{H}^2$ by reflections in the sides of an ideal hyperbolic triangle. In this case, the input is the result of Bordenave and Collins for random homomorphisms from the free product $\mathbf{Z}/2\mathbf{Z} * \mathbf{Z}/2\mathbf{Z} * \mathbf{Z}/2\mathbf{Z}$ to S_n. This is not exactly the same as the Brooks–Makover gluing, however.

[6] We emphasize that the methods of the present paper are completely independent of those used in [MNP22] to obtain the suboptimal bound $\lambda_1 \geqslant \frac{3}{16} - o(1)$ for the uniform random cover model, with the exception of the self-contained Proposition 4.1 that appears in [MNP22, Prop. 4.6].

For $f, g : \mathbf{N} \to \mathbf{R}$, we write $f = O(g)$ if there is a universal constant $C > 0$ — depending on no other parameters — such that for all $n \in \mathbf{N}$, $|f(n)| \leqslant Cg(n)$. If C depends on other parameters this is indicated with a subscript to the big O.

We denote by $f^{(m)}$ the mth derivative of a univariate function f, and denote by $\|f\|_{[a,b]} = \sup_{z\in[a,b]} |f(z)|$ the uniform norm on the interval $[a, b]$.

Groups and C^*-algebras

Let Λ be a finitely generated group with a finite generating set S. We denote by $|\gamma|$ the word length of $\gamma \in \Lambda$ with respect to S.

The group algebra $\mathbf{C}[\Lambda]$ has a natural involution $(\sum_\gamma \alpha_\gamma \gamma)^* = \sum_\gamma \bar{\alpha}_\gamma \gamma^{-1}$. Thus we can speak of self-adjoint elements of the group algebra; these become self-adjoint operators in any unitary representation. For any univariate polynomial h and $x \in \mathbf{C}[\Lambda]$, we interpret $h(x)$ in the obvious algebraic sense. We denote by $|x|$ the maximum word length of a group element in the support of x. We define

$$\|x\|_{C^*(\Lambda)} = \sup_\pi \|\pi(x)\|$$

for $x \in \mathbf{C}[\Lambda]$, where the supremum is taken over all unitary representations of Λ.

Let λ be the regular representation of Λ, and extend it linearly to $\mathbf{C}[\Lambda]$. The reduced C^*-algebra $C^*_{\mathrm{red}}(\Lambda)$ of Λ is the norm-closure of $\{\lambda(x) : x \in \mathbf{C}[\Lambda]\}$.

For $M \in \mathrm{M}_n(\mathbf{C})$, let $\mathrm{Tr}\, M$ be the trace and $\mathrm{tr}\, M = \frac{1}{n} \mathrm{Tr}\, M$ be the normalized trace. We denote by τ the canonical trace on $C^*_{\mathrm{red}}(\Lambda)$, i.e., $\tau(a) = \langle \delta_e, a\delta_e \rangle$, and by

$$\|a\|_{L^2(\tau)} = \tau(a^*a)^{1/2}$$

for $a \in C^*_{\mathrm{red}}(\Lambda)$. In particular, $\|\lambda(x)\|_{L^2(\tau)} = (\sum_\gamma |\alpha_\gamma|^2)^{1/2}$ for $x = \sum_\gamma \alpha_\gamma \gamma$.

Acknowledgments

We thank Bram Petri and Joe Thomas for discussions related to this work.

This work was conducted while D.P. and R.v.H. were members of the Institute of Advanced Study in Princeton, NJ, which is gratefully acknowledged for providing a fantastic mathematical environment.

Funding (M.M.): This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 949143).

Funding (D.P.): This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 850956), by the Israel Science Foundation, ISF grants 1140/23, as well as National Science Foundation under Grant No. DMS-1926686.

Funding (R.v.H.): This work was supported in part by NSF grant DMS-2347954.

2 Analysis of polynomials and rational functions

We begin by recalling an elaboration of the Markov brothers inequality that is the basic principle used in the polynomial method [CGVTvH24]. The following convenient formulation may be found in [MdlS24b, Lem. 4.2].

Lemma 2.1. *For every real polynomial P of degree at most q and every $k \in \mathbf{Z}_+$,*

$$\sup_{t\in\left[0,\frac{1}{2q^2}\right]} |P^{(k)}(t)| \leqslant \frac{2^{2k+1}q^{4k}}{(2k-1)!!} \sup_{n\geqslant q^2} \left|P\left(\frac{1}{n}\right)\right|.$$

We will require a variant of this principle for rational functions.

Lemma 2.2. *Let P, Q be real polynomials of degree at most $q \in \mathbf{N}$, and let*

$$\Phi(t) = \frac{P(t)}{Q(t)}.$$

Assume that there is a constant $C > 1$ so that

$$\left|P\left(\frac{1}{n}\right)\right| \leqslant C, \qquad\qquad C^{-1} \leqslant Q\left(\frac{1}{n}\right) \leqslant C$$

for all $n \in \mathbf{N}$ with $n \geqslant (Cq)^C$. Then there is $C' = C'(C) > 1$ so that for all $k \leqslant Cq$

$$\sup_{t\in[0,(C'q)^{-C'}]} |\Phi^{(k)}(t)| \leqslant (C'q)^{C'k}.$$

Proof. Applying Lemma 2.1 with $q \leftarrow (Cq)^C$ yields

$$\sup_{t\in[0,(2Cq)^{-2C}]} |P^{(k)}(t)| \leqslant 2C(2Cq)^{4Ck}, \qquad \sup_{t\in[0,(2Cq)^{-2C}]} |Q^{(k)}(t)| \leqslant 2C(2Cq)^{4Ck},$$

for all $k \in \mathbf{N}$. On the other hand, every $t \in [0, (2Cq)^{-2C-1}]$ is within distance at most $(2Cq)^{-4C-2}$ from some point $\frac{1}{n}$ with $n \geqslant (Cq)^C$. Thus

$$\inf_{t\in[0,(2Cq)^{-2C-1}]} Q(t) \geqslant C^{-1} - (2Cq)^{-4C-2} \sup_{t\in[0,(2Cq)^{-2C-1}]} |Q^{(1)}(t)| \geqslant (2C)^{-1}.$$

The conclusion now follows from the chain rule. Indeed, we can first estimate

$$\sup_{t\in[0,(2Cq)^{-2C-1}]} \left|\left(\frac{1}{Q}\right)^{(k)}(t)\right| \leqslant (2C)^{2k+1}(2Cq)^{4Ck}k!$$

as in the last part of [MdlS24b, proof of Lemma 7.2]. Thus

$$|\Phi^{(k)}(t)| = \left|\sum_{i=0}^{k} \binom{k}{i} P^{(k-i)}(t) \left(\frac{1}{Q}\right)^{(i)} (t)\right| \leqslant 2^k (2C)^{2k+2} (2Cq)^{4Ck} k!$$

for $t \in [0, (2Cq)^{-2C-1}]$. The conclusion follows as $k! \leqslant (Cq)^k$ for $k \leqslant Cq$. $\square$

We finally recall a simple fact about trigonometric polynomials.

Lemma 2.3. *Let $f(\theta) = \sum_{k=0}^{q} a_k \cos(k\theta)$ and $m \in \mathbf{N}$. Then*

$$\sum_{k=1}^{q} k^{m-1} |a_k| \leqslant 4 \|f^{(m)}\|_{[0,2\pi]}.$$

Proof. As $(k^m a_k)_{0 \leqslant k \leqslant q}$ are the Fourier coefficients of $f^{(m)}$, we have

$$\sum_{k=1}^{q} k^{m-1} |a_k| \leqslant \left(\sum_{k=1}^{q} \frac{1}{k^2}\right)^{1/2} \left(\sum_{k=1}^{q} |k^m a_k|^2\right)^{1/2} \leqslant \sqrt{\frac{\pi^2}{6}} \|f^{(m)}\|_{L^2[0,2\pi]}$$

by Cauchy-Schwarz and Parseval. We conclude using $\|f\|_{L^2[0,2\pi]} \leqslant \sqrt{2\pi} \|f\|_{[0,2\pi]}$. $\square$

3 The generalized polynomial method

The aim of this section is to prove Theorem 1.5. The reader who is new to the polynomial method is encouraged to review [CGVTvH24, §2] for an introduction, before proceeding to the details of the proof.

In this section, we always assume that Assumptions 1.3 and 1.4 hold. We denote by C the constant in Assumption 1.3, and by $r = |S|$ the size of the generating set of Λ.

The key step in the proof is to establish the following master inequality.

Lemma 3.1. *Fix a self-adjoint $x \in \mathbf{C}[\Lambda]$ and let $K = \|x\|_{C^*(\Lambda)}$. Then*

$$|u_1(h(x))| \leqslant r(6Cq|x|)^{6C} \|h\|_{[-K,K]}$$

and

$$\left|\mathbb{E}\big[\operatorname{tr} h(\pi_n(x))\big] - \tau\big(h(\lambda(x))\big) - \frac{1}{n} u_1(h(x))\right| \leqslant \frac{2r^2(6Cq|x|)^{12C}}{n^2} \|h\|_{[-K,K]}$$

for all $n \geqslant 1$ and every real polynomial h of degree at most q.

Proof. We assume without loss of generality that $|x| \geqslant 1$ and $q \geqslant 1$ (otherwise $h(x)$ is a multiple of the identity, so both sides of the inequalities vanish). Let us write

$$h(x) = \sum_{\gamma \in \Lambda} \alpha_\gamma \gamma.$$

As $|h(x)| \leqslant q|x|$, there are at most $(2r+1)^{q|x|} \leqslant (3r)^{q|x|}$ nonzero coefficients α_g. Applying Assumption 1.3 separately to each element of the support of $h(x)$ yields

$$\left|\mathbb{E}\big[\operatorname{tr} h(\pi_n(x))\big] - f_h\left(\frac{1}{n}\right)\right| \leqslant \frac{(3Cq|x|)^{3Cq|x|}}{n^{q|x|+2}} \sum_{\gamma\in\Lambda} |\alpha_\gamma|$$

for all $n \geqslant C(3q|x|)^C$, where we define the polynomial

$$f_h(t) = \sum_{k=0}^{q|x|+1} u_k(h(x))\, t^k$$

and we used that $q|x| + 2 \leqslant 3q|x|$. As

$$\sum_\gamma |\alpha_\gamma| \leqslant (3r)^{q|x|/2} \|h(\lambda(x))\|_{L^2(\tau)} \leqslant (3r)^{q|x|/2} \|h\|_{[-K,K]} \tag{3.1}$$

by Cauchy-Schwarz and as $\|\lambda(x)\| \leqslant K$, we readily estimate

$$\left|\mathbb{E}\big[\operatorname{tr} h(\pi_n(x))\big] - f_h\left(\frac{1}{n}\right)\right| \leqslant \frac{1}{n^2} \|h\|_{[-K,K]} \tag{3.2}$$

for $n \geqslant (3r)^{1/2}(3Cq|x|)^{3C}$. Now note that (3.2) implies the *a priori* bound

$$\left|f_h\left(\frac{1}{n}\right)\right| \leqslant \big|\mathbb{E}\big[\operatorname{tr} h(\pi_n(x))\big]\big| + \frac{1}{n^2}\|h\|_{[-K,K]} \leqslant 2\|h\|_{[-K,K]}.$$

Moreover, $u_k(h(x))$ are real as $h(x)$ is self-adjoint, so f_h is a real polynomial. We can therefore apply Lemma 2.1 with $q^2 \leftarrow (3r)^{1/2}(3Cq|x|)^{3C}$ to estimate

$$\begin{aligned}
\|f_h'\|_{[0,r^{-1/2}(6Cq|x|)^{-3C}]} &\leqslant r(6Cq|x|)^{6C}\|h\|_{[-K,K]},\\
\|f_h''\|_{[0,r^{-1/2}(6Cq|x|)^{-3C}]} &\leqslant r^2(6Cq|x|)^{12C}\|h\|_{[-K,K]}.
\end{aligned}$$

We can now readily conclude the proof. Indeed, as $u_1(h(x)) = f_h'(0)$, the first part of the statement is immediate. On the other hand, Taylor expanding f_h yields

$$\left|f_h\left(\frac{1}{n}\right) - u_0(h(x)) - \frac{1}{n}u_1(h(x))\right| \leqslant \frac{1}{n^2}\|f_h''\|_{[0,\frac{1}{n}]} \leqslant \frac{r^2(6Cq|x|)^{12C}}{n^2}\|h\|_{[-K,K]}$$

for $n \geqslant r^{1/2}(6Cq|x|)^{3C}$, so the second part follows in this case using (3.2) and that

$u_0(x) = \tau(\lambda(x))$ by Assumption 1.3. Finally, when $n < r^{1/2}(6Cq|x|)^{3C}$, we have

$$\left|\mathbb{E}\big[\operatorname{tr} h(\pi_n(x))\big] - \tau\big(h(\lambda(x))\big) - \frac{1}{n}u_1(h(x))\right| \leqslant 2\|h\|_{[-K,K]} + \frac{1}{n}|u_1(h(x))| \leqslant \frac{2r^2(6Cq|x|)^{12C}}{n^2}\|h\|_{[-K,K]}$$

using the triangle inequality, the bound on $|u_1(h(x))|$, and $1 \leqslant \frac{r^{1/2}(6Cq|x|)^{3C}}{n}$. □

With the statement of Lemma 3.1 in hand, the remainder of the argument is now precisely as in [CGVTvH24]. We first extend Lemma 3.1 to smooth h.

Corollary 3.2. *Fix a self-adjoint $x \in \mathbf{C}[\Lambda]$ and let $K = \|x\|_{C^*(\Lambda)}$. Then $\nu(h) \overset{def}{=} u_1(h(x))$ extends to a compactly supported distribution, and*

$$\left|\mathbb{E}\big[\operatorname{tr} h(\pi_n(x))\big] - \tau\big(h(\lambda(x))\big) - \frac{1}{n}\nu(h)\right| \leqslant \frac{8r^2(6C|x|)^{12C}}{n^2}\|f^{(m)}\|_{[0,2\pi]}$$

for all $n \geqslant 1$ and $h \in C^\infty(\mathbf{R})$. Here $m = 1 + \lceil 12C \rceil$ and $f(\theta) = h(K\cos\theta)$.

Proof. That ν extends to a compactly supported distribution follows directly from the first part of Lemma 3.1 and [CGVTvH24, Lem. 4.7].

Now assume first that h is a polynomial of degree q. Then we can uniquely write

$$h(t) = \sum_{k=0}^{q} a_k T_k\left(\frac{t}{K}\right),$$

where T_k is the Chebyshev polynomial defined by $T_k(\cos(\theta)) = \cos(k\theta)$. Applying the second part of Lemma 3.1 separately to each Chebyshev polynomial yields

$$\left|\mathbb{E}\big[\operatorname{tr} h(\pi_n(x))\big] - \tau\big(h(\lambda(x))\big) - \frac{1}{n}u_1(h(x))\right| \leqslant \frac{2r^2(6C|x|)^{12C}}{n^2}\sum_{k=1}^{q} k^{12C}|a_k|.$$

The conclusion now follows from Lemma 2.3 in the case that h is a polynomial. Since the resulting bound does not depend on the degree of h and as polynomials are dense in $C^\infty(\mathbf{R})$, the conclusion extends by continuity to any $h \in C^\infty(\mathbf{R})$. □

We can now state a quantitative version of Theorem 1.5.

Theorem 3.3. *Fix a self-adjoint $x \in \mathbf{C}[\Lambda]$ and let $K = \|x\|_{C^*(\Lambda)}$. Then*

$$\mathbb{P}\big[\|\pi_n(x)\| \geqslant \|\lambda(x)\| + \varepsilon\big] \leqslant \frac{r^2}{n}\left(\frac{c|x|K}{\varepsilon}\right)^c$$

for every $n \geqslant 1$ and $\varepsilon > 0$. Here c is a constant that depends only on C.

Proof. By [CGVTvH24, Lem. 4.10], there exists $h \in C^\infty(\mathbf{R})$ with values in $[0,1]$ so that

1. $h(z) = 0$ for $|z| \leqslant \|\lambda(x)\| + \frac{\varepsilon}{2}$ and $h(z) = 1$ for $|z| \geqslant \|\lambda(x)\| + \varepsilon$;
2. $\|f^{(m)}\|_{[0,2\pi]} \leqslant (\frac{cK}{\varepsilon})^c$ for $f(\theta) = h(K\cos\theta)$.

Moreover, Assumption 1.4 and [CGVTvH24, Lem. 4.9] yield $\operatorname{supp}\nu \subseteq [-\|\lambda(x)\|, \|\lambda(x)\|]$, where ν is defined in Corollary 3.2. Thus $\tau(h(\lambda(x))) = \nu(h) = 0$, so that

$$\mathbb{P}\big[\|\pi_n(x)\| \geqslant \|\lambda(x)\| + \varepsilon\big] \leqslant \mathbb{E}\big[\operatorname{Tr} h(\pi_n(x))\big] \leqslant \frac{8r^2(6C|x|)^{12C}}{n}\left(\frac{cK}{\varepsilon}\right)^c$$

using Corollary 3.2. This concludes the proof. □

The statement of Theorem 1.5 now follows immediately from Theorem 3.3 (if x is not self-adjoint, we simply apply Theorem 3.3 to x^*x). However, Theorem 3.3 provides much stronger information, since it even yields a polynomial rate

$$\|\pi_n(x)\| \leqslant \|\lambda(x)\| + O_{\mathrm{P}}(n^{-1/c}).$$

Remarks

1. Assumption 1.3 ensures that $\mathbb{E}[\operatorname{tr} h(\pi_n(x))]$ can be approximated by a polynomial of $\frac{1}{n}$. At first sight, however, it is somewhat surprising that this suffices for the polynomial method. As is explained in [CGVTvH24, §2.3], the key feature of this method is that it captures certain delicate cancellations between the coefficients of $h(x) = \sum_\gamma \alpha_\gamma \gamma$. On the other hand, Assumption 1.3 only applies to each γ separately and is therefore unable to preserve cancellations.

   The reason this does not matter is that the price for disregarding cancellations is exponential in $q|x|$ (see (3.1)), and can therefore be absorbed by the error term in Assumption 1.3 by expanding to order $\sim q|x|$. We can therefore first approximate the spectral statistics by high degree polynomials in the most naive manner, and then proceed to capture the cancellations in the latter using the polynomial method. This is crucial for applications, since we do not know how to establish *a priori* bounds as in Assumption 1.3 that capture cancellations.

2. While we developed the argument above in the group setting, the method is very robust and applies to many other situations. The only part of the argument that relied on the group structure is (3.1). In other situations, one can achieve a similar conclusion by applying a classical result of V. Markov which states that

   $$\sum_{k=0}^{q} |a_k| \leqslant e^{q/K} \|h\|_{[-K,K]}$$

   for every real polynomial $h(t) = \sum_{k=0}^q a_k t^k$ and $K > 0$ [Tim63, §2.6, Eq. (9)], at the expense of a less explicit dependence of the constants on the choice of x.

4 Effective approximation by polynomials

The aim of this section is to prove Theorem 1.7.

Throughout this section, $C > 0$ will denote a constant that may depend on the genus g, which can change from line to line but will only increase finitely many times.

4.1 Background

Representation theory of the symmetric group

A (number[7]) partition of n is a non-increasing sequence of non-negative integers $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_{\ell(\lambda)})$ with

$$\sum_{i=1}^{\ell(\lambda)} \lambda_i = n.$$

In this case we write $\lambda \vdash n$ and say λ has size n, $|\lambda| = n$. We think of partitions and Young diagrams (YDs) interchangeably in the sequel. If λ is a YD we write $\lambda^\vee$ for the conjugate YD obtained by interchanging rows and columns. We write b_λ for the number of boxes outside the first row of λ and $b_\lambda^\vee \stackrel{\text{def}}{=} b_{\lambda^\vee}$ the number of boxes outside the first column.

The equivalence classes of irreducible representations (irreps) of S_n are parameterized (using Young symmetrizers, for example) by the YDs $\lambda \vdash n$. Given $\lambda \vdash n$, we write χ_λ for the character of the associated irrep and $d_\lambda = \chi_\lambda(1)$ for its dimension. We write V^λ for the corresponding S_n-module. As $\chi_{\lambda^\vee} = \text{sign} \cdot \chi_\lambda$ we have $d_\lambda = d_{\lambda^\vee}$.

If one YD μ is contained in another λ, we say $\mu \subset \lambda$, and if they differ by k boxes, we write $\mu \subset_k \lambda$. For YDs $\mu \subset_k \lambda$, a (standard) tableau of shape λ/μ is a filling of the k boxes of λ not in μ by the numbers in $[k]$ such that the numbers in each row (resp. column) are increasing from left to right (resp. top to bottom). We write $\text{Tab}(\lambda/\mu)$ for the collection of such tableaux. It is obvious that when $\mu \subset_k \lambda$,

$$|\text{Tab}(\lambda/\mu)| \leqslant k!.$$

We use the fact that every irrep of S_n can be realized over the reals and therefore as orthogonal matrices (so no complex conjugates appear in our calculations).

The Witten zeta function

The Witten zeta function of S_n is defined by

$$\zeta(s; S_n) \stackrel{\text{def}}{=} \sum_{\lambda \vdash n} \frac{1}{d_\lambda^s}.$$

This appears as a normalization factor in the uniform measure on $\text{Hom}(\Gamma_g, S_n)$ and hence is important in the sequel. By a result of Liebeck and Shalev [LS04, Prop. 2.5]

[7] As opposed to a set partition.

(independently, Gamburd [Gam06, Prop. 4.2]), for real $s > 0$,

$$\zeta(s; S_n) = 2 + O_s\left(n^{-s}\right) \tag{4.1}$$

as $n \to \infty$. We also have the following effective tail bound [MNP22, Prop. 4.6].

Proposition 4.1. *For all $s > 0$, there is $\kappa = \kappa(s) > 1$ such that*

$$\sum_{\substack{\lambda \vdash n \\ b_\lambda, b_\lambda^\vee \geqslant b}} \frac{1}{d_\lambda^s} \leqslant \left(\frac{\kappa b^{2s}}{(n-b^2)^s}\right)^b \tag{4.2}$$

for all $b, n \in \mathbf{N}$ with $b^2 \leqslant \frac{n}{3}$.

Expansion of $\mathbb{E}_{g,n}[\mathsf{fix}_\gamma]$ by expected numbers of embeddings

In the following, ϕ is a uniformly random element of $\mathrm{Hom}(\Gamma_g, S_n)$. We now recap on the language of [MP23]. This is just a formalism for expressing the facts:

- $\mathbb{E}_{g,n}[\mathsf{fix}_\gamma]$ is n times the probability that $\phi(\gamma)$ fixes a given point, say $1 \in [n]$, by invariance under S_n conjugation.
- The event $\mathcal{E}_\gamma$ that $\phi(\gamma)$ fixes 1 can be partitioned into sub-events in a variety of different ways.

In [MP23] these sub-events are chosen more carefully than we have to here.

We use the standard set of generators $a_1, \ldots, b_g$ of Γ_g from (1.2). The partitioning is described by reference to the Schreier graph

$$X_\phi \stackrel{\text{def}}{=} \text{Schreier}\left(\{\phi(a_1), \ldots, \phi(b_g)\}, [n]\right)$$

defined by the action of $\phi(a_1), \ldots, \phi(b_g)$ on $[n]$, i.e., there is a directed edge from i to j for each $f \in \{a_1, \ldots, b_g\}$ such that

$$\phi(f)[i] = j.$$

Every directed edge is colored by the generator from which it arises.[8]

Given $\gamma \in \Gamma_g$, let γ' be a conjugate of γ whose shortest representing word in the generators is cyclically reduced. We have $|\gamma'| \leqslant |\gamma|$ and

$$\mathbb{E}_{g,n}[\mathsf{fix}_\gamma] = \mathbb{E}_{g,n}[\mathsf{fix}_{\gamma'}],$$

so without loss of generality in the following suppose γ is cyclically reduced.

[8]The definition of X_ϕ given here is the 1-skeleton of the X_ϕ of [MP23], which includes n additional 2-cells to form a closed surface.

Let F_{2g} denote the free group on $a_1, \ldots, b_g$. Let C_γ denote a circle subdivided into $|\gamma|$ intervals (viewed as an undirected graph). Direct and label these intervals by $a_1, \ldots, b_g$ according to an expression of γ of length $|\gamma|$ in these generators — the direction indicates whether a generator or its inverse appears at a given location.

Let $\mathcal{R}$ denote the collection of surjective labeled-graph morphisms

$$r : C_\gamma \twoheadrightarrow W_r$$

such that:

1. W_r is *folded* in the sense that every vertex has at most one incoming f-labeled half-edge and at most one outgoing f-labeled half-edge, for each $f \in \{a_1, \ldots, b_g\}$. (Note that the assumption γ is cyclically reduced means C_γ itself is folded.)

2. Every path in W_r spelling out an element of F_{2g} that is in the kernel of $F_{2g} \to \Gamma_g$ is closed.

In the language of [MP22, MP23] these W_r are (very special cases of) *tiled surfaces*.[9]

Any homomorphism of folded labeled graphs $C_\gamma \to X_\phi$ factors uniquely as a surjective morphism followed by an injective one, namely

$$C_\gamma \overset{r}{\twoheadrightarrow} W_r \hookrightarrow X_\phi,$$

for unique $r \in \mathcal{R}$. In the language of [MP23], this factoring property means $\mathcal{R}$ is a *resolution* of C_γ.

Remark 4.2 (Comparison to [MP23]). Although the resolution we define here might be the most natural, the methods of [MP23] leading to [MP23, Thm. 1.2] — not proved here — only work on W_r who have, very roughly speaking, geodesic boundary. So in [MP23] great care is taken over finding a finer resolution whose boundary components have desired properties. For our purposes, the simplest resolution suffices.

We have

$$|\mathcal{R}| \leqslant |\gamma|! \tag{4.3}$$

since each element is defined by the partition of the vertices of C_γ given by the fibers of the surjective immersion, and there are $|\gamma|$ vertices. By [MP23, Lem. 2.7 and 2.9]

$$\mathbb{E}_{g,n}[\mathsf{fix}_\gamma] = \sum_{r \in \mathcal{R}} \mathbb{E}_n^{\mathrm{emb}}(W_r), \tag{4.4}$$

where

$$\mathbb{E}_n^{\mathrm{emb}}(W_r) = \mathbb{E}_{\phi \in \mathbb{X}_{g,n}}\{\,\#\text{ injective morphisms } W_r \to X_\phi\,\}.$$

In the sequel, for W_r as above we write $\mathfrak{v}(W_r), \mathfrak{e}(W_r)$ for the number of vertices and edges of W_r. For $f \in \{a_1, \ldots, b_g\}$ we write $\mathfrak{e}_f(W_r)$ for the number of f-labeled edges of

[9] In (ibid.) it is a canonical thickening of the graph that is thought of as the surface.

W_r. With things as before, clearly

$$\mathfrak{v}(W_r), \mathfrak{e}(W_r), \mathfrak{e}_f(W_r) \leqslant |\gamma|.$$

Expected number of embeddings, prior results

In the remainder of §4, we restrict to $g = 2$ for simplicity of exposition. We now use a, b, c, d for a_1, b_1, a_2, b_2. The only things that depend on g in a non-obvious way are constants and we indicate how these depend on g throughout. All integrals in the rest of this section are taken with respect to uniform probability measures.

Let Y denote some fixed W_r from our resolution $\mathcal{R}$. Let $\mathfrak{v} = \mathfrak{v}(Y)$, $\mathfrak{e}_f = \mathfrak{e}_f(Y)$. In the following we label the vertices of Y injectively by $1, \ldots, \mathfrak{v}$.

For $g_f \in S_n$, $f \in \{a, b, c, d\}$, we say g_f *obey* Y if the fixed labeling of the vertices of Y induces an embedding

$$Y \hookrightarrow \text{Schreier}\left(\{g_f\}, [n]\right),$$

or in simple terms, if Y has an f-colored directed edge from a vertex labeled i to vertex labeled j, then $g_f(i) = j$.

We have by [MP23, (2.1), (5.6), and proof of Prop. 5.1], since $\mathfrak{v}(Y) \leqslant q$, for $n \geqslant q$

$$\mathbb{E}_n^{\text{emb}}(Y) = \frac{1}{\zeta(2; S_n)} \frac{(n)_{\mathfrak{v}}}{\prod_{f \in a,b,c,d} (n)_{\mathfrak{e}_f}} \sum_{\lambda \vdash n} d_\lambda \Theta_\lambda\left(Y\right),$$

where

$$\Theta_\lambda\left(Y\right) \stackrel{\text{def}}{=} \left(\prod_{f \in a,b,c,d} (n)_{\mathfrak{e}_f}\right) \int_{g_f \in S_n} \mathbf{1}\{g_f \text{ obey } Y\} \chi_\lambda\left([g_a, g_b][g_c, g_d]\right) \tag{4.5}$$

(there is a change of normalization since we pass between integrals over different groups: in the notation of *(ibid.)* $|G_f| = (n - \mathfrak{e}_f)!$).

Remark 4.3. In [MP23] the definition of $\Theta_\lambda\left(Y\right)$ depends on an auxiliary labeling $\mathcal{J}_n : Y^{(0)} \to \{n - \mathfrak{v} + 1, \ldots, n\}$. This is not needed here, as the definition of $\Theta_\lambda\left(Y\right)$ in (4.5) is the same for any labeling by invariance of the uniform measure on S_n by conjugation.

From [MP23, Prop. 5.8]

$$\Theta_\lambda(Y) = \sum_{\nu \subset_{\mathfrak{v}} \lambda} d_\nu \sum_{\substack{\nu \subset \mu_f \subset_{\mathfrak{e}_f} \lambda \\ f \in \{a,b,c,d\}}} \frac{1}{d_{\mu_a} d_{\mu_b} d_{\mu_c} d_{\mu_d}} \Upsilon_n\left(\left\{\sigma_f^{\pm}, \tau_f^{\pm}\right\}, \nu, \left\{\mu_f\right\}, \lambda\right), \tag{4.6}$$

where

$$\Upsilon_n\left(\left\{\sigma_f^{\pm}, \tau_f^{\pm}\right\}, \nu, \left\{\mu_f\right\}, \lambda\right) \stackrel{\text{def}}{=} \sum_{\substack{r_f^+, r_f^- \in \text{Tab}\left(\mu_f / \nu\right) \\ s_f, t_f \in \text{Tab}\left(\lambda / \mu_f\right)}} \mathcal{M}\left(\left\{\sigma_f^{\pm}, \tau_f^{\pm}, r_f^{\pm}, s_f, t_f\right\}\right) \tag{4.7}$$

and $\mathcal{M}\left(\left\{\sigma_f^\pm,\tau_f^\pm,r_f^\pm,s_f,t_f\right\}\right)$ is a product of matrix coefficients of unitary operators on unit vectors as in [MP23, eq. (5.15)]. All we need here about this product is that

$$\left|\mathcal{M}\left(\left\{\sigma_f^\pm,\tau_f^\pm,r_f^\pm,s_f,t_f\right\}\right)\right| \leqslant 1. \tag{4.8}$$

4.2 Tail estimate for $\mathbb{E}_n^{\mathrm{emb}}(W_r)$

In this section we prove:

Proposition 4.4. *There is $C>0$ such that if $\mathfrak{v}(Y)\leqslant q$ and $n\geqslant 28q^2$ then*

$$\mathbb{E}_n^{\mathrm{emb}}(Y) = \frac{1}{\zeta(2;S_n)}\frac{(n)_{\mathfrak{v}}}{\prod_{f\in a,b,c,d}(n)_{\mathfrak{e}_f}}\sum_{\substack{\lambda\vdash n\\ \min(b_\lambda,b_\lambda^\vee)\leqslant 4q}} d_\lambda\Theta_\lambda(Y) + O\left((Cq)^{Cq}n^{-q}\right).$$

Proof. Using $|\mathrm{Tab}\left(\mu_f/\nu\right)|,|\mathrm{Tab}\left(\lambda/\mu_f\right)|\leqslant \mathfrak{v}!\leqslant q!$ together with (4.8) in (4.6) we obtain

$$\Theta_\lambda(Y)\leqslant (Cq)^{Cq}\sum_{\nu\subset_{\mathfrak{v}}\lambda}\frac{1}{d_\nu^3}.$$

Note if $b_\lambda>4q$ then $b_\nu\geqslant b_\lambda-\mathfrak{v}>3q$. We have then

$$\sum_{\substack{\lambda\vdash n\\ b_\lambda,b_\lambda^\vee>4q}} d_\lambda\Theta_\lambda(Y)\leqslant (Cq)^{Cq}\sum_{\lambda\vdash n}\sum_{\substack{\nu\subset_{\mathfrak{v}}\lambda\\ b_\nu,b_\nu^\vee>3q}}\frac{d_\lambda}{d_\nu^3}.$$

As in [MP23, proof of Lemma 5.23] the above is

$$\leqslant (n)_{\mathfrak{v}}(Cq)^{Cq}\sum_{\substack{\nu\vdash n-\mathfrak{v}\\ b_\nu,b_\nu^\vee>3q}}\frac{1}{d_\nu^2}. \tag{4.9}$$

Hence by (4.2)

$$\sum_{\substack{\nu\vdash n-\mathfrak{v}\\ b_\nu,b_\nu^\vee>3q}}\frac{1}{d_\nu^2}\leqslant C\left(\frac{(Cq)^4}{(n-\mathfrak{v}-9q^2)^2}\right)^{3q}\leqslant C(Cq)^{Cq}n^{-6q} \tag{4.10}$$

in e.g. $n\geqslant 28q^2$. (Note that to use (4.2) for $b=3q$ as above we need $\mathfrak{v}+3(3q)^2\leqslant n$.)

Combining all previous arguments gives

$$\mathbb{E}_n^{\text{emb}}(Y) = \frac{1}{\zeta(2;S_n)} \frac{(n)_{\mathfrak{v}}}{\prod_{f\in a,b,c,d}(n)_{\mathfrak{e}_f}} \sum_{\substack{\lambda \vdash n \\ \min(b_\lambda, b_\lambda^\vee) \leqslant 4q}} d_\lambda \Theta_\lambda(Y) + O\left(\frac{1}{\zeta(2;S_n)} \frac{(n)_{\mathfrak{v}}^2}{\prod_{f\in a,b,c,d}(n)_{\mathfrak{e}_f}} (Cq)^{Cq} n^{-6q} \right).$$

The denominator Pochhammer symbols in the error are $\geqslant 1$ and numerators $\leqslant n^{2q}$. Hence (also using (4.1)) the whole error is on the order of

$$(Cq)^{Cq} n^{-q}$$

as required.

(For general $g \geqslant 2$, the number of tableaux being summed over in the combination of (4.6) and (4.7) is $8g$, hence C depends linearly on g, the summand of (4.9) is $\frac{1}{d_\nu^{2g-2}} \leqslant \frac{1}{d_\nu^2}$ and the estimate (4.10) can be kept the same. The rest of the proof is the same, changing the product $\prod_{f\in a,b,c,d}(n)_{\mathfrak{e}_f}$ to a product over $2g$ generators and $\zeta(2;S_n)$ to $\zeta(2g-2;S_n)$.) □

4.3 Rationality of contributions from fixed representations

Given a YD $\lambda \vdash b$, for $n \geqslant 2b$ we write $\lambda^+(n)$ for the YD of size n with λ outside the first row. Our aim here is to prove the following rationality result.

Proposition 4.5. *Let $b \in \mathbf{Z}_+$, $n \geqslant \mathfrak{v}(Y) + 9b$, and $\lambda \vdash b$. As a function of n, $\Theta_{\lambda^+(n)}(Y)$ agrees with a rational function of n with coefficients in $\mathbf{Q}$ whose denominator can be taken to be*

$$(n)_{\mathfrak{v}+2b}^5.$$

The proof of Proposition 4.5 uses a result of Cassidy [Cas23]. Let $\{e_i\}_{i=1}^n$ denote the standard orthonormal basis of $\mathbf{C}^n$ with its standard inner product. Given a set partition π of $[2b]$, define the following endomorphism P_π of $(\mathbf{C}^n)^{\otimes b}$ by

$$\langle P_\pi(e_{i_1} \otimes \cdots \otimes e_{i_b}), e_{i_{b+1}} \otimes \cdots \otimes e_{i_{2b}} \rangle \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } i_p = i_q \text{ iff } p \text{ and } q \text{ in same block of } \pi \\ 0 & \text{else.} \end{cases}$$

Denote by $\mathrm{Part}([2b])$ the set partitions of $[2b]$, and by ι the inclusion $S_b \hookrightarrow \mathrm{Part}([2b])$ defined by viewing an element of S_b as a matching between two horizontal rows of b elements, the first row labeled by $[b]$ and the second row labeled by $[2b]\backslash[b]$. Say $\pi_1 \leqslant \pi_2$ if the blocks of π_1 subdivide those of π_2. Let $\mathrm{SubPerm}(b) \subseteq \mathrm{Part}([2b])$ denote "sub-permutations": set partitions π of $[2b]$ such that $\pi \leqslant \iota(\sigma)$ for some $\sigma \in S_b$, namely, such that every block of π contains at most one element from each of the two horizontal rows. We write $|\pi|$ for the number of blocks of π. Here is Cassidy's formula.[10] (This differs from the formula in [Cas23] by Möbius inversion and orthogonality.)

Theorem 4.6 (Cassidy [Cas23, Thm. 1.2])**.** *For $b \in \mathbf{Z}_+$, $n \geqslant 2b$, and $\lambda \vdash b$,*

$$\mathfrak{p}_\lambda = (-1)^b d_{\lambda^+(n)} \sum_{\pi \in \mathrm{SubPerm}(b)} \frac{(-1)^{|\pi|}}{(n)_{|\pi|}} \left(\sum_{\substack{\tau \in S_b \\ \pi \leqslant \iota(\tau)}} \chi_\lambda(\tau) \right) P_\pi \tag{4.11}$$

is the orthogonal projection in $(\mathbf{C}^n)^{\otimes b}$ onto d_λ copies of $V^{\lambda^+(n)}$ as an S_n-representation.

Proof of Proposition 4.5. Here we evaluate $\Theta_{\lambda^+(n)}(Y)$ differently to [MP23]. If

$$\rho_b : S_n \to \mathrm{End}\left((\mathbf{C}^n)^{\otimes b} \right)$$

is the b^{th} tensor power of the standard permutation representation of S_n, we can rewrite (4.5) as

$$\Theta_{\lambda^+(n)}(Y) = \frac{\prod_{f \in a,b,c,d} (n)_{\mathfrak{e}_f}}{d_\lambda} \int_{g_f \in S_n} \mathbf{1}\{g_f \text{ obey } Y\} \mathrm{Tr}_{(\mathbf{C}^n)^{\otimes b}} \left(\mathfrak{p}_\lambda \rho_b \left([g_a, g_b][g_c, g_d] \right) \right)$$

where $\mathfrak{p}_\lambda$ is the projection operator from Theorem 4.6.

From (4.11), $\mathfrak{p}_\lambda$ is a linear combination of partition operators P_π; all the coefficients are rational functions of n in $n \geqslant 2b$ and can be put over a common denominator of $(n)_{2b}$. Here we used that $d_{\lambda^+(n)}$ is a polynomial of n in this range by the hook-length formula [FRT54]. From this we learn $\Theta_{\lambda^+(n)}(Y)$ is a linear combination of

$$\int_{g_f \in S_n} \mathbf{1}\{g_f \text{ obey } Y\} \mathrm{Tr}_{(\mathbf{C}^n)^{\otimes b}} \left(P_\pi \rho_b \left([g_a, g_b][g_c, g_d] \right) \right) \tag{4.12}$$

with rational coefficients in n, with common denominator $(n)_{2b}$.

We can now expand

$$\mathrm{Tr}_{(\mathbf{C}^n)^{\otimes b}} \left(P_\pi \rho_b \left([g_a, g_b][g_c, g_d] \right) \right) = \sum_{I,J,K,L,M,N,O,Q,R} (P_\pi)_{RI} (g_a)_{IJ} (g_b)_{JK} (g_a)_{LK} (g_b)_{ML} (g_c)_{MN} (g_d)_{NO} (g_c)_{QO} (g_d)_{RQ} \tag{4.13}$$

[10] We point out that we do not use the full force of Cassidy's work in this paper, and it can be bypassed by the (stable) character theory of S_n. In particular, all we need here is to know that for large n, $\chi_{\lambda^+(n)}(g)$ is a fixed polynomial of $\mathsf{fix}(g^k)$ for finitely many k (depending on λ) where the polynomial has rational coefficients of n with denominators that are 'not too bad'. The more classical formula [HP23, (B.1)] would accomplish the same thing. Using Cassidy's result is handy here, however, and we believe that this formalism will also be useful for some future investigations in this area.

with $I, J, K, L, M, N, O, Q, R \in [n]^b$, where by slight abuse of notation we write $(g_f)_{IJ} \stackrel{\text{def}}{=} \rho_b(g_f)_{IJ}$. We view $I, J, K, L, M, N, O, Q, R$ as well as the fixed labeling of $Y^{(0)}$ concatenated together as a function

$$\Omega : Y^{(0)} \cup [9b] \to [n].$$

The function Ω and all these indices contain exactly the same information. We now partition the summation in (4.13) according to the partition p_Ω of $Y^{(0)} \cup [9b]$ that the function Ω defines: two locations are in the same block of p_Ω exactly when they have the same image. Write $\pi \to \Omega$ if and only if the values of R, I encoded by Ω satisfy $(P_\pi)_{RI} = 1$. We obtain

$$\begin{aligned} &\mathbf{1}\{g_f \text{ obey } Y\}\mathrm{Tr}_{(\mathbf{C}^n)^{\otimes b}}\left(P_\pi \rho_b\left([g_a, g_b][g_c, g_d]\right)\right) \\ &= \sum_{\pi \to \Omega} \mathbf{1}\{g_f \text{ obey } Y\}(g_a)_{IJ}(g_b)_{JK}(g_a)_{LK}(g_b)_{ML}(g_c)_{MN}(g_d)_{NO}(g_c)_{QO}(g_d)_{RQ}. \end{aligned} \tag{4.14}$$

In the above sum the labeling of $Y^{(0)}$ is fixed as usual. Now we use the following elementary integration formula:

$$\begin{aligned} &\int_{g \in S_n} g_{i_1 j_1} g_{i_2 j_2} \cdots g_{i_k j_k} \\ &= \begin{cases} \frac{1}{(n)_{|\theta|}} & \text{if the maps } \ell \mapsto i_\ell \text{ and } \ell \mapsto j_\ell \text{ define the same partition } \theta \in \mathrm{Part}([k]) \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Notice that $\mathbf{1}\{g_f \text{ obey } Y\}$ is a product of matrix entries of the g_f:

$$\mathbf{1}\{g_f \text{ obey } Y\} = \prod_{f \in \{a,b,c,d\}} \prod_{\{(i,j)\,:\,Y \text{ has } f\text{-colored edge from } i \to j\}} (g_f)_{ij}.$$

Hence for $n \geqslant \mathfrak{v} + 9b$ the integral of the summand of (4.14) depends only on p_Ω and is either zero or equal to

$$\frac{1}{\prod_{f \in \{a,b,c,d\}} (n)_{|(p_\Omega)_f|}} \tag{4.15}$$

where $(p_\Omega)_f$ is the partition induced by p_Ω on

- vertices of Y with an outgoing f-colored edge, together with
- elements of $[9b] \backslash Y^{(0)}$ who are in the domain of indices at the start of the matrix coefficient, e.g. for $f = a$, domain of I or domain of L.

Note that each

$$|(p_\Omega)_f| \leqslant \mathfrak{v} + 2b.$$

Therefore the total contribution from a fixed p_Ω to (4.12) is (4.15) times the number of Ω inducing p_Ω. If a block of p_Ω contains an element of $Y^{(0)}$ then the corresponding value of Ω is fixed by the fixed labeling of $Y^{(0)}$. The other blocks have values in $[n]$ that must be disjoint from each other, hence the number of Ω for a given p_Ω is

$$(n)_{\text{num. blocks of } p_\Omega \text{ disjoint from } Y^{(0)}}.$$

Putting the previous arguments together proves the result.

(For general g the only change is the number of factors $(n)_{|(p_\Omega)_f|}$ in the denominator and the end result is a denominator $(n)_{\mathfrak{v}+2b}^{1+2g}$.) $\square$

4.4 Proof of Theorem 1.7

We can now complete the proof.

Proof of Theorem 1.7. Suppose that $|\gamma| \leqslant q$. We write $\zeta(n) = \zeta(2; S_n)$. Now by (4.3), (4.4), and Proposition 4.4, for $n \geqslant 28q^2$

$$\mathbb{E}_{g,n}[\mathsf{fix}_\gamma] = \frac{1}{\zeta(n)} \sum_{r\in\mathcal{R}} \frac{(n)_{\mathfrak{v}(W_r)}}{\prod_{f\in a,b,c,d}(n)_{\mathfrak{e}_f(W_r)}} \sum_{\substack{\lambda\vdash n \\ \min(b_\lambda, b_\lambda^\vee)\leqslant 4q}} d_\lambda \Theta_\lambda(W_r) \\ + O\left((Cq)^{Cq} n^{-q}\right).$$

Since above either $b_\lambda \leqslant 4q$ or $b_\lambda^\vee \leqslant 4q$, for $n \geqslant Cq$ these events cannot happen simultaneously. Moreover, it is easy to check from (4.5) that $d_\lambda \Theta_\lambda(W_r)$ is invariant under $\lambda \mapsto \lambda^\vee$ and hence our main term can be written

$$\frac{2}{\zeta(n)} \sum_{r\in\mathcal{R}} \frac{(n)_{\mathfrak{v}(W_r)}}{\prod_{f\in a,b,c,d}(n)_{\mathfrak{e}_f(W_r)}} \sum_{|\lambda|\leqslant 4q} d_{\lambda^+(n)} \Theta_{\lambda^+(n)}(W_r).$$

By the hook-length formula [FRT54] each $d_{\lambda^+(n)}$ is a divisor in $\mathbf{Q}[n]$ of

$$(n)_{8q} = n(n-1)\cdots(n-8q+1)$$

and in particular, a polynomial of n in the range we consider. Since for everything above, $\mathfrak{v}(W_r), \mathfrak{e}_f(W_r) \leqslant q$, for $n \geqslant Cq$ Proposition 4.5 tells us that

$$\Phi_\gamma(n) \stackrel{\text{def}}{=} \sum_{r\in\mathcal{R}} \frac{(n)_{\mathfrak{v}(W_r)}}{\prod_{f\in a,b,c,d}(n)_{\mathfrak{e}_f(W_r)}} \sum_{|\lambda|\leqslant 4q} d_{\lambda^+(n)} \Theta_{\lambda^+(n)}(W_r)$$

is a rational function of n whose denominator can be taken to be $(n)_{9q}^9$. (For general g, this is replaced by $(n)_{9q}^{1+4g}$.)

So we write

$$\Phi_\gamma(n) = \frac{p(n)}{(n)_{9q}^9} = \frac{1}{2}\zeta(n)\mathbb{E}_{g,n}[\mathsf{fix}_\gamma] + O\left((Cq)^{Cq}n^{-q}\right)$$

using $\zeta \to 2$ as $n \to \infty$ (see (4.1)) to remove ζ from the error. Using the same fact again with a priori bound $|\mathbb{E}_{g,n}[\mathsf{fix}_\gamma]| \leqslant n$ we learn that $|\frac{1}{n}\frac{p(n)}{(n)_{9q}^9}.|$ is bounded as $n \to \infty$ and hence

$$\deg(p) \leqslant 81q + 1.$$

Let

$$g_q(t) = \prod_{k=0}^{9q-1}(1-kt)^9$$

and for $t \stackrel{\text{def}}{=} n^{-1}$ write

$$\Phi_\gamma(n) = \frac{p(n)}{(n)_{9q}^9} = \frac{n^{\deg(p)}}{n^{81q}}\frac{P(t)}{g_q(t)} = t^{81q-\deg(p)}\frac{P(t)}{g_q(t)} \stackrel{\text{def}}{=} \frac{1}{t}\frac{Q(t)}{g_q(t)}.$$

where $P(t)$ is a polynomial of t of degree $\leqslant \deg(p)$. For $\tau \in [0, \frac{1}{Cq^2}]$ we have

$$g_q(\tau) \geqslant C^{-1} \tag{4.16}$$

so we have

$$Q(t) = t^{81q-\deg(p)+1}P(t) = O(1) \tag{4.17}$$

for $t \in \left[0, \frac{1}{Cq^2}\right] \cap \mathbf{N}^{-1}$ (using the a priori bound again). Note that $\deg(Q) \leqslant 81q + 1$. (By the previous remarks, for general g, C can be taken to be linear in g.)

To conclude, we begin with

$$\mathbb{E}_{g,n}[\mathsf{fix}_\gamma] = \frac{2}{t\zeta(n)}\frac{Q(t)}{g_q(t)} + O\left((Cq)^{Cq}n^{-q}\right). \tag{4.18}$$

We now need to deal with the rogue $\zeta(n)$ factor. We can use a sort of trick to do this efficiently. If $\gamma = \mathrm{id}$ then all the previous arguments apply to give

$$t^{-1} = n = \mathbb{E}_{g,n}[\mathsf{fix}_{\mathrm{id}}] = \frac{2}{\zeta(n)}t^{-1}\frac{Q_{\mathrm{id}}(t)}{g_q(t)} + O\left((Cq)^{Cq}n^{-q}\right)$$

where Q_{id} is a polynomial that also satisfies (4.17). Rearranging,

$$\frac{Q_{\mathrm{id}}(t)}{g_q(t)} = \zeta(n)\left(\frac{1}{2} + O\left((Cq)^{Cq}n^{-q-1}\right)\right). \tag{4.19}$$

So when $n \geqslant (Cq)^{2C}$ we have

$$\frac{2}{\zeta(n)} = \frac{g_q(t)}{Q_{\mathrm{id}}(t)} \left(1 + O\left((Cq)^{Cq} n^{-q-1}\right)\right).$$

Therefore

$$\mathbb{E}_{g,n}[\mathsf{fix}_\gamma] = t^{-1} \frac{Q(t)}{Q_{\mathrm{id}}(t)} \left(1 + O\left((Cq)^{Cq} n^{-q-1}\right)\right) + O\left((Cq)^{Cq} n^{-q}\right).$$

Note that $Q_{\mathrm{id}} \geqslant C^{-1}$ in $n \geqslant q^C$ from (4.19) and (4.16). So by the bound on Q from (4.17) (also for Q_{id}) we obtain from Lemma 2.2 that for $\tau \in [0, \frac{1}{Cq^{2C}}]$

$$\left| \left(\frac{Q}{Q_{\mathrm{id}}}\right)^{(i)} (\tau) \right| \leqslant C(Cq)^{Ci}.$$

Therefore by Taylor's theorem there are $a_{-1}, a_0, \ldots, a_q$ (depending on γ) with

$$\left(\frac{Q}{Q_{\mathrm{id}}}\right)(\tau) = a_{-1} + a_0 \tau + \cdots + a_{q-1}\tau^q + O\left((Cq)^{Cq} \tau^{q+1}\right)$$

for $\tau \in [0, \frac{1}{Cq^{2C}}]$. Thus for $n \geqslant Cq^C$

$$\mathbb{E}_{g,n}[\mathsf{fix}_\gamma] = a_{-1} n + a_0 + a_1 n^{-1} + \cdots + a_{q-1} n^{-(q-1)} + O\left((Cq)^{Cq} n^{-q}\right).$$

This concludes the proof. □

5 Geometry of proper powers

The aim of this section is to prove Theorem 1.8.

5.1 Embedding the Cayley graph of Γ in $\mathbb{H}^2$

Denote by $\mathrm{Cay}(\Gamma)$ the Cayley graph of Γ with respect to the generators $\{a_1, b_1, \ldots, a_g, b_g\}$ in (1.2). We use the following standard embedding of $\mathrm{Cay}(\Gamma)$ in the hyperbolic plane $\mathbb{H}^2$, which is useful, inter alia, as it agrees with the description in [BS87].

Consider the tiling $\mathcal{T}$ of $\mathbb{H}^2$ by regular $4g$-gons with geodesic sides and interior angles $\frac{2\pi}{4g}$. The Cayley graph is then the tiling dual to $\mathcal{T}$. More precisely, the vertices of $\mathrm{Cay}(\Gamma)$ are the centers of the $4g$-gons, and the edges are geodesic arcs between the centers of any two $4g$-gons sharing a side. Every edge of $\mathrm{Cay}(\Gamma)$ is directed and labeled by one of the $2g$ generators. For any generator x, when traversing an x-edge against its direction, one reads x^{-1}. At each vertex, there is exactly one edge directed outward

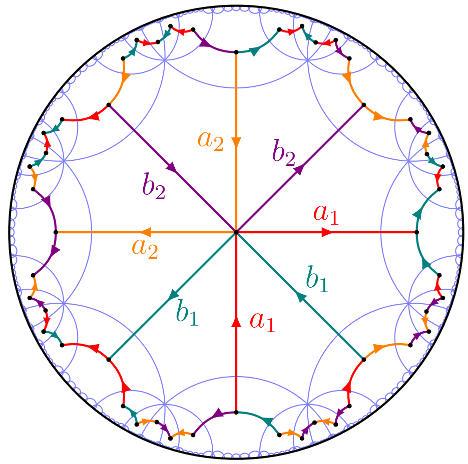

Figure 5.1: Illustration of the tiling $\mathcal{T}$ and embedded Cayley graph of Γ for $g = 2$.

and one directed inward with every given label. The cyclic order of the $4g$ outgoing edges, say clockwise, is

$$a_1, b_1^{-1}, a_1^{-1}, b_1, a_2, b_2^{-1}, a_2^{-1}, b_2, \ldots, b_g.$$

(Up to isometry, there is basically a unique way to label the edges of the dual tiling in this way in a compatible fashion.) The boundary of every dual $4g$-gon now reads cyclically, counterclockwise, the relation $[a_1, b_1] \cdots [a_g, b_g]$. See Figure 5.1.

We pick one of the vertices to be the identity element and mark it by o. Any other vertex v of $\mathrm{Cay}(\Gamma)$ now corresponds to the element of Γ represented by any of the paths from o to v. The left action of Γ on $\mathrm{Cay}(\Gamma)$ extends to an action by isometries on $\mathbb{H}^2$. The isometry corresponding to each non-trivial element of Γ is hyperbolic, which means that it admits a (unique, bi-infinite) geodesic axis in $\mathbb{H}^2$: the isometry acts on this axis by translation. These facts can be found, e.g., in [FM11, §1].

5.2 $\prod$-shaped paths for elements in Γ

The following definition will play a key role in the sequel. Roughly speaking, it is the appropriate replacement for surface groups of the fact that any element γ of a free group can be expressed as bhb^{-1} where h is the cyclic reduction of γ.

Definition 5.1 ($\prod$-path)**.** Let $1 \neq \gamma \in \Gamma$. Consider the unique (bi-infinite) geodesic $\rho \subset \mathbb{H}^2$ which is the axis of the isometry given by γ. The $\prod$-path corresponding to γ, which we denote by $\prod_\gamma$, is the path in $\mathbb{H}^2$ made out of three geodesics arcs: a geodesic arc $[o, x]$ from o to $x \in \rho$ which is perpendicular to ρ, the geodesic arc $[x, \gamma.x] \subset \rho$, and the geodesic arc $[\gamma.x, \gamma.o]$ (which is, too, perpendicular to ρ).

See Figure 5.3 below. Note that for any hyperbolic isometry of $\mathbb{H}^2$, each half-space at either side of the axis is invariant under the isometry. This means that o and $\gamma.o$ lie on the same side of the axis, so that $\prod_\gamma$ is made of three geodesic arcs with two right-turns, or two left-turns, between them (unless o happens to lie on the axis of γ, in which case $\prod_\gamma$ is simply a geodesic arc).

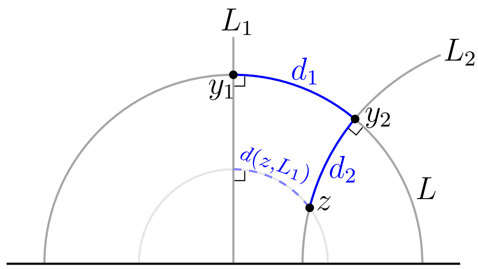

Figure 5.2: Illustration of Fact 5.2 (half-plane model of $\mathbb{H}^2$)

We will need the following easy fact from hyperbolic geometry:

Fact 5.2. *Let L be a geodesic in $\mathbb{H}^2$ and L_1, L_2 be two geodesic rays perpendicular to L, emanating from two distinct points y_1 and y_2, respectively, on L, and at the same side of L. Denote $d_1 = d(y_1, y_2) > 0$. Let z be a point on L_2 at distance $d_2 \geqslant 0$ from y_2. Then the distance $d(z, L_1)$ of z from L_1 grows monotonically with each of d_1 and d_2, and for every fixed $d_1 > 0$,*

$$d(z, L_1) \overset{d_2 \to \infty}{\to} \infty.$$

Fact 5.2 is illustrated in Figure 5.2. It follows immediately from the explicit formula $\sinh d(z, x) = \sinh(d_1)\cosh(d_2)$, that may be found in [Mar75, Thm. 32.21(a')]. For completeness, we give an elementary argument.

Proof of Fact 5.2. Assume without loss of generality that L_1 is the geodesic $x = 0$ in the upper half plane model of $\mathbb{H}^2$. Then the distance between z and L_1 is determined by the argument of z, and monotonically increases and tends to ∞ as the argument in $\left[0, \frac{\pi}{2}\right]$ decreases. This proves the claims when $d_1 > 0$ is fixed.

Now fix $d_2 \geqslant 0$. Then $d(z, L_1)$ is continuous as a function of d_1. Clearly $d(z, L_1) \overset{d_1 \searrow 0}{\to} 0$ and $d(z, L_1) \overset{d_1 \to \infty}{\to} \infty$. So if it is not monotonically increasing, there must be two different values $\alpha_1 < \alpha_2$ of d_1 with two corresponding points $z_1 \neq z_2$ with $d(z_1, L) = d(z_2, L)$. Then z_1 must lie on the Euclidean interval from z_2 to 0. So the hyperbolic isometry $\Phi\colon t \mapsto ct$ maps z_1 to z_2 for some $c > 1$. But this isometry shows that z_2 is at distance d_2 from the geodesic $\Phi(L)$ which lies strictly above L — a contradiction. $\square$

The following lemma is illustrated in Figure 5.3.

Lemma 5.3. *There is a constant $c_1 = c_1(g) > 0$ so that $\prod_\gamma$ and the geodesic arc $[o, \gamma.o]$ are each contained in a c_1-neighborhood of the other. Moreover, there are two points $z_1, z_2 \in [o, \gamma.o]$, with z_2 not closer to o than z_1, such that the two corners x and $\gamma.x$ of $\prod_\gamma$ satisfy $d_{\mathbb{H}^2}(x, z_1) < c_1$ and $d_{\mathbb{H}^2}(\gamma.x, z_2) < c_1$.*

Proof. $\mathbb{H}^2$ has δ-thin triangles in the sense of Rips and Gromov: every side of a geodesic triangle is contained in a δ-neighborhood of the union of the other two sides [BH99, §III.H.1]. It follows that every side of a geodesic quadrilateral is contained in a 2δ-neighborhood of the union of the other three sides. Hence, $[o, \gamma.o] \subset \mathcal{N}_{2\delta}\left(\prod_\gamma\right)$.

Conversely, denote by $\rho_1 = [o, x]$, $\rho_2 = [x, \gamma.x]$ and $\rho_3 = [\gamma.x, \gamma.o]$ the three geodesic arcs composing $\prod_\gamma$, so $x = \rho_1 \cap \rho_2$ and $\gamma.x = \rho_2 \cap \rho_3$. Note that $|\rho_2|$ is bounded

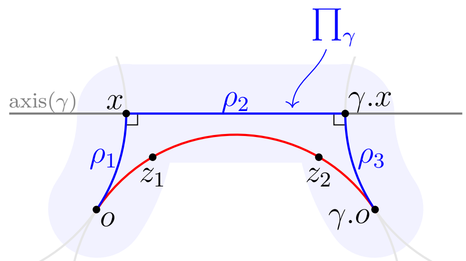

Figure 5.3: Illustration of Definition 5.1 and Lemma 5.3

from below by the injectivity radius of the genus-g hyperbolic surface $\Gamma\backslash\mathbb{H}^2$. Hence, by Fact 5.2, there is some $c' \geqslant 2\delta$ such that if $y \in \rho_1$ satisfies $d(y, x) > c'$, then $d(y, \rho_3) > 4\delta$. If $|\rho_1| < 2c'$ we may simply take $z_1 = o$ and $z_2 = \gamma.o$ (recall that $|\rho_3| = |\rho_1|$). Otherwise, we reason as follows. By construction, the sets $\mathcal{N}_{2\delta}(\{y \in \rho_1 : d(y, x) > c'\})$ and $\mathcal{N}_{2\delta}(\rho_3)$ are disjoint. Thus $[o, \gamma.o]$ must visit a point $z_1 \in \mathcal{N}_{2\delta}(\{y \in \rho_1 : d(y, x) \leqslant 2c'\})$ before it reaches $\mathcal{N}_{2\delta}(\rho_3)$. Clearly, $d(z_1, x) \leqslant 2c' + 2\delta$. Symmetrically, the inverse geodesic $[\gamma.o, o]$ must visit a point $z_2 \in \mathcal{N}_{2\delta}(\{y \in \rho_3 : d(y, \gamma.x) \leqslant 2c'\})$ before it reaches $\mathcal{N}_{2\delta}(\rho_1)$ and $d(z_2, \gamma.x) \leqslant 2c' + 2\delta$. Finally,

$$\rho_1 \subset \mathcal{N}_\delta([o, z_1] \cup [z_1, x]) \subseteq \mathcal{N}_{3\delta+2c'}([o, z_1]),$$

likewise $\rho_3 \subset \mathcal{N}_{3\delta+2c'}([z_2, \gamma.o])$, and

$$\rho_2 \subset \mathcal{N}_{2\delta}([x, z_1] \cup [z_1, z_2] \cup [z_2, \gamma.x]) \subseteq \mathcal{N}_{4\delta+2c'}([z_1, z_2]).$$

The lemma is now proven with $c_1 = 4\delta + 2c'$. □

We now follow the terminology from [BS87]. Recall the tiling $\mathcal{T}$ of $\mathbb{H}^2$ that was introduced in §5.1. Denote by hO the $4g$-gon with center $h.o$.

Definition 5.4 (Geodesic edge path [BS87, p. 455])**.** Let ρ be a finite-length, oriented geodesic arc in $\mathbb{H}^2$. Assume it begins in the interior of the $4g$-gon h_1O, and then visits h_2O, h_3O and so on until it terminates at the interior of h_kO, and that it does not visit any vertex of the tiling $\mathcal{T}$ and does not coincide with any edge of $\mathcal{T}$. Then the corresponding *geodesic edge path* is the path $h_1.o - h_2.o - \ldots - h_k.o$ in Cay(Γ). The corresponding geodesic word is the (reduced) word in $F(a_1, \ldots, b_g)$

$$\left(h_1^{-1}h_2\right)\left(h_2^{-1}h_3\right)\ldots\left(h_{k-1}^{-1}h_k\right)$$

(recall that if hO and $h'O$ share a common side, then the edge in Cay(Γ) from hO to $h'O$ is labeled by $h^{-1}h' \in \left\{a_1^{\pm 1}, \ldots, b_g^{\pm 1}\right\}$). If ρ goes through a vertex of $\mathcal{T}$ or coincides with an edge of $\mathcal{T}$, we deform it slightly as in [Ibid., Fig. 1], and then define the geodesic edge path as above using the deformed arc.[11]

[11] To avoid a vertex, the deformation consists of taking a detour on either side of the vertex. To avoid an edge, it consists of moving the entire geodesic arc to either side of the edge. The geodesic

Using geodesic edge paths, we have an analog of Lemma 5.3 in $\mathrm{Cay}(\Gamma)$. It basically says that in $\mathrm{Cay}(\Gamma)$ there are, too, a geodesic path from o to $\gamma.o$ and a $\prod_\gamma$-like path which are close to each other. We focus on proper powers and on the precise properties that will be needed below. We view $\mathrm{Cay}(\Gamma)$ as a metric space with edge-length 1, and a geodesic path in it is any shortest path between two vertices (here and in the following lemma, vertices are vertices of $\mathrm{Cay}(\Gamma)$ and not of the tiling, unless stated otherwise).

Lemma 5.5. *There is a constant $c_2 = c_2(g) > 0$ so that the following holds. Let $k \in \mathbf{Z}_{\geqslant 2}$ and $1 \neq \gamma \in \Gamma$ be a proper k^{th}-power. Then there are $b, h \in \Gamma$ with $\gamma = bh^k b^{-1}$, a geodesic path $\tau_\gamma \subset Cay(\Gamma)$ from o to $\gamma.o$, and four vertices $v_0, v_1, v_2, v_k \in \tau_\gamma$, with v_j not closer to o than v_i when $i < j$, such that $d_{Cay(\Gamma)}(bh^j.o, v_j) < c_2$ for all $j = 0, 1, 2, k$.*

Proof. Let $\gamma_0 \in \Gamma$ with $\gamma = \gamma_0^k$. Consider the geodesic arc $[o, \gamma.o]$ in $\mathbb{H}^2$, (deform it a bit if need be), and let τ_γ be the associated geodesic edge path as in Definition 5.4. By [BS87, Thm. 2.8(b)], every geodesic edge path, and τ_γ in particular, is a geodesic in $\mathrm{Cay}(\Gamma)$.

Now consider the $\prod$-path $\prod_\gamma$ with vertices $o, x, \gamma.x, \gamma.o$. Let $b \in \Gamma$ satisfy that $x \in bO$ (if x lies on a vertex or an edge of $\mathcal{T}$, pick some neighboring $4g$-gon arbitrarily), and let $h \overset{\text{def}}{=} b^{-1}\gamma_0 b \in \Gamma$. Note that $\gamma = bh^k b^{-1}$. Note also that $\gamma_0.x \in \gamma_0 bO = bhO$, and likewise ${\gamma_0}^j.x \in bh^j O$ for all $j \in \mathbf{Z}$. As the axes of γ and of γ_0 are identical, the points $x, \gamma_0.x, \ldots, {\gamma_0}^k.x = \gamma.x$ all lie on $[x, \gamma.x]$. By Lemma 5.3, there are points z_0 and then z_k along $[o, \gamma.o]$ at distance $< c_1(g)$ from x and $\gamma.x$, respectively. Now let $\delta = \delta_{\mathbb{H}^2}$ be such that triangles in $\mathbb{H}^2$ are δ-thin, and consider the geodesic quadrilateral in $\mathbb{H}^2$ with corners $x, \gamma.x, z_k, z_0$. The point $\gamma_0.x \in [x, \gamma.x]$ is at distance $< 2\delta$ from the union of the other three sides, and thus at distance $< 2\delta + c_1(g)$ from $[z_0, z_k]$. We mark a point $z_1 \in [z_0, z_k]$ with $d_{\mathbb{H}^2}(z_1, \gamma_0.x) < 2\delta + c_1(g)$. Similarly, we find a point $z_2 \in [z_1, z_k]$ with $d_{\mathbb{H}^2}(z_2, \gamma_0^2.x) < 4\delta + c_1(g)$ (if $k = 2$ we may simply set $z_2 = z_k$).

Denote by $D > 0$ the diameter of the $4g$-gons. Every point on a geodesic arc in $\mathbb{H}^2$ lies in the D-neighborhood of its corresponding geodesic edge path, and even in the D-neighborhood of the vertices of the geodesic edge path. So we may find vertices $v_0, v_1, v_2, v_k \in \tau_\gamma$ with $d_{\mathbb{H}^2}(z_j, v_j) < D$ for $j = 0, 1, 2, k$. It is clear, by the definition of a geodesic edge path, that we may choose the v_j's so that v_j is not closer than v_i to o when $i < j$. As $d_{\mathbb{H}^2}({\gamma_0}^j.x, bh^j.o) < D$ for all $j = 0, 1, 2, k$, we obtain that

$$d_{\mathbb{H}^2}\left(bh^j.o, v_j\right) < 4\delta + c_1\left(g\right) + 2D.$$

Finally, by the Švarc-Milnor Lemma (e.g., [BH99, Prop. I.8.19]), the embedding above of $\mathrm{Cay}(\Gamma)$ in $\mathbb{H}^2$ is a quasi-isometry, and bounded distances remain bounded. In particular, there is some $c_2(g) > 0$ so that $d_{\mathrm{Cay}(\Gamma)}(bh^j.o, v_j) < c_2(g)$ for $j = 0, 1, 2, k$. □

The following lemma considers words with letters in an arbitrary (finite) subset $\mathcal{S} \subset \Gamma$. Let w be such a word of length p. We let $w_{[t_1, t_2)}$ denote the subword consisting of the letters at positions $t_1, t_1 + 1, \ldots, t_2 - 1$ (here $1 \leqslant t_1 \leqslant t_2 \leqslant p + 1$), and $w_{[t_1, t_2]}$ the subword consisting of the letters at positions $t_1, \ldots, t_2$ (here $1 \leqslant t_1 \leqslant t_2 \leqslant p$).

arcs we use here never coincide with edges of $\mathcal{T}$. If ρ needs to be deformed, the resulting geodesic edge path is not unique, but this is immaterial for our purposes.

Lemma 5.6. *There is a constant $c_3 = c_3(g) > 0$ so that the following holds. Fix any finite subset $\mathcal{S} \subset \Gamma$ so that $\max_{\gamma\in\mathcal{S}} |\gamma| \leqslant d$, and let w be any not-necesssarily-reduced word of length p with letters in $\mathcal{S}$ representing a non-trivial k^{th} power in Γ for some $k \geqslant 2$. Then there are $1 \leqslant t_1 \leqslant t_2 \leqslant t_3 \leqslant t_4 \leqslant p+1$, and elements $b, h, u_1, \ldots, u_4 \in \Gamma$ with $|u_i| \leqslant c_3 (d + \log p)$, such that*

$$\left(w_{[1,t_1)}u_1\right) \cdot \left(u_1^{-1}w_{[t_1,t_2)}u_2\right) \cdot \left(u_2^{-1}w_{[t_2,t_3)}u_3\right) \cdot \left(u_3^{-1}w_{[t_3,t_4)}u_4\right) \cdot \left(u_4^{-1}w_{[t_4,p]}\right)$$

is of the form $b \cdot h \cdot h \cdot h^{k-2} \cdot b^{-1}$. Namely,

$$\begin{aligned} w_{[1,t_1)}u_1 &\equiv_\Gamma b, & u_1^{-1}w_{[t_1,t_2)}u_2 &\equiv_\Gamma u_2^{-1}w_{[t_2,t_3)}u_3 \equiv_\Gamma h \\ u_3^{-1}w_{[t_3,t_4)}u_4 &\equiv_\Gamma h^{k-2}, & u_4^{-1}w_{[t_4,p]} &\equiv_\Gamma b^{-1}. \end{aligned}$$

Here $\equiv_\Gamma$ denotes equality in Γ.

Proof. All the distances in this proof are measured in $\mathrm{Cay}(\Gamma)$. Let $1 \neq \gamma \in \Gamma$ denote the element represented by w, and let $b, h \in \Gamma$, τ_γ and $v_0, v_1, v_2, v_k \in \tau_\gamma$ be the elements of Γ, the geodesic and the vertices given by Lemma 5.5. We may expand w to a word of length at most dp in the generators $\left\{a_1^\pm, \ldots, b_g^\pm\right\}$ of Γ, and the expanded word then represents a path $\mathfrak{p}_w$ of length at most dp from o to $\gamma.o$. The Cayley graph $\mathrm{Cay}(\Gamma)$ is a hyperbolic space (in the sense of Rips-Gromov) with δ-thin triangles for some $\delta > 0$. A standard fact in hyperbolic geometry (e.g., [BH99, Prop. III.H.1.6]) gives that as τ_γ is a geodesic in $\mathrm{Cay}(\Gamma)$ between o and $\gamma.o$, every point $x \in \tau_\gamma$ is at distance at most $\delta\left|\log_2|\mathfrak{p}_w|\right| + 1 \leqslant \delta\left|\log_2(pd)\right| + 1$ from $\mathfrak{p}_w$. Together with Lemma 5.5, this shows that each of the points $b.o, bh.o, bh^2.o$ and $bh^k.o$ are rather close to $\mathfrak{p}_w$. It remains to show we can find four points along $\mathfrak{p}_w$ which are close to $b.o, bh.o, bh^2.o, bh^k.o$ and ordered correctly.

First, as every vertex on $\mathfrak{p}_w$ is at distance at most d from the beginning of a letter of w, there is a vertex y_0 on $\mathfrak{p}_w$ which is at the beginning of a letter of w and with $d(v_0, y_0) < \delta\left|\log_2(pd)\right| + 1 + d$, so

$$d(b.o, y_0) \leqslant d(b.o, v_0) + d(v_0, y_0) < c_2 + \delta\left|\log_2(pd)\right| + 1 + d < c'(d + \log p),$$

for some $c' > 0$ depending only on g. Set u_1 to be a word representing a shortest path in $\mathrm{Cay}(\Gamma)$ from y_0 to $b.o$, and let t_1 be the smallest integer so that $w_{[1,t_1)}$ ends at y_0. Now consider the path from v_0 to $\gamma.o$ given by $\mathfrak{q} \cdot w_{[t_1,p]}$, where $\mathfrak{q}$ is a geodesic path from v_0 to y_0 (so $|\mathfrak{q}| < \delta\left|\log_2(pd)\right| + 1 + d$). As v_1 lies on a geodesic from v_0 to $\gamma.o$, there is a vertex y_1 on $\mathfrak{p}_w$, at the beginning of a letter of $w_{[t_1,p]}$, such that

$$d(v_1, y_1) < \delta\left|\log_2(d(p+1-t_1) + |\mathfrak{q}|)\right| + 1 + d + |\mathfrak{q}|,$$

so $d(bh.o, y_1) < c''(d + \log p)$ for some $c'' > 0$ depending only on g. We let u_2 be a word representing a shortest path from y_1 to $bh.o$ and t_2 the smallest integer so that $t_2 \geqslant t_1$ and $w_{[1,t_2)}$ ends at y_1.

In the same manner, and enlarging the constant c'' as needed at each step, we find points y_2 and y_k on the path of w, words u_3 and u_4 and times t_3 and t_4 so that the statement of the lemma holds. $\square$

5.3 Proof of Theorem 1.8

As is explained in [MdlS24b, §6], the proof of Theorem 1.8 can be reduced to bounding the exponential growth rate of the probability that a product $\boldsymbol{\gamma}_1 \cdots \boldsymbol{\gamma}_p$ of i.i.d. finitely supported random group elements $\boldsymbol{\gamma}_i \in \Gamma$ (i.e., a random walk on Γ with a general finitely supported generating measure) lands on a proper power.

In the case of the free group, the requisite bound follows by a simple spectral argument that dates back to [Fri03, Lem. 2.4]. This argument can be adapted to the setting of surface groups due to Lemma 5.6.

Proof of Theorem 1.8. Fix a self-adjoint $x = \sum_{\gamma \in \Gamma} \alpha_\gamma \gamma$ with $|x| = d$. By [MdlS24b, Prop. 6.3] and as surface groups satisfy the rapid decay property [Jol90, Thm. 3.2.1], we can assume that x has positive coefficients, and by homogeneity we can assume the coefficients sum to one. Thus we can write $x^p = \mathbb{E}[\boldsymbol{\gamma}_1 \cdots \boldsymbol{\gamma}_p]$ for every $p \in \mathbf{N}$, where $\boldsymbol{\gamma}_i$ are i.i.d. random elements of Γ with $\mathbb{P}[\boldsymbol{\gamma}_i = \gamma] = \alpha_\gamma$. In particular, (1.3) yields

$$u_1(x^p) = -\tau(\lambda(x)^p) + \sum_{k=2}^{pd} (\omega(k) - 1) \sum_{v \in \Gamma_{\mathrm{np}}} \mathbb{E}\left[1_{\boldsymbol{\gamma}_1 \cdots \boldsymbol{\gamma}_p \equiv_\Gamma v^k}\right],$$

where Γ_{np} denotes the non-powers in Γ. Here we used that $\boldsymbol{\gamma}_1 \cdots \boldsymbol{\gamma}_p \equiv_\Gamma v^k$ implies that $k \leqslant |v^k| \leqslant |\boldsymbol{\gamma}_1| + \cdots + |\boldsymbol{\gamma}_p| \leqslant pd$, and that $\mathbb{E}[1_{\boldsymbol{\gamma}_1 \cdots \boldsymbol{\gamma}_p \equiv_\Gamma 1}] = \mathbb{E}[\tau(\lambda(\boldsymbol{\gamma}_1 \cdots \boldsymbol{\gamma}_p))] = \tau(\lambda(x)^p)$.

By Lemma 5.6, we have for every $k \geqslant 2$ and $\gamma_1, \ldots, \gamma_p \in \Gamma$

$$\begin{aligned} \sum_{v \in \Gamma_{\mathrm{np}}} 1_{\gamma_1 \cdots \gamma_p \equiv_\Gamma v^k} \quad \leqslant \quad & \sum_{1 \leqslant t_1 \leqslant \cdots \leqslant t_4 \leqslant p+1} \sum_{\substack{b,h,u_1,\ldots,u_4 \in \Gamma \\ |u_i| \leqslant c_3(d + \log p)}} 1_{\gamma_1 \cdots \gamma_{t_1 - 1} \equiv_\Gamma b u_1^{-1}} 1_{\gamma_{t_1} \cdots \gamma_{t_2 - 1} \equiv_\Gamma u_1 h u_2^{-1}} \times \\ & 1_{\gamma_{t_2} \cdots \gamma_{t_3 - 1} \equiv_\Gamma u_2 h u_3^{-1}} 1_{\gamma_{t_3} \cdots \gamma_{t_4 - 1} \equiv_\Gamma u_3 h^{k-2} u_4^{-1}} 1_{\gamma_{t_4} \cdots \gamma_p \equiv_\Gamma u_4 b^{-1}}. \end{aligned}$$

As

$$1_{\gamma_1 \cdots \gamma_p \equiv_\Gamma \gamma} = \langle \delta_\gamma, \lambda(\gamma_1 \cdots \gamma_p) \delta_e \rangle,$$

this yields

$$\begin{aligned} \sum_{v \in \Gamma_{\mathrm{np}}} \mathbb{E}\left[1_{\boldsymbol{\gamma}_1 \cdots \boldsymbol{\gamma}_p \equiv_\Gamma v^k}\right] \leqslant & \sum_{1 \leqslant t_1 \leqslant \cdots \leqslant t_4 \leqslant p+1} \sum_{\substack{b,h,u_1,\ldots,u_4 \in \Gamma \\ |u_i| \leqslant c_3(d + \log p)}} \left\langle \delta_{b u_1^{-1}}, \lambda(x)^{t_1 - 1} \delta_e \right\rangle \times \\ & \left\langle \delta_{u_1 h u_2^{-1}}, \lambda(x)^{t_2 - t_1} \delta_e \right\rangle \left\langle \delta_{u_2 h u_3^{-1}}, \lambda(x)^{t_3 - t_2} \delta_e \right\rangle \|\lambda(x)\|^{t_4 - t_3} \left\langle \delta_{u_4 b^{-1}}, \lambda(x)^{p+1-t_4} \delta_e \right\rangle. \end{aligned}$$

Now note that the sum of $t_1, \ldots, t_4$ has at most $(p+1)^4$ terms, while the sum over

$u_1, \dots, u_4$ has at most $(4g+1)^{4c_3(d+\log p)}$ terms. Moreover, as

$$\sum_{v\in\Gamma} |\langle \delta_v, \lambda(x)^t \delta_e\rangle|^2 = \|\lambda(x)^t\delta_e\|^2 \leqslant \|\lambda(x)\|^{2t},$$

we can apply Cauchy-Schwarz to the remaining sums over b, h to estimate

$$\begin{aligned} u_1(x^p) &\leqslant |\tau(\lambda(x)^p)| + (pd)^2 \max_{2\leqslant k\leqslant pd} \sum_{v\in\Gamma_{\mathrm{np}}} \mathbb{E}\big[1_{\gamma_1\cdots\gamma_p\equiv v^k}\big] \\ &\leqslant \big[1 + (pd)^2(p+1)^4(4g+1)^{4c_3(d+\log p)}\big]\,\|\lambda(x)\|^p. \end{aligned}$$

The conclusion follows directly. □

6 Proof of Theorem 1.1

Before we complete the proof, we must dispense with a cosmetic issue. Let

$$\pi_n = \mathrm{std}\circ\phi_n,$$

where ϕ_n is a uniformly distributed random element of $\mathrm{Hom}(\Gamma_g, S_n)$. In Theorem 1.7, we have proved an effective asymptptic expansion of $\mathbb{E}[\mathrm{Tr}\,\pi_n(\gamma)]$ in powers of $\frac{1}{n}$. However, since $\pi_n : \Gamma \to \mathcal{U}(n-1)$ is a representation of dimension $n-1$, this differs slightly from the setting of section 1.1 where π_n is n-dimensional.

This minor discrepancy is readily resolved by noting that the statement and proof of Theorem 1.5 extend *verbatim* to the case that π_n is $(n-1)$-dimensional, provided that we still understand the normalized trace to be defined by

$$\mathrm{tr}\,\pi_n \stackrel{\mathrm{def}}{=} \frac{1}{n}\,\mathrm{Tr}\,\pi_n.$$

We will therefore impose this convention in the remainder of this section.[12]

We can now proceed to assemble all the pieces of the proof.

Proof of Theorem 1.1. The effective asymptotic expansion of Assumption 1.3 follows immediately from Theorem 1.7 and (1.1) with

$$u_1(\gamma) = a_0(\gamma) - 1, \qquad u_k(\gamma) = a_{k-1}(\gamma) \quad \text{for} \quad k \neq 1.$$

Moreover, it is shown in [MP23, Thm. 1.2] that $a_{-1}(\gamma) = 1_{\gamma=1}$ and $a_0(\gamma) = \omega(\gamma)$. Thus

$$u_0(\gamma) = 1_{\gamma=1} = \tau(\lambda(\gamma)), \qquad u_1(\gamma) = \omega(\gamma) - 1.$$

[12] The issue is purely cosmetic in nature, since the proof of Theorem 1.7 could also be readily adapted to yield an effective asymptotic expansion in powers of $\frac{1}{n-1}$.

Thus Assumption 1.3 holds, and Assumption 1.4 follows from Theorem 1.8. Consequently, Theorem 1.5 yields the strong convergence upper bound

$$\|\pi_n(x)\| \leqslant \|\lambda(x)\| + o(1) \quad \text{with probability} \quad 1 - o(1)$$

as $n \to \infty$ for every $x \in \mathbf{C}[\Gamma]$. As surface groups have the unique trace property [dlH88], the corresponding strong convergence lower bound

$$\|\pi_n(x)\| \geqslant \|\lambda(x)\| - o(1) \quad \text{with probability} \quad 1 - o(1)$$

follows from the upper bound, see, e.g., [MdlS24b, §5.3]. □

References

[Alo86] N. Alon, *Eigenvalues and expanders*, Combinatorica **6** (1986), 83–96. 8

[AM23] N. Anantharaman and L. Monk, *Friedman-Ramanujan functions in random hyperbolic geometry and application to spectral gaps*, 2023, arXiv:2304.02678. 8

[AM25] ______, *Friedman-Ramanujan functions in random hyperbolic geometry and application to spectral gaps II*, 2025, arXiv:2502.12268. 9

[Aus24] T. Austin, *Notions of entropy for unitary representations*, 2024, arXiv:2412.13751. 8

[BC19] C. Bordenave and B. Collins, *Eigenvalues of random lifts and polynomials of random permutation matrices*, Ann. of Math. (2) **190** (2019), 811–875. 2, 7

[BC22] S. Belinschi and M. Capitaine, *Strong convergence of tensor products of independent G.U.E. matrices*, 2022, arXiv:2205.07695. 8

[BC24a] C. Bordenave and B. Collins, *Norm of matrix-valued polynomials in random unitaries and permutations*, 2024, arXiv:2304.05714. 8

[BC24b] ______, *Strong asymptotic freeness for independent uniform variables on compact groups associated to nontrivial representations*, Invent. Math. **237** (2024), 221–273. 8

[BH99] M. R. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, vol. 319, Springer Science & Business Media, 1999. 28, 30, 31

[BM04] R. Brooks and E. Makover, *Random construction of Riemann surfaces*, J. Differ. Geom. **68** (2004), 121–157. 8

[Bor20] C. Bordenave, *A new proof of Friedman's second eigenvalue theorem and its extension to random lifts*, Ann. Sci. Éc. Norm. Supér. (4) **53** (2020), 1393–1439. 8

[BS87] J. S. Birman and C. Series, *Dehn's algorithm revisited, with applications to simple curves on surfaces*, Combinatorial group theory and topology (Alta, Utah, 1984), Ann. of Math. Stud., vol. 111, Princeton Univ. Press, 1987, pp. 451–478. 26, 29, 30

[Bus84] P. Buser, *On the bipartition of graphs*, Discrete Appl. Math. **9** (1984), 105–109. 8

[Cas23] E. Cassidy, *Projection formulas and a refinement of Schur–Weyl–Jones duality for symmetric groups*, 2023, arXiv:2312.01839. 21, 22

[Cas25] ———, *Random permutations acting on k–tuples have near–optimal spectral gap for $k = \mathrm{poly}(n)$*, 2025, arXiv:2412.13941. 8

[CGVTvH24] C. F. Chen, J. Garza-Vargas, J. A. Tropp, and R. van Handel, *A new approach to strong convergence*, Ann. of Math. (2) (2024), to appear, arXiv:2405.16026. 3, 4, 5, 6, 8, 11, 12, 14, 15

[CGVvH24] C. F. Chen, J. Garza-Vargas, and R. van Handel, *A new approach to strong convergence II. The classical ensembles*, 2024, arXiv:2412.00593. 3, 6, 8

[CM14] B. Collins and C. Male, *The strong asymptotic freeness of Haar and deterministic matrices*, Ann. Sci. Éc. Norm. Supér. (4) **47** (2014), 147–163. 7

[dlH88] P. de la Harpe, *Groupes hyperboliques, algèbres d'opérateurs et un théorème de Jolissaint*, C. R. Acad. Sci. Paris Sér. I Math. **307** (1988), 771–774. 34

[FM11] B. Farb and D. Margalit, *A primer on mapping class groups*, Princeton Mathematical Series, vol. 49, Princeton university press, 2011. 6, 27

[Fri03] J. Friedman, *Relative expanders or weakly relatively Ramanujan graphs*, Duke Math. J. **118** (2003), 19. 7, 32

[Fri08] ———, *A proof of Alon's second eigenvalue conjecture and related problems*, Mem. Amer. Math. Soc. **195** (2008). 8

[FRT54] J. S. Frame, G. de B. Robinson, and R. M. Thrall, *The hook graphs of the symmetric group*, Canad. J. Math. **6** (1954), 316–324. 22, 24

[Gam06] A. Gamburd, *Poisson-Dirichlet distribution for random Belyi surfaces*, Ann. Probab. **34** (2006), 1827–1848. 17

[Hay22] B. Hayes, *A random matrix approach to the Peterson-Thom conjecture*, Indiana Univ. Math. J. **71** (2022), 1243–1297. 8

[Hid25] W. Hide, *Effective lower bounds for spectra of random covers and random unitary bundles*, Israel J. Math. (2025+), to appear, available at arXiv:2305.04584. 3

[HM23] W. Hide and M. Magee, *Near optimal spectral gaps for hyperbolic surfaces*, Ann. Math. (2) **198** (2023), 791–824. 3, 9

[HMY25] J. Huang, T. McKenzie, and H.-T. Yau, *Ramanujan property and edge universality of random regular graphs*, 2025, arXiv:2412.20263. 8

[HP23] L. Hanany and D. Puder, *Word measures on symmetric groups*, Int. Math. Res. Not. IMRN **2023** (2023), 9221–9297. 22

[HT05] 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** (2005), 711–775. 7

[Hub74] H. Huber, *Über den ersten Eigenwert des Laplace-Operators auf kompakten Riemannschen Flächen*, Comment. Math. Helv. **49** (1974), 251–259. 3

[Jol90] P. Jolissaint, *Rapidly decreasing functions in reduced C^*-algebras of groups*, Trans. Amer. Math. Soc. **317** (1990), 167–196. 7, 32

[LM25] L. Louder and M. Magee, *Strongly convergent unitary representations of limit groups*, J. Funct. Anal. **288** (2025), no. 6, 110803. 3, 8, 9

[LS04] M. W. Liebeck and A. Shalev, *Fuchsian groups, coverings of Riemann surfaces, subgroup growth, random quotients and random walks*, J. Algebra **276** (2004), 552–601. 16

[LW24] M. Lipnowski and A. Wright, *Towards optimal spectral gaps in large genus*, Ann. Probab. **52** (2024), 545–575. 8

[Mag25] M. Magee, *Strong convergence of unitary and permutation representations of discrete groups*, 2025, arXiv:2503.21619. 2

[Mar75] G. E. Martin, *The foundations of geometry and the non-Euclidean plane*, Springer, 1975. 28

[MdlS24a] M. Magee and M. de la Salle, $\mathrm{SL}_4(\mathbf{Z})$ *is not purely matricial field*, C. R., Math., Acad. Sci. Paris **362** (2024), 903–910. 8

[MdlS24b] ———, *Strong asymptotic freeness of Haar unitaries in quasi-exponential dimensional representations*, 2024, arXiv:2409.03626. 3, 4, 6, 7, 8, 11, 12, 32, 34

[Mir13] M. Mirzakhani, *Growth of Weil-Petersson volumes and random hyperbolic surface of large genus*, J. Differ. Geom. **94** (2013), 267–300. 8

[MNP22] M. Magee, F. Naud, and D. Puder, *A random cover of a compact hyperbolic surface has relative spectral gap* $\frac{3}{16} - \varepsilon$, Geom. Funct. Anal. **32** (2022), 595–661. 8, 9, 17

[MP22] M. Magee and D. Puder, *Core surfaces*, Geom. Dedicata **216** (2022), 46. 5, 18

[MP23] ______, *The asymptotic statistics of random covering surfaces*, Forum Math. Pi **11** (2023), e15. 3, 5, 6, 7, 17, 18, 19, 20, 22, 33

[MT23] M. Magee and J. Thomas, *Strongly convergent unitary representations of right-angled Artin groups*, 2023, arXiv:2308.00863. 8

[Par24] F. Parraud, *The spectrum of a tensor of random and deterministic matrices*, 2024, arXiv:2410.04481. 8

[Pud15] D. Puder, *Expansion of random graphs: New proofs, new results*, Invent. math. **201** (2015), 845–908. 8

[Sel65] A. Selberg, *On the estimation of Fourier coefficients of modular forms*, Proc. Sympos. Pure Math. 8, 1-15 (1965)., 1965. 3

[Tim63] A. F. Timan, *Theory of approximation of functions of a real variable. Translated by J.Berry*, International Series of Monographs on Pure and Applied Mathematics. 34. Oxford etc.: Pergamon Press, 1963. 15

[WX22] Y. Wu and Y. Xue, *Random hyperbolic surfaces of large genus have first eigenvalues greater than* $\frac{3}{16} - \varepsilon$, Geom. Funct. Anal. **32** (2022), 340–410. 8

Michael Magee
Department of Mathematical Sciences, Durham University, DH1 3LE Durham, UK
`michael.r.magee@durham.ac.uk`

Doron Puder
School of Mathematical Sciences, Tel Aviv University, Tel Aviv, 6997801, Israel
and IAS Princeton, School of Mathematics, 1 Einstein Drive, Princeton 08540, USA
`doronpuder@gmail.com`

Ramon van Handel
Department of Mathematics, Princeton University, Princeton, NJ 08544, USA
`rvan@math.princeton.edu`