Electronic Colloquium on Computational Complexity, Report No. 163 (2022)

Random Walks on Rotating Expanders

Gil Cohen* Gal Maor†

November 16, 2022

Abstract

Random walks on expanders are a powerful tool which found applications in many areas of theoretical computer science, and beyond. However, they come with an inherent cost – the spectral expansion of the corresponding power graph deteriorates at a rate that is exponential in the length of the walk. As an example, when G is a d-regular Ramanujan graph, the power graph G^t has spectral expansion $2^{\Omega(t)}\sqrt{D}$, where $D = d^t$ is the regularity of G^t, thus, G^t is $2^{\Omega(t)}$ away from being Ramanujan. This exponential blowup manifests itself in many applications.

In this work we bypass this barrier by permuting the vertices of the given graph after each random step. We prove that there exists a sequence of permutations for which the spectral expansion deteriorates by only a *linear* factor in t. In the Ramanujan case this yields an expansion of $O(t\sqrt{D})$. We stress that the permutations are tailor-made to the graph at hand and require no randomness to generate.

Our proof, which holds for all sufficiently high girth graphs, makes heavy use of the powerful framework of finite free probability and interlacing families that was developed in a seminal sequence of works by Marcus, Spielman and Srivastava.

*Tel Aviv University. gil@tauex.tau.ac.il. Supported by ERC starting grant 949499 and by the Israel Science Foundation grant 1569/18.

†Tel Aviv University. galmaor@mail.tau.ac.il. Supported by ERC starting grant 949499.

ISSN 1433-8092

Contents

1 Introduction **1**
1.1 Step-permute-step random walks 2
1.2 Our results 3

2 Proof Overview **5**
2.1 Analyzing $\mathbf{A}^{\circlearrowleft}(x)$: the Haar-expected characteristic polynomial 7
2.2 Quadrature: from Haar to random permutations 10
2.3 Interlacing: from random permutations to a tailor-made sequence 11

3 Preliminaries **12**
3.1 Haar distribution on the orthogonal group 12
3.2 Distributions and transforms 13
3.3 The Kesten-McKay distribution 15

4 Haar Analysis **16**
4.1 The adapter: formalizing the cheat and uncheat idea 17
4.2 Haar rotated Ramanujan graphs 22
4.3 Haar rotated expander graphs 27

5 Quadrature: From Haar to Random Permutations **28**
5.1 Proof of Lemma 5.7 33
5.2 Proof of Theorem 5.2 35

6 Interlacing: From Random Permutations to a Tailor-Made Sequence **36**

7 Proof of Theorem 1.1 and Theorem 1.2 **44**

A Missing Proofs **48**

B The construction of $\mathcal{T}_1$ **49**

1 Introduction

Random walks on expanders are a powerful tool which found applications in many areas of theoretical computer science, and beyond. This is due to the expander's highly desired pseudorandom properties such as the hitting property [AKS87, Kah95] and the expander Chernoff bound [AKS87, Gil98, Hea08]. The extent to which random walks are pseudorandom is studied to this day (e.g., [TS17, GK21, CPTS21, JM21, CMP+22, GV22]). Random walks are key primitives in many works in theoretical computer science, including several seminal results such as Ta-Shama's state-of-the-art construction of small-bias sets [TS17], Reingold's undirected connectivity in log-space [Rei08] and Dinur's proof of the PCP Theorem by gap amplification [Din07]. More intrinsically, expander random walks are used in several constructions of expander graphs [RVW00, BATS11]. We refer the reader to the wonderful texts [HLW06, Vad12, Tre17, Spi19] for an extensive treatment of expander graphs.

While a random walk is a key primitive, it has an inherent cost we wish to address. To this end, we first recall some basic definitions and set some notation. Let $G = (V, E)$ be a d-regular undirected graph with adjacency matrix $\mathbf{A}$. As $\mathbf{A}$ is symmetric, its eigenvalues are all real, and are denoted by $d = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The *spectral expansion*[1] of G, denoted $\lambda(G)$, is defined by $\lambda(G) = \max(\lambda_2(G), |\lambda_n(G)|)$. A graph G is called a λ-*spectral expander* if $\lambda(G) \leq \lambda$.

To illustrate the inherent cost eluded to above, consider for example the case where G is a d-regular Ramanujan graph, that is, $\lambda(G) \leq 2\sqrt{d-1}$. A length-$t$ random walk on G is studied by analyzing the operator $\mathbf{A}^t$ which is the adjacency matrix of the $D \triangleq d^t$ regular graph G^t that encodes length-t walks on G. As

$$\lambda(G^t) = \lambda(G)^t = \left(2\sqrt{d-1}\right)^t = 2^{\Omega(t)}\sqrt{D},$$

taking a length-t random walk has the effect of deteriorating the spectral expansion of the graph in a rate that is *exponential* in t. Indeed, the bound $2^{\Omega(t)}\sqrt{D}$ should be compared with $2\sqrt{D}$ which is, roughly, the spectral expansion of a D-regular Ramanujan graph. This deterioration is a real phenomena - it is not an artifact of some loose analysis. [2]

The question that we consider in this work is whether there is an alternative to random walks which has a slower deterioration of the spectral expansion and yet this alternative

[1]There is some harmless inconsistency in the literature regarding the definition of spectral expansion. Some sources refer to $d - \lambda(G)$ as the spectral expansion. Others consider $1 - \frac{1}{d}\lambda(G)$. In some cases, it is only λ_2 that is considered.

[2]It is worth noting that a small improvement can be obtained by removing the self loops in the power graph G^t. This, however, has little effect on the deterioration which, in particular, remains exponential in t.

is similar enough to taking a random walk so that many of the analyses that use a random walk "go through" under this variant.

Before we proceed to investigate this idea, we remark that random walks on directed graphs, and also non-backtracking random walks on undirected graphs, do not suffer this exponential blowup. We, however, focus on (unconstrained) random walks on undirected graphs as these have myriad applications to theoretical computer science. Indeed, in many scenarios, even if one considers the adjacency matrix $\mathbf{A}$ of a directed graph, the analysis typically boils down to bounding $\|\mathbf{A}\|$ [3] or a variant thereof. As $\|\mathbf{A}\| = \sqrt{\|\mathbf{A}^\mathsf{T}\mathbf{A}\|}$, one ends up studying the corresponding symmetrization. Put differently, more often than not the singular values of $\mathbf{A}$ rather than its eigenvalues that are of interest.

1.1 Step-permute-step random walks

Influenced by the seminal sequence of works by Marcus, Spielman and Srivastava [MSS13, MSS18, MSS22] on the existence of bipartite Ramanujan graphs, the proposal that we put forth in this work is to permute the vertices of the given graph G after each random step, where the permutations to be used are *tailor-made* to G and require *no randomness* to generate. This is a key point as in many applications the randomness used for the walk is the expensive resource. We turn to explore this idea for Ramanujan graphs, and start by considering the case $t = 2$. We will use this toy example also for refining our suggestion.

1.1.1 What does one permutation buy us?

Given a d-regular Ramanujan graph $G = (V, E)$ and a permutation matrix $\mathbf{P}$ on V, instead of considering a length-2 random walk, we consider permuting the vertices after the first step according to the permutation, namely, we look at the operator $\mathbf{PAP}^\mathsf{T} \cdot \mathbf{A}$. This suggestion has the significant drawback that the graph which corresponds to the resulted matrix is directed. Thus, we refine our suggestion and instead consider the operator $\mathbf{A} \cdot \mathbf{PAP}^\mathsf{T} \cdot \mathbf{A}$ which corresponds to a length-3 random walk, where the first step is done according to G, the second according to the permuted G, and the third is again according to G. For technical reasons, we will in fact consider the operator

$$\mathbf{A_P} \triangleq \mathbf{A} \cdot \mathbf{PA}^2\mathbf{P}^\mathsf{T} \cdot \mathbf{A}.$$

$\mathbf{A_P}$ is the adjacency matrix of a $D = d^4$-regular graph, which we denote by $G_\mathbf{P}$, and so for every permutation matrix $\mathbf{P}$, the value $\lambda(G_\mathbf{P})$ is somewhere between $2\sqrt{D}$–the

[3] Throughout the paper, $\|\cdot\|$ means the induced 2-norm $\|\mathbf{A}\|_2 = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2$.

bound for a D-regular Ramanujan graph–and $\lambda(G^4) \approx 16\sqrt{D}$. Although we care about the asymptotic behavior in t, we cannot help ourselves from digressing slightly and ask: What does one permutation buy us?

From our more general result stated below (see Theorem 1.1) it follows that for every d-regular Ramanujan graph G there exists a permutation $\mathbf{P} = \mathbf{P}(G)$ such that

$$\lambda(G_{\mathbf{P}}) \leq \frac{27}{4}\sqrt{D} + \varepsilon.$$

The ε term should be thought of as negligible since it vanishes at an exponential rate with the girth of G. In particular, if G has a sufficiently large (constant) girth then the bound obtained is below $7\sqrt{D}$. It remains open whether the constant $\frac{27}{4}$ is tight (up to a vanishing error term), that is, whether there exists a permutation $\mathbf{P}$ such that $\lambda(G_{\mathbf{P}})$ is significantly smaller than $\frac{27}{4}\sqrt{D}$. Simulations suggest that the $\frac{27}{4}$ bound from our analysis represents the typical behavior.

1.1.2 The general case

More generally, we ask, given a d-regular Ramanujan graph $G = (V, E)$ and an integer $t \geq 2$, does there exist a sequence of permutation matrices $\mathbf{P} = (\mathbf{P}_1, \ldots, \mathbf{P}_{t-1})$ on V such that the graph $G_{\mathbf{P}}$, whose adjacency matrix is given by

$$\mathbf{A}_{\mathbf{P}} \triangleq \mathbf{A}\mathbf{P}_{t-1} \cdots \mathbf{A}\mathbf{P}_1\mathbf{A}^2\mathbf{P}_1^{\mathsf{T}}\mathbf{A} \cdots \mathbf{P}_{t-1}^{\mathsf{T}}\mathbf{A},$$

has a spectral expansion that avoids the exponential deterioration in t? Note that $G_{\mathbf{P}}$ is $D = d^{2t}$-regular, and so we should compare $\lambda(G_{\mathbf{P}})$ with the optimal possible value of $\approx 2\sqrt{D}$ on the one hand, and with $\approx 4^t\sqrt{D}$ on the other hand. Indeed, the latter is an upper bound for every choice of $\mathbf{P}$ and is attained by a standard random walk, namely, by picking $\mathbf{P}_1 = \cdots = \mathbf{P}_{t-1} = \mathbf{I}$.

1.2 Our results

Our first result deals with the particular case of Ramanujan graphs. Roughly speaking, we prove that a sequence of permutation matrices, tailor-made to the graph at hand, exists such that the deterioration of the spectral expansion in t, when permuted accordingly, is *linear* rather than exponential. More precisely though still somewhat informally, we prove

Theorem 1.1 (Main result for Ramanujan graphs; informal)**.** For every d-regular Ramanujan graph G and for every integer $t \geq 2$ there exists a sequence of permutation matrices $\mathbf{P} = (\mathbf{P}_1, \dots, \mathbf{P}_{t-1})$ such that

$$\lambda(G_{\mathbf{P}}) \leq \left(1 + \frac{1}{t}\right)^t (t+1)\sqrt{D} + \varepsilon$$
$$< e(t+1)\sqrt{D} + \varepsilon,$$

where the reader should think of ε as a vanishing term and, as before, $D = d^{2t}$.

Theorem 1.1 is stated in a somewhat informal manner. The complete and formal statement is the content of Theorem 7.1, though already here we wish to say that $\varepsilon = 2^{-\Omega(g)} \cdot (dt)^{O(t)}$ where g is the girth of G. Thus, indeed, for sufficiently high girth, constant-degree graphs [4], ε is not only bounded but in fact vanishes, and so it can be ignored.

For general d-regular λ-spectral expanders we obtain the following result.

Theorem 1.2 (Main result for general graphs; informal)**.** For every d-regular λ-spectral expander G and for every integer $t \geq 2$ there exists a sequence of permutation matrices $\mathbf{P} = (\mathbf{P}_1, \dots, \mathbf{P}_{t-1})$ such that

$$\lambda(G_{\mathbf{P}}) \leq \begin{cases} O(\lambda^2 d^{t-1}) = O(\lambda\sqrt{D}), & t < \frac{8\lambda^2}{d}; \\ \left(1 + \frac{1}{t}\right)^t (t+1)d^t + \varepsilon < e(t+1)\sqrt{D} + \varepsilon, & \text{otherwise.} \end{cases}$$

Here ε is similarly bounded as in Theorem 1.1.

Again, Theorem 1.2 is stated somewhat informally. The complete and formal statement is the content of Theorem 7.2. The proof of Theorem 1.2 is similar to that of Theorem 1.1 and, up to constants, imply the latter.

Theorem 1.2, and in particular the split of the bound to the two cases, can be interpreted as follows: Taking the first $O(\frac{\lambda^2}{d})$ steps, the improvement made to the initially poor spectral expansion, λ, by the tailor-made permutations outweigh the product structure of the operator $\mathbf{A}_{\mathbf{P}}$. Indeed, the improvement is quite dramatic – a $1 \cdot \sqrt{d}$ factor to the spectral expansion per step – a value that cannot be attained without the permutations even if the initial graph would have been Ramanujan. Reaching the threshold value, the product structure takes it toll, though, on the positive side, with no reference whatsoever to the fact that we started with an initial poor spectral expansion of λ.

It is interesting to compare the guarantee of Theorem 1.2, in the early interval $t = O(\frac{\lambda^2}{d})$, with the recent work of Jeronimo, Mittal, Roy, and Wigderson [JMRW22] who showed how the spectral expansion of a graph can be improved by local operations.

[4] For example, the seminal Lubotzky-Phillips-Sarnak construction [LPS88] is of d-regular Ramanujan graphs having girth $\Omega(\log_d n)$.

Theorem 1.2 accomplishes that as well since short random walks can be computed locally. Theorem 1.2, however, only guarantees the existence of a sequence of permutations whereas [JMRW22] gives an efficient algorithm for performing the local operations. This leads us to highlight what we consider to be an interesting open problem.

Open Problem 1.3. Devise an efficient algorithm that, given a graph G, computes a sequence of permutations $\mathbf{P} = \mathbf{P}(G)$ as guaranteed to exists by Theorem 1.2 (or even for the special case of Ramanujan graphs, Theorem 1.1).

In the context of expander graphs, by the word "efficient" that appears in Open Problem 1.3, one can either mean in time $\mathrm{poly}(n)$, n being the size of the graph or, more ambitiously, that the permutations can be computed in way that will allow one to compute the neighbors of a vertex in time which is polynomial in the corresponding input's length, namely, in $\mathrm{poly}(\log n)$ time. While the second interpretation seems to be out of reach of current techniques, the first may be possible to obtain using ideas from the paper by Cohen [Coh16] who constructed Ramanujan graphs in polynomial time based on the same techniques we are using [MSS18].

2 Proof Overview

In this section we give an informal, yet comprehensive, overview of our proofs. The reader may freely skip this section, moving on to the formal sections.

Let us take a closer look at the problem with expander random walks. If G is an undirected graph on n vertices with adjacency matrix $\mathbf{A}$ then, by the spectral decomposition theorem, we can write $\mathbf{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i^\mathsf{T}$, where the $\boldsymbol{\psi}_i$-s are corresponding orthonormal eigenvectors of $\mathbf{A}$. As the $\boldsymbol{\psi}_i$-s are orthonormal,

$$\mathbf{A}^t = \left(\sum_{i=1}^{n} \lambda_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i^\mathsf{T} \right)^t = \sum_{i=1}^{n} \lambda_i^t \boldsymbol{\psi}_i \boldsymbol{\psi}_i^\mathsf{T}.$$

This simple calculation sheds light on the reason for the equality $\lambda(G^t) = \lambda(G)^t$. Indeed, each eigenvector, in particular the "heavy" ones (e.g., $\boldsymbol{\psi}_2$ or $\boldsymbol{\psi}_n$, but potentially also $\boldsymbol{\psi}_3$ etc.) is in a perfect alignment with itself and so, say, a stretch of magnitude λ_2 occurs t times in the *same* direction. This suggests that one may benefit by "rotating" the expander, or rather its eigenvectors, so as to break these alignments. Note that it does not suffice to merely break the alignment of $\boldsymbol{\psi}_2$ with itself as, say, λ_3 may be as large, or almost as large, as λ_2.

The problem with just rotating the eigenvectors of $\mathbf{A}$ is that we will lose the graph structure. More precisely, the graph that corresponds to some rotated version of $\mathbf{A}$ is likely to be a complete graph with both positive as well as negative edge weights. Still, we will proceed with this thought experiment and, for the time being, play loose with the graph structure.

As always, a good first step is to try to understand what can be said in expectation. To be precise, we will take $\mathbf{Q} = (\mathbf{Q}_1, \ldots, \mathbf{Q}_{t-1})$ to be independent Haar distributed orthogonal matrices [5] and would like to explore $\mathbb{E}_{\mathbf{Q}} \|\mathbf{A}_{\mathbf{Q}}\|$ as this quantity captures the expected bound on the support of the spectrum of $\mathbf{A}_{\mathbf{Q}}$. However, it turns out that it will be extremely beneficial to work with more information by tracking down the entire spectrum. Inspired by the seminal sequence of works by Marcus, Spielman and Srivastava [MSS13, MSS18, MSS22], we do this by studying the expected characteristic polynomial

$$\mathbf{A}^{\circlearrowleft}(x) \triangleq \mathop{\mathbb{E}}_{\mathbf{Q}} \chi_x(\mathbf{A}_{\mathbf{Q}}).$$

The first thing worth emphasizing is that while each of the polynomials that participate in the expectation, $\chi_x(\mathbf{A}_{\mathbf{Q}})$, is real-rooted, the fact that $\mathbf{A}^{\circlearrowleft}(x)$ is real-rooted is a nontrivial, though true, statement. Even with this in mind, whatever we conclude by studying the roots of $\mathbf{A}^{\circlearrowleft}(x)$ is insufficient by itself for two reasons:

1. It is unclear how to deduce anything from the Haar-expected behavior on a particular choice of $\mathbf{Q}$, let alone on a sequence of permutation matrices. That is, the expectation is taken with respect to the coefficient space whereas we are interested in the eigenvalues which are highly non-linear in the coefficients.

2. As mentioned, typically, rotating (or, more precisely, applying an orthogonal operator to) the eigenvectors of a d-regular graph does not yield a d-regular graph.

Still, in the following section (Section 2.1) we will proceed with analyzing $\mathbf{A}^{\circlearrowleft}(x)$. Both our analysis of $\mathbf{A}^{\circlearrowleft}(x)$ as well as the solution to the above two problems make heavy use of the powerful framework of finite free probability and interlacing families, as well as a certain quadrature result that was introduced by Marcus, Spielman and Srivastava [MSS13, MSS15, MSS18, MSS22] in which the authors prove the existence of bipartite Ramanujan graphs of every degree and size, and further prove Weaver's conjecture [Wea04] which, by extension, resolves the Kadison-Signer problem [KS59].

[5] The reader unfamiliar with the Haar measure on the group of orthogonal matrices is referred to Section 4. The reader that is familiar with this Haar measure will observe that "rotation" does not precisely capture what is suggested as also reflections, and compositions of rotations and reflections, are involved. Nonetheless, we choose simplicity over accuracy in our name-giving. A remark regarding our notation is in place: Throughout the paper, we will denote permutation matrices by $\mathbf{P}$-s and Haar distributed orthogonal matrices by $\mathbf{Q}$-s.

With these caveats in mind, we turn to analyze $\mathbf{A}^{\circlearrowright}(x)$. The other aspects mentioned above are treated afterwards in Section 2.2 and Section 2.3.

2.1 Analyzing $\mathbf{A}^{\circlearrowright}(x)$: the Haar-expected characteristic polynomial

The problem with analyzing $\mathbf{A}^{\circlearrowright}(x)$ is that it is difficult to work out with a given graph G. On the positive side, note that due to the Haar random matrices that we apply, $\mathbf{A}^{\circlearrowright}(x)$ depends only on the spectrum of $\mathbf{A}$, namely, it does not depend on the latter's eigenvectors. This leads us to our strategy, which is to *cheat*.

Instead of analyzing $\mathbf{A}^{\circlearrowright}(x)$, we are going to replace the distribution of eigenvalues of $\mathbf{A}$ with the Kesten-McKay distribution. The latter is a limit object that does not correspond to any particular finite graph. It does, however, correspond to the spectrum of the d-regular infinite tree [6], and so one might hope that this analysis will shed light on high girth finite graphs. At any rate, our starting point is in observing that $\mathbf{A}^{\circlearrowright}(x)$ can be expressed neatly using the multiplicative convolution.

2.1.1 The multiplicative convolution

We make use of the following key definition and result which can be found in [MSS22].

Definition 2.1 (Multiplicative convolution)**.** Let $\mathbf{A}, \mathbf{B}$ be real symmetric matrices of equal order with characteristic polynomials $a(x), b(x)$. The multiplicative convolution $a \boxtimes b$ is the polynomial defined by

$$(a \boxtimes b)(x) = \mathop{\mathbb{E}}_{\mathbf{Q}} \chi_x(\mathbf{AQBQ}^{\mathsf{T}}),$$

where $\mathbf{Q}$ is Haar random orthogonal matrix.

It should be noted that although the right hand side seems to depend on the eigenvectors of $\mathbf{A}, \mathbf{B}$ it in fact depends solely on the spectrum of these matrices, due to the Haar random $\mathbf{Q}$, and so the $\boxtimes$ operator, which note receives only information about the spectrum of the matrices, is well-defined.

Using the multiplicative convolution, and the cyclic-invariant property of the characteristic polynomial, we can write

$$\mathbf{A}^{\circlearrowright}(x) = \chi_x(\mathbf{A}^2)^{\boxtimes t}, \tag{2.1}$$

[6] To formally define the spectrum of an infinite operator, even one with countable dimension, requires some background and we anyhow won't be needing this point of view.

where the notation on the RHS means we take the multiplicative convolution on t copies of $\chi_x(\mathbf{A}^2)$, noting that the multiplicative convolution is both associative as well as commutative. E.g., for $t = 2$,

$$\mathbf{A}^{\circlearrowright}(x) = \mathop{\mathbb{E}}_{\mathbf{Q}_1} \chi_x(\mathbf{A}\mathbf{Q}_1\mathbf{A}^2\mathbf{Q}_1^{\mathsf{T}}\mathbf{A}) = \mathop{\mathbb{E}}_{\mathbf{Q}_1} \chi_x(\mathbf{A}^2\mathbf{Q}_1\mathbf{A}^2\mathbf{Q}_1^{\mathsf{T}}) = \chi_x(\mathbf{A}^2) \boxtimes \chi_x(\mathbf{A}^2) = \chi_x(\mathbf{A}^2)^{\boxtimes 2}.$$

The multiplicative convolution was studied by Marcus, Spielman and Srivastava around ten years ago, and has its origins in the "infinite" or limit case in free probability theory, and before that, in the study of random matrices. Analyzing the multiplicative convolution is done by studying several analytic transform which we turn to discuss.

2.1.2 Transforms

Let μ be a distribution on $[0, a]$ for some real $a > 0$. The *Cauchy transform* of μ is defined by

$$\mathcal{G}_\mu(x) = \int_0^a \frac{1}{x-t}\mu(t)dt.$$

Following Marcus, Spielman and Srivastava, we will study $\mathcal{G}_\mu$ as a function on (a, ∞) though we remark that if one wish to extract more information about μ than bounding it support, it is beneficial to study $\mathcal{G}_\mu$ as a function on $\mathbb{C}^+$. We further note that $\mathcal{G}_\mu$ is essentially the moment generating function of μ around the point at infinity. More precisely,

$$\mathcal{G}_\mu(x) = \sum_{r=0}^{\infty} \frac{m_r(\mu)}{x^{r+1}},$$

where $m_r(\mu)$ is the r-th moment of μ.

We extend the definition of $\mathcal{G}_\mu$ to real-rooted polynomials $p(x)$ by defining μ as the uniform distribution over roots of $p(x)$, where repeated roots are sampled accordingly. We further extend $\mathcal{G}_\mu$ to real symmetric matrices by considering the corresponding characteristic polynomial. E.g., if $\mathbf{A}$ is positive semidefinite with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ then

$$\mathcal{G}_{\mathbf{A}}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{x-\lambda_i}.$$

The $\mathcal{M}_\mu$ transform is defined by

$$\mathcal{M}_\mu(x) = x\mathcal{G}_\mu(x) - 1 = \sum_{r=1}^{\infty} \frac{m_r(\mu)}{x^r}. \tag{2.2}$$

We define $\mathcal{N}_\mu(y)$ to be the largest x such that $\mathcal{M}_\mu(x) = y$, assuring the reader that this is well-defined in our case, namely, when μ is supported on $[0, a]$. A simple yet key observation is that for every $y > 0$, $\mathcal{N}_{\mathbf{A}}(y)$ is an upper bound on λ_1. Thus, to bound $\|\mathbf{A}\| = \lambda_1$ for some PSD matrix $\mathbf{A}$ of interest, we will find a good choice y_0 for y, ideally a minimizer of $\mathcal{N}_{\mathbf{A}}(y)$, and conclude that $\|\mathbf{A}\| \leq \mathcal{N}_{\mathbf{A}}(y_0)$. Lastly, it will be convenient to define the $\mathcal{S}_\mu$-transform by $\mathcal{S}_\mu(y) = \frac{y}{y+1}\mathcal{N}_\mu(y)$. We make use of the following powerful result.

Theorem 2.2 (Theorem 1.12 of [MSS22])**.** For all polynomials $p(x), q(x)$ with non-negative real roots and for every $y > 0$, $\mathcal{S}_{p \boxtimes q}(y) \leq \mathcal{S}_p(y) \cdot \mathcal{S}_q(y)$.

2.1.3 Analyzing $\mathbf{A}^{\circlearrowleft}(x)$ by cheating

Recall that $\mathcal{N}_{\mathbf{A}}(y)$ is the "max-inverse" of $\mathcal{M}_{\mathbf{A}}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{\lambda_i}{x-\lambda_i}$ which is hard to get a handle on. Instead, we work with the Kesten-McKay distribution [McK81] which is a continuous measure that is given by

$$\mu_{\mathsf{km}}(t) = \begin{cases} \dfrac{d\sqrt{4(d-1)-t^2}}{2\pi(d^2-t^2)}, & \text{for } |t| \leq 2\sqrt{d-1}; \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

Observing Equation (2.1), we will in fact be interested in the square of the Kesten-McKay distribution, denoted here km^2. One can calculate the corresponding $\mathcal{S}$-transform,

$$\mathcal{S}_{\mathsf{km}^2}(y) = d^2 \cdot \frac{y+1}{y+d}.$$

Had $\mathcal{S}_{\mathsf{km}^2}(y)$ been the $\mathcal{S}$-transform of a polynomial (rather than of a continuous measure), we could have deduce, using Theorem 2.2, that

$$\mathcal{S}_{(\mathsf{km}^2)^{\boxtimes t}}(y) \leq d^{2t} \cdot \left(\frac{y+1}{y+d}\right)^t.$$

Hence,

$$\mathcal{N}_{(\mathsf{km}^2)^{\boxtimes t}}(y) \leq \left(\frac{d^2}{y+d}\right)^t \cdot \frac{(y+1)^{t+1}}{y}. \tag{2.4}$$

We will perform this illegal step as part of the "cheat" we are anyhow to blame for.

Had it not been for the y in the denominator of Equation (2.4), we could have plugged in $y=0$ to get a bound of $d^t = \sqrt{D}$ which is too much to hope for as it is off, even for a Ramanujan graph, by a factor of two. However, one can show that by taking $y_{\mathsf{min}} = \frac{d}{dt-t-1}$ which minimizes the bound in Equation (2.4), one gets

$$\mathcal{N}_{(\mathsf{km}^2)^{\boxtimes t}}(y_{\mathsf{min}}) \leq \left(1+\frac{1}{t}\right)^t (t+1)d^t < e(t+1)d^t = O(t\sqrt{D}).$$

Thus, modulo the cheating, one can deduce the desired linear deterioration in t, namely,

$$\mathsf{maxroot}\, \mathbf{A}^{\circlearrowleft}(x) = O(t\sqrt{D}).$$

2.1.4 Uncheat

To obtaining the result about the graph at hand, we use the observation [McK81] that the first $\frac{g}{2}$ moments of $\mathbf{A}$ and km are equal, g being the girth of G. Hence, by Equation (2.2), $\mathcal{M}_{\mathbf{A}^2}(x)$ and $\mathcal{M}_{\mathsf{km}^2}(x)$ differ only on their respective tails. In particular, $\mathcal{M}_{\mathbf{A}^2}(x) \approx \mathcal{M}_{\mathsf{km}^2}(x)$ for a sufficiently large x. From this, with some technical work, we are able to show that $\mathcal{N}_{\mathbf{A}^2}(y) \approx \mathcal{N}_{\mathsf{km}^2}(y)$ when y is taken from some interval (see Corollary 4.10). Thus, the bound computed for $\mathcal{N}_{\mathsf{km}^2}(y_{\mathsf{min}})$ can be used to bound $\mathcal{N}_{\mathbf{A}^2}(y_{\mathsf{min}})$ losing only the error term $\varepsilon = |\mathcal{N}_{\mathbf{A}^2}(y_{\mathsf{min}}) - \mathcal{N}_{\mathsf{km}^2}(y_{\mathsf{min}})|$. We refer to this part of the proof as the "adapter" as it allows us to plug the spectrum of a graph to the analysis that works given the Kesten-McKay distribution.

We remark that the above is an oversimplified overview. In particular, our analysis requires a somewhat delicate technical work if we wish the analysis to hold for all $t \geq 2$. Moreover, to prove the result for general λ-spectral expanders, not necessarily Ramanujan graphs, some technical complications occur, in particular, we cannot simply take the minimizer y_{min} and instead we need to choose the best y_0 possible under the various constraints that present themselves throughout the analysis. These, however, are technical details that we choose to omit from this informal proof sketch.

2.2 Quadrature: from Haar to random permutations

The next step in the proof, following the MSS framework, is to move from the Haar-expected analysis that was discussed in Section 2.1 to a statement about expectation with respect to random permutations. The key ingredient in accomplishing this is the following lemma.

Lemma 2.3. *Let $\mathbf{A}, \mathbf{B}$ be real $n \times n$ symmetric matrices such that $\mathbf{A}\mathbf{1} = a\mathbf{1}$ and $\mathbf{B}\mathbf{1} = b\mathbf{1}$. Denote by $p_{\mathbf{A}}, p_{\mathbf{B}}$ the polynomials satisfying $\chi_x(\mathbf{A}) = (x-a)p_{\mathbf{A}}(x)$, $\chi_x(\mathbf{B}) = (x-b)p_{\mathbf{B}}(x)$. Let $\mathbf{P}$ be a uniformly random $n \times n$ permutation matrix. Then,*

$$\mathop{\mathbb{E}}_{\mathbf{P}} \chi_x(\mathbf{A}\mathbf{P}\mathbf{B}\mathbf{P}^{\mathsf{T}}) = (x - ab)\,(p_{\mathbf{A}} \boxtimes p_{\mathbf{B}})\,(x).$$

Note that

$$(p_{\mathbf{A}} \boxtimes p_{\mathbf{B}})\,(x) = \mathop{\mathbb{E}}_{\mathbf{Q}} \chi_x(\widehat{\mathbf{A}}\mathbf{Q}\widehat{\mathbf{B}}\mathbf{Q}^{\mathsf{T}}),$$

where $\widehat{\mathbf{A}}$ is the operator induced by $\mathbf{A}$ when restricted to $\mathbf{1}^{\perp}$, and similarly for $\widehat{\mathbf{B}}$. Therefore, Lemma 2.3 should be understood as relating the Haar-expected behavior of the "error operator" [7] to the expected behavior with respect to random permutations. Thus,

$$\lambda\left(\mathop{\mathbb{E}}_{\mathbf{P}} \chi_x(\mathbf{A}_{\mathbf{P}})\right) = \mathsf{maxroot}\, \widehat{\mathbf{A}}^{\circlearrowleft}(x)$$

(see Lemma 5.9). Therefore, the bound obtained for the Haar-expected behavior serves as a bound on the largest eigenvalue of the expected characteristic polynomial with respect to random permutations.

The last step in the proof, discussed in Section 2.3 below, allows one to deduce a bound on $\lambda(\mathbf{A}_{\mathbf{P}})$ for some *specific* permutation sequence $\mathbf{P}$ from the bound obtained for the expected characetristic polynomial. Before moving on, we make some remarks.

Lemma 2.3 is the multiplicative analog of Theorem 4.1 of [MSS18] who considered the additive case, namely, $\mathbb{E}_{\mathbf{Q}}\, \chi_x(\mathbf{A} + \mathbf{Q}\mathbf{B}\mathbf{Q}^{\mathsf{T}})$. Our proof mimics the latter though it requires a bit more technical work (see Lemma 5.7). This part of the proof is dubbed, by MSS, the *quadrature* step as it expresses an integral (the Haar expectation) as a finite sum (expectation with respect to permutations).

The proof, which will occupy us in Section 5, proceeds roughly as follows. First, one observes that when a permutation $\mathbf{P}$ is sampled uniformly at random, its orthogonal projection to $\mathbf{1}^{\perp}$, $\widehat{\mathbf{P}}$, is a random element in the symmetry group of the n-vertex regular simplex, embedded in $\mathbb{R}^{n-1}$. It is a well-known geometric fact that the group of Haar random orthogonal matrices is generated by the two-dimensional Haar random matrices in the sense that every such matrix can be decomposed to matrices acting on the faces of the simplex. Thus, the problem essentially boils down to the plane. A Haar random orthogonal matrix in the plane is nothing more than rotations and reflections, and so, with some work, the desired statement can be shown to hold for dimension two and, by the above, to any dimension.

2.3 Interlacing: from random permutations to a tailor-made sequence

So far we discussed how to obtain a bound on the largest root of the expected characteristic polynomial $\mathbb{E}[\chi_x(\mathbf{A}_{\mathbf{P}})]$, excluding the trivial root. It is generally false that a bound on the (second) largest root of the expectation of polynomials can be used to deduce a bound on the (second) largest root of one of the polynomials that participate in the expectation. It is not even true that the expectation is necessarily real-rooted given that all polynomials in the expectation are.

[7] To clarify the term "error operator", recall that when working with an expander, we can decompose the corresponding adjacency matrix $\mathbf{A} = \mathbf{J} + \mathbf{E}$, with $\mathbf{J}$ being the normalized all-ones matrix, and think of $\mathbf{E}$ as the error operator. Here $\mathbf{E}$ plays the role of $\widehat{\mathbf{A}}$.

The key observation of MSS with regards to this issue is that the polynomials that participate in the expectation form an *interlacing family*. Informally, this means that there exists a binary tree whose nodes are labeled by polynomials in such a way that:

1. The leaves are labeled with the polynomials that participate in the expectation.
2. The root is labeled by the expected characteristic polynomial $\mathbb{E}_{\mathbf{P}}\, \chi_x(\mathbf{A}_{\mathbf{P}})$.
3. The polynomial corresponding to an internal node is a convex combination of those corresponding to its sons.
4. The polynomials $p_u(x), p_v(x)$ that correspond to siblings u, v have a common interlacing. That is, there exists a third real-rooted polynomial $q(x)$ such that between every two consecutive roots of $q(x)$ there is precisely one root of each of the polynomials $p_u(x), p_v(x)$.

MSS proved that Properties (3) and (4) guarantee that the (second) largest root of a node bounds from above the (second) largest root of one of its two sons. By proceeding from the root downwards to the leaves one can deduce, using Properties (1) and (2), a bound for a specific sequence of permutations $\mathbf{P}$ from the bound on the second largest root of the expected characteristic polynomial.

Our proof for the existence of one good sequence $\mathbf{P}$ builds on this part of the proof of MSS. We prove that the interlacing property holds in our multiplicative case in a way that is very similar to the proof of the existing interlacing based proofs.

3 Preliminaries

Throughout the paper we denote matrices by capital bold letters, e.g., $\mathbf{A}, \mathbf{B}, \mathbf{W}$. The normalized (with respect to $\|\cdot\|_2$) all-ones vector is denoted by $\mathbf{1}$, where the dimension of the vector is always implicit. The characteristic polynomial of a matrix $\mathbf{A}$, in variable x, is denoted by $\chi_x(\mathbf{A})$. For a real-rooted polynomial $p(x)$ we let $\mathsf{maxroot}(p(x))$ denote its largest root.

3.1 Haar distribution on the orthogonal group

Denote the group of $n \times n$ orthogonal matrices by $\mathcal{O}(n)$. The *Haar distribution* is the unique distribution over $\mathcal{O}(n)$ which is invariant under multiplication (from the right or from the left) with any arbitrary orthogonal matrix from $\mathcal{O}(n)$. We call a matrix drawn from this distribution a *Haar random matrix*. Although not required for our proof, it is illuminating to picture at least one way of how such a matrix can be drawn: one can pick the first column uniformly at random (normalized), next picking the second column uniformly as well, conditioned one being orthogonal to the first, and so on.

An important characteristic of the Haar distribution, which we rely on in this work, can be formalized as follows. Let $\mathbf{A}, \mathbf{B}$ be two arbitrary $n \times n$ symmetric matrices, and $\mathbf{Q}$ be a Haar random matrix. Then the random rotation of either $\mathbf{A}$ or $\mathbf{B}$ according to $\mathbf{Q}$ removes any dependence between the respective eigenvectors of $\mathbf{A}$ and $\mathbf{B}$. More formally, if $\chi_x(\mathbf{A}) = p(x)$ and $\chi_x(\mathbf{B}) = q(x)$, then both expected characteristic polynomials $\mathbb{E}_{\mathbf{Q}}\, \chi_x(\mathbf{A} + \mathbf{QBQ}^\intercal)$ and $\mathbb{E}_{\mathbf{Q}}\, \chi_x(\mathbf{AQBQ}^\intercal)$ depend only on p and q, and not on the eigenvectors of either $\mathbf{A}$ or $\mathbf{B}$.

3.2 Distributions and transforms

We identify a distribution μ on a set S with a function $\mu : S \to [0,1]$ in the natural way. Let $\alpha \in [0,1]$, and note that if $\mu, \nu : S \to [0,1]$ are two distributions then the function $\alpha\mu + (1-\alpha)\nu$ corresponds to the distribution which is obtained by sampling from μ with probability α and sampling from ν with probability $1-\alpha$. If $\mu : S \to [0,1]$ is a distribution and S is a multiplicative group then for $t \in S$ we define the distribution $t\mu : S \to [0,1]$ that is given by $(t\mu)(s) = \mu(t^{-1}s)$. As suggested by the notation, to sample from the distribution $t\mu$ one first sample $s \sim \mu$ and then return ts. We similarly define μt.

Let μ be a probability distribution over $\mathbb{R}$. The distribution μ^2 is defined in the natural way, namely, to sample from μ^2, one samples x from μ and returns x^2.

3.2.1 Transforms

Let μ be a probability distribution over $\mathbb{R}$. The *Cauchy transform* of μ is defined as the function

$$\mathcal{G}_\mu(x) = \int_{\mathbb{R}} \frac{1}{x-t}\mu(t)dt.$$

We remark that in many settings it is instructive to study the Cauchy transform as a function whose domain is $\mathbb{C}^+$. However, we will consider the Cauchy transform as a function on $\mathbb{R}$. More accurately, the distributions that we consider will be of bounded positive support, namely $\mathsf{supp}(\mu) \subseteq [0,a]$, and we will always evaluate the Cauchy transform outside of that support. The Cauchy transform is also related to the moments of a distribution. If $m_r(\mu)$ is the r-th moment of μ then for every $x > a$ we have

$$\mathcal{G}_\mu(x) = \sum_{r=0}^{\infty} \frac{m_r(\mu)}{x^{r+1}}. \tag{3.1}$$

(see Remark 2.19 in [NS06].) Accompanied to the Cauchy transform is the $\mathcal{M}$-*transform* [8] which is defined by

$$\mathcal{M}_\mu(x) = x\mathcal{G}_\mu(x) - 1 = \int_{\mathbb{R}} \frac{t}{x-t}\mu(t)dt.$$

We define the inverse of this transform, $\mathcal{N}_\mu(y)$, to be the largest x so that $\mathcal{M}_\mu(x) = y$. For the aforementioned distributions, this is well-defined, as $\mathcal{M}_\mu$ can be shown to be monotone decreasing in (a, ∞).

The $\mathcal{S}$-transform of μ, denoted $\mathcal{S}_\mu$ is defined by $\mathcal{S}_\mu(y) = \frac{y}{y+1}\mathcal{N}_\mu(y)$. We make use of the following powerful result which bounds the $\mathcal{S}$-transform of the multiplicative convolution (recall Definition 2.1) of two polynomials by the product of their $\mathcal{S}$-transforms.

Theorem 3.1 ([MSS22], Theorem 4.7)**.** *Let $p(x), q(x)$ be polynomials with non-negative real roots. Then, for every $y > 0$, $\mathcal{S}_{p\boxtimes q}(y) \le \mathcal{S}_p(y) \cdot \mathcal{S}_q(y)$.*

As a direct corollary one gets the following result.

Corollary 3.2. *Let $p(x)$ be a polynomial with non-negative real roots. Then, for every $y > 0$,*

$$\mathcal{N}_{p^{\boxtimes t}}(y) \le \left(\frac{y}{y+1}\right)^{t-1} \mathcal{N}_p(y)^t.$$

3.2.2 Distributions and transforms of polynomials, matrices, and graphs

Let $p(x)$ be a degree n real rooted polynomial with roots $\alpha_1 \ge \alpha_2 \ge \cdots \ge \alpha_n$. To $p(x)$ we associate the distribution μ_p that is uniform over its roots, namely, to sample from μ_p one first samples $i \in [n]$ uniformly at random and then returns α_i. Note that

$$\mathcal{G}_{\mu_p}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{x-\alpha_i}.$$

For ease of readability we write $\mathcal{G}_p$ instead of the more cumbersome $\mathcal{G}_{\mu_p}$. We similarly define $\mathcal{M}_p$ and $\mathcal{N}_p$ as a shorthand for $\mathcal{M}_{\mu_p}$ and $\mathcal{N}_{\mu_p}$, respectively. The r-th moment of μ_p is denoted by $m_r(p)$. Furthermore, for the characteristic polynomial $\chi_x(\mathbf{A})$ of a real symmetric matrix $\mathbf{A}$, we denote $\mu_{\chi_x(\mathbf{A})}$ by $\mu_\mathbf{A}$ for short. Similarly, we write $\mathcal{G}_\mathbf{A}$, $\mathcal{M}_\mathbf{A}$ and $\mathcal{N}_\mathbf{A}$ for $\mathcal{G}_{\chi_x(\mathbf{A})}$, $\mathcal{M}_{\chi_x(\mathbf{A})}$, and $\mathcal{N}_{\chi_x(\mathbf{A})}$, respectively.

For an undirected graph G we write μ_G for $\mu_{\mathbf{A}_G}$ where $\mathbf{A}_G$ is the adjacency matrix of G. We write $\mathcal{G}_G$ for $\mathcal{G}_{\mathbf{A}_G}$ and similarly define $\mathcal{M}_G$ and $\mathcal{N}_G$. Throughout the paper

[8] The reader should note that [MSS22] denotes what we refer to as the $\mathcal{M}$-transform by $\widetilde{\mathcal{M}}$. The $\mathcal{N}$-transform and $\mathcal{S}$-transform that we define next are denoted in [MSS22] by $\widetilde{\mathcal{M}}^{(-1)}$ and $\widetilde{\mathcal{S}}$, respectively. We do so for ease of readability.

we will consider d-regular graphs, and in these cases it will be convenient to denote by $\mathcal{G}_{\widehat{G}}$, $\mathcal{M}_{\widehat{G}}$ and $\mathcal{N}_{\widehat{G}}$, the corresponding transforms of the polynomial $\frac{1}{x-d}\chi_x(\mathbf{A})$, where $\mathbf{A}$ is the adjacency matrix of G. We remark that $\mathcal{N}_{\mathbf{A}}$ is well-defined also for a real symmetric matrix $\mathbf{A}$ that is not positive semidefinite.

3.3 The Kesten-McKay distribution

The probability measure of the Kesten-McKay distribution with parameter d is given by

$$\mu_{\mathsf{km}}(t) = \begin{cases} \dfrac{d\sqrt{4(d-1)-t^2}}{2\pi(d^2-t^2)}, & \text{for } |t| \leq 2\sqrt{d-1}; \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

Note that we suppress the parameter d from the notation as it will always be clear from context. The following is a well-known fact from free probability theory (see, e.g., Example 12.8 in [NS06]). It can also be verified by a straightforward calculation.

Claim 3.3. *The Cauchy transform of the Kesten-McKay distribution with parameter d is given by*

$$\mathcal{G}_{\mu_{\mathsf{km}}}(x) = \frac{d\sqrt{x^2-4(d-1)} - x(d-2)}{2(x^2-d^2)}.$$

For ease of readability we denote the Cauchy transform of the Kesten-McKay distribution by $\mathcal{G}_{\mathsf{km}}$, and similarly denote the transforms $\mathcal{M}_{\mathsf{km}}$ and $\mathcal{N}_{\mathsf{km}}$. We will mostly be working with the distribution of μ_{km}^2, which we write as μ_{km^2} and denote its Cauchy transform by $\mathcal{G}_{\mathsf{km}^2}$ rather than by the cumbersome $\mathcal{G}_{\mu_{\mathsf{km}}^2}$. We similarly write $\mathcal{M}_{\mathsf{km}^2}$, $\mathcal{N}_{\mathsf{km}^2}$ for the respective transforms of this distribution.

Claim 3.4. *For μ_{km}^2 with parameter d we have*

$$\mathcal{G}_{\mathsf{km}^2}(x) = \frac{d\sqrt{x-4(d-1)} - \sqrt{x}(d-2)}{2\sqrt{x}(x-d^2)}, \tag{3.3}$$

$$\mathcal{M}_{\mathsf{km}^2}(x) = \frac{2d}{\sqrt{x(x-4(d-1))} + x - 2d}, \tag{3.4}$$

and

$$\mathcal{N}_{\mathsf{km}^2}(y) = \frac{d^2(y+1)^2}{y(y+d)}.$$

The proof of the first two equalities readily follows by Claim 3.5, stated below, and Claim 3.3, and the assertion regarding the $\mathcal{N}$-transform can be verified.

Claim 3.5. *For every symmetric probability distribution μ supported on $(-a,a)$ it holds that $\mathcal{G}_{\mu^2}(x) = \frac{1}{\sqrt{x}}\mathcal{G}_\mu(\sqrt{x})$ for all $x > a^2$.*

For completeness, the easy proof of Claim 3.5 appears in Appendix A. Lastly, the following claim, relating the moments of a d-regular graph with the moments of the Kesten-McKay distribution with parameter d is implicit in [McK81].

Claim 3.6. *Let G be a d-regular graph with girth g. Then, for every $0 \le r < \frac{g}{2}$, $m_r(G) = m_r(\mu_{\mathsf{km}})$.*

4 Haar Analysis

The purpose of this section is proving a bound on the largest non-trivial root of the Haar-expected characteristic polynomial, formalizing the overview given in Section 2.1. The main two results, corresponding to Ramanujan graphs and to general expander graphs, are the following:

Proposition 4.1. *Let G be a d-regular Ramanujan graph on n vertices with girth g. Denote $\bar{g} = \min(g, \frac{1}{6} \cdot \log_d n)$. Then, for every $t \ge 2$,*

$$\mathsf{maxroot}\left((\widehat{G^2})^{\boxtimes t}(x)\right) \le \left(1 + \frac{1}{t}\right)^t (t+1)d^t + (8td)^{t+3} \cdot 2^{-\Omega(\bar{g})}.$$

Proposition 4.2. *Let G be a d-regular λ-spectral expander on n vertices having girth g. Denote $\bar{g} = \min(g, \frac{1}{6} \cdot \log_d n)$. Then,*

$$\mathsf{maxroot}\left((\widehat{G^2})^{\boxtimes t}(x)\right) \le \begin{cases} 5e^2 \cdot \lambda^2 d^{t-1} + (5d\lambda^2)^{t+3} \cdot 2^{-\Omega(\bar{g})}, & t < \frac{8\lambda^2}{d}; \\ \left(1 + \frac{1}{t}\right)^t (t+1)d^t + (t\lambda^2)^{t+4} \cdot 2^{-\Omega(\bar{g})}, & \textit{otherwise.} \end{cases}$$

The proofs of the above are given in Section 4.2 and Section 4.3, respectively. In the sections preceding these, we build the tools needed. For ease of readability, we repeat the definition of the multiplicative convolution as given in Definition 2.1.

Definition 4.3. *Let $\mathbf{A}, \mathbf{B}$ be real symmetric matrices of equal order with characteristic polynomials $a(x)$ and $b(x)$, respectively. The* multiplicative convolution *of $a(x)$ and $b(x)$, denoted $(a \boxtimes b)(x)$ is defined to be*

$$(a \boxtimes b)(x) = \mathop{\mathbb{E}}_{\mathbf{Q}}\left[\chi_x\left(\mathbf{AQBQ}^\mathsf{T}\right)\right],$$

where $\mathbf{Q}$ is Haar random orthogonal matrix.

We remark that there is a more general definition of the multiplicative convolution to polynomials which are not necessarily characteristic polynomials of matrices (see Definition 1.4 in [MSS22]) though we will not need it here. One can show that $\boxtimes$ is associative, namely, $(a \boxtimes b) \boxtimes c = a \boxtimes (b \boxtimes c)$ and so the $(t-1)$-fold multiplicative convolution of a with itself is well-defined, and is denoted by $a^{\boxtimes t}$ (so that $a^{\boxtimes 1} = a$). Moreover, $\boxtimes$ is commutative.

4.1 The adapter: formalizing the cheat and uncheat idea

The purpose of this section is to formalize the cheat and uncheat idea that was informally discussed in Section 2.1. Recall that we wish to bound the $\mathcal{N}$-transform of the graph at hand but doing so directly seems difficult as it is unclear how to get a handle on the max-inverse of the $\mathcal{M}$-transform of a graph. To bypass this difficulty, we show that the $\mathcal{N}$-transform of the Kesten-McKay distribution is a good approximation to the $\mathcal{N}$-transform of any d-regular Ramanujan graph with sufficiently large girth. We refer to this part of the proof, which is formalized in Corollary 4.10, as *the adapter.* Invoking our adapter with the well-known properties of the Kesten-McKay distribution yields the main result of this section.

Proposition 4.4. *Let G be a d-regular graph on n vertices having girth g whose eigenvalues are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Denote* [9] $\lambda = \max\left(|\lambda_2|, |\lambda_n|, 2\sqrt{d-1}\right)$, *and assume that $3 \leq d \leq \frac{n}{10}$. Then, for every $t \geq 2$ the following holds. For every $y > 0$ for which $\mathcal{N}_{\mathsf{km}^2}(y), \mathcal{N}_{\widehat{G^2}}(y) \in [\beta\lambda^2, \gamma\lambda^2]$, we have that*

$$\mathcal{N}_{(\widehat{G^2})^{\boxtimes t}}(y) \leq \frac{(y+1)^{t+1}}{y(y+d)^t} \cdot d^{2t} + \varepsilon_n t \cdot (\gamma\lambda^2)^t,$$

where

$$\varepsilon_n \triangleq \frac{4\gamma^2\lambda^4}{d} \cdot \min_{h \in [g/2]} \left(\frac{d^{3h}}{n} + \frac{1}{1-\frac{1}{\beta}} \left(\frac{1}{\beta}\right)^{h+1} \right). \tag{4.1}$$

Throughout this section we consider the following setting. Let $a \geq 0$ be a real number, μ a probability measure supported on $[-a, a]$, and set $d = m_2(\mu)$. Let $\mathbf{A}$ be an $n \times n$ real symmetric matrix whose eigenvalues are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Denote $\lambda = \max(|\lambda_2|, |\lambda_n|)$ and set $b = \max(a, \lambda)$. We make the following assumptions: $b \geq 1$, $\lambda_1 = d$, $3 \leq d \leq \frac{n}{10}$ and $\lambda \leq d$. Let $p(x)$ be the polynomial that satisfies $\chi_x(\mathbf{A}^2) = (x - d^2)p(x)$. Assume that the first $2h$ moments of μ and $\mu_{\mathbf{A}}$ match, namely,

$$m_r(\mu) = m_r(\mathbf{A}) \quad \text{for} \quad r = 0, 1, \ldots, 2h. \tag{4.2}$$

Claim 4.5. *With the notation and under the assumptions above, for every $x > b^2$,*

$$|\mathcal{G}_{\mu^2}(x) - \mathcal{G}_p(x)| \leq \frac{2}{nx} \cdot d^{3h} + \frac{2}{x - b^2} \left(\frac{b^2}{x}\right)^{h+1}.$$

Proof. By Equation (3.1) and since $x > b^2$, recalling that the r-th moment of a polynomial $p(x)$, denoted $m_r(p)$, is the corresponding moment of the uniform distribution over its

[9] Note that, technically, λ defined here is not precisely $\lambda(G)$ which recall is defined by $\max\left(|\lambda_2|, |\lambda_n|\right)$. Indeed, it can be the case that $\lambda(G) < 2\sqrt{d-1}$.

roots, we have that

$$
\begin{aligned}
|\mathcal{G}_{\mu^2}(x) - \mathcal{G}_p(x)| &= \left| \sum_{r=0}^{\infty} \frac{m_r(\mu^2) - m_r(p)}{x^{r+1}} \right| \\
&\leq \left| \sum_{r=0}^{h} \frac{m_r(\mu^2) - m_r(p)}{x^{r+1}} \right| + \left| \sum_{r=h+1}^{\infty} \frac{m_r(\mu^2) - m_r(p)}{x^{r+1}} \right| \\
&= \left| \sum_{r=0}^{h} \frac{m_r(\mathbf{A}^2) - m_r(p)}{x^{r+1}} \right| + \left| \sum_{r=h+1}^{\infty} \frac{m_r(\mu^2) - m_r(p)}{x^{r+1}} \right|, \qquad (4.3)
\end{aligned}
$$

where the last equality follows per our assumption as given by Equation (4.2). Using that $\lambda_1 = d$,

$$
m_r(\mathbf{A}^2) - m_r(p) = \frac{1}{n} \sum_{i=1}^{n} \lambda_i^{2r} - \frac{1}{n-1} \sum_{i=2}^{n} \lambda_i^{2r} = \frac{1}{n} d^{2r} - \frac{1}{n(n-1)} \sum_{i=2}^{n} \lambda_i^{2r},
$$

and so, since $\lambda \leq d$,

$$
\left| m_r(\mathbf{A}^2) - m_r(p) \right| \leq \frac{1}{n} d^{2r} + \frac{1}{n(n-1)} \sum_{i=2}^{n} \lambda_i^{2r} \leq \frac{1}{n} d^{2r} + \frac{\lambda^{2r}}{n} \leq \frac{2}{n} d^{2r}.
$$

Thus, we can bound the first summand on the right hand side of Equation (4.3) by

$$
\left| \sum_{r=0}^{h} \frac{m_r(\mathbf{A}^2) - m_r(p)}{x^{r+1}} \right| \leq \frac{1}{x} \sum_{r=0}^{h} \frac{|m_r(\mathbf{A}^2) - m_r(p)|}{x^r} \leq \frac{2}{nx} \sum_{r=0}^{h} \left(\frac{d^2}{x} \right)^r \leq \frac{2}{nx} \cdot d^{3h}, \qquad (4.4)
$$

where for the last inequality we used our assumption $x > b^2 \geq 1$. Moving forward to the higher moments which appear on the second summand on the right hand side of Equation (4.3), we have that

$$
|m_r(\mu^2) - m_r(p)| \leq m_r(\mu^2) + m_r(p) \leq a^{2r} + \lambda^{2r} \leq 2b^{2r},
$$

where we used the fact that μ is supported on $[-a, a]$ (hence $m_r(\mu^2) = m_{2r}(\mu) \leq a^{2r}$) and similarly for μ_p. Thus,

$$
\left| \sum_{r=h+1}^{\infty} \frac{m_r(\mu^2) - m_r(p)}{x^{r+1}} \right| \leq \sum_{r=h+1}^{\infty} \frac{|m_r(\mu^2) - m_r(p)|}{x^{r+1}} \leq \frac{2}{x} \cdot \sum_{r=h+1}^{\infty} \left(\frac{b^2}{x} \right)^r = \frac{2}{x - b^2} \left(\frac{b^2}{x} \right)^{h+1}.
$$

The proof follows by substituting Equation (4.4) and the above bound to Equation (4.3). $\square$

Claim 4.5 tells us that when the two distributions μ^2 and p agree on the low moments, the corresponding $\mathcal{G}$-transforms are close for sufficiently large input x. We wish to prove closeness of the $\mathcal{N}$-transforms which, recall, are the inverses of the $\mathcal{M}$-transforms. To this end, we need the following technical claim.

Claim 4.6. *Let $f, g : (c, \infty) \to \mathbb{R}$ be differentiable, strictly decreasing convex functions. Note that $f^{-1} : \mathrm{Im}(f) \to (c, \infty)$ and $g^{-1} : \mathrm{Im}(g) \to (c, \infty)$ are well-defined functions. Let $\varepsilon : (c, \infty) \to \mathbb{R}$ be a function such that for every $x > c$, $|f(x) - g(x)| \le \varepsilon(x)$. Then, for every $y \in \mathrm{Im}(f) \cap \mathrm{Im}(g)$,*

$$\left|f^{-1}(y) - g^{-1}(y)\right| \le \max\left(\frac{\varepsilon(f^{-1}(y))}{|g'(f^{-1}(y))|}, \frac{\varepsilon(g^{-1}(y))}{|f'(g^{-1}(y))|}\right).$$

Proof. Let $x = g^{-1}(y)$ and $x' = f^{-1}(y)$. Consider the case that $x' \ge x$ and denote $\Delta = x' - x = |f^{-1}(y) - g^{-1}(y)|$. To conclude the proof, it suffices to prove that in this case, namely, $x' \ge x$,

$$\Delta \le \frac{\varepsilon(f^{-1}(y))}{|g'(f^{-1}(y))|}.$$

Indeed, by exchanging the roles of f, g this will imply that in the second case the bound $\Delta \le \frac{\varepsilon(g^{-1}(y))}{|f'(g^{-1}(y))|}$ holds and the proof will follow. Since $f(x + \Delta) = f(x') = y = g(x)$,

$$\varepsilon(x + \Delta) \ge |f(x + \Delta) - g(x + \Delta)| = |g(x) - g(x + \Delta)|.$$

Since g is convex, differentiable and strictly decreasing, we have that

$$\frac{g(x) - g(x + \Delta)}{\Delta} \ge |g'(x + \Delta)|,$$

and so

$$\Delta \le \frac{\varepsilon(x + \Delta)}{|g'(x + \Delta)|} = \frac{\varepsilon(f^{-1}(y))}{|g'(f^{-1}(y))|}.$$

□

In light of the hypothesis of Claim 4.6, the next two claims prepare the grounds by proving that $\mathcal{M}_{\mu^2}$ and $\mathcal{M}_p$ are strictly decreasing convex functions, and give bounds on their derivatives.

Claim 4.7. *In the interval (a^2, ∞), the function $\mathcal{M}_{\mu^2}(x)$ is strictly decreasing and convex. Moreover, $|\mathcal{M}'_{\mu^2}(x)| \ge \frac{d}{x^2}$.*

Proof. As $\frac{t}{x-t}$ is differential (in x) wherever $x > t$, we can write

$$\mathcal{M}'_{\mu^2}(x) = \frac{d}{dx}\int_0^{a^2} \frac{t}{x-t}\mu^2(t)dt = -\int_0^{a^2} \frac{t}{(x-t)^2}\mu^2(t)dt < 0,$$

establishing that $\mathcal{M}_{\mu^2}(x)$ is strictly decreasing in (a^2, ∞). Moreover, by the above equation,

$$\left|\mathcal{M}'_{\mu^2}(x)\right| = \int_0^{a^2} \frac{t}{(x-t)^2}\mu^2(t)dt \ge \frac{1}{x^2}\int_0^{a^2} t\mu^2(t)dt = \frac{m_1(\mu^2)}{x^2} = \frac{m_2(\mu)}{x^2} = \frac{d}{x^2}.$$

As for the convexity of $\mathcal{M}_{\mu^2}(x)$ in (a^2, ∞),

$$\mathcal{M}''_{\mu^2}(x) = \frac{d^2}{dx^2} \int_0^{a^2} \frac{t}{x-t} \mu^2(t) dt = 2 \int_0^{a^2} \frac{t}{(x-t)^3} \mu^2(t) dt > 0.$$

$\square$

Claim 4.8. *In the interval (λ^2, ∞), the function $\mathcal{M}_p(x)$ is strictly decreasing and convex. Moreover, $|\mathcal{M}'_p(x)| \geq \frac{d}{2x^2}$.*

Proof. We have that

$$\mathcal{M}_p(x) = \frac{1}{n-1} \sum_{i=2}^{n} \frac{\lambda_i^2}{x - \lambda_i^2}.$$

Thus, for $x > \lambda^2$,

$$\mathcal{M}'_p(x) = -\frac{1}{n-1} \sum_{i=2}^{n} \frac{\lambda_i^2}{(x - \lambda_i^2)^2} < 0,$$

which shows that $\mathcal{M}'_p(x)$ is strictly decreasing in (λ^2, ∞). As

$$\sum_{i=2}^{n} \lambda_i^2 = \mathrm{Tr}(\mathbf{A}^2) - d^2 = d(n-d),$$

we have that

$$\left|\mathcal{M}'_p(x)\right| = \frac{1}{n-1} \sum_{i=2}^{n} \frac{\lambda_i^2}{(x - \lambda_i^2)^2} \geq \frac{1}{n-1} \sum_{i=2}^{n} \frac{\lambda_i^2}{x^2} = \frac{n-d}{n-1} \cdot \frac{d}{x^2} > \frac{d}{2x^2},$$

as per our assumption, $d \leq \frac{n}{2}$. To conclude the proof, note that

$$\mathcal{M}''_p(x) = \frac{2}{n-1} \sum_{i=2}^{n} \frac{\lambda_i^2}{(x - \lambda_i^2)^3}$$

which is strictly positive for $x > \lambda^2$. $\square$

We invoke the results from this section to obtain the following.

Claim 4.9. *Let $1 < \beta < \gamma$ be parameters. With the notation and under the assumptions listed at the beginning of the section, for every y that satisfies $\mathcal{N}_{\mu^2}(y), \mathcal{N}_p(y) \in [\beta b^2, \gamma b^2]$ it holds that*

$$|\mathcal{N}_{\mu^2}(y) - \mathcal{N}_p(y)| \leq \frac{4\gamma^2 b^4}{d} \cdot \left(\frac{d^{3h}}{n} + \frac{1}{1 - \frac{1}{\beta}} \left(\frac{1}{\beta}\right)^{h+1} \right).$$

Proof. By Claim 4.5, for every $x \geq \beta b^2$,

$$\begin{aligned}
|\mathcal{M}_{\mu^2}(x) - \mathcal{M}_p(x)| &= x\,|\mathcal{G}_{\mu^2}(x) - \mathcal{G}_p(x)| \\
&\leq \frac{2}{n} \cdot d^{3h} + \frac{2x}{x-b^2}\left(\frac{b^2}{x}\right)^{h+1} \\
&\leq \frac{2}{n} \cdot d^{3h} + \frac{2}{1-\frac{1}{\beta}}\left(\frac{1}{\beta}\right)^{h+1} \triangleq \varepsilon.
\end{aligned}$$

By Claim 4.7 and Claim 4.8, the functions $\mathcal{M}_{\mu^2}, \mathcal{M}_p$ are differentiable, strictly decreasing and convex when restricted to the domain $(a^2, \infty) \cap (\lambda^2, \infty) = (b^2, \infty)$. We can thus invoke Claim 4.6 with f, g taken to be $\mathcal{M}_{\mu^2}$ and $\mathcal{M}_p$, respectively, on the domain (c, ∞) where $c = \beta b^2$, and conclude that

$$|\mathcal{N}_{\mu^2}(y) - \mathcal{N}_p(y)| \leq \varepsilon \cdot \max\left(\frac{1}{\mathcal{M}'_p\left(\mathcal{N}_{\mu^2}(y)\right)}, \frac{1}{\mathcal{M}'_{\mu^2}\left(\mathcal{N}_p(y)\right)}\right).$$

Per our assumption, $\mathcal{N}_p(y) \geq \beta b^2 > a^2$, and so we can invoke Claim 4.7 to get

$$\mathcal{M}'_{\mu^2}\left(\mathcal{N}_p(y)\right) \geq \frac{d}{\mathcal{N}_p(y)^2} \geq \frac{d}{\gamma^2 b^4},$$

where for the last inequality, we used our assumption $\mathcal{N}_p(y) \leq \gamma b^2$. Similarly, by Claim 4.8,

$$\mathcal{M}'_p\left(\mathcal{N}_{\mu^2}(y)\right) \geq \frac{d}{2\mathcal{N}_{\mu^2}(y)^2} \geq \frac{d}{2\gamma^2 b^4}.$$

Combining the above we get that

$$|\mathcal{N}_{\mu^2}(y) - \mathcal{N}_p(y)| \leq \varepsilon \cdot \frac{2\gamma^2 b^4}{d},$$

which concludes the proof. ☐

We conclude

Corollary 4.10. *Let G be a d-regular graph on n vertices having girth g whose eigenvalues are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Denote $\lambda = \max\left(|\lambda_2|, |\lambda_n|, 2\sqrt{d-1}\right)$, and assume that $3 \leq d \leq \frac{n}{2}$. Then, for every $y > 0$ for which $\mathcal{N}_{\mathsf{km}^2}(y), \mathcal{N}_{\widehat{G^2}}(y) \in [\beta\lambda^2, \gamma\lambda^2]$, [10] we have that*

$$\left|\mathcal{N}_{\mathsf{km}^2}(y) - \mathcal{N}_{\widehat{G^2}}(y)\right| \leq \varepsilon_n,$$

where ε_n is given by Equation (4.1).

[10] We remind the reader that the notation used, $\mathcal{N}_{\widehat{G^2}}(y)$, is defined in Section 3.2.2.

Proof. The proof readily follows by Claim 4.9 once one verifies the hypothesis of the latter. To this end, as km is supported on $[-2\sqrt{d-1}, 2\sqrt{d-1}]$ (recall Equation (3.2)), the parameter a in the notation used by Claim 4.9 can be taken to be $a = 2\sqrt{d-1}$.

It is a matter of a calculation to show that $m_2(\mathsf{km}) = d$. Recall that b was defined by $b = \max(a, \lambda)$, and indeed $b \geq 1$ as required. As G is d-regular we have that $\lambda_1 = d$. Moreover, $\max(|\lambda_2|, |\lambda_n|) \leq d$ is a standard fact about symmetric matrices. □

We are finally in a position to prove Proposition 4.4.

Proof of Proposition 4.4. Note that for every pair of real numbers a, b such that $|a-b| \leq \varepsilon$ and every $t \in \mathbb{N}$, $|a^t - b^t| \leq \varepsilon t \cdot \max(a,b)^t$. Thus, Corollary 4.10 implies that

$$\left|\mathcal{N}_{\mathsf{km}^2}(y)^t - \mathcal{N}_{\widehat{G^2}}(y)^t\right| \leq \varepsilon_n t \cdot (\gamma\lambda^2)^t.$$

Therefore, by Claim 3.4,

$$\mathcal{N}_{\widehat{G^2}}(y)^t \leq \left(\frac{d^2(y+1)^2}{y(y+d)}\right)^t + \varepsilon_n t \cdot (\gamma\lambda^2)^t.$$

Corollary 3.2 then implies that

$$\mathcal{N}_{(\widehat{G^2})^{\boxtimes t}}(y) \leq \left(\frac{y}{y+1}\right)^{t-1} \left(\mathcal{N}_{\widehat{G^2}}(y)\right)^t \leq \frac{(y+1)^{t+1}}{y(y+d)^t} \cdot d^{2t} + \varepsilon_n t \cdot (\gamma\lambda^2)^t.$$

□

4.2 Haar rotated Ramanujan graphs

In this section we prove Proposition 4.1 below though only for $t \geq 4$. Interestingly, to handle the smaller values $t = 2, 3$ some more technical work is required. As this additional effort is quite technical and not too illuminating, the proof for these cases is deferred to Section 4.2.1.

Proof of Proposition 4.1 for $t \geq 4$. We start by briefly describing the proof strategy.

Proof strategy: using the $\mathcal{N}$-transform and the adapter. First note that given a polynomial $p(x)$ with non-negative roots, for any $y > 0$, $\mathcal{N}_p(y)$ is an upper bound on $\mathsf{maxroot}(p(x))$. That is,

$$\mathsf{maxroot}(p(x)) \leq \inf_{y>0} \mathcal{N}_p(y).$$

Indeed, if we denote $x = \mathcal{N}_p(y)$ then, by the definition of $\mathcal{N}_p$, x is the largest real number for which $\mathcal{M}_p(x) = y$. Now, denote $\alpha = \mathsf{maxroot}(p(x))$ and note that the function $\mathcal{M}_p$

restricted to the domain (α, ∞) has precisely $(0, \infty)$ as its image. By the intermediate value theorem, there is some $x' > \alpha$ for which $\mathcal{M}_p(x') = y$. Therefore $x \geq x'$, and so

$$\mathcal{N}_p(y) = x \geq x' > \alpha = \mathsf{maxroot}(p(x)).$$

From this we see that to prove the corollary it suffices to bound $\mathcal{N}_{(\widehat{G^2})^{\boxtimes t}}(y)$ where we have the freedom to choose $y > 0$ as desired. The straightforward way of doing that is to find a y that minimizes $\mathcal{N}_{(\widehat{G^2})^{\boxtimes t}}(y)$. As we do not have a good handle on the later polynomial (namely, we do not have concrete, easy to work with, information on the spectrum of G), we use Proposition 4.4 which essentially studies the latter by considering the Kesten-McKay distribution instead. However, the technical delicate point in using Proposition 4.4 is that we are restricted to work with y-s such that

$$\mathcal{N}_{\mathsf{km}^2}(y), \mathcal{N}_{\widehat{G^2}}(y) \in \left[\beta\lambda^2, \gamma\lambda^2\right], \tag{4.5}$$

where $\gamma > \beta > 1$ affect the quantitative bound we get. Hence, we will proceed by finding the value y_{min} that minimizes $\mathcal{N}_{\mathsf{km}^2}^{\boxtimes t}$ in the relevant domain.

Applying the adapter. Denote

$$f(y) = \frac{(y+1)^{t+1}}{y(y+d)^t}.$$

It is easy to verify that

$$f'(y) = \frac{(y+1)^t}{y^2(y+d)^{t+1}} \left((td - t - 1)y - d\right),$$

and so

$$y_{\mathsf{min}} = \frac{d}{dt - t - 1} \tag{4.6}$$

is an extreme, in fact a global minimum, of $f(y)$ in $(0, \infty)$. Substituting, we get

$$f(y_{\mathsf{min}}) = (t+1)\left(1 + \frac{1}{t}\right)^t \cdot \frac{d-1}{d^{t+1}} < \frac{e(t+1)}{d^t}. \tag{4.7}$$

Thus, assuming the hypothesis of Proposition 4.4 holds, as we will verify shortly with suitably chosen β, γ, the latter implies that

$$\mathcal{N}_{(\widehat{G^2})^{\boxtimes t}}(y_{\mathsf{min}}) \leq \left(1 + \frac{1}{t}\right)^t (t+1)d^t + \varepsilon_n t \cdot (\gamma \cdot 4(d-1))^t, \tag{4.8}$$

where recall that ε_n (see Equation (4.1)), set with $\lambda = 2\sqrt{d-1}$, equals

$$\varepsilon_n \triangleq \frac{64\gamma^2(d-1)^2}{d} \cdot \min_{h \in [g/2]} \left(\frac{d^{3h}}{n} + \frac{1}{1 - \frac{1}{\beta}} \left(\frac{1}{\beta}\right)^{h+1}\right). \tag{4.9}$$

Verifying the adapter's hypothesis for our choice of β, γ. Before proceeding, we verify that the hypothesis of Proposition 4.4 holds. More accurately, we show that for $\gamma = t + 1$ and $\beta = \frac{9}{8}$ it holds that

$$\mathcal{N}_{\mathsf{km}^2}(y_{\min}), \mathcal{N}_{\widehat{G^2}}(y_{\min}) \in \left[\beta\lambda^2, \gamma\lambda^2\right] = \left[4\beta(d-1), 4\gamma(d-1)\right].$$

Starting with km^2, by Claim 3.4,

$$\mathcal{N}_{\mathsf{km}^2}(y_{\min}) = \frac{d^2(y_{\min}+1)^2}{y_{\min}(y_{\min}+d)} = \frac{(t+1)^2}{t}(d-1), \tag{4.10}$$

and so by our choice of β, γ, we have that $\mathcal{N}_{\mathsf{km}^2}(y_{\min})$ is in the required interval. This in fact holds for all $t \geq 2$.

As for $\mathcal{N}_{\widehat{G^2}}(y)$, we have that $\mathcal{M}_{\widehat{G^2}}(x) = \frac{1}{n-1}\sum_{i=2}^{n} \frac{\lambda_i^2}{x-\lambda_i^2}$, and so for every $x > \lambda^2$, $\mathcal{M}_{\widehat{G^2}}(x) \leq \frac{\lambda^2}{x-\lambda^2}$. This then implies that for every $y > 0$,

$$\mathcal{N}_{\widehat{G^2}}(y) \leq \lambda^2\left(1 + \frac{1}{y}\right). \tag{4.11}$$

Thus,

$$\mathcal{N}_{\widehat{G^2}}(y_{\min}) \leq \lambda^2\left(1 + \frac{1}{y_{\min}}\right) = \frac{4(d-1)^2}{d}(t+1), \tag{4.12}$$

and so, per our choice $\gamma = t+1$, we have that $\mathcal{N}_{\widehat{G^2}}(y_{\min}) \leq 4\gamma(d-1)$ holds for all $t \geq 2$.

As for the lower bound, for any fixed x, we invoke Jensen's inequality to the function $g_x(z) = \frac{1}{x-z}$ with Z being the random variable in which we sample $i \sim \{2, \dots, n\}$ uniformly at random and return λ_i^2. Alongside the fact that

$$\mathbb{E}[Z] = \frac{1}{n-1}\sum_{i=2}^{n} \lambda_i^2 = \frac{nd-d^2}{n-1} = \frac{n-d}{n-1} \cdot d \geq \frac{9d}{10}, \tag{4.13}$$

where the last inequality follows per our assumption $d \leq \frac{n}{10}$, we conclude that for every $x > \lambda^2$,

$$\mathcal{M}_{\widehat{G^2}}(x) = x \cdot \mathbb{E}\left[g_x(Z)\right] - 1 \geq x \cdot g_x(\mathbb{E}\left[Z\right]) - 1 = \frac{9d}{10x - 9d}.$$

Hence, for every $y > 0$,

$$\mathcal{N}_{\widehat{G^2}}(y) \geq \frac{9d}{10} \cdot \left(1 + \frac{1}{y}\right), \tag{4.14}$$

and so

$$\mathcal{N}_{\widehat{G^2}}(y_{\min}) \geq \frac{9}{10}(t+1)(d-1). \tag{4.15}$$

Therefore, for $\mathcal{N}_{\widehat{G^2}}(y_{\min}) \geq 4\beta(d-1) = \frac{9}{2}(d-1)$ to hold, it suffices that $t \geq 4$.

Concluding the proof. Now that we have verified that the hypothesis of the adapter holds with the setting of parameters $\beta = \frac{9}{8}$, $\gamma = t+1$, we make some final calculations and conclude the proof. Substituting the chosen β, γ to Equation (4.9), we get

$$\varepsilon_n \le 128t^2 d \cdot \min_{h \in [g/2]} \left(\frac{d^{3h}}{n} + 9 \left(\frac{8}{9} \right)^{h+1} \right) = O(t^2 d) \cdot 2^{-\Omega(\bar{g})}.$$

Substituting this to Equation (4.8),

$$\begin{aligned} \mathcal{N}_{(\widehat{G^2})^{\boxtimes t}}(y_{\mathsf{min}}) &\le \left(1 + \frac{1}{t}\right)^t (t+1)d^t + \varepsilon_n t \cdot (\gamma \cdot 4(d-1))^t \\ &\le e(t+1)d^t + 2^{-\Omega(\bar{g})} \cdot (8td)^{t+3}, \end{aligned}$$

which concludes the proof. □

4.2.1 Proof of Proposition 4.1 for $t = 2, 3$

By inspection, the only part of the proof of Proposition 4.1 in which we make use of the assumption $t \ge 4$ is in establishing a lower bound for $\mathcal{N}_{\widehat{G^2}}(y_{\mathsf{min}})$ (see Equation (4.15)). Recall that this is achieved by using Jensen's inequality. In this section we prove that with a certain strenghening of Jensen's inequality, given by Lemma 4.11 below, the assumption $t \ge 4$ can be removed.

Lemma 4.11. *Let Z be a random variable, and $f(z)$ be a 4-times differentiable function. Denote $\mu = \mathbb{E}[Z]$, and $\Delta_i = \mathbb{E}\,(Z-\mu)^i$ (e.g., Δ_2 is the variance of Z). Define the function*

$$K(z) = \frac{f(z) - f(\mu)}{(z-\mu)^4} - \frac{f'(\mu)}{(z-\mu)^3} - \frac{f''(\mu)}{2(z-\mu)^2} - \frac{f'''(\mu)}{6(z-\mu)}.$$

Then,

$$\mathbb{E}[f(Z)] - f(\mu) \ge \frac{f''(\mu)}{2} \cdot \Delta_2 + \frac{f'''(\mu)}{6} \cdot \Delta_3 + \inf_z K(z) \cdot \Delta_4.$$

The proof of Lemma 4.11 can be found in Appendix A. With this we turn to complete the proof of Proposition 4.1 for $t = 2, 3$ by establishing the following claim.

Claim 4.12. *For $t \ge 2$ there exist $\gamma > \beta > 1$ and $x \in [\beta\lambda^2, \gamma\lambda^2]$ such that $\mathcal{M}_{\widehat{G^2}}(x) = y_{\mathsf{min}}$, where recall $y_{\mathsf{min}} = \frac{d}{dt-t-1}$.*

Proof. Let $g_x(z) = \frac{1}{x-z}$ and Z be as defined in the proof of Proposition 4.1. We proceed as in Equation (4.13) though without bounding it from below to get

$$\mu \triangleq \mathbb{E}[Z] = \frac{1}{n-1} \sum_{i=2}^{n} \lambda_i^2 = \frac{nd - d^2}{n-1} = \frac{n-d}{n-1} \cdot d = (1-\delta)d,$$

where $\delta = \frac{d-1}{n-1}$.

Using the notation of Lemma 4.11, by calculating the derivatives of $g_x(z)$, we get

$$\mathbb{E}[g_x(Z)] - g_x(\mu) \geq \frac{\Delta_2}{(x-\mu)^3} + \frac{\Delta_3}{(x-\mu)^4} + \inf_z K_x(z) \cdot \Delta_4, \tag{4.16}$$

where

$$K_x(z) = \frac{1}{(x-\mu)^4(x-z)},$$

and so, since we are in the regime $x > z$, we get $\inf_z K_x(z) = \frac{1}{(x-\mu)^4 x}$.

For calculating the Δ_i-s, we compute the first few moments of Z, which we denote by $(\widehat{m}_i)_i$. To compute $\widehat{m}_i$ we consider first the moments, $m_i = \frac{1}{n} \cdot \mathrm{Tr}(\mathbf{A}^i)$, of the adjacency matrix of G^2. Note that it suffices to prove the theorem for a graph of girth at least 9 due to the hidden constant under the Ω in the bound we wish to prove. For such a graph, we know from [McK81] (page 3) that

$$\begin{aligned}
m_1 &= d, \\
m_2 &= d^2 + d(d-1), \\
m_3 &= d^3 + 2d^2(d-1) + 2d(d-1)^2, \\
m_4 &= d^4 + 3d^3(d-1) + 5d^2(d-1)^2 + 5d(d-1)^3.
\end{aligned}$$

Since

$$\widehat{m}_i = \frac{1}{n-1}\left(\mathrm{Tr}(\mathbf{A}^i) - d^{2i}\right) = m_i + \varepsilon_i,$$

where $\varepsilon_i = \frac{m_i - d^{2i}}{n-1}$, we have that

$$\begin{aligned}
\Delta_2 &= \widehat{m}_2 - 2\mu\widehat{m}_1 + \mu^2 = d(d-1) + \delta_2, \\
\Delta_3 &= \widehat{m}_3 - 3\mu\widehat{m}_2 + 3\mu^2\widehat{m}_1 - \mu^3 = d^3 - 3d^2 + 2d + \delta_3, \\
\Delta_4 &= \widehat{m}_4 - 4\mu\widehat{m}_3 + 6\mu^2\widehat{m}_2 - 4\mu^3\widehat{m}_1 + \mu^4 = 3d^4 - 10d^3 + 12d^2 - 5d + \delta_4,
\end{aligned}$$

where the δ_i-s are bounded, in absolute value, by $\frac{1}{n} \cdot \mathrm{poly}(d)$. Plugging the above to Equation (4.16), we get that

$$\mathbb{E}[g_x(Z)] \geq \frac{1}{x-d} + \frac{d(d-1)(3d^2 - 7d + x^2 - 2x + 5)}{(x-d)^4 x} + \delta,$$

for some δ which is bounded by $\frac{1}{n} \cdot \mathrm{poly}(d)$ in absolute value for all $x \geq 2d$. Therefore,

$$\begin{aligned}
\mathcal{M}_{\widehat{G^2}}(x) &= x\,\mathbb{E}[g_x(Z)] - 1 \\
&\geq \frac{d\left(2d^3 + 3d^2x - 10d^2 - 2dx^2 - 2dx + 12d + x^3 - x^2 + 2x - 5\right)}{(x-d)^4} + \delta'.
\end{aligned}$$

for some δ' which too is bounded by $\frac{1}{n} \cdot \mathrm{poly}(d)$ in absolute value for all $x \geq 2d$. One can verify that for all $d \geq 3$, setting $\beta = 1.04$ and $x_0 = \beta\lambda^2 = 4\beta(d-1)$ we get that $\mathcal{M}_{\widehat{G^2}}(x_0) \geq y_{\mathsf{min}}$. The fact that $\mathcal{M}_{\widehat{G^2}}(x)$ is decreasing concludes the proof. $\square$

4.3 Haar rotated expander graphs

In this section we prove Proposition 4.2 which recall is the Haar-random analog of Theorem 1.2. The proof follows the same line of reasoning as the proof of Proposition 4.1 from the previous section, though it diverges at a certain technical point regarding the choice of y_{min}. The reason for the change is that given a graph which is not Ramanujan, we do not have the guarantee that the choice of y_{min} will adhere with our restriction $\mathcal{N}_{\widehat{G^2}}(y_{\mathsf{min}}) \in [\beta\lambda^2, \gamma\lambda^2]$.

Proof of Proposition 4.2. We consider two cases, according to the value of t.

The case $t \geq \frac{8\lambda^2}{d}$. For this case we will use the same choice of y_{min} as in the proof of Proposition 4.1 (see Equation (4.6)). Set $\beta = 2$, $\gamma = t+1$, and recall (Equation (4.10)) that $\mathcal{N}_{\mathsf{km}^2}(y_{\mathsf{min}}) = \frac{(t+1)^2}{t}(d-1)$. We wish to invoke Proposition 4.4 and so we must verify that with our choices above, $\beta\lambda^2 \leq \mathcal{N}_{\mathsf{km}^2}(y_{\mathsf{min}}) \leq \gamma\lambda^2$. The lower bound follows since $\mathcal{N}_{\mathsf{km}^2}(y_{\mathsf{min}}) \geq \frac{td}{2} \geq 2\lambda^2$. As for the upper bound, since $\lambda \geq 2\sqrt{d-1}$, $\mathcal{N}_{\mathsf{km}^2}(y_{\mathsf{min}}) \leq 4t(d-1) \leq \gamma\lambda^2$. Moving on to the transform of the graph at hand, in the proof of Proposition 4.1 (see Equation (4.12)), we showed that

$$\mathcal{N}_{\widehat{G^2}}(y_{\mathsf{min}}) \leq \lambda^2\left(1+\frac{1}{y_{\mathsf{min}}}\right) = \lambda^2\frac{(t+1)(d-1)}{d} < \lambda^2(t+1) = \gamma\lambda^2.$$

As for the lower bound, by Equation (4.15),

$$\mathcal{N}_{\widehat{G^2}}(y_{\mathsf{min}}) \geq \frac{9}{10}(t+1)(d-1) > \frac{72\lambda^2}{10}\cdot\frac{d-1}{d} > 2\lambda^2 = \beta\lambda^2.$$

Therefore, we may invoke Proposition 4.4, also using Equation (4.7), to conclude that

$$\begin{aligned}\mathcal{N}_{(\widehat{G^2})^{\boxtimes t}}(y_{\mathsf{min}}) &\leq \frac{(y_{\mathsf{min}}+1)^{t+1}}{y_{\mathsf{min}}(y_{\mathsf{min}}+d)^t}\cdot d^{2t} + \varepsilon_n t\cdot(\gamma\lambda^2)^t \\ &= \left(1+\frac{1}{t}\right)^t(t+1)d^t + (t\lambda^2)^{t+4}\cdot 2^{-\Omega(\bar{g})}.\end{aligned}$$

The case $t < \frac{8\lambda^2}{d}$. We now take $y_0 = \frac{d}{4\lambda^2}$ rather than y_{min} as defined in Equation (4.6), and set $\beta = 2$, $\gamma = 5d$. Again, we wish to invoke Proposition 4.4 and so we must verify that with our choices above, $\beta\lambda^2 \leq \mathcal{N}_{\mathsf{km}^2}(y_0) \leq \gamma\lambda^2$. Indeed,

$$\mathcal{N}_{\mathsf{km}^2}(y_0) = \frac{d^2\left(\frac{d}{4\lambda^2}+1\right)^2}{\frac{d}{4\lambda^2}\left(\frac{d}{4\lambda^2}+d\right)} = \frac{(d+4\lambda^2)^2}{1+4\lambda^2} \geq \frac{(d+4\lambda^2)^2}{8\lambda^2} \geq 2\lambda^2 = \beta\lambda^2.$$

As for the upper bound,

$$\mathcal{N}_{\mathsf{km}^2}(y_0) = \frac{(d+4\lambda^2)^2}{1+4\lambda^2} \leq \frac{(5\lambda^2)^2}{4\lambda^2} \leq 7\lambda^2 \leq \gamma\lambda^2.$$

Moving on to the transform of the graph at hand, by Equation (4.14), we have that

$$\mathcal{N}_{\widehat{G^2}}(y_0) \geq \frac{9d}{10} \cdot \left(1 + \frac{1}{y_0}\right) > 2\lambda^2 = \beta\lambda^2.$$

Further, by Equation (4.11),

$$\mathcal{N}_{\widehat{G^2}}(y_0) \leq \lambda^2 \left(1 + \frac{1}{y_0}\right) = \lambda^2 \left(1 + \frac{4\lambda^2}{d}\right) \leq \frac{5\lambda^4}{d} \leq \gamma\lambda^2.$$

Therefore, we may invoke Proposition 4.4 to conclude that

$$\mathcal{N}_{(\widehat{G^2})^{\boxtimes t}}(y_0) \leq \frac{(y_0+1)^{t+1}}{y_0(y_0+d)^t} \cdot d^{2t} + 4t(5d\lambda^2)^{t+2} 2^{-\Omega(\bar{g})}.$$

Now,

$$\frac{(y_0+1)^{t+1}}{y_0(y_0+d)^t} = \left(1 + \frac{1}{y_0}\right) \left(\frac{y_0+1}{y_0+d}\right)^t \leq \frac{5\lambda^2}{d} \left(\frac{y_0+1}{d}\right)^t \leq \frac{5\lambda^2}{d^{t+1}} \cdot e^{y_0 t} \leq \frac{5\lambda^2}{d^{t+1}} \cdot e^2,$$

where for the last inequality we used that $t < \frac{8\lambda^2}{d}$. $\square$

5 Quadrature: From Haar to Random Permutations

Our main goal in this section is to reduce the analysis with respect to a sequence of random permutations to a sequence of Haar random orthogonal matrices which, recall, has been analyzed in Section 4. Equivalently, we wish to prove the relation between permutation of matrices, and the multiplicative convolution of their respective characteristic polynomials. The main result proved in this section is Theorem 5.2 which appears right after the following definition.

Definition 5.1. *Let $t \geq 1$ be an integer. Given an $n \times n$ real symmetric matrix $\mathbf{A}$ and $n \times n$ permutation matrices $\mathbf{P}_1, \dots, \mathbf{P}_{t-1}$, we define the $n \times n$ matrices $\Psi_{\mathbf{A}}(\mathbf{P}_1, \dots, \mathbf{P}_j)$ for $j = 0, 1, 2, \dots, t$ recursively as follows: $\Psi_{\mathbf{A}}(\perp) = \mathbf{A}^2$, where $\perp$ denotes the empty sequence. For $j \geq 1$,*

$$\Psi_{\mathbf{A}}(\mathbf{P}_1, \dots, \mathbf{P}_j) = \mathbf{A}\mathbf{P}_j \Psi_{\mathbf{A}}(\mathbf{P}_1, \dots, \mathbf{P}_{j-1}) \mathbf{P}_j^\top \mathbf{A}.$$

Theorem 5.2. *Let $\mathbf{A}$ be an $n \times n$ real symmetric matrix with $\mathbf{A}\mathbf{1} = a\mathbf{1}$, and $p_{\mathbf{A}^2}(x)$ be the polynomial satisfying $\chi_x(\mathbf{A}^2) = (x - a^2) p_{\mathbf{A}^2}(x)$. Let $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_{t-1}$ be independent random permutation matrices. Then,*

$$\mathop{\mathbb{E}}_{\mathbf{P}_1,\dots,\mathbf{P}_{t-1}} \chi_x\left(\Psi_{\mathbf{A}}(\mathbf{P}_1, \dots, \mathbf{P}_{t-1})\right) = \left(x - a^{2t}\right) \left(p_{\mathbf{A}^2}(x)\right)^{\boxtimes t}.$$

The idea and most of the claims in our proof of Theorem 5.2 come from [MSS18]. In particular, Lemma 5.3 below (which is essentially Lemma 2.3, reproduced here for ease of reading) is the key ingredient of the proof which is a multiplicative analog of Theorem 4.1 of [MSS18], who considered the additive case. Our proof mimics the proof for the additive case though it requires a bit more technical work.

Lemma 5.3. *Let $\mathbf{A}, \mathbf{B}$ be real $n \times n$ symmetric matrices such that $\mathbf{A1} = a\mathbf{1}$ and $\mathbf{B1} = b\mathbf{1}$. Denote by $p_{\mathbf{A}}, p_{\mathbf{B}}$ the polynomials satisfying $\chi_x(\mathbf{A}) = (x-a)p_{\mathbf{A}}(x)$, $\chi_x(\mathbf{B}) = (x-b)p_{\mathbf{B}}(x)$. Let $\mathbf{P}$ be a uniformly random $n \times n$ permutation matrix. Then,*

$$\mathop{\mathbb{E}}_{\mathbf{P}} \chi_x(\mathbf{APBP}^\intercal) = (x - ab)\,(p_{\mathbf{A}} \boxtimes p_{\mathbf{B}})\,(x).$$

Proof. The proof of Lemma 5.3 closely follows [MSS18]. We start by setting the ground for "working" orthogonal to $\mathbf{1}$. By basic linear algebra, there exists an orthonormal change of basis matrix $\mathbf{V}$ such that $\mathbf{VAV}^\intercal = \widehat{\mathbf{A}} \oplus a$, where $\widehat{\mathbf{A}} \oplus a$ denotes the direct sum

$$\widehat{\mathbf{A}} \oplus a = \begin{pmatrix} \widehat{\mathbf{A}} & 0 \\ 0 & a \end{pmatrix}.$$

Clearly then, $\chi_x(\mathbf{A}) = (x - a)\chi_x(\widehat{\mathbf{A}})$. As $\mathbf{1}$ is also an eigenvector of $\mathbf{B}$ we have that $\mathbf{VBV}^\intercal = \widehat{\mathbf{B}} \oplus b$ for some matrix $\widehat{\mathbf{B}}$. Since a characteristic polynomial is invariant to a change of basis, we have that

$$\mathop{\mathbb{E}}_{\mathbf{P}} \chi_x(\mathbf{APBP}^\intercal) = \mathop{\mathbb{E}}_{\mathbf{P}} \chi_x\left((\mathbf{VAV}^\intercal)(\mathbf{VPV}^\intercal)(\mathbf{VBV}^\intercal)(\mathbf{VP}^\intercal\mathbf{V}^\intercal)\right).$$

Since for every permutation matrix $\mathbf{P}$ it holds that $\mathbf{P1} = \mathbf{1}$, we have that $\mathbf{VPV}^\intercal = \widehat{\mathbf{P}} \oplus 1$ for some matrix $\widehat{\mathbf{P}}$, and so

$$\begin{aligned} \mathop{\mathbb{E}}_{\mathbf{P}} \chi_x(\mathbf{APBP}^\intercal) &= \mathop{\mathbb{E}}_{\mathbf{P}} \chi_x\left((\widehat{\mathbf{A}} \oplus a)(\widehat{\mathbf{P}} \oplus 1)(\widehat{\mathbf{B}} \oplus b)(\widehat{\mathbf{P}}^\intercal \oplus 1)\right) \\ &= (x - ab) \mathop{\mathbb{E}}_{\mathbf{P}} \chi_x(\widehat{\mathbf{A}}\widehat{\mathbf{P}}\widehat{\mathbf{B}}\widehat{\mathbf{P}}^\intercal), \end{aligned}$$

where note that we used the fact that $\widehat{\mathbf{P}}^\intercal = \widehat{\mathbf{P}^\intercal}$. We observe that when $\mathbf{P}$ is sampled uniformly at random from the set of permutation matrices, $\widehat{\mathbf{P}}$ is a random element in the symmetry group of the n-vertex regular simplex, embedded in $\mathbb{R}^{n-1}$. With this in mind, according to Definition 4.3, it suffices to prove that

$$\mathop{\mathbb{E}}_{\mathbf{P}} \chi_x(\widehat{\mathbf{A}}\widehat{\mathbf{P}}\widehat{\mathbf{B}}\widehat{\mathbf{P}}^\intercal) = \mathop{\mathbb{E}}_{\mathbf{Q}} \chi_x\left(\widehat{\mathbf{A}}\mathbf{Q}\widehat{\mathbf{B}}\mathbf{Q}^\intercal\right), \tag{5.1}$$

where $\mathbf{Q}$ is drawn according to the Haar measure.

Note that both $\widehat{\mathbf{P}}$ and $\mathbf{Q}$ in Equation (5.1) are of dimension $n-1$. However, in the remainder of the proof it will be more convenient to work with these matrices, forgetting that they are induced by $n \times n$ matrices, and so we redefine n as the size of the matrices that follow. To prove Equation (5.1), we make use of the following simple lemma by [MSS18]. We remind the reader that we denote the group of $n \times n$ orthogonal matrices by $\mathcal{O}(n)$.

Lemma 5.4 ([MSS18] Lemma 4.3)**.** *Let $f : \mathcal{O}(n) \to \mathbb{R}$ and let H be a subgroup of $\mathcal{O}(n)$. Assume that for all $\mathbf{Q} \in \mathcal{O}(n)$,*

$$\mathop{\mathbb{E}}_{\mathbf{P}\sim H} f(\mathbf{P}) = \mathop{\mathbb{E}}_{\mathbf{P}\sim H} f(\mathbf{PQ}). \tag{5.2}$$

Then, $\mathbb{E}_{\mathbf{P}\sim H} f(\mathbf{P}) = \mathbb{E}_{\mathbf{Q}} f(\mathbf{Q})$, where $\mathbf{Q}$ is Haar random.

We will establish Equation (5.1) by showing that the condition that is given by Equation (5.2) holds for $H = \mathcal{A}_{n-1}$, the symmetry group of the n-vertex regular simplex, with respect to the function $f_{\mathbf{A},\mathbf{B}}(\mathbf{Q}) = \chi_x\left(\mathbf{AQBQ}^{\intercal}\right)$ for every choice of $\mathbf{A}, \mathbf{B}$. We mimic the proof strategy of [MSS18], namely, we prove the above for the group $\mathcal{A}_2$ of symmetries on the 3-vertex simplex, fixing the remaining $n-2$ dimensions, and then show that this suffices as all orthogonal Haar matrices can be constructed as a product of such two-dimensional transformations. For ease of notation, throughout the proof, for a 2×2 matrix $\mathbf{M}$ we let $\widetilde{\mathbf{M}} = \mathbf{M} \oplus \mathbf{I}_{n-2}$.

Lemma 5.5. *Let $\mathbf{A}, \mathbf{B}$ be a pair of $n \times n$ real symmetric matrices. Then, for every $\mathbf{Q} \in \mathcal{O}(2)$,*

$$\mathop{\mathbb{E}}_{\mathbf{P}\sim\mathcal{A}_2} f_{\mathbf{A},\mathbf{B}}(\widetilde{\mathbf{P}}) = \mathop{\mathbb{E}}_{\mathbf{P}\in\mathcal{A}_2} f_{\mathbf{A},\mathbf{B}}(\widetilde{\mathbf{PQ}}). \tag{5.3}$$

Proof. Let $\mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ be the reflection around the horizontal axis and $F = \{\mathbf{I}, \mathbf{H}\}$ be the group generated by $\mathbf{H}$. For an angle θ define $\mathbf{R}_\theta$ to be the rotation of the plane by angle θ, that is,

$$R_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

It is a basic geometric fact that every $\mathbf{Q} \in \mathcal{O}(2)$ can be written as $\mathbf{R}_\theta\mathbf{D}$ for $\mathbf{D} \in F$. Moreover, the six elements of $\mathcal{A}_2$ can be thought of as the three permutations $\mathbf{R}_\tau$ where $\tau \in T \triangleq \{0, \frac{2\pi}{3}, \frac{4\pi}{3}\}$, multiplied by a matrix $\mathbf{D} \in F$.

We start by proving that for every θ,

$$\mathop{\mathbb{E}}_{\tau\sim T}\left[\chi_x\left(\mathbf{A}\widetilde{\mathbf{R}_\tau}\mathbf{B}\widetilde{\mathbf{R}_\tau}^{\intercal}\right)\right] = \mathop{\mathbb{E}}_{\tau\sim T}\left[\chi_x\left(\mathbf{A}(\widetilde{\mathbf{R}_\tau\mathbf{R}_\theta})\mathbf{B}(\widetilde{\mathbf{R}_\tau\mathbf{R}_\theta})^{\intercal}\right)\right]. \tag{5.4}$$

To see this, note that the right hand side of this equation can be written as

$$\frac{1}{3}\sum_{\tau\in T}\chi_x\left(\mathbf{A}\widetilde{\mathbf{R}_{\tau+\theta}}\mathbf{B}\widetilde{\mathbf{R}_{\tau+\theta}}^\mathsf{T}\right),$$

and according to Lemma 5.7, whose proof is deferred to Section 5.1, this equals to

$$\begin{aligned}\frac{1}{3}\sum_{\tau\in T}\sum_{k=-2}^{2}c_k e^{ik(\theta+\tau)} &= \frac{1}{3}\sum_{k=-2}^{2}\sum_{\tau\in T}c_k e^{ik(\theta+\tau)}\\ &= \frac{1}{3}\sum_{k=-2}^{2}c_k e^{ik\theta}\sum_{\tau\in T}e^{ik\tau}\\ &= \frac{1}{3}\sum_{k=-2}^{2}c_k e^{ik\theta}(e^0+e^{ik\frac{2\pi}{3}}+e^{ik\frac{4\pi}{3}})\\ &= c_0.\end{aligned}$$

Therefore, this expression is independent of θ, and in particular it coincides with the expression obtained for $\theta = 0$, proving Equation (5.4).

We proceed by showing that

$$\mathop{\mathbb{E}}_{\tau\sim T}\left[\chi_x\left(\mathbf{A}(\widetilde{\mathbf{R}_\tau\mathbf{H}})\mathbf{B}(\widetilde{\mathbf{R}_\tau\mathbf{H}})^\mathsf{T}\right)\right] = \mathop{\mathbb{E}}_{\tau\sim T}\left[\chi_x\left(\mathbf{A}(\widetilde{\mathbf{R}_\tau\mathbf{H}\mathbf{R}_\theta})\mathbf{B}(\widetilde{\mathbf{R}_\tau\mathbf{H}\mathbf{R}_\theta})^\mathsf{T}\right)\right]. \tag{5.5}$$

To see this, note that for every θ, $\mathbf{H}\mathbf{R}_\theta = \mathbf{R}_{-\theta}\mathbf{H}$, and so the matrix on the right hand-side

$$\mathbf{A}(\widetilde{\mathbf{R}_\tau\mathbf{H}\mathbf{R}_\theta})\mathbf{B}(\widetilde{\mathbf{R}_\tau\mathbf{H}\mathbf{R}_\theta})^\mathsf{T} = \mathbf{A}\widetilde{\mathbf{R}_{\tau-\theta}}\left(\widetilde{\mathbf{H}}\mathbf{B}\widetilde{\mathbf{H}}^\mathsf{T}\right)\widetilde{\mathbf{R}_{\tau-\theta}}^\mathsf{T},$$

whereas the matrix that appear on the left hand-side

$$\mathbf{A}(\widetilde{\mathbf{R}_\tau\mathbf{H}})\mathbf{B}(\widetilde{\mathbf{R}_\tau\mathbf{H}})^\mathsf{T} = \mathbf{A}\widetilde{\mathbf{R}_\tau}\left(\widetilde{\mathbf{H}}\mathbf{B}\widetilde{\mathbf{H}}^\mathsf{T}\right)\widetilde{\mathbf{R}_\tau}^\mathsf{T}.$$

Thus, Equation (5.5) follows from Equation (5.4) by taking $\mathbf{B}$ in the notation of the latter equation to be $\widetilde{\mathbf{H}}\mathbf{B}\widetilde{\mathbf{H}}^\mathsf{T}$.

To recap, Equation (5.4) and Equation (5.5) together show that

$$\mathop{\mathbb{E}}_{\mathbf{P}\sim\mathcal{A}_2}\left[\chi_x\left(\mathbf{A}\widetilde{\mathbf{P}}\mathbf{B}\widetilde{\mathbf{P}}^\mathsf{T}\right)\right] = \mathop{\mathbb{E}}_{\mathbf{P}\in\mathcal{A}_2}\left[\chi_x\left(\mathbf{A}\widetilde{\mathbf{P}\mathbf{Q}}\mathbf{B}\widetilde{\mathbf{P}\mathbf{Q}}^\mathsf{T}\right)\right] \tag{5.6}$$

for every $\mathbf{Q}$ of the form $\mathbf{Q} = \mathbf{R}_\theta$. Equivalently, Equation (5.3) holds for every such $\mathbf{Q}$. To conclude the proof, we need to prove the same statement for $\mathbf{Q}$ of the form $\mathbf{Q} = \mathbf{R}_\theta\mathbf{H}$. To this end, note that

$$\mathbf{A}(\widetilde{\mathbf{P}\mathbf{R}_\theta\mathbf{H}})\mathbf{B}(\widetilde{\mathbf{P}\mathbf{R}_\theta\mathbf{H}})^\mathsf{T} = \mathbf{A}(\widetilde{\mathbf{P}\mathbf{H}\mathbf{R}_{-\theta}})\mathbf{B}(\widetilde{\mathbf{P}\mathbf{H}\mathbf{R}_{-\theta}})^\mathsf{T}.$$

Therefore, as $\mathcal{A}_2\mathbf{H}$ has the same distribution as $\mathcal{A}_2$, we have that

$$\begin{aligned}\mathop{\mathbb{E}}_{\mathbf{P}\sim\mathcal{A}_2}\left[\chi_x\left(\mathbf{A}(\widetilde{\mathbf{P}\mathbf{R}_\theta\mathbf{H}})\mathbf{B}(\widetilde{\mathbf{P}\mathbf{R}_\theta\mathbf{H}})^\intercal\right)\right] &= \mathop{\mathbb{E}}_{\mathbf{P}\sim\mathcal{A}_2}\left[\chi_x\left(\mathbf{A}(\widetilde{\mathbf{P}\mathbf{H}\mathbf{R}_{-\theta}})\mathbf{B}(\widetilde{\mathbf{P}\mathbf{H}\mathbf{R}_{-\theta}})^\intercal\right)\right]\\ &= \mathop{\mathbb{E}}_{\mathbf{P}\sim\mathcal{A}_2}\left[\chi_x\left(\mathbf{A}(\widetilde{\mathbf{P}\mathbf{R}_{-\theta}})\mathbf{B}(\widetilde{\mathbf{P}\mathbf{R}_{-\theta}})^\intercal\right)\right]\\ &= \mathop{\mathbb{E}}_{\mathbf{P}\sim\mathcal{A}_2}\left[\chi_x\left(\mathbf{A}\widetilde{\mathbf{P}}\mathbf{B}\widetilde{\mathbf{P}}^\intercal\right)\right],\end{aligned}$$

where the last equality follows by Equation (5.6). $\square$

With Lemma 5.5 in hand, we proceed with the proof of Lemma 5.3 and prove the following.

Claim 5.6. *For every pair of $n\times n$ real symmetric matrices $\mathbf{A},\mathbf{B}$ and every $\mathbf{Q}$ in the support of $\mathcal{O}(n)$,*

$$\mathop{\mathbb{E}}_{\mathbf{P}\sim\mathcal{A}_n} f_{\mathbf{A},\mathbf{B}}(\mathbf{P}) = \mathop{\mathbb{E}}_{\mathbf{P}\sim\mathcal{A}_n} f_{\mathbf{A},\mathbf{B}}(\mathbf{P}\mathbf{Q})$$

Proof. For every distinct $i,j,k\in[n]$ we denote by $\mathcal{A}_{i,j,k}$ the group of symmetries on the 3-vertex simplex on the respective vertices, fixing the remaining $n-2$ dimensions. Clearly, Lemma 5.5 holds for every choice triplet i,j,k. With this in mind, observing also that $\mathcal{A}_n$ has the same distribution as $\mathcal{A}_n\widetilde{\mathbf{P}}_2$, where $\widetilde{\mathbf{P}}_2$ is sampled from $\mathcal{A}_{i,j,k}$, we have that for every such triplet

$$\begin{aligned}\mathop{\mathbb{E}}_{\mathbf{P}\sim\mathcal{A}_n} f_{\mathbf{A},\mathbf{B}}(\mathbf{P}) &= \mathop{\mathbb{E}}_{\substack{\mathbf{P}\sim\mathcal{A}_n\\ \widetilde{\mathbf{P}}_2\in\mathcal{A}_{i,j,k}}} f_{\mathbf{A},\mathbf{B}}(\mathbf{P}\widetilde{\mathbf{P}}_2)\\ &= \mathop{\mathbb{E}}_{\mathbf{P},\widetilde{\mathbf{P}}_2}\left[\chi_x\left(\mathbf{A}\mathbf{P}\widetilde{\mathbf{P}}_2\mathbf{B}(\widetilde{\mathbf{P}}_2)^\intercal\mathbf{P}^\intercal\right)\right]\\ &= \mathop{\mathbb{E}}_{\mathbf{P},\widetilde{\mathbf{P}}_2}\left[\chi_x\left(\mathbf{P}^\intercal\mathbf{A}\mathbf{P}\widetilde{\mathbf{P}}_2\mathbf{B}(\widetilde{\mathbf{P}}_2)^\intercal\right)\right]\\ &= \mathop{\mathbb{E}}_{\mathbf{P}}\mathop{\mathbb{E}}_{\widetilde{\mathbf{P}}_2} f_{\mathbf{P}^\intercal\mathbf{A}\mathbf{P},\mathbf{B}}(\widetilde{\mathbf{P}}_2).\end{aligned}$$

Let $\mathcal{O}_{i,j,k}$ be the Haar random measure on the plane corresponding to vertices i,j,k leaving the remaining $n-2$ dimensions intact. Fix $\mathbf{P}$ and note that by Lemma 5.5, for every $\widetilde{\mathbf{Q}}_2$ in the support of $\mathcal{O}_{i,j,k}$,

$$\mathop{\mathbb{E}}_{\widetilde{\mathbf{P}}_2} f_{\mathbf{P}^\intercal\mathbf{A}\mathbf{P},\mathbf{B}}(\widetilde{\mathbf{P}}_2) = \mathop{\mathbb{E}}_{\widetilde{\mathbf{P}}_2} f_{\mathbf{P}^\intercal\mathbf{A}\mathbf{P},\mathbf{B}}(\widetilde{\mathbf{P}_2\mathbf{Q}_2}),$$

and so

$$\begin{aligned}\mathop{\mathbb{E}}_{\mathbf{P}\sim\mathcal{A}_n} f_{\mathbf{A},\mathbf{B}}(\mathbf{P}) &= \mathop{\mathbb{E}}_{\mathbf{P}}\mathop{\mathbb{E}}_{\widetilde{\mathbf{P}_2}} f_{\mathbf{P}^\intercal\mathbf{A}\mathbf{P},\mathbf{B}}(\widetilde{\mathbf{P}_2}) \\ &= \mathop{\mathbb{E}}_{\mathbf{P}}\mathop{\mathbb{E}}_{\widetilde{\mathbf{P}_2}} f_{\mathbf{P}^\intercal\mathbf{A}\mathbf{P},\mathbf{B}}(\widetilde{\mathbf{P}_2\mathbf{Q}_2}) \\ &= \mathop{\mathbb{E}}_{\mathbf{P}\sim\mathcal{A}_n} f_{\mathbf{A},\mathbf{B}}(\mathbf{P}\widetilde{\mathbf{Q}_2}). \end{aligned} \tag{5.7}$$

A basic fact about the orthogonal group is that given $\mathbf{Q}\in\mathcal{O}(n)$, one can write $\mathbf{Q}$ as the product $\mathbf{Q}=\widetilde{\mathbf{Q}^{(1)}}\cdots\widetilde{\mathbf{Q}^{(m)}}$ where $\widetilde{\mathbf{Q}^{(\ell)}}$ is in the support of $\mathcal{O}_{i_\ell,j_\ell,k_\ell}$ (see Lemma 4.7 in [MSS18]). The proof follows by applying Equation (5.7) m times. □

Modulo the proof of Lemma 5.7 which is given in the next section, the proof of Lemma 5.3 follows.

□

5.1 Proof of Lemma 5.7

As mentioned, the proof of Lemma 5.7, which we turn to now, is the point in which our proof diverge from the original proof of the analog of statement of Lemma 5.3 to the additive convolution case. [11]

Lemma 5.7. *Let $\mathbf{A},\mathbf{B}$ be a pair of $n\times n$ real symmetric matrices. Then, one can write*

$$\chi_x\left(\mathbf{A}\widetilde{\mathbf{R}_\theta}\mathbf{B}(\widetilde{\mathbf{R}_\theta})^\intercal\right)=\sum_{k=-2}^{2} c_k e^{ik\theta}$$

for some $c_{-2},c_{-1},c_0,c_1,c_2\in\mathbb{R}[x]$ that are independent of θ.

For the proof of Lemma 5.7 we need a well-known result about the determinant of a sum of two matrices in terms of their respective adjugates and compounds. We first recall these definitions, and begin by introducing the following standard notation: For an $n\times n$ matrix $\mathbf{A}$ and $S,T\subseteq[n]$ of equal size, we denote the submatrix of $\mathbf{M}$ on the row set S and column set T by $\mathbf{A}[S,T]$. We further let $[\mathbf{A}]_{S,T}=\det\mathbf{A}[S,T]$. For $r\geq 0$, the r-th *adjugate* matrix $\mathrm{Adj}_r(\mathbf{A})$ is the $\binom{n}{r}\times\binom{n}{r}$ matrix whose rows and columns are indexed by subsets of $[n]$ of size r which is defined by

$$(\mathrm{Adj}_r(\mathbf{A}))_{S,T}=(-1)^{p(S,T)}\,[\mathbf{A}]_{T^c,S^c}\,,$$

[11] Interestingly, for an invertible $\mathbf{A}$, one can proceed more or less in the lines of the proof of [MSS18]. While this is not the case in general, we note that one can reduce to that case by adding self loops to the graph at hand (when we turn to use the lemma, we will take $\mathbf{A}\succeq 0$). However, we rather work a bit harder in the analysis rather than making unnecessary changes to the construction.

where $p(S,T) = \sum_{i\in S} i + \sum_{j\in T} j$. The r-th *compound* matrix $\mathrm{C}_r(\mathbf{A})$ is the $\binom{n}{r}\times\binom{n}{r}$ matrix that is also indexed by subsets of $[n]$ of size r, which is given by $(\mathrm{C}_r(\mathbf{A}))_{S,T} = [\mathbf{A}]_{S,T}$.

Lemma 5.8 (See Chapter 0, Section 8.12 of [HJ12])**.** *Let $\mathbf{A}, \mathbf{B}$ be a pair of $n \times n$ matrices. Then, for every $s, t \in \mathbb{R}$,*

$$\det(s\mathbf{A} + t\mathbf{B}) = \sum_{r=0}^{n} s^{n-r}t^r \cdot \operatorname{Tr}\left(\mathrm{Adj}_r(\mathbf{A})\mathrm{C}_r(\mathbf{B})\right).$$

With this we are ready to prove Lemma 5.7.

Proof of Lemma 5.7. The matrix $\mathbf{R}_\theta$ can be written as

$$\mathbf{R}_\theta = \mathbf{U}\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta}\end{pmatrix}\mathbf{U}^\mathsf{T},$$

where $\mathbf{U}$ is independent of the angle θ. Hence we can write $\mathbf{R}_\theta \oplus \mathbf{I}_{n-2} = \mathbf{V}\mathbf{D}_\theta\mathbf{V}^\mathsf{T}$ where $\mathbf{V}$ is independent of θ and $\mathbf{D}_\theta$ is the diagonal matrix whose top two entries are $e^{i\theta}$, $e^{-i\theta}$, and the remaining entries on the diagonal are 1. By Lemma 5.8, we can write

$$\chi_x\left(\mathbf{A}\widetilde{\mathbf{R}_\theta}\mathbf{B}(\widetilde{\mathbf{R}_\theta})^\mathsf{T}\right) = \det\left(x\mathbf{I} - \mathbf{A}\widetilde{\mathbf{R}_\theta}\mathbf{B}(\widetilde{\mathbf{R}_\theta})^\mathsf{T}\right) = \sum_{r=0}^{n} x^{n-r}\operatorname{Tr}\left(\mathrm{C}_r(\mathbf{A}\widetilde{\mathbf{R}_\theta}\mathbf{B}(\widetilde{\mathbf{R}_\theta})^\mathsf{T})\right).$$

Thus we are left with showing that for every $S \subseteq [n]$ one can write

$$\mathrm{C}_{|S|}(\mathbf{A}\widetilde{\mathbf{R}_\theta}\mathbf{B}(\widetilde{\mathbf{R}_\theta})^\mathsf{T})_{S,S} = \sum_{k=-2}^{2} c_k(S)e^{ik\theta},$$

for some $c_{-2}(S), c_{-1}(S), c_0(S), c_1(S), c_2(S) \in \mathbb{R}$ that are independent of θ. To show this, fix a set $S \subseteq [n]$ and denote $r = |S|$. By the Cauchy-Binet formula (see Chapter 0, Section 8.7 of [HJ12]),

$$\mathrm{C}_r(\mathbf{A}\widetilde{\mathbf{R}_\theta}\mathbf{B}(\widetilde{\mathbf{R}_\theta})^\mathsf{T})_{S,S} = \left[\mathbf{A}\widetilde{\mathbf{R}_\theta}\mathbf{B}(\widetilde{\mathbf{R}_\theta})^\mathsf{T}\right]_{S,S} = \sum_{\substack{T\subseteq[n]\\|T|=r}} [\mathbf{A}]_{S,T}\left[\widetilde{\mathbf{R}_\theta}\mathbf{B}(\widetilde{\mathbf{R}_\theta})^\mathsf{T}\right]_{T,S}.$$

As $\mathbf{A}$ is independent of θ we may focus our attention on $\left[\widetilde{\mathbf{R}_\theta}\mathbf{B}(\widetilde{\mathbf{R}_\theta})^\mathsf{T}\right]_{T,S}$. We have that

$$\widetilde{\mathbf{R}_\theta}\mathbf{B}(\widetilde{\mathbf{R}_\theta})^\mathsf{T} = \mathbf{V}\mathbf{D}_\theta\mathbf{V}^\mathsf{T}\mathbf{B}\mathbf{V}\mathbf{D}_\theta\mathbf{V}^\mathsf{T},$$

and so, again by the Cauchy-Binet formula

$$\left[\widetilde{\mathbf{R}_\theta}\mathbf{B}(\widetilde{\mathbf{R}_\theta})^\mathsf{T}\right]_{T,S} = \sum_{\substack{R,P\subseteq[n]\\|R|=|P|=r}} [\mathbf{V}]_{T,R}\left[\mathbf{D}_\theta\mathbf{V}^\mathsf{T}\mathbf{B}\mathbf{V}\mathbf{D}_\theta\right]_{R,P}\left[\mathbf{V}^\mathsf{T}\right]_{P,S}.$$

Since $\mathbf{V}$ is independent of θ, it suffices to prove that for every R, P as above, one can write

$$\left[\mathbf{D}_\theta \mathbf{V}^\top \mathbf{B} \mathbf{V} \mathbf{D}_\theta\right]_{R,P} = \sum_{k=-2}^{2} c_k(R,P) e^{ik\theta},$$

where $c_{-2}(R,P), \ldots, c_2(R,P) \in \mathbb{R}$ are independent of θ. This is straightforward to verify as $\mathbf{C} = \mathbf{V}^\top \mathbf{B} \mathbf{V}^\top$ is independent of θ and since

$$(\mathbf{D}_\theta \mathbf{C} \mathbf{D}_\theta)_{r,s} = c_{r,s} e^{i\theta(\delta_{s,1} - \delta_{s,2} + \delta_{r,1} - \delta_{r,2})},$$

where $\delta_{i,j}$ is the Kronecker delta function. This completes the proof of Lemma 5.7.

□

5.2 Proof of Theorem 5.2

With Lemma 5.3 in hand we are ready to prove the following lemma that slightly generalizes Theorem 5.2.

Lemma 5.9. *Let $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_t$ be $n \times n$ real symmetric matrices with $\mathbf{A}_i \mathbf{1} = a_i \mathbf{1}$, and $p_{\mathbf{A}_i^2}(x)$ the polynomial satisfying $\chi_x(\mathbf{A}_i^2) = (x - a_i^2) p_{\mathbf{A}_i^2}(x)$. Let $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_{t-1}$ be independent random permutation matrices. Define*

$$\begin{aligned}
\mathbf{B}_1 &= \mathbf{A}_1^2, \\
\mathbf{B}_j &= \mathbf{A}_j \mathbf{P}_{j-1} \mathbf{B}_{j-1} \mathbf{P}_{j-1}^\top \mathbf{A}_j \quad \textit{for } j = 2, \ldots, t.
\end{aligned}$$

Then,

$$\mathop{\mathbb{E}}_{\mathbf{P}_1, \ldots, \mathbf{P}_{t-1}} \chi_x(\mathbf{B}_t) = \left(x - \prod_{i=1}^{t} a_i^2\right) \left(p_{\mathbf{A}_1^2} \boxtimes p_{\mathbf{A}_2^2} \boxtimes \cdots \boxtimes p_{\mathbf{A}_t^2}\right)(x).$$

Proof. For $j = 1, \ldots, t$ denote $\alpha_j = \prod_{i=1}^{j} a_i^2$. We proceed by induction on j, where the base case $j = 1$ trivially holds. It is easy to see that $\mathbf{B}_{j-1} \mathbf{1} = \alpha_{j-1} \mathbf{1}$. Moreover, by the induction hypothesis,

$$\mathop{\mathbb{E}}_{\mathbf{P}_1, \ldots, \mathbf{P}_{j-2}} \chi_x(\mathbf{B}_{j-1}) = (x - \alpha_{j-1}) \left(p_{\mathbf{A}_1^2} \boxtimes \cdots \boxtimes p_{\mathbf{A}_{j-1}^2}\right)(x).$$

Therefore,

$$\mathop{\mathbb{E}}_{\mathbf{P}_1, \ldots, \mathbf{P}_{j-2}} \chi_x(\widehat{\mathbf{B}}_{j-1}) = \left(p_{\mathbf{A}_1^2} \boxtimes \cdots \boxtimes p_{\mathbf{A}_{j-1}^2}\right)(x), \tag{5.8}$$

where we remind the reader of the $\widehat{\cdot}$ notation that was defined at the beginning of the proof of Lemma 5.3. We have that

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathbf{P}_1, \ldots, \mathbf{P}_{j-1}} \chi_x(\mathbf{B}_j) &= \mathop{\mathbb{E}}_{\mathbf{P}_1, \ldots, \mathbf{P}_{j-1}} \chi_x(\mathbf{A}_j \mathbf{P}_{j-1} \mathbf{B}_{j-1} \mathbf{P}_{j-1}^\top \mathbf{A}_j) \\
&= \mathop{\mathbb{E}}_{\mathbf{P}_1, \ldots, \mathbf{P}_{j-2}} \mathop{\mathbb{E}}_{\mathbf{P}_{j-1}} \chi_x\left(\mathbf{A}_j \mathbf{P}_{j-1} \mathbf{B}_{j-1} \mathbf{P}_{j-1}^\top \mathbf{A}_j\right).
\end{aligned} \tag{5.9}$$

Note that $\mathbf{B}_{j-1}$ is independent of $\mathbf{P}_{j-1}$ and so by Lemma 5.3, for every fixing of $\mathbf{P}_1, \ldots, \mathbf{P}_{j-2}$

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathbf{P}_{j-1}} \chi_x\left(\mathbf{A}_j \mathbf{P}_{j-1} \mathbf{B}_{j-1} \mathbf{P}_{j-1}^{\mathsf{T}} \mathbf{A}_j\right) &= \mathop{\mathbb{E}}_{\mathbf{P}_{j-1}} \chi_x\left(\mathbf{A}_j^2 \mathbf{P}_{j-1} \mathbf{B}_{j-1} \mathbf{P}_{j-1}^{\mathsf{T}}\right) \\
&= (x - a_j^2 \alpha_{j-1}) p_{\mathbf{A}_j^2}(x) \boxtimes \chi_x(\widehat{\mathbf{B}}_{j-1}) \\
&= (x - \alpha_j) p_{\mathbf{A}_j^2}(x) \boxtimes \chi_x(\widehat{\mathbf{B}}_{j-1}), \qquad (5.10)
\end{aligned}$$

where we slightly abuse notation in that $\mathbf{B}_{j-1}$ above is with respect to the specific fixing of $\mathbf{P}_1, \ldots, \mathbf{P}_{j-2}$. At any rate, by Equations (5.8) to (5.10) and using the bi-linearity of the multiplicative convolution,

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathbf{P}_1,\ldots,\mathbf{P}_{j-1}} \chi_x(\mathbf{B}_j) &= (x - \alpha_j) p_{\mathbf{A}_j^2}(x) \boxtimes \mathop{\mathbb{E}}_{\mathbf{P}_1,\ldots,\mathbf{P}_{j-2}} \chi_x(\widehat{\mathbf{B}}_{j-1}) \\
&= (x - \alpha_j) p_{\mathbf{A}_j^2}(x) \boxtimes \left(p_{\mathbf{A}_1^2} \boxtimes \cdots \boxtimes p_{\mathbf{A}_{j-1}^2}\right)(x).
\end{aligned}$$

The proof then follows by the commutativity and associativity of the multiplicative convolution. $\square$

6 Interlacing: From Random Permutations to a Tailor-Made Sequence

As mentioned in the introduction, we would like to use our analysis for the expected characteristic polynomial to deduce something about a specific polynomial. That is, we would like to prove that there exists a polynomial (a characteristic polynomial of some graph) whose second largest root can be bounded. This is not true in general. Given a real-rooted polynomial which is the sum of two polynomials, each of the two polynomials may not even be real-rooted, let alone having bounded roots. However, there is a case in which such a deduction can be made, and this case being if the polynomials over which we take the expectation form an *interlacing family*, a term that will be formalized below. The main theorem that we will use for relating the expectation to a specific polynomial is Lemma 6.3, which was used similarly in [MSS18].

Let $p(x)$ and $q(x)$ be real-rooted polynomials of the same degree n, and let $\alpha_1 \geq \cdots \geq \alpha_n$ be the roots of $p(x)$ and $\beta_1 \geq \cdots \geq \beta_n$ be the roots of $q(x)$. We say that $q(x)$ *interlaces* $p(x)$, and write $q \to p$, if $\beta_n \leq \alpha_n \leq \beta_{n-1} \leq \cdots \leq \beta_1 \leq \alpha_1$. If $q(x)$ is of degree $n-1$, we use the same condition without the left most inequality. In this case, as there is no confusion, we sometimes say that p and q interlace. A *common interlacing* for polynomials $p(x), q(x)$ of the same degree n is a third polynomial $r(x)$ that interlaces both $p(x)$ and $q(x)$. We remark that it does not matter if the common interlacing is of degree n or $n-1$, though in our proofs, the degree will be n.

Definition 6.1 (Interlacing families [MSS13]). Let $T = (V, E)$ be a full rooted ordered binary tree[12] such that every $v \in V$ is labeled by a polynomial $p_v(x) \in \mathbb{R}[x]$. Denote $\mathcal{P}_T = \{p_v(x)\}_{v \in V}$. The collection $(T, \mathcal{P}_T)$ is called an *interlacing family* if the following holds:

1. For every $v \in V$, $p_v(x)$ is a monic, real-rooted polynomial. Furthermore, all polynomials in $\mathcal{P}_T$ have the same degree.

2. For every internal node v with sons u, w it holds that $p_v(x) = \alpha p_u(x) + \beta p_w(x)$ for some non-negative $\alpha = \alpha(v)$, $\beta = \beta(v)$.

3. For every two siblings $u, w \in V$, the corresponding polynomials $p_u(x), p_w(x)$ have a common interlacing $r_{u,w}(x)$.

Definition 6.2 ([MSS13]). *A set of polynomials $\mathcal{P}$ is said to* form an interlacing family *if there exists an interlacing family $(T, \mathcal{P}_T)$ such that for every leaf v, $p_v(x) \in \mathcal{P}$.*

Lemma 6.3. *Let $p_1(x), \ldots, p_m(x)$ be a set of polynomials of degree n which forms an interlacing family. Let T be the corresponding tree having root r. Then, for every $j \in [n]$ there exists $i \in [m]$ such that the j-th largest root of $p_i(x)$ is at most the j-th largest root of $p_r(x)$.*

For proving Lemma 6.3, we introduce the following notation: For a real-rooted polynomial $p(x)$ of degree n, and $j \in [n]$, we let $\lambda_j(p(x))$ denotes its j-th largest root. We make use of the following simple claim from [Spi19].

Claim 6.4. *Let $p_1(x), p_2(x)$ be two polynomials of degree n with a positive leading coefficient, having a common interlacing. Then, for every $\alpha \in [0, 1]$ and every $j \in [n]$ there exists $i \in \{1, 2\}$ such that*

$$\lambda_j(p_i(x)) \le \lambda_j(\alpha p_1(x) + (1 - \alpha)p_2(x)).$$

Proof of Lemma 6.3. Fix $j \in [n]$. By Claim 6.4 and Property (2) of Definition 6.1, every non-leaf v in T has at least one son u for which $\lambda_j(p_u(x)) \le \lambda_j(p_v(x))$. Starting from the root, we can proceed down the tree, maintaining this invariant, until we reach a leaf. $\square$

With these definitions in place, also denoting the set of $n \times n$ permutation matrices by $\mathbf{S}_n$, we are in a position to state the main result of this section.

[12] Recall that in an ordered tree, the sons of a given node are ordered. In particular, in a full ordered binary tree, every internal node has a left son and a right son.

Theorem 6.5. *Let $\mathbf{A}$ be a real symmetric $n \times n$ matrix. Then, the polynomials*

$$\{\chi_x (\Psi_{\mathbf{A}}(\mathbf{P}_1, \dots, \mathbf{P}_{t-1})) \mid \mathbf{P}_1, \dots, \mathbf{P}_{t-1} \in \mathbf{S}_n\} \tag{6.1}$$

form an interlacing family $(T, \mathcal{P}_T)$ such that the root of T is labeled by the polynomial

$$\mathop{\mathbb{E}}_{\mathbf{P}_1,\dots,\mathbf{P}_{t-1}} \chi_x (\Psi_{\mathbf{A}}(\mathbf{P}_1, \dots, \mathbf{P}_{t-1})), \tag{6.2}$$

where, in the expectation above, $\mathbf{P}_1, \dots, \mathbf{P}_{t-1}$ are sampled uniformly and independently at random from $\mathbf{S}_n$.

Towards proving Theorem 6.5 we set some notation. Given $i, j \in [n]$ we denote the $n \times n$ matrix that corresponds to the involution $(i\,j)$ by $\mathbf{\Gamma}_{i,j}$. That is, $\mathbf{\Gamma}_{i,j}\mathbf{e}_i = \mathbf{e}_j$, $\mathbf{\Gamma}_{i,j}\mathbf{e}_j = \mathbf{e}_i$ and for every $k \in [n] \setminus \{i, j\}$, $\mathbf{\Gamma}_{i,j}\mathbf{e}_k = \mathbf{e}_k$. An $n \times n$ matrix $\mathbf{\Gamma}$ is called a *swap matrix* if $\mathbf{\Gamma} = \mathbf{\Gamma}_{i,j}$ for some $i, j \in [n]$. For $\alpha \in [0, 1]$, a *random swap* $\mathbf{S} = \mathbf{S}(\alpha, i, j)$ is a random $n \times n$ matrix which is equal to $\mathbf{\Gamma}_{i,j}$ with probability $1 - \alpha$ and to the identity matrix with probability α.

The basic building block for describing the tree underlying the proof of Theorem 6.5 is the full, rooted, ordered binary tree $\mathcal{T}_1$. With every node of the tree we associate a distribution over $\mathbf{S}_n$ such that the following properties hold:

1. To a leaf v that is labeled by $\mathbf{Q} \in \mathbf{S}_n$, we have that $\nu_v = \mathbf{Q}$. That is, the distribution associated with v is the distribution in which $\mathbf{Q}$ is sampled with probability 1.

2. Let v be an internal node with left son u and right son w. Then, there exist $i_v, j_v \in [n]$ and $\alpha_v \in [0, 1]$ such that

$$\begin{aligned}\nu_w &= \nu_u \mathbf{\Gamma}_{i_v, j_v},\\ \nu_v &= \alpha_v \nu_u + (1 - \alpha_v)\nu_w.\end{aligned}$$

3. To the root r corresponds the distribution ν_r which is uniform over $\mathbf{S}_n$.

The description of a tree $\mathcal{T}_1$ with the above properties was given in [MSS18]. For completeness, and since we use a somewhat different notation, we sketch the construction in Appendix B. Moving forward, the depth of $\mathcal{T}_1$ is denoted by N.

We wish to construct a tree $\mathcal{T}_m$ representing a product distribution of m permutations on n elements; That is, we draw each of the permutations uniformly and independently at random. The construction is the intuitive one, and formally goes as follows: starting with $\mathcal{T}_1$, which we think of as the first drawn permutation, to each of its leaves, we attach another copy of $\mathcal{T}_1$, and repeat this process for m times (the overall depth of the tree is mN).

Let x be a node of $\mathcal{T}_1$ with corresponding distribution ν_x, and let v be an occurrence of x in a copy of $\mathcal{T}_1$, after $\mathcal{T}_1$ has been attached to a leaf r times. Denote those intermediate leaves by $l_1, \ldots, l_r$, and their corresponding permutations by $\mathbf{Q}_1, \ldots, \mathbf{Q}_r$. Then, the (product) distribution of v is defined by

$$\nu_v = \mathbf{Q}_1 \times \cdots \times \mathbf{Q}_r \times \nu_x \times \mathbf{S}_n \times \cdots \times \mathbf{S}_n,$$

where $\mathbf{S}_n$ repeats $m-(r+1)$ times and denotes here, with a slight abuse of notation, the uniform distribution over the set $\mathbf{S}_n$.

For proving the main result of this section, Theorem 6.5, we need to prove a more general result, which considers the distributions of the tree not only with respect to the matrix $\mathbf{A}$, but rather for all matrices of a certain form. To this end, we need further notations as given by the following definitions.

Definition 6.6. *Given an $n \times n$ real matrix $\mathbf{M}$ and $n \times n$ permutation matrices $\mathbf{P}_1, \ldots, \mathbf{P}_m$, we define the $n \times n$ matrices $\mathbf{M}_{\mathbf{P}_1,\ldots,\mathbf{P}_j}$ for $j = 0, 1, \ldots, m$ recursively as follows: For $j = 0$ we set $\mathbf{M}_{\perp} = \mathbf{I}$, and for $j \geq 1$,*

$$\mathbf{M}_{\mathbf{P}_1,\ldots,\mathbf{P}_j} = \mathbf{M}\mathbf{P}_j\mathbf{M}_{\mathbf{P}_1,\ldots,\mathbf{P}_{j-1}}.$$

Definition 6.7 (Property $\star$ for distributions)**.** *A probability distribution ν on $\mathbf{S}_n^m$ is said to satisfy* property $\star$ *if for every pair of $n \times n$ real matrices $\mathbf{M}, \mathbf{N}$, with $\mathbf{N}$ symmetric, the polynomial*

$$\varphi(\nu, \mathbf{M}, \mathbf{N})(x) \triangleq \mathop{\mathbb{E}}_{\mathbf{P}\sim\nu}\left[\chi_x(\mathbf{M}_{\mathbf{P}}\mathbf{N}\mathbf{M}_{\mathbf{P}}^{\mathsf{T}})\right]$$

is real-rooted.

Definition 6.8 (Property $\star$ for trees)**.** *Let $\mathcal{T} = (V, E)$ be a rooted ordered binary tree such that with every $v \in V$ we associate a distribution ν_v. $\mathcal{T}$ is said to satisfy* property $\star$ *if the following holds:*

1. *$\forall v \in V$ the distribution ν_v satisfies property $\star$; and*
2. *For every pair of matrices $\mathbf{M}, \mathbf{N}$ such that $\mathbf{N}$ is symmetric, and every pair of siblings u, w in $\mathcal{T}$, the polynomials $\varphi(\nu_u, \mathbf{M}, \mathbf{N})(x)$ and $\varphi(\nu_w, \mathbf{M}, \mathbf{N})(x)$ have a common interlacing.*

The following straightforward assertion, left without a proof, will be useful.

Claim 6.9. *For every pair of distributions ν, μ on $\mathbf{S}_n$, $\alpha \in [0,1]$, and every pair of $n \times n$ real matrices $\mathbf{M}, \mathbf{N}$,*

$$\varphi(\alpha\nu + (1-\alpha)\mu, \mathbf{M}, \mathbf{N}) = \alpha\varphi(\nu, \mathbf{M}, \mathbf{N}) + (1-\alpha)\varphi(\mu, \mathbf{M}, \mathbf{N}). \tag{6.3}$$

Claim 6.10. *Let $m \geq 1$ be an integer, and ν a distribution over $\mathbf{S}_n^{m-1}$. For every distribution μ over $\mathbf{S}_n$, every $\mathbf{Q} \in \mathbf{S}_n$ and for all $n \times n$ real matrices $\mathbf{M}, \mathbf{N}$,*

$$\varphi((\mu\mathbf{Q}) \times \nu, \mathbf{M}, \mathbf{N}) = \varphi(\mu \times \nu, \mathbf{M}, \mathbf{Q}\mathbf{N}\mathbf{Q}^{\mathsf{T}}).$$

Proof. We have that

$$\begin{aligned}
\varphi((\mu\mathbf{Q}) \times \nu, \mathbf{M}, \mathbf{N}) &= \mathop{\mathbb{E}}_{\substack{\mathbf{P}\sim\mu \\ \mathbf{R}\sim\nu}} \chi_x \left(\mathbf{M}_{\mathbf{R}}\mathbf{M}(\mathbf{P}\mathbf{Q})\mathbf{N}(\mathbf{P}\mathbf{Q})^{\mathsf{T}}\mathbf{M}^{\mathsf{T}}\mathbf{M}_{\mathbf{R}}^{\mathsf{T}}\right) \\
&= \mathop{\mathbb{E}}_{\substack{\mathbf{P}\sim\mu \\ \mathbf{R}\sim\nu}} \chi_x \left(\mathbf{M}_{\mathbf{R}}\mathbf{M}\mathbf{P}(\mathbf{Q}\mathbf{N}\mathbf{Q}^{\mathsf{T}})\mathbf{P}^{\mathsf{T}}\mathbf{M}^{\mathsf{T}}\mathbf{M}_{\mathbf{R}}^{\mathsf{T}}\right) \\
&= \varphi(\mu \times \nu, \mathbf{M}, \mathbf{Q}\mathbf{N}\mathbf{Q}^{\mathsf{T}}).
\end{aligned}$$

□

Lemma 6.11. *Let $m \geq 1$ be an integer, and ν a distribution over $\mathbf{S}_n^{m-1}$. Let ν_1, ν_2 be distributions over $\mathbf{S}_n^m$ such that $\nu_1 = \mu_1 \times \nu$ and $\nu_2 = \mu_2 \times \nu$ for some distributions μ_1, μ_2 over $\mathbf{S}_n$. Assume that*

1. *$\mu_2 = \mu_1\mathbf{\Gamma}$ for some swap matrix $\mathbf{\Gamma}$; and*

2. *Both ν_1 and ν_2 satisfy property $\star$.*

Then, for every $n \times n$ real matrices $\mathbf{M}, \mathbf{N}$ such that $\mathbf{N}$ is symmetric, $\varphi(\nu_1, \mathbf{M}, \mathbf{N})(x)$ and $\varphi(\nu_2, \mathbf{M}, \mathbf{N})(x)$ have a common interlacing, and every convex combination of ν_1 and ν_2 satisfies property $\star$.

For the proof of Lemma 6.11 we need the following sequence of lemmata that relate interlacing and real rooted-ness. The proofs can be found in, e.g., Chapter 42 of [Spi19].

Lemma 6.12. Let $p(x), q(x)$ be polynomials of degree $n, n-1$ respectively with a positive leading coefficient, and let $p_t(x) = p(x) - tq(x)$. If $p(x), q(x)$ interlace then for every $t > 0$, $p_t(x)$ is real rooted and $p \to p_t$.

Lemma 6.13. Let $p(x), q(x)$ be polynomials of degree $n, n-1$, respectively, both of which have a positive leading coefficient. Let $p_t(x) = p(x) - tq(x)$. If $p_t(x)$ is real-rooted for every $t \in \mathbb{R}$ then $p(x)$ and $q(x)$ interlace.

Lemma 6.14. Let $\mathbf{A}$ be an $n \times n$ real symmetric matrix and $\mathbf{v} \in \mathbb{R}^n$. For $t \in \mathbb{R}$ define $p_t(x) = \chi_x(\mathbf{A} + t \cdot \mathbf{v}\mathbf{v}^\intercal)$. Then, there exists a degree $n-1$ monic polynomial $q(x)$ such that $p_t(x) = \chi_x(\mathbf{A}) - t \cdot q(x)$.

Lemma 6.15. Let $p_0(x), p_1(x)$ be polynomials of degree n with a positive leading coefficient, having a common interlacing. Then, $q(x) = tp_0(x) + (1-t)p_1(x)$ is real-rooted for every $t \in [0,1]$.

We also make use of the following lemma which states that applying a swap to a matrix is a rank two update of a specific form.

Lemma 6.16 (Lemma 3.10 in [MSS18])**.** *Let $\mathbf{A}$ be a real $n \times n$ symmetric matrix and $\mathbf{\Gamma} \in \mathbf{S}_n$ a swap matrix. Then, there exist vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ such that $\mathbf{\Gamma}\mathbf{A}\mathbf{\Gamma}^\intercal = \mathbf{A} - \mathbf{u}\mathbf{u}^\intercal + \mathbf{v}\mathbf{v}^\intercal$.*

Proof of Lemma 6.11. Fix $\mathbf{M}, \mathbf{N}$ as in the statement of the lemma. By Claim 6.10 and Lemma 6.16, as $\mu_2 = \mu_1\mathbf{\Gamma}$, we have that

$$\varphi(\nu_2, \mathbf{M}, \mathbf{N}) = \varphi(\nu_1, \mathbf{M}, \mathbf{\Gamma}\mathbf{N}\mathbf{\Gamma}^\intercal) = \varphi(\nu_1, \mathbf{M}, \mathbf{N} - \mathbf{u}\mathbf{u}^\intercal + \mathbf{v}\mathbf{v}^\intercal)$$

for some $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. For $t \in \mathbb{R}$ define

$$p_t(x) = \varphi(\nu_1, \mathbf{M}, \mathbf{N} + t \cdot \mathbf{v}\mathbf{v}^\intercal)(x). \tag{6.4}$$

We have that

$$\begin{aligned} p_t(x) &= \mathop{\mathbb{E}}_{\substack{\mathbf{P}\sim\mu_1 \\ \mathbf{R}\sim\nu}} \chi_x\left(\mathbf{M}_{\mathbf{PR}}(\mathbf{N} + t \cdot \mathbf{v}\mathbf{v}^\intercal)\mathbf{M}_{\mathbf{PR}}^\intercal\right) \\ &= \mathop{\mathbb{E}}_{\substack{\mathbf{P}\sim\mu_1 \\ \mathbf{R}\sim\nu}} \chi_x\left(\mathbf{M}_{\mathbf{PR}}\mathbf{N}\mathbf{M}_{\mathbf{PR}}^\intercal + t \cdot \mathbf{w}_{\mathbf{PR}}\mathbf{w}_{\mathbf{PR}}^\intercal\right), \end{aligned}$$

where $\mathbf{M}_{\mathbf{PR}} = \mathbf{M}_{\mathbf{R}}\mathbf{M}\mathbf{P}$ and $\mathbf{w}_{\mathbf{PR}} = \mathbf{M}_{\mathbf{PR}}\mathbf{v}$. By Lemma 6.14, for every term in the expectation we can write

$$\chi_x\left(\mathbf{M}_{\mathbf{PR}}\mathbf{N}\mathbf{M}_{\mathbf{PR}}^\intercal + t \cdot \mathbf{w}_{\mathbf{PR}}\mathbf{w}_{\mathbf{PR}}^\intercal\right) = \chi_x\left(\mathbf{M}_{\mathbf{PR}}\mathbf{N}\mathbf{M}_{\mathbf{PR}}^\intercal\right) - t \cdot q_{\mathbf{PR}}(x),$$

where $q_{\mathbf{PR}}(x)$ is monic of degree $n-1$. We can therefore write

$$p_t(x) = \varphi(\nu_1, \mathbf{M}, \mathbf{N})(x) - t \cdot q(x), \tag{6.5}$$

where $q(x) = \mathbb{E}_{\mathbf{P},\mathbf{R}}[q_{\mathbf{PR}}(x)]$ is a monic polynomial of degree $n-1$.

As ν_1 satisfies property $\star$, observing Equation (6.4), we see that $p_t(x)$ is real-rooted for every $t \in \mathbb{R}$. Hence, looking also on Equation (6.5), we have by Lemma 6.13 that

$q \to \varphi(\nu_1, \mathbf{M}, \mathbf{N})$. Therefore, Lemma 6.12 (which, with its notation, is applied with $p(x) = \varphi(\nu_1, \mathbf{M}, \mathbf{N})(x)$ and $q(x)$ as itself), when specialized to $t = 1$, yields

$$\varphi\left(\nu_1, \mathbf{M}, \mathbf{N}\right) \to \varphi\left(\nu_1, \mathbf{M}, \mathbf{N} + \mathbf{v}\mathbf{v}^\mathsf{T}\right).$$

A similar argument, invoked by defining

$$\widehat{p}_t(x) = \varphi(\nu_1, \mathbf{M}, \mathbf{N} - \mathbf{u}\mathbf{u}^\mathsf{T} + \mathbf{v}\mathbf{v}^\mathsf{T} + t \cdot \mathbf{u}\mathbf{u}^\mathsf{T})(x). \tag{6.6}$$

instead of $p_t(x)$ from Equation (6.4), can be carried out to show that

$$\varphi(\nu_2, \mathbf{M}, \mathbf{N}) \to \varphi(\nu_1, \mathbf{M}, \mathbf{N} + \mathbf{v}\mathbf{v}^\mathsf{T}).$$

Thus, $\varphi(\nu_1, \mathbf{M}, \mathbf{N} + \mathbf{v}\mathbf{v}^\mathsf{T})$ is a common interlacing of $\varphi(\nu_1, \mathbf{M}, \mathbf{N})$ and $\varphi(\nu_2, \mathbf{M}, \mathbf{N})$, establishing the first part of the lemma. To conclude the proof, we invoke Lemma 6.15 which ensures that every convex combination of $\varphi(\nu_1, \mathbf{M}, \mathbf{N})$ and $\varphi(\nu_2, \mathbf{M}, \mathbf{N})$ is real-rooted, and so, by Claim 6.9, every convex combination of ν_1 and ν_2 satisfies property $\star$. □

Claim 6.17. *Let ν_1, ν_2 be distributions over $\mathbf{S}_n^m$ that satisfy property $\star$. Assume that for every real $n \times n$ matrices $\mathbf{M}, \mathbf{N}$, with $\mathbf{N}$ symmetric, we have that $\varphi(\nu_1, \mathbf{M}, \mathbf{N})$ and $\varphi(\nu_2, \mathbf{M}, \mathbf{N})$ have a common interlacing. Then, for every $\mathbf{Q} \in \mathbf{S}_n$, both $\mathbf{Q} \times \nu_1$ and $\mathbf{Q} \times \nu_2$ have the same properties.*

Proof. Fix a pair of matrices $\mathbf{M}$, $\mathbf{N}$ as above. Then for $i \in \{1, 2\}$,

$$\begin{aligned}\varphi\left(\mathbf{Q} \times \nu_i, \mathbf{M}, \mathbf{N}\right)(x) &= \mathop{\mathbb{E}}_{\mathbf{P} \sim \nu_i} \chi_x\left(\mathbf{M}_\mathbf{P}\left(\mathbf{M}\mathbf{Q}\mathbf{N}\mathbf{Q}^\mathsf{T}\mathbf{M}^\mathsf{T}\right)\mathbf{M}_\mathbf{P}^\mathsf{T}\right) \\ &= \varphi\left(\nu_i, \mathbf{M}, \mathbf{M}\mathbf{Q}\mathbf{N}\mathbf{Q}^\mathsf{T}\mathbf{M}^\mathsf{T}\right)(x).\end{aligned}$$

Denote $\mathbf{N}' = \mathbf{M}\mathbf{Q}\mathbf{N}\mathbf{Q}^\mathsf{T}\mathbf{M}^\mathsf{T}$. Since $\mathbf{N}'$ is symmetric, by the assumption on ν_1, ν_2 we get that $\mathbf{Q} \times \nu_1$ and $\mathbf{Q} \times \nu_2$ satisfy property $\star$ and that $\varphi(\mathbf{Q} \times \nu_1, \mathbf{M}, \mathbf{N})$ and $\varphi(\mathbf{Q} \times \nu_2, \mathbf{M}, \mathbf{N})$ have a common interlacing. □

Proposition 6.18. *$\mathcal{T}_m$ satisfies property $\star$.*

Proof. The proof is by induction on $m \geq 1$, where in each induction step there will be an inner inductive argument from the leaves upwards to the root. Starting with the base case $m = 1$, to every leaf l of the tree we associate a permutation $\mathbf{Q}$, and so the expectation is over a single element, hence

$$\varphi(\nu_l, \mathbf{M}, \mathbf{N})(x) = \chi_x(\mathbf{M}\mathbf{Q}\mathbf{N}\mathbf{Q}^\mathsf{T}\mathbf{M}^\mathsf{T})$$

which is real-rooted. Therefore, for every leaf l, ν_l satisfies property $\star$.

We proceed by induction from the leaves of $\mathcal{T}_1$ upwards. Let v be a node in $\mathcal{T}_1$ with left son u and right son w, for which ν_u and ν_w satisfy property $\star$. By Property (2) of $\mathcal{T}_1$, we can write $\nu_w = \nu_u \mathbf{\Gamma}$ for some swap matrix $\mathbf{\Gamma}$. Moreover, ν_v is a convex combination of ν_w, ν_u. Lemma 6.11 then implies that ν_v satisfies property $\star$. Furthermore, by Lemma 6.11, for every pair of matrices $\mathbf{M}, \mathbf{N}$, with $\mathbf{N}$ symmetric, the polynomials $\varphi(\nu_u, \mathbf{M}, \mathbf{N})(x)$ and $\varphi(\nu_w, \mathbf{M}, \mathbf{N})(x)$ have a common interlacing. We therefore conclude that $\mathcal{T}_1$ satisfies property $\star$.

Let $m \geq 2$, and assume $\mathcal{T}_{m-1}$ satisfies property $\star$. By Claim 6.17, for every $\mathbf{Q} \in \mathbf{S}_n$ and every node v in $\mathcal{T}_{m-1}$ it holds that $\mathbf{Q} \times \nu_v$ satisfies property $\star$. Furthermore, for every pair of siblings u, w and for all matrices $\mathbf{M}, \mathbf{N}$ as above, $\varphi(\mathbf{Q} \times \nu_u, \mathbf{M}, \mathbf{N})$ and $\varphi(\mathbf{Q} \times \nu_w, \mathbf{M}, \mathbf{N})$ have a common interlacing. By the construction of $\mathcal{T}_m$, this applies to the entire tree except for the first $(N-1)$ layers, and so we are left to handle only the part from the root to the first concatenation of $\mathcal{T}_1$.

The nodes of depth N - the "leaves" of the first appearance of $\mathcal{T}_1$ - satisfy property $\star$. In the first $(N-1)$ layers, to each node v we associated a distribution $\mu_v \times \nu$, where ν is the uniform distribution over $\mathbf{S}_n^{m-1}$. Furthermore, for every pair of siblings u, w we have $\mu_w = \mu_u \mathbf{\Gamma}$ for some swap matrix $\mathbf{\Gamma}$. By Claim 6.9 we get that the distribution associated with any parent of these leaves is a convex combination of the corresponding distributions of its children. We can now use induction from the "leaves" of depth N to the root, using Lemma 6.11, to get the desired property over all of the tree. $\square$

Theorem 6.5 readily follows by Proposition 6.18 as follows.

Proof of Theorem 6.5. Consider the tree $T = \mathcal{T}_{t-1}(\mathbf{A})$ that has the same tree structure as $\mathcal{T}_{t-1}$ though we ignore the distribution ν_v attached to a node v in $\mathcal{T}_{t-1}$ and only associate the polynomial $\varphi(\nu_v, \mathbf{A}, \mathbf{A}^2)(x)$ to the node. Note that the polynomials in $\mathcal{P}_T \triangleq \{\varphi(\nu_v, \mathbf{A}, \mathbf{A}^2)\}_{v \in V}$ are all monic, real-rooted, and have same degree, n. Moreover, to every leaf l corresponds some distribution which outputs $\mathbf{P} = (\mathbf{P}_{t-1}, \ldots, \mathbf{P}_1)$ with probability 1 for some permutation matrices $\mathbf{P}_{t-1}, \ldots, \mathbf{P}_1$. Recalling Definition 5.1 and Definition 6.7, note that

$$\varphi(\nu_\ell, \mathbf{A}, \mathbf{A}^2)(x) = \chi_x\left(\mathbf{A}_{\mathbf{P}} \mathbf{A}^2 \mathbf{A}_{\mathbf{P}}^\top\right) = \chi_x\left(\Psi_{\mathbf{A}}(\mathbf{P}_1, \ldots, \mathbf{P}_{t-1})\right). \tag{6.7}$$

Namely, the leaves of $\mathcal{T}_m(\mathbf{A})$ are labeled by polynomials from Equation (6.1). Observe that, as $\mathcal{T}_m$ satisfies property $\star$, $(T, \mathcal{P}_T)$ is an interlacing family. Thus, the polynomials that are given by Equation (6.1) form an interlacing family. Moreover, since the distribution that is associated to the root of $\mathcal{T}_{t-1}$ is the uniform distribution over $\mathbf{S}_n^{t-1}$, we have that the polynomial associated with the root of T is the one given by Equation (6.2), concluding the proof. $\square$

7 Proof of Theorem 1.1 and Theorem 1.2

In this section we put it all together and prove the following two theorems which formalize Theorem 1.1 and Theorem 1.2. We recall that for a graph G with adjacency matrix $\mathbf{A}$ and a sequence of permutation matrices $\mathbf{P} = (\mathbf{P}_1, \ldots, \mathbf{P}_{t-1})$, we let $G_\mathbf{P}$ be the graph whose adjacency matrix is

$$\mathbf{A}_\mathbf{P} \triangleq \mathbf{A}\mathbf{P}_{t-1} \cdots \mathbf{A}\mathbf{P}_1 \mathbf{A}^2 \mathbf{P}_1^\mathsf{T} \mathbf{A} \cdots \mathbf{P}_{t-1}^\mathsf{T} \mathbf{A}.$$

Theorem 7.1. *Let G be a d-regular Ramanujan graph on n vertices with girth g. Denote $\bar{g} = \min(g, \frac{1}{6} \cdot \log_d n)$. Then, for every $t \geq 2$ there exists a sequence of permutation matrices $\mathbf{P} = (\mathbf{P}_1, \ldots, \mathbf{P}_{t-1})$ such that*

$$\lambda(G_\mathbf{P}) \leq \left(1 + \frac{1}{t}\right)^t (t+1)d^t + 2^{-\Omega(\bar{g})} \cdot (td)^{t+3}.$$

Theorem 7.2. *Let G be a d-regular λ-spectral expander on n vertices with girth g. Denote $\bar{g} = \min(g, \frac{1}{6} \cdot \log_d n)$. Then, for every $t \geq 2$ there exists a sequence of permutation matrices $\mathbf{P} = (\mathbf{P}_1, \ldots, \mathbf{P}_{t-1})$ such that*

$$\lambda(G_\mathbf{P}) \leq \begin{cases} O(\lambda^2 d^{t-1}), & t < \frac{8\lambda^2}{d}; \\ \left(1 + \frac{1}{t}\right)^t (t+1) \cdot d^t + (t\lambda^2)^{t+4} \cdot 2^{-\Omega(\bar{g})}, & \textit{otherwise}. \end{cases}$$

We turn to prove Theorem 7.1. The proof of Theorem 7.2 is identical but for invoking Proposition 4.2 instead of Proposition 4.1.

Proof of Theorem 7.1. Let $\mathbf{A}$ be the adjacency matrix of G, and write $\chi_x(\mathbf{A}^2) = (x - a^2)p_{\mathbf{A}^2}(x)$. By Proposition 4.1,

$$\mathsf{maxroot}\left((p_{\mathbf{A}^2}(x))^{\boxtimes t}\right) \leq \left(1 + \frac{1}{t}\right)^t (t+1)d^t + 2^{-\Omega(\bar{g})} \cdot (td)^{t+3}.$$

Theorem 5.2 states that

$$\mathop{\mathbb{E}}_{\mathbf{P}_1, \ldots, \mathbf{P}_{t-1}} \chi_x\left(\Psi_\mathbf{A}(\mathbf{P}_1, \ldots, \mathbf{P}_{t-1})\right) = \left(x - a^{2t}\right) p_{\mathbf{A}^2}^{\boxtimes t}(x),$$

and so

$$\mathsf{secmaxroot}\left(\mathop{\mathbb{E}}_{\mathbf{P}_1, \ldots, \mathbf{P}_{t-1}} \chi_x\left(\Psi_\mathbf{A}(\mathbf{P}_1, \ldots, \mathbf{P}_{t-1})\right)\right) \leq \left(1 + \frac{1}{t}\right)^t (t+1)d^t + 2^{-\Omega(\bar{g})} \cdot (td)^{t+3},$$

where $\mathsf{secmaxroot}(p(x))$ denotes the second largest root of the real-rooted polynomial $p(x)$. By Theorem 6.5, the polynomials

$$\left\{\chi_x\left(\Psi_\mathbf{A}(\mathbf{P}_1, \ldots, \mathbf{P}_{t-1})\right) \mid \mathbf{P}_1, \ldots, \mathbf{P}_{t-1} \in \mathbf{S}_n\right\}$$

form an interlacing family $(T, \mathcal{P}_T)$ such that the root of T is labeled by the polynomial

$$\mathop{\mathbb{E}}_{\mathbf{P}_1,\ldots,\mathbf{P}_{t-1}} \chi_x\left(\Psi_{\mathbf{A}}(\mathbf{P}_1,\ldots,\mathbf{P}_{t-1})\right).$$

By Lemma 6.3, invoked with $j = 2$, there exists a sequence of permutations $\mathbf{P} = (\mathbf{P}_1,\ldots,\mathbf{P}_{t-1}) \in \mathbf{S}_n^{t-1}$ such that

$$\begin{aligned}\lambda_2(\Psi_{\mathbf{A}}(\mathbf{P}_1,\ldots,\mathbf{P}_{t-1})) &\le \mathsf{secmaxroot}\left(\mathop{\mathbb{E}}_{\mathbf{P}_1,\ldots,\mathbf{P}_{t-1}} \chi_x\left(\Psi_{\mathbf{A}}(\mathbf{P}_1,\ldots,\mathbf{P}_{t-1})\right)\right)\\ &\le \left(1+\frac{1}{t}\right)^t (t+1)d^t + 2^{-\Omega(\bar{g})}\cdot (td)^{t+3}.\end{aligned}$$

Observing that $\Psi_{\mathbf{A}}(\mathbf{P}_1,\ldots,\mathbf{P}_{t-1})$ is a PSD matrix, meaning in particular that

$$\lambda_n(\Psi_{\mathbf{A}}(\mathbf{P}_1,\ldots,\mathbf{P}_{t-1})) \ge 0,$$

implies that

$$\lambda(\Psi_{\mathbf{A}}(\mathbf{P}_1,\ldots,\mathbf{P}_{t-1})) = \lambda_2(\Psi_{\mathbf{A}}(\mathbf{P}_1,\ldots,\mathbf{P}_{t-1})),$$

which concludes the proof. □

References

[AKS87] Miklós Ajtai, János Komlós, and Endre Szemerédi. Deterministic simulation in logspace. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 132–140, 1987.

[BATS11] Avraham Ben-Aroya and Amnon Ta-Shma. A combinatorial construction of almost-Ramanujan graphs using the zig-zag product. *SIAM J. Comput.*, 40(2):267–290, 2011.

[CMP+22] Gil Cohen, Dor Minzer, Shir Peleg, Aaron Potechin, and Amnon Ta-Shma. Expander random walks: The general case and limitations. In *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022.

[Coh16] Michael B. Cohen. Ramanujan graphs in polynomial time. In *57th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2016*, pages 276–281. IEEE Computer Soc., Los Alamitos, CA, 2016.

[CPTS21] Gil Cohen, Noam Peri, and Amnon Ta-Shma. Expander random walks: a Fourier-analytic approach. In *STOC '21—Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1643–1655. ACM, New York, [2021] ©2021.

[Din07] Irit Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3):Art. 12, 44, 2007.

[Gil98] David Gillman. A chernoff bound for random walks on expander graphs. *SIAM Journal on Computing*, 27(4):1203–1220, 1998.

[GK21] Venkatesan Guruswami and Vinayak M. Kumar. Pseudobinomiality of the sticky random walk. In *12th Innovations in Theoretical Computer Science Conference*, volume 185 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 48, 19. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2021.

[GV22] Louis Golowich and Salil Vadhan. Pseudorandomness of expander random walks for symmetric functions and permutation branching programs. In *37th Computational Complexity Conference (CCC 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022.

[Hea08] Alexander D Healy. Randomness-efficient sampling within NC. *Computational Complexity*, 17(1):3–37, 2008.

[HJ12] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[JM21] Akhil Jalan and Dana Moshkovitz. Near-optimal Cayley expanders for abelian groups. In *41st IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science*, volume 213 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 24, 23. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2021.

[JMRW22] Fernando Granha Jeronimo, Tushant Mittal, Sourya Roy, and Avi Wigderson. Almost Ramanujan expanders from arbitrary expanders via operator amplification. *arXiv preprint arXiv:2209.07024*, 2022.

[Kah95] Nabil Kahale. Eigenvalues and expansion of regular graphs. *J. Assoc. Comput. Mach.*, 42(5):1091–1106, 1995.

[KS59] Richard V. Kadison and I. M. Singer. Extensions of pure states. *Amer. J. Math.*, 81:383–400, 1959.

[LPS88] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[McK81] Brendan D McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra and its Applications*, 40:203–216, 1981.

[MSS13] Adam Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families I: Bipartite ramanujan graphs of all degrees. In *2013 IEEE 54th Annual Symposium on Foundations of computer science*, pages 529–537. IEEE, 2013.

[MSS15] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. Interlacing families II: Mixed characteristic polynomials and the Kadison-Singer problem. *Ann. of Math. (2)*, 182(1):327–350, 2015.

[MSS18] Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families IV: Bipartite ramanujan graphs of all sizes. *SIAM Journal on Computing*, 47(6):2488–2509, 2018.

[MSS22] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. Finite free convolutions of polynomials. *Probab. Theory Related Fields*, 182(3-4):807–848, 2022.

[NS06] Alexandru Nica and Roland Speicher. *Lectures on the combinatorics of free probability*, volume 13. Cambridge University Press, 2006.

[Rei08] Omer Reingold. Undirected connectivity in log-space. *J. ACM*, 55(4):Art. 17, 24, 2008.

[RVW00] Omer Reingold, Salil Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 3–13. IEEE, 2000.

[Spi19] Daniel A Spielman. Spectral and algebraic graph theory, 2019. `http://cs-www.cs.yale.edu/homes/spielman/sagt/`.

[Tre17] Luca Trevisan. Lecture notes on graph partitioning, expanders and spectral methods. *University of California, Berkeley, https://people. eecs. berkeley. edu/luca/books/expanders-2016. pdf*, 2017.

[TS17] Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 238–251, 2017.

[Vad12] Salil P Vadhan. Pseudorandomness. *Foundations and Trends® in Theoretical Computer Science*, 7(1–3):1–336, 2012.

[Wea04] Nik Weaver. The Kadison-Singer problem in discrepancy theory. *Discrete Math.*, 278(1-3):227–239, 2004.

A Missing Proofs

For the proof Claim 3.5, we prove the following easy claim.

Claim A.1. *For every $t \geq 0$, $\mu^2(t) = \frac{\mu(\sqrt{t})+\mu(-\sqrt{t})}{2\sqrt{t}}$. In particular, if μ is symmetric around 0 then $\mu^2(t) = \frac{\mu(\sqrt{t})}{\sqrt{t}}$.*

Proof. Let X be a random variable with probability density function μ, and define the random variable $Y = X^2$. Clearly Y has probability density function μ^2. For every integer $k \geq 0$ we have that

$$\mathbb{E}[Y^k] = \mathbb{E}[X^{2k}] = \int_{-\infty}^{\infty} t^{2k}\mu(t)dt = \int_{-\infty}^{0} t^{2k}\mu(t)dt + \int_{0}^{\infty} t^{2k}\mu(t)dt.$$

By substituting variables $y = t^2$, and noting that $dy = 2tdt$, we get that

$$\mathbb{E}[Y^k] = \int_{\infty}^{0} y^k \mu(-\sqrt{y})\frac{dy}{-2\sqrt{y}} + \int_{0}^{\infty} y^k \mu(\sqrt{y})\frac{dy}{2\sqrt{y}} = \int_{0}^{\infty} y^k \frac{\mu(\sqrt{y})+\mu(-\sqrt{y})}{2\sqrt{y}}dy.$$

On the other hand, $\mathbb{E}[Y^k] = \int_0^\infty y^k\mu^2(y)dy$, and so the proof follows. □

Proof of Claim 3.5.

$$\begin{aligned}\mathcal{G}_\mu(x) &= \int_{\mathbb{R}} \frac{1}{x-t}\mu(t)dt = \int_{-\infty}^{0} \frac{1}{x-t}\mu(t)dt + \int_{0}^{\infty} \frac{1}{x-t}\mu(t)dt \\ &= \int_{0}^{\infty} \frac{1}{x+t}\mu(t)dt + \int_{0}^{\infty} \frac{1}{x-t}\mu(t)dt,\end{aligned}$$

where in the last equality we used the symmetry assumption. Thus,

$$\mathcal{G}_\mu(x) = 2x \int_0^\infty \frac{1}{x^2 - t^2}\mu(t)dt = x \int_0^\infty \frac{1}{x^2 - t^2}\frac{\mu(t)}{t}2tdt = x \int_0^\infty \frac{1}{x^2 - s}\mu^2(s)ds,$$

where in the last equality we replaced $s = t^2$ and used Claim A.1. The proof then follows as the RHS equals to $x\mathcal{G}_{\mu^2}(x^2)$. $\square$

Proof of Lemma 4.11. We can write the Tailor expansion of $f(z)$ around μ in the following way:

$$f(z) = f(\mu) + f'(\mu)(z - \mu) + \frac{f''(\mu)}{2}(z - \mu)^2 + \frac{f'''(\mu)}{6}(z - \mu)^3 + K(z)(z - \mu)^4, \quad \text{(A.1)}$$

noting that $K(z)$ as defined above can be extracted from Equation (A.1). Taking the expectation of both sides, we get

$$\mathbb{E}[f(Z)] - f(\mu) = \frac{f''(\mu)}{2}\Delta_2 + \frac{f'''(\mu)}{6}\Delta_3 + \mathbb{E}[K(Z)(Z - \mu)^4].$$

Since $(z - \mu)^4$ is a non-negative function,

$$\mathbb{E}[K(Z)(Z - \mu)^4] \geq \inf_z K(z) \cdot \mathbb{E}\left[(Z - \mu)^4\right] = \inf_z K(z) \cdot \Delta_4,$$

which concludes the claim. $\square$

B The construction of $\mathcal{T}_1$

In this section, for completeness, we sketch the construction of the tree $\mathcal{T}_1$ that is used by the proof of Theorem 6.5. Recall that $\mathcal{T}_1$ is a full rooted ordered binary tree such that with every node of the tree we associate a distribution over $\mathbf{S}_n$ such that the following holds:

1. To a leaf v that is labeled by $\mathbf{Q} \in \mathbf{S}_n$, we have that $\nu_v = \mathbf{Q}$. That is, the distribution associated with v is the distribution in which $\mathbf{Q}$ is sampled with probability 1.

2. Let v be an internal node with left son u and right son w. Then, there exist $i_v, j_v \in [n]$ and $\alpha_v \in [0, 1]$ such that

$$\begin{aligned} \nu_w &= \nu_u \mathbf{\Gamma}_{i_v, j_v}, \\ \nu_v &= \alpha_v \nu_u + (1 - \alpha_v)\nu_w. \end{aligned}$$

3. To the root r corresponds the distribution ν_r which is uniform over $\mathbf{S}_n$.

To construct $\mathcal{T}_1$ we make use of the following lemma.

Lemma B.1 (Lemma 3.5 in [MSS18]). *Let* $\mathbf{P}$ *be a random permutation matrix of order* $n \times n$. *Then, there exist random swaps* $\mathbf{T}_1, \ldots, \mathbf{T}_N$, *where* $N = N(n)$, *such that the distribution of* $\mathbf{P}$ *is equal to the distribution of* $\mathbf{T}_1 \mathbf{T}_2 \cdots \mathbf{T}_N$.

We construct $\mathcal{T}_1$ from the root downwards recursively, where the proof of correctness is by induction on the number of random swaps that participate in the distribution of a node. Recall that, for every $k \in [N]$, we can write the matrix corresponding to $\mathbf{T}_k$ by $\alpha_k \mathbf{I} + \beta_k \mathbf{\Gamma}_k$, where $\mathbf{\Gamma}_k$ is some swap matrix $\mathbf{\Gamma}_{i_k, j_k}$, $\alpha_k \in [0,1]$ and $\beta_k = 1 - \alpha_k$. We slightly abuse notation and denote the above matrix also by $\mathbf{T}_k$. For $k \in [N]$ denote $\mathbf{M}_k = \mathbf{T}_1 \cdots \mathbf{T}_k$. To the root of the tree we associate the distribution corresponding to $\mathbf{M}_N$ which, by Lemma B.1, is the uniform distribution over $\mathbf{S}_n$, establishing Property (3) above. Now,

$$
\begin{aligned}
\mathbf{M}_N &= \mathbf{M}_{N-1} \mathbf{T}_N \\
&= \mathbf{M}_{N-1} \left(\alpha_N \mathbf{I} + \beta_N \mathbf{\Gamma}_N \right) \\
&= \alpha_N \mathbf{M}_{N-1} + \beta_N \mathbf{M}_{N-1} \mathbf{\Gamma}_N.
\end{aligned}
$$

The sub-tree that is rooted at the left son of the root is constructed in a recursive manner so that the distribution associated with this son corresponds to $\mathbf{M}_{N-1}$. To the right son corresponds the distribution $\mathbf{M}_{N-1}\mathbf{\Gamma}_N$, and so Property (2) above holds for the root and its two sons.

To the right son we associate the distribution that corresponds to $\mathbf{M}_{N-1}\mathbf{\Gamma}_N$ which is not of the form suitable for recursion. That is, while the number of random swaps has decreased, it is not immediately clear how to proceed as before as the rightmost term is not of the form $\alpha \mathbf{I} + \beta \mathbf{\Gamma}$. To remedy this, observe that

$$
\begin{aligned}
\mathbf{M}_{N-1}\mathbf{\Gamma}_N &= \mathbf{M}_{N-2} \mathbf{T}_{N-1} \mathbf{\Gamma}_N \\
&= \mathbf{M}_{N-2} \left(\alpha_{N-1} \mathbf{I} + \beta_{N-1} \mathbf{\Gamma}_{N-1} \right) \mathbf{\Gamma}_N \\
&= \mathbf{M}_{N-2} \left(\alpha_{N-1} \mathbf{\Gamma}_N + \beta_{N-1} \mathbf{\Gamma}_{N-1} \mathbf{\Gamma}_N \right) \\
&= \mathbf{M}_{N-2} \mathbf{\Gamma}_N \left(\alpha_{N-1} \mathbf{I} + \beta_{N-1} \mathbf{\Gamma}_N^{-1} \mathbf{\Gamma}_{N-1} \mathbf{\Gamma}_N \right).
\end{aligned}
$$

Note that $\mathbf{\Gamma}_N^{-1}\mathbf{\Gamma}_{N-1}\mathbf{\Gamma}_N$ is also a swap matrix. Proceeding in the manner, we can "push" $\mathbf{\Gamma}_N$ all the way so that it appears to the left of $\mathbf{M}_{N-2}$, making the necessary conjugation to all the swaps in the product. This results with $\mathbf{M}_{N-1}\mathbf{\Gamma}_N = \mathbf{\Gamma}_N \widehat{\mathbf{M}}_{N-1}$ for $\widehat{\mathbf{M}}_{N-1}$ which is the product of at most $N-1$ random swaps. As the number of random swaps has decreased (note that $\mathbf{\Gamma}_N$ to the left is a deterministic permutation), by recursion, we can construct the sub-tree rooted at the right son.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il