Eigenvalue Fluctuations of Random Matrices beyond the Gaussian Universality Class

Elliot Paquette

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Ioana Dumitriu, Chair

Christopher Hoffman

Soumik Pal

Program Authorized to Offer Degree:
Mathematics

University of Washington

Abstract

Eigenvalue Fluctuations of Random Matrices beyond the Gaussian Universality Class

Elliot Paquette

Chair of the Supervisory Committee:
Associate Professor Ioana Dumitriu
Mathematics

The goal of this thesis is to develop one of the threads of what is known in random matrix theory as universality, which essentially is that a large class of matrices generalizing the Gaussian matrices (certain Wigner matrices and β-ensembles) show the same limiting behavior as the Gaussian ensembles. The more well known thread is the universality of local eigenvalue statistics, which is most directly tied to the physical roots of the theory. The other thread is the universality of the fluctuations of global eigenvalue statistics, which occurs in the same classes of matrices that show the local universality. We focus on this second type of universality, and we analyze in detail some features of the limiting Gaussian process that arises this way. Beyond that, the mathematical contributions of this thesis are in three different models of random matrices, twice visiting linear statistics and once visiting the correct order scaling of the spectrum.

In all cases, the ensembles studied do not fall into a matrix class covered by current, broad universality theorems. The closest of these models to the Gaussian class is the β-Jacobi ensemble. It lies firmly in the environment of classical random matrix theory in that it can be defined by a log-gas with potential $V(x) = -(p-1)\log(x) - (q-1)\log(1-x)$ for some p and q. From the standpoint of Johansson's results, the interest

here is to see if the singular constraining potential is strong enough to disrupt the global fluctuations. We see that this is not the case, and the same formula for CLTs of linear statistics holds. We additionally find the limiting level density, and we find the first order correction to the limiting level density, which is also obtained by Johansson.

The next of these models is the adjacency matrix of a *permutation model* regular graph. This matrix can be defined by sampling independently and uniformly permutation matrices $P_1, P_2, \ldots, P_d$ and defining

$$\mathcal{P}_{n,d} := P_1 + P_1^t + P_2 + P_2^t + \cdots + P_d + P_d^t.$$

For this matrix we will show how to derive a uniform bound on the eigenvalues that holds for all n and d and show how this bound can be used in conjunction with estimates on the number of non-backtracking walks to derive the law of the global fluctuations. In this setting, the dependence of d on n is seen to govern whether or not the d-regular graph has Gaussian-type global fluctuations. If $d \to \infty$ slowly as $n \to \infty$, then the fluctuations are like the GOE. On the other hand, if d remains fixed, the fluctuations are a Poissonian analogue.

Finally, we will investigate the normalized Laplacian $\mathscr{L}$ of the Erdős–Rényi graph model, a graph on n vertices with edges included independently and with probability $p = p(n)$. The behavior of this object depends strongly on whether or not the degrees are strongly concentrated around their means. Indeed, the most interesting features of this graph from the spectral point of view arise precisely when the degrees stop being strongly concentrated. When all the degrees are concentrated, (the $p = \Omega(\log n/n)$ regime), then the nontrivial eigenvalues of $\mathscr{L}$ are within a $1/\sqrt{np}$ window of 1, consistent with GOE predictions. Outside of this regime, the answer is less clear. If p is exactly $\log n/n$, the graph has isolated vertices with probability tending to $1/e$. If the graph has isolated vertices, then $\mathscr{L}$ has eigenvalue 0 with multiplicity greater than 1, violating GOE predictions. This thesis will estimate the nontrivial eigenvalues of $\mathscr{L}$

as p descends to $\log n/n$ and attempt to determine their order.

Each of these problems has additional motivations outside of the scope of exploring the Gaussian universality class. Each problem will be presented with its own history and motivation in addition to its contribution to understanding the broader picture.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

This thesis was largely written over a two month period in early 2013, starting from previously completed collaborative work. Once I finally began writing it, I found – much to my surprise – that the thesis almost wrote itself; the lessons and experiences of the past five years at UW and before were enough to propel me through to the finish.

Chris Hoffman taught my first probability course in 2008. I started with such a poor understanding of the subject I thought that independent meant disjoint. Chris got me on my feet, and then through the course of reading ergodic theory with him and then working with him, he has continually been a fount of intuition and wisdom without which I would be so much poorer.

Summer 2009, I was preparing for my linear analysis prelim exam, and I wouldn't have wanted to learn all those problems with Huy Vo Tran. Huy and I, as best I remember, did every problem from the history of the exam. I have no idea how we got that momentum, but we passed it together, and he has been a great mathematical resource ever since.

Soumik Pal taught me my second probability course, the other kind of probability, stochastic analysis. Even though this thesis has exactly zero such tools in it, I'm using it this very day on other projects, and I wouldn't be in the position to do that without his tutelage. Even before the course, Soumik was my greatest resource for the analytic side of probability, which finds its way much of the work in this thesis.

My academic brother Toby Johnson and I started working together a year and some back, and I only wish we had worked more together sooner. I'm constantly embarrassed by the number of mistakes I make that Toby catches, because I'm pretty certain it doesn't go the other way around. We have collaborated since the work in this thesis was completed, and I hope we continue for a long time.

Matt Kahle has been a constant source of inspiration to me. More than anyone I know, he's packed with incredible conjectures about the world that are instantaneously recognizable as his, and they inspire and impel me to more and greater things.

Jim Morrow and I started bicycling together in the Summer 2011, and then I lived with him for about two years subsequently. Living with him enriched my life in ways I'll never be able to list or even understand, but that list includes items ranging from the joys of pressure cooking to conformal mapping to marital advice. He has been a constant source of wisdom and energy, advising me in a way that only he can.

Finally, I would be remiss if I didn't thank Ioana, my adviser. I'm pretty certain there is nothing I can write here to adequately do justice to my relationship with Ioana. Ioana suggested to me my first project, and that turned into a great odyssey. On that journey, I experienced the entirety of human emotion, but with Ioana as my captain, we made it back alive. I'm all the wiser and all the smarter for having done it, and I would do it again.

There are also people whose effect on me is profound and ineffable, but without whom I could not be where I am today: my mother and father, Courtney, and a long list of friends here in Seattle and in places all around the world.

DEDICATION

This thesis is dedicated to all those who study. May you enlighten the world.

Chapter 1

INTRODUCTION

Perhaps the central effort of modern random matrix theory is to prove what is called *universality*. The precise meaning of the term varies from paper to paper, but the philosophical underpinnings are the same: the limiting eigenvalue behaviors of many models of random matrices are the same as a matrix of independent Gaussians.

So many diverse matrix models display this behavior that it is the expectation, rather than the exception, that a given matrix model should behave in this way. This holds even in situations in which the matrix model is not a generalization of Gaussian matrices.

This thesis is concerned with studying the extent of this intuition in three situations that, in varying degrees and capacities, significantly differ from Gaussian matrices. Each displays the degree to which these universal laws are sensitive or immune to the varying alterations.

To contextualize the contributions of this thesis properly, it is necessary to first understand the eigenvalue theory of Gaussian matrices and the matrix models that generalize them. Further, to understand why the mathematical theory of universality has developed, it is necessary to understand *how* this eigenvalue theory of Gaussian matrices came to be.

1.1 The Eigenvalues of Self-Adjoint Gaussian matrices and Heavy Nuclei

In the early history of random matrices, there are two threads along which the theory developed. The first begins with Wishart [126] coming from the statistical side, and the second begins with Wigner [124, 125] coming from the physical side. The eigenvalues of self-adjoint Gaussian matrices arise through the physical connection.

Wigner and Dyson [36] viewed large self-adjoint Gaussian matrices as a first approximation to the Schrödinger operator of a highly disordered nucleus. In essence, a large nucleus is so chaotic that laws of nuclear interaction might be adequately replaced by a typical sampling from appropriately chosen random laws of interaction. This amounts to placing a distribution on self-adjoint operators and asking if the energy levels of these self-adjoint operators bear any resemblance to the measurable energy levels of nuclei.

To study the eigenvalues of a random self-adjoint, infinite-dimensional operator, the approach taken by these researchers was to take a large random self-adjoint matrix, appropriately rescale, and take a limit. For their computational convenience, this is the way that Gaussian matrices came to be studied.

The most easily defined such matrix is the Gaussian Orthogonal Ensemble (GOE). It is defined by taking the symmetric part of an $n \times n$ matrix $Z = (Z_{i,j})$ whose entries are independent standard normals; thus $G_1 := \frac{Z+Z^t}{2}$. It is an easy calculation to see that the eigenvalues of this matrix are not at unit order. To wit, the square-Frobenius norm of the matrix is given by

$$\|G_1\|_F^2 = \sum_{i>j} \frac{(Z_{i,j} + Z_{j,i})^2}{2} + \sum_i Z_{i,i}^2,$$

which has expectation n^2. As there are n eigenvalues, the proper scaling of the matrix to put a typical eigenvalue at unit order is $\frac{1}{\sqrt{n}}$.

The law of the GOE is absolutely continuous with respect to Lebesgue measure on $\mathbb{M}_{n,SA}(\mathbb{R})$, the space of $n \times n$ self-adjoint matrices over $\mathbb{R}$. Thus it possible to write the density of the ensemble with respect to this measure. From the joint independence of the entries,

$$p_{G_1}(M) := \frac{1}{Z} e^{-\sum_{i>j} m_{i,j}^2 + \sum_i m_{i,i}^2/2} = \frac{1}{Z} e^{-\|M\|_F^2/2}. \tag{1.1}$$

In this form, and from the orthogonal invariance of the Frobenius norm, it follows that the eigenvectors of G_1 are uniformly distributed over the orthogonal group and are independent of the eigenvalues. Furthermore, it is possible to compute the explicit Jacobian of eigenvalue factorization (see [92] or [33]) to conclude that

$$p_{G_1}(Q\Lambda Q^t) = \frac{1}{Z} e^{-\sum_{i=1}^{n} \lambda_i^2/2} \prod_{i>j} |\lambda_i - \lambda_j| \,. \tag{1.2}$$

Besides the GOE, there are also two other canonical Gaussian matrices, the Gaussian Unitary Ensemble (GUE) and the Gaussian Symplectic Ensemble (GSE). The GUE can be defined by filling an $n \times n$ matrix Z with i.i.d. standard, complex normals,[1] and taking $G_2 := \frac{Z+Z^*}{2}$. Likewise, the GSE can be defined by filling a matrix Z with i.i.d. standard, quaternion normals [2], and taking $G_4 := \frac{Z+Z^A}{2}$. Both of these matrix models show the same decoupling of eigenvalues and eigenvectors, with the eigenvectors being uniformly distributed over the appropriate compact group (Orthogonal, Unitary or Symplectic). The eigenvalue densities are concisely summarized as

$$p_\beta^{\frac{x^2}{2}}(\lambda_1, \ldots, \lambda_n) = \frac{1}{Z_\beta} e^{-\sum_{i=1}^{n} \beta\lambda_i^2/2} \prod_{i>j} |\lambda_i - \lambda_j|^\beta \,, \tag{1.3}$$

where $\beta = 1, 2,$ or 4 corresponds to the GOE, GUE, or GSE respectively.

The existence of the joint eigenvalue densities opens one line of approach to studying these matrix models, by analytically manipulating the expression and taking appropriate limits. Similarly, it is possible to realize an explicit connection between these matrix models and orthogonal polynomial theory.

Regardless, the first approach considered was to work with these matrices through their entrywise distributions. This was the approach taken by Wigner in proving his semicircle law [124]. In fact, he first proved the semicircle law for random $\{-1, +1\}$-entry matrices, and later published a note that the proof extends to a larger class of matrices including the GOE [125]. His proof would also show that the semicircle law

[1] A standard complex normal has real and imaginary parts that are independent centered normals with variance $1/2$.

[2] Similarily, a standard quaternion normal has $1, i, j,$ and k parts that are independent centered normals with variance $1/4$.

holds for the GUE and GSEs although he does not explicitly state it. Precisely, he does show that for each positive integer k,

$$\frac{1}{n}\mathbb{E}\,\mathrm{tr}\left(\frac{G_1}{\sqrt{2n}}\right)^k \to \frac{2}{\pi}\int_{-1}^{1} x^k\sqrt{1-x^2}\; dx, \tag{1.4}$$

as $n \to \infty$. From this statement, it follows from an approximation argument that the expected eigenvalue counting function converges to the distribution function of the semicircle law, i.e.

$$\mathbb{E}\frac{\left|\{\lambda_i(G_1) \leq t\sqrt{2n}\}\right|}{n} \to \frac{2}{\pi}\int_{-1}^{t} \sqrt{(1-x^2)_+}\; dx,$$

for every fixed t.

The more usual mathematical formulation of this statement is in terms of weak convergence. For a matrix $M \in \mathbb{M}_{n,SA}(\mathbb{A})$, with $\mathbb{A} = \mathbb{R}, \mathbb{C}$, or $\mathbb{J}$, let $\lambda_1(M) \leq \ldots \leq \lambda_n(M)$, denote the ordered eigenvalues of these matrices. Define the *empirical spectral measure* $\hat{\mu}_M$ to be the random measure

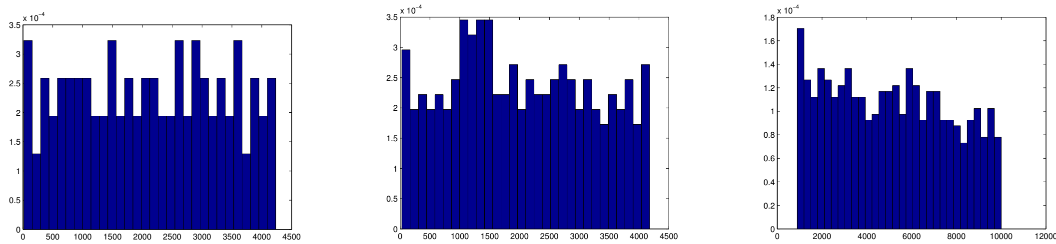
$$\hat{\mu}_M(x) := \sum_{i=1}^{n} \delta_{\lambda_i(M)}(x),$$

and define the semicircle law to be

$$d\mu_{\mathrm{sc}}(x) := \frac{2}{\pi}\left(\sqrt{1-x^2}\right)_+ dx.$$

Finally, define the 1-point correlation function $R_M^{(1)}$ to be the deterministic measure induced by the linear functional $f \mapsto \mathbb{E}\int_{\mathbb{R}} f(x)\; d\hat{\mu}_(x)$. Wigner's moment calculation (1.4) shows that $\frac{1}{n}R_{G_1}^{(1)} \Rightarrow \mu_{\mathrm{sc}}$.

The global semicircle law does not capture much of physical reality. This is to say, the distribution of energy levels of neutron resonances from heavy nuclei is not semicircular (see Figure 1.1), and so it cannot be said that a typical energy level of a heavy nucleus is well-modeled by a uniformly sampled eigenvalue of a large self-adjoint Gaussian matrix. Instead, a nuclear energy level sequence looks something like a small chunk of the eigenvalues extracted from somewhere in the inside of the



(a) Er166 spectral sequence, eV; from [80]
(b) U236 spectral sequence, eV; from [86]
(c) U238 spectral sequence, eV; from [96]

Figure 1.1: ESDs from three neutron resonance experiments.

spectrum; otherwise stated, the scaling in the semicircle law is very different from the one required for physicality.

To understand the sense in which the GOE does shed light on the behavior of nuclear energy levels, we must scale the matrices so the mean eigenvalue spacing is 1 (known as the *microscopic* scaling). Further, we will formalize this scaling by looking at the *k-point correlation functions*.

For any random $M \in \mathbb{M}_{n,SA}(r)(\mathbb{A})$ with ($\mathbb{A} = \mathbb{R}, \mathbb{C}$, or $\mathbb{J}$) whose law is absolutely continuous to Lebesgue measure and any fixed natural number k, define the function $R_M^{(k)} : \mathbb{R}^k \to \mathbb{R}^+$ which is the unique symmetric function so that

$$\int_{\mathbb{R}^k} F(x_1, \ldots, x_k) R_M^{(k)} dx_1 \cdots dx_k = k!\mathbb{E} \sum_{1 \le i_1 < \cdots < i_k \le n} F(\lambda_{i_1}(M), \ldots, \lambda_{i_k}(M))$$

for any symmetric continuous, compactly supported function $F : \mathbb{R}^k \to \mathbb{R}$ (see [116])[3]. Of particular note, the n-point correlation function is a multiple of the joint probability density function for the ordered eigenvalues. Also, the 1-point correlation function expresses the ensemble level density, i.e.

$$\mathbb{E}N_I(M) = \int_I R_M^{(1)}(x)dx,$$

[3]These are the *unnormalized* k-point correlation functions. Note that for matrices with entries singular to Lebesgue measure, R would be a measure.

where $N_I(M)$ is the number of eigenvalues in I.

The first successful calculation of these correlation functions was for the GUE, due to Gaudin and Mehta [54, 91]. Similar calculations followed for the GOE and GSE. The approach to these calculations is through the theory of determinantal point processes. Explicitly, one has that

$$R_{G_2}^{(k)} = \det(K_n(x_i, x_j))_{1\leq i,j\leq k}, \tag{1.5}$$

where K_n is given by $K_n(x,y) = \sum_{k=0}^n \psi_k(x)\psi_k(y)$ and where

$$\psi_k(x) = (-1)^n(2^n n!\sqrt{\pi})^{-1/2} e^{x^2/2}\frac{d^n}{dx^n}e^{-x^2}$$

is the Hermite function. In this normalization, the Hermite functions are orthonormal on $\mathbb{R}$. Equation (1.5) is easily established by comparing with the explicit eigenvalue density (1.3) and using the multilinearity of the determinant (see Chapter 5 of [92] for details).

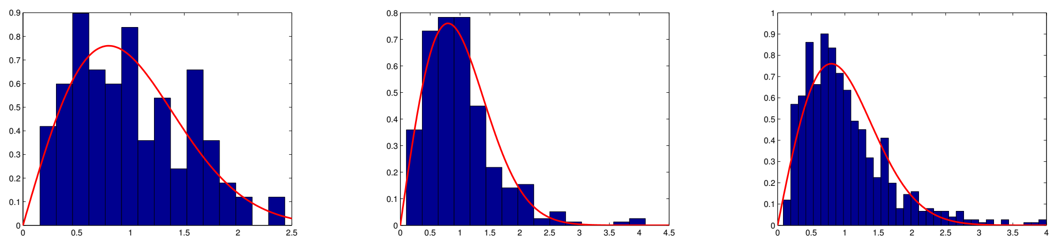
If we scale the eigenvalues so that the mean eigenvalue spacing is 1, then we find

$$\frac{\pi t}{\sqrt{2n}} K_n\left(\frac{\pi t\xi}{\sqrt{2n}}, \frac{\pi t\eta}{\sqrt{2n}}\right) \to \frac{\sin(\pi t(\xi-\eta))}{\pi(\xi-\eta)},$$

with the convergence uniform on compact sets of $|\eta| < 1$ and $|\xi| < 1$ [92, 3].[4] This in turn allows one to deduce exact expressions for the distribution of the number of eigenvalues in an interval whose length is on the same order as the eigenvalue spacing. Similar, albeit more complicated, expressions exist for the correlation functions of the GOE and GSE.

We will recast the study of correlation functions in terms of *linear statistics*, which will additionally allow us to formalize the notion of fluctuations about the semicircle

[4]This amounts to looking at the spectrum in a neighborhood of 0. It is also possible to look at the spectrum in the neighborhood of any point strictly within the support of the semicircle law, for which the same limit is seen (see, e.g. [116]). This lends further credence to the ansatz that a sequence of nuclear energy levels looks like a small block of contiguous GOE eigenvalues, as the choice of block is unimportant. That said, in a neighborhood of the edge of the semicircle law, the microscopic behavior is much different.



(a) Normalized Er166 spacings; from [80]

(b) Normalized U236 spacings; from [86]

(c) Normalized U238 spacings; from [96]

Figure 1.2: Spacing distributions from the same three spectral sequences, with superimposed GOE spacing distribution. Two serious complications arise in comparing actual nuclear data to RMT predictions. First, the sequences are not always *complete*, meaning they miss some resonances. Secondly, there are so-called s-wave resonances and p-wave resonances, corresponding to neutrons with different angular momenta. Part of *Wigner's surmise* (see [92] for a discussion) is that the s and p sequences are independent, so a polluted sequence of s-wave data can be expected to have more compressed spacings.

law. A linear statistic is any random variable of the form

$$\int_{\mathbb{R}} \phi(x) d\hat{\mu}_{G_1}(x)$$

for a fixed test function ϕ. The study of linear statistics is formally equivalent to the study of correlation functions. Recall that

$$\mathbb{E} \int_{\mathbb{R}} \phi(x) d\hat{\mu}_{G_1}(x) = \int_{\mathbb{R}} \phi(x) R_M^{(1)}(x) dx.$$

Thus, to know the first moments of all linear statistics of M is to know the first correlation function of M and vice versa. The same statement holds in general: to know the first k moments of all linear statistics of M is to know the first k correlation functions and vice versa.

For example, for a fixed test function ψ, one has

$$\mathbb{E}\left(\int_{\mathbb{R}} \psi(x) d\hat{\mu}_M(x)\right)^2 = 2\sum_{i<j} \mathbb{E}\psi(\lambda_i)\psi(\lambda_j) + \sum_i \mathbb{E}\psi(\lambda_i)^2$$
$$= \int_{\mathbb{R}^2} \psi(x)\psi(y) R_M^{(2)}(x,y)\ dxdy + \int_{\mathbb{R}} \psi(x)^2 R_M^{(1)}\ dx,$$

and similar relations hold for all higher k.

The study of linear statistics also allows us to study the fluctuations of the empirical spectral measure about the semicircle law. Generally speaking this means one would like to analyze the random signed measure $\hat{\mu}_{G_1}(x) - R_{G_1}^{(1)}(x)$, as $n \to \infty$. To formalize this study, one needs to give a collection of test functions on which this signed measure is to act. Depending on the class of test functions chosen, a different type of limit arises. Regardless the class of function ϕ, we will denote the centered linear statistic as

$$X_\phi^{G_1} := \int_{\mathbb{R}} \phi(x) \left[d\hat{\mu}_{G_1}(x) - dR_{G_1}^{(1)}(x)\right] = \operatorname{tr}\left(\phi(G_1)\right) - \mathbb{E}\operatorname{tr}\left(\phi(G_1)\right).$$

The most technically simple test functions to study are polynomials, as they can be understood combinatorially. For the GOE, one sees that

$$X_\phi^{G_1/\sqrt{2n}} \Rightarrow N(0, V_\phi),$$

where V_ϕ is given by

$$V_\phi = \frac{1}{2\pi^2} \oint_{\substack{|z|=1 \\ \Im z>0}} \oint_{\substack{|w|=1 \\ \Im w>0}} \phi'(\Re z)\phi'(\Re w) \log\left|\frac{1-zw}{1-z\bar{w}}\right| \frac{\Im z}{z}\frac{\Im w}{w} dzdw, \tag{1.6}$$

see [3] and [18] for the particular variance form. Once the weak convergence is established for polynomial test functions, it is possible to use concentration inequalities to show that the same theorem holds for C^1 test functions (see e.g. [3] and Section 2.4).

One of the most interesting features of this theorem are the scalings involved. The $1/\sqrt{n}$ scaling of the G_1 matrix puts the eigenvalues at unit order. Yet, no scaling of the linear statistic is required to produce a limit theorem. Recent results of O'Rourke [97] in the case of the GOE/GSE and Gustavsson [61] in the case of the GUE show that the variance of the individual eigenvalues at this scaling decays like $(\log n)/n$. Thus, the fact that no scaling is required shows that the eigenvalue correlations are sufficiently long range to cancel the logarithmic term.

From the point of view of studying fluctuations from the semicircle law, perhaps the most sensible choice of test functions is some class of indicator functions. Consider the case that $\phi = \mathbf{1}\{[y, y+L]\}$ for some fixed $y \in (-1,1)$ and some $L = L(n) > 0$ nonincreasing so that $y + L \in (-1,1)$.

In the case of *macroscopic* statistics, i.e. where $\operatorname{Var} X_\phi^{G_1/\sqrt{2n}} \to \infty$, it is a theorem of Costin and Lebowitz [31] that the standardized linear statistic converges to a normal. At the *microscopic* scaling, i.e. where $L = r/n$, the limiting distribution can be understood by the connection to the microscopically scaled k-point correlation functions.

The great validation for all these computations is that experimental data show that in the microscopic regime, nuclear data has statistics that follow GOE behavior. In particular, many heavy nuclei show neutron resonance spectra that agree with the GOE spacing (see Chapter 1 of [92]). Further, the variance of X_ϕ^M with M coming from the *nuclear data ensemble* (NDE)[5] agrees strongly with that of the GOE, and microscopically scaled k-point correlation functions match for the GOE and NDE for $k = 2, 3, 4$ [16].

1.2 Universality of the Gaussian Ensembles

Stepping back, the result of all this work is that a model chosen essentially for its computational convenience (the GOE/GUE/GSE) appears to have properly modeled some features of a purely physical phenomenon at certain scalings. The model itself,

[5]This is data from many well-studied heavy nuclei, the energy levels of which are considered as coming from a single data set. These energy levels are considered as a single sampling of a point process to be compared with the eigenvalues of the GOE. See [16] for more. More recent discussions of the validity of the dataset are also in print, see both [75, 107].

though, raises some serious mathematical questions: these matrices were filled with i.i.d. standard normals, which did not have any *a priori* justification. Nonetheless, the resulting models in large n seem to say something about the universe. Either it was a miracle that we chose just the right model, or some of the arbitrary modeling choices did not matter.

This latter alternative is the impetus for *universality*: informally, can one find ensembles of random matrices that generalize the Gaussian ensembles and produce the same limiting features? Because of their physicality, the most pressing question is: what ensembles of random matrices produce the same microscopically scaled k-point correlation functions?

One way to generalize the Gaussian ensembles is simply to let the entry distributions vary. This leads to *Wigner* matrices. In Wigner's second paper [125], he notes that the semicircle law did not need the Gaussian assumption on the entries. It would be enough that the entries of the matrix are independent, centered, and have variance agreeing with the Gaussian matrices.[6] Such matrices are called called Wigner matrices. A more specific version of the universality question for Wigner matrices is the Wigner-Dyson-Mehta conjecture. Loosely, it states that the microscopic (or *local*) statistics of the eigenvalues should be independent of the entry distributions of the matrix, subject to their mutual independence and standardized variance. Usually, this is formulated as showing that the k-point correlation functions scaled so that the mean eigenvalue spacing is 1 converges to the same as limit as does the corresponding self-adjoint Gaussian matrix (see [116, 39] for more on this conjecture and the great strides made towards its resolution).

The subtler assumption to remove is the independence of the entries. After all, if the entries are not chosen to be independent, what dependency structure *should* they

[6] He make some additional assumptions for his proof, which can be removed. Beyond independence, centering and identical variance, it is necessary to assume an additional uniform integrability condition on the entries; the lindeberg condition suffices.

have? The answer lies in the symmetry classes of the matrix. The GOE/GUE/GSE can be uniquely defined by asking that the entries of the matrices have (i) standardized variance, (ii) independent entries, and (iii) orthogonal/unitary/symplectic invariance. Furthermore, this symmetry invariance was the abstract motivation for Wigner's original program, whose intuition was that the Hamiltonian governing these systems had no other structure besides its symmetry class [123].

Recall that the Gaussian matrices could be defined by their density with respect to Lebesgue measure on $\mathbb{M}_{n,SA}(\mathbb{A})$ by

$$p_\beta(M) = \frac{1}{Z_\beta} e^{-\beta\|M\|_F^2/2}, \tag{1.7}$$

with $\beta = 1, 2, 4$ corresponding to $\mathbb{A} = \mathbb{R}, \mathbb{C}, \mathbb{J}$. If the quadratic term in the exponent,

$$\|M\|_F^2 = \sum_{i=1}^n |\lambda_i|^2$$

were replaced by an alternative, group-invariant function V, one could produce a whole family of orthogonally, unitarily, or symplectically invariant random matrix ensembles. As these are self-adjoint matrices, this invariance criterion forces

$$V(M) = \sum_{i=1}^n V(\lambda_i),$$

and leads to the eigenvalue density

$$p_\beta^V(\lambda_1, \ldots, \lambda_n) = \frac{1}{Z_{V,\beta}} \exp\left(-\beta \sum_{i=1}^n V(\lambda_i)\right) \prod_{i<j} |\lambda_i - \lambda_j|^\beta, \tag{1.8}$$

by the Jacobian formula for eigenvalue factorization. These are called *β-ensembles* or *log-gases* [44].

The universality of this ensemble would now mean that the local statistics of the ensemble are independent of the choice of constraining potential, subject to sanity conditions such as the integrability of the density. This problem has been studied in greater and greater generality, with many techniques being applied, to the point that the universality of the k-point correlation functions is known for any β and a large class of real analytic potentials V [20].

1.3 Global Fluctuations and the Extent of the Gaussian Universality Class

It is tempting to say that universality is a phenomenon of local eigenvalue statistics, so that microscopic statistics are universal, and macroscopic statistics (such as the limiting level density) are not. This, however, disregards some of the more amazing universal features of centered global statistics of random matrices, specifically those relating to fluctuations of the empirical spectral measure about the semicircle law.

One of the key contributions in this direction is work by Johansson on linear statistics of smooth test functions of β-ensembles [68]. He begins by scaling the potentials so that the eigenvalues are at unit order; thus the actual potentials in consideration are $nV(x)$. First, he shows that for essentially all V, the empirical spectral measure converges as $n \to \infty$ to the unique minimizer μ of the energy functional

$$I_V[\mu] = \int_{\mathbb{R}} \left(\log |t-s|^{-1} + \frac{1}{2}V(t) + \frac{1}{2}V(s) \right) d\mu(t) d\mu(s).$$

This is the analogue of the Wigner semicircle law, and in fact this portion of this study could be guessed from classical potential theory. More interestingly, Johansson is able to study the global fluctuations for a large class of polynomial potentials V.

Let $\mathscr{V}$ denote the set of polynomials of even degree with positive leading coefficient that are convex.[7] Johansson shows that the limiting eigenvalue distribution is supported on an interval $[a, b]$, and he shows that smooth test functions exhibit a central limit theorem (CLT). With $V \in \mathscr{V}$ and M_V drawn from the β-ensemble with potential V,

$$X_\phi^{(2M_V-(a+b))/(b-a)} \Rightarrow N(0, \tfrac{1}{\beta}V_\phi),$$

where V_ϕ (see (1.6)) is the same variance as seen in the GOE. Thus, if the eigenvalues are standardized to lie on an interval $[-1, 1]$, their fluctuations are once again inde-

[7] Johansson works in somewhat greater generality than this, but we will present the theorems just for this case, which affords a simpler exposition.

pendent of the matrix model. This is an exact way in which the fluctations for the Gaussian matrices generalize to log-gas models.

The generalization to Wigner matrices produces the same statement: with the eigenvalues linearly transformed to lie asymptotically on $[-1, 1]$, the linear statistics of smooth test functions ϕ show a CLT with limiting variance $\frac{1}{\beta}V_\phi$ and with $\beta = 1, 2, 4$ corresponding to the appropriate symmetry class (see [3]).

Thus the global fluctuations, up to standardizing the support of the eigenvalues, appear so far to go hand-in-hand with the microscopic fluctuations of the eigenvalues. This is not to say that the CLT for linear statistics implies or is implied by the scaling limits at the microscopic level; it would be very interesting to find random matrix models that broke this symmetry. Further, there is a limitation to this universality in the sense that it presumes a single interval of support for the limiting eigenvalue distribution. In the case that there are multiple intervals of support, some other fluctuation expressions arise (see for example [99]). Nonetheless, global fluctuations give one tool for the study of the extent of the Gaussian universality class.

1.4 Contributions of this Thesis

This thesis will cover three different models of random matrices, twice visiting linear statistics and once visiting the correct order scaling of the spectrum. The closest of these models to the Gaussian class is the β-Jacobi ensemble. It lies firmly in the environment of classical random matrix theory in that it can be defined by a log-gas with potential $V(x) = -(p-1)\log(x) - (q-1)\log(1-x)$ for some p and q. From the standpoint of Johansson's results, the interest here is to see if the singular constraining potential is strong enough to disrupt the global fluctuations. We see that this is not the case, and the same formula for CLTs of linear statistics holds. We additionally find the limiting level density, and we find the first order correction to the limiting level density (referred to as the deviation, which is also obtained by Johansson for $V \in \mathscr{V}$ [68]).

The next of these models comes from an alternate mathematical universe, with completely different initial motivations. It is the adjacency matrix of a *permutation model* regular graph. This matrix can be defined by sampling independently and uniformly permutation matrices $P_1, P_2, \ldots, P_d$ and defining

$$\mathcal{P}_{n,d} := P_1 + P_1^t + P_2 + P_2^t + \cdots + P_d + P_d^t.$$

For this matrix we will show how to derive a uniform bound on the eigenvalues that holds for all n and d and show how this bound can be used in conjunction with estimates on the number of non-backtracking walks to derive the law of the global fluctuations. In this setting, the dependence of $d = d(n)$ is seen to govern whether or not the d-regular graph has Gaussian-type global fluctuations. If $d \to \infty$ slowly as $n \to \infty$, then the fluctuations are like the GOE. On the other hand, if d remains fixed, the fluctuations are a Poissonian analogue.

Finally, we will investigate the Erdős–Rényi graph model, i.e. a graph on n vertices with edges included independently and with probability $p = p(n)$. Beyond the adjacency matrix for this graph, we will investigate its normalized Laplacian. Letting T be the diagonal matrix of degrees and letting A be the adjacency matrix, the normalized Laplacian is defined as

$$\mathscr{L} = I - T^{-1/2} A T^{-1/2},$$

provided all degrees are positive. If some degrees are 0, the Laplacian is defined for the subgraph formed by removing the isolated vertices, and extending it to be 0 on the isolated vertices.

If the vertices of the graph are labelled $v_1, v_2, \ldots, v_n$, the matrix entries are

$$\mathscr{L}_{i,j} = \delta_{i,j} \mathbf{1}\left\{\deg v_i > 0\right\} - \frac{\mathbf{1}_{v_i \leftrightarrow v_j}}{\sqrt{\deg v_i \deg v_j}}.$$

Note that if $\deg v_i = \deg v_j = 0$, then we take $\mathscr{L}_{i,j} = 0$.

In both of these graph models, it is already a question of what the correct scaling should be to put a typical eigenvalue at unit order. Both matrix models have some trivial eigenvalues that need to be disregarded for this to make sense, namely $\mathcal{P}_{n,d}$ has the eigenvalue $2d$ with eigenvector $\mathbf{1}$, the all 1 vector. Likewise $\mathscr{L}$ has the eigenvalue 0 and eigenvector which is $T^{1/2}\mathbf{1}$. For the remainder of the spectrum, the order of a typical eigenvalue can be suggested by once again investigating the Frobenius norm. For the d-regular graph matrix, one has that $(\mathcal{P}_{n,d})_{i,j} \geq (P_1)_{i,j} + \cdots + (P_d)_{i,j}$, and $\mathbb{E}\left(\mathcal{P}_{n,d}\right)_{i,j}^2 \leq 2\mathbb{E}\left((P_1)_{i,j} + \cdots + (P_d)_{i,j}\right)^2$, from convexity and the symmetry of the entries. Noting that $\mathbb{E}(P_1)_{i,j} = 1/n$, then one has

$$\frac{d}{n} \leq \mathbb{E}\left(\mathcal{P}_{n,d}\right)_{i,j}^2 \leq 2\frac{d}{n} + 2\frac{d^2}{n^2}.$$

Thus provided that $d = o(n)$, the correct order of $\mathbb{E}\|\mathcal{P}_{n,d}\|^2$ is nd, so that one should expect a typical eigenvalue to be about $\sqrt{d}$.

The same reasoning can be applied to the normalized Laplacian of an Erdős–Rényi graph, although it reveals the greater complexities of that situation. If the vertices of the graph are labelled $v_1, v_2, \ldots, v_n$, then one has

$$\left\|T^{-1/2}AT^{-1/2}\right\|_F^2 = \sum_{i,j} \frac{\mathbf{1}_{v_i \leftrightarrow v_j}}{\deg v_i \deg v_j}.$$

The behavior of this object depends strongly on whether or not the degrees are strongly concentrated around their means. Indeed, the most interesting features of this graph from the spectral point of view arise precisely when the degrees stop being strongly concentrated. When all the degrees are concentrated, (the $p = \Omega(\log n/n)$ regime), then one has

$$\left\|T^{-1/2}AT^{-1/2}\right\|_F^2 = \sum_{i,j} \frac{\mathbf{1}_{v_i \leftrightarrow v_j}}{\deg v_i \deg v_j} \approx \sum_{i,j} \frac{\mathbf{1}_{v_i \leftrightarrow v_j}}{(np)^2} \approx \frac{1}{p}.$$

This suggests that eigenvalues of $\mathscr{L}$ are within a $1/\sqrt{np}$ window of 1. This is consistent with what one would expect by comparison with the GOE. Namely, if the variance of the entries is scaled to be order $1/n$, a typical eigenvalue should be at unit order. Further, the comparison with a Wigner matrix is especially good in this regime as the degrees strongly concentrate and hence the dependence of the entries becomes relatively weak.

Outside of this regime, the answer is less clear. If p is exactly $\log n/n$, the graph has isolated vertices with probability tending to $1/e$ [112]. If the graph has isolated vertices, then $\mathscr{L}$ has eigenvalue 0 with multiplicity greater than 1. In fact, it is easily checked that the dimension of the kernel of $\mathscr{L}$ is the number of connected components of the graph. Thus as p decreases to the connectivity threshold ($\log n/n$), the extreme parts of the spectra appear to disobey this $1/\sqrt{np}$ prediction. This thesis will estimate the nontrivial eigenvalues of $\mathscr{L}$ as p descends to $\log n/n$ and attempt to determine their order.

Each of these problems has additional motivations outside of the scope of exploring the Gaussian universality class. Each problem will be presented with its own history and motivation in addition to its contribution to understanding the broader picture.

1.5 Understanding the Variance Expression arising from the CLT for Linear Statistics

We recall for convenience the limiting variance expression that arises in the GOE (1.6), given by

$$V_\phi = \frac{1}{2\pi^2} \oint_{\substack{|z|=1 \\ \Im z>0}} \oint_{\substack{|w|=1 \\ \Im w>0}} \phi'(\Re z)\phi'(\Re w) \log\left|\frac{1-zw}{1-z\bar{w}}\right| \frac{\Im z}{z}\frac{\Im w}{w} dz dw,$$

for C^1 test functions ϕ. This expresses the variance in terms of the *Green's function for the Laplacian in the upper half plane with Dirichlet boundary conditions* (which is, up to scaling, $\log\left|\frac{1-zw}{1-z\bar{w}}\right|$). This expression reveals a connection between these CLTs and the *Gaussian Free Field*, which is elaborated upon more in [18]. However, this particular choice of expression is not widely adopted, and there are many others in frequent use.

We will present a broad sampling of these expressions and show their equivalence, as well as show a few simple properties of the quadratic form, which can be connected to the $H^{1/2}(S)$ fractional Sobolev norm on the unit circle.

Another equivalent expression is

$$V_\phi^J := \frac{1}{\pi^2}\int_{-1}^{1}\frac{\phi(y)}{\sqrt{1-y^2}}\int_{-1}^{1}\frac{\phi'(x)\sqrt{1-x^2}}{x-y}\,dxdy. \tag{1.9}$$

This appears in [68], and it can be obtained by integrating the first expression by parts.[8] Note that it is stated in terms of a singular integral. This is properly defined in the terms of the Cauchy principal value, i.e.

$$\text{p. v.}\int_{-1}^{1}\frac{\phi'(x)\sqrt{1-x^2}}{x-y}\,dx := \lim_{\epsilon\to 0}\int_{\substack{[-1,y-\epsilon]\\ \cup[y+\epsilon,1]}}\frac{\phi'(x)\sqrt{1-x^2}}{x-y}\,dx.$$

This shows that V_ϕ can be understood as the Dirichlet form associated to a weighted Hilbert transform of the derivative. For more on the Hilbert transform, see for example [111] or [56].

Yet another important expression for the variance is

$$V_\phi^S = \frac{1}{2\pi^2}\int_{-1}^{1}\int_{-1}^{1}\left(\frac{\phi(x)-\phi(y)}{x-y}\right)^2\frac{1-xy}{\sqrt{1-x^2}\sqrt{1-y^2}}\,dxdy. \tag{1.10}$$

This appears in [106]. It is useful in that it allows one to determine for which ϕ the variance is finite.

Most importantly, all of these variance expressions agree when the test functions are sufficiently regular.

Proposition 1. *For ϕ that are $C^1[-1,1]$, $V_\phi = V_\phi^J = V_\phi^S$.*

This is not the sharpest possible condition. For example, if ϕ is absolutely continuous and ϕ' is in $L^p[-1,1]$ for some $p > 1$, then all of these expressions can be seen to agree by a limiting argument.

---

[8] Because of various conventions taken, various factors of 2 are easily lost when comparing this expression to Johansson's. That this is correct is most easily checked in following the proofs here.

Proof. Using the expression (1.6), we parameterize the curves to get that

$$V_\phi = \frac{1}{2\pi^2}\int_0^\pi\int_0^\pi \phi'(\cos s)\phi'(\cos t)\log\left|\frac{1-e^{it-is}}{1-e^{it+is}}\right|\sin s\sin t\; dsdt. \tag{1.11}$$

We now make the changes of variables $x = \cos s$ and $y = \cos t$. Doing so produces the equation

$$V_\phi = \frac{-1}{2\pi^2}\int_{-1}^1\int_{-1}^1 \phi'(x)\phi'(y)\log\left[\frac{1-xy+\sqrt{1-x^2}\sqrt{1-y^2}}{1-xy-\sqrt{1-x^2}\sqrt{1-y^2}}\right]\; dxdy. \tag{1.12}$$

We will now integrate by parts with respect to y, integrating $\phi'(y)$ and differentiating the log expression. It can be checked that

$$\frac{\partial}{\partial y}\log\left[\frac{1-xy+\sqrt{1-x^2}\sqrt{1-y^2}}{1-xy-\sqrt{1-x^2}\sqrt{1-y^2}}\right] = 2\frac{\sqrt{1-x^2}}{\sqrt{1-y^2}(x-y)}.$$

Let I_ϵ denote $[-1, x-\epsilon]\cup[x+\epsilon, 1]$, then we have that for all x,

$$\int_{-1}^1 \phi'(y)\log\left[\frac{1-xy+\sqrt{1-x^2}\sqrt{1-y^2}}{1-xy-\sqrt{1-x^2}\sqrt{1-y^2}}\right]\; dy$$
$$= \lim_{\epsilon\to 0}\int_{I_\epsilon}\phi'(y)\log\left[\frac{1-xy+\sqrt{1-x^2}\sqrt{1-y^2}}{1-xy-\sqrt{1-x^2}\sqrt{1-y^2}}\right]\; dy.$$

By applying the integration by parts to this integral, we get a boundary term

$$-\partial := \phi(y)\log\left[\frac{1-xy+\sqrt{1-x^2}\sqrt{1-y^2}}{1-xy-\sqrt{1-x^2}\sqrt{1-y^2}}\right]\Bigg|_{x-\epsilon}^{x+\epsilon},$$

as the contributions at $y = \pm 1$ vanish. Most importantly, we have that $\lim_{\epsilon\to 0}\partial = 0$ so that

$$\int_{I_\epsilon}\phi'(y)\log\left[\frac{1-xy+\sqrt{1-x^2}\sqrt{1-y^2}}{1-xy-\sqrt{1-x^2}\sqrt{1-y^2}}\right]\; dy \;=\; \partial - \int_{I_\epsilon}\phi(y)2\frac{\sqrt{1-x^2}}{\sqrt{1-y^2}(x-y)}.$$

In fact, the antiderivative of $\phi'(y)$ may be taken to be $\phi(y)+C(x)$ for any continuous function $C(x)$. Doing the integration by parts produces the expression

$$V_\phi = \frac{-1}{\pi^2}\int_{-1}^1\phi'(x)\left[\text{p. v.}\int_{-1}^1(\phi(y)+C(x))\frac{\sqrt{1-x^2}}{\sqrt{1-y^2}(x-y)}\; dy\right]dx. \tag{1.13}$$

We need to change the order of integration to establish that $V_\phi^J = V_\phi$. We will justify this by using the so-called maximal Hilbert transform. Let

$$H^*(x) := \sup_{\epsilon>0}\left|\int_{I_\epsilon} \frac{\phi(y)+C(x)}{\sqrt{1-y^2}(x-y)}\,dy\right|.$$

It is now consequence of standard theory that for any $1<p<\infty$, there is a constant C_p so that

$$\|H^*(x)\|_p \le C_p\left(\left\|\phi(y)(1-y^2)^{-1/2}\right\|_p + \|C(x)\|_\infty \left\|(1-y^2)^{-1/2}\right\|_p\right),$$

where these are the standard $L^p[-1,1]$ norms (see Theorem 4.1.12 of [56]). From the continuity of ϕ and $C(x)$, this is finite for $p<2$. By Hölder's inequality, we now have an integrable dominator, so that we can write

$$\int_{-1}^1 \phi'(x)\left[\text{p. v.}\int_{-1}^1 (\phi(y)+C(x))\frac{\sqrt{1-x^2}}{\sqrt{1-y^2}(x-y)}\,dy\right]dx \\ = \lim_{\epsilon\to 0}\int_{-1}^1 \phi'(x)\left[\int_{I_\epsilon}(\phi(y)+C(x))\frac{\sqrt{1-x^2}}{\sqrt{1-y^2}(x-y)}\,dy\right]dx. \quad (1.14)$$

We can change the order of the integration as the integrand is continuous on this domain. Switching the order of the integration and applying a similar argument on the dominator, we conclude that

$$V_\phi = \frac{-1}{\pi^2}\int_{-1}^1\left[\text{p. v.}\int_{-1}^1 \phi'(x)(\phi(y)+C(x))\frac{\sqrt{1-x^2}}{\sqrt{1-y^2}(x-y)}\,dx\right]dy. \quad (1.15)$$

Taking $C(x)=0$ proves $V_\phi^J = V_\phi$. To prove that $V_\phi^J = V_\phi^S$, we take $C(x) = -\phi(x)$ and integrate by parts again. This time, we integrate with respect to x and take $(\phi(x)-\phi(y))^2$ as the antiderivative of $2\phi(x)\phi'(x) - 2\phi'(x)\phi(y)$. The x derivative of the remainder is given by

$$\frac{\partial}{\partial x}\frac{\sqrt{1-x^2}}{\sqrt{1-y^2}(x-y)} = \frac{xy-1}{\sqrt{1-y^2}\sqrt{1-x^2}(x-y)^2},$$

which after a similar argument as with the first integration by parts leads to

$$V_\phi = \frac{1}{2\pi^2}\int_{-1}^1\left[\text{p. v.}\int_{-1}^1\left(\frac{\phi(x)-\phi(y)}{x-y}\right)^2\frac{1-xy}{\sqrt{1-x^2}\sqrt{1-y^2}}\,dx\right]dy.$$

Note that the integral is absolutely convergent, and so the principal value may be dropped, proving Proposition 1. □

The expression V_ϕ^S is the most robust in that its definition requires no regularity assumptions to make sense. With this definition, we turn to analyzing those ϕ for which V_ϕ^S is finite. We let $w(x)$ denote the measure on $[-1,1]$ given by density $1/\sqrt{1-x^2}$. We will define $\mathscr{H}$ to be the collection of all $L^1(w)$ functions $\phi(x)$ for which V_ϕ^S is finite. The extra condition that the test functions are in $L^1(w)$ imposes essentially no additional conditions on ϕ. In fact, as it is easily checked that $\frac{1-xy}{(x-y)^2} \geq \frac{1}{2}$ for Lebesgue-a.e. $(x,y) \in [-1,1]^2$ we have that

$$\frac{1}{4\pi^2}\int_{-1}^{1}\int_{-1}^{1}\frac{(\phi(x)-\phi(y))^2}{\sqrt{1-x^2}\sqrt{1-y^2}}\,dxdy \leq \frac{1}{2\pi^2}\int_{-1}^{1}\int_{-1}^{1}\left(\frac{\phi(x)-\phi(y)}{x-y}\right)^2\frac{1-xy}{\sqrt{1-x^2}\sqrt{1-y^2}}\,dxdy. \quad (1.16)$$

The left hand side is one half the variance of ϕ with respect to the normalization of the measure $w(x)$, and hence $\mathscr{H} \subseteq L^2(w)$. Note also that $\phi \mapsto \sqrt{V_\phi^S}$ is a seminorm, which when combined with the $L^2(w)$ norm can be used to metrize $\mathscr{H}$.

We now turn to our first characterization of $\mathscr{H}$, which shows that it contains many smooth functions.

Proposition 2. *For any $\alpha > 1/2$, there is a constant C_α so that $V_\phi^S \leq C_\alpha \left|\phi\right|_\alpha^2$, where*

$$\left|\phi\right|_\alpha = \sup_{x,y\in[-1,1]}\frac{|\phi(x)-\phi(y)|}{|x-y|^\alpha}.$$

Proof. We use the given bound $|\phi(x)-\phi(y)| \leq C_\alpha |x-y|^\alpha$ to conclude that

$$V_\phi^S \leq \frac{|\phi|_\alpha^2}{2\pi^2}\int_{-1}^{1}\int_{-1}^{1}|x-y|^{2(\alpha-1)}\frac{1-xy}{\sqrt{1-x^2}\sqrt{1-y^2}}\,dxdy.$$

For $\alpha > \frac{1}{2}$, this is integrable, as can be seen by taking the rotated coordinate system $u = x-y$ and $v = x+y$. □

On the other hand, we will see that $\mathscr{H}$ does not contain *only* continuous functions. To establish this property though, we will need to develop a Fourier type characterization of it.

As V_ϕ^S is a quadratic form, it is possible to diagonalize it. By polarization, we can define the real symmetric bilinear form

$$\langle \phi_1, \phi_2 \rangle_V = \frac{1}{2\pi^2} \int_{-1}^{1} \int_{-1}^{1} \frac{\phi_1(x) - \phi_1(y)}{x-y} \frac{\phi_2(x) - \phi_2(y)}{x-y} \frac{1-xy}{\sqrt{1-x^2}\sqrt{1-y^2}}\, dxdy,$$

for any ϕ_1 and $\phi_2 \in \mathscr{H}$. It will turn out that the Chebyhsev polynomials provide a countable orthogonal set with dense span. The Chebyshev T_k polynomials of the first kind are defined by the relation $T_k(\cos\theta) = \cos(k\theta)$. Equivalently, this is stated in the complex plane as $T_k((z+z^{-1})/2) = \frac{z^k+z^{-k}}{2}$, where the equivalence can be seen by taking z on the unit circle.

The orthogonality properties of these Chebyshev polynomials are better expressed for the inner product given by

$$\langle \phi_1, \phi_2 \rangle_T := \frac{2}{\pi} \int_{-1}^{1} \frac{\phi_1(x)\phi_2(x)}{\sqrt{1-x^2}}\, dx,$$

for which the Chebyshev polynomials offer a complete orthonormal system.[9] Using this inner product, the Chebyshev coefficients $\hat{f}(k)$ can be defined by $\hat{f}(k) := \langle f, T_k \rangle_T$. Much of the theory of the convergence of Chebyshev series transfers immediately from well established theory for the convergence of Fourier series on the unit circle. On making the substitution $x = \cos\theta$, we have that $f(\cos\theta)$ is in $L^2(\mathbb{T})$ if and only if f is in $L^2(w)$. Further, $f(\cos\theta)$ is continuous on the circle if and only if $f(x)$ is continuous on $[-1, 1]$, and most importantly, the Fourier series for $f(\cos\theta)$ is given by

$$f(\cos\theta) = \sum_{k=0}^{\infty} \hat{f}(k) \cos k\theta.$$

[9]The constant function T_0 must be rescaled, but this can be essentially be ignored, as $V_\phi = V_{\phi+C}$ for any scalar C.

Thus, convergence properties about the Fourier series transfer directly to the Chebyshev series. Our main result connecting V_ϕ^S and the Chebyshev polynomials is the following.

Proposition 3. *For any $\phi \in \mathscr{H}$, we have that*

$$V_\phi^S = \frac{1}{2}\sum_{k=1}^{\infty} k|\hat{f}(k)|^2.$$

Moreover, for any $\phi \in L^1(w)$, $\phi \in \mathscr{H}$ if and only if the right hand side is finite.

If we assume additional regularity on ϕ, such as $\phi \in C^1[-1,1]$, then this proposition is relatively straightforward. To prove it for every $\phi \in \mathscr{H}$, we require a fair bit more care. We begin with an integration by parts lemma whose proof is similar to that of Proposition 1.

Lemma 1. *For $\phi \in \mathscr{H}$ and for $g \in C^2[-1,1]$,*

$$\langle \phi, g\rangle_V = \frac{1}{\pi^2}\int_{-1}^{1} \frac{\phi(y)}{\sqrt{1-y^2}}\left[\text{p. v.}\int_{-1}^{1} \frac{g'(x)\sqrt{1-x^2}}{x-y}\ dx\right] dy.$$

To prove this statement, we need a simple technical lemma. For any $y \in (-1,1)$ let $I_\epsilon = [-1, y-\epsilon] \cup [y+\epsilon, 1]$.

Lemma 2. *For any $p > 1$, there is a constant $C_p > 0$ so that for any f absolutely continuous on $[-1,1]$ having $f(1) = f(-1) = 0$ and $f' \in L^p[-1,1]$,*

$$\left|\int_{I_\epsilon} \frac{f(x)}{x-y}\ dx\right| \le C_p\|f'\|_p.$$

Proof. This follows directly from integration by parts. Doing so, we get

$$\int_{I_\epsilon} \frac{f(x)}{x-y}\ dx = \log|x-y| f(x)\ |_{\partial I_\epsilon} - \int_{I_\epsilon} f'(x)\log|x-y|\ dx.$$

We control each term separately. For the boundary term, we have

$$\left|\log|x-y|f(x)\ |_{\partial I_\epsilon}\right| \le C(1+\log_+(1/\epsilon))\omega(|2\epsilon|)),$$

where ω is the modulus of continuity of f. By Hölder's inequality, we have that $\omega(t) \leq t^{1-1/p}\|f'\|_p$, so that this expression can be bounded uniformly in ϵ. For the integral term, we simply apply Hölder's inequality

$$\int_{I_\epsilon} |f'(x)\log|x-y|| \ dx \leq \|f'\|_p \left|\int_{-2}^{2} |\log|x||^{p/(p-1)} \ dx\right|^{1-1/p},$$

and note that the log is always integrable. □

Proof of Lemma 1 . By the Schwarz inequality, the integral defining $\langle\phi, g\rangle_V$ is well defined, and so we can cut away a strip around $x = y$ to conclude that

$$\langle\phi, g\rangle_V = \lim_{\epsilon\to 0}\frac{1}{2\pi^2}\int_{-1}^{1}\int_{I_\epsilon}\frac{\phi(x)-\phi(y)}{x-y}\frac{g(x)-g(y)}{x-y}\frac{1-xy}{\sqrt{1-x^2}\sqrt{1-y^2}}\ dxdy.$$

As the integrand is now well-behaved, we may split the integral into two pieces I_1 and I_2. One such piece is given by

$$I_1 = \lim_{\epsilon\to 0}\frac{1}{2\pi^2}\int_{-1}^{1}\int_{I_\epsilon}\frac{-\phi(y)}{x-y}\frac{g(x)-g(y)}{x-y}\frac{1-xy}{\sqrt{1-x^2}\sqrt{1-y^2}}\ dxdy,$$

and I_2 given so that $\langle\phi, g\rangle_V = I_1 + I_2$. We note that by symmetry of the integrand $I_1 = I_2$, and it suffices to compute I_1. Integrating the interior integral by parts in x, we conclude (by analogous reasoning to what was done in Proposition 1)

$$I_1 = \lim_{\epsilon\to 0}\frac{1}{2\pi^2}\int_{-1}^{1}\frac{\phi(y)}{\sqrt{1-y^2}}\left[\int_{I_\epsilon}\frac{g'(x)\sqrt{1-x^2}}{x-y}\ dx\right]dy.$$

It remains to justify commuting the limit and integral. Note that $g'(x)\sqrt{1-x^2}$ has a derivative in $L^p[-1,1]$ for every $1 \leq p < 2$. Hence by Lemma 2, there is a constant C sufficiently large that

$$\left|\int_{I_\epsilon}\frac{g'(x)\sqrt{1-x^2}}{x-y}\ dx\right| \leq C$$

Thus, we have that the integrand is dominated by $C\left|\frac{\phi(y)}{\sqrt{1-y^2}}\right|$, which is integrable, as $\phi(y) \in L^1(w)$. Thus, we may apply dominated convergence, which completes the proof. □

The advantage of Lemma 1 is that we can now compute inner products with respect to Chebyshev polynomials. These are due to a set of equations that go by the name *Aerofoil equations* [13]. To state them, we must recall the Chebyshev polynomials of the second kind U_k, which can be defined by the relationship $U_k(\cos\theta) = \frac{\sin k\theta}{\sin\theta}$. We also have the identity that $\frac{d}{dx}T_k(x) = kU_{k-1}(x)$. The Aerofoil equations are given by

$$T_k(x) = \text{p. v.}\,\frac{1}{\pi}\int_{-1}^{1} \frac{U_{k-1}(y)\sqrt{1-y^2}}{x-y}\,dy, \tag{1.17}$$

$$U_{k-1}(x) = -\,\text{p. v.}\,\frac{1}{\pi}\int_{-1}^{1} \frac{T_k(y)}{(x-y)\sqrt{1-y^2}}\,dy,$$

see Lemma 4.5.2 of [13] for a proof. Combining this first equation with Lemma 1, we get that $\langle \phi, T_k\rangle_V = k\hat{\phi}(k)$. We are now in a position to prove Proposition 3.

Proof of Proposition 3. Let V_ϕ^F be the square seminorm given in terms of the Chebyshev coefficients, i.e.

$$V_\phi^F := \frac{1}{2}\sum_{k=1}^{\infty} k|\hat{\phi}(k)|^2.$$

Let $\phi \in L^1(w)$ be arbitrary, and define $\phi_m(x) = \sum_{k=0}^{m}\hat{\phi}(k)T_k(x)$. Note that from Lemma 1 and (1.17), we have that $V_{\phi_m}^F = V_{\phi_m}^S$ for every $m \geq 0$. Further, note that if either V_ϕ^S or V_ϕ^F is finite, then $\phi \in L^2(w)$. Thus, we may assume this is the case from here on out. As this is the case, then by Carleson's theorem on Fourier series, we have $\phi_m \to \phi$ Lebesgue-a.e. (Note we need not even appeal to this theorem, as it would be enough to pass to a subsequence). By Fatou's Lemma, $\liminf_{m\to\infty} V_{\phi_m}^S \geq V_\phi^S$. Note that we have $\liminf_{m\to\infty} V_{\phi_m}^S = \lim_{m\to\infty} V_{\phi_m}^F = V_\phi^F$, and hence $V_\phi^S \leq V_\phi^F$. On the other hand, we have, by orthogonality and Lemma 1,

$$V_\phi^S = \langle \phi, \phi\rangle_V = \langle \phi - \phi_m, \phi - \phi_m\rangle_V + \langle \phi_m, \phi_m\rangle_V \geq V_{\phi_m}^S.$$

Note that this equation is the only place that we actually require the extra technical work of Lemma 1. As this holds for all m, we get that

$$V_\phi^F = \sup_{m\geq 1} V_{\phi_m}^S \leq V_\phi^S.$$

□

This Fourier interpretation allows us to make a strong connection between $\mathscr{H}$ and the so-called *Dirichlet space* $\mathscr{D}$ of analytic functions (See [5, 103] for a survey of this space). These can be defined as those analytic functions in the unit disc $\mathbb{D} = \{z \in \mathbb{C} \ : \ |z| < 1\}$ so that

$$\|f\|_{\mathscr{D}}^2 := \int_{\mathbb{D}} |f'(z)|^2 \, dA(z),$$

where dA is the standard Lebesgue area element. If $f(z)$ has power series $f(z) = \sum_{n=0}^{\infty} a_n z^n$, then we have

$$\|f\|_{\mathscr{D}}^2 = \sum_{n=1}^{\infty} n|a_n|^2.$$

Further, all such analytic functions have non-tangential limits at almost every boundary point, and thus we can form the following correspondence. Let $\mathscr{D}^*$ be those functions in the Dirichlet space that have reflection symmetry, i.e. $\mathscr{D}^* := \{f(z) \in \mathbb{D} \ : \ f(z) = \overline{f(\bar{z})}\}$. This becomes a Banach space with the norm

$$\|f\|_{\mathscr{D}^*} = \frac{1}{2}\left(\|f\|_{\mathscr{D}} + \|\Re f(e^{i\theta})\|_{L^2}\right).$$

This is isometrically isomorphic to $\mathscr{H}$ as a Banach space under norm $\sqrt{V_\phi^S} + \|f\|_{L^2(w)}$. Indeed, for any $g \in \mathscr{H}$ we can define

$$T[g](z) := \frac{1}{2\pi}\int_0^{2\pi} \frac{e^{i\theta}+z}{e^{i\theta}-z} g(\cos\theta) \, d\theta.$$

It is straightforward to verify $T[g]$ is analytic in the disc and that its power series coefficients at 0 are the Chebyshev coefficients of $g(x)$, so that $2\|T[g]\|_{\mathscr{D}} = \sqrt{V_g^S}$. Further, the real parts of the boundary values of $\Re T[g](e^{i\theta})$ are almost everywhere equal to $g(\cos\theta)$. Thus we have that $\|\Re T[g](e^{i\theta})\|_{L^2} = 2\|g\|_{L^2(w)}$. An inverse to $g \mapsto T[g]$ is given by mapping the function $f(z) \in \mathscr{D}$ to $\Re f(\cos^{-1}(x))$ in $\mathscr{H}$.

This affords us two powerful tools. First, we are able to construct many examples of pathological functions in $\mathscr{H}$.

Proposition 4. *Let $U \subset \mathbb{C}$ be any connected, simply-connected, open set with finite area, which is symmetric with respect to reflection over the real axis. Then there exists a conformal isomorphism $\phi(z) : \mathbb{D} \to U$ so that $\Re\phi(\cos^{-1}(x)) \in \mathscr{H}$.*

Proof. The primary realization is that for a conformal map such as ϕ, $\|\phi\|_{\mathscr{D}}$ is the area of the image, as $|\phi'|^2$ is the Jacobian for the map. As U is reflection symmetric, we may choose ϕ to take $\phi(0)$ to some given value of $U \cap \mathbb{R}$. It then follows from Schwartz reflection that $\phi(z) = \overline{\phi(\bar{z})}$. As a consequence we have that $\Re\phi(\cos^{-1}(x)) \in \mathscr{H}$. □

Thus, as we can construct domains in the plane that are unbounded but have finite area, we have that we can construct functions in $\mathscr{H}$ that are discontinuous at any desired point or even densely discontinuous.

Second, we immediately conclude that functions in $\mathscr{H}$ cannot be large too frequently. Specifically we have that

$$\sup \left\{ \int_{-1}^{1} \frac{e^{|f(x)|^2}}{\sqrt{1-x^2}}\, dx \;\middle|\; \hat{f}(0) = 0, V_f^S \leq 1 \right\} < \infty,$$

which is an immediate consequence of the Chang-Marshall theorem for $\mathscr{D}$ [25, 88].

Chapter 2

GLOBAL FLUCTUATIONS OF THE β-JACOBI ENSEMBLE

ADAPTED FROM JOINT WORK WITH IOANA DUMITRIU [98].

2.1 Background, Definitions and Motivation

The Jacobi ensemble, in the context of what has been described so far, can be succinctly given as a log-gas; hence, it has eigenvalue density

$$d\mu_J(\lambda_1,\ldots,\lambda_n) := \frac{1}{Z}\prod_i \lambda_i^{\frac{\beta}{2}[n_1-n+1]-1}(1-\lambda_i)^{\frac{\beta}{2}[n_2-n+1]-1}\prod_{i<j}|\lambda_i-\lambda_j|^{\beta}, \tag{2.1}$$

where $Z = Z(n, n_1, n_2, \beta)$ is a normalization constant. In full generality, $\beta > 0$, while n_1 and n_2 need not be positive integers; in fact, the only constraints (which relate to the integrability of the measure) are that $n_1, n_2 > n - 1$. The Jacobi ensemble arises in a variety of contexts and applications, over which we will give a brief sampling.

In the case that $\beta \in \{1, 2, 4\}$ and $n_1, n_2 \in \mathbb{N}$, they admit full matrix models as $J = W_1^{1/2}(W_1 + W_2)^{-1}W_1^{1/2}$, where W_1, W_2 are Wishart matrices.[1] Because of this connection, they sometimes go by the name "Double Wishart" or Beta matrices.

The MANOVA Connection

These full matrix models arise most prominently through a connection to statistics.[2] In greatest generality, they pertain to the general linear model fitting problem. In the classical one dimensional case, one has a vector Y of observations which one would like

[1] A Wishart matrix $W(n, n_1)$ is defined by taking an $n \times n_1$ matrix G of independent standard normals, be they real, imaginary or quaternion, and defining $W = GG^A$. Here W_1 is $W(n, n_1)$ and $W(n, n_2)$. For a proof, see [95, 44].

[2] We follow a development of this material that can be found in Chapter 10 of [95].

to correlate to a deterministic vector X. Thus one seeks to find regression coefficient ρ so that

$$Y = X\rho + E$$

for the smallest possible error E. Once one chooses a good estimator for ρ in terms of X and Y – most commonly using least squares – one can speak of testing the null hypothesis that $\rho = 0$.

The multivariate linear model is a natural higher dimensional analogue of this one dimensional case. There, the observations themselves are m-dimensional, which one represents as row vectors. Making l observations, one forms an $l \times m$ matrix Y of observations. Then, one seeks to find a correlation with an $l \times p$ matrix of deterministic data X; thus the model has the form

$$Y = X\mathbb{B} + E,$$

for some matrix of coefficients $\mathbb{B}$. Under the assumption of normal variates, so that each row of Y is $N(X\mathbb{B}_{i\cdot}, \Sigma)$, the maximum likelihood estimators for $\mathbb{B}$ and the covariance matrix are easily derived (see Theorem 10.1.1 of [95]). A general null hypothesis in this situation takes the form that $C\mathbb{B} = 0$ for some fixed matrix C.

The Jacobi ensemble appears in a special case of this general framework, the multivariate analysis of variance (MANOVA), which is instructive to outline. Suppose one has p classes of subjects. Among these p classes, one measures m different characteristics, and we assume that there are no correlations between these m characteristics over the entire population. One would like test the null hypothesis that all the classes have the same means.

From each class i for $1 \leq i \leq p$, one makes q_i independent samplings. Let y_{ij} be the j^{th} $m \times 1$ sample from class i, so that one has made $l = \sum_{i=1}^{p} q_i$ total samples. We assume that all the samples are normally distributed. Let $\bar{y}_i = \frac{1}{q_i}\sum_{j=1}^{q_i} y_{ij}$ be the sample mean of the q_i observations from the i^{th} class, and let $\bar{y} = \frac{1}{l}\sum_{i=1}^{p} q_i\bar{y}_i$ be the sample mean of the population.

In terms of these data, one forms two $m \times m$ matrices. The *matrix due to hypothesis* is given by

$$A = \sum_{i=1}^{p} q_i(\bar{y}_i - \bar{y})(\bar{y}_i - \bar{y})^t,$$

and it measures the variance between different classes. Note that it has small singular values when all of $\bar{y}_i \approx \bar{y}$. Under the null hypothesis, A has the Wishart distribution $W(m, p-1)$.

Likewise, the *matrix due to error* is given by

$$B = \sum_{i=1}^{p} \sum_{j=1}^{q_i} (y_{ij} - \bar{y}_i)(y_{ij} - \bar{y}_i)^t,$$

which measures the variance within classes. Note that this matrix has small singular values when $y_{ij} \approx \bar{y}_i$ for each j. Like A, B carries the distribution $W(m, l-p)$, although unlike A, this remains true even under a non-null hypothesis (i.e. that the classes have different means). Furthermore, A and B are always independent.

Intuitively, one would like to reject the null hypothesis when A is much larger than B. Formally, we should choose a statistic $T(A, B)$ and a rejection threshold, and reject the null hypothesis when a sample statistic $\hat{T}(A, B)$ exceeds that threshold. The choice of statistic is not obvious, and in this case, and there are multiple statistics that are in frequent use.

However, there are some symmetries in the problem that greatly reduce the class of desirable statistics. In particular, if instead of sampling y_{ij}, one samples Qy_{ij} for some fixed orthogonal matrix Q, then the matrices of hypothesis and error become QAQ^t and QBQ^t respectively. Under the hypotheses of this model, both of these matrices have the same distribution as A and B. Thus, one would like for the statistic $T(A, B)$ to be invariant under orthogonal change of base, i.e. $T(A, B) = T(QAQ^t, QBQ^t)$.

One natural choice of statistics are functions of the eigenvalues of AB^{-1}. The most common statistics are the Wilks' likelihood ratio $1/\det(1 + AB^{-1})$, Hotelling's $T_0^2 = \operatorname{tr} AB^{-1}$, Pillai's trace $\operatorname{tr}(AB^{-1})(1 + AB^{-1})^{-1}$, and finally Roy's largest root

$\|AB^{-1}\|_2$. Note that two of these are linear statistics, and Wilk's likelihood ratio is the exponential of a linear statistic. [3]

As A and B follow independent Wishart distributions, we can connect these eigenvalues to the eigenvalue of the Jacobi ensemble. Specifically, with J as before we have that $\sigma(J) = \sigma\left((1 + AB^{-1})^{-1}\right)$ where the Jacobi parameters are $n_1 = p-1$, $n_2 = l-p$ and $n = m$. Thus, the common eigenstatistics for testing the null hypothesis in this particular case of MANOVA can be expressed in terms of the Jacobi ensemble.

The Random Projection Connection

The Jacobi ensemble also can be connected to Haar matrix ensembles in the $\beta = 1, 2, 4$ cases. Define an invariant orthogonal projection π to be one whose law is invariant in law under $Q\pi Q^A$ for any Q orthogonal, unitary or symplectic, as the case $\beta = 1, 2$ or 4 demands.

Proposition 5. *Let π be a fixed $l \times l$ self-adjoint projection of rank q, and let $\tilde{\pi}$ be a fixed $l \times l$ invariant orthogonal projection of rank $\tilde{q}$ with $q \leq \tilde{q}$. Then*

$$\mathcal{L}\sigma\left(\pi\tilde{\pi}\pi\right) = \mathcal{L}\sigma(J),$$

where J is a β-Jacobi ensemble with parameters $n_1 = \tilde{q}$, $n_2 = l - \tilde{q}$ and $n = q$.

This proof is originally due to Collins [30], and it was rediscovered by Edelman and Sutton [38]. We follow the proof stated in [30], which has some typos which we correct here. See also the proof due to Forrester [43].

Proof. Let G be an $l \times l$ matrix of independent standard Gaussians over $\mathbb{R}, \mathbb{C}$, or $\mathbb{J}$,

---

[3] In the general case that one does not assume the sample covariance is the identity, the problem becomes invariant under any change of basis, from which point one can conclude that the statistic must be a function of the eigenvalues of AB^{-1}. See Theorem 10.2.1 of [95].

independent of $\tilde{\pi}$. Define

$$X_1 = \pi G \tilde{\pi} G^A \pi$$
$$X_2 = \pi G (I - \tilde{\pi}) G^A \pi.$$

Note that this makes X_1 and X_2 independent. Further, each restricts to a map from $\mathrm{Im}(\pi) \to \mathrm{Im}(\pi)$. Picking an orthonormal basis for this space, each becomes a Wishart matrix, with laws $W(q, \tilde{q})$ and $W(q, l - \tilde{q})$ respectively. In the same way, $X_1 + X_2$ has the law of $W(q, l)$.

By the polar decomposition (or, more generally, the Cartan decomposition), we have that there exists a Q orthogonal, unitary or symplectic that depends measurably on G so that $\pi G Q$ is self-adjoint positive semidefinite. Hence

$$(\pi G Q)^2 = \pi G Q Q^A G^A \pi = X_1 + X_2,$$

so that $\pi G Q$ this is the positive semidefinite square root of $X_1 + X_2$. Note that we have that

$$\pi \left(X_1 + X_2\right)^{1/2} = \left(X_1 + X_2\right)^{1/2} = \left(X_1 + X_2\right)^{1/2} \pi.$$

Thus, we now have the exact relationship that

$$X_1 = \left(X_1 + X_2\right)^{1/2} \pi Q^A \tilde{\pi} Q \pi \left(X_1 + X_2\right)^{1/2}.$$

Using the invariance of $\tilde{\pi}$ and the almost sure invertibility of $X_1 + X_2$, we have the conclusion of the proposition. □

This gives the eigenvalues of a Jacobi ensemble the alternative interpretation as taking a $q \times \tilde{q}$ submatrix of a orthogonal, unitary, or symplectic matrix and taking its squared singular values.

Mesoscopic Universal Conductance Fluctuations

The Jacobi ensemble also plays a role in understanding the physics behind a certain simple model of electrical conductance.[4] We consider placing a 1-dimensional wire between two external reservoirs of electrons which are at different potentials. The wire provides resistance if an electron travelling through the wire *scatters* off of impurities. On sufficiently short scales (known as the *phase coherence scale*, which for most materials is up to the order of a micrometer [121]), then the scattering event is adequately modeled by a single coin flip. Namely, there is some probability T that an electron entering the wire from the left transmits through the wire, and the conductance between the two reservoirs is $G = \frac{e^2}{\pi\hbar}T$, where e and $\hbar$ are physical constants (the charge of an electron and Planck's constant, respectively).

In 1-dimension and with fixed energy, an electron in the wire has only two possible states, travelling to the left or to the right, as speed is determined by the energy. When one generalizes to a wire which is not purely 1-dimensional, one allows an electron to have momentum in both the left-right direction and the orthogonal direction. With given energy, the quantization of momentum effectively implies there are some larger, finite number of states, corresponding to travelling with varying quantities of left-right momentum and orthogonal momentum. These different modes of transit are referred to as *channels*, and we consider the situation that there are n left-to-right channels and m right-to-left channels.

At the phase coherence scale, resistance in the wire is still determined by a single scattering event. An electron passing through the wire from left to right in some channel is scattered according with certain fixed probabilities to some other channels, some representing leftward motion and some representing rightward motion. Formally, we define I_l and I_r to be two n and m dimensional complex vectors representing the state of the travelling electron before scattering; here we have that I_l (respectively

[4]We follow a development of this material available in [44, 43, 2]. For more on the physical side of the problem, see [37, 121].

I_r) are the states corresponding to entering the wire from the left (respectively right). Let O_l and O_r be the states after scattering. If $|I_l|^2 + |I_r|^2 = 1$, then each entry of these vectors has squared modulus that represents the probability of the electron being in the given channel before scattering. The scattering event is represented by a matrix S for which

$$S \begin{bmatrix} I_l \\ I_r \end{bmatrix} = \begin{bmatrix} O_l \\ O_r \end{bmatrix}.$$

As no electrons are destroyed or created during the scattering, if we have that $|I_l|^2 + |I_r|^2 = 1$, then we must also have that $|O_l|^2 + |O_r|^2 = 1$. As this must be the case for any possible input states, we have that S is unitary.

Further S may be decomposed into pieces corresponding to transmitted electrons and reflected electrons. Specifically, we have that

$$S = \begin{bmatrix} R_{ll} & T_{lr} \\ T_{rl} & R_{rr} \end{bmatrix},$$

where R and T are the reflection and transmission matrices. The conductance G of the wire is now given by the *two probe Landauer formula*,

$$G/G_0 = \operatorname{tr} T_{lr}^* T_{lr} = \operatorname{tr} T_{rl}^* T_{rl},$$

for a physical constant G_0.

A similar random matrix ansatz is now posed as was done for the nuclear models. For a disordered wire, S should have no preserved structure besides being unitary, and hence a generic wire's scattering matrix should look like a randomly chosen unitary matrix. This is born out in experiment [121] and goes by the name universal conductance fluctuations. They are universal in that their magnitude is independent of the sample and even the length of the sample, provided that the length is kept short enough to be in the phase coherence regime.[5]

[5] It also must be longer than the *mean free path*, which is the average distance to a scattering event; if the wire length is too short, the electron will transmit with so high a probability that the impurities are irrelevant. This in the nanometer range and varies with the material and the electron energy.

Furthermore, as $R_{ll}R_{ll}^* + T_{lr}T_{lr}^* = I_n$, we have that this trace is given in terms of the squared singular values of R_{ll}, the upper $n \times n$ block of a unitary matrix. Thus, this trace is a linear statistic of the eigenvalues of a β-Jacobi matrix.

Tridiagonal Models

The most general matrix model for β-Jacobi ensemble is the tridiagonal model [74, 38], which covers any $\beta > 0$, and removes the condition that $n_1, n_2 \in \mathbb{N}$. We give the model below (hereafter referred to as the Edelman-Sutton model, as it appears most clearly in their work [38]). Given the matrix B_β defined as

$$B_\beta = \begin{pmatrix} c_n s'_{n-1} & & & & \\ -s_{n-1}c'_{n-1} & c_{n-1}s'_{n-2} & & & \\ & -s_{n-2}c'_{n-2} & c_{n-2}s'_{n-3} & & \\ & & \ddots & \ddots & \\ & & & -s_1c'_1 & c_1 \end{pmatrix}, \tag{2.2}$$

with the variables $c_i,\ s_i,\ i = 1, \ldots, n$, and $c'_j,\ s'_j,\ j = 1, \ldots, (n-1)$ obeying the distribution laws and relationships

$$\begin{aligned} &\left\{c_1, c_2, \ldots, c_n, c'_1, c'_2, \ldots, c'_{n-1}\right\} \text{ mutually independent,} \\ &c_i \sim \sqrt{\mathrm{Beta}(\tfrac{\beta}{2}(n_1 - n + i), \tfrac{\beta}{2}(n_2 - n + i))}, \\ &c'_j \sim \sqrt{\mathrm{Beta}(\tfrac{\beta}{2}j, \tfrac{\beta}{2}(n_1 + n_2 - 2n + 1 + j))}, \\ &s_i = \sqrt{1 - c_i^2} \text{ and } s'_j = \sqrt{1 - c_j'^2}\ , \end{aligned} \tag{2.3}$$

the eigenvalues of $A = B_\beta B_\beta^T$ are distributed according to (2.1) (see [38]).

In our work going forward, we will rely heavily on this model for understanding the eigenvalues of the β-Jacobi ensemble.

2.1.1 Scaling Regimes

We are interested in the behavior of X_f^A for some subclass of continuous functions as $n \to \infty$ with $(n_1 + n_2 - 2n)$ growing linearly in n and with β fixed. This is the only scaling regime in which the Jacobi ensemble looks distinct from other models. If either $n_1 \gg n$ or $n_2 \gg n$, in the case when $\beta = 1, 2, 4$, the Wishart matrices in the full models have $W_1 \approx \beta n_1 I_n$, respectively, $W_2 \approx \beta n_2 I_n$. For example if $n_2 \gg n$ and $n_2 \gg n_1$, this heuristic predicts that the Double Wishart model behaves like

$$W_1(W_1 + W_2)^{-1} \approx W_1(W_1 + \beta n_2 Id)^{-1} \approx W_1/(\beta n_2),$$

so that appropriately rescaling, Wishart behavior should appear. These heuristics are studied rigorously in Jiang [66]. (The symmetric regime, $n_1 \gg n$ and $n_1 \gg n_2$, predicts Wishart behavior with a huge shift in eigenvalues.)

Conversely, in the sublinear growth cases, i.e. where $(n_1 + n_2 - 2n) \ll n$, the Jacobi ensemble takes on behavior that looks much more like the classical compact groups. This connection is explicit for $\beta = 1, 4$ and fixed values of $n_1 - n$ and $n_2 - n$ (see Proposition 3.1 of [67]). Furthermore, the Haar projection formula in the $\beta = 1, 2, 4$ case shows that these Jacobi ensembles are the squared singular values of a principal submatrix of a Haar matrix whose dimensions are both $n - o(1)$, and thus it is expected that they should look strongly like the squared singular values of the whole Haar matrix.

These heuristics predict the correct limiting spectral distributions. In the superlinear case, the limiting spectral distribution is a point mass (easily seen also from 2.3, which shows that the matrix $B_\beta B_\beta^T$is very close to a multiple of the identity), while in the sublinear case, the limiting spectral distribution is the arcsine law. These statements about the limiting spectral distributions are straightforward exercises following the approach of Trotter [118]. We sketch this approach in the proof of the following theorem.

Theorem 1. *Let f be a continuous test function on* $[0, 1]$.

1. *If $n_1 + n_2 - 2n = o(n)$, then*
$$\frac{1}{n}\sum_{i=1}^{n} f(\lambda_i) \to_{\mathbb{P}} \frac{1}{\pi}\int_0^1 \frac{f(x)}{\sqrt{x(1-x)}}\, dx.$$

2. If $n_1/n \to p$ *and* $n_2/n \to q$*, then*

$$\frac{1}{n}\sum_{i=1}^{n} f(\lambda_i) \to_{\mathbb{P}} \int_0^1 f(x)\ d\mu(x),$$

where μ *has density*

$$d\mu(x) := \frac{p+q}{2\pi}\frac{\sqrt{-(x-\lambda_-)(x-\lambda_+)}}{x(1-x)}\mathbf{1}_{[\lambda_-,\lambda_+]}\ dx,$$

and

$$\lambda_{\pm} := \left[\sqrt{\tfrac{p}{p+q}(1-\tfrac{1}{p+q})} \pm \sqrt{\tfrac{1}{p+q}(1-\tfrac{p}{p+q})}\right]^2.$$

3. If $n_1+n_2-2n=\omega(n)$ *and if* $(n_1-n)/(n_1+n_2-2n)\to\lambda$*, then*

$$\frac{1}{n}\sum_{i=1}^{n} f(\lambda_i) \to_{\mathbb{P}} f(\lambda).$$

Proof. Regardless of the scales of n_1-n and n_2-n, the limiting eigenvalue distribution can be understood by computing $A_\infty = B_\infty B_\infty^T$. (Note that on taking the β parameter to infinity, the Beta$(\beta x, \beta y)$ variable in the matrix model converges in probability to $\frac{x}{x+y}$. Replacing the Beta variables by these limits in B_β gives the matrix B_∞.)

By applying Stirling's approximation, it can be shown that there is a constant C depending only on β so that

$$\mathbb{E}\left|c_i - \sqrt{\frac{n_1-n+i}{n_1+n_2-2n+2i}}\right|^2 \le \frac{C}{i}.$$

A similar bound holds for c_i' and for $c_i s_i$. Applying all these bounds, it follows that

$$\mathbb{E}\left\|B_\beta B_\beta^T - B_\infty B_\infty^T\right\|_F^2 = O(\log n). \tag{2.4}$$

From the fundamental realization of Trotter [118], any $o(n)$ bound on the expected-square Frobenius norm suffices to show that the ESDs of two matrix models are converging together as $n\to\infty$.

It is now elementary to check that the limiting spectral distribution for $B_\infty B_\infty^T$ is that which is stated in the theorem in the sublinear and superlinear cases. In the linear case, we compute the limiting distribution by way of the Jacobi differential recurrence formula, which we do in proving Theorem 4 (see (2.42)). □

2.1.2 Our Approach

In our study of the linear scaling regime, we apply a wide array of methods, starting with the method of moments (which often boils down to path-counting), special functions (orthogonal polynomial) theory and generating functions, as well as one important result from the work of Anderson and Zeitouni [4] (more details in Section 2.4).

The study of global fluctuations of linear statistics for random matrices spans a wide literature, and covers a broad spectrum of models. We will only mention here a few works that are either closely related in scope, in model, or those that have served as inspiration for our study.

The method of moments, introduced by Wigner himself [124, 125] and used for proving central limit theorems for polynomials of Wishart matrices by Jonsson [69], has been employed with great success by Sinai and Soshnikov [108], Soshnikov [109], Péché and Soshnikov [100], etc., to obtain both central limit theorems for traces of large powers of random matrices and universality results for the fluctuations of the extremal eigenvalues in the case of Wigner and Wishart matrices. The method of moments has also been used by Dumitriu and Edelman [34] to calculate the fluctuations in the case of β-Hermite and β-Laguerre ensembles (generalizations of the Gaussian and central Wishart ensembles for $\beta = 1, 2, 4$), in the case of polynomial functions. It is also one essential ingredient in the work of Anderson-Zeitouni [4] on band matrices.

It is worth mentioning that the method of moments is formally related to the Stieltjes transform methods used by Bai and Silverstein (e.g., [7]) to calculate central limit theorems for generalized Wishart matrices; for a good reference on the methodology involved, we recommend [8].

Another method for computing fluctuations of linear statistics involves a stochastic calculus approach introduced by Cabanal-Duvillard [23] to prove a central limit theorem for Wishart matrices in the case $\beta = 2$; stochastic calculus was also used by Guionnet [59] in computing fluctuations for a class of band matrices and sample covariance matrices, and by Guionnet and Zeitouni [60] to calculate large deviations for a wide class of random matrices.

Other approaches to calculating fluctuations for linear functionals for β-ensembles include the Capitaine and Casalis work [24], which, through free probability, obtains results for both Wishart and Double Wishart matrices in the case $\beta = 2$. The later work of Kusalik, Mingo, and Speicher [76] builds on [24] and on results obtained by Mingo and Nica [94] to obtain fluctuations (second-order asymptotics) for random matrices (also in the case $\beta = 2$). Finally, Chatterjee [26] has introduced Stein's method to computing central limit theorems for a wide class of Gaussian and quasi-Gaussian matrices and entire test functions.

Specifically in the case of β-Jacobi ensembles, for an "extremal" class of β-Jacobi ensembles (when $n_1 = o(\sqrt{n_2})$ and $n = o(\sqrt{n_2})$), as mentioned before, Jiang [66] has established a series of important results, among which are the calculations of fluctuations, through approximation methods.

For all β-Jacobi ensembles of fixed parameters, Killip [73] proved that the fluctuations of *macroscopic* statistics obey a CLT; this result is similar to the one we obtain, but in the case that $f = \chi_I$ where I is a (fixed, independent of n) finite union of intervals in $[0, 1]$ and under a different normalization. It is unclear how Killip's result changes if the parameters of the ensemble scale with n, which is the regime studied here. In addition, while our method does not allow us to obtain any results for discontinuous functions, it seems that going in the opposite direction – using Killip's results to obtain fluctuation theorems for smooth functions – would need *microscopic* statistics, i.e. where the lengths of the intervals shrink with n. As Killip notes, the microscopic regime is much more difficult and is not covered in [73].

2.1.3 Our results

Our purpose is to calculate the global fluctuations for β-Jacobi ensembles, for as large a class of functions f as possible. By using concentration properties of the Jacobi ensemble, we were able to obtain the fluctuations for all β in the case of C^1 test functions on $[0,1]$. We only obtain the deviation from the mean for polynomial test functions, and conjecture the deviation should extend to a larger class of functions.

Our asymptotic analysis will occur in the proportional scaling regime, and so we will make the following assumptions on the growth of n_1 and n_2.

Assumption 1. Let $n_1 = pn$ and $n_2 = qn$ for some fixed $p, q \geq 1$ having $p + q > 2$.

We are now in a position to formulate the CLT result, which we formulate in terms of the universal results laid out earlier. From Theorem 1, we have the limiting ESD is supported on $[\lambda_-, \lambda_+]$. Thus, we define a shifted, scaled matrix $\tilde{A}$ by

$$\tilde{A} = \frac{2A - (\lambda_+ + \lambda_-)I_n}{\lambda_+ - \lambda_-},$$

chosen so that its limiting ESD is supported on $[-1, 1]$.

Theorem 2. *Let A be an $n \times n$ β-Jacobi matrix, with n_1, n_2 satisfying Assumption 1, and let $\tilde{A}$ be centered and scaled so the ESD is asymptotically supported on $[-1, 1]$. For any $f_1, f_2, \ldots, f_k$ continuously differentiable on $[-\frac{\lambda_+ + \lambda_-}{\lambda_+ - \lambda_-}, \frac{2 - \lambda_+ - \lambda_-}{\lambda_+ - \lambda_-}]$, the variable $(X_{f_1}^{\tilde{A}}, X_{f_2}^{\tilde{A}}, \ldots, X_{f_k}^{\tilde{A}})$ converges in distribution to a centered normal variable $(Y_1, \ldots, Y_k)$ with variance given by*

$$\operatorname{Cov}(Y_i, Y_j) := \tfrac{1}{2\beta} \sum_{n=1}^{\infty} n \hat{f}_i(n) \hat{f}_j(n).$$

While we state the results in this universal form, we will work directly with the unstandardized matrix model A, so that regularity condition in this theorem becomes $f \in C^1[0, 1]$. We will keep this convention throughout the proof.

Our second result concerns the deviation from the mean, and is restricted to polynomial functions.

Theorem 3. *For any polynomial* ϕ,

$$\mathbb{E}\,\mathrm{tr}(\phi(A)) = n\int_{\lambda_-}^{\lambda_+}\phi(x)d\mu(x) + (\tfrac{2}{\beta}-1)\int_{\lambda_-}^{\lambda_+}\phi(x)d\nu(x) + o(\tfrac{1}{n}),$$

where μ *is as defined in Theorem 1 and* ν *is the signed measure with density*

$$d\nu := \tfrac{1}{4}\delta_{\lambda_-} + \tfrac{1}{4}\delta_{\lambda_+} - \frac{1}{2\pi\sqrt{-(x-\lambda_+)(x-\lambda_-)}}\mathbf{1}_{(\lambda_-,\lambda_+)}\;dx.$$

The approach to the proof of Theorem 2 is as follows.

Step 1. Prove a central limit theorem for polynomials;

Step 2. Find the class of polynomials which diagonalizes the covariance matrix for the resulting Gaussian process;

Step 3. Use concentration techniques to show that $C^1[0,1]$ linear statistics can be approximated by polynomial test functions in such a way that the variance of the difference of the two is small for all n.

Step 4. Prove that the approximation works asymptotically.

The rest of the chapter is structured as follows: after a reparameterization of the model (Section 2.1.4), Section 2.2 covers Step 1 in the above recipe: show that the fluctuations are Gaussian when the test functions are the monomials. The proof extends the mechanism that was employed in [34] for the β-Hermite and β-Laguerre ensembles. In Section 2.3 we show that the limiting covariance is diagonalized in shifted Chebyhsev basis; the method employed is original and has to do with the generating function of the covariance matrix. Section 2.4 contains the proof that the matrix model satisfies the necessary conditions to apply the Anderson-Zeitouni theorem. Section 2.5 contains the proof of Theorem 3 (calculating the deviation from the mean for analytic functions). Section 2.6 contains experimental results for the case that $p = q = 1$. Section 2.7 contains the symmetric function theory results necessary for the calculation of the deviation (Section 2.5); more explicitly, it contains the proof that the series expansion of the functional $\mathcal{F}(A)$ for monomials has a "palindromic" quality (the mechanism here is similar to the one employed in [34]). Section 2.8 shows the existence of a Poincaré inequality for Beta variables that is stronger than what can be proven using general log-concave theory. Finally Section 2.9 shows a theorem of independent interest, which we proved in the course of an unsuccessful attempt to obtain our main result by a different approximation method: that "square root of beta" variables can be coupled to Gaussian variables in such a way as to have small variance.

2.1.4 Reparameterization

While the parameters given naturally arise in the full matrix model as the size ratios of the two Wishart matrices involved, we choose to work with a slightly different set of parameters for the purposes of this problem. Define parameters a and b by

$$a := \frac{1}{p+q} \quad \text{and} \quad b := \frac{p}{p+q}.$$

As we shall see, a and b allow us to express the results in a cleaner form. They also expose symmetries of the asymptotics, which are invariant under the involution $a \mapsto 1-b$, $b \mapsto 1-a$. This can be anticipated as these parameters would arise naturally in the Haar projection model, where $q \sim an$ and $\tilde{q} \sim bn$ for $n \times n$ projections.

For the regime of consideration of Theorem 2 the parameters a and b take on values in the triangle $0 < a < \frac{1}{2}$, and $a < b < 1-a$. The limiting spectral distribution will have support given by

$$\lambda_{\pm} = \left[\sqrt{b(1-a)} \pm \sqrt{a(1-b)}\right]^2.$$

The reciprocal expression $\frac{2}{\beta}$ appears frequently, with some terms having polynomial dependence upon it. Thus in the proofs we have used α in place of $\frac{2}{\beta}$. The Jacobi ensemble density, with these parameters, is expressed as

$$d\mu_J(\lambda_1,\ldots,\lambda_n) = \tfrac{1}{Z}\prod_i \lambda_i^{\frac{n}{\alpha}\left[\frac{b}{a}-1\right]+\frac{1}{\alpha}-1}(1-\lambda_i)^{\frac{n}{\alpha}\left[\frac{1-b}{a}-1\right]+\frac{1}{\alpha}-1}\prod_{i<j}|\lambda_i-\lambda_j|^{\frac{2}{\alpha}}. \qquad (2.5)$$

The tridiagonal matrix model with these parameters is given $A = B_\beta B_\beta^t$ where

$$B_\beta = \begin{pmatrix} c_n s'_{n-1} & & & & \\ -s_{n-1}c'_{n-1} & c_{n-1}s'_{n-2} & & & \\ & -s_{n-2}c'_{n-2} & c_{n-2}s'_{n-3} & & \\ & & \ddots & \ddots & \\ & & & -s_1 c'_1 & c_1 \end{pmatrix},$$

$$\begin{aligned} &\left\{c_1, c_2, \ldots, c_n, c'_1, c'_2, \ldots, c'_{n-1}\right\} \text{ mutually independent,} \\ &c_i \sim \sqrt{\mathrm{Beta}(\tfrac{nb}{\alpha a} + \alpha^{-1}(i-n), \tfrac{n(1-b)}{\alpha a} + \alpha^{-1}(i-n)}, \\ &c'_i \sim \sqrt{\mathrm{Beta}(\alpha^{-1}i, \tfrac{n}{\alpha a} + \alpha^{-1}(i-2n+1)}, \\ &s_i = \sqrt{1-c_i^2} \text{ and } s'_i = \sqrt{1-c_i'^2}. \end{aligned} \qquad (2.6)$$

2.2 Polynomial Fluctuations

2.2.1 Traces of powers and path counting

When the test function f is a polynomial, the linear statistic X_f^A can be computed explicitly using powers of the matrix model. By linearity, this reduces to the study of monomials $\mathrm{tr}(A^k)$, and by the tridiagonality of A, there is a simple combinatorial expansion for this trace. In particular, these traces can be expressed in terms of certain lattice paths. In this section we will study these lattice paths and develop their combinatorial properties. We will use these combinatorial properties to compute the covariance of the limiting Gaussian process for polynomial test functions. Their properties are not needed for the proof that the limiting fluctuations are Gaussian.

Definition 1. An *alternating bridge* is a lattice path from $(0,0)$ to $(2k,0)$ using only the steps $(1,1)$, $(1,0)$, and $(1,-1)$ none of whose odd steps travel up and none of whose even steps travel down. Let $\mathcal{A}_{2k}$ denote the collection of all such lattice paths. Likewise, let $\mathcal{L}_k$ denote the collection of all lattice paths of length k without the alternating property.

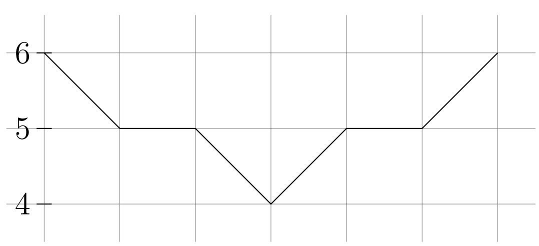
Remark 1. These paths bear some similarity to the *alternating Motzkin Paths* which have been used to study the Laguerre Ensemble [33]. These paths differ in that Motzkin paths are restricted to stay above the x-axis, while these are allowed to go above and below the axis.

For a lattice path $\bar{w}$ starting at $(0, k)$ with sequence of vertical coordinates $\{w_0 = k, w_1, w_2, \ldots\}$ and an $n \times n$ matrix M, define $M_{\bar{w}}$ to be the product

$$M_{\bar{w}} = \prod_{i=0} M_{w_{2i},w_{2i+1}} M^T_{w_{2i+1},w_{2i+2}} = \prod_{i=0} M_{w_{2i},w_{2i+1}} M_{w_{2i+2},w_{2i+1}},$$

provided that all $n \leq w_i \leq 1$. If the lattice path $\bar{w}$ walks off the edge of the matrix, in the sense that either some $w_i > n$ or $w_i < 1$, then define $M_{\bar{w}} = 0$.

Example 1. A lattice path $\bar{w}$ and its associated product $M_{\bar{w}}$.



Provided the matrix M is at least 6×6, this lattice path $\bar{w}$ would produce the product $M_{\bar{w}} = M_{6,5} M^T_{5,5} M_{5,4} M^T_{4,5} M_{5,5} M^T_{5,6}$.

Expanding the trace,

$$\operatorname{tr} A^k = \sum_{i=1}^{n} \left[\left(B_\beta B_\beta^T \right)^k \right]_{i,i}.$$

The diagonal entries $[(B_\beta B_\beta^T)^k]_{i,i}$ can be written in terms of alternating bridges, since for all $1 \leq i \leq n$,

$$\left[\left(B_\beta B_\beta^T \right)^k \right]_{i,i} = \sum_{\bar{w} \in \mathcal{A}_{2k}} (B_\beta)_{\bar{w}+i},$$

where $\bar{w} + i$ is the lattice path $\bar{w}$ shifted up by i. For convenience, define $\tilde{\mathcal{A}}_{2k,n}$ to be all alternating bridges that are shifted up to start at coordinates between 1 and n; we will refer to these lattice paths as *tridiagonal trace paths*. In terms of these paths, we can write the trace of a power of a matrix as

$$\operatorname{tr} A^k = \sum_{\bar{w} \in \tilde{\mathcal{A}}_{2k,n}} A_{\bar{w}}.$$

When n is large and k is fixed, each $A_{\bar{w}}$ is approximated by a substantially simpler quantity: every entry in a $2k \times 2k$ principal submatrix on the diagonal of A is strongly approximated by a deterministic tridiagonal band matrix (c.f. Lemmas 8 and 8). Thus, endow an alternating bridge with a weight by giving each horizontal edge weight x and each inclined edge weight y. Define the weight of the bridge to be the product of the weights of its edges, and define $p_k(x, y)$ to be the sum of all the weights over all the paths in $\mathcal{A}_{2k}$. If we let $h(\bar{w})$ denote the number of horizontal steps taken by path $\bar{w}$, then

$$p_k(x, y) = \sum_{\bar{w} \in \mathcal{A}_{2k}} x^{h(\bar{w})} y^{2k - h(\bar{w})}.$$

We are interested in finding the exponential generating function for these p_k, i.e. we will compute

$$\mathscr{P}(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} p_k(x, y),$$

and show that

$$\mathscr{P}(t) = e^{t(x^2+y^2)} I_0(2xyt), \tag{2.7}$$

where I_0 is the modified Bessel function of the first kind.

These polynomials exhibit some nice combinatorial properties. Suppose that a path $\bar{w} \in \mathcal{A}_{2k}$ has i up-steps. Because the path returns to 0, it must also have i down-steps. Down-steps must be placed in odd positions, and up-steps must be placed in even positions; as a result, the placement of the up-steps is independent from the placement of the down-steps. Thus, there are exactly $\binom{k}{i}\binom{k}{k-i}$ paths in $\mathcal{A}_{2k}$

having $2i$ inclined steps. Note, this argument also shows that the number of inclined steps must be even. Consequently, the number of horizontal steps is even as well, and we have shown

$$p_k(x,y) = \sum_{l=0}^{k} \binom{k}{l}^2 x^{2l} y^{2(k-l)} = y^{2k}{}_2F_1(-k,-k,1;(x/y)^{2k}). \tag{2.8}$$

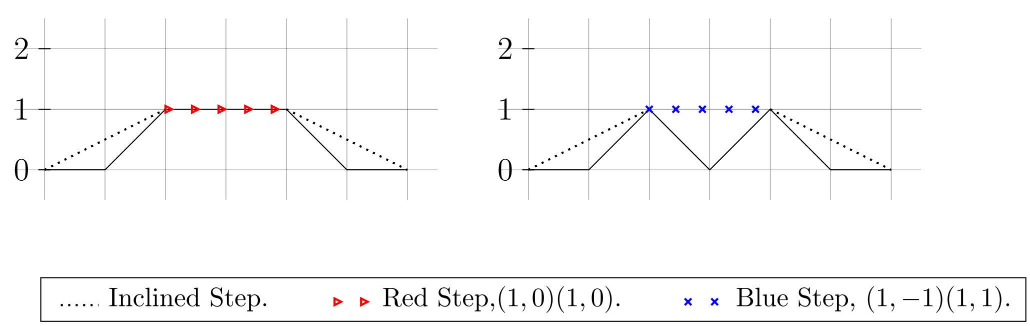
For definitions and properties of the hypergeometric function ${}_2F_1$, see [1, page 556]. As a consequence, we are able to compute the size of $\mathcal{A}_{2k}$ by simply evaluating this polynomial at $x = y = 1$,

$$|\mathcal{A}_{2k}| = \sum_{i=0}^{k} \binom{k}{i}^2 = \binom{2k}{k}.$$

While the alternating structure naturally lends itself to describing traces of A, there is another way to view $\mathcal{A}_{2k}$ which lends itself better to computing $\mathscr{P}(t)$. If $\bar{w} = w_1 w_2 \cdots w_{2k-1} w_{2k}$, for steps w_i, then the concatenation of the steps $w_{2i-1}w_{2i}$ is one of $(2,1)$, $(2,-1)$ or $(2,0)$. Moreover, if it is either of the first two, then by the alternating structure, $w_{2i-1}w_{2i}$ must have been $(1,0)(1,1)$ or $(1,-1)(1,0)$ respectively. If it was a horizontal step, then there are two possibilities, either $(1,0)(1,0)$ or $(1,-1)(1,1)$.

Definition 2. By concatenating pairs of steps, alternating bridges $\bar{w}$ are in bijective correspondence with lattice paths in $\mathcal{L}_k$ *whose horizontal steps are* 2*-colored.* Let those horizontal steps corresponding to $(1,0)(1,0)$ be colored *red*, and let those horizontal steps corresponding to $(1,-1)(1,1)$ be colored *blue.*

Example 2. Two alternating bridges with the overlaid $\mathcal{L}_k$ path.



Lemma 3. *Let $\bar{w} \in \tilde{\mathcal{A}}_{2k,n}$ be given, and define $S_{\uparrow}^{\bar{w}}(m)$ to be the number of times $\bar{w}$ walks from height m to height $m+1$ or back, and let $S_{\rightarrow}^{\bar{w}}(m)$ be the number of times that $\bar{w}$ walks horizontally at height m. Both $S_{\rightarrow}^{\bar{w}}(m)$ and $S_{\uparrow}^{\bar{w}}(m)$ are even.*

Proof. Let $\bar{u}$ be the colored lattice path from $\mathcal{L}_k$ that corresponds to $\bar{w}$. Let v be the number of steps that $\bar{u}$ makes between height m and height $m+1$ and back. Because $\bar{u}$ returns to its starting height, v is even. Let R be the number of red horizontal steps (i.e. those resulting from a $(1,0)(1,0)$ pattern) that $\bar{u}$ makes at height m, and let B be the number of blue horizontal steps (those resulting from a $(1,-1)(1,1)$ pattern) that $\bar{u}$ makes at height $m+1$. Because $S_{\rightarrow}^{\bar{w}}(m) = v + 2R$ and $S_{\uparrow}^{\bar{w}}(m) = v + 2B$, both are always even. □

The correspondence between colored $\mathcal{L}_k$ and $\mathcal{A}_{2k}$ allows the polynomials $p_k(x,y)$ to be represented in a third way. We will define the weight of an *uncolored* path $p \in \mathcal{L}_k$ to equal the sum of the weights over all alternating bridges $\bar{w}$ to which its colorings correspond. Suppose that an alternating bridge $\bar{w}$ is in correspondence with a colored path p, one with r red edges and b blue edges. Recall that $h(p)$ is the number of horizontal steps the path takes, and therefore the weight of $\bar{w}$ is $(xy)^{k-h(p)}x^{2r}y^{2b}$. There are $\binom{h(p)}{r}$ ways of placing the r red edges on the path (after which the placement of the b blue edges is determined). As the possible colorings of a fixed path p are in bijective correspondence with $\{1,0\}^{h(p)}$, it follows that the sum of the weights corresponding to all different colorings of a given path p is $(xy)^{k-h(p)}(x^2+y^2)^{h(p)}$. In conclusion, $p_k(x,y)$ can be written as

$$p_k(x,y) = \sum_{p \in \mathcal{L}_k} (xy)^{k-h(p)}(x^2+y^2)^{h(p)}.$$

The subset of the lattice paths $\mathcal{L}_k$ that fixes a given horizontal edge is in bijective correspondence with $\mathcal{L}_{k-1}$, simply by removing the given edge. By inclusion-exclusion, it follows immediately that the lattice paths in $\mathcal{L}_k$ that have no horizontal steps are counted by

$$|\mathcal{L}_k| - \binom{k}{1}|\mathcal{L}_{k-1}| + \binom{k}{2}|\mathcal{L}_{k-2}| - \binom{k}{3}|\mathcal{L}_{k-3}| + \cdots = \begin{cases} \binom{k}{\frac{k}{2}} & k \text{ even} \\ 0 & k \text{ odd .} \end{cases}$$

The correspondence between $\mathcal{L}_k$ with a fixed horizontal edge and $\mathcal{L}_{k-1}$ decreases the statistic $h(p)$ by exactly 1, and so this inclusion-exclusion formula carries over to p_k as

$$p_k(x,y) - \binom{k}{1}(x^2+y^2)p_{k-1}(x,y) + \binom{k}{2}(x^2+y^2)^2 p_{k-2}(x,y) + \cdots$$
$$= \begin{cases} (xy)^k\binom{k}{\frac{k}{2}} & k \text{ even} \\ 0 & k \text{ odd .} \end{cases}$$

This recurrence can be recast in terms of the exponential generating function $\mathscr{P}(t)$ to read

$$\mathscr{P}(t)e^{-t(x^2+y^2)} = 1 + \frac{x^2y^2t^2}{2!}\binom{2}{1} + \frac{x^4y^4t^4}{4!}\binom{4}{2} + \frac{x^6y^6t^6}{6!}\binom{6}{3} + \cdots = I_0(2xyt).$$

Thus, we have shown (2.7),

$$\mathscr{P}(t) = e^{t(x^2+y^2)}I_0(2xyt).$$

Working with this function proves to be somewhat complicated, and it will be convenient to instead use the Laplace transform of $\mathscr{P}(t)$. Let $\mathcal{L}_t[f(t)](\omega)$ denote the Laplace transform in the variable t

$$\mathcal{L}_t[f(t)](\omega) = \int_0^\infty e^{-\omega t}f(t)dt.$$

When applicable, $\mathcal{L}_{s,t}$ will denote the Laplace transform in both variables. The calculation of the Laplace transform of $\mathscr{P}(t)$ is simplified greatly by some elementary properties of the Laplace transform and the known Laplace transforms of modified Bessel functions. All of these properties are available for reference in [1, Chapter

29]; properties of the modified Bessel functions are available in [1, Chapter 9]. The Laplace transform of the modified Bessel functions I_n is given by

$$\mathcal{L}_t[I_n(ct)](\omega) = \frac{c^n}{\left(\omega + \sqrt{\omega^2 - c^2}\right)^n} \frac{1}{\sqrt{\omega^2 - c^2}}, \qquad \omega > c. \tag{2.9}$$

If for some real value of ω_0, the Laplace transform is finite, then for any ω in the half plane $\Re\omega > \omega_0$, the Laplace transform is finite. Further, the transform satisfies the following identities

$$\mathcal{L}_t[e^{kt} f(t)](\omega) = \mathcal{L}_t[f(t)](\omega - k), \tag{2.10}$$

$$\mathcal{L}_t[tf(t)](\omega) = -\frac{d}{d\omega}\mathcal{L}_t[f(t)](\omega). \tag{2.11}$$

We will show that *a priori*, the Laplace transform of $\mathscr{P}(t)$ is finite in the half plane $\Re\omega > (x+y)^2$. This follows as $I_n(2xyt)$ satisfies the simple estimate

$$0 \leq I_n(2xyt) \leq e^{2xyt},$$

for $t > 0, 2xy > 0$, and thus

$$0 \leq \mathscr{P}(t) \leq e^{t(x+y)^2}.$$

Identity (2.10) makes computing the Laplace transform of $\mathscr{P}(t)$ a simple substitution into (2.9), as

$$\mathcal{L}_t[\mathscr{P}(t)](\omega) = \mathcal{L}_t[I_0(2xyt)](\omega - x^2 - y^2) = \frac{1}{\sqrt{(\omega - x^2 - y^2)^2 - 4x^2y^2}}.$$

Using (2.11), it is possible to compute the Laplace transform of $\partial_x \mathscr{P}(t)$, which arises later.

Lemma 4.

$$\mathcal{L}_t[\partial_x \mathscr{P}(t)](\omega) = \frac{2x(\omega + y^2 - x^2)}{\left((\omega - x^2 - y^2)^2 - 4x^2y^2\right)^{\frac{3}{2}}}, \qquad \omega > (x+y)^2.$$

Proof. This is a straightforward application of (2.10), (2.11) and the identity $I_0(t)' = I_1(t)$.

$$\begin{aligned}\mathcal{L}_t[\partial_x \mathscr{P}(t)](\omega) &= \mathcal{L}_t[2xte^{t(x^2+y^2)}I_0(2xyt) + 2yte^{t(x^2+y^2)}I_1(2xyt)](\omega)\\ &= -\partial_\omega \mathcal{L}_t[2xe^{t(x^2+y^2)}I_0(2xyt) + 2ye^{t(x^2+y^2)}I_1(2xyt)](\omega)\\ &= -\partial_\omega\left[\frac{2x}{\sqrt{\tilde{\omega}^2 - 4x^2y^2}} + \frac{2y}{\sqrt{\tilde{\omega}^2 - 4x^2y^2}}\frac{2xy}{\tilde{\omega} + \sqrt{\tilde{\omega}^2 - 4x^2y^2}}\right],\end{aligned}$$

where $\tilde{\omega}$ is $\omega - x^2 - y^2$. Thus

$$\begin{aligned}\mathcal{L}_t[\partial_x \mathscr{P}(t)](\omega) &= \frac{2x(\tilde{\omega} + 2y^2)}{(\tilde{\omega}^2 - 4x^2y^2)^{\frac{3}{2}}}\\ &= \frac{2x(\omega + y^2 - x^2)}{\left((\omega - x^2 - y^2)^2 - 4x^2y^2\right)^{\frac{3}{2}}}, \qquad \omega > (x+y)^2.\end{aligned}$$

□

Remark 2. In a manner of speaking, we have circuitously arrived at the regular generating function for $p_k(x, y)$, since it is possible to deduce the generating function from the exponential generating function by way of the Laplace transform, as follows. Let $\mathscr{P}^R(t)$ denote the generating function,

$$\mathscr{P}^R(t) = \sum_{k=0}^{\infty} t^k p_k(x, y).$$

The effect of taking the Laplace transform on an exponential generating function can be understood using the Gamma function.

$$\begin{aligned}\mathcal{L}_t[\mathscr{P}(t)](\omega) &= \int_0^\infty e^{-\omega t}\mathscr{P}(t)dt\\ &= \int_0^\infty e^{-\omega t}\sum_{k=0}^{\infty}\frac{t^k}{k!}p_k(x, y)dt.\end{aligned}$$

The order of summation and integration can be interchanged because $\frac{t^k}{k!}p_k(x, y)$ is always positive for $t > 0$, $x, y \in \mathbf{R}$,

$$\mathcal{L}_t[\mathscr{P}(t)](\omega) = \sum_{k=0}^{\infty}\int_0^\infty e^{-\omega t}\frac{t^k}{k!}p_k(x, y)dt.$$

Make the change of variables $s = \omega t$, so that

$$\begin{aligned}\mathcal{L}_t[\mathscr{P}(t)](\omega) &= \sum_{k=0}^{\infty} \int_0^{\infty} e^{-s} \frac{s^k}{\omega^{k+1} k!} p_k(x, y) ds \\ &= \sum_{k=0}^{\infty} \omega^{-k-1} p_k(x, y) ds \\ &= \mathscr{P}^R(\omega^{-1})\omega^{-1}.\end{aligned}$$

Thus, putting everything together,

$$\mathscr{P}^R(\omega) = \frac{1}{\sqrt{(1 - \omega(x^2 + y^2))^2 - 4\omega^2 x^2 y^2}}.$$

2.2.2 Asymptotic normality of fluctuations

We show in this subsection that polynomial test functions asymptotically have jointly normal fluctuations. This is the first component of Theorem 2, and we summarize the precise claim in the following proposition.

Proposition 6. *Let A be an $n \times n$ β-Jacobi matrix, with parameters as described in Section 2.1.4. For any fixed $k \in \mathbb{N}$, the k-tuple $\left(X_{x^1,A}, X_{x^2,A}, \ldots, X_{x^k,A}\right)$ converges in distribution to a centered multivariate normal random variable.*

The method of proof will be the computation of the moments. Recall that a multivariate normal variable has mixed moments characterized by the Wick formula, which we will state precisely.

Proposition 7. *A centered random vector $(Z_1, Z_2, \ldots, Z_k)$ is a multivariate normal if and only if for each word $m \in [k]^l$, the mixed moments satisfy*

$$\mathbb{E} \prod_{i \in m} Z_i = \begin{cases} 0 & \text{if } l \text{ is odd,} \\ \sum_G \prod_{\{a,b\} \in \mathcal{E}(G)} \mathbb{E} Z_{m_a} Z_{m_b} & \text{if } l \text{ is even,} \end{cases}$$

where the sum is over all graphs G that are perfect matchings on the vertices $[k]$, and where $\mathcal{E}(G)$ is the edge set of this graph.

To prove Proposition 6, it suffices to show that all the mixed moments asymptotically obey the Wick formula. Thus, our first goal is to show that the moments have the correct form.

Proposition 8. *For a fixed word $m \in [k]^l$,*

$$\mathbb{E}\prod_{i\in m} X_{x^i,A} = \begin{cases} O(n^{-1/2}) & \text{if } l \text{ is odd,} \\ \sum_G \prod_{\{a,b\}\in\mathcal{E}(G)} \mathbb{E}X_{x^{m_a},A}X_{x^{m_b},A} + O(n^{-1/2}) & \text{if } l \text{ is even,} \end{cases}$$

where the sum is over all graphs G that are perfect matchings on the vertices $[k]$, and where $\mathcal{E}(G)$ is the edge set of this graph.

This nearly proves Proposition 6, but it remains to show that the covariances have a limit. We will delay this proof as we will identify the limiting covariance explicitly, and we begin in the direction of proving Proposition 8. In the sequel, fix some word $m \in [k]^l$. We will write the mixed moment indicated by m in a way that exposes its asymptotically relevant terms. The first step is to write the mixed moment in terms of tridiagonal trace paths.

$$\begin{aligned}\mathbb{E}\prod_{u\in m} X_{x^k,A} &= \mathbb{E}\prod_{u\in m}\left[\sum\nolimits_{\bar{w}\in\tilde{\mathcal{A}}_{2u,n}} A_{\bar{w}} - \mathbb{E}A_{\bar{w}}\right] \\ &= \sum_{\bar{w}_1,\ldots,\bar{w}_l} \mathbb{E}\prod_{i=1}^{l}\left[A_{\bar{w}_i} - \mathbb{E}A_{\bar{w}_i}\right], \end{aligned} \tag{2.12}$$

where the sum is over all tridiagonal trace paths $(\bar{w}_1,\ldots,\bar{w}_l) \in \tilde{\mathcal{A}}_{2m_1,n} \times \tilde{\mathcal{A}}_{2m_2,n} \times \cdots \times \tilde{\mathcal{A}}_{2m_l,n}$.

Each nonzero random variable $A_{\bar{w}}$ is a product of terms of matrix entries. More specifically, by Lemma 3 trace paths visit each matrix entry an even number of times, and so $A_{\bar{w}}$ is a polynomial in the random variables $\{c_i^2\}$ and $\{(c_i')^2\}$. Thus for each tridiagonal trace path $\bar{w}_i$ for which $A_{\bar{w}_i} \not\equiv 0$, it is possible to define random variables $q_j^{\bar{w}_i}$ with $1 \le j \le 2n-1$ so that

1. $q_j^{\bar{w}_i}$ is a polynomial in c_j^2 for $1 \leq j \leq n$;

2. $q_{j+n}^{\bar{w}_i}$ is a polynomial in $(c_j')^2$ for $1 \leq j \leq n-1$;

3. $A_{\bar{w}_i} = \prod_{j=1}^{2n-1} q_j^{\bar{w}_i}$;

4. The smallest nonzero coefficient of each $q_j^{\bar{w}_i}$ is 1.

We will write $q_j^{\bar{w}_i}(x)$ for the corresponding polynomial in x, while when no argument is provided, we mean the random variable defined above. This decomposition breaks a random variable $A_{\bar{w}_i}$ into a product of independent random variables. Further, each polynomial has the form $q_j^{\bar{w}_i}(x) = x^{a_{i,j}}(1-x)^{b_{i,j}}$ for some non-negative integer powers. Note, however, that most of these polynomials are identically 1.

We will use these polynomials to alternately express the difference $A_{\bar{w}} - \mathbb{E}A_{\bar{w}}$. Specifically, we telescope in the following way.

$$
\begin{aligned}
A_{\bar{w}_i} - \mathbb{E}A_{\bar{w}_i} &= \prod_{j=1}^{2n-1} \left[(q_j^{\bar{w}_i} - \mathbb{E}q_j^{\bar{w}_i}) + \mathbb{E}q_j^{\bar{w}_i} \right] - \prod_{j=1}^{2n-1} \mathbb{E}q_j^{\bar{w}_i} \\
&= \sum_{\substack{S \subset [2n-1] \\ S \neq \emptyset}} \left[\prod_{j \in S} (q_j^{\bar{w}_i} - \mathbb{E}q_j^{\bar{w}_i}) \prod_{j \notin S} \mathbb{E}q_j^{\bar{w}_i} \right]. \qquad (2.13)
\end{aligned}
$$

In this last step we omit the empty set precisely because it is the term canceled by $\mathbb{E}A_{\bar{w}_i}$.

Note that in (2.12) we require a product of l of these terms. Thus, by applying the (2.13) multiple times, we can write

$$
\prod_{i=1}^{l} [A_{\bar{w}_i} - \mathbb{E}A_{\bar{w}_i}] = \sum_{S_1 \ldots S_l} \prod_{i=1}^{l} \prod_{j \in S_i} (q_j^{\bar{w}_i} - \mathbb{E}q_j^{\bar{w}_i}) \prod_{j \notin S_i} \mathbb{E}q_j^{\bar{w}_i}, \qquad (2.14)
$$

where it is important to note that the sum is over *nonempty* subsets of $[2n-1]$.

In expectation, we will see that each difference term $q_j^{\bar{w}_i} - \mathbb{E}q_j^{\bar{w}_i}$ that appears in the product contributes a factor of $n^{-1/2}$, and thus that the magnitude of (2.14) is at most $O(n^{-l/2})$. To show this, we require the ability to estimate moments of the terms that appear in the right hand side of (2.14). This is expressed in the following lemma.

Lemma 5. *Fix a polynomial $q(x) = x^{a_1}(1-x)^{a_2}$, and fix an $n \in \mathbb{N}$. There is a constant $C = C(m, a_1, a_2)$ so that*

$$\max_{1 \le i \le n} \mathbb{E}\left|q(c_i^2) - \mathbb{E}q(c_i^2)\right|^m \le Cn^{-m/2}, \text{ and}$$
$$\max_{1 \le i \le n-1} \mathbb{E}\left|q((c_i')^2) - \mathbb{E}q((c_i')^2)\right|^m \le Cn^{-m/2}.$$

Proof. In the current parameterization, we recall that c_i^2 and $(c_i')^2$ are mutually independent Beta random variables with parameters

$$c_i^2 \sim \mathrm{Beta}(\tfrac{nb}{\alpha a} + \alpha^{-1}(i-n), \tfrac{n(1-b)}{\alpha a} + \alpha^{-1}(i-n)), \text{ and}$$
$$(c_i')^2 \sim \mathrm{Beta}(\alpha^{-1}i, \tfrac{n}{\alpha a} + \alpha^{-1}(i-2n+1)).$$

The primary tool in this proof is the Poincaré inequality for Beta random variables. From Lemma 16, a Beta variable $X \sim \mathrm{Beta}(p_1, p_2)$ satisfies a Poincaré inequality

$$\mathrm{Var}\, f(X) \le \frac{1}{4(p_1+p_2)} \mathbb{E}\left|f'(X)\right|^2,$$

for any Lipschitz function f on $[0,1]$. Let $\mathcal{M}$ denote the collection of all Beta variables appearing in the matrix model. We note that for all these variables, the sum of their parameters is at least $\frac{n}{\alpha}\left[\frac{1}{a} - 2\right]$. By hypothesis on the parameters of the matrix, $a < 1/2$, and thus there is a constant C so that

$$\max_{X \in \mathcal{M}} \sup_{\|f\|_{Lip} < \infty} \left[\frac{\mathrm{Var}\, f(X)}{\mathbb{E}\left|f'(X)\right|^2}\right] \le \frac{C}{n}.$$

Further, by applying each of these inequalities to $q(X)$ for any $X \in \mathcal{M}$, we see that for any Lipschitz f,

$$\mathrm{Var}\, f(q(X)) \le \frac{C}{n} \mathbb{E}\left|f'(q(X))q'(X)\right|^2.$$

Note that $|q'(x)| \leq (a_1 + a_2)$ on $[0,1]$, and thus

$$\operatorname{Var} f(q(X)) \leq \frac{C(a_1+a_2)^2}{n} \mathbb{E}\,|f'(q(X))|^2\,,$$

for all Lipschitz functions on the interval and any $X \in \mathcal{M}$. It is well known that a Poincaré inequality implies exponential integrability (see [19]). Precisely,

$$\mathbb{E} \exp\left[\frac{|g(X) - \mathbb{E}g(X)|\,\sqrt{n}}{12(a_1+a_2)\sqrt{C}}\right] \leq 2,$$

for every $X \in \mathcal{M}$. By expanding the exponential in its series, the claim follows. □

As a consequence of Lemma 5, it is possible to estimate the contribution of any product of terms as in (2.14).

Lemma 6. *There is a constant $C = C(l, \max_{1\leq i\leq l} m_i)$ so that for any l-tuple $(\bar{w}_1, \ldots, \bar{w}_l) \in \tilde{\mathcal{A}}_{2m_1,n} \times \tilde{\mathcal{A}}_{2m_2,n} \times \cdots \times \tilde{\mathcal{A}}_{2m_l,n}$,*

$$\left|\mathbb{E} \prod\nolimits_{i=1}^{l} [A_{\bar{w}_i} - \mathbb{E}A_{\bar{w}_i}]\right| \leq Cn^{-l/2}.$$

Furthermore, the dominant contribution is given by

$$D_{(\bar{w}_i)_i} := \sum_{s_1\ldots s_l} \prod_{i=1}^{l} \left[(q_{s_i}^{\bar{w}_i} - \mathbb{E}q_{s_i}^{\bar{w}_i}) \prod_{j\neq s_i} \mathbb{E}q_j^{\bar{w}_i}\right],$$

with the sum over all l-tuples $(s_1, \ldots, s_l) \in [l]^{2n-1}$, and

$$\left|\mathbb{E} \prod_{i=1}^{l} [A_{\bar{w}_i} - \mathbb{E}A_{\bar{w}_i}] - \mathbb{E}D_{(\bar{w}_i)_i}\right| \leq Cn^{-(l+1)/2}.$$

Proof. We recall (2.14):

$$\prod_{i=1}^{l} [A_{\bar{w}_i} - \mathbb{E}A_{\bar{w}_i}] = \sum_{S_1\ldots S_l} \prod_{i=1}^{l} \prod_{j\in S_i} (q_j^{\bar{w}_i} - \mathbb{E}q_j^{\bar{w}_i}) \prod_{j\notin S_i} \mathbb{E}q_j^{\bar{w}_i},$$

where the sum is over nonempty subsets $S_i \subset [2n-1]$. Taking expectations, most of these of summands will be 0. This is because for each word $\bar{w}_i$, there are at most $4m_i$ nontrivial polynomials $q_j^{\bar{w}_i}$, where $\bar{w}_i \in \tilde{\mathcal{A}}_{2m_i,n}$. Thus, there are at most $2^{4m_1}2^{4m_2}\cdots 2^{4m_l}$ nonzero summands of the form

$$P_{S_1,\ldots,S_l} := \mathbb{E}\prod_{i=1}^{l}\prod_{j\in S_i}(q_j^{\bar{w}_i} - \mathbb{E}q_j^{\bar{w}_i})\prod_{j\notin S_i}\mathbb{E}q_j^{\bar{w}_i}, \tag{2.15}$$

and thus it suffices to show the desired bound for an arbitrary term such as this. From each S_i, pick an arbitrary j_i. Each $q_j^{\bar{w}_i}$ is a random variable supported on $[0,1]$, and thus both $|q_j^{\bar{w}_i} - \mathbb{E}q_j^{\bar{w}_i}| \leq 1$ and $|\mathbb{E}q_j^{\bar{w}_i}| \leq 1$. Therefore, the term in (2.15) can be bounded by

$$|P_{S_1,\ldots,S_l}| \leq \mathbb{E}\left|\prod_{i=1}^{l}(q_{j_i}^{\bar{w}_i} - \mathbb{E}q_{j_i}^{\bar{w}_i})\right| \leq \frac{1}{l}\sum_{i=1}^{l}\mathbb{E}\left|q_{j_i}^{\bar{w}_i} - \mathbb{E}q_{j_i}^{\bar{w}_i}\right|^l,$$

where we have applied the arithmetic-geometric mean inequality. By applying Lemma 5, we conclude that there is a constant C that depends only on $\max_{1\leq i\leq l} m_i$ and l so that

$$|P_{S_1,\ldots,S_l}| \leq Cn^{-l/2}.$$

Summing over all possible nonzero summands, the first conclusion follows. Note that the same argument shows that if $\sigma := |S_1| + |S_2| + \cdots + |S_l| > l$, then the same argument (with the same constant no less) shows

$$|P_{S_1,\ldots,S_l}| \leq Cn^{-\sigma/2},$$

from which the second conclusion follows. □

Having established these bounds, we introduce the notion of a dependency graph.

Definition 3. For any tuple of tridiagonal trace paths $(\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_l)$, define the *dependency graph* $\mathcal{G}$ to be a graph with vertex set $[l]$ and $i \not\leftrightarrow j$ if and only if $A_{\bar{w}_i}$ and $A_{\bar{w}_j}$ are functions of mutually independent random variables.

The family of vector variables

$$\Xi := \left\{(A_{\bar{w}_i})_{i\in S}\right\}_S,$$

where S ranges over all connected components of $\mathcal{G}$, is a mutually independent family of random variables. The importance of these connected components is that there are very few l-tuples of tridiagonal trace paths that have few connected components in their dependency graph. Moreover, it is possible to estimate exactly how many trace paths have such dependency graphs. This motivates the following definition.

Definition 4. For any $\chi \in \{1, 2, \ldots, \lfloor l/2 \rfloor\}$, let $\mathcal{B}_\chi$ be the collection of all l-tuples in $\tilde{\mathcal{A}}_{2m_1,n} \times \tilde{\mathcal{A}}_{2m_2,n} \times \cdots \times \tilde{\mathcal{A}}_{2m_l,n}$ whose dependency graphs have χ connected components and no isolated vertices. For any such word tuple of words, let $\mathcal{E} = \mathcal{E}(\bar{w}_1, \ldots, \bar{w}_l)$ denote the edge set of the dependency graph.

When l is even, $\mathcal{B}_{l/2}$ is the collection of all l-tuples of trace paths whose dependency graphs are perfect matchings. With this definition, we can count the number of l-tuples of trace paths having a particular number of connected components.

Lemma 7. *For any $\chi \in \mathbb{N}$, there is a constant $C = C(\chi, \max_{1 \le i \le l} m_i)$ so that $|\mathcal{B}_\chi| \le Cn^\chi$.*

Proof. This ultimately stems from the observation that there are only finitely many entries in the matrix that depend on a given entry. Thus, once any arbitrary trace path in a connected component has been chosen, the remainder of the trace paths must start nearby. Formally, we begin by bounding the number of ways to construct a connected component on s vertices.

Without loss of generality, suppose these s-tuples are chosen from $\tilde{\mathcal{A}}_{2m_1,n} \times \tilde{\mathcal{A}}_{2m_2,n} \times \cdots \times \tilde{\mathcal{A}}_{2m_s,n}$. As we would like choices having a connected dependency graph, we overcount by first choosing a desired spanning tree and then filling out the graph. As there are only s^{s-2} such spanning trees, we lose at most a constant factor.

Let $M = \max_{1 \le i \le l} m_i$, and choose the first trace path in the tuple arbitrarily; there are $\left|\tilde{\mathcal{A}}_{2m_1,n}\right|$ possible choices for this path. Traversing the vertices of the tree in a depth first search, each vertex traversed must depend on the previously chosen path $\bar{w}_{prev} \in$

$\tilde{\mathcal{A}}_{2m_{prev},n}$. This forces the choice of $\bar{w}_{new} \in \tilde{\mathcal{A}}_{2m_{new},n}$ to have that $A_{\bar{w}_{new}}$ depends on $A_{\bar{w}_{prev}}$, and thus the starting point of $\bar{w}_{new}$ must be no more than $m_{new} + m_{prev}$ steps from the starting point of the previous. Thus there are at most $4M\,|\mathcal{A}_{2M}|$ ways to choose the new path. This bound holds for every vertex explored in the depth first search, and we arrive at the bound that there are at most $[4M\,|\mathcal{A}_{2M}|]^s \cdot n$ ways to choose trace paths having dependency graph spanned by a given tree.

Summing over all possible partitions of l with χ parts, i.e. all multisets of naturals $\{s_i\}$ so that $s_1 + s_2 + \cdots + s_\chi = l$, and choosing components of these sizes for each, we arrive at the bound that there is a constant C so that $|\mathcal{B}_\chi| \leq Cn^\chi$.

□

It is now possible to identify the asymptotically relevant portions of an arbitrary mixed moment, and hence prove Proposition 8.

Proof of Proposition 8. In terms of the notation B_χ, we recall (2.12) and rewrite it as

$$\begin{aligned}\mathbb{E}\prod_{u\in m} X_{x^k,A} &= \sum_{\bar{w}_1,\ldots,\bar{w}_l} \mathbb{E}\prod_{i=1}^{l}\left[A_{\bar{w}_i} - \mathbb{E}A_{\bar{w}_i}\right] \\ &= \sum_{\chi=1}^{\lfloor l/2\rfloor} \sum_{(\bar{w}_1,\ldots,\bar{w}_l)\in B_\chi} \mathbb{E}\prod_{i=1}^{l}\left[A_{\bar{w}_i} - \mathbb{E}A_{\bar{w}_i}\right], \qquad (2.16)\end{aligned}$$

noting that this sum contains no l-tuples of words with isolated vertices in their dependency graphs, as these vanish identically on taking expectations. By Lemma 6, there is a constant C_1 sufficiently large that

$$\left|\mathbb{E}\prod_{i=1}^{l}\left[A_{\bar{w}_i} - \mathbb{E}A_{\bar{w}_i}\right]\right| \leq C_1 n^{-l/2},$$

for every word in the sum. Also, by Lemma 7 there is a constant C_2 sufficiently large that for all $1 \leq \chi \leq l/2$, $|B_\chi| \leq C_2 n^\chi$. It is immediate that if l is odd, then by (2.16),

$$\left|\mathbb{E}\prod_{u\in m} X_{x^k,A}\right| \leq \sum_{\chi=1}^{\lfloor l/2\rfloor} C_1 n^{-l/2} C_2 n^\chi = O(n^{-1/2}).$$

If l is even, however, then applying the same bound to terms for which $\chi < l/2$,

$$\mathbb{E}\prod_{u\in m} X_{x^k,A} = \sum_{(\bar{w}_1,\ldots,\bar{w}_l)\in B_{l/2}} \mathbb{E}\prod_{i=1}^{l}\left[A_{\bar{w}_i} - \mathbb{E}A_{\bar{w}_i}\right] + O(n^{-1/2})$$
$$= \sum_{(\bar{w}_i)_i\in B_{l/2}} \prod_{\{a,b\}\in\mathcal{E}} \mathbb{E}\left[A_{\bar{w}_a} - \mathbb{E}A_{\bar{w}_a}\right]\left[A_{\bar{w}_b} - \mathbb{E}A_{\bar{w}_b}\right] + O(n^{-1/2}). \quad (2.17)$$

It only remains to show that the Wick word has the same form, i.e. it should be shown that

$$W := \sum_{G} \prod_{\{a,b\}\in\mathcal{E}(G)} \mathbb{E}\left[X_{x^{m_a},A}X_{x^{m_b},A}\right], \quad (2.18)$$

where G ranges over all perfect matchings of $[l]$, has the same asymptotically relevant terms as (2.17). We recall (2.12), due to which we may rewrite

$$W = \sum_{G} \prod_{\{a,b\}\in\mathcal{E}(G)} \sum_{\tilde{\mathcal{A}}_{2m_a,n}\times\tilde{\mathcal{A}}_{2m_b,n}} \mathbb{E}\left[A_{\bar{w}_a} - \mathbb{E}A_{\bar{w}_a}\right]\left[A_{\bar{w}_b} - \mathbb{E}A_{\bar{w}_b}\right],$$

where the inner sum may be taken over all pairs of l-tuples. For a fixed perfect matching G, every possible tuple $(\bar{w}_1,\ldots,\bar{w}_l)$ is represented exactly once. After commuting the inner sum and the product, we may write

$$W = \sum_{(\bar{w}_1,\ldots,\bar{w}_l)} \sum_{G} \prod_{\{a,b\}\in\mathcal{E}(G)} \mathbb{E}\left[A_{\bar{w}_a} - \mathbb{E}A_{\bar{w}_a}\right]\left[A_{\bar{w}_b} - \mathbb{E}A_{\bar{w}_b}\right].$$

As before, we may ignore l-tuples whose dependency graphs have an isolated vertex, and thus we write

$$W = \sum_{\chi=1}^{l/2} \sum_{(\bar{w}_i)_i\in B_\chi} \sum_{G} \prod_{\{a,b\}\in\mathcal{E}(G)} \mathbb{E}\left[A_{\bar{w}_a} - \mathbb{E}A_{\bar{w}_a}\right]\left[A_{\bar{w}_b} - \mathbb{E}A_{\bar{w}_b}\right].$$

We will bound the contribution of terms having $\chi < l/2$, and we note that there is a constant C_3 so that for any pairing G and any tuple of paths $(\bar{w}_i)_i$,

$$\left|\prod_{\{a,b\}\in\mathcal{E}(G)} \mathbb{E}\left[A_{\bar{w}_a} - \mathbb{E}A_{\bar{w}_a}\right]\left[A_{\bar{w}_b} - \mathbb{E}A_{\bar{w}_b}\right]\right| \le C_3 n^{-l/2},$$

which follows from applying Lemma 6. Writing $C_4 = (2l)!/2^l/l!$ for the number of perfect matchings on $[l]$, we have

$$\sum_{\chi=1}^{l/2-1} \sum_{(\bar{w}_i)_i \in B_\chi} \sum_{G} \prod_{\{a,b\} \in \mathcal{E}(G)} \left| \mathbb{E} \left[A_{\bar{w}_a} - \mathbb{E} A_{\bar{w}_a} \right] \left[A_{\bar{w}_b} - \mathbb{E} A_{\bar{w}_b} \right] \right| \\ \leq \sum_{\chi=1}^{l/2-1} C_2 n^{\chi} \cdot C_4 \cdot C_3 n^{-l/2} = O(n^{-1/2}).$$

For each tuple of words $(\bar{w}_i)_i \in B_{l/2}$, there is exactly one choice of pairing G so that so that the product is nonzero, and thus

$$W = \sum_{(\bar{w}_i)_i \in B_{l/2}} \prod_{\{a,b\} \in \mathcal{E}} \mathbb{E} \left[A_{\bar{w}_a} - \mathbb{E} A_{\bar{w}_a} \right] \left[A_{\bar{w}_b} - \mathbb{E} A_{\bar{w}_b} \right] + O(n^{-1/2}),$$

which completes the proof on comparison with (2.17). □

2.2.3 Computing the covariance

We now turn to showing that all possible the pairwise covariances $\mathrm{Cov}(X_{x^k,A}, X_{x^l,A})$ have limits and produce an expression for that limiting covariance. We will use $C_{k,l}$ to denote the covariance we eventually show to be the limit. These covariances can be described in terms of the polynomials $p_k(x, y)$ introduced in Section 2.2.1. The exact form of the covariance is given by an integral against a parameter σ. In terms of σ, define the expressions

$$x := \frac{\sqrt{(b+\sigma)(1-a+\sigma)}}{1+2\sigma}, \quad \text{and} \quad y := \frac{\sqrt{(1-b+\sigma)(a+\sigma)}}{1+2\sigma}. \tag{2.19}$$

The matrix $C_{k,l}$ for $k, l \geq 1$ can now be defined by

$$C_{k,l} := \frac{\alpha}{4} \int_{-a}^{0} \frac{1}{1+2\sigma} \left[\left(\partial_x p_k \partial_x p_m + \partial_y p_k \partial_y p_m \right) \left(1 - x^2 - y^2\right) \right. \\ \left. - \left(\partial_x p_k \partial_y p_m + \partial_y p_k \partial_x p_m \right) (2xy) \right] d\sigma. \tag{2.20}$$

Remark 3. In this form, the integrand is separated into positive and negative parts. We can check that $x^2 + y^2 < 1$ for all $-a \le \sigma \le 0$. Furthermore, because p_k have all positive coefficients, x is nonnegative, and y is nonnegative, it follows that

$$\left(\partial_x p_k \partial_x p_m + \partial_y p_k \partial_y p_m\right)\left(1 - x^2 - y^2\right) \ge 0 \text{ , and}$$
$$\left(\partial_x p_k \partial_y p_m + \partial_y p_k \partial_x p_m\right)(2xy) \ge 0,$$

for all $-a \le \sigma \le 0$. To check that $x^2 + y^2 < 1$, we clear the denominator and expand the terms to show that this is equivalent to

$$b(1-a) + (1-b)a < 1 + 2\sigma + 2\sigma^2.$$

The quadratic on the right is increasing for $-1/2 < \sigma < 0$, and thus to show the inequality, it suffices to show that

$$b(1-a) + (1-b)a = a + b(1-2a) < 1 - 2a + 2a^2.$$

Using that $1 - 2a > 0$ and $b < 1 - a$, the inequality follows.

Our primary purpose in this section is to prove the following Proposition.

Proposition 9. *For each fixed $k, l \in \mathbb{N}$, as $n \to \infty$,*

$$\operatorname{Cov}(X_{x^k,A}, X_{x^l,A}) = \mathbb{E}\left[X_{x^k,A} X_{x^l,A}\right] = C_{k,l} + O(n^{-1/2}).$$

Note that combining this Proposition with Proposition 8, we have proven Proposition 6. We turn immediately towards proving Proposition 9. We recall that by (2.12), we have

$$\mathbb{E}\left[X_{x^k,A} X_{x^l,A}\right] = \sum_{\bar{w}_k, \bar{w}_l} \mathbb{E}\left[A_{\bar{w}_k} - \mathbb{E}A_{\bar{w}_k}\right]\left[A_{\bar{w}_l} - \mathbb{E}A_{\bar{w}_l}\right].$$

By Lemma 7, there is a constant K_χ so that there are at most $K_\chi \cdot n$ such words. Applying the second portion of Lemma 6, we have that there is a constant $K_{k \vee l}$ so that

$$\left|\mathbb{E}\left[X_{x^k,A} X_{x^l,A}\right] - \sum\nolimits_{\bar{w}_k, \bar{w}_l} \mathbb{E} D_{(\bar{w}_k, \bar{w}_l)}\right| \le K_\chi n \cdot K_{k \vee l} \cdot n^{-3/2},$$

where we recall that $D_{(\bar{w}_k,\bar{w}_l)}$ is given by

$$D_{(\bar{w}_k,\bar{w}_l)} = \sum_{s_k,s_l} \prod_{i\in\{k,l\}} \left[(q_{s_i}^{\bar{w}_i} - \mathbb{E}q_{s_i}^{\bar{w}_i}) \prod\nolimits_{j\neq s_i} \mathbb{E}q_j^{\bar{w}_i} \right],$$

with the sum over all choices of $s_k, s_l \in [2n-1]$. Thus, it suffices to analyze the quantity $\sum_{\bar{w}_k,\bar{w}_l} \mathbb{E}D_{(\bar{w}_k,\bar{w}_l)}$ and show it has the desired limit. Note that by the construction of $q_{s_i}^{\bar{w}_i}$, each of $q_{s_k}^{\bar{w}_k}$ and $q_{s_l}^{\bar{w}_l}$ are independent if $s_k \neq s_l$, and thus we have

$$\begin{aligned} &\mathbb{E}\left[X_{x^k,A}X_{x^l,A}\right] \\ &\quad = \sum_{\bar{w}_k,\bar{w}_l} \sum_{t=1}^{2n-1} \mathbb{E}\left[q_t^{\bar{w}_k} - \mathbb{E}q_t^{\bar{w}_k}\right]\left[q_t^{\bar{w}_l} - \mathbb{E}q_t^{\bar{w}_l}\right]\left[\prod\nolimits_{j\neq t} \mathbb{E}q_j^{\bar{w}_k}\mathbb{E}q_j^{\bar{w}_l}\right] + O(n^{-1/2}). \end{aligned}$$

We define r_t so that

$$r_t := \sum\nolimits_{\bar{w}_k,\bar{w}_l} \mathbb{E}\left[q_t^{\bar{w}_k} - \mathbb{E}q_t^{\bar{w}_k}\right]\left[q_t^{\bar{w}_l} - \mathbb{E}q_t^{\bar{w}_l}\right]\left[\prod\nolimits_{j\neq t} \mathbb{E}q_j^{\bar{w}_k}\mathbb{E}q_j^{\bar{w}_l}\right], \tag{2.21}$$

and note that by commuting sums in the previous equation, we have

$$\mathbb{E}\left[X_{x^k,A}X_{x^l,A}\right] = \sum_{t=1}^{2n-1} r_t + O(n^{-1/2}). \tag{2.22}$$

Let $\{z_i\}_{i=1}^{2n-1}$ be the enumeration of all the Beta variables in $\mathcal{M}$, where $z_i = (c_i)^2$ for $1 \leq i \leq n$ and $z_i = (c'_{i-n})^2$ when $n+1 \leq i \leq 2n-1$. This makes each $q_j^{\bar{w}_i}$ a polynomial in z_j. The first step in the analysis amounts to using Taylor approximation to pull the expectations inside the $q_j^{\bar{w}_i}$ polynomials.

Lemma 8. *There is a constant $K = K(k,l)$ so that for all $1 \leq t \leq 2n-1$,*

$$|r_t| \leq Kn^{-1}.$$

Moreover, it is possible to identify the dominant contribution r_t^D, which is given by

$$r_t^D := \mathrm{Var}(z_t) \sum\nolimits_{\bar{w}_k,\bar{w}_l} \left[{q_t^{\bar{w}_k}}'(\mathbb{E}z_t)\right]\left[{q_t^{\bar{w}_l}}'(\mathbb{E}z_t)\right]\left[\prod\nolimits_{j\neq t} q_j^{\bar{w}_k}(\mathbb{E}z_j)q_j^{\bar{w}_l}(\mathbb{E}z_j)\right],$$

and which has

$$\left|r_t - r_t^D\right| \leq Kn^{-3/2}.$$

Proof. The first claim follows from Lemma 5 and from the fact that the number of trace paths that depend on z_t is bounded by some $K = K(k,l)$. The second claim will follow from Taylor approximation. For any polynomial $q_j^{\bar{w}_k}$ or $q_j^{\bar{w}_l}$, it is possible to bound the maximums of the derivatives over $[0,1]$ in terms of k and l. Each polynomial has the form $q_j^{\bar{w}_i}(x) = x^{a^1_{i,j}}(1-x)^{a^2_{i,j}}$, and hence its first and second derivatives can be bounded by $a^1_{i,j} + a^2_{i,j}$ and $(a^1_{i,j} + a^2_{i,j})^2$. These parameters a^1 and a^2 can in turn be bounded by i, to yield

$$\max_{x\in[0,1]} \left| q_j^{\bar{w}_i \prime} \right| \leq 4i \quad \text{and} \quad \max_{x\in[0,1]} \left| q_j^{\bar{w}_i \prime\prime} \right| \leq (4i)^2,$$

for either $i \in \{k,l\}$. These imply that the 0^{th} order approximation has error

$$\left| q_j^{\bar{w}_i}(z_j) - q_j^{\bar{w}_i}(\mathbb{E}z_j) \right| \leq 4i \left| z_j - \mathbb{E}z_j \right|,$$

and the 1^{st} order approximation has error

$$\left| q_j^{\bar{w}_i}(z_j) - q_j^{\bar{w}_i}(\mathbb{E}z_j) - q_j^{\bar{w}_i \prime}(\mathbb{E}z_j)(z_j - \mathbb{E}z_j) \right| \leq 8i^2 \left| z_j - \mathbb{E}z_j \right|^2.$$

We recall the definition of r_t, which was given by

$$r_t = \sum\nolimits_{\bar{w}_k, \bar{w}_l} \underbrace{\mathbb{E}\left[q_t^{\bar{w}_k} - \mathbb{E}q_t^{\bar{w}_k}\right]\left[q_t^{\bar{w}_l} - \mathbb{E}q_t^{\bar{w}_l}\right]}_{\text{(i)}} \underbrace{\left[\prod\nolimits_{j\neq t} \mathbb{E}q_j^{\bar{w}_k}\mathbb{E}q_j^{\bar{w}_l}\right]}_{\text{(ii)}}.$$

Using 1^{st} order approximation for term (i), we bound

$$D_{(i)} := \left| \mathbb{E}\left[q_t^{\bar{w}_k} - \mathbb{E}q_t^{\bar{w}_k}\right]\left[q_t^{\bar{w}_l} - \mathbb{E}q_t^{\bar{w}_l}\right] - \mathbb{E}\left[q_t^{\bar{w}_k \prime}(\mathbb{E}z_t) q_t^{\bar{w}_l \prime}(\mathbb{E}z_t)(z_t - \mathbb{E}z_t)^2\right] \right|$$
$$\leq K_1 n^{-3/2}, \quad (2.23)$$

with the constant implicitly depending on k, l, and the constants assured by Lemma 5. Using the 0^{th} order approximation for term (ii), we will bound the difference between (ii) and its approximation. This will be done by replacing each z_j by $\mathbb{E}z_j$ one term at a time. As there are at most $2k + 2l$ non-constant polynomials $q_j^{\bar{w}_k}$ and $q_j^{\bar{w}_l}$, this reduces bounding (ii) to bounding, for any fixed u,

$$\Delta_u := [q_u^{\bar{w}_k}(\mathbb{E}z_u)q_u^{\bar{w}_l}(\mathbb{E}z_u) - \mathbb{E}q_u^{\bar{w}_k}(z_u)\mathbb{E}q_u^{\bar{w}_l}(z_u)] \\ \cdot \prod_{\substack{j<u \\ j\neq t}} q_j^{\bar{w}_k}(\mathbb{E}z_j)q_j^{\bar{w}_l}(\mathbb{E}z_j) \prod_{\substack{j>u \\ j\neq t}} \mathbb{E}q_j^{\bar{w}_k}(z_j)\mathbb{E}q_j^{\bar{w}_l}(z_j).$$

Recalling that all $q_j^{\bar{w}_i}$ are almost surely less than 1, this can be bounded by

$$|\Delta_u| \leq |q_u^{\bar{w}_k}(\mathbb{E}z_u)q_u^{\bar{w}_l}(\mathbb{E}z_u) - \mathbb{E}q_u^{\bar{w}_k}(z_u)\mathbb{E}q_u^{\bar{w}_l}(z_u)| \leq (4k+4l)\mathbb{E}\,|z_u - \mathbb{E}z_u| \leq K_2 n^{-1/2}.$$

These bounds applied to the difference of (ii) and its approximation show

$$D_{(ii)} := \left|\prod\nolimits_{j\neq t} \mathbb{E}q_j^{\bar{w}_k}(z_j)\mathbb{E}q_j^{\bar{w}_l}(z_j) - \prod\nolimits_{j\neq t} q_j^{\bar{w}_k}(\mathbb{E}z_j)q_j^{\bar{w}_l}(\mathbb{E}z_j)\right| \\ \leq \sum_{\substack{1\leq u\leq 2n-1 \\ q_u^{\bar{w}_k} q_u^{\bar{w}_l} \neq 1}} |\Delta_u| \leq (2k+2l)\cdot K_2 n^{-1/2}. \quad (2.24)$$

By combining Lemma 5 with Cauchy-Schwarz, one has that (i) is at most $K_3 n^{-1}$. Therefore, we can combine both of (2.23) and (2.24) to show

$$\left|r_t - r_t^D\right| \leq \sum\nolimits_{\bar{w}_k,\bar{w}_l} |(i)|\left|D_{(ii)}\right| + \left|D_{(i)}\right|\left|\left[\prod\nolimits_{j\neq t} q_j^{\bar{w}_k}(\mathbb{E}z_j)q_j^{\bar{w}_l}(\mathbb{E}z_j)\right]\right| \\ \leq \sum\nolimits_{\bar{w}_k,\bar{w}_l} K_3 n^{-1}\cdot(2k+2l)\cdot K_2 n^{-1/2} + K_1 n^{-3/2}\cdot 1.$$

As the sum is only over paths that depend upon t, the proof is complete. □

All the expectations in r_t^D are approximately equal to one of two values, $\mathbb{E}z_t$ and $\mathbb{E}z_{t+n}$ (or $t-n$ in the case $t > n$), on account of the trace paths being forced to overlap. Thus, this can be expressed in terms of the polynomials $p_k(x,y)$ for values of t for which the trace paths are sufficiently far from the matrix edge. The values of x and y are given in terms of the expectations of matrix entries. Put $s(t) = t$ if $1 \leq t \leq n$, and put $s(t) = t - n$ if $n+1 \leq t \leq 2n-1$. The values of x and y are given by

$$x(t) := \sqrt{\mathbb{E}\left[c_s^2(1-(c_s')^2)\right]} \quad \text{and} \quad y(t) := \sqrt{\mathbb{E}\left[(c_s')^2(1-(c_s)^2)\right]}. \quad (2.25)$$

Note that these are not exactly the expressions for x and y given in (2.19), but we will show that these two quantities are strongly related. In what follows, we unequivocally mean the x and y given in (2.25).

Lemma 9. *Define ξ_t^D to be*

$$\xi_t := r_t^D - \frac{\mathrm{Var}(c_s^2)}{4}\left(\frac{x\partial_x p_k(x,y)}{\mathbb{E}c_s^2} - \frac{y\partial_y p_k(x,y)}{1-\mathbb{E}c_s^2}\right)\left(\frac{x\partial_x p_m(x,y)}{\mathbb{E}c_s^2} - \frac{y\partial_y p_m(x,y)}{1-\mathbb{E}c_s^2}\right),$$

when $1 \le t \le n$ and

$$\xi_t := r_t^D - \frac{\mathrm{Var}(c_s'^2)}{4}\left(\frac{y\partial_y p_k(x,y)}{\mathbb{E}c_s'^2} - \frac{x\partial_x p_k(x,y)}{1-\mathbb{E}(c_s')^2}\right)\left(\frac{y\partial_y p_m(x,y)}{\mathbb{E}c_s'^2} - \frac{x\partial_x p_m(x,y)}{1-\mathbb{E}(c_s')^2}\right)$$

when $n+1 \le t \le 2n-1$. There is a constant $K = K(k,l)$ so that for all $k+l \le t \le n-k-l$ and $n+k+l \le t \le 2n-k-l-1$, $\left|\xi_t^D\right| \le Kn^{-2}$.

Proof. We show the proof for $1 \le t \le n$. The proof for $t > n$ is identical. We recall that r_t^D is given by

$$r_t^D = \mathrm{Var}(z_t)\sum\nolimits_{\bar{w}_k,\bar{w}_l}\left[q_t^{\bar{w}_k\prime}(\mathbb{E}z_t)\right]\left[q_t^{\bar{w}_l\prime}(\mathbb{E}z_t)\right]\left[\prod\nolimits_{j\neq t} q_j^{\bar{w}_k}(\mathbb{E}z_j)q_j^{\bar{w}_l}(\mathbb{E}z_j)\right].$$

This splits nicely as $r_t^D = \mathrm{Var}(z_t)M_t(\bar{w}_k)M_t(\bar{w}_l)$, where we define

$$M_t(\bar{w}_i) := \sum\nolimits_{\bar{w}_i}\left[q_t^{\bar{w}_i\prime}(\mathbb{E}z_t)\right]\left[\prod\nolimits_{j\neq t} q_j^{\bar{w}_i}(\mathbb{E}z_j)\right].$$

This $M_t(\bar{w}_i)$ is essentially computable from just two expectations, $\mathbb{E}z_t$ and $\mathbb{E}z_{t+n}$. Letting

$$M_t(\bar{w}_i)^D := \sum\nolimits_{\bar{w}_i}\left[q_t^{\bar{w}_i\prime}(\mathbb{E}z_t)\right]\left[\prod\nolimits_{\substack{j\neq t\\ 1\le j\le n}} q_j^{\bar{w}_i}(\mathbb{E}z_t)q_{j+n}^{\bar{w}_i}(\mathbb{E}z_{t+n})\right],$$

we show that $\left|M_t(\bar{w}_i) - M_t^D(\bar{w}_i)\right|$ is $O(n^{-1})$. We will require the formulae for $\mathbb{E}z_t = \mathbb{E}c_t^2$ and $\mathbb{E}z_{t+n} = \mathbb{E}(c_t')^2$, and so we recall the precise distributions of these entries,

$$c_i^2 \sim \mathrm{Beta}(\tfrac{nb}{\alpha a} + \alpha^{-1}(i-n), \tfrac{n(1-b)}{\alpha a} + \alpha^{-1}(i-n)) \quad \text{and}$$
$$(c_i')^2 \sim \mathrm{Beta}(\alpha^{-1}i, \tfrac{n}{\alpha a} + \alpha^{-1}(i-2n+1)).$$

Their expectations are given by

$$\mathbb{E}z_t = \mathbb{E}c_t^2 = \frac{\frac{nb}{\alpha a} + \alpha^{-1}(t-n)}{\frac{n}{\alpha a} + \alpha^{-1}(2t-n)} = \frac{b - a + 2a\frac{t}{n}}{1 - 2a + 2a\frac{t}{n}} \tag{2.26}$$

$$\mathbb{E}z_{t+n} = \mathbb{E}(c_t')^2 = \frac{\alpha^{-1}t}{\frac{n}{\alpha a} + \alpha^{-1}(2t-n)} = \frac{a\frac{t}{n}}{1 - 2a + 2a\frac{t}{n}}. \tag{2.27}$$

Each of these expectations, as a function of t, is uniformly Lipschitz continuous over $0 \le t \le n$ with constant $K_1 \cdot n^{-1}$ for some K_1 depending only on the ensemble parameters. By the same method used in the proof of Lemma 8, it is straightforward to show that there is a constant $K_2 = K_2(k, l)$ so that

$$\left| M_t(\bar{w}_i) - M_t^D(\bar{w}_i) \right| \le K_2 n^{-1}.$$

We recall the notation of Lemma 3, where we defined $S_{\rightarrow}^{\bar{w}_i}(t)$ to be the number of horizontal steps of $\bar{w}_i$ from level i to i and $S_{\uparrow}^{\bar{w}_i}(t)$ to be the number of steps of $\bar{w}_i$ from level i to $i+1$ or vice versa. The polynomial $q_t^{\bar{w}_i}$ may be identified precisely in terms of these counts. Recalling the matrix model (2.6), the variables c_t and s_t appear only in the t^{th} row from the bottom of the matrix. It follows that

$$q_t^{\bar{w}_i}(c_t^2) = c_t^{S_{\rightarrow}^{\bar{w}_i}(t)} (1 - c_t^2)^{S_{\uparrow}^{\bar{w}_i}(t)/2},$$

and thus, differentiating,

$$\begin{aligned} q_t^{\bar{w}_i}(z)' &= \frac{S_{\rightarrow}^{\bar{w}_i}(t) z^{S_{\rightarrow}^{\bar{w}_i}(t)/2-1} (1-z)^{S_{\uparrow}^{\bar{w}_i}(t)/2}}{2} - \frac{S_{\uparrow}^{\bar{w}_i}(t) z^{S_{\rightarrow}^{\bar{w}_i}(t)/2} (1-z)^{S_{\uparrow}^{\bar{w}_i}(t)/2-1}}{2} \\ &= \frac{S_{\rightarrow}^{\bar{w}_i}(t) q_t^{\bar{w}_i}(z)}{2z} - \frac{S_{\uparrow}^{\bar{w}_i}(t) q_t^{\bar{w}_i}(z)}{2(1-z)}. \end{aligned} \tag{2.28}$$

We now relate $M_t^D(\bar{w}_i)$ to expressions containing $p_i(x, y)$. The essential realization is that

$$\begin{aligned} &\sum\nolimits_{\bar{w}_i} S_{\rightarrow}^{\bar{w}_i}(t) \left[\prod\nolimits_{1 \le j \le n} q_j^{\bar{w}_i}(\mathbb{E}z_t) q_{j+n}^{\bar{w}_i}(\mathbb{E}z_{t+n}) \right] \\ &\qquad = \sum\nolimits_{\bar{w} \in \mathcal{A}_{2i}} h(\bar{w}) x^{h(\bar{w})} y^{2i - h(\bar{w})} = x \partial_x p_i(x, y), \end{aligned} \tag{2.29}$$

where $h(\bar{w}_i)$ is the number of horizontal steps $\bar{w}_i$ makes, and x and y are defined earlier. This is a direct consequence of the bijection between paths $\bar{w}_1 \in \tilde{\mathcal{A}}_{2i,n}$ that have a single marked horizontal edge at level t and paths $\bar{w}_2 \in \mathcal{A}_{2i}$ having a single marked horizontal edge. This is given by the map that simply vertically shifts $\bar{w}_1$ to start at 0; note that this is invertible on account of the mark being forced to lie at level t. For $n-k-l \geq t \geq k+l$, every summand on the left hand side of (2.29) is exactly the summand given on the right when identifying paths via this bijection (note that for t too close to the matrix edge, some of the paths on the left hand side will be 0, destroying the identity). Similar reasoning shows

$$\sum\nolimits_{\bar{w}_i} S_{\uparrow}^{\bar{w}_i}(t)\left[\prod\nolimits_{1\leq j\leq n} q_j^{\bar{w}_i}(\mathbb{E}z_t)q_{j+n}^{\bar{w}_i}(\mathbb{E}z_{t+n})\right] = y\partial_y p_i(x,y). \tag{2.30}$$

By combining (2.28), (2.29),and (2.30), it follows that

$$M_t^D(\bar{w}_i) = \frac{x\partial_x p_i(x,y)}{2\mathbb{E}z_t} - \frac{y\partial_y p_i(x,y)}{2\mathbb{E}[1-z_t]}.$$

The conclusion of the lemma follows more or less immediately. By Lemma 5, the variance of z_t can be controlled by $K_3 n^{-1}$, with K_3 depending only on the matrix parameters. The moduli of $M_t(\bar{w}_i)$ and $M_t(\bar{w}_i)^D$ can be controlled by some $K_4 = K_4(k,l)$, and so

$$|\xi_t| = \left|\mathrm{Var}(z_t)M_t(\bar{w}_k)M_t(\bar{w}_l) - \mathrm{Var}(z_t)M_t^D(\bar{w}_k)M_t^D(\bar{w}_l)\right| \leq K_3 n^{-1}\cdot 2K_4\cdot K_2 n^{-1},$$

completing the proof. □

On account of the variance being of the order of n^{-1}, summing these expressions takes the form of a Riemann sum. We thus conclude the proof of the limiting covariance formula by showing that this Riemann sum converges to the integral given by $C_{k,l}$.

Proof of Proposition 9. By Lemma 8 and (2.22),

$$\mathbb{E}X_{x^k,A}X_{x^l,A} = \left[\sum\nolimits_{t=1}^{2n-1} r_t^D\right] + O(n^{-1/2}).$$

Let $\Sigma := \sum_{t=1}^{n} r_t^D$ and $\Sigma' := \sum_{t=n+1}^{2n-1} r_t^D$. Lemma 9 shows that

$$\Sigma = \sum_{t=k+l}^{n-k-l} \frac{\mathrm{Var}(c_t^2)}{4} \left(\frac{x\partial_x p_k(x,y)}{\mathbb{E}c_t^2} - \frac{y\partial_y p_k(x,y)}{1-\mathbb{E}c_t^2} \right)\left(\frac{x\partial_x p_m(x,y)}{\mathbb{E}c_t^2} - \frac{y\partial_y p_m(x,y)}{1-\mathbb{E}c_t^2} \right) + O(n^{-1/2}). \quad (2.31)$$

We will show that the variance of these Beta variables is of order n^{-1}. To concisely describe the integrand that results in the limit, put τ to be the variable over which the integral is taken, and define $e(\tau)$ and $e'(\tau)$ as

$$e(\tau) := \frac{b-a+2a\tau}{1-2a+2a\tau} \quad \text{and} \quad e'(\tau) := \frac{a\tau}{1-2a+2a\tau} \quad 0 \le \tau \le 1, \quad (2.32)$$

so that for $\tau = t/n$, $e(\tau) = \mathbb{E}c_t^2$ and $e'(\tau) = \mathbb{E}(c_t')^2$ (see (2.26)). We will reuse the notation x and y by putting

$$x(\tau) := \sqrt{e(\tau)(1-e'(\tau))} \quad \text{and} \quad y(\tau) := \sqrt{e'(\tau)(1-e(\tau))}. \quad (2.33)$$

This definition is now consistent with (2.19), after making a change of variables. We recall the variances of these Beta variables,

$$\begin{aligned} \mathrm{Var}\, c_t^2 &= \frac{\alpha a}{n} \frac{\left(b-a+a\frac{t}{n}\right)\left(1-b-a+a\frac{t}{n}\right)}{\left(1-2a+2a\frac{t}{n}\right)^3} + O(n^{-2}), \\ \mathrm{Var}(c_t')^2 &= \frac{\alpha a}{n} \frac{\left(a\frac{t}{n}\right)\left(1-2a+a\frac{t}{n}\right)}{\left(1-2a+2a\frac{t}{n}\right)^3} + O(n^{-2}), \end{aligned} \quad (2.34)$$

where we may choose the constants in the error terms to depend only on the ensemble parameters (and not t). By virtue of the $\frac{\alpha}{n}$ factor, the sum $\sum_{t=1}^{n} r_t^D$ takes the form of a Riemann sum. The integrand, exposed on the right hand side of (2.31), is Lipschitz continuous in t/n, and thus the convergence of the Riemann sum to the integral occurs with rate $O(n^{-1})$. This shows

$$\Sigma = \frac{\alpha a}{4} \int_0^1 \frac{(b-a+a\tau)(1-b-a+a\tau)}{(1-2a+2a\tau)^3} \left(\frac{x\partial_x p_k}{e(\tau)} - \frac{y\partial_y p_k}{1-e(\tau)} \right)\left(\frac{x\partial_x p_m}{e(\tau)} - \frac{y\partial_y p_m}{1-e(\tau)} \right) d\tau + O(n^{-1/2}). \quad (2.35)$$

Applying the same reasoning to $n+1 \le t \le 2n-1$, it follows that

$$\Sigma' = \frac{\alpha a}{4} \int_0^1 \frac{(a\tau)\,(1-2a+a\tau)}{(1-2a+2a\tau)^3} \left(\frac{y\partial_y p_k}{e'(\tau)} - \frac{x\partial_x p_k}{1-e'(\tau)} \right) \left(\frac{y\partial_y p_m}{e'(\tau)} - \frac{x\partial_x p_m}{1-e'(\tau)} \right) d\tau + O(n^{-1/2}). \quad (2.36)$$

The sum of these two integrals (2.35) and (2.36) and the associated error bounds show that the limiting covariance exists, and their sum provides an expression for the limit. The remainder of the proof will show that this expression can be alternately expressed in the form given by $C_{k,l}$ (defined in (2.20)). The primary difference is a change of variables. Take $\sigma = a(\tau - 1)$. The integrals become

$$\Sigma = \frac{\alpha}{4} \int_{-a}^0 \frac{e(\sigma)(1-e(\sigma))}{(1+2\sigma)} \left(\frac{x\partial_x p_k}{e(\sigma)} - \frac{y\partial_y p_k}{1-e(\sigma)} \right)\left(\frac{x\partial_x p_m}{e(\sigma)} - \frac{y\partial_y p_m}{1-e(\sigma)} \right) d\sigma + O(n^{-1/2}), \quad (2.37)$$

$$\Sigma' = \frac{\alpha}{4} \int_{-a}^0 \frac{e'(\sigma)(1-e'(\sigma))}{(1+2\sigma)} \left(\frac{y\partial_y p_k}{e'(\sigma)} - \frac{x\partial_x p_k}{1-e'(\sigma)} \right)\left(\frac{y\partial_y p_m}{e'(\sigma)} - \frac{x\partial_x p_m}{1-e'(\sigma)} \right) d\sigma + O(n^{-1/2}). \quad (2.38)$$

The sum of these integrals can be shown to equal $C_{k,l}$ by checking the coefficients in front of the terms $\partial_x p_k \partial_x p_m$, $\partial_y p_k \partial_y p_m$, $\partial_x p_k \partial_y p_m$ and $\partial_y p_k \partial_x p_m$. The coefficient on $\partial_x p_k \partial_x p_m$ in the sum of the integrands (2.37) and (2.38) is given by

$$\left[\frac{e(\sigma)(1-e(\sigma))}{(1+2\sigma)} \frac{x^2}{e(\sigma)^2} + \frac{e'(\sigma)(1-e'(\sigma))}{(1+2\sigma)} \frac{x^2}{1-e'(\sigma)^2} \right] = \frac{1-x^2-y^2}{1+2\sigma}.$$

Similar manipulations show that the coefficients on each of the other terms agree with the coefficients in the integrand of $C_{k,l}$, completing the proof. □

2.3 Diagonalizing the Covariance Matrix

We proceed by showing that the covariances are diagonalized by the appropriate Chebyshev polynomial basis. This will be done by verifying that certain generating functions agree. We would like to show that the infinite covariance matrix can be decomposed as

$$C = L\Lambda L^t,$$

for the diagonal matrix $\Lambda = \text{diag}(0, 1, 2, 3, 4, \ldots)$, and some lower triangular matrix L. The $L_{n,k}$ entry of this matrix is the coefficient of the k^{th} Chebyshev polynomial $\Gamma_k(x)$ in the expansion of x^n. Define the exponential covariance generating function $\mathscr{C}(s,t)$ as

$$\mathscr{C}(s,t) = \sum_{k,l>0} \frac{s^k}{k!}\frac{t^l}{l!} C_{k,l},$$

and define the exponential generating function of $L\Lambda L^t$ analogously,

$$\mathscr{T}(s,t) = \sum_{k,l>0} \frac{s^k}{k!}\frac{t^l}{l!} \left[L\Lambda L^t\right]_{k,l}.$$

We will show that these generating functions are equal by computing their bivariate Laplace transforms and showing they are the same, from which it follows that $C = L\Lambda L^t$.

Computing $\mathcal{L}_{s,t}[\mathscr{T}]$

The coefficients $L_{n,k}$ can be computed by a recursive formula, but they have a useful Fourier-like expansion. Define θ in terms of x so that

$$\cos(\theta) = \frac{2x - \lambda_- - \lambda_+}{\lambda_+ - \lambda_-},$$

from which it follows that

$$2\cos(n\theta) = 2T_n(\cos\theta) = 2T_n\left(\frac{2x - \lambda_- - \lambda_+}{\lambda_+ - \lambda_-}\right).$$

Expand $\frac{1}{2}e^{tx}$ as a series in t,

$$\tfrac{1}{2}e^{tx} = \tfrac{1}{2}\sum_{n=0}^{\infty} \frac{t^n}{n!}\left(c + r\cos\theta\right)^n = \tfrac{1}{2}\sum_{n=0}^{\infty}\sum_{k=0}^{k} L_{n,k}\frac{t^n}{n!} 2\cos k\theta = \sum_{n=0}^{\infty}\sum_{k=0}^{n} L_{n,k}\frac{t^n}{n!}\cos k\theta,$$

where we have used the definition of $L_{n,k}$ as the coefficient of the k^{th} Chebyshev polynomial in the expansion of x^n and where $c = \left(\frac{\lambda_+ + \lambda_-}{2}\right)$ and $r = \left(\frac{\lambda_+ - \lambda_-}{2}\right)$.

The Fourier interpretation allows for the matrix multiplication $L\Lambda L^t$ to be carried out by an integral. Consider the kernel $K_N(\theta, \phi)$, which will formally play the role of Λ, given by

$$K_N(\theta, \phi) = \sum_{k=1}^{N} k \cos k\theta \cdot \cos k\phi.$$

This allows for $\mathscr{T}$ to be given by

$$\mathscr{T}(s,t) = \lim_{N\to\infty} \frac{1}{4\pi^2} \int_0^\pi \int_0^\pi e^{tc + r\cos\theta} K_N(\theta,\phi) e^{sc + r\cos\theta} d\phi d\theta,$$

as the coefficient on $t^k s^l$ would be

$$\lim_{N\to\infty} \frac{1}{\pi^2} \int_0^\pi \int_0^\pi \left(\sum_{j=0}^{k} L_{k,j} \frac{t^k}{k!} \cos j\theta\right) \left(\sum_{k=1}^{N} k \cos k\theta \cdot \cos k\phi\right) \left(\sum_{j=0}^{l} L_{l,j} \frac{s^l}{l!} \cos j\phi\right) d\phi d\theta,$$

which by the orthogonality of $\{\cos j\theta\}_{j=0}^{\infty}$ on $[0, \pi]$, is exactly $(L\Lambda L^t)_{k,l}$ when $N > \min(k, l)$. Further, these integrals can be evaluated, as the expression $e^{z\cos\theta}$ has an expansion in terms of Bessel functions. Namely,

$$e^{z\cos\theta} = I_0(z) + 2\sum_{k=1}^{\infty} I_k(z) \cos k\theta,$$

(see [1, p. 376]). This defines the Fourier coefficients of $e^{z\cos\theta}$, from which it follows that $\mathscr{T}(s,t)$ can be rewritten as

$$\begin{aligned}\mathscr{T}(s,t) &= e^{c(t+s)} \lim_{N\to\infty} \frac{1}{4\pi^2} \int_0^\pi \int_0^\pi e^{rt\cos\theta} K_N(\theta,\phi) e^{rs\cos\phi} d\theta d\phi \\ &= e^{c(t+s)} \sum_{k=1}^{\infty} k I_k(rt) I_k(rs).\end{aligned}$$

Again, we will require the Laplace transform of this generating function. Each summand $kI_k(rt)I_k(rs)$ is positive for $s, t > 0$, and so commuting the sum and the

Laplace transform is justified.

$$\begin{aligned}\mathcal{L}_{s,t}[\mathscr{T}(s,t)](\eta,\omega) &= \sum_{k=1}^{\infty} k\mathcal{L}_{s,t}\left[e^{c(t+s)}I_k(rt)I_k(rs)\right](\eta,\omega)\\ &= \sum_{k=1}^{\infty} k\frac{r^k}{\left(\tilde{\omega}+\sqrt{\tilde{\omega}^2-r^2}\right)^k}\frac{1}{\sqrt{\tilde{\omega}^2-r^2}}\frac{r^k}{(\tilde{\eta}+\sqrt{\tilde{\eta}^2-r^2})^k}\frac{1}{\sqrt{\tilde{\eta}^2-r^2}},\end{aligned}$$

where $\tilde{\omega}=\omega-c$. This has the form for the series expansion of $\frac{x}{(1-x)^2}$. After simplifying, this expression is

$$\mathcal{L}_{s,t}[\mathscr{T}(s,t)](\eta,\omega) = \frac{r^2}{\sqrt{\tilde{\eta}^2-r^2}\sqrt{\tilde{\omega}^2-r^2}\left[\sqrt{(\tilde{\omega}+r)(\tilde{\eta}-r)}+\sqrt{(\tilde{\omega}-r)(\tilde{\eta}+r)}\right]^2}. \tag{2.39}$$

Computing $\mathcal{L}_{s,t}[\mathscr{C}]$

We will now turn to computing the Laplace transform of $\mathscr{C}$. The integrand of $C_{k,l}$ is not positive, but it can be split into two integrals whose integrands are positive (see Remark 3)

$$C_{k,l} = L_{k,l} - R_{k,l},$$

with

$$L_{k,l} = \frac{\alpha}{4}\int_{-a}^{0}\frac{1}{1+2\sigma}\left[(\partial_x p_k \partial_x p_m + \partial_y p_k \partial_y p_m)\right]d\sigma,$$

and

$$\begin{aligned}R_{k,l} = \frac{\alpha}{4}\int_{-a}^{0}\frac{1}{1+2\sigma}\big[&(\partial_x p_k \partial_x p_m + \partial_y p_k \partial_y p_m)(x^2+y^2)\\ &+(\partial_x p_k \partial_y p_m + \partial_y p_k \partial_x p_m)(2xy)\big]d\sigma.\end{aligned}$$

As $p_l(x,y)$ has all positive coefficients, and both x and y are positive on the domain of integration, each of these integrands is positive. Defining generating functions for each array,

$$\mathscr{L}(s,t) = \sum_{k,l>0}\frac{s^k}{k!}\frac{t^l}{l!}L_{k,l} \quad \text{and} \quad \mathscr{R}(s,t) = \sum_{k,l>0}\frac{s^k}{k!}\frac{t^l}{l!}R_{k,l},$$

we can write

$$\mathscr{L}(s,t) = \frac{\alpha}{4}\int_{-a}^{0} \frac{1}{1+2\sigma}\left[(\partial_x \mathscr{P}(s)\partial_x \mathscr{P}(t) + \partial_y \mathscr{P}(s)\partial_y \mathscr{P}(t))\right] d\sigma, \qquad \text{and}$$

$$\mathscr{R}(s,t) = \frac{\alpha}{4}\int_{-a}^{0} \frac{1}{1+2\sigma}\Big[(\partial_x \mathscr{P}(s)\partial_x \mathscr{P}(t) + \partial_y \mathscr{P}(s)\partial_y \mathscr{P}(t))\,(x^2+y^2) + (\partial_x \mathscr{P}(s)\partial_y \mathscr{P}(t) + \partial_y \mathscr{P}(s)\partial_x \mathscr{P}(t))\,(2xy)\Big] d\sigma, \tag{2.40}$$

where we have commuted sum and integral by the positivity of the integrands. Recall that $\mathscr{P}(t) = \mathscr{P}(t,x,y)$ is the exponential generating function for the polynomials $p_k(x,y)$, and from (2.7), it is jointly analytic in all variables. As $-a > -1/2$, it follows that the integrands are continuous for all $-a \le \sigma \le 0$, and all s,t. In particular, each of $\mathscr{L}$ and $\mathscr{R}$ is finite for all s,t, and it follows that we can write $\mathscr{C}(s,t)$ as the sum of these two functions, so

$$\mathscr{C}(s,t) = \mathscr{L}(s,t) - \mathscr{R}(s,t).$$

The joint Laplace transforms in s and t will be computed for both of these expressions. This makes heavy use of Lemma 4. Additionally, it requires that the order of integration be switched, which requires an argument. We prove a simplified statement, by whose method it is easily seen that these integrals can be exchanged.

Lemma 10. *Suppose that $\omega > \lambda_+$ and that $\eta > \lambda_+$, then*

$$\int_0^\infty\int_0^\infty\int_{-a}^0 \frac{e^{-\omega t-\eta s}}{1+2\sigma}\partial_x \mathscr{P}(s)\partial_x \mathscr{P}(t)d\sigma ds dt = \int_{-a}^0\int_0^\infty\int_0^\infty \frac{e^{-\omega t-\eta s}}{1+2\sigma}\partial_x \mathscr{P}(s)\partial_x \mathscr{P}(t)d\sigma ds dt,$$

and each is finite.

Proof. We begin by maximizing $x+y$ over $\sigma \in [-a,0]$, where it is seen that the maximum is attained at $\sigma = 0$, at which point,

$$(x(0)+y(0))^2 = \left(\sqrt{b(1-a)} + \sqrt{a(1-b)}\right)^2 = \lambda_+.$$

Thus, it follows that $(x+y)^2 < \lambda_+ < \omega$ for all $-a \le \sigma \le 0$. Recall that $\mathscr{P}(t)$ is given by $e^{t(x^2+y^2)}I_0(2xyt)$, and thus

$$\partial_x \mathscr{P}(t) = 2xte^{t(x^2+y^2)}I_0(2xyt) + 2yte^{t(x^2+y^2)}I_1(2xyt).$$

Using that $0 \le I_n(2xyt) \le e^{2xyt}$ for all n, it follows that

$$0 \le \partial_x \mathscr{P}(t) \le 2(x+y)e^{t(x+y)^2},$$

for $x, y \ge 0$. It follows that there is a constant C so that for all $-a \le \sigma \le 0$,

$$0 \le \frac{e^{-\omega t - \eta s}}{1+2\sigma}\partial_x \mathscr{P}(s)\partial_x \mathscr{P}(t) < C(x+y)^2 e^{-(\omega-(x+y)^2)t-(\eta-(x+y)^2)s}.$$

Using the bound on $x+y$ derived above,

$$0 \le \frac{e^{-\omega t - \eta s}}{1+2\sigma}\partial_x \mathscr{P}(s)\partial_x \mathscr{P}(t) < C\lambda_+^2 e^{-(\omega-\lambda_+)t-(\eta-\lambda_+)s}.$$

Thus, provided that $\omega > \lambda_+$ and $\eta > \lambda_+$, the order of integration may be reversed by Fubini. □

We can now compute the bivariate Laplace transform of $\mathscr{C}(s,t)$.

Lemma 11.

$$\mathcal{L}_{s,t}[\mathscr{C}(s,t)](\eta,\omega) = \frac{\alpha}{8}\int_{(1-2a)^2}^{1} \frac{n_1 - n_2\rho}{(p_1 - p_2\rho)^{\frac{3}{2}}(r_1 - r_2\rho)^{\frac{3}{2}}}d\rho,$$

where these parameters are given by

$$n_1 = -\omega\eta(1-\lambda_- - \lambda_+)^2 - (\omega+\eta)\lambda_-\lambda_+(1-\lambda_- - \lambda_+) - \lambda_-\lambda_+(1-\lambda_- - \lambda_+ + 2\lambda_-\lambda_+)$$
$$n_2 = -\omega\eta[1-\lambda_- - \lambda_+ + 2\lambda_-\lambda_+] + (\omega+\eta-1)\lambda_-\lambda_+$$
$$p_1 = \omega(1-\lambda_- - \lambda_+) + \lambda_-\lambda_+$$
$$p_2 = \omega - \omega^2$$
$$r_1 = \eta(1-\lambda_- - \lambda_+) + \lambda_-\lambda_+$$
$$r_2 = \eta - \eta^2.$$

Proof. We start by commuting the integration in σ and the Laplace transform in (2.40). To evaluate these Laplace transforms, we recall Lemma 4, where the Laplace transform $\mathcal{L}_t[\partial_x \mathscr{P}(t)]$ was computed to be

$$\mathcal{L}_t[\partial_x \mathscr{P}(t)](\omega) = \frac{2x(\omega + y^2 - x^2)}{\left((\omega - x^2 - y^2)^2 - 4x^2y^2\right)^{\frac{3}{2}}}, \qquad \omega > (x+y)^2.$$

The quantity $x^2 - y^2$ simplifies to

$$x^2 - y^2 = \frac{b-a}{1+2\sigma}.$$

The Laplace transform of $\partial_x \mathscr{P}(t)$ can be rewritten as

$$\mathcal{L}_t[\partial_x \mathscr{P}(t)](\omega) = \frac{2x(1+2\sigma)^2(\omega(1+2\sigma)-(b-a))}{((\omega^2-\omega)(1+2\sigma)^2-(1-2a)(1-2b)\omega+(b-a)^2)^{\frac{3}{2}}}, \qquad \omega > \lambda_+,$$

for $\sigma \in [-a, 0]$. By symmetry, the Laplace transform $\partial_y \mathscr{P}(t)$ is

$$\mathcal{L}_t[\partial_y \mathscr{P}(t)](\omega) = \frac{2y(1+2\sigma)^2(\omega(1+2\sigma)+(b-a))}{((\omega^2-\omega)(1+2\sigma)^2-(1-2a)(1-2b)\omega+(b-a)^2)^{\frac{3}{2}}}, \qquad \omega > \lambda_+.$$

Define $\Delta(\omega)$ to be

$$\Delta(\omega) = \left((\omega^2-\omega)(1+2\sigma)^2-(1-2a)(1-2b)\omega+(b-a)^2\right),$$

and define $\rho = (1+2\sigma)^2$. We will now split the computation of $\mathscr{C}(s,t)$ into two pieces for simplicity's sake. The first piece is

$$\begin{aligned}\mathcal{L}_{s,t}&\left[\partial_x \mathscr{P}(s)\partial_x \mathscr{P}(t) + \partial_y \mathscr{P}(s)\partial_y \mathscr{P}(t)\right](\eta,\omega) \\ &= 4\rho^2 \frac{(x^2+y^2)[\omega\eta\rho+(b-a)^2]-(b-a)^2(\omega+\eta)}{\Delta(\omega)^{\frac{3}{2}}\Delta(\eta)^{\frac{3}{2}}}.\end{aligned}$$

The second piece is

$$\begin{aligned}\mathcal{L}_{s,t}&\left[\partial_x \mathscr{P}(s)\partial_y \mathscr{P}(t) + \partial_y \mathscr{P}(s)\partial_x \mathscr{P}(t)\right](\eta,\omega) \\ &= \frac{8xy\rho^2[\omega\eta\rho-(b-a)^2]}{\Delta(\omega)^{\frac{3}{2}}\Delta(\eta)^{\frac{3}{2}}}.\end{aligned}$$

Combining these two pieces,

$$\begin{aligned}&\mathcal{L}_{s,t}[\mathscr{C}](\eta,\omega) = \\ &\quad \alpha\int_{-a}^{0} \frac{\rho^2\left[((x^2+y^2)(1-x^2-y^2)-4x^2y^2)\omega\eta\rho+((x^2+y^2)(1-x^2-y^2)+4x^2y^2)(b-a)^2\right]}{\sqrt{\rho}\Delta(\omega)^{\frac{3}{2}}\Delta(\eta)^{\frac{3}{2}}} \\ &\qquad + \frac{\rho^2\left[-(b-a)^2(\omega+\eta)(1-x^2-y^2)\right]}{\sqrt{\rho}\Delta(\omega)^{\frac{3}{2}}\Delta(\eta)^{\frac{3}{2}}} d\sigma.\end{aligned}$$

We simplify some of these expressions,

$$((x^2+y^2)(1-x^2-y^2)-4x^2y^2)$$
$$=\tfrac{1}{2}\rho^{-2}\left[(-4a(1-a)+1)(-4b(1-b)+1)\right]+\tfrac{1}{2}\rho^{-1}\left[1-2b(1-b)-2a(1-a)\right],$$
$$((x^2+y^2)(1-x^2-y^2)+4x^2y^2)$$
$$=\tfrac{1}{2}\rho^{-1}\left[\rho-1+2b(1-b)+2a(1-a)\right],$$
$$1-x^2-y^2$$
$$=\tfrac{1}{2}\rho^{-1}\left[\rho+(1-2a)(1-2b)\right].$$

After changing the integration to be over ρ, we produce the desired formula. □

We will explicitly evaluate the integral in Lemma 11 to conclude that

Lemma 12.

$$\mathcal{L}_{s,t}[\mathscr{C}(s,t)]=\alpha\frac{d^2}{\left[\sqrt{(\tilde{\omega}+d)(\tilde{\eta}-d)}+\sqrt{(\tilde{\omega}-d)(\tilde{\eta}+d)}\right]^2\sqrt{\tilde{\omega}^2-d^2}\sqrt{\tilde{\eta}-d^2}},$$

where $d=\left(\frac{\lambda_++\lambda_-}{2}\right)$, $\tilde{\omega}=\omega-\left(\frac{\lambda_++\lambda_-}{2}\right)$ *and* $\tilde{\eta}=\eta-\left(\frac{\lambda_++\lambda_-}{2}\right)$.

By comparing with the expression for $\mathcal{L}_{s,t}[\mathscr{T}(s,t)]$ derived in (2.39), this lemma completes the proof of the diagonalization of the covariances.

Proof. Differentiating both sides, it can be shown that

$$\int\frac{n_1-n_2\rho}{(p_1-p_2\rho)^{\frac{3}{2}}(r_1-r_2\rho)^{\frac{3}{2}}}d\rho$$
$$=\frac{-2}{(p_1r_2-p_2r_1)^2}\frac{2n_2p_1r_1-n_1(p_1r_2+p_2r_1)+\rho(2n_1p_2r_2-n_2(p_1r_2+p_2r_1))}{(p_1-p_2\rho)^{\frac{1}{2}}(r_1-r_2\rho)^{\frac{1}{2}}}.$$

The indefinite integral can be greatly simplified, plugging in some of the n, p, and r terms.

$$\int\frac{n_1-n_2\rho}{(p_1-p_2\rho)^{\frac{3}{2}}(r_1-r_2\rho)^{\frac{3}{2}}}d\rho$$
$$=\frac{2((1-\lambda_--\lambda_+)(\eta+\omega)+2\lambda_-\lambda_+)-2\rho(\omega+\eta-2\omega\eta)}{(\eta-\omega)^2(p_1-p_2\rho)^{\frac{1}{2}}(r_1-r_2\rho)^{\frac{1}{2}}}.$$

The antiderivative will now be evaluated at both endpoints. At $\rho = 1$, it becomes

$$2\frac{2\omega\eta - (\omega+\eta)(\lambda_- + \lambda_+) + 2\lambda_-\lambda_+}{(\eta-\omega)^2\sqrt{(\omega-\lambda_-)(\omega-\lambda_+)}\sqrt{(\eta-\lambda_-)(\eta-\lambda_+)}}.$$

To evaluate at $\rho = (1-2a)^2$, it is helpful to work with a and b instead of $\lambda_\pm$. Using the formulae

$$\lambda_-\lambda_+ = (b-a)^2 \quad \text{and} \quad \lambda_- + \lambda_+ = 2(a+b-2ab),$$

the antiderivative evaluated at $\rho = (1-2a)^2$ is simply

$$\frac{4}{(\eta-\omega)^2}.$$

At last we can give a single expression for the Laplace transform of the covariance function:

$$\mathcal{L}_{s,t}[\mathscr{C}(s,t)] = \frac{\alpha}{4}\frac{\left[\sqrt{(\omega-\lambda_-)(\eta-\lambda_+)} - \sqrt{(\omega-\lambda_+)(\eta-\lambda_-)}\right]^2}{(\eta-\omega)^2\sqrt{(\omega-\lambda_-)(\omega-\lambda_+)}\sqrt{(\eta-\lambda_-)(\eta-\lambda_+)}}.$$

Recall that $r = \left(\frac{\lambda_+ + \lambda_-}{2}\right)$, $\tilde{\omega} = \omega - \left(\frac{\lambda_+ + \lambda_-}{2}\right)$ and $\tilde{\eta} = \eta - \left(\frac{\lambda_+ + \lambda_-}{2}\right)$. We rewrite this expression in terms of these modified parameters to get

$$\mathcal{L}_{s,t}[\mathscr{C}(s,t)] = \alpha\mathcal{L}_{s,t}[\mathscr{T}(s,t)].$$

□

2.4 Extension to Continuously Differentiable Test Functions

We learned the idea for the extending the CLT from the appendix of Anderson-Zeitouni [4]. Roughly speaking, one would like to extend a CLT for polynomial test functions to a CLT for a larger class of functions, the hope being to invoke the density of the polynomials. However, it needs to be assured that error-in-approximation produces small error in the fluctuations when evaluated on the empirical process. The property of a matrix ensemble that allows one to execute this is a type of global concentration of eigenvalues. See also Proposition 11.6 in [4] and Lemma 1 of [105] for related approaches.

Proposition 10. *Let* $\{A_n\}$ *be an ensemble of matrices with compact spectral support* S*, and let* $V : C^1(S) \to \mathbb{R}$ *be a positive semidefinite quadratic form for which there is constant* C_1 *so that* $V(f) \leq C_1^2 \|f\|_{Lip}^2$ *for all* $f \in C^1(S)$*. Suppose that* $\{A_N\}$ *satisfies a polynomial-type CLT, i.e. for all polynomials* g*,*

$$\operatorname{tr} g(A_n) - \mathbb{E} \operatorname{tr} g(A_n) \Rightarrow N(0, V(g))$$

and additionally $\operatorname{Var} \operatorname{tr} g(A_n) \to V(g)$*. If the ensemble satisfies a Poincaré type* concentration inequality*, i.e.*

$$\operatorname{Var}(\operatorname{tr} f(A_n)) \leq C_2^2 \|f\|_{Lip}^2. \tag{2.41}$$

for some constant C_2 *independent of* n *and any Lipschitz* f *on* S*, then the polynomial CLT extends to all* C^1 *functions* $f : S \to \mathbb{R}$*, as*

$$\operatorname{tr} f(A_n) - \mathbb{E} \operatorname{tr} f(A_n) \Rightarrow N(0, V(f)).$$

Proof. We recall the quadratic Wasserstein metric

$$W_2(\mu, \nu)^2 = \inf \mathbb{E}\left(X - Y\right)^2,$$

with the infimum over all couplings (X, Y) with marginals μ and ν respectively. For a random variable X, we let $\mathcal{L}X$ denote its law. It is well known that $W_2(\mathcal{L}X_n, \mathcal{L}X) \to 0$ if and only if $X_n \Rightarrow X$ and $\mathbb{E}X_n^2 \to \mathbb{E}X^2$ (see Theorem 7.12 of [119]). For any $f \in C^1(S)$, let Z_f denote a centered normal random variable with variance $V(f)$. Thus for any polynomial g, $W_2(\mathcal{L}(\operatorname{tr} g(A_n) - \mathbb{E} \operatorname{tr} g(A_n)), \mathcal{L}Z_g) \to 0$.

Let f be any $C^1(S)$ function. By Weierstrass approximation of the derivative of f, there is a sequence of polynomials p_k so that $\|f - p_k\|_{Lip} \to 0$ as $k \to \infty$. It follows that $V(p_k) \to V(f)$ from its continuity with respect to the Lipschitz seminorm, and hence that $W_2(\mathcal{L}Z_{p_k}, \mathcal{L}Z_f) \to 0$ as $k \to \infty$. For any k we can bound,

$$\begin{aligned} W_2\left(\mathcal{L}(\operatorname{tr} f(A_n) - \mathbb{E}\operatorname{tr} f(A_n)), \mathcal{L}Z_f\right) \leq \\ & W_2\left(\mathcal{L}(\operatorname{tr} f(A_n) - \mathbb{E}\operatorname{tr} f(A_n)), \mathcal{L}(\operatorname{tr} p_k(A_n) - \mathbb{E}\operatorname{tr} p_k(A_n))\right) \\ & + W_2\left(\mathcal{L}(\operatorname{tr} p_k(A_n) - \mathbb{E}\operatorname{tr} p_k(A_n)), \mathcal{L}Z_{p_k}\right) \\ & + W_2\left(\mathcal{L}Z_{p_k}, \mathcal{L}Z_f\right). \end{aligned}$$

By the concentration inequality, it is possible to bound

$$\mathbb{E}\left[\operatorname{tr} f(A_n) - \mathbb{E}\operatorname{tr} f(A_n) - \operatorname{tr} p_k(A_n) - \mathbb{E}\operatorname{tr} p_k(A_n)\right]^2 \leq C_1^2 \|f\|_{Lip}^2,$$

from which it follows that $W_2\left(\mathcal{L}(\operatorname{tr} f(A_n) - \mathbb{E}\operatorname{tr} f(A_n)), \mathcal{L}(\operatorname{tr} p_k(A_n) - \mathbb{E}\operatorname{tr} p_k(A_n))\right) \leq C_1\|f\|_{Lip}$ by the definition of the Wasserstein metric as the infimum over couplings. Likewise

$$W_2\left(\mathcal{L}Z_{p_k}, \mathcal{L}Z_f\right) = \left|\sqrt{V(p_k)} - \sqrt{V(f)}\right| \leq \sqrt{V(p_k - f)} \leq C_2\|f\|_{Lip}.$$

Therefore, from the polynomial CLT,

$$\limsup_{n\to\infty} W_2\left(\mathcal{L}(\operatorname{tr} f(A_n) - \mathbb{E}\operatorname{tr} f(A_n)), \mathcal{L}Z_f\right) \leq (C_1 + C_2)\|f - p_k\|_{Lip}.$$

Taking $k \to \infty$ completes the proof. □

To show that linear statistics of the Jacobi ensemble satisfy a Poincaré inequality, we will work directly with the joint eigenvalue density function. Recall (2.5), which stated

$$d\mu_J(\lambda_1, \ldots, \lambda_n) = \frac{1}{Z}\prod_i \lambda_i^{\frac{n}{\alpha}\left[\frac{b}{a}-1\right]+\frac{1}{\alpha}-1}(1-\lambda_i)^{\frac{n}{\alpha}\left[\frac{1-b}{a}-1\right]+\frac{1}{\alpha}-1}\prod_{i<j}|\lambda_i - \lambda_j|^{\frac{2}{\alpha}}.$$

We first show that the Jacobi ensemble satisfies a log-Sobolev inequality, which is strictly stronger than the Poincaré inequality. Define the entropy of a non-negative measurable function f with respect to a probability measure μ by

$$\operatorname{Ent}\mu(f) := \int f\log f d\mu - \left(\int f d\mu\right)\left(\log\int f d\mu\right),$$

if $\int f \log(1+f) d\mu < \infty$ and $+\infty$ otherwise. Our tool in this direction is a consequence of the well-known Bakry-Emery condition, the content of which is contained in the following theorem (see Proposition 3.1 of [15]).

Proposition 11. *Suppose that $d\mu = e^{-U}dx$ is supported on a convex set Ω. If there is a $c > 0$ so that for all $x \in \operatorname{int}(\Omega)$, $\operatorname{Hess} U(x) \geq c \operatorname{Id}$, where Id is the identity matrix and $\geq$ is the partial ordering on positive semidefinite matrices, then for all smooth functions f on $\mathbb{R}^n$,*

$$\operatorname{Ent} \mu(f^2) \leq \frac{2}{c} \int |\nabla f|^2 \, d\mu.$$

To prove the log-Sobolev inequality with the appropriate constant, we need only check that the condition of Proposition 11 is satisfied. This we do in showing the following lemma.

Lemma 13. *The Jacobi ensemble satisfies a log-Sobolev inequality*

$$\operatorname{Ent} \mu_J(f^2) \leq \frac{2}{c} \int |\nabla f|^2 \, d\mu_J,$$

with $c = 4\frac{n}{\alpha} \min\left\{\frac{b}{a} - 1, \frac{1-b}{a} - 1\right\}$.

Proof. We will employ Proposition 11, and thus we begin by computing the Hessian of the logarithm of the density. Let $p := \frac{n}{\alpha}\left[\frac{b}{a} - 1\right] + \frac{1}{\alpha} - 1$, and let $q := \frac{n}{\alpha}\left[\frac{1-b}{a} - 1\right] + \frac{1}{\alpha} - 1$. The first derivative is given by

$$\frac{d}{d\lambda_i}(\log(d\mu_J)) = \frac{p}{\lambda_i} - \frac{q}{1-\lambda_i} + \frac{1}{\alpha}\sum_{j \neq i} \frac{1}{|\lambda_i - \lambda_j|}.$$

The second derivative is thus

$$\frac{d^2}{d\lambda_i^2}(\log(d\mu_J)) = -\frac{p}{\lambda_i^2} - \frac{q}{(1-\lambda_i)^2} - \frac{1}{\alpha}\sum_{j \neq i} \frac{1}{(\lambda_i - \lambda_j)^2}.$$

The mixed partials are just

$$\frac{d}{d\lambda_j}\frac{d}{d\lambda_i}(\log(d\mu_J)) = -\frac{1}{\alpha}\frac{1}{(\lambda_i - \lambda_j)^2}.$$

By the method of Gershgorin discs we conclude that the smallest eigenvalue of $\text{Hess}(-\log d\mu_J)$ is at least

$$\min_{\substack{1\le i\le n\\ 0\le\lambda_i\le 1}}\left[\frac{p}{\lambda_i^2}+\frac{q}{(1-\lambda_i)^2}\right]\ge 4\min\{p,q\}\ge\frac{4n}{\alpha}\min\left\{\tfrac{b}{a}-1,\tfrac{1-b}{a}-1\right\}.$$

□

It is now a simple manner to show the needed concentration inequality and prove Theorem 2.

Proof of Theorem 2. From Proposition 6 and Proposition 10, it suffices to demonstrate a constant C so that $\operatorname{Var}\operatorname{tr} f\le C\|f\|_{Lip}^2$, with the Lipschitz norm on $[0,1]$, for all Lipschitz f. This is turn follows from the somewhat sharper inequality that

$$\operatorname{Var}\operatorname{tr} f\le C\int\left|\partial_{\lambda_i}(f(\lambda_i))\right|^2 d\mu_J(\lambda_1,\ldots,\lambda_n)=\frac{C}{n}\int\left|\nabla\operatorname{tr} f\right|^2 d\mu_J(\lambda_1,\ldots,\lambda_n),$$

where in the last step we have used the symmetry of the linear statistic. It is a standard fact that the log-Sobolev inequality implies the Poincaré inequality with half the constant (see [77, Chapter 5]). Thus by Lemma 13 we have that for all smooth functions f,

$$\operatorname{Var}\operatorname{tr} f\le\frac{\alpha}{4n\min\left\{\frac{b}{a}-1,\frac{1-b}{a}-1\right\}}\int\left|\nabla\operatorname{tr} f\right|^2 d\mu_J(\lambda_1,\ldots,\lambda_n).$$

Extension to Lipschitz functions follows from the density of smooth functions in L^2, and the proof is complete. □

2.5 Computing the Expectation

In this section, we will prove Theorem 3. To establish the theorem for polynomial linear statistics ϕ, a proof will be given that follows a similar tract to the analogous statement proven for the Laguerre and Hermite ensembles in [34]. The key to this method of proof is establishing a certain palindromy. Recall that a polynomial $p(z)=a_nz^n+a_{n-1}z^{n-1}+\cdots+a_1z+a_0$ is palindromic in z if $a_nz^n+a_{n-1}z^{n-1}+\cdots+a_1z+a_0=a_0z^n+a_1z^{n-1}+\cdots+a_{n-1}z+a_n$, or equivalently that $p(z)=z^np(z^{-1})$.

Theorem 4. *The scaled moment* $\frac{1}{n}\mathbb{E}\,\mathrm{tr}(A^k)$ *has a series expansion*

$$\tfrac{1}{n}\mathbb{E}\,\mathrm{tr}(A^k) = \sum_{j=0}^{\infty} \eta_k(j,\alpha) n^{-j}$$

whose coefficients $\eta_k(j,\alpha)$ *are palindromic polynomials in* $(-\alpha)$ *of degree* j.

While the proof of this palindromy works for all of these coefficients η simultaneously, only the palindromy of $\eta_k(0,\alpha)$ and $\eta_k(1,\alpha)$ are required for Theorem 3. Especially, palindromy forces $\eta_k(0,\alpha)$ to have no α dependence, and it forces $\eta_k(1,\alpha)$ to be a multiple of $1-\alpha$. As will be seen, this allows the $\alpha = 0$ case to be used to study the arbitrary α case. As the proof of Theorem 4 requires symmetric function theory, we delay the proof to Section 2.7 to allow a brief introduction to the relevant symmetric function theory.

Proof of Theorem 3 for polynomial ϕ. Formally, let $\tilde{m}(x)$ be the moment generating function for the ensemble, and expand each moment asymptotically around $n = \infty$, i.e.

$$\tilde{m}(x) = \frac{1}{n}\sum_{k=0}^{\infty} \frac{\mathbb{E}\,\mathrm{tr}(A^k)}{x^k} = \sum_{k=0}^{\infty} x^{-k} \sum_{j=0}^{\infty} \eta_k(j,\alpha) n^{-j},$$

then one has, to order $\frac{1}{n}$,

$$\tilde{m}(x) = \sum_{k=0}^{\infty} x^{-k} \left(\eta_k(0,\alpha) + \frac{\eta_k(1,\alpha)}{n} \right) + O(n^{-2}).$$

The α-dependence of either of these terms is completely determined by Theorem 4, as $\eta_k(0,\alpha)$ can have no α dependence, and $\eta_k(1,\alpha)$ is a multiple of $(1-\alpha)$. Define $m_0(x)$ and $m_1(x)$ so that

$$\tilde{m}(x)\big|_{\alpha=0} = m_0(x) + \tfrac{1}{n} m_1(x) + O(n^{-2}).$$

In this notation, the palindromy shows that

$$\tilde{m}(x) = m_0(x) + (1-\alpha)\tfrac{1}{n} m_1(x) + O(n^{-2}).$$

Further, the $\alpha = 0$ case, for fixed n, is relatively simple. As observed by Sutton [113], the Jacobi matrix model tends to a deterministic one as $\alpha \to 0$; precisely, it has eigenvalues that are the roots of $J_n^{r,s}$, the Jacobi polynomial of degree n and parameters

$$r = n\left(\tfrac{b}{a} - 1\right), \quad s = n\left(\tfrac{1-b}{a} - 1\right).$$

Suppose that the roots of $J_n^{r,s}$ are given by $\{\lambda_i\}_{i=1}^n$. Then for $\alpha = 0$, the moment generating function takes on the form

$$\tilde{m}(x) = \frac{1}{n}\sum_{k=0}^{\infty}\sum_{i=1}^{n}\frac{\lambda_i^k}{x^k} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x-\lambda_i} = \frac{1}{n}\left(\ln J_n^{r,s}(x)\right)' = \frac{1}{n}\frac{J_n^{r,s\prime}(x)}{J_n^{r,s}(x)}.$$

Using the differential recurrence for Jacobi polynomials, it follows that $\tilde{m}(x)$ satisfies a formal power series equation

$$\tilde{m}^2 + \frac{\frac{r+1}{n} - x\frac{r+s+2}{n}}{x(1-x)}\tilde{m} + \frac{1 + \frac{r+s+1}{n}}{x(1-x)} + \frac{\tilde{m}'}{n} = 0. \tag{2.42}$$

It follows that the constant-order term m_0 satisfies

$$am_0^2 + \frac{b-a-(1-2a)x}{x(1-x)}m_0 + \frac{1-a}{x(1-x)} = 0.$$

This leads to an explicit form for m_0,

$$\begin{aligned} m_0 &= \frac{(a-b)+(1-2a)x - \sqrt{(b-a-(1-2a)x)^2 - 4a(1-a)x(1-x)}}{2ax(1-x)} \\ &= \frac{(a-b)+(1-2a)x - \sqrt{(x-\lambda_-)(x-\lambda_+)}}{2ax(1-x)}, \end{aligned}$$

where

$$\lambda_\pm = \left[\sqrt{b(1-a)} \pm \sqrt{a(1-b)}\right]^2.$$

Note that $\lambda_\pm$ are always real, and that they are always on $[0,1]$. They are 0 and 1 exactly when $a = b$ and when $a = 1-b$, respectively. Taking an inverse Stieltjes transform gives absolutely continuous part

$$d\mu(x) = \frac{\sqrt{-(x-\lambda_-)(x-\lambda_+)}}{2\pi a x(1-x)}\mathbf{1}_{[\lambda_-,\lambda_+]}.$$

This integrates to 1, as it can be shown that

$$\int_{\lambda_-}^{\lambda_+} \frac{\sqrt{-(x-\lambda_-)(x-\lambda_+)}}{x(1-x)} = \pi\left[1-\sqrt{\lambda_-\lambda_+}-\sqrt{(1-\lambda_-)(1-\lambda_+)}\right] = 2\pi a.$$

Note that this implies that the distribution has no discrete part.

In the same fashion, one can also derive an explicit form for m_1. Pulling out the $\frac{1}{n}$ terms from (2.42), one is left with

$$2am_0m_1 + \frac{(b-a)-(1-2a)x}{x(1-x)}m_1 + \frac{1-2x}{x(1-x)}am_0 + \frac{a}{x(1-x)} + am_0' = 0.$$

Solving for m_1,

$$m_1 = \frac{-x+\frac{1}{2}(\lambda_-+\lambda_+)+\sqrt{(x-\lambda_+)(x-\lambda_-)}}{2(x-\lambda_+)(x-\lambda_-)}.$$

To recover the density, one again applies the inverse Stieltjes transform. When x is neither λ_+ nor λ_-, the limit $\lim_{\epsilon\to 0} m_1(x+i\epsilon)$ exists, and

$$\lim_{\epsilon\to 0} m_1(x+i\epsilon) = -\frac{1}{2\pi\sqrt{-(x-\lambda_+)(x-\lambda_-)}}\mathbf{1}_{(\lambda_-,\lambda_+)}(x).$$

Computing the inverse Stieltjes transform at either of the poles, it is seen that there are point masses, so that the entire signed measure is

$$\nu(x) = \tfrac{1}{4}\delta_{\lambda_-}(x) + \tfrac{1}{4}\delta_{\lambda_+}(x) - \frac{1}{2\pi\sqrt{-(x-\lambda_+)(x-\lambda_-)}}\mathbf{1}_{(\lambda_-,\lambda_+)}(x).$$

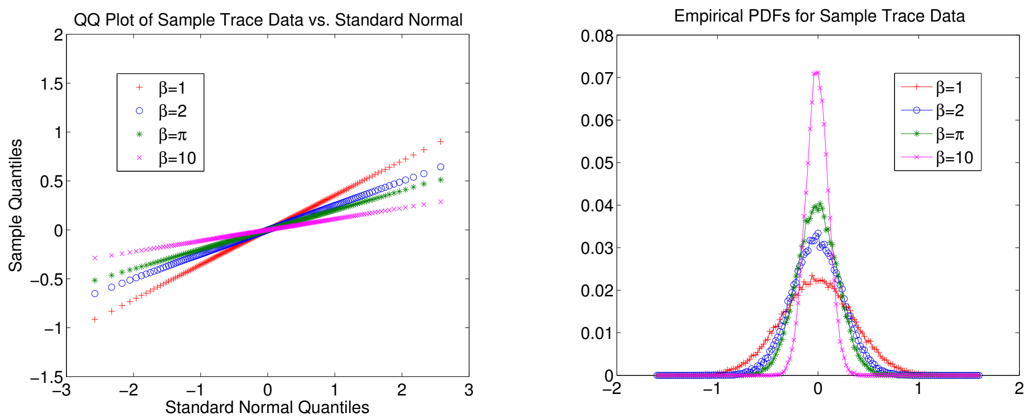
□

2.6 Numerics for the Extremal Case

In this section, we investigate the choice $p = q = 1$, which was not covered by Theorem 2. The method of proof breaks down in this extreme case, and so we have run a numerical simulation to help conjecture if the theorem extends.

In the alternate parameterization we have that $a = \frac{1}{2}$ and $b = \frac{1}{2}$. The density of the Jacobi ensemble becomes

$$d\mu_J(\lambda_1,\ldots,\lambda_n) = \tfrac{1}{Z}\prod_i \lambda_i^{\frac{\beta}{2}-1}(1-\lambda_i)^{\frac{\beta}{2}-1}\prod_{i<j}|\lambda_i-\lambda_j|^\beta. \tag{2.43}$$



(a) Quantile plots for $tr(A)$ experiments (b) PDF plots for $tr(A)$ experiments

Figure 2.1: Experimental data for different values of β, with $n = 5000$ and 50000 samples of each. All experiments were run in Matlab R2010B, using the Edelman-Sutton matrix model.

Note that the constraining potential no longer carries any dependence on n. However, because the particles are forced to lie on $[0, 1]$ (physically speaking, they are trapped in an infinite potential well), it is likely that we have some limiting behavior. For polynomial test functions and $\beta = 2$, this case is covered by a theorem of Johansson (see Theorem 3.1 of [67]).

However, the method of proof used here breaks down in the case $a < \frac{1}{2}$, as it requires the entries of the sparse matrix model to have uniform variance estimates on the order of n^{-1}. When $a = \frac{1}{2}$, the matrix model entries are

$$c_i \sim \sqrt{\text{Beta}(\tfrac{\beta}{2}i, \tfrac{\beta}{2}i)} \quad \text{and} \quad c_i' \sim \sqrt{\text{Beta}(\tfrac{\beta}{2}i, \tfrac{\beta}{2}(i+1))}.$$

The variances of entries c_i and s_i are on the order of i^{-1}, for which reason many of the arguments in later sections are no longer valid. To see how different the $a = b = \frac{1}{2}$ case is from the $a < \frac{1}{2}$ case, consider taking $f(x) = x$. It is easily seen that

$$X_{x^1,A} \to \sum_{i=1}^{\infty}(c_i^2 - \tfrac{1}{2})(1 - (c_i')^2 - (c_{i-1}')^2),$$

with the convergence in L^2. Note that while a normal limit is expected if the summands are becoming infinitesimal (and this is what happens when $a < \frac{1}{2}$), the normal limit here must follow from something else; in particular, the staircase dependency structure of the variables can not be ignored. We invite the reader to check that the variable is symmetric and to note how much cancellation occurs in computing the second and fourth moments (they are $1/(8\beta)$ and $3/(64\beta^2)$ respectively). Again, the fact that this variable is normally distributed follows from the mentioned theorem of Johansson.

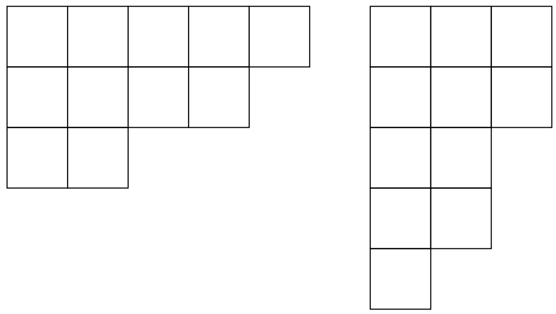
2.7 Symmetric Functions

To find the asymptotic distribution of the traces, we will appeal to Kadell's integral formula [70]. This formula makes use of Jack functions, and so we will provide a skeletal introduction to the relevant portions of symmetric function theory. A more expansive treatment is available in Macdonald's book [85], whose notation we will follow.

By a partition λ, we mean a non-increasing sequence of positive integers. The notation $\lambda \vdash n$, read 'λ partitions n,' means that the sum of the parts of λ equal n. There is an important pictorial representation of a partition called a Young diagram. The diagram representation of a partition $(\lambda_1, \ldots, \lambda_n)$ is drawn by placing λ_1 boxes horizontally in a row, placing λ_2 boxes horizontally below that, continuing through n and left justifying each row. Having drawn a diagram representation, we can easily define the *conjugate*[6] partition λ' to be that partition represented by reflecting the diagram across the vertical axis and rotating counterclockwise by a quarter turn.

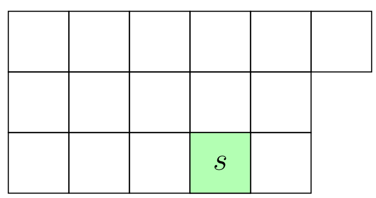
Example 3. The partition $\lambda = (5, 4, 1)$ is to the left, and its conjugate $\lambda' = (3, 3, 2, 2, 1)$ is to the right.

[6] This is also called the *transpose*.



Many formulas in symmetric function theory have sums or products computed from statistics of the diagram representation. For our purposes, we will need the arm length a, arm co-length a', leg length l, and leg co-length l' of a box s. The statistics $a(s)$ and $a'(s)$ are the number of boxes to the right and to the left of box s, respectively. Likewise, the statistics $l(s)$ and $l'(s)$ are the number of boxes below and above box s.

Example 4. This is $\lambda = (6, 5, 5)$.



$a(s) = 1$
$a'(s) = 3$
$l(s) = 0$
$l'(s) = 2$

The ring of symmetric functions Λ, are all those formal power series with complex coefficients[7] in the indeterminates $\{x_1, x_2, \ldots\}$, that are symmetric under permutation of the indices. In this application, the symmetric functions will be evaluated at some point $y = (y_1, y_2, \ldots, y_n) \in \mathbf{C}^n$, where it is understood that $f(y) = f(y_1, y_2, \ldots, y_n, 0, 0, \ldots)$. In this way, symmetric functions specialize to symmetric polynomials.

The symmetric functions of interest here are the power sums, as they describe traces. For an integer k, define p_k by

$$p_k = x_1^k + x_2^k + x_3^k + \cdots ,$$

[7] More often in the literature on Jack functions, these coefficients are defined to be from $\mathbf{Q}(\alpha)$, but the distinction here is immaterial.

and for a partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$, define p_λ by

$$p_\lambda = p_{\lambda_1} p_{\lambda_2} \cdots p_{\lambda_n}.$$

These are called the *power sum symmetric functions,* and $\{p_\lambda\}_\lambda$ are a basis for Λ. Note that the trace of a power of a matrix $\operatorname{tr} A^k$ can alternately be expressed as p_k evaluated at the eigenvalues of A.

The second basis we require are the *Jack symmetric functions* P_λ^α. For those interested, there is a concise introduction available in Stanley's paper [110]. By virtue of being a basis, it is possible to write p_k as a finite linear combination of $\{P_\lambda^\alpha\}_{\lambda \vdash k}$.

There are multiple normalizations for the Jack functions in the literature. In citing some theorems, we will require a second normalization, J_λ^α. The two are related, as $J_\lambda^\alpha = c(\lambda, \alpha) P_\lambda^\alpha$, where

$$c(\lambda, \alpha) = \prod_{s \in \lambda} (\alpha a(s) + l(s) + 1), \tag{2.44}$$

using the arm length $a(s)$ and leg length $l(s)$.

One final tool we will use is the Macdonald automorphism ω_α. It is defined in terms of the symmetric power functions by $\omega_\alpha p_k = \alpha p_k$; it is extended to each p_λ as a multiplicative homomorphism; and at last it is extended to all Λ as a $\mathbf{C}$-linear transformation. This automorphism acts on the Jack functions in a nice way as well, as by a formula of Stanley [110],

$$\omega_{-1/\alpha} J_{\lambda'}^{\alpha^{-1}} = (-\alpha)^{|\lambda|} J_\lambda^\alpha. \tag{2.45}$$

2.7.1 Kadell's integral

Kadell's integral (see [70]) is a generalization of Selberg's integral [104], which states the following

$$\int_{[0,1]^n} \prod_{i<j} |x_i - x_j|^{2/\alpha} \prod_{i=1}^n x_i^{r-1} (1 - x_i)^{s-1} dx = \prod_{i=1}^n \frac{\Gamma(1 + \frac{i}{\alpha}) \Gamma(r + \frac{i-1}{\alpha}) \Gamma(s + \frac{i-1}{\alpha})}{\Gamma(1 + \frac{1}{\alpha}) \Gamma(r + s + \frac{n+i-2}{\alpha})}.$$

It was generalized to include the Jack function $P_\lambda^{1/\alpha}(x)$ in the integrand. Letting $W(n,\alpha,r,s)$ be the integrand of Selberg's integral, Kadell's integral is

$$\int_{[0,1]^n} P_\lambda^{1/\alpha}(x)W(n,\alpha,r,s)dx = n!v_\lambda^\alpha \prod_{i=1}^{n} \frac{\Gamma(\lambda_i + r + \frac{n-i}{\alpha})\Gamma(s+\frac{n-i}{\alpha})}{\Gamma(\lambda_i + r + s + \frac{2n-i-1}{\alpha})}, \tag{2.46}$$

where the term v_λ^α is defined as

$$v_\lambda^\alpha = \prod_{i<j} \frac{\Gamma(\lambda_i - \lambda_j + \frac{j-i+1}{\alpha})}{\Gamma(\lambda_i - \lambda_j + \frac{j-i}{\alpha})}. \tag{2.47}$$

Our goal is to show that

$$\int_{[0,1]^n} \frac{P_\lambda^{1/\alpha}}{P_\lambda^{1/\alpha}(I_n)} W(n,\alpha,r,s)dx,$$

where $I_n = (1,1,\dots 1)$ has n $1's$, has a quasi-palindromic property. The constant $P_\lambda^{1/\alpha}(I_n)$ is computable in terms of diagram statistics. From formula VI.10.20 of [85],

$$P_\lambda^{1/\alpha}(I_n) = \prod_{s\in\lambda}\left(\frac{n+\alpha a'(s)-l'(s)}{\alpha a(s)+l(s)+1}\right) = \frac{1}{c(\lambda,\alpha)}\prod_{s\in\lambda}(n+\alpha a'(s)-l'(s)) \tag{2.48}$$

where $c(\lambda,\alpha)$ is the constant that relates J_λ^α and P_λ^α (see (2.44)). To compare the two, we will convert Kadell's expression using Γ functions into a Young diagram formula.

Recall that a quotient of Γ functions, also known as the Pochhammer symbol $(x)_k$, may be expressed alternately as

$$\frac{\Gamma(x+k)}{\Gamma(x)} = (x)_k = (x)(x+1)\cdots(x+k-1),$$

when k is a natural number. Define the generalized Pochhammer symbol $(t)_\mu$ (also known as the shifted factorial) to be

$$(t)_\mu = \prod_{s\in\mu}\left(t + a'(s) - \tfrac{1}{\alpha}l'(s)\right). \tag{2.49}$$

In terms of these expressions, (2.48) can be rewritten as

$$P_\lambda^{1/\alpha}(I_n) = \frac{(\frac{n}{\alpha})_\lambda(\alpha)^{|\lambda|}}{c(\lambda,\alpha)}. \tag{2.50}$$

We will need a closely related quantity to $c(\lambda,\alpha)$, so define $c'(\lambda,\alpha)$ to be

$$\prod_{s\in\lambda}(\alpha a(s)+l(s)+\alpha).$$

Both $c(\lambda,\alpha)$ and $c'(\lambda,\alpha)$ can be expressed as products of Γ terms, which we will need to rewrite Kadell's integral. Write out the terms in $\alpha^{-|\lambda|}c'(\lambda,\alpha)$ by going from right to left along the first row of the diagram of λ. There are $\lambda_1-\lambda_2$ terms that have $l(s)=0:$

$$(\tfrac{0}{\alpha}+1+0)(\tfrac{0}{\alpha}+1+1)\cdots(\tfrac{0}{\alpha}+1+\lambda_1-\lambda_2-1)=\frac{\Gamma(\lambda_1-\lambda_2+1)}{\Gamma(1)}.$$

There are then $\lambda_2-\lambda_3$ terms that have $l(s)=1:$

$$(\tfrac{1}{\alpha}+1+\lambda_1-\lambda_2)(\tfrac{1}{\alpha}+1+\lambda_1-\lambda_2+1)\cdots(\tfrac{1}{\alpha}+1+\lambda_1-\lambda_3-1)=\frac{\Gamma(\lambda_1-\lambda_3+1+\frac{1}{\alpha})}{\Gamma(\lambda_1-\lambda_2+1+\frac{1}{\alpha})}.$$

This pattern continues until at last there are λ_n terms that have $l(s)=n-1:$

$$(\tfrac{n-1}{\alpha}+1+\lambda_1-\lambda_n)(\tfrac{n-1}{\alpha}+1+\lambda_1-\lambda_n+1)\cdots(\tfrac{n-1}{\alpha}+\lambda_1)=\frac{\Gamma(\lambda_1+1+\frac{n-1}{\alpha})}{\Gamma(\lambda_1-\lambda_n+1+\frac{n-1}{\alpha})}.$$

Writing out all the terms in the first row gives

$$\frac{\Gamma(\lambda_1-\lambda_2+1)}{\Gamma(1)}\frac{\Gamma(\lambda_1-\lambda_3+1+\frac{1}{\alpha})}{\Gamma(\lambda_1-\lambda_2+1+\frac{1}{\alpha})}\frac{\Gamma(\lambda_1-\lambda_4+1+\frac{2}{\alpha})}{\Gamma(\lambda_1-\lambda_3+1+\frac{2}{\alpha})}\cdots\frac{\Gamma(\lambda_1+1+\frac{n-1}{\alpha})}{\Gamma(\lambda_1-\lambda_n+1+\frac{n-1}{\alpha})}.$$

Inducting over the rows, it follows that $c'(\lambda,\alpha)$ can be written as

$$c'(\lambda,\alpha)=(\alpha)^{|\lambda|}\prod_{i<j}\frac{\Gamma(\lambda_i-\lambda_j+1+\frac{j-i-i}{\alpha})}{\Gamma(\lambda_i-\lambda_j+1+\frac{j-i}{\alpha})}\prod_{i=1}^{n}\Gamma(\lambda_i+1+\tfrac{n-i}{\alpha}) \tag{2.51}$$

If one does the same expansion along the first row for $c(\lambda,\alpha)$, one gets

$$\frac{\Gamma(\lambda_1-\lambda_2+\frac{1}{\alpha})}{\Gamma(\frac{1}{\alpha})}\frac{\Gamma(\lambda_1-\lambda_3+\frac{2}{\alpha})}{\Gamma(\lambda_1-\lambda_2+\frac{2}{\alpha})}\frac{\Gamma(\lambda_1-\lambda_4+\frac{3}{\alpha})}{\Gamma(\lambda_1-\lambda_3+\frac{3}{\alpha})}\cdots\frac{\Gamma(\lambda_1+\frac{n}{\alpha})}{\Gamma(\lambda_1-\lambda_n+\frac{n}{\alpha})}.$$

Repeating the analogous procedure for the rest of the rows, we eventually conclude

$$c(\lambda,\alpha)=(\alpha)^{|\lambda|}\prod_{i<j}\frac{\Gamma(\lambda_i-\lambda_j+\frac{j-i}{\alpha})}{\Gamma(\lambda_i-\lambda_j+\frac{j-i+1}{\alpha})}\prod_{i=1}^{n}\frac{\Gamma(\lambda_i+\frac{n-i+1}{\alpha})}{\Gamma(\frac{1}{\alpha})}. \tag{2.52}$$

Equations (2.51) and (2.52) allow (2.47) to be rewritten as

$$v_\lambda^\alpha = \tfrac{(\alpha)^{|\lambda|}}{c(\lambda,\alpha)} \prod_{i=1}^{n} \frac{\Gamma(\lambda_i + \frac{n-i+1}{\alpha})}{\Gamma(\frac{1}{\alpha})}. \tag{2.53}$$

We can repeat the same procedure as used for c and c' to show that $(t)_\lambda$ can be computed by

$$(t)_\lambda = \prod_{i=1}^{n} \frac{\Gamma(t - \frac{i-1}{\alpha} + \lambda_i)}{\Gamma(t - \frac{i-1}{\alpha})}. \tag{2.54}$$

This allows the expression in (2.53) for v_λ^α to be replaced by

$$v_\lambda^\alpha = \tfrac{(\alpha)^{|\lambda|}}{c(\lambda,\alpha)} \prod_{i=1}^{n} \frac{\Gamma(\lambda_i + \frac{n}{\alpha} - \frac{i-1}{\alpha})}{\Gamma(\frac{1}{\alpha})} \frac{\Gamma(\frac{n}{\alpha} - \frac{i-1}{\alpha})}{\Gamma(\frac{n}{\alpha} - \frac{i-1}{\alpha})} = \tfrac{(\alpha)^{|\lambda|}}{c(\lambda,\alpha)} \left(\tfrac{n}{\alpha}\right)_\lambda \prod_{i=1}^{n} \frac{\Gamma(\frac{i}{\alpha})}{\Gamma(\frac{1}{\alpha})}. \tag{2.55}$$

Combine this expression for v_λ^α with Kadell's integral formula (2.46) and the simplified expression (2.50) for $P_\lambda^{1/\alpha}(I_n)$ to get

$$\begin{aligned}
&\int_{[0,1]^n} \frac{P_\lambda^{1/\alpha}(x)}{P_\lambda^{1/\alpha}(I_n)} W(n,\alpha,r,s)dx \\
&\quad = \frac{c(\lambda,\alpha)}{(\frac{n}{\alpha})_\lambda (\alpha)^{|\lambda|}} n! \frac{(\alpha)^{|\lambda|}}{c(\lambda,\alpha)} \left(\tfrac{n}{\alpha}\right)_\lambda \prod_{i=1}^{n} \frac{\Gamma(\frac{i}{\alpha})}{\Gamma(\frac{1}{\alpha})} \prod_{i=1}^{n} \frac{\Gamma(\lambda_i + r + \frac{n-i}{\alpha})\Gamma(s + \frac{n-i}{\alpha})}{\Gamma(\lambda_i + r + s + \frac{2n-i-1}{\alpha})} \\
&\quad = \prod_{i=1}^{n} \frac{\Gamma(1+\frac{i}{\alpha})}{\Gamma(1+\frac{1}{\alpha})} \frac{\Gamma(\lambda_i + r + \frac{n-i}{\alpha})\Gamma(s + \frac{n-i}{\alpha})}{\Gamma(\lambda_i + r + s + \frac{2n-i-1}{\alpha})}.
\end{aligned}$$

Let μ_J be the $(\frac{1}{\alpha}, r, s)$-Jacobi ensemble measure on $[0,1]^n$. This has density function proportional to $W(n,\alpha,r,s)$, but it is appropriately renormalized to be a probability measure. This normalization is given by Selberg's integral.

The integral expression above can be rewritten as

$$\begin{aligned}
&\int_{[0,1]^n} \frac{P_\lambda^{1/\alpha}(x)}{P_\lambda^{1/\alpha}(I_n)} d\mu_J(x) \\
&\quad = \prod_{i=1}^{n} \frac{\Gamma(1+\frac{i}{\alpha})}{\Gamma(1+\frac{1}{\alpha})} \frac{\Gamma(\lambda_i + r + \frac{n-i}{\alpha})\Gamma(s + \frac{n-i}{\alpha})}{\Gamma(\lambda_i + r + s + \frac{2n-i-1}{\alpha})} \frac{1}{\int W(n,\alpha,r,s)dx} \\
&\quad = \prod_{i=1}^{n} \frac{\Gamma(\lambda_i + r + \frac{n-i}{\alpha})}{\Gamma(r + \frac{n-i}{\alpha})} \frac{\Gamma(r + s + \frac{2n-i-1}{\alpha})}{\Gamma(\lambda_i + r + s + \frac{2n-i-1}{\alpha})} \\
&\quad = \frac{\left(r + \frac{n-1}{\alpha}\right)_\lambda}{\left(r + s + \frac{2n-2}{\alpha}\right)_\lambda}.
\end{aligned} \tag{2.56}$$

2.7.2 Palindromy

Lemma 14. *Let*

$$\int_{[0,1]^n} \frac{P_\lambda^{1/\alpha}(x)}{P_\lambda^{1/\alpha}(I_n)} d\mu_J(x) = \sum_{k=0}^{\infty} \rho(k,\lambda,\alpha) n^{-k}$$

be the series expansion about $n = \infty$. The coefficients $\rho(k,\lambda,\alpha)$ are skew-palindromic in that

$$\rho(k,\lambda,\alpha) = (-\alpha)^k \rho(k,\lambda',\tfrac{1}{\alpha})$$

Proof. In the calculation that follows, let $f(\lambda,\alpha,t) = f(t) = \alpha a'(t) - l'(t)$, for tableau block $t \in \lambda$. Starting from the formula computed in (2.56), and applying formula (2.49) gives

$$\begin{aligned}
\frac{\left(\frac{nb}{a\alpha}\right)_\lambda}{\left(\frac{n}{a\alpha}\right)_\lambda} &= \prod_{t\in\lambda} \frac{nb + af(t)}{n + af(t)} \\
&= \prod_{t\in\lambda} \frac{b + \frac{a}{n} f(t)}{1 + \frac{a}{n} f(t)} \\
&= \prod_{t\in\lambda} \left(b\left(1 + \tfrac{a}{bn} f(t)\right) \sum_{k=0}^{\infty} \left(-\tfrac{a}{n} f(t)\right)^k \right) \\
&= b^{|\lambda|} \prod_{t\in\lambda} \left(1 + (1 - \tfrac{1}{b}) \sum_{k=1}^{\infty} \left(-\tfrac{a}{n} f(t)\right)^k \right).
\end{aligned}$$

Let $M(\lambda,k)$ be the collection of all k-element multisets sampled from λ. If $\tau \in M(\lambda,k)$ is such a multiset, let $m_\tau(t)$ denote the multiplicity of $t \in \tau$ and let $\epsilon_\tau(t)$ be the characteristic function for $t \in \tau$. The sum can be written as:

$$= b^{|\lambda|} \sum_{k=0}^{\infty} n^{-k} \left[\sum_{\tau\in M(\lambda,k)} \prod_{t\in\lambda} (-af(t))^{m_\tau(t)} \left(1 - \tfrac{1}{b}\right)^{\epsilon_\tau(t)} \right].$$

This gives an explicit form for the coefficients $f(k,\lambda,\alpha)$. Mapping λ to λ' induces a bijection mapping the collection $M(\lambda,k)$ to $M(\lambda',k)$. In the conjugate, the arm co-length a' and leg co-length l' are reversed, so that $f(t)$ becomes $\alpha l'(t) - a'(t)$. Thus

$f(\lambda, \alpha, t) = -\alpha f(\lambda', \alpha^{-1}, t)$, so that

$$\sum_{\tau \in M(\lambda,k)} \prod_{t \in \lambda} (-af(\lambda, \alpha, t))^{m_\tau(t)} \left(1 - \tfrac{1}{b}\right)^{\epsilon_\tau(t)}$$
$$= (-\alpha)^k \sum_{\tau \in M(\lambda',k)} \prod_{t \in \lambda'} \left(-af(\lambda', \alpha^{-1}, t)\right)^{m_\tau(t)} \left(1 - \tfrac{1}{b}\right)^{\epsilon_\tau(t)}$$

□

Let $J_\lambda^{1/\alpha}$ be the Jack functions renormalized by

$$J_\lambda^{1/\alpha} = c(\lambda, \alpha) P_\lambda^{1/\alpha}. \tag{2.57}$$

Expand the symmetric power function p_k as

$$p_k = \sum_{\lambda \vdash k} \xi(\lambda, \alpha) J_\lambda^{1/\alpha}.$$

By applying Stanley's formula (see (2.45)), it follows (see [34]) that

$$\xi(\lambda, \alpha) = (-\alpha)^{1-|\lambda|} \xi(\lambda', \alpha^{-1}). \tag{2.58}$$

One last piece is needed. The normalization factor $J_\lambda^{1/\alpha}(I_n)$ can be computed by relating (2.50) and the definition of J_λ^α in (2.57). These two combined give that

$$J_\lambda^{1/\alpha(I_n)} = \left(\tfrac{n}{\alpha}\right)_\lambda (\alpha)^{|\lambda|} = \prod_{t \in \lambda} (n + \alpha a'(t) - l'(t))\,;$$

expand this as a polynomial in n, i.e. put

$$\prod_{t \in \lambda} (n + \alpha a'(t) - l'(t)) = \sum_{j=0}^{|\lambda|} \zeta(j, \lambda, \alpha) n^j.$$

Because the product can be expressed as

$$\prod_{t \in \lambda} (n + \alpha a'(t) - l'(t)) = \prod_{t \in \lambda'} \left(n + (-\alpha)\left(\alpha^{-1} a'(t) - l'(t)\right)\right),$$

it follows that

$$\zeta(j, \lambda, \alpha) = (-\alpha)^{|\lambda|-j} \zeta(j, \lambda', \alpha^{-1}). \tag{2.59}$$

Proof of Theorem 4. Expand $p_{[k]}$ in the Jack function basis:

$$\begin{aligned}\tfrac{1}{n}\mathbb{E}_\alpha p_{[k]} &= \tfrac{1}{n}\sum_{\lambda\vdash k}\xi(\lambda,\alpha)\mathbb{E}_\alpha J_\lambda^{1/\alpha}\\ &= \tfrac{1}{n}\sum_{\lambda\vdash k}\xi(\lambda,\alpha)J_\lambda^{1/\alpha}(I_n)\mathbb{E}_\alpha\frac{J_\lambda^{1/\alpha}}{J_\lambda^{1/\alpha}(I_n)}.\end{aligned}$$

Apply Lemma 14, and expand $J_\lambda^{1/\alpha}(I_n)$. Note that the alternative normalization used in the Lemma cancels out.

$$\begin{aligned}\tfrac{1}{n}\mathbb{E}_\alpha p_{[k]} &= \tfrac{1}{n}\sum_{\lambda\vdash k}\xi(\lambda,\alpha)\left(\sum_{j=0}^{k}\zeta(j,\lambda,\alpha)n^j\right)\left(\sum_{j=0}^{\infty}\rho(j,\lambda,\alpha)n^{-j}\right)\\ &= \sum_{j=-\infty}^{k} n^{j-1}\left(\sum_{\lambda\vdash k}\xi(\lambda,\alpha)\sum_{l=0}^{k}\zeta(l,\lambda,\alpha)\rho(l-j,\lambda,\alpha)\right),\end{aligned}$$

with $\rho(l-j,\lambda,\alpha)=0$ for negative $l-j$.

This gives a formula for $\eta_k(j,\alpha)$, namely that

$$\eta_k(j,\alpha)=\sum_{\lambda\vdash k}\xi(\lambda,\alpha)\sum_{l=0}^{k}\zeta(l,\lambda,\alpha)\rho(l+j-1,\lambda,\alpha).$$

The $j<0$ terms vanish, which can be seen because the trace can naturally be bounded as

$$\tfrac{1}{n}|\mathbb{E}_\alpha p_k| \le \tfrac{1}{n}\mathbb{E}_\alpha\sum_{i=0}^{n}|x_i|^k \le \tfrac{1}{n}n = 1,$$

as the Jacobi distribution is supported on $[0,1]^n$.

We will show that each $\eta_k(j,\alpha)$ is palindromic. Applying Lemma 14, (2.58), and (2.59), these can be written as

$$\begin{aligned}\eta_k(j,\alpha) &= \sum_{\lambda\vdash k}\xi(\lambda,\alpha)\sum_{l=0}^{k}\zeta(l,\lambda,\alpha)\rho(l{+}j{-}1,\lambda,\alpha)\\ &= \sum_{\lambda\vdash k}(-\alpha)^{1-k}\xi(\lambda',\alpha^{-1})\sum_{l=0}^{k}(-\alpha)^{k-l}\zeta(l,\lambda',\alpha^{-1})(-\alpha)^{l+j-1}\rho(l{+}j{-}1,\lambda',\alpha^{-1})\\ &= (-\alpha)^{j}\sum_{\lambda\vdash k}\xi(\lambda',\alpha^{-1})\sum_{l=0}^{k}\zeta(l,\lambda',\alpha^{-1})\rho(l{+}j{-}1,\lambda',\alpha^{-1}).\end{aligned}$$

The sum is over all partitions of k, so taking conjugates makes no difference. Thus,

$$\eta_k(j,\alpha) = (-\alpha)^j\eta_k(j,\alpha^{-1}).$$

The last claim we make is that $\eta_k(j,\alpha)$ is a polynomial in α of degree j. This is more involved, and requires that we appeal to Edelman and Sutton's tridiagonal matrix model (see the start of Section 3). The moment $\frac{1}{n}\mathbb{E}p_k = \frac{1}{n}\mathbb{E}\,\mathrm{tr}(A^k)$ can be written in terms of a sum over alternating bridges (see Section 3.1),

$$\tfrac{1}{n}\mathbb{E}\,\mathrm{tr}\,A^k = \tfrac{1}{n}\sum_{\bar{w}\in\mathcal{A}_{2k}}\mathbb{E}\left(B_\beta\right)_{\bar{w}+i}.$$

A priori, these expectations are moments of random variables distributed as the square root of a Beta random variable. However, by Lemma 3, the alternating bridge visits each matrix entry an even number of times. Thus, any term in the sum takes the form

$$\mathbb{E}\prod_{i=1}^{k}c_{\omega_{2i-1}}^{2m_{2i-1}}s_{\omega_{2i-1}}^{2n_{2i-1}}c'^{2m_{2i}}_{\omega_{2i}}s'^{2n_{2i}}_{\omega_{2i}},$$

where ω_i ranges over the matrix entries referenced by the bridge $\bar{w}$ and $\sum_0^{2k} m_i = k$. By independence, this expectation is a product of terms of the form

$$\mathbb{E}c_\omega^{2m}s_\omega^{2n} \quad\text{and}\quad \mathbb{E}c'^{2m}_\omega s'^{2n}_\omega$$

By Lemma 15, each such Beta moment admits a series expansion around $n=\infty$ and a K so that

$$\mathbb{E}\left(B_\beta\right)_{\bar{w}+i} = \sum_{m=0}^{\infty}n^{-m}\alpha^m\Omega_{\bar{w}+i,m}(n),$$

where $0 < \Omega_{\bar{w}+i,m}(n) < K^m$ for all n. Moreover, this constant K can be chosen independently of $\bar{w}+i$. Thus the entire trace admits such a series expansion,

$$\begin{aligned}\tfrac{1}{n}\mathbb{E}\,\mathrm{tr}(A^k) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{\bar{w}\in\mathcal{A}_{2k}}\sum_{m=0}^{\infty} n^{-m}\alpha^m\Omega_{\bar{w}+i,m}(n)\\ &= \sum_{m=0}^{\infty} n^{-m}\alpha^m\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{\bar{w}\in\mathcal{A}_{2k}}\Omega_{\bar{w}+i,m}(n)\right).\end{aligned}$$

Because the cardinality of $\mathcal{A}_{2k}$ is at most $\binom{2k}{k}$, the sum $\Omega_m(n) = \frac{1}{n}\sum_{i=k+1}^{n-k}\sum_{\bar{w}\in\mathcal{A}_{2k}}\Omega_{\bar{w}+i,m}(n)$ satisfies an estimate $0 < \Omega_m(n) < \binom{2k}{k}K^m = CK^m$. Thus there are two expansions for the trace, valid for all n sufficiently large, i.e.

$$\sum_{j=0}^{\infty}\eta(j,\alpha)n^{-j} = \frac{1}{n}\mathbb{E}\,\mathrm{tr}(A^k) = \sum_{j=0}^{\infty}\Omega_j(n)\alpha^j n^{-j} \tag{2.60}$$

The left hand side expansion shows that the $n\to\infty$ limit must exist. Thus

$$\eta(0,\alpha) = \lim_{n\to\infty}\sum_{j=0}^{\infty}\eta(j,\alpha)n^{-j} = \lim_{n\to\infty}\sum_{j=0}^{\infty}\Omega_j(n)\alpha^j n^{-j} = \lim_{n\to\infty}\Omega_0(n).$$

In particular, $\eta(0,\alpha)$ has no α dependence. The proof now proceeds by induction. Suppose that for all $j < l$, the term $\eta(j,\alpha)$ is a polynomial in α of degree j. It should be shown that $\eta(l,\alpha)$ is a polynomial in α of degree l. The limit

$$\lim_{n\to\infty} n^l\left[\tfrac{1}{n}\mathbb{E}\,\mathrm{tr}(A^k) - \sum_{j=0}^{l-1}\eta(j,\alpha)n^{-j}\right] = \lim_{n\to\infty}\sum_{j=0}^{\infty}\eta(j+l)n^{-j} = \eta(l,\alpha)$$

exists by virtue of the η expansion, and by substituting the right hand side of (2.60), it follows that

$$\begin{aligned}\eta(l,\alpha) = \lim_{n\to\infty} n^l&\left[\sum_{j=0}^{\infty}\Omega_j(n)\alpha^j n^{-j} - \sum_{j=0}^{l-1}\eta(j,\alpha)n^{-j}\right]\\ &= \lim_{n\to\infty}\sum_{j=0}^{l-1}\left[\Omega_j(n)\alpha^j - \eta(j,\alpha)\right]n^{l-j} + \Omega_l(n)\alpha^l.\end{aligned}$$

By the inductive hypothesis, this limit can be written in the form

$$\eta(l,\alpha) = \lim_{n\to\infty} f_0(n) + f_1(n)\alpha + f_2(n)\alpha^2 + \cdots + f_l(n)\alpha^l,$$

and the limit exists for each fixed α. Take $l+1$ distinct values of α. The convergence is uniform on this finite set $\alpha_0, \dots, \alpha_l$, and so each $f_i(n)$ converges, where $0 \le i \le l$. Thus $\eta(j, \alpha)$ is a polynomial of degree l in α, concluding the proof. □

Lemma 15. *Let $f_r(n)$ and $f_s(n)$ be positive real-valued functions defined on $\mathbb{N}$ so that*

$$0 < f_r(n) \le C_2, \qquad 0 < f_s(n) \le C_2 i, \qquad C_1 < f_r(n) + f_s(n),$$

where C_i are some positive constants. Let $r = \alpha^{-1} f_r(n) n$, $s = \alpha^{-1} f_s(n) n$, and let $\mathbf{X} \sim \mathrm{Beta}(r, s)$. There is an asymptotic expansion

$$\mathbb{E}\left[\mathbf{X}^k (1-\mathbf{X})^l\right] = \sum_{m=0}^{\infty} n^{-m} \alpha^m p_m(n),$$

and a constant K depending only on $k, l, C_1,$ and C_2 so that $0 < p_m(n) < K^m$.

Proof. The expectation, which can be computed using Euler's Beta integral formula, gives that

$$\mathbb{E}\left[\mathbf{X}^k (1-\mathbf{X})^l\right] = \frac{(r)_k (s)_l}{(r+s)_{k+l}}.$$

Substituting in the definitions for r and s and writing out the Pochhammer symbols gives

$$\mathbb{E}\left[\mathbf{X}^k (1-\mathbf{X})^l\right] = \prod_{i=0}^{k-1} \frac{\alpha^{-1} f_r n + i}{\alpha^{-1}(f_r + f_s) n + i} \prod_{i=0}^{l-1} \frac{\alpha^{-1} f_s n + i}{\alpha^{-1}(f_r + f_s) n + k + i}.$$

All rational terms in this product produce similar asymptotic series expansions, and so we will only examine one. Working with a term from the left hand product,

$$\frac{\alpha^{-1} f_r n + i}{\alpha^{-1}(f_r + f_s) n + i} = \left(\frac{\alpha^{-1} f_r n + i}{\alpha^{-1}(f_r + f_s) n}\right) \left(\frac{1}{1 + \frac{i}{\alpha^{-1}(f_r + f_s) n}}\right).$$

Provided that n is sufficiently large (depending on C_1 and α), this can be expanded as a series.

$$\begin{aligned}\frac{\alpha^{-1} f_r n + i}{\alpha^{-1}(f_r + f_s)n + i} &= \frac{\alpha^{-1} f_r n + i}{\alpha^{-1}(f_r + f_s)n} \sum_{m=0}^{\infty} (f_r + f_s)^{-m} n^{-m} \alpha^m \\ &= \frac{f_r}{f_r + f_s} + \sum_{m=1}^{\infty} \left(\tfrac{f_r}{f_r + f_s} + i \right) (f_r + f_s)^{-m} n^{-m} \alpha^m \\ &= \sum_{m=1}^{\infty} \tilde{p}_m(n) n^{-m} \alpha^m.\end{aligned}$$

The coefficients $\tilde{p}_m(n)$ satisfy an estimate

$$0 < \tilde{p}_m(n) < (C_1 C_2 + k)(C_1)^{-m}.$$

□

2.8 *Poincaré Inequality for* Beta

Lemma 16. *Let* $Y \sim \mathrm{Beta}(p, q)$*. For any Lipschitz function* f *on* $[0, 1]$*,*

$$\operatorname{Var} f(Y) \le \frac{1}{4(p+q)} \mathbb{E} \left| f'(Y) \right|^2.$$

We note that in the case that both p and q are greater than 1, the density is log-concave, and it is possible to use the general theory outlined by Bobkov in [14] to produce an equivalent bound, but we require the inequality to hold for all p and q positive, and thus we use an alternative technique.

Proof. We begin by showing the analogous bound for the translated random variable $X = 2Y - 1$, and write $Y = T(X) := \frac{1}{2}(X + 1)$. The density of Y is given by

$$\frac{d\mu_\beta}{dx} = \frac{1}{Z_{p,q}} (1 - x)^{p-1} (1 + x)^{q-1}.$$

We will show that for any Lipschitz function f on $[-1, 1]$, that

$$\operatorname{Var} f(X) \le \frac{1}{p+q} \mathbb{E} \left[(1 - X^2) \left| f'(X) \right|^2 \right]. \tag{2.61}$$

As will be seen in the proof, this inequality is attained taking f to be a multiple of the linear Jacobi polynomial (for definitions, see [114]). The proof follows from (2.61), as

$$\begin{aligned}
\operatorname{Var} f(Y) &= \operatorname{Var}(f \circ T)(X) \\
&\leq \frac{1}{p+q}\mathbb{E}\left[(1-X^2)\left|(f\circ T)'(X)\right|^2\right] \\
&\leq \frac{1}{p+q}\mathbb{E}\left|(f\circ T)'(X)\right|^2 \\
&= \frac{1}{p+q}\mathbb{E}\left|(f'\circ T)(X)\right|^2 \frac{1}{2}^2 \\
&= \frac{1}{4(p+q)}\mathbb{E}\left|f'(Y)\right|^2.
\end{aligned}$$

The method of proof follows the general outline in the notes of Bakry [9]. Define the Jacobi differential operator L to be

$$Lf = (1-x^2)f''(x) + (q-p-(p+q)x)f'(x),$$

and define the *carré du champ* operator Γ by

$$\Gamma(f,g) = (1-x^2)f'(x)g'(x).$$

It can be checked by integration by parts that for all C^2 functions on $[-1,1]$ that the Dirichlet form $\mathcal{E}(f,g)$ associated to L satisfies

$$\mathcal{E}(f,g) := -\int_{-1}^{1} f(x)(Lg)(x)d\mu_\beta(x) = \int_{-1}^{1}\Gamma(f(x),g(x))d\mu_\beta(x).$$

The spectrum of L restricted to $L^2(\mu_\beta)$ is non-positive, with eigenvalues $y_n = -n(n+p+q-1)$ for non-negative integers n. Further, its eigenfunctions are given by the Jacobi polynomials $P_n^{p-1,q-1}(x)$, which when normalized form a complete orthonormal system for $L^2(\mu_\beta)$. From the density of the polynomials in $L^2(\mu_\beta)$, it is an immediate consequence that

$$p+q = -y_1 = \inf_{\substack{f\in L^2(\mu)\\ \mathbb{E}f(X)=0}} \frac{\mathcal{E}(f,f)}{\operatorname{Var} f(X)} = \inf_{\substack{f\in L^2(\mu)\\ \mathbb{E}f(X)=0}} \frac{\int_{-1}^{1}\Gamma(f(x),g(x))d\mu_\beta(x)}{\operatorname{Var} f(X)},$$

which upon rewriting, gives (2.61). □

2.9 Coupling Bound for $\sqrt{\mathrm{Beta}}$

We provide an auxiliary lemma regarding the square root of Beta variables that appear in the matrix entries. Note that because one of the parameters of the c_i' family is not $\Omega(n)$ for all i, this approximation can not be applied to every matrix entry with uniform error.

Lemma 17. *If Y is distributed as $\sqrt{Beta(np,nq)}$, then*

$$\frac{2(p+q)}{\sqrt{q}}\sqrt{n}\left(Y-\sqrt{\frac{p}{p+q}}\right)\Rightarrow N(0,1),$$

as $n\to\infty$, where p,q are fixed positive constants. Moreover, it is possible to couple Y to a standard normal X so that

$$\mathrm{Var}\left(Y-\frac{\sqrt{q}}{2(p+q)\sqrt{n}}X\right)\leq\frac{K_{p,q}}{n^2},$$

for some $K_{p,q}>0$, independent of n, and continuous in p,q positive, provided that $n>\max\{\frac{1}{p},\frac{1}{q}\}$.

Proof of 17. Let Y be distributed as $\sqrt{\mathrm{Beta}(np,nq)}$. Put

$$\mu=\sqrt{\frac{p}{p+q}},\qquad\sigma=\sqrt{\frac{q}{2n(p+q)}}.$$

Note that these are not exactly the mean or standard deviation of Y, however,

$$\tilde{Y}=\frac{Y-\mu}{\sigma}\Rightarrow N(0,1).$$

Moreover, it will be shown that there is an X distributed as $N(0,1)$ so that

$$\mathbb{E}(\tilde{Y}-X)^2\leq\frac{K}{n}.$$

for some $K=K(p,q)$ depending continuously on p,q positive. Note that this implies Lemma 17 after dividing through by σ.

The primary machinery here is Talagrand's transport inequality, which bounds the squared L^2-Wasserstein distance of $\tilde{Y}$ and X, with X distributed as $N(0,1)$. We use a special case of Theorem 1.1 of [115], which states

Proposition 12 (Talagrand)**.** *Let $\tilde{Y}$ be a random variable given by probability measure $\tilde{\nu}$, which is absolutely continuous with Lebesgue measure, and let γ be a standard Gaussian measure. There is a standard normal random variable X so that*

$$\mathbb{E}(\tilde{Y}-X)^2 \leq 2\int \log \tfrac{d\tilde{\nu}}{d\gamma} d\tilde{\nu}.$$

The density $\frac{d\nu}{dy}$ of Y can be computed to be

$$\frac{d\nu}{dy} = 2y(y^2)^{np-1}(1-y^2)^{nq-1}\frac{\Gamma(np+nq)}{\Gamma(np)\Gamma(nq)}$$

for $y \in [0,1]$. It follows that density of $\tilde{Y}$ is given by

$$\frac{d\tilde{\nu}}{dy} = 2\sigma(\mu+y\sigma)^{2np-1}(1-(\mu+y\sigma)^2)^{nq-1}\frac{\Gamma(np+nq)}{\Gamma(np)\Gamma(nq)},$$

and thus the Radon-Nikodym derivative $\frac{d\tilde{\nu}}{d\gamma}(y)$ is a product of four terms

$$\frac{d\tilde{\nu}}{d\gamma}(y) = \underbrace{(\mu+y\sigma)^{2np-1}}_{\text{(i)}}\,\underbrace{(1-(\mu+y\sigma)^2)^{nq-1}}_{\text{(ii)}}\,\underbrace{e^{y^2/2}}_{\text{(iii)}}\,\underbrace{2\sigma\frac{\Gamma(np+nq)}{\Gamma(np)\Gamma(nq)}\sqrt{2\pi}}_{(iv)}.$$

The logs of terms (i) and (ii) can be controlled by Taylor expansion. Explicitly,

$$\ln[1+y]^g = g\ln(1+y) \leq g\left[y - \frac{y^2}{2} + \frac{y^3}{3}\right],$$

for all $y > -1$, and all $g > 0$. Note that both produce a nonzero constant term, by virtue of the relationship $\ln(a+y) = \ln(a) + \ln(1+y/a)$. This bound is applied to the logs of both (i) and (ii) after suitable rearrangement. This bounds the sum of the logs by a polynomial in y of degree 6. We can bound the log of term (i) as

$$\begin{aligned}\ln\left[(\mu+y\sigma)^{2np-1}\right] &= (2np-1)\ln\mu + (2np-1)\ln\left[1+\tfrac{y\sqrt{q}}{\gamma\sqrt{2pn}}\right]\\ &\leq (2np-1)\left[\ln\mu + \tfrac{y\sqrt{q}}{\gamma\sqrt{2pn}} - \tfrac{1}{2}\left(\tfrac{y\sqrt{q}}{\gamma\sqrt{2pn}}\right)^2 + \tfrac{1}{3}\left(\tfrac{y\sqrt{q}}{\gamma\sqrt{2pn}}\right)^3\right].\end{aligned}$$

Applying the same to term (ii),

$$\begin{aligned}&\ln\left[(1-(\mu+y\sigma)^2)^{nq-1}\right] \leq\\ &\quad (nq-1)\left[\ln(1-\mu^2) + u - \tfrac{1}{2}u^2 + \tfrac{1}{3}u^3\right],\end{aligned}$$

where $u = 2\frac{y\sqrt{p}}{\sqrt{2qn}} + \left[\frac{y}{\sqrt{2n}}\right]^2$.

From this form, it is easy to see that the coefficients of this polynomial depend continuously on p and q. Further, the coefficients of $y^4, y^5,$ and y^6 already decay at least as fast as $1/n$. The coefficient of y^3 decays like $n^{-1/2}$, so some amount of control over $\mathbb{E}\tilde{Y}^3$ will need to be gained. The coefficients of the lower order terms to do not *a priori* decay at all, but there is strong cancellation. The constant term is

$$C_0(p,q) := (2\,n\,p - 1)\ln(\mu) + (n\,q - 1)\ln\left(1 - \mu^2\right),$$

the linear term has coefficient

$$\frac{C_1(p,q)}{\sqrt{n}} := \frac{(2\,n\,p-1)\,\sigma}{\mu} - 2\,\frac{(n\,q-1)\,\mu\,\sigma}{1-\mu^2},$$

and the quadratic term has coefficient

$$-\frac{1}{2} + \frac{C_2(p,q)}{n} := -1/2\,\frac{(2\,n\,p-1)\,\sigma^2}{\mu^2} + (n\,q-1)\left(-\frac{\sigma^2}{1-\mu^2} - 2\,\frac{\mu^2\sigma^2}{(1-\mu^2)^2}\right).$$

The $-\frac{1}{2}$ in the quadratic term represents the asymptotically Gaussian portion, and it annihilates term (iii). This leaves four sources of error that need to be controlled to show the desired $O(n^{-1})$ bound:

1. $|\mathbb{E}\tilde{Y}| \leq C(p,q)n^{-\frac{1}{2}}$ to control the linear term.

2. $|\mathbb{E}\tilde{Y}^3| \leq C(p,q)n^{-\frac{1}{2}}$ to control the cubic term.

3. $|\mathbb{E}(\tilde{Y})^k| \leq C(p,q)$ to control the second, fourth, fifth, and sixth terms.

4. The constants from the Taylor approximation and the constants from part (iv) of the Radon-Nikodym derivative need to cancel to order $O(n^{-1})$.

The raw moments of Y are easily computable, and their formula follows immediately from Euler's Beta integral,

$$\mathbb{E}\,(Y)^k = \frac{\Gamma(n(p+q))\Gamma(np+\frac{k}{2})}{\Gamma(n(p+q)+\frac{k}{2})\Gamma(np)}.$$

Appropriate control over the first 6 raw moments could be achieved by taking sufficiently many terms from the Stirling approximation and canceling terms. To some extent, doing such a procedure is necessary, as this is necessary to get the precise control over the first and third raw moments. However, we will not need to do this for all 6 moments, because we can appeal to a Poincaré inequality. Provided that $n > \max\{\frac{1}{p}, \frac{1}{q}\}$, the density $\frac{d\tilde{\nu}}{dy}$ is log-concave. Thus if it can be shown that $\tilde{Y}$ has constant order variance, we can use the Poincaré inequality to bound higher moments by lower moments, i.e.

$$\operatorname{Var} f(\tilde{Y}) \leq C\mathbb{E}|f'(\tilde{Y})|^2,$$

applied to $f(\tilde{Y}) = (\tilde{Y})^k$, gives

$$\mathbb{E}\tilde{Y}^{2k} \leq \left(\mathbb{E}\tilde{Y}^k\right)^2 + Ck^2\mathbb{E}\tilde{Y}^{2k-2}.$$

Because of the log-concavity, C can be taken to be $12\mathbb{E}|\tilde{Y}|^2$ (see Corr 4.3 of [14]), which is continuous in p and q. Thus provided that $\mathbb{E}|\tilde{Y}|$ can be bounded by some continuous function in p and q, iterating the Poincaré inequality gives constant order bounds that are continuous in p and q for all absolute moments. Further,

$$\mathbb{E}|\tilde{Y}| \leq \sqrt{\mathbb{E}\left(\tilde{Y}\right)^2},$$

so the problem has been reduced to finding good bounds for the first three raw moments of $\tilde{Y}$.

By appealing to Stirling's formula, and using that the error-in-approximation is bounded by the first omitted term in the asymptotic expansion, the first three moments of $\tilde{Y}$ can be bounded by

$$\begin{aligned}
\left|\mathbb{E}\left(\tilde{Y}\right)\right| &\leq \frac{\sqrt{q}}{4\sqrt{p(p+q)}} n^{-\frac{1}{2}}, \\
\left|\mathbb{E}\left(\tilde{Y}\right)^2\right| &\leq 1, \\
\left|\mathbb{E}\left(\tilde{Y}\right)^3\right| &\leq \frac{8p+q}{4\sqrt{qp(p+q)}} n^{-\frac{1}{2}}.
\end{aligned}$$

It only remains to control the constant terms. The log of (iv) can be approximated by Stirling's formula:

$$\left|\ln\left[2\sigma\frac{\Gamma(n(p+q)}{\Gamma(np)\Gamma(nq)}\sqrt{2\pi}\right] - \left[-np\ln\mu^2 - nq\ln(1-\mu^2) + \ln\frac{q\sqrt{p}}{(p+q)^{\frac{3}{2}}}\right]\right| \leq \frac{1}{12}\ \frac{1}{n}\ \frac{1}{\sqrt{p}\sqrt{q}\sqrt{p+q}}.$$

Comparing this with the constants produced by the Taylor approximation on terms (i) and (ii), it is seen that only the $O(n^{-1})$ term remains.

□

Chapter 3

FLUCTUATIONS OF d-REGULAR GRAPHS

ADAPTED FROM JOINT WORK WITH IOANA DUMITRIU, TOBIAS JOHNSON, AND SOUMIK PAL [35].

3.1 Introduction

While the results for fluctuations of linear statistics in the case of the β-Jacobi ensemble agreed marvelously with that which could be predicted from Gaussian behavior, this story becomes a bit more nuanced in the case of d-regular graphs.

Recall that a graph is called regular if every vertex has the same degree; a sparse regular graph is typically one for which the degree d is either constant or growing logarithmically in the number of vertices n. As we shall see, sparseness plays a critical role in how d-regular graphs behave.

As for the model of regular graph, the classical model is the uniform distribution over all d-regular graphs on n labeled vertices; a thorough survey on properties of the uniform model can be found in [127]. Our model of choice is the *permutation model*: consider d many i.i.d. uniformly chosen permutations $\{\pi_1, \ldots, \pi_d\}$ on n vertices labeled $\{1, 2, \ldots, n\}$. A graph can be constructed by adding one edge between each pair $(i, \pi_j(i))$; thus every vertex i has edges to $\pi_j(i)$ and $\pi_j^{-1}(i)$ for every permutation π_j, for a total degree of $2d$. As the reader will note, this allows multiples edges and self-loops, with each self-loop contributing two to the degree of its vertex.

Another way to represent this graph is by its adjacency matrix, which is an $n \times n$ matrix whose (i, j) entry is the number of edges between i and j, with self-loops counted twice. In this case, the adjacency matrix of the permutation model A_n can be defined by sampling d i.i.d. uniformly chosen permutation matrices $P_1, P_2, \ldots, P_d$

and forming

$$A_n := P_1 + P_1^t + P_2 + P_2^t + \cdots + P_d + P_d^t.$$

Note that the top eigenvalue is trivially $2d$; the distribution of the rest of the eigenvalues is an interesting question. For the uniform model of random regular graphs (or Erdős-Rényi graphs) such questions have been studied since the pioneering work [89]. Many results about the permutation model are transferable to other models by virtue of various contiguity results when d is fixed; see [127, Section 4] and [57]. These state that a property that holds with high probability in one model holds with high probability in another.

Among the more recent work, see [42], [117], and [98]. We refer the reader to [98] for a more exhaustive review of the vast related literature. Most importantly, we would like to remark that for a single permutation matrix, such a study has been approached in [122] and completed in [6]; our results share several features with the latter paper.

3.1.1 Scaling

We would like to scale the eigenvalues of the matrix to be at unit order. Unlike before, we have two parameters to choose, d and n. We would like to allow for d to either stay fixed in n or for $d = d(n)$ to grow slowly with n. We will assume throughout that $d \geq 2$; the reason for this is that the $d = 1$ case has been dealt with (in a larger context) by [6]. By considering the square Frobenius norm, we see that we must scale the off-diagonal entries to have variance $O(1/n)$. Each off-diagonal entry of $P_i + P_i^t$ is either $0, 1, 2$ but with probabilities so that it is very nearly Bernoulli$(2/n)$ distributed. Thus the distribution of an off-diagonal entry of the adjacency matrix A_n is very nearly Binom$(d, 2/n)$ for a variance of about $2d/n$. From the usual $\operatorname{tr}(A_n^2)$ estimate we should expect the nontrivial portion of the spectrum to be about $O(\sqrt{d})$.

In fact, as we will see, to scale the eigenvalues to lie on $[-1, 1]$, we define

$$\mathcal{P}_{n,d} := \frac{A_n}{2\sqrt{2d-1}}.$$

3.1.2 Our results

We shall see that for d fixed and $n \to \infty$, the centered linear statistics converge to linear functionals of a field of independent Poisson variables. In the case of $d = d(n) \to \infty$ slowly with n, these linear statistics converge to the normals one would anticipate by analogy with real Wigner matrices. This transition is expected, as can be seen from the behavior of the limiting ESD. In the case that d is fixed and $n \to \infty$, the ESD of the adjacency matrix converges to the *Kesten-McKay* law, given by density

$$f_d(x) = \frac{2d\sqrt{4(2d-1) - x^2}}{2\pi(4d^2 - x^2)}, \qquad -2\sqrt{2d-1} \leq x \leq 2\sqrt{2d-1}.$$

On the other hand, the ESD of $\mathcal{P}_{n,d}$ is seen to converge to the semicircle law if $d = d(n) \to \infty$ with n (See [89, 98, 117]).

To state the results, begin by defining the coefficients

$$a(d, 2k) = (2d-1)^{2k} - 1 + 2d,$$
$$a(d, 2k+1) = (2d-1)^{2k+1} + 1,$$

whose combinatorial significance will be explained later. In terms of these, define $(C_k^{(\infty)};\ k \geq 1)$ to be independent Poisson random variables, with $C_k^{(\infty)}$ having mean $a(d, k)/2k$.

To state the non-polynomial fluctuation results, we need to adequately handle the deterministic largest eigenvalue of A_n which is equal to $2d$. Additionally, we need to be able to evaluate the expectations of these linear statistics in order to center them. As this is a discrete ensemble, the 1-point function is atomic, and thus we must have that the test functions are well-defined and bounded on $[\frac{-d}{\sqrt{2d-1}}, \frac{d}{\sqrt{2d-1}}]$, the support of the spectrum of $\mathcal{P}_{n,d}$. If we also have that $d \to \infty$, then we assume the functions are well defined on all $\mathbb{R}$ and locally bounded (in fact we only are able to prove theorems about subclasses of entire functions, so this is no loss), so that their expectations with respect to $\mathcal{P}_{n,d}$ are well defined.

Theorem 5. *Suppose that $d \geq 2$ is fixed and that $n \to \infty$. There is an open neighborhood U_d of $[-1, 1]$ in the complex plane so that if $f_1, f_2, \dots f_k$ are analytic in U_d, then*

$$\left(X_{f_1}^{\mathcal{P}_{n,d}}, X_{f_2}^{\mathcal{P}_{n,d}}, \dots, X_{f_k}^{\mathcal{P}_{n,d}}\right) \Rightarrow (Y_1, Y_2, \dots Y_k),$$

where

$$Y_i = \sum_{j=1}^{\infty} j \left[\sum_{r=1}^{\infty} \hat{f}_i(jr)(2d-1)^{-jr/2}\right] \left[C_j^{(\infty)} - \mathbb{E}C_j^{(\infty)}\right].$$

See Definitions 5 and 6 for a precise definition of U_d.

These limiting variables can now easily be verified to converge to the correct normal variable as $d \to \infty$ in the case of polynomial f. Recalling that for $Z \sim \text{Poisson}(\lambda)$, $\frac{Z-\lambda}{\sqrt{\lambda}} \Rightarrow N(0,1)$, as $\lambda \to \infty$, we get that

$$\frac{C_j^{(\infty)} - \mathbb{E}C_j^{(\infty)}}{(2d-1)^{j/2}} \Rightarrow N\left(0, \frac{1}{2j}\right),$$

so that $Y_i \Rightarrow N(0, V_{f_i})$, where V_{f_i} the variance appearing in the GOE CLT.

We show that in fact, we may take $d = d(n) \to \infty$ with n and produce the desired limit. If the rate at which $d = d(n) \to \infty$ is not important, then we can establish the theorem for any entire functions.

Theorem 6. *Suppose that $f_1, f_2, \dots f_k$ are entire functions. There is a sequence of $d = d(n) \to \infty$ with n sufficiently slowly so that*

$$\left(X_{f_1}^{\mathcal{P}_{n,d}}, X_{f_2}^{\mathcal{P}_{n,d}}, \dots, X_{f_k}^{\mathcal{P}_{n,d}}\right) \Rightarrow (Z_1, Z_2, \dots Z_k),$$

where $(Z_i)_i$ is jointly normal with covariance matching the GOE, i.e.

$$\text{Cov}(Z_i, Z_j) = \frac{1}{2} \sum_{n=1}^{\infty} n \hat{f}_i(n) \hat{f}_j(n).$$

On the other hand, if one would like d to grow to infinity at some prescribed rate, then we require some additional regularity over the test functions.

Theorem 7. *Suppose that $d = (\log n)^\gamma$ for some $\gamma > 0$. There is a class $\mathscr{O}_\gamma$ of entire functions so that if $f_1, f_2, \ldots f_k$ are in $\mathscr{O}_\gamma$ then*

$$\left(X_{f_1}^{\mathcal{P}_{n,d}}, X_{f_2}^{\mathcal{P}_{n,d}}, \ldots, X_{f_k}^{\mathcal{P}_{n,d}}\right) \Rightarrow (Z_1, Z_2, \ldots Z_k),$$

where $(Z_i)_i$ is jointly normal with covariance matching the GOE, i.e.

$$\mathrm{Cov}(Z_i, Z_j) = \frac{1}{2}\sum_{n=1}^{\infty} n\hat{f}_i(n)\hat{f}_j(n).$$

See Definition 8 for a precise definition of $\mathscr{O}_\gamma$.

3.2 Polynomial Fluctuations

We will begin by establishing these theorems in the case that the test functions are polynomials. Recall from Proposition 3 that the variance of a linear statistic could be expressed in terms of Chebyshev coefficients. The principal realization is that traces of Chebyshev polynomials of $\mathcal{P}_{n,d}$ can be expressed combinatorially. The key concept is that of a *non-backtracking walk.*

A non-backtracking walk is one that begins and ends at the same vertex, and that never follows an edge and immediately follows that same edge backwards. Let $\mathrm{NBW}_k^{(n)}$ denote the number of closed non-backtracking walks of length k on G_n.

If the last step of a closed non-backtracking walk is anything other than the reverse of the first step, we say that the walk is *cyclically non-backtracking.* Cyclically non-backtracking walks are easier to analyze than plain non-backtracking walks because every cyclic and inverted cyclic shift of a cyclically non-backtracking walk remains cyclically non-backtracking. Let $\mathrm{CNBW}_k^{(n)}$ denote the number of closed cyclically non-backtracking walks of length k on G_n.

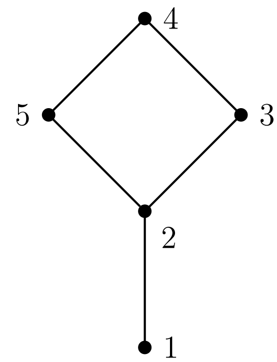


Figure 3.1: The walk $1 \to 2 \to 3 \to 4 \to 5 \to 2 \to 1$ is non-backtracking, but not cyclically non-backtracking. Note that such walks have a "lollipop" shape.

These notions sometimes go by different names. In [50], non-backtracking walks are called irreducible, and $\mathrm{NBW}_k^{(n)}$ is called $\mathrm{IrredTr}_k(G)$. Cyclically non-backtracking walks are called strongly irreducible, and $\mathrm{CNBW}_k^{(n)}$ is called $\mathrm{SIT}_k(G)$.

The main connection between these cyclically non-backtracking walks and the matrix is the following explicit connection. Recall that $\{T_n(x)\}_{n\in\mathbb{N}}$ are the Chebyshev polynomials of the first kind on the interval $[-1, 1]$. We define a set of polynomials

$$\Gamma_0(x) = 1 \ , \tag{3.1}$$

$$\Gamma_{2k}(x) = T_{2k}\left(x\right) + \frac{2d-2}{2(2d-1)^k} \ , \quad \forall\, k \geq 1 \ , \tag{3.2}$$

$$\Gamma_{2k+1}(x) = T_{2k+1}\left(x\right) \ , \quad \forall\, k \geq 0 \ . \tag{3.3}$$

These are modified so that we may exactly express a trace in terms of the non-backtracking walks.

Proposition 1. *Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of $\mathcal{P}_{n,d}$. Then*

$$\sum_{i=1}^{n} \Gamma_k(\lambda_i) = \frac{\mathrm{CNBW}_k^{(n)}}{2(2d-1)^{k/2}}.$$

Proof. See Proposition 32 of [35]. □

3.2.1 *Non-backtracking walks and cycles*

For any cycle in G_n of length $j|k$, we obtain $2j$ non-backtracking walks of length k by choosing a starting point and direction and then walking around the cycle repeatedly. However, these are not *all* possible cyclically non-backtracking walks; we refer the remainder as "bad" walks, and we denote these as $B_k^{(n)}$. These now allow us to write

$$\mathrm{CNBW}_k^{(n)} = \sum_{j|k} 2jC_j^{(n)} + B_k^{(n)}. \tag{3.4}$$

A non-backtracking walk is bad if and only if the *trail* of the walk has more than one cycle.[1] It turns out that these are relatively rare.

Proposition 2. *There is an absolute constant C such that for all n, $r \leq n^{1/4}$, and $d \geq 2$,*

$$\mathbb{E}[B_k^{(n)}] \leq \frac{Cr^4(2d-1)^r}{n}.$$

Proof. See Proposition 15 and the proof of Corollary 16 of [35]. □

In view of (3.4) and this Proposition, it follows that to understand the distribution of non-backtracking walks, it suffices to understand the limiting distribution of cycles. The number of cycles of short length, meanwhile, are known to be asymptotically Poisson in many graph models. See [17] or [127] for an account of older results, or [90] for the best result in this direction

By using Stein's method, it is possible to show a joint Poisson approximation with an error bound. In preparation, suppose that $w = w_1 \cdots w_k$ is a word on the letters $\pi_1, \ldots, \pi_d$ and $\pi_1^{-1}, \ldots, \pi_d^{-1}$. We call w *cyclically reduced* if $w_1 \neq w_k^{-1}$ and $w_i \neq w_{i+1}^{-1}$ for $1 \leq i < k$. Let $a(d, k)$ denote the number of cyclically reduced words of length k on this alphabet, which we will count explicitly.

[1] The trail of a walk is the simple graph whose vertices are those of the graph that are visited by the walk and whose edges are present if and only if the walk traversed that edge.

Lemma 18.

$$a(d, 2k) = (2d-1)^{2k} - 1 + 2d,$$
$$a(d, 2k+1) = (2d-1)^{2k+1} + 1.$$

Proof. This is a quick exercise in inclusion-exclusion. The proof requires some notation, but this should not obscure the simplicity of the ideas. Define

$$\Pi_k = \left\{\pi_1, \pi_2, \ldots, \pi_d, \pi_1^{-1}, \pi_2^{-2} \ldots, \pi_d^{-1}\right\}^k$$

to be all words of length k in these letters. Let $G = \mathbb{Z}/k\mathbb{Z}$ denote the cyclic group of order k, and for any subset $S \subseteq G$, define

$$V_S = \left\{w = w_0 w_1 \cdots w_{k-1} \in \Pi_k \;\middle|\; w_s = w_{s+1}^{-1} \; s \in S\right\},$$

where the addition is performed in G. The essential observation is that

$$|V_S| = \begin{cases} (2d)^{k-|S|} & k > |S| \\ 2d & k = |S|, k \text{ even} \\ 0 & k = |S|, k \text{ odd.} \end{cases}$$

To see the formula for $k > |S|$, note that each w_i with $i \neq S$ can be chosen freely from the alphabet. Moreover, once these are chosen, the word can be completed uniquely by the rules of V_S. The $k = |S|$ formula follow as in these cases, the word must be a single letter that alternates with its inverse, and this is only possible if the length of the word is even.

Having established these formulae, we can compute $a(d, k)$ by inclusion-exclusion,

$$a(d,k) = \sum_{S \subseteq G} (-1)^{|S|} |V_S| = \sum_{l=0}^{|S|-1} \binom{k}{l} (-1)^l (2d)^{k-l} + \begin{cases} 2d & k \text{ even} \\ 0 & k \text{ odd.} \end{cases}$$

Noting that this is nearly the binomial formula, the desired expressions follow. $\square$

The essential Poisson approximation can now be attacked using Stein's method. Recall that $(C_k^{(\infty)};\ k \geq 1)$ are independent Poisson random variables, with $C_k^{(\infty)}$ having mean $a(d,k)/2k$. Let $d_{TV}(X,Y)$ denote the total variation distance between the laws of X and Y.

Proposition 3. *There is a constant C such that for all n, k, and $d \geq 2$,*

$$d_{TV}\left(\left(C_1^{(n)}, \dots, C_r^{(n)}\right),\ \left(C_1^{(\infty)}, \dots, C_r^{(\infty)}\right) \right) \leq \frac{C(2d-1)^{2r}}{n}.$$

Proof. See Theorem 11 of [35]. □

Define

$$\mathrm{CNBW}_k^{(\infty)} = \sum_{j|k} 2jC_j^{(\infty)}.$$

It is now an immediate corollary of Proposition 3 and Proposition 2 that $\mathrm{CNBW}_k^{(n)}$ is approximately $\mathrm{CNBW}_k^{(\infty)}$.

Proposition 4. *There is a constant C such that for all n, r, and $d \geq 2$,*

$$d_{TV}\left(\left(\mathrm{CNBW}_k^{(n)};\ 1 \leq k \leq r\right), \left(\mathrm{CNBW}_k^{(\infty)};\ 1 \leq k \leq r\right) \right) \leq \frac{C(2d-1)^{2r}}{n}.$$

This allows us to formulate a polynomial fluctuation theorem for d fixed, $n \to \infty$.

Proposition 13. *Let $f_1, f_2, \dots f_k$ be polynomials. Suppose that d is fixed and that $n \to \infty$, then*

$$\left(X_{f_1}^{\mathcal{P}_{n,d}}, X_{f_2}^{\mathcal{P}_{n,d}}, \dots, X_{f_k}^{\mathcal{P}_{n,d}} \right) \Rightarrow \left(Y_1, Y_2, \dots Y_k\right),$$

where

$$Y_i = \sum_{j=1}^{\infty} j \left[\sum_{r=1}^{\infty} \hat{f}_i(jr)(2d-1)^{-jr/2} \right] \left[C_j^{(\infty)} - \mathbb{E}C_j^{(\infty)} \right].$$

Proof. We simply combine Proposition 1 and Proposition 4. □

Likewise, we are able to formulate a polynomial fluctuation theorem for slowly growing d and polynomial test functions.

Proposition 14. *Let $f_1, f_2, \dots f_k$ be polynomials of degree at most l. If $d = d(n) \to \infty$ as $n \to \infty$, but slowly enough that $\limsup_{n\to\infty} \frac{\log d(n)}{\log n} < \frac{1}{2l}$, then*

$$\left(X_{f_1}^{\mathcal{P}_{n,d}}, X_{f_2}^{\mathcal{P}_{n,d}}, \dots, X_{f_k}^{\mathcal{P}_{n,d}}\right) \Rightarrow (Z_1, Z_2, \dots Z_k),$$

where $(Z_i)_i$ is jointly normal with covariance matching the GOE, i.e.

$$\mathrm{Cov}(Z_i, Z_j) = \frac{1}{2}\sum_{n=1}^{\infty} n\hat{f}_i(n)\hat{f}_j(n).$$

To prove this, we require a small technical lemma on coupling of Poisson variables to normals.

Lemma 19. *Let $X \sim \mathrm{Poisson}(\lambda)$, and let $W = \frac{X-\lambda}{\sqrt{\lambda}}$. There is a coupling of X to a standard normal variable Z so that $\mathbb{E}|W - Z| \leq \frac{1}{\sqrt{\lambda}}$.*

Proof. As X is Poisson, its size bias transform X^s has the law of $1 + X$. Hence, we may trivially couple X^s to X. The desired statement now follows immediately from Theorem 3.13 of [102]. □

Proof of Proposition 14. Let $r = \max_{1\leq i\leq k} \deg(f_i)$ for the remainder of the proof, so that

$$2r \log d(n) - \log n \to -\infty.$$

This implies that the error in Proposition 4 goes to 0. For each n, by Proposition 3 and Lemma 19, we can construct a probability space which supports the permutation model $G(n, d(n))$, the idealized cycle count vector $(C_k^{(\infty)};\ k \geq 1)$, and a vector $(Z_i;\ i \geq 1)$ of independent standard normals; on this space, we have

$$\mathbb{P}\left[\left(C_1^{(n)}, \dots, C_r^{(n)}\right) \neq \left(C_1^{(\infty)}, \dots, C_r^{(\infty)}\right)\right] \leq \frac{C(2d-1)^{2r}}{n},$$

as well as $\left|C_j^{(\infty)} - \mathbb{E}C_j^{(\infty)} - \sqrt{\mathbb{E}C_j^{(\infty)}}Z_j\right| \leq 1$ for all $j \geq 1$. Note that by Proposition 2

$$\mathbb{P}\left[\left(\mathrm{CNBW}_k^{(n)};\ 1 \leq k \leq r\right) \neq \left(\mathrm{CNBW}_k^{(\infty)};\ 1 \leq k \leq r\right)\right] \leq \frac{C(2d-1)^{2r}}{n},$$

for some other constant C. Let Y_i be defined as

$$Y_i = \sum_{j=1}^{\infty} j \left[\sum_{r=1}^{\infty} \hat{f}_i(jr)(2d-1)^{-jr/2} \right] \left[C_j^{(\infty)} - \mathbb{E} C_j^{(\infty)} \right],$$

and note that we have that

$$\mathbb{P}\left[\left(X_{f_1}^{\mathcal{P}_{n,d}}, X_{f_2}^{\mathcal{P}_{n,d}}, \dots, X_{f_k}^{\mathcal{P}_{n,d}} \right) \neq (Y_1, Y_2, \dots Y_k) \right] \\ \leq \mathbb{P}\left[\left(\mathrm{CNBW}_k^{(n)};\ 1 \leq k \leq r \right) \neq \left(\mathrm{CNBW}_k^{(\infty)};\ 1 \leq k \leq r \right) \right] \leq \frac{C(2d-1)^{2r}}{n}.$$

Furthermore, define $W_i = \sum_{j=1}^{\infty} \sqrt{\frac{j}{2}} \hat{f}_i(j) Z_j$. Note that we have $\mathbb{E}|W_i - Y_i| \to 0$ as $d \to \infty$, as all the sums are finite, and hence we have established in-probability convergence of the linear statistics to the vector $(W_1, \dots, W_k)$, as desired. □

3.3 Extension to non-polynomial test functions

3.3.1 Approach

In contrast to polynomial test functions, applying non-polynomial test functions requires some amount of control over the regularity of the spectrum. The basic approach will be the following approximation scheme. For f continuous on $[-1,1]$, define $f_r = \sum_{i=0}^{r} \hat{f}(i) T_i(x)$. For the classes of test functions considered, we have that

$$\sup_{x \in [-1,1]} |f_r(x) - f(x)| \to 0,$$

very rapidly in r. In fact, more is true. As we look at classes of analytic functions, their behavior inside the interval $[-1,1]$ determines their behavior in any open interval containing $[-1,1]$, and so we can actually establish estimates for the rate at which

$$\sup_{x \in [-R,R]} |f_r(x) - f(x)| \to 0$$

for some $R > 0$. Now, we would like to approximate $X_f^{\mathcal{P}_{n,d}}$ by $X_{f_r}^{\mathcal{P}_{n,d}}$ by setting $r = r(n)$ to grow at some rate so that they converge together and so that $X_{f_r}^{\mathcal{P}_{n,d}}$ converges to what it should. To do this approximation, it would suffice to be able to control

$$\xi_n := \sum_{i=2}^{n} f(\lambda_i) - f_r(\lambda_i),$$

noting it would suffice to show that $\mathbb{E}|\xi_n| \to 0$ to establish the desired co-convergence.

There are two types of control that are desirable here, control over the support of the spectrum and control over spacings of the eigenvalues. Control over the support of the spectrum is the more important of these two, as it allows f and f_r to be compared on an interval of constant order instead of order $O(\sqrt{d})$. Ideally, the spectrum could be localized to $|x| \leq 1+\epsilon_n$ where $\epsilon_n \to 0$ with n, but this is out of reach. In the realm of d fixed, we may appeal to Friedman's work on the eigenvalues of the permutation model.

Theorem 8. *For $d \geq 2$ fixed there is a $\delta \geq 0$ depending on d so that for all $\epsilon > 0$ there is a constant $C = C(\epsilon, d)$ so that*

$$\mathbb{P}\left[\exists i \neq 1 \ : \ |\lambda_i| \geq 1+\epsilon\right] \leq Cn^{-1-\delta}.$$

For $d \geq 6$, we may take $\delta = 1$.

Proof. See [50]. □

This allows us to get immediate control over ξ_n of the form

$$\xi_n \leq n \sup_{|x| \leq 1+\epsilon} \left| f_r(x) - f(x) \right|, \tag{3.5}$$

except for on an event of vanishing probability. If we additionally show that

$$\mathbb{E} \sum_{i=2}^{n} \left[f(\lambda_i) - f_r(\lambda_i) \right] \to 0,$$

then we will be able to conclude that the two linear statistics converge together.

3.3.2 The possible benefits of additional rigidity

The previous analysis ignores that the eigenvalues of $\mathcal{P}_{n,d}$ are expected to be concentrated around their limiting positions. In the prior analysis, the error term ξ_n was simply bounded as (3.5). For this to tend to 0, we require that this supremum decay faster than n^{-1}. As we are only able to take $r(n) \to \infty$ at some logarithmic rate, we require that $\sup_{|x|\leq 1+\epsilon} |f_r(x) - f(x)|$ decay exponentially in r.

If all the eigenvalues clustered around where this supremum is attained, this would be the best bound on ξ_n for which one could hope. However, this is very much *not* expected. Quite to the contrary, the eigenvalues are expected to be very regularly distributed according to the limiting level density, the semicircle law. Suppose for a moment that the spectra were localized around the Chebyshev nodes, $\sigma_{i,n} = \cos\left(\frac{(2i-1)\pi}{n}\right)$.[2] Then we can observe the following lemma,

Lemma 20. *For f sufficiently smooth ($f \in C^2[-1,1]$ suffices),*

$$\sum_{i=1}^{n} f(\sigma_{i,n}) = \sum_{r=0}^{\infty} (-1)^r n \hat{f}(rn).$$

Proof. The proof follows immediately from the identity

$$\sum_{i=1}^{n} T_k(\sigma_{i,n}) = \Re\left[\sum_{i=1}^{n} \exp(i\pi \frac{k}{n}(2i-1))\right],$$

which can be seen to be 0 except for when k/n is an integer. □

In particular, as $f - f_r$ has many of its Chebyshev coefficients zeroed out, we would have $\xi_n \to 0$ for f in some class such as $C^2[-1,1]$ instead of f analytic. As the asymptotic distribution of the eigenvalues should look like the Chebyshev nodes, one would hope to be able to leverage this extra cancellation that should appear in ξ_n. However, at this time, not much is known in terms of the *rigidity* of the eigenvalues, that is to say, how nearly they are localized around the Chebyshev nodes. This rigidity has been observed in both log-gases and Wigner matrices.

---

[2]These nodes have density which is asymptotically arcsine distributed, not semicircularly distributed. Still, for the point of intuition, this is a helpful exercise.

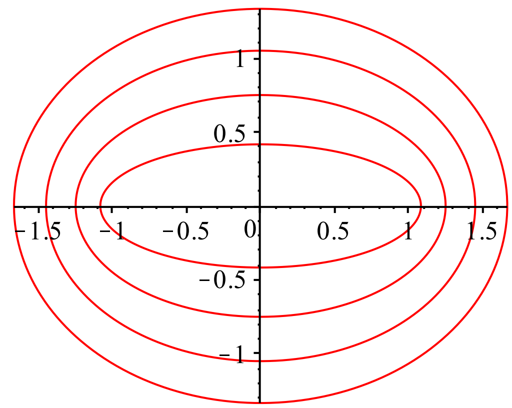


Figure 3.2: Bernstein ellipses $\mathcal{E}_B(\rho)$ with $\rho = 1.5, 2, 2.5, 3$.

3.3.3 Fixed d extension

We now turn to proving Theorem 5 by the procedure outlined, and we begin by defining the Bernstein ellipse $\mathcal{E}_B(\rho)$ of radius ρ.

Definition 5. Let $\rho > 1$, and let $\mathcal{E}_B(\rho)$ be the image of the annulus $A_\rho = \{z : 1/\rho < |z| < \rho\}$ under the map $\Theta(z) = \frac{z+z^{-1}}{2}$. We call $\mathcal{E}_B(\rho)$ the Bernstein ellipse of radius ρ. The ellipse has foci at ± 1, and the sum of the major semiaxis and the minor semiaxis is exactly ρ.

Bernstein ellipses are useful for understanding the Chebyshev approximation of analytic functions. The heart of this connection is the following formula for the Chebyshev coefficients

$$\hat{f}(k) = \frac{2}{2\pi i} \oint_{|z|=1} f(\Theta(z)) z^{k-1}\, dz,$$

which is easily verified by parameterization. This gives the Chebyshev coefficients the interpretation of being the Laurent coefficients of $f \circ \Theta$. As $f \circ \Theta$ is invariant under $z \mapsto 1/z$, the positive and negative parts of the series are identical. Now, it is easily established that a geometric rate of decay of the Chebyshev coefficients is equivalent to f being analytic in some appropriately large Bernstein ellipse. Precisely, $\limsup_{k\to\infty} |\hat{f}(k)|^{1/k} = \frac{1}{\rho}$ if and only if f is analytic in a Bernstein ellipse of radius ρ.

We can also bound the Chebyshev coefficients explicitly in terms of the growth rate of the analytic function. Define

$$M_f(\rho) := \sup_{z\in\mathcal{E}_B(\rho)} |f(\Theta(z))|.$$

From the Cauchy estimates, we have that $|\hat{f}(k)| \leq \frac{2M_f(\rho)}{\rho^k}$. Further, it is easily verified that for $R \geq 1$,

$$\sup_{|x|\leq R} |T_k(x)| \leq R + \sqrt{R^2-1}, \tag{3.6}$$

as follows from the identity

$$T_k(x) = \frac{\left(x-\sqrt{x^2-1}\right)^k + \left(x+\sqrt{x^2-1}\right)^k}{2}.$$

Hence, we have the following conclusion more suited to our purposes.

Lemma 21. *If f is analytic in a Bernstein ellipse $\mathcal{E}_B(\rho)$, for any $R \geq 1$ and any $\epsilon > 0$,*

$$\sup_{|x|\leq R} |f(x) - f_r(x)| = O\left(\left(\frac{R+\sqrt{R^2-1}}{\rho} - \epsilon\right)^r\right).$$

We can now state precisely where f need be analytic for the fixed d central limit theorem.

Definition 6. For $d \geq 6$, let U_d be an open set in the complex plane containing the closed Bernstein ellipse of radius $(2d-1)^2$. For $2 \leq d \leq 5$, let U_d be an open neighborhood of the closed Bernstein ellipse of radius $(2d-1)^{5/2}$.

In particular, for f analytic in U_d, we have that for some $\epsilon > 0$ sufficiently small, there is a $\delta > 0$ so that

$$\sup_{|x|\leq 1+\epsilon} |f(x) - f_r(x)| = O\left(((2d-1)^2+\delta)^{-r}\right).$$

For $d \geq 6$ Friedman's bound holds with high enough probability that we choose a sequence $r(n) \to \infty$ so that $n^{-1}(2d-1)^{2r(n)} \to 0$ and $n\left(((2d-1)^2+\delta)^{-r(n)}\right) \to 0$. When $d \leq 5$, we choose $r(n) \to \infty$ so that $n^{-1}(2d-1)^{2r(n)} \to 0$ and $n \sup_{|x| \leq \frac{d}{\sqrt{2d-1}}} |f(x) - f_r(x)| \to 0$, thus removing the need to know bounds for the support of the eigenvalues. Note that in this case, we have that $|\xi_n| \to 0$ considered as a deterministic function of a sequence of $2d$-regular graphs with $n \to \infty$.

We are now set up to prove Theorem 5 with very little difficulty.

Proof Theorem 5. We prove only the case $d \geq 6$ as the other is easier. Suppose f is analytic in U_d. From how $r(n)$ is chosen, it is immediate to check that $X_{f_r}^{\mathcal{P}_{n,d}} \Rightarrow Y$, where Y is

$$Y = \sum_{j=1}^{\infty} j \left[\sum_{r=1}^{\infty} \hat{f}(jr)(2d-1)^{-jr/2}\right] \left[C_j^{(\infty)} - \mathbb{E}C_j^{(\infty)}\right].$$

Likewise, it holds that any finite dimensional vector converges as well. To conclude this we need to show that $X_f^{\mathcal{P}_{n,d}} - X_{f_r}^{\mathcal{P}_{n,d}} \Rightarrow 0$ in probability. We establish the theorem by showing that $\mathbb{E}|\xi_n| \to 0$, proceeding by a standard truncation approach. With $\epsilon > 0$ and $r(n)$ as chosen, we have that

$$\mathbb{E}\sum_{i=2}^{n} |f(\lambda_i) - f_r(\lambda_i)| \mathbf{1}\{|\lambda_i| < 1+\epsilon\} \leq n \sup_{|x| \leq 1+\epsilon} |f(x) - f_r(x)| \to 0.$$

It remains to show that the contribution of the remainder of the eigenvalues is small. On the one hand, by the assumed boundedness of the test function on the support,

$$\mathbb{E}\sum_{i=2}^{n} |f(\lambda_i)| \mathbf{1}\{|\lambda_i| \geq 1+\epsilon\} \leq n \sup_{|x| \leq \frac{d}{\sqrt{2d-1}}} |f(x)| \mathbb{P}\left[\exists i \neq 1 \ : \ |\lambda_i| \geq 1+\epsilon\right].$$

From Theorem 8, this probability is $o(1/n)$, and so this contribution vanishes. Noting the bound (3.6) and that the Chebyshev coefficients for all the f must decay like $|\hat{f}(k)| = O((2d-1)^{-2+\epsilon})$, we have that $\sum_{i=1}^{n} |f_r(\lambda_i)| \leq Cn$, for some universal constant C and all $d \geq 6$. The proof now follows from Friedman's eigenvalue bound, Theorem 8,

$$\mathbb{E}\sum_{i=2}^{n} |f_r(\lambda_i)| \mathbf{1}\{|\lambda_i| \geq 1+\epsilon\} \leq Cn\mathbb{P}\left[\exists i \neq 1 \ : \ |\lambda_i| \geq 1+\epsilon\right] \to 0.$$

$\square$

3.3.4 Growing d extension

The case of growing d uses essentially the same machinery as was developed for the fixed d case. As there, we need to show that $X_f^{\mathcal{P}_{n,d}} - X_{f_r}^{\mathcal{P}_{n,d}} \to 0$ in probability. The fact that $X_{f_r}^{\mathcal{P}_{n,d}}$ converges to the normal to which it should follows from an identical argument as in Proposition 14 provided $r(n)$ and $d(n)$ satisfy

$$2r(n)\log d(n) - \log n \to -\infty. \tag{3.7}$$

The principal missing ingredient is Friedman's eigenvalue bound, which holds only for d fixed. For this reason, we prove an alternate eigenvalue concentration bound.

Theorem 7. *For any $m > 0$, there is a constant $C = C(m)$ and universal constants K and c so that*

$$\mathbb{P}\left[\exists i \neq 1 \ : \ |\lambda_i| \geq C\right] \leq n^{-m} + K\exp(-cn).$$

Further, the constant C may be taken to be $36000 + 2400m$.

We will prove and discuss this bound in Section 3.4. For now, we will turn to proving Theorems 6 and 7, which have very similar proofs. We need to show that for a given entire function, we can choose $r(n) \to \infty$ sufficiently quickly that $\mathbb{E}|\xi_n| \to 0$ but sufficiently slowly that (3.7) is still possible. Choose K sufficiently large that

$$\mathbb{P}\left[\exists i \neq 1 \ : \ |\lambda_i| \geq K\right] = O(n^{-2}).$$

Following the truncation approach used in Theorem 5, we have that

$$E_1 := \mathbb{E}\sum_{i=2}^{n} |f(\lambda_i) - f_r(\lambda_i)|\mathbf{1}\left\{|\lambda_i| < K\right\} \leq \frac{2nM_f(\rho)}{1 - \frac{2K}{\rho}}\left(\frac{2K}{\rho}\right)^r \tag{3.8}$$

for any $\rho > 2K$. We also have

$$E_2 := \mathbb{E}\sum_{i=2}^{n} |f(\lambda_i)|\mathbf{1}\left\{|\lambda_i| \geq K\right\} \leq n^{-2}M_f\left(\frac{d}{\sqrt{2d-1}}\right), \tag{3.9}$$

and

$$E_3 := \mathbb{E}\sum_{i=1}^{n} |f_r(\lambda_i)| \mathbf{1}\left\{|\lambda_i| \geq K\right\} = O(n^{-2}(2K)^r). \tag{3.10}$$

The goal is to be able to choose all these errors to go to 0. If all of them do, we have that $\mathbb{E}|\xi_n| \to 0$ implying the needed convergence.

Proof of Theorem 6. Here we are given arbitrary entire functions $f_1, \ldots, f_k$. We need only show the approximation for a single one of these functions as the desired vector convergence is identical. Fix $\delta > 0$ arbitrary. For a given entire function f, there is some rate $d(n) \to \infty$ sufficiently slow that $M_f(d^{2+3\delta}(n)) \leq \log n$. We also assume that $d(n) = o(\log n)$, as must be necessary if f is transcendental.

Define $r(n)$ to be a sequence of integers so $r(n) \sim \frac{1}{2(1+\delta)} \frac{\log n}{\log d(n)}$, noting that this ensures that (3.7) is satisfied. This choice also ensures that $E_2 \to 0$ and $E_3 \to 0$. To bound E_1, we take $\rho = d^{2+3\delta}$ to get

$$E_1 = O\left(n \log n \left(\frac{2K}{d^{2+3\delta}}\right)^r\right).$$

Note that $d^{(2+3\delta)r} = n^{1+\epsilon}$ for some small $\epsilon > 0$, so that $E_1 \to 0$ as well. □

Following this proof, we can determine more specifically how $M_f(\rho)$ must behave to produce the desired limits.

Definition 8. Define $\mathscr{O}_\gamma$ to be the class of entire functions f for which there exist $C > 0$ and $\epsilon > 0$ so that

$$M_f(x) \leq C \exp\left(x^{\frac{1}{2\gamma}-\epsilon}\right)$$

for all $x > 0$.

This is in fact identical to specifying the *order* of the entire function, a classical subject. This is defined as

$$\rho(f) = \limsup_{r \to \infty} \frac{\log(\log \sup_{|z| \leq r} |f(z)|)}{\log r}.$$

Lemma 22. *An entire function $f \in \mathscr{O}_\gamma$ if and only if it has order less than $\frac{1}{2\gamma}$.*

Proof. This follows immediately from the observation that for large r, the Bernstein ellipse $\mathcal{E}_B(r)$ is nearly a disc of radius $r/2$. All that is necessary for the proof is to notice that there is an r_0 sufficiently large so that

$$\mathbb{D}(0, r/4) \subseteq \mathcal{E}_B(r) \subseteq \mathbb{D}(0, r),$$

for all $r \geq r_0$. $\square$

Remark 4. The effect of having the eigenvalue bound Theorem 7 is to adjust the dependence of the exponent $x^{\frac{1}{2\gamma}-\epsilon}$. Without the eigenvalue bound, we would need the order to be at most $\frac{2}{5\gamma} - \epsilon$.

Proof of Theorem 7. In this case we take $d = (\log n)^\gamma$. For entire functions in $\mathscr{O}_\gamma$, we therefore have that $M_f(d^{2+3\delta}) = O(n^{o(1)})$ for some $\delta > 0$. Take $r(n) \sim \frac{1}{2(1+\delta)} \frac{\log n}{\log d(n)}$, so that we have $E_2 \to 0$ and $E_3 \to 0$. Take $\rho = d^{2+3\delta}$ in (3.8) to conclude that $d^{(2+3\delta)r} = n^{1+\epsilon}$ for some small $\epsilon > 0$, so that $E_1 \to 0$ as well $\square$

3.4 Spectral concentration

The problem of estimating the magnitude of all but the first eigenvalue of a d-regular graph has been approached primarily in two ways, the method of moments and the counting method of Kahn and Szemerédi, presented in [49]. The method of moments has been developed in the work of Broder and Shamir [21] and very extensively by Friedman [48], [50]. In his work, Friedman, relying on d being fixed independently of n, developed extremely fine control over the magnitude of the second eigenvalue. On the other hand in [49], Kahn and Szemerédi only show that the second largest eigenvalue has magnitude $O(\sqrt{d})$. While weaker than Friedman's bound, their techniques readily extend to the case where d is allowed to grow as a function of n; this observation has been informally made by others, and communicated to us by Vu and Friedman. Here we will formalize it, and present the Kahn-Szemerédi argument in the context of growing d to demonstrate the method's validity, as well as to develop some handle on the constants in the bound.

Both [22] and [81] provide examples of how the Kahn-Szemerédi argument can be used to control the second eigenvalue when d grows with n. In [22], the authors work in the configuration model to obtain the $O(\sqrt{d})$ bound for $d = O(\sqrt{n})$, essentially the largest d for which the configuration model represents the uniform d-regular graph well enough to prove eigenvalue concentration. In [81], the authors study the spectra of random covers. The permutation model is an example of such a cover, where the base graph is a single point with d self loops. Using the Kahn-Szemerédi machinery, they are able to show an $O(\sqrt{d}\log d)$ bound with $d(n) = \text{poly}(n)$ growth. The adaptations to the original Kahn-Szemerédi argument made in [81], especially the usage of Freedman's martingale inequality, are similar to the ones made here. However, as we do not need to consider the geometry of the base graph, we are able to push this argument to prove a non-asymptotic bound of the correct order.

Recall that the adjacency matrix A_n can be realized by sampling independently and uniformly d permutation matrices $P_1, P_2, \ldots, P_d$ and forming

$$A_n = P_1 + P_1^t + P_2 + P_2^t + \cdots + P_d + P_d^t.$$

We will drop the n and refer to the adjacency matrix simply as A. The starting point is the variational characterization of the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ of A, which states that

$$\max\{\lambda_2, |\lambda_n|\} = \sup_{\substack{w \perp \mathbf{1} \\ \|w\|=1}} \left| w^t A w \right|.$$

We will restate the theorem we wish to prove in terms of the eigenvalues of the adjacency matrix.

Theorem 9. *For any $m > 0$, there is a constant $C = C(m)$ and universal constants K and c so that*

$$\mathbb{P}\left[\exists i \neq 1 \;:\; |\lambda_i| \geq C\sqrt{d}\right] \leq n^{-m} + K\exp(-cn).$$

Further, the constant C may be taken to be $36000 + 2400m$.

Additional flexibility is provided by replacing the symmetric Rayleigh quotient by the asymmetric version,

$$\sup_{\substack{w,v\perp \mathbf{1}\\ \|v\|=\|w\|=1}} |v^t A w|.$$

The random variables v^tAw, for fixed w and v, are substantially more tractable than the supremum. To be able to work with these random variables instead of the supremum, we will pass to a finite set of vectors which approximate the sphere $\mathcal{S} = \{w \perp \mathbf{1} \ : \ \|w\| = 1\}$. More specifically, we will only consider those w and v lying on the subset of the lattice $\mathcal{T}$ defined as

$$\mathcal{T} := \left\{ \frac{\delta z}{\sqrt{n}} \ : \ z \in \mathbb{Z}^n, \|z\|^2 \le \frac{n}{\delta^2}, z \perp \mathbf{1} \right\},$$

for a fixed $\delta > 0$.

Vectors from $\mathcal{T}$ approximate vectors from $\mathcal{S}$ in the sense that every $v \in (1-\delta)\mathcal{S}$ is a convex combination of points in $\mathcal{T}$. (See Lemma 2.3 of [42].) Thus

$$\frac{1}{(1-\delta)^2} \sup_{\substack{w,v\perp \mathbf{1}\\ \|v\|=\|w\|=1}} \left|[1-\delta]v^t A[1-\delta]w\right| \le \frac{1}{(1-\delta)^2} \sup_{x,y\in\mathcal{T}} \left|x^t A y\right|.$$

Furthermore, by a volume argument, it is possible to bound the cardinality of $\mathcal{T}$ as

$$|\mathcal{T}| \left(\frac{\delta}{\sqrt{n}}\right)^n \le \mathrm{Vol}\left[x \in \mathbb{R}^n \ : \ \|x\| \le 1 + \tfrac{\delta}{2}\right] = \frac{(1+\frac{\delta}{2})^n \sqrt{\pi}^n}{\Gamma(\frac{n}{2}+1)}.$$

Employing Stirling's approximation, this shows $|\mathcal{T}| \le C\left[\frac{(1+\frac{\delta}{2})\sqrt{2e\pi}}{\delta}\right]^n$. for some universal constant C.

The breakthrough of Kahn and Szemerédi was to realize that x^tAy can be controlled by virtue of a split into two types of terms. If x^tAy is written as a sum

$$x^tAy = \sum_{\substack{(u,v)\\ |x_u y_v| < \frac{\sqrt{d}}{n}}} x_u A_{uv} y_v + \sum_{\substack{(u,v)\\ |x_u y_v| \ge \frac{\sqrt{d}}{n}}} x_u A_{uv} y_v,$$

then the contribution of the first sum turns out to be very nearly its mean because of the Lipschitz dependence of the sum on the edges of the graph. The contribution of the second sum turns out to never be too large for a very different reason: the number of edges between any two sets in the graph is on the same order as its mean. Following Feige and Ofek, for a fixed pair of vectors $(x, y) \in \mathcal{T}^2$, define the *light couples* $\mathcal{L} = \mathcal{L}(x, y)$ to be all those ordered pairs (u, v) so that $|x_u y_v| \leq \frac{\sqrt{d}}{n}$, and let the *heavy couples* $\mathcal{H}$ be all those pairs that are not light.

3.4.1 Controlling the contribution of the light couples.

Part of the advantage of having selected only the light couples is that their expected contribution is of the correct order, as the lemma below shows.

Lemma 23.

$$\left| \mathbb{E} \sum_{(u,v)\in\mathcal{L}} x_u A_{uv} y_v \right| \leq 2\sqrt{d}.$$

Proof. By symmetry, $\mathbb{E}A_{uv}$ is simply equal to $\frac{2d}{n}$, so that

$$\mathbb{E} \sum_{\{u,v\}\in\mathcal{L}} x_u A_{uv} y_v = \frac{2d}{n} \sum_{\{u,v\}\in\mathcal{L}} x_u y_v.$$

Because each of x_u and y_v sum to 0, the sum over light couples is equal in magnitude to the sum over heavy couples. Thus, it suffices to estimate

$$\begin{aligned} \left| \sum_{\{u,v\}\in\mathcal{H}} x_u y_v \right| &\leq \sum_{\{u,v\}\in\mathcal{H}} |x_u y_v| = \sum_{\{u,v\}\in\mathcal{H}} \frac{x_u^2 y_v^2}{|x_u y_v|} \\ &\leq \frac{n}{\sqrt{d}} \sum_{\{u,v\}\in\mathcal{H}} x_u^2 y_v^2, \quad \text{by the defining property of heavy couples,} \\ &\leq \frac{n}{\sqrt{d}}. \end{aligned}$$

In the last step we recall that both $\|x\|, \|y\| \leq 1$. □

To show that not only the expectation, but the sum itself is of the correct order, we must prove a concentration estimate for this sum. For technical reasons, it is helpful if we deal with sums over fewer terms. To this end, define

$$M = A_1 + A_2 + \cdots + A_d.$$

In terms of M it is enough to insist that for every $x, y \in \mathcal{T}$

$$\left| \sum_{(u,v) \in \mathcal{L}} x_u M_{uv} y_v \right| \leq t\sqrt{d}$$

for then by symmetry,

$$\left| \sum_{(u,v) \in \mathcal{L}} x_u A_{uv} y_v \right| \leq 2t\sqrt{d},$$

for all $x, y \in \mathcal{T}$. As a further simplification, we will not prove a tail estimate for the whole quantity $\sum\limits_{(u,v)\in\mathcal{L}} x_u M_{uv} y_v$; instead, fix an arbitrary collection U of vertices of size at most $\lceil \frac{n}{2} \rceil$. Having fixed this collection, we will show a tail estimate for $\sum_{(u,v)\in\mathcal{L}\cap U\times[n]} x_u M_{uv} y_v$. This truncation is made to simplify a variance estimate (see (3.12)), and it might be possible to avoid it entirely.

Theorem 10. *For every $x, y \in \mathcal{T}$, and every $U \subset [n]$ with $|U| \leq \lceil \frac{n}{2} \rceil$,*

$$\mathbb{P}\left[\left| \sum_{(u,v)\in\mathcal{L}\cap U\times[n]} x_u M_{uv} y_v - \mathbb{E} x_u M_{uv} y_v \right| > t\sqrt{d} \right] \leq C_0 \exp\left(-\frac{nt^2}{C_1 + C_2 t} \right)$$

for some universal constants C_0, C_1 and C_2. These constants can be taken as 2, 64, *and* 8/3 *respectively.*

Proof. Let $\tilde{\mathcal{L}}$ be $\mathcal{L} \cap U \times [n]$. We will estimate tail probabilities for $\sum\limits_{(u,v)\in\tilde{\mathcal{L}}} x_u M_{uv} y_v$.

The main tool needed to establish this result is Freedman's martingale inequality [46]. Let $X_1, X_2, \ldots$ be martingale increments. Write $\mathscr{F}_k$ for the natural filtration induced by these increments, and define $V_k = \mathbb{E}\left[X_k^2 \mid \mathscr{F}_{k-1}\right]$. If S_n is the partial sum $S_n = \sum_{i=1}^n X_i$ (with $S_0 = 0$) and T_n is the sum $T_n = \sum_{i=1}^n V_i$ (with $T_0 = 0$), then by analogy with the continuous case, one expects S_n to be a Brownian motion at time T_n (a discretization of the bracket process). The analogy requires, however, that the increments have some *a priori* bound. Namely, if $|X_k| \leq R$,

$$\mathbb{P}\left[\exists\, n \leq \tau \text{ so that } S_n \geq a \text{ and } T_n \geq b\right] \leq 2\exp\left(-\frac{a^2/2}{\frac{Ra}{3} + b} \right).$$

Remark 11. The constants quoted here are slightly better than the constants that appear in Freedman's original paper. This statement of the theorem follows from Proposition 2.1 of [46] and the calculus lemma

$$(1+u)\log(1+u) - u \geq \frac{u^2/2}{1+u/3},$$

for $u \geq 0$.

Reorder and relabel the vertices from U as $x_1, x_2, \ldots, x_r$, with $r \leq \lceil \frac{n}{2} \rceil$ so that $|x_j|$ decreases in j. Order pairs $(i,j) \in [d] \times \{0, 1, 2, \ldots r\}$ lexicographically, and enumerate $\pi_i(j)$ in this order as $f_1, f_2, \ldots, f_{rd}$. Define a filtration of σ-algebras $\{\mathscr{F}_k\}_{k=1}^{rd}$ by revealing these pieces of information, i.e. $\mathscr{F}_k = \mathscr{F}_{k-1} \vee \pi(f_k)$. According to this filtration, let

$$S_k = \mathbb{E}\left[\sum_{(u,v) \in \tilde{\mathcal{L}}} x_u M_{uv} y_v \middle| \mathscr{F}_k \right]$$

define a martingale and let $X_k = X_{(i,j)}$ be the associated martingale increments.

The desired deviation bound can now be cast in terms of S_k as

$$\begin{aligned} \mathbb{P}\left[\left| \sum_{\tilde{\mathcal{L}}} x_u M_{uv} y_v - \mathbb{E} x_u M_{uv} y_v \right| \geq t \right] &\leq \mathbb{P}\left[\exists\ k \leq rd \text{ so that } |S_k - S_0| = |S_k| \geq t \text{ and } T_n \geq b \right] \\ &\leq 2 \exp\left(\frac{-t^2/2}{(\frac{Rt}{3} + b)} \right), \end{aligned}$$

provided that b satisfies $\sum_{k=1}^{rd} \mathbb{E}\left[X_k^2 \;\middle|\; \mathscr{F}_{k-1} \right] \leq b$.

This reduces the problem to finding suitable R and b. The starting point for finding any such bound is simplifying the expression for the martingale increments $X_{(i,k)}$. To this end, let π be a fixed permutation of $[n]$, and define Π_k to be the collection of all permutations that agree with π in the first k entries, i.e.

$$\Pi_k = \{\sigma\ :\ \sigma(i) = \pi(i)\ i = 1, 2, \ldots, k\}.$$

Further let $T : \Pi_{k-1} \to \Pi_k$ be the map which maps a permutation to its nearest neighbor in Π_k, in the sense of transposition distance, i.e.

$$T[\sigma](i) = \begin{cases} \pi(k) & i = k \\ \sigma(k) & i = \sigma^{-1}(\pi(k)) \\ \sigma(i) & \text{else} \end{cases} .$$

Note that this map is the identity upon restriction to Π_k. Let $L_{[u,v]}$ be the characteristic function for $(u, v) \in \tilde{\mathcal{L}}$. In terms of these notation, it is possible to express $X_{(i,k)}$ as

$$X_{(i,k)} = \frac{1}{|\Pi_{k-1}|} \sum_{\tau \in \Pi_{k-1}} \sum_{u \in U} x_u L_{[u,T[\tau](u)]} y_{T[\tau](u)} - x_u L_{[u,\tau(u)]} y_{\tau(u)},$$

where $\pi = \sigma_i$, and the contributions of the other σ_j all cancel. As $\tau(u) = T[\tau](u)$ except for when $u = k$ or $u = \tau^{-1}(\pi(k))$, this simplifies to

$$\begin{aligned} X_{(i,k)} =& \frac{1}{|\Pi_{k-1}|} \sum_{\tau \in \Pi_{k-1}} \big(x_u L_{[u,\pi(k)]} y_{\pi(k)} - x_u L_{[u,\tau(k)]} y_{\tau(k)} \\ & + x_{\tau^{-1}(\pi(k))} L_{[\tau^{-1}(\pi(k)),\tau(k)]} y_{\tau(k)} - x_{\tau^{-1}(\pi(k))} L_{[\tau^{-1}(\pi(k)),\pi(k)]} y_{\pi(k)}\big) . \end{aligned}$$

This can be recast probabilistically. Define two random variables v and u as

$$v \sim \text{Unif}\,\{[n] \setminus \pi[k]\} \ ,$$
$$u \sim \text{Unif}\,\{[n] \setminus [k]\} \,,$$

(where $[n] = \{1, 2, \ldots, n\}$) so that

$$\begin{aligned} &\tfrac{n-k+1}{n-k} X_k \\ &\quad = \mathbb{E}\big[x_k L_{[k,v]} y_v - x_k L_{[k,\pi(k)]} y_{\pi(k)} + x_u L_{[u,\pi(k)]} y_{\pi(k)} - x_u L_{[u,v]} y_v \ \big| \mathscr{F}_k\big]. \quad (3.11) \end{aligned}$$

Terms for which $\pi(k) = \tau(k)$ again cancel, and so we have disregarded these terms from the right hand side. It is also for this reason that the small correction appears in front of X_k. From here it is possible to immediately deduce a sufficient *a priori* bound on X_k, as each term in this expectation is at most $\frac{\sqrt{d}}{n}$, so that $|X_k| \leq 4\frac{\sqrt{d}}{n}$.

The conditional variance $\mathbb{E}\left[X_k^2 \mid \mathscr{F}_{k-1}\right]$ is not much more complicated. Effectively, we take $\pi(k)$ to be uniformly distributed over $[n] \setminus \pi[k-1]$ and bound $\mathbb{E}\left[X_k^2 \mid \mathscr{F}_{k-1}\right]$ by

$$\begin{aligned}
&\mathbb{E}\left[X_k^2 \mid \mathscr{F}_{k-1}\right] \\
&\leq 4\mathbb{E}\left[x_k^2(L_{[k,v]}y_v)^2 + x_k^2(L_{[k,\pi(k)]}y_{\pi(k)})^2 + x_u^2(L_{[u,\pi(k)]}y_{\pi(k)})^2 + x_u^2(L_{[u,v]}y_v)^2 \mid \mathscr{F}_{k-1}\right].
\end{aligned}$$

As we have ordered the x_i, $x_u^2 \leq x_k^2$. Further, by bounding all the $L_{[a,b]}$ terms by 1, and using that v is marginally distributed as $\mathrm{Unif}\left\{[n] \setminus \pi[k-1]\right\}$, this bound becomes

$$\mathbb{E}\left[X_k^2 \mid \mathscr{F}_{k-1}\right] \leq 16\mathbb{E}\left[x_k^2 y_v^2 \mid \mathscr{F}_{k-1}\right].$$

Upon explicit calculation, we see that

$$\mathbb{E}\left[y_v^2 \mid \mathscr{F}_{k-1}\right] = \frac{1}{n-k}\sum_{[n]\setminus\pi[k-1]} y_v^2 \leq \frac{1}{n-k},$$

where it has been used that $\|y\| \leq 1$. Combining the above with (3.11), we see that

$$\mathbb{E}\left[X_k^2 \mid \mathscr{F}_{k-1}\right] \leq \left[\frac{n-k}{n-k+1}\right]^2 \frac{16x_k^2}{n-k} \leq \frac{32x_k^2}{n} \tag{3.12}$$

where it has been used that $k \leq r \leq \lceil \frac{n}{2} \rceil$. Summing over all martingale increments,

$$\sum_{i=1}^{d}\sum_{k=1}^{r} \frac{32x_k^2}{n} \leq \frac{32d}{n}.$$

Thus the Freedman martingale bound becomes

$$\mathbb{P}\left[\left|\sum_{\tilde{\mathcal{L}}} x_u M_{uv} y_v - \mathbb{E}x_u M_{uv} y_v\right| > t\sqrt{d}\right] \leq 2\exp\left(\frac{-nt^2}{64+8/3t}\right).$$

□

Let $\mathcal{L}_{\text{left}}$ be the set of vertices that appear in the first coordinate of some light couple, and choose $U \subseteq \mathcal{L}_{\text{left}}$ arbitrarily so that $|U| = \lceil |\mathcal{L}_{\text{left}}|/2 \rceil$. It follows then that, if $U_1 := U$, and $U_2 := \mathcal{L}_{\text{left}} \setminus U_1$,

$$\mathbb{P}\left[\left|\sum_{(u,v)\in\mathcal{L}} x_u A_{uv} y_v - \mathbb{E} x_u A_{uv} y_v\right| > t\sqrt{d}\right]$$
$$\leq 2\mathbb{P}\left[\max_{i=1,2}\left|\sum_{(u,v)\in\mathcal{L}\cap U_i\times[n]} x_u A_{uv} y_v - \mathbb{E} x_u A_{uv} y_v\right| > \frac{t}{2}\sqrt{d}\right].$$

From this point, it is possible to estimate

$$\mathbb{P}\left[\exists\, x, y \in \mathcal{T} \;:\; \left|\sum_{\mathcal{L}} x_u M_{uv} y_v\right| > 2(2t+1)\sqrt{d}\right]$$

by

$$\mathbb{P}\left[\exists\, x, y \in \mathcal{T} \;:\; \left|\sum_{\mathcal{L}\cap U\times[n]} x_u [A_{uv} - \mathbb{E} A_{uv}] y_v\right| > t\sqrt{d}\right]$$

Applying the union bound and Theorem 10, we see now that

$$\mathbb{P}\left[\exists\, x, y \in \mathcal{T} \;:\; \left|\sum_{\mathcal{L}} x_u M_{uv} y_v\right| > 2(2t+1)\sqrt{d}\right] \leq C\left[\frac{(2+\delta)\sqrt{2e\pi}}{2\delta}\right]^{2n} \exp\left(\frac{-nt^2}{64+8t/3}\right),$$

so that taking $e - 2 \geq \delta \geq \frac{1}{2}$ and $t = 27$, it is seen that this probability decays exponentially fast, and we have proven

Theorem 12. *There are universal constants C and K sufficiently large and $c > 0$ so that for $e - 2 \geq \delta \geq \frac{1}{2}$ and except for with probability at most*

$$K \exp(-cn),$$

there is no pair of vectors $x, y \in \mathcal{T}$ having

$$\left|\sum_{(u,v)\in\mathcal{L}} x_u M_{uv} y_v\right| \geq C\sqrt{2d}.$$

It is possible to take $C = 110$.

3.4.2 Controlling the contribution of the heavy couples.

Lemma 24 (Discrepancy)**.** *For any two vertex sets A and B, let $e(A,B)$ denote the number of directed edges from A to B that result as a form $\pi_i(a) = b$ for some $1 \leq i \leq d$, $a \in A$ and $b \in B$. Let $\mu(A,B) = |A||B|\frac{d}{n}$. For every $m > 0$, there are constants $c_1 \geq e$ and c_2 so that for every pair of vertex sets A and B, except with probability n^{-m}, exactly one of the following properties holds*

1. *either* $\frac{e(A,B)}{\mu(A,B)} \leq c_1$,
2. *or* $e(A,B)\log\frac{e(A,B)}{\mu(A,B)} \leq c_2(|A| \vee |B|)\log\frac{n}{|A|\vee|B|}$

It is possible to take $c_1 = e^4$ and $c_2 = 2e^2(6+m)$.

To prove this lemma, we rely on a standard type of large deviation inequality shown below, which mirrors the large deviation inequalities available for sums of i.i.d. indicators.

Lemma 25. *For any $k \geq e$,*

$$\mathbb{P}\left[e(A,B) \geq k\mu(A,B)\right] \leq \exp(-k[\log k - 2]\mu).$$

Proof. Let $e_\pi(A,B)$ denote the number $a \in A$ so that $\pi(a) \in B$. It is possible to bound

$$\mathbb{P}\left[e_\pi(A,B) = t\right] \leq \frac{[a]_t[b]_t}{t![n]_t},$$

where we recall that $[a]_t = a(a-1)\dots(a-t+1)$ is the falling factorial or Pochhammer symbol. Using the fact that $[n]_t \geq e^{-t}n^t$, this may be bounded as

$$\mathbb{P}\left[e_\pi(A,B) = t\right] \leq \frac{a^t b^t e^t}{t! n^t},$$

so that the Laplace transform of $e_\pi(A,B)$ can be estimated as

$$\mathbb{E}\left[\exp(\lambda e_\pi(A,B))\right] \leq \sum_{t=0}^{\infty} e^{\lambda t}\frac{a^t b^t e^t}{t! n^t} = \exp\left[\frac{abe^{1+\lambda}}{n}.\right]$$

Thus by Markov's inequality, we have

$$\begin{aligned}\mathbb{P}\left[e(A,B) \geq k\mu(A,B)\right] &\leq \frac{\mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^{d} e_{\sigma_i}(A,B)\right)\right]}{e^{-k\lambda\mu}} \\ &\leq \exp\left[\mu e^{1+\lambda} - k\lambda\mu\right],\end{aligned}$$

where $\lambda > 0$ is any positive number and $\mu = \mu(A,B)$. Taking $1 + \lambda = \log k$, valid for $k > e$, it follows that

$$\mathbb{P}\left[e(A,B) \geq k\mu(A,B)\right] \leq \exp\left[-k(\log k - 2)\mu\right],$$

for $k \geq e$. □

Armed with Lemma 25, we can proceed with the proof of Lemma 24.

Proof of Lemma 24. If either of $|A|$ or $|B|$ is greater than $\frac{n}{e}$, then $e(A,B) \leq (|A| \vee |B|)d$, so that

$$\frac{e(A,B)}{\mu(A,B)} \leq \frac{nd(|A| \vee |B|)}{|A||B|d} = \frac{n}{|A| \wedge |B|} \leq e.$$

Thus, it suffices to deal with the case that both A and B are less than $\frac{n}{e}$. In what follows, we will think of a and b as being the sizes of $|A|$ and $|B|$ in preparation to use a union bound. Let $k = k(a,b,n)$ be defined as $k = \max\{k^*, \frac{1}{e}\}$, where k^* satisfies

$$k^* \log k^* = \frac{(6+m)(a \vee b)n}{abd} \log \frac{n}{a \vee b},$$

or $\frac{1}{e}$, whichever is larger. When $a \vee b \leq \frac{n}{e}$, it follows that

$$(6+m)(a \vee b) \log \tfrac{n}{a \vee b} \geq 2a \log \tfrac{n}{a} + 2b \log \tfrac{n}{b} + (2+m)(a \vee b) \log \tfrac{n}{a \vee b},$$

where we have used the monotonicity of $x \log \frac{n}{x}$ on $[1, \frac{n}{e}]$; thus

$$(6+m)(a \vee b) \log \tfrac{n}{a \vee b} \geq a(1 + \log \tfrac{n}{a}) + b(1 + \log \tfrac{n}{b}) + (2+m) \log n.$$

Exponentiating,

$$\exp\left[k \log k \tfrac{abd}{n}\right] \geq \left(\tfrac{ea}{n}\right)^n \left(\tfrac{eb}{n}\right)^n n^{2+m},$$

if $k \geq \frac{1}{e}$. It follows that

$$\mathbb{P}\left[\exists A, B \text{ with } |A| = a,\ |B| = b,\ \text{so that } e(A,B) \geq e^2 k(a,b)\mu(A,B)\right] \\ \leq \binom{n}{a}\binom{n}{b} \exp(-e^2 k[\log k]\mu) \leq n^{-2-m}.$$

Moreover, applying this bound to all a and b, it follows that

$$e(A,B) \leq e^2 k(|A|,|B|)\mu(A,B),$$

except with probability smaller than n^{-m}. If for two sets A and B, $k = \frac{1}{e}$, then

$$e(A,B) \leq e\mu(A,B),$$

and we are in the first case of the discrepancy property, for $c_1 \geq e$. Otherwise,

$$e(A,B)\log k \leq e^2 k \log k \mu(A,B) = e^2(6+m)(a \vee b)\log \frac{n}{a \vee b},$$

and noting that $k \geq \frac{e(A,B)}{e^2\mu(A,B)}$, it follows that

$$\tfrac{1}{2}e(A,B)\log\frac{e(A,B)}{\mu(A,B)} \leq e(A,B)\log\frac{e(A,B)}{e^2\mu(A,B)} \leq e^2(6+m)(a\vee b)\log\frac{n}{a\vee b},$$

when $\frac{e(A,B)}{\mu(A,B)} \geq e^4$. If this is not the case, then we are again in the first case of the discrepancy property, taking $c_1 \geq e^4$. Taking $c_1 = e^4$, it follows that we may take $c_2 = 2e^2(6+m)$. □

The discrepancy property implies that there are no dense subgraphs, and thus the contribution of the heavy couples is not too large.

Lemma 26. *If the discrepancy property holds, with associated constants c_1 and c_2, then*

$$\sum_{\{u,v\}\in\mathcal{H}} |x_u A_{u,v} y_v| \leq C\sqrt{d},$$

for some constant C depending on $c_1, c_2,$ and δ.

Proof. The method of proof here is essentially identical to Kahn and Szemerédi or Feige and Ofek (see [49] or [42]). We provide a proof of this lemma for completeness as well as to establish the constants involved. We will partition the summands into blocks where each term x_u or y_v has approximately the same magnitude. Thus let $\gamma_i = 2^i\delta$, and put

$$A_i = \left\{u \;\middle|\; \tfrac{\gamma_{i-1}}{\sqrt{n}} \leq |x_u| < \tfrac{\gamma_i}{\sqrt{n}}\right\}, \qquad 1 \leq i \leq \log\lceil\sqrt{n}\rceil.$$

$$B_i = \left\{u \;\middle|\; \tfrac{\gamma_{i-1}}{\sqrt{n}} \leq |y_u| < \tfrac{\gamma_i}{\sqrt{n}}\right\}, \qquad 1 \leq i \leq \log\lceil\sqrt{n}\rceil.$$

Let $\hat{\mathcal{H}}$ denote those pairs (i,j) so that $\gamma_i\gamma_j \geq \sqrt{d}$. The contribution of the absolute sum can, in these terms, be bounded by

$$\sum_{(u,v)\in\mathcal{H}} |x_u A_{u,v} y_v| \leq \sum_{(i,j)\in\hat{\mathcal{H}}} \frac{\gamma_i\gamma_j}{n} e(A_i, B_j).$$

Let $\lambda_{i,j} = \frac{e(A_i,B_j)}{\mu(A_i,B_j)}$ denote the discrepancy, which can be controlled using Lemma 24. In terms of this quantity, the bound becomes

$$\sum_{(u,v)\in\mathcal{H}} |x_u A_{u,v} y_v| \leq \sum_{(i,j)\in\hat{\mathcal{H}}} \frac{\gamma_i\gamma_j}{n} \lambda_{i,j}|A_i||B_j|\tfrac{d}{n}.$$

In this form, the magnitudes of each of the quantities are somewhat opaque. Consider the sum $\sum_i |A_i|\frac{\gamma_i^2}{n}$; it is at most $4\|x\|^2 = 4$. In particular, it is of constant order. Thus let $\alpha_i = |A_i|\frac{\gamma_i^2}{n}$ and $\beta_j = |B_j|\frac{\gamma_j^2}{n}$. This allows the bound to be rewritten as

$$d \sum_{(i,j)\in\hat{\mathcal{H}}} \frac{\gamma_i^2|A_i|}{n}\frac{\gamma_j^2|B_j|}{n}\frac{\lambda_{i,j}}{\gamma_i\gamma_j} = \tfrac{d}{\sqrt{d}} \sum_{(i,j)\in\hat{\mathcal{H}}} \alpha_i\beta_j \frac{\lambda_{i,j}\sqrt{d}}{\gamma_i\gamma_j}.$$

This exposes the quantity $\sigma_{i,j} = \frac{\lambda_{i,j}\sqrt{d}}{\gamma_i\gamma_j}$ as having some special importance. In effect, we will show that either for fixed i, $\sum_j \sigma_{i,j}\beta_j$ has constant order, or for fixed j, $\sum_i \sigma_{i,j}\alpha_i$ has constant order.

In what follows, we will bound the contribution of the summands where $|A_i| \geq |B_j|$. By symmetry, the contribution of the other summands will have the same bound.

The heavy couples will now be partitioned into 6 classes $\{\hat{\mathcal{H}}_i\}_{i=1}^6$ where their contribution is bounded in a different way. Let $\hat{\mathcal{H}}_i \subseteq \hat{\mathcal{H}}$ be those pairs (i,j) which satisfy the i^{th} property from the following list but none of the prior properties:

1. $\sigma_{i,j} \le c_1$.
2. $\lambda_{i,j} \le c_1$.
3. $\gamma_j > \frac{1}{4}\sqrt{d}\gamma_i$.
4. $\log \lambda_{i,j} > \frac{1}{4}\left[2\log\gamma_i + \log\frac{1}{\alpha_i}\right]$.
5. $2\log\gamma_i \ge \log\frac{1}{\alpha_i}$.
6. $2\log\gamma_i < \log\frac{1}{\alpha_i}$.

The last properties are better understood when the second case of the discrepancy property is expressed in present notation. In its original form, it states

$$e(A_i, B_j)\log\lambda_{i,j} \le c_2|A_i|\log\tfrac{n}{|A_i|}.$$

Substituting γ_i^2/α_i for $n/|A_i|$ and multiplying both sides of this equation through by $\frac{\gamma_i}{|B_j|\gamma_j\sqrt{d}\log\lambda_{i,j}}$ produces the equivalent form

$$\sigma_{i,j}\beta_j \le c_2\frac{\gamma_j}{\sqrt{d}\gamma_i}\frac{\left[2\log\gamma_i + \log\frac{1}{\alpha_i}\right]}{\log\lambda_{i,j}}.$$

Thus, the last 3 cases cover each of the possible dominant log terms in this bound.

Bounding the contribution of $\hat{\mathcal{H}}_1$ and $\hat{\mathcal{H}}_2$.

In either of these situations, we have a bound on $\sigma_{i,j}$. Especially, either $\sigma_{i,j} \le c_1$ or, all the discrepancies $\lambda_{i,j}$ are uniformly bounded by c_1. As

$$\sigma_{i,j} = \frac{\lambda_{i,j}\sqrt{d}}{\gamma_i\gamma_j},$$

and $\gamma_i\gamma_j \geq \sqrt{d}$,

$$\sigma_{i,j} \leq c_1$$

for both cases.

Bounding the contribution of $\hat{\mathcal{H}}_3$.

For these terms, we fix j. In this case, the magnitudes of the entries corresponding to j of y_v dominate those of the entries corresponding to i of x_u. However, by regularity $e(A_i, B_j) \leq |B_j|d$, so that the discrepancy $\lambda_{i,j}$ is at most $\frac{n}{|A_i|} = \frac{\gamma_i^2}{\alpha_i}$.

$$\sum_{i\,:\,(i,j)\in\hat{\mathcal{H}}_3} \alpha_i\sigma_{i,j} = \sum_{i\,:\,(i,j)\in\hat{\mathcal{H}}_3} \alpha_i \frac{\lambda_{i,j}\sqrt{d}}{\gamma_i\gamma_j} \leq \sum_{i\,:\,(i,j)\in\hat{\mathcal{H}}_3} \frac{\gamma_i\sqrt{d}}{\gamma_j} \leq 8,$$

where in the last step it has been used that the sum is geometric with leading term less than $4\gamma_j/\sqrt{d}$.

Bounding the contribution of $\hat{\mathcal{H}}_4$.

For these terms, we fix i. We are not in case (2), and it follows that the second case of the discrepancy property holds. In present notation

$$\sigma_{i,j}\beta_j \leq c_2 \frac{\sqrt{d}\gamma_j}{d\gamma_i} \frac{\left[2\log\gamma_i + \log\frac{1}{\alpha_i}\right]}{\log\lambda_{i,j}} \leq \frac{4c_2\gamma_j}{\gamma_i\sqrt{d}},$$

where the hypothesis has been used. As we are not in case (3), the sum of these terms is bounded as

$$\sum_{j\,:\,(i,j)\in\hat{\mathcal{H}}_4} \beta_j\sigma_{i,j} \leq 2c_2,$$

where it has been used that the sum above has a geometric dominator with leading term at most $\frac{1}{4}\gamma_i\sqrt{d}$.

Bounding the contribution of $\hat{\mathcal{H}}_5$.

For these terms, we fix i. Again, the second case of the discrepancy property holds. Now, in addition,

$$\log \lambda_{i,j} \leq \tfrac{1}{4}\left[2\log\gamma_i + \log\tfrac{1}{\alpha_i}\right] \leq \log\gamma_i,$$

i.e. that $\lambda_{i,j} \leq \gamma_i$. Furthermore, we are not in case (1) so $c_1 \leq \sigma_{i,j} = \frac{\lambda_{i,j}\sqrt{d}}{\gamma_i\gamma_j} \leq \frac{\sqrt{d}}{\gamma_j}$. Thus the second discrepancy bound becomes

$$\sigma_{i,j}\beta_j \leq c_2 \frac{\sqrt{d}\gamma_j}{d\gamma_i}\frac{\left[2\log\gamma_i + \log\frac{1}{\alpha_i}\right]}{\log\lambda_{i,j}} \leq c_2 \frac{\gamma_j 4\log\gamma_i}{\sqrt{d}\gamma_i\log c_1} \leq \frac{4c_2}{c_1}\frac{\gamma_j}{\sqrt{d}},$$

where it has been used that $\gamma_i \geq \lambda_{i,j} \geq c_1 \geq e$, and that $\log x/x$ is monotonically decreasing for $x > e$. Thus,

$$\sum_{j\,:\,(i,j)\in\hat{\mathcal{H}}_5} \beta_j\sigma_{i,j} \leq \sum_{j\,:\,(i,j)\in\hat{\mathcal{H}}_5} \frac{4c_2}{c_1}\frac{\gamma_j}{\sqrt{d}} \leq \frac{8c_2}{c_1^2},$$

where it has been used that the second sum above is geometric with largest term $\sqrt{d}/c_1$.

Bounding the contribution of $\hat{\mathcal{H}}_6$.

For these terms, we fix j. The second case of the discrepancy property holds and in addition,

$$\log \lambda_{i,j} \leq \tfrac{1}{4}\left[2\log\gamma_i + \log\tfrac{1}{\alpha_i}\right] \leq \tfrac{1}{2}\log\tfrac{1}{\alpha_i}.$$

This implies that σ satisfies the asymmetric bound $\sigma_{i,j} \leq \frac{1}{\alpha_i}\frac{\sqrt{d}}{\gamma_i\gamma_j}$. Thus,

$$\sum_{i\,:\,(i,j)\in\hat{\mathcal{H}}_6} \alpha_i\sigma_{i,j} \leq \sum_{i\,:\,(i,j)\in\hat{\mathcal{H}}_6} \frac{\sqrt{d}}{\gamma_i\gamma_j} \leq 2,$$

where it has been used that the sum above is geometric with leading term $\frac{1}{\sqrt{d}}$ (which follows as $\gamma_i\gamma_j \geq \sqrt{d}$).

Assembling the bound

We must sum the contributions of each of the classes of couples. Recall that we must double the contribution here because we have only considered couples where $|A_i| \geq |B_j|$. In each of the cases outlined above, it only remains to sum over the α_i or β_j in each bound. Doing so contributes a factor of 4 to each bound, so that the constant can be given by

$$2\left[16c_1 + 32 + 8c_2 + \frac{32c_2}{c_1^2} + 8\right]$$

□

3.4.3 Finalizing the proof of Theorem 7

Proof. We will take $\delta = \frac{1}{2}$. With m given, it follows the discrepancy property (Lemma 24) holds with probability at least $1 - n^{-m}$, and with constants $c_1 = e^4$ and $c_2 = 2e^2(6+m)$. Therefore, by Lemma 26, for any two $x, y \in \mathcal{T}$, the contribution of the heavy couples to $x^t A y$ (which is at most twice the contribution of $x^t M y$, given that the bounds hold for *all* x and y) is at most

$$4\left[16c_1 + 32 + 8c_2 + \frac{32c_2}{c_1^2} + 8\right]\sqrt{d} \leq (8854 + 585m)\sqrt{d}.$$

By Theorem 12, with probability at least $(1 - C\exp(-cn)$ for some universal constants $C > 0$ and $c > 0$, the contribution of the light couples is never more than $110\sqrt{d}$. Hence

$$\sup_{x,y\in\mathcal{T}} |x^t A y| \leq (8964 + 585m)\sqrt{d},$$

except with probability at most $n^{-m} + C\exp(-cn)$. At last, this implies that $\lambda_2 \vee |\lambda_n| \leq 4(8964 + 585m)\sqrt{d}$, except with probability at most $n^{-m} + C\exp(-cn)$. □

Chapter 4

SPECTRAL GAP OF THE LAPLACIAN OF ERDŐS–RÉNYI GRAPHS

ADAPTED FROM JOINT WORK WITH CHRIS HOFFMAN AND MATT KAHLE [63].

4.1 Introduction

The notion of graph *expansion* has received an enormous amount of study over the past decades; they have been useful across a wide range of fields in mathematics and computer science including algorithm design, sorting networks, computational complexity theory and cryptography. An excellent introduction is available in [64]. Suffice it so say that a graph is an expander if the top eigenvalue of its adjacency matrix is much larger in magnitude than all the others. The eigenvalue bound of the previous chapter can be understood through this lens as showing that d-regular graphs are good expanders with high probability. In fact, various kinds of random graphs have played a major role in showing the existence of expanders — see for example Pinsker [101] and Barzdin–Kolmogorov [11].

In fact, we will not deal with the adjacency matrix, but instead we will focus on the *normalized Laplacian* $\mathscr{L}$. If the vertices of the graph are labeled $1, 2, \ldots, n$ the normalized Laplacian $\mathscr{L}$ is defined entrywise as

$$\mathscr{L}_{i,j} = \delta_{i,j}\mathbf{1}\{\deg i > 0\} - \frac{\mathbf{1}_{i \leftrightarrow j}}{\sqrt{\deg i \deg j}}.$$

If either degree is 0, we take the fraction to be 0. When all degrees in the graph are positive, we may write

$$\mathscr{L} = I - T^{-1/2}AT^{-1/2},$$

where T is the diagonal matrix of degrees, and A is the adjacency matrix. This becomes the normalized Laplacian of an Erdős–Rényi graph when we let each of $\mathbf{1}_{v_i \leftrightarrow v_j}$ be i.i.d. Bernoulli(p) variables. The matrix is easily checked to positive semidefinite, and 0 is always an eigenvalue of the matrix. We label the eigenvalues of $\mathscr{L}$ by $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.

A good introduction to the properties of the normalized Laplacian are available in [28] (note this is Chung's $\mathcal{L}$). We note that some authors use an alternate definition of normalized Laplacian, always putting a 1 on the diagonal. The two definitions agree when the graph has no isolated vertices.

To compare this with Wigner matrices, we will work with what is essentially $I - \mathscr{L}$. Define $M_{i,j} = \frac{\mathbf{1}_{i \leftrightarrow j}}{\sqrt{\deg i \deg j}}$. Thus if all degrees are positive we have $M = T^{-1/2} A T^{-1/2} = I - \mathscr{L}$, and it is easily checked that $T^{1/2}\mathbf{1}$ is an eigenvector with eigenvalue one. However, even in the case that some degrees are 0, it can be checked that $T^{1/2}\mathbf{1}$ is an eigenvector with eigenvalue 1 so long as the graph has at least one edge. We label the eigenvalues of M by $1 = \tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \cdots \geq \tilde{\lambda}_n$.

4.1.1 Spectral gap

The basic random matrix question we will study here is $\max_{i \geq 2} |1 - \lambda_i|$. When the graph has no isolated vertices, this is the same as $\tilde{\lambda}_2 \vee |\tilde{\lambda}_n|$. This is the principal spectral quantity of interest in a graph, where its applications include determining the mixing time of Markov chains and quantifying the expansion properties of a graph, to name a few (see [28] for further applications).

In the case of Wigner matrices, the largest eigenvalue distribution has been studied and understood in very great generality. For a large class of Wigner matrices ([109]), it is seen that

$$\lambda_1 \sim 2\sigma\sqrt{n} + X n^{1/6}$$

where σ^2 is the variance of an entry of the matrix, and X follows the Tracy-Widom law. Likewise, the joint distribution of the k largest eigenvalues has been studied and understood, where k is fixed and $n \to \infty$.[1] One might wonder if appropriately subtracting the means of the entries of the matrix, one can recover Wigner-like behavior for $\tilde{\lambda}_2 \vee |\tilde{\lambda}_n|$. This is known for the adjacency matrix for $p \geq n^{-2/3}$ [40].

Depending on how $p = p(n)$ varies with n, the matrices which we study here can be paradigmatically quite different from Wigner matrices. Many of the results about Wigner matrices, such as the semicircle law, require hypotheses that exclude adjacency matrices of very sparse random graphs. Along this line of development, the *given expected degree distribution* model of random graphs, which generalizes Erdős–Rényi, can have empirical spectral distributions that are approximately power law distributed, with appropriate choice of parameters [27]. This can be understood as the effect on the spectrum of having heavy-tailed entry distributions.

Beyond the difficulties with having lopsided entries, the normalized Laplacian we study differs from Wigner matrices in that it has globally dependent entries. This causes certain effects on the spectrum that are not present in Wigner matrices. In particular, we note that the dimension of the kernel of $\mathscr{L}$ is equal to the number of components of G. An immediate consequence is that for a graph with multiple nontrivial components, there is no spectral gap; in particular, for $p \ll \log n/n$, the spectral gap of the normalized Laplacian is 0.

Furthermore, many of the tools which were developed for estimation of the spectral gap were designed with Wigner matrices in mind, and thus their usefulness for matrices of sparse graphs is limited. In particular, we recall the bound of Füredi and Komlós [51] which can be extended to show that when $p \gg \log^6 n/n$, the spectral gap is of order $O(1/\sqrt{np})$. Improvements along this line of reasoning bring the range of feasible p to as low as $p \gg \log^2 n/n$ [120, 27]. Note that this rate is what should be anticipated by the same heuristics that work for Wigner matrices (see the discussion in Section 1.3).

---

[1] Note that the k smallest eigenvalues can be studied by negating the matrix, which in the Wigner case gives another Wigner matrix.

The method used to prove these bounds is the vaunted trace method, ubiquitous in random matrix theory, of estimating the expectation of high powers of the trace of a matrix. The state-of-the-art for this method is precisely what produces the spectral gap bound in the $p \gg \log^2 n/n$ regime. This method starts to lose effectiveness as the density of edges decreases to the connectivity threshold, where the cost of estimating a single eigenvalue by a trace becomes too high. In sparser regimes, alternative methods have been developed for controlling the spectral gap. In particular, we would like to once again draw attention the method of Kahn and Szemerédi [49].

Beyond its use in bounding the spectral gap of d-regular graphs, this method has already been used quite successfully for estimating the spectral gap of non-regular graphs in the sparse regime. Coja-Oghlan [29] shows that with $p \geq c \log n/n$, the gap is $O(1/\sqrt{np})$, and Feige and Ofek [42] show an analogous claim for the adjacency matrix. Our main contribution in this direction is to establish that for $p - \log n/n$ going to ∞ sufficiently quickly, the spectral gap goes to 1. More precisely, we are able to show the gap is $O((np)^{-1/4+\epsilon})$; it is an open question whether the $1/\sqrt{np}$ rate persists for p very near to the connectivity threshold. See Section 4.5 for simulations pertinent to this question.

Our main theorem on the spectral gap is the following.

Theorem 9. *Fix $k \geq 0$ and $0 < \epsilon \leq 1/4$, and let $G \sim G(n,p)$ where*

$$p \geq \frac{(k+1)\log n + C[\log n]^{1-2\epsilon}\log\log n}{n}.$$

Let $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of the normalized Laplacian $\mathscr{L}[G]$. Then G is connected and for each fixed $c > 0$,

$$\max_{i \neq 1} |1 - \lambda_i| < cd^{-1/4+\epsilon}$$

with probability $1 - o(n^{-k})$.

Here $d = (n-1)p$ denotes the expected degree of every vertex, and $C = C(\epsilon, k)$ is a constant which only depends on ϵ and k.

4.2 Higher dimensional expansion

Given the interest and broad applicability of expander graphs, the study of their higher-dimensional analogues is natural. For example Ramanujan complexes [78, 83, 84] are natural analogues of Ramanujan graphs [82]. Gromov et al. have discussed simplicial complex expanders in terms of "geometric overlap" properties [58, 45], with explicit examples as well as random ones.

Linial and Meshulam [41] define random 2-complexes $Y(n,p)$ based on Erdős–Rényi random graphs $G(n,p)$ and find that $p = 2\log n/n$ is a sharp threshold for vanishing of first cohomology [79], analogous to the Erdős–Rényi theorem which gives the threshold for connectivity of the random graph $G(n,p)$. The proof introduces a higher-dimensional analogue of edge expansion as an essential tool. In order to put our results in context, we first state the main result of [79].

Define $Y(n,p)$ to be the probability distribution on all simplicial complexes with n vertices, $\binom{n}{2}$ edges, and with each 2-dimensional face included independently with probability p. We use the notation $Y \sim Y(n,p)$ to mean that Y is chosen according to the distribution $Y(n,p)$. Let $H^1(Y,A)$ denote the first cohomology with coefficients in group A.

Linial–Meshulam theorem. Let $Y \sim Y(n,p)$.

1. If
$$p \leq \frac{2\log n - \omega}{n}$$
then $\mathbb{P}[H^1(Y,\mathbf{Z}_2) = 0] \to 0$, and

2. if
$$p \geq \frac{2\log n + \omega}{n}$$
then $\mathbb{P}[H^1(Y,\mathbf{Z}_2) = 0] \to 1$, as $n \to \infty$.

Here $\omega \to \infty$ is any function such that $\omega \to \infty$ as $n \to \infty$.

Meshulam and Wallach generalize this to d-dimensional random complexes, with arbitrary finite field coefficients [93]. (The same statement is true with cohomology H^1 replaced by homology H_1, by the Universal Coefficient Theorem.)

4.2.1 Property (T)

A group-theoretic notion closely related to expander graphs is Kazhdan's property (T), which can be considered as a type of strong non-amenability. The first explicit examples of expanders, due to Margulis, are constructed using Cayley graphs on quotients of (T) groups such as $SL(3, \mathbf{Z})$ [87]. Conversely, expansion properties of some graphs associated to the generating set of a group can imply property (T) (see [128]).

Property (T) has found use in many different areas of mathematics. For example, groups with property (T) lead to good mixing properties in ergodic theory — a process which mixes slowly must leave some subsets almost invariant. In particular, if a group Γ has property (T), then every ergodic Γ system is also strongly ergodic [55]. See the monograph [12] for a comprehensive overview of property (T), whose treatment we follow here. We also would like to draw attention to the Master's thesis (in French) [53] which contains a concise introduction to the material we need here.

The standard definition of property (T) is given in terms of Hilbert space representation theory of a group. A topological group G has Property (T) if there exist a compact subset Q and a real number $\varepsilon > 0$ so that for any continuous unitary representation of G on a Hilbert space $\mathcal{H}$ for which

$$\inf_{\|\xi\|=1} \sup_{q \in Q} \|\pi(q)\xi - \xi\| < \varepsilon,$$

there exists a vector $\eta \neq 0$ in $\mathcal{H}$ so that $\pi(g)\eta = \eta$ for all $g \in G$. We will work entirely with countable groups having discrete topology, and hence we may replace "compact" by finite in the above definition. The pair (Q, ε) is referred to as a *Kazhdan pair.*

Our main result is that the threshold for the random fundamental group $\pi_1(Y)$

to have property (T) agrees with the vanishing threshold for $H^1(Y, \mathbf{Z}_2)$. Most of the work lies in showing that, with high probability, certain normalized graph Laplacians have sufficiently large spectral gaps. We then can appeal to a Theorem of Żuk [129] (see Section 4.2.2) that connects properties of these Laplacians to the fundamental group — see also the more general statement in Ballman–Świątkowski [10], and earlier work by Garland [52].

Theorem 10. *Let $\omega \to \infty$ as $n \to \infty$, and $Y \sim Y(n,p)$.*

1. *If*
$$p \leq \frac{2\log n - \omega}{n}$$
then $\mathbb{P}[\pi_1(Y)$ has property (T)$] \to 0$, and

2. *if*
$$p \geq \frac{2\log n + \omega\sqrt{\log n}\log\log n}{n}$$
then $\mathbb{P}[\pi_1(Y)$ has property (T)$] \to 1$.

The threshold for property (T) is *sharp*, in the technical sense that the transition from probability $\to 0$ to probability $\to 1$, happens in a very narrow window, as in Section 3 of Friedgut and Kalai [47].

For our property (T) results we only require the case $k = 1$ of Theorem 9. The case $k = 0$ is also of particular interest as it gets closer to the connectivity threshold $p = \log n/n$ than earlier results. Other values of k have interest as well – see for example the main results of the concurrent papers [71] and [32], which use Theorem 9 with infinitely many values of k.

Both the Linial–Meshulam theorem and our Theorem 10 imply the following corollary.

Corollary 13. *Let $\epsilon > 0$ be fixed and $Y \sim Y(n,p)$. Then as $n \to \infty$,*

$$\mathbb{P}[H_1(Y,\mathbb{Q}) = 0] \to \begin{cases} 1 & : p \geq (2+\epsilon)\log n/n \\ \\ 0 & : p \leq (2-\epsilon)\log n/n \end{cases}$$

This follows from the Linial–Meshulam theorem by universal coefficients for homology and cohomology, and follows from our Theorem 10 since a (T) group has finite abelianization. Both these theorems depend on some kind of expansion property — for the first, a combinatorial "global" generalization of edge expansion, and for the second, a "local" spectral expansion condition.

4.2.2 Cohomological property (T) and Żuk's criterion

Since its definition, many properties have been shown to be equivalent to property (T). We will ultimately connect property (T) to properties of the Laplacian, and so it should not be surprising that property (T) has a cohomological connection. We suppose that G has a unitary action π on Hilbert space $\mathcal{H}$, and we define the cochain complex $C^k(G)$ by letting $C^k(G)$ denote the space of continuous functions from $G^k \to \mathcal{H}$. Property (T) can be understood by looking at the first cohomology group, and so we need only consider $k = \{0,1,2\}$.

We will define the coboundary maps on this chain, and we note that we may identify $C^0(G)$ with $\mathcal{H}$. The 0^{th} coboundary $d_0(v)$ is defined as $g \mapsto \pi(g)v - v$. The first $d_1(f)$ is given as $(g,h) \mapsto \pi(g)f(h) - f(gh) + f(g)$.[2] It is easily verified that $d_1 \circ d_0 = 0$. We write $Z^1 = \ker d_1$, the space of 1-cocycles and $B^1 = \operatorname{Im} d_0$ for the space of 1-coboundaries. Finally, we write $H^1(G) = Z^1/B^1$ for the first cohomology group. The important revelation is that a σ-compact, locally compact group has property (T) if and only if $H^1(G) = 0$ for all unitary representations π of G.[3]

[2] This is the standard Eilenberg-Maclane group cohomology with coefficients in the G-module $\mathcal{H}$.

[3] This follows from the theorem of Delorme-Guichardet on Properties (T) and (FH), see Proposition 2.2.10 and Theorem 2.12.4 of [12].

This cohomology can be directly connected to a certain simplicial cohomology theory, which we will now describe. Let X be a finite 2-dimensional simplicial complex, and let G be its fundamental group. Let $\tilde{X}$ be the universal cover of X, which we can actually take to be a 2-dimensional simplicial complex. Further, we can endow G with a *free, simplicial* action on $\tilde{X}$, meaning

1. Only the identity of G has a fixed point.

2. G acts by simplicial isomorphisms on $\tilde{X}$. [4]

We can identify X with the quotient space $\tilde{X}/G$. This allows us to orient each simplex $\tilde{X}$ in such a way that it is invariant under the action of G (by simply defining one for X and then extending it to be so for all $\tilde{X}$). Let $S^i(\tilde{X})$ be the collection of i-dimensional simplices, and define the cochain complex $C^k(\tilde{X})$ to be those maps

$$C^k(\tilde{X}) = \left\{ f : S^k(\tilde{X}) \to \mathcal{H} \ \mid f(g \cdot \triangle) = \pi(g) f(\triangle) \right\}.$$

We can then define the usual simplicial coboundary maps d_0 and d_1 on this cochain complex. Again we write $Z^1(\tilde{X}) = \ker d_1$, the space of 1-cocycles and $B^1(\tilde{X}) = \operatorname{Im} d_0$ for the space of 1-coboundaries, and define the first cohomology group $H^1(\tilde{X}) = Z^1(\tilde{X})/B^1(\tilde{X})$. The important point is that these two cohomologies are isomorphic, i.e. $H^1(G) \cong H^1(\tilde{X})$ (see [53], whose treatment of the proof we will follow). The most important tool used to make this connection is the *Cayley complex* (see Section 1.3 of [62]), which provides an alternate explicit construction of the universal cover as a 2-dimensional cellular complex over a Cayley graph of the fundamental group. As the universal cover is unique up to homeomorphism, it follows that the $H^1(\tilde{X})$ cohomology is isomorphic to the first cellular cohomology group of the Cayley complex (as both are realizations of the universal cover of X). On the other hand, it is possible to construct an explicit isomorphism of this cellular cohomology to $H^1(G)$.

[4] A map is *simplicial* if it maps simplices to simplices.

Thus, we are able to study property (T) by studying $H^1(\tilde{X})$. We define the *link* of a vertex $\mathrm{lk}(v)$ in complex X to be a graph with vertex set $\mathcal{V}(\mathrm{lk}(v))$ given by $\{e \in S^1(X) \mid v \in e\}$ and edge set $\{(e_1e_2) \mid \exists t \in S^2(X), e_1 \in t, e_2 \in t\}$. Further note that as $\tilde{X}$ is locally homeomorphic to X, the $\mathrm{lk}(\tilde{v})$ is a relabeling of $\mathrm{lk}(v)$, where $\tilde{v}$ is a lift of v. Note further that for $Y(n,p)$, a link becomes an Erdős–Rényi graph $G(n-1,p)$. Żuk's criterion can now be stated.

Żuk's criterion. If X is a pure 2-dimensional locally-finite simplicial complex so that for every vertex v, the vertex link $\mathrm{lk}(v)$ is connected and the normalized Laplacian $\Delta_v =$ satisfies $\lambda_2(\Delta_v) + \lambda_2(\Delta_w) > 1$ for all adjacent vertices v and w, then $\pi_1(X)$ has property (T).

The proof goes by defining an inner product on $C^r(\tilde{X})$ given by

$$\langle f, g \rangle = \sum_{s \in S^r(X)} \langle f(s), g(s) \rangle \, n(s), \tag{4.1}$$

where $n(s)$ is the number of triangles containing s. By G-equivariance, we have that this sum does not depend on the choice of representatives of $S^r(X)$. With respect to these inner products, we can define the adjoint maps d^*.

We would still like to relate this to the expansion properties of the links. Thus, we also define a localized inner product on each link:

$$\langle f, g \rangle_{\mathrm{lk}(v)} = \sum_{s \in \mathcal{V}(\mathrm{lk}(v))} \langle f(s), g(s) \rangle \, N(s), \tag{4.2}$$

where $N(s)$ is the degree of the vertex s. For $f \in C^1(\tilde{X})$ define f_v^* to be the induced map $f_v^* \in C_0(\mathrm{lk}(v))$ given by $f_v^*(e) = f(e)$, recalling the vertices of $\mathrm{lk}(v)$ are edges of $\tilde{X}$ based at v.

It is an elementary computation to show that

$$\langle df, df \rangle = \sum_{v \in S^0(X)} \left\langle (\Delta_v - \tfrac{1}{2}) f_v^*, f_v^* \right\rangle_{\mathrm{lk}(v)}, \tag{4.3}$$

where Δ_v is the Laplacian on the link given by $(\Delta_v f)(s) = f(s) - \frac{1}{N(s)} \sum_{t \sim s} f(t)$. A further computation shows that

$$\langle df, df \rangle + \langle d^* f, d^* f \rangle \geq (c-1) \langle f, f \rangle ,$$

where $c = \min_{(vw) \in S^1(X)} \{\lambda_2(\Delta_v) + \lambda_2(\Delta_w)\}$.

The conclusion of all of this development is that the operator $dd^* : C^1(\tilde{X}) \to B^1(\tilde{X})$ restricts to an injection of Z^1, as for $f \in Z^1 = \ker d$, we have

$$\langle dd^* f, f \rangle = \langle d^* f, d^* f \rangle \geq C \langle f, f \rangle .$$

As these are finite dimensional spaces, and we have $B^1(\tilde{X}) \subseteq Z^1(\tilde{X})$, and we conclude $H^1(\tilde{X}) = 0$.

4.2.3 Part (1) of Theorem 10

Let I denote the number of isolated edges of Y (i.e. edges not contained in any 2-face). By linearity of expectation

$$\mathbb{E}[I] = \binom{n}{2} (1-p)^{n-2}.$$

If $p = (2 \log n + c)/n$ with $c \in \mathbb{R}$ constant then $\mathbb{E}[I] \to e^{-c}/2$ as $n \to \infty$. Moreover by considering higher moments one can show that I approaches a Poisson distribution with mean $e^{-c}/2$ as $n \to \infty$ [72]. So

$$\mathbb{P}[I = 0] \to \exp[-e^{-c}/2],$$

as $n \to \infty$, so if $p \leq (2 \log n - \omega)/n$ where $\omega \to \infty$ arbitrarily slowly, $\mathbb{P}[I = 0] \to 0$.

An isolated edge in Y generates a free **Z** factor in $\pi_1(Y)$, but a (T) group cannot have a free **Z** factor.[5] So if

$$p \leq \frac{2 \log n - \omega}{n},$$

and $Y \sim Y(n,p)$ then $\pi_1(Y)$ a.a.s. does not have property (T).

[5] For if a group has a free **Z** factor, its abelianization includes a factor of **Z**, and (T) groups must have finite abelianization.

4.2.4 Part (2) of Theorem 10

Assume that

$$p \geq \frac{2\log n + \omega\sqrt{\log n}\log\log n}{n},$$

and $Y \sim Y(n,p)$.

Every vertex link $\mathrm{lk}(v)$ has the same distribution as an Erdős-Rényi random graph $G(n-1,p)$. Applying Theorem 9 with $k=1$ gives that $\lambda_2[\mathrm{lk}(v)] > 1/2$ with probability $1-o(n^{-1})$. Applying a union bound,

$$\mathbb{P}\left[\text{there exists a vertex } v \text{ such that } \lambda_2[\mathrm{lk}(v)] \leq 1/2\right] = o(1)$$

So a.a.s. every vertex link has large spectral gap. A large spectral gap implies in turn connectivity, and also that Y is pure 2-dimensional. Indeed, if edge $\{u,v\}$ were isolated, then v would be an isolated vertex in $\mathrm{lk}(u)$, contradicting connectivity.

Then Y is pure 2-dimensional, and every vertex link is connected with spectral gap $> 1/2$. Applying Żuk's criterion, we have that $\pi_1(Y)$ a.a.s. has property (T).

4.3 Proof of the spectral gap theorem

The approach of this proof is identical to the approach to Theorem 7. However, as we are not dealing with an adjacency matrix, many of the exact technical details are more complicated. The starting point of our argument is the variational characterization of the eigenvalues which states that

$$\tilde{\lambda}_2 = \inf_{v\neq 0} \sup_{\substack{w\perp v\\ \|w\|=1}} w^t M w.$$

The first bound is made by estimating the infimum by

$$\tilde{\lambda}_2 \leq \sup_{\substack{w\perp \mathbf{1}\\ \|w\|=1}} w^t M w,$$

which, when the degrees of the graph are concentrated around their means, is very nearly the minimum. Additional flexibility is provided by replacing this symmetric version of the Rayleigh quotient by the asymmetric version,

$$\tilde{\lambda}_2 \leq \sup_{\substack{w \perp \mathbf{1}, v \perp \mathbf{1} \\ \|v\|=\|w\|=1}} v^t M w. \tag{4.4}$$

With a little extra work, it is possible to also estimate $\tilde{\lambda}_2 \vee |\tilde{\lambda}_n|$. If $\tilde{\lambda}_n \geq 0$, then this is simply $\tilde{\lambda}_2$ and there is nothing to do. Otherwise, we would like to bound $-\tilde{\lambda}_n$ from above. Using the variational characterization once more, this has the exact value

$$-\tilde{\lambda}_n = -\inf_{\substack{w \\ \|w\|=1}} w^t M w = \sup_{\substack{w \\ \|w\|=1}} (-w^t M w).$$

For each w, orthogonally decompose it as $w = u + v$, where $v \perp \mathbf{1}$. As each entry of M is non-negative, $-u^t M u \leq 0$, and thus

$$-w^t M w \leq -v^t M u - u^t M v - v^t M v = -2v^t M u - v^t M v,$$

where the symmetry of M has been used. Taking absolute values, this can be bounded by

$$-2v^t M u - v^t M v \leq 2|v^t M u| + |v^t M v| \leq \left[\sup_{\substack{x \perp \mathbf{1}, y \\ \|x\|=\|y\|=1}} |x^t M y| \right] \left[2\|u\|\|v\| + \|v\|^2 \right].$$

Using that $\|u\|^2 + \|v\|^2 = 1$, it follows that

$$-\tilde{\lambda}_n \leq 2 \left[\sup_{\substack{x \perp \mathbf{1}, y \\ \|x\|=\|y\|=1}} |x^t M y| \right].$$

Recalling that $\tilde{\lambda}_2$ had a similar bound (4.4), it follows that the expression $\tilde{\lambda}_2 \vee |\tilde{\lambda}_n|$ may be bounded by

$$\tilde{\lambda}_2 \vee |\tilde{\lambda}_n| \leq 2 \left[\sup_{\substack{x \perp \mathbf{1}, y \\ \|x\|=\|y\|=1}} |x^t M y| \right]. \tag{4.5}$$

4.3.1 Outline

We begin with a relaxation lemma. This turns the problem of finding the spectral gap, which requires finding the supremum over an infinite set, to a problem that involves finding the maximum over a finite set. The cost of this relaxation is just a constant factor.

Lemma 27. *Define*

$$\mathcal{U} = \left\{ \frac{\delta z}{\sqrt{n}} \ : \ z \in \mathbf{Z}^n, \|z\|^2 \le \frac{n}{\delta^2} \right\} \qquad \textit{and} \qquad \mathcal{T} = \{ z \in \mathcal{U} \ : \ z \perp \mathbf{1} \},$$

Then

$$\tilde{\lambda}_2 \vee |\tilde{\lambda}_n| \le 2 \sup_{\substack{v \perp \mathbf{1}, w \\ \|v\| = \|w\| = 1}} v^t M w \le 2(1-\delta)^{-2} \sup_{x \in \mathcal{T}, y \in \mathcal{U}} x^t M y.$$

and

$$|\mathcal{T}| \le |\mathcal{U}| \le C \left[\frac{(2+\delta)\sqrt{2e\pi}}{2\delta} \right]^n.$$

Proof. Every v with $\|v\| \le 1-\delta$ and $v \perp \mathbf{1}$ is a convex combination of points in $\mathcal{T}$ (See Lemma 2.3 of [42]). Likewise, it is easily shown that every w with $\|w\| \le 1-\delta$ is a convex combination of $\mathcal{U}$. Thus

$$\sup_{\substack{v \perp \mathbf{1}, w \\ \|v\| = \|w\| = 1}} \left| [(1-\delta)v]^t M [(1-\delta)w] \right| \le \sup_{x \in \mathcal{T}, y \in \mathcal{U}} x^t M y.$$

Furthermore, by a volume argument, it is possible to bound the cardinality of $\mathcal{U}$ as

$$|\mathcal{U}| \left(\frac{\delta}{\sqrt{n}} \right)^n \le \mathrm{Vol} \left[x \in \mathbb{R}^n \ : \ \|x\| \le 1 + \delta/2 \right] = \frac{(1+\delta/2)^n \sqrt{\pi}^n}{\Gamma(\frac{n}{2}+1)}.$$

Employing Stirling's approximation, this shows

$$|\mathcal{U}| \le C \left[\frac{(2+\delta)\sqrt{2e\pi}}{2\delta} \right]^n.$$

for some universal constant C. □

Next we define some regularity conditions on our graph G which facilitate a large spectral gap. Fix $m \ge 0$, and let $c_*(m) = 18 + 3m$. Further, fix an ϵ with $1/4 \ge \epsilon > 0$, and define $f(d) = c_3 d^{1/4+\epsilon}$ where $c_3 \le 1$ is a constant to be determined later. In terms of these constants, define the following three regularity conditions.

1. **Bounded degree condition (b.d.c)** Every vertex has degree at most $c_* d$.

2. **Minimum degree condition (m.d.c)** Every vertex has degree at least

$$\min_u [\deg u] \geq \frac{d^{3/2}}{f(d)^2}.$$

3. **Discrepancy** For every pair of vertex sets A and B, letting $e(A,B)$ denote the number of edges between the sets and $\mu(A,B) = \frac{|A||B|d}{n}$, one of

   (a) $\frac{e(A,B)}{\mu(A,B)} \leq ec_*$

   (b) $e(A,B) \log \frac{e(A,B)}{\mu(A,B)} \leq c_*(|A| \vee |B|) \log \frac{n}{|A| \vee |B|}$

   occurs.

In the $p \gg \log n/n$ regime, all of these criteria hold with overwhelming probability. Of great importance here is verifying each of these conditions with high probability when p is very nearly this threshold. The most difficult of these to satisfy is the **m.d.c.**, which requires $p - \log n/n = \omega(1/n)$. The other two properties are actually satisfied for any $p = \Omega(\log n/n)$ with sufficiently large choice of c_*. Regardless, all of these regularity properties are consequences of tail bounds for binomial random variables.

Lemma 28. *Let G be an Erdős–Rényi random graph chosen from $G(n,p)$ where $p \geq \frac{\log n}{n}$. For every $m \geq 0$ both the* **b.d.c** *and* **discrepancy** *properties hold with probability at least $1 - o(n^{-m})$.*

Lemma 29. *For all $k \geq 0$ and $1/4 \geq \epsilon > 0$ there exists $C = C(k,\epsilon)$ such that if G is an Erdős–Rényi random graph with*

$$p \geq \frac{(k+1)\log n + C(\log n)^{1-2\epsilon} \log\log n}{n}$$

then

$$\mathbb{P}(\textbf{m.d.c.}\ \textit{fails}) = o(n^{-k}).$$

Following Feige and Ofek we make the following definition.

Definition 5. For a fixed pair of vectors $(x, y) \in \mathcal{T} \times \mathcal{U}$, define the **light couples** $\mathcal{L} = \mathcal{L}(x, y)$ to be all those ordered pairs $(u, v) \in \{1, 2, \ldots, n\}^2$ so that $|x_u y_v| \le \frac{f(d)}{n}$, and let the **heavy couples** $\mathcal{H} = \mathcal{H}(x, y)$ be all those pairs that are not light.

Using this logical division, we get

$$\sum_{(u,v)\in\{1,\ldots,n\}^2} x_u M_{uv} y_v = \sum_{(u,v)\in\mathcal{L}} x_u M_{uv} y_v + \sum_{(u,v)\in\mathcal{H}} x_u M_{uv} y_v.$$

If the spectral gap is small then there must be an x and y such that one of the two terms on the right hand side is large. To show this happens with small probability we get a bound on the probability that these sums are large for each fixed pair x and y. After that, we apply the union bound.

First, we get control of the probability that the light couples contribute a large quantity to the sum. The main tool needed to establish this result is Freedman's martingale inequality. Let $X_1, X_2, \ldots$ be martingale increments. Write $\mathscr{F}_k$ for the natural filtration induced by these increments, and define $V_k = \mathbb{E}\left[X_k^2 \mid \mathscr{F}_{k-1}\right]$. If S_n is the partial sum $S_n = \sum_{i=1}^n X_i$ and T_n is the sum $T_n = \sum_{i=1}^n V_i$, then by analogy with the continuous case, one expects S_n to be a Brownian motion at time T_n (a discretization of the bracket process). The analogy requires, however, that the increments have some a priori bound. In its strongest form, put τ to be a stopping time so that for $k \le \tau$, $|X_k| \le R$. Then

$$\mathbb{P}\left[\exists\ n \le \tau \text{ so that } S_n \ge a \text{ and } T_n \ge b\right] \le 2\exp(-a^2/2(Ra+b)).$$

(see Proposition 2.1 of [46]). With this tool in hand, we form a martingale by revealing edges one at a time and taking conditional expectations of the random variable $\sum_{(u,v)\in\mathcal{L}} x_u M_{uv} y_v$. By virtue of having selected only the light couples, it is possible to bound the increments $|X_k|$ to have the right order.

Lemma 30. *There exist constants C_0, C_1 and C_2 so that for every $(x,y) \in \mathcal{T} \times \mathcal{U}$, $t \geq 2$ and all n sufficiently large,*

$$\mathbb{P}\left[\left|\sum_{(u,v)\in\mathcal{L}} x_u M_{uv} y_v\right| > \tfrac{t(1+c_*)f(d)}{\sqrt{d}} \text{ and } \mathbf{b.d.c.}\right] \leq C_0 \exp\left(\frac{-nt^2}{C_1 + C_2\frac{t}{\sqrt{d}}}\right).$$

To control the heavy couples, it would suffice that with high probability, the edge density between two vertex sets is no more than a constant multiple of its expectation. However, to require this of every pair of vertex sets is too restrictive in the $d \sim \log n$ regime. Instead, we use the **discrepancy** property, which allows larger discrepancies for small vertex sets. The three regularity properties are then used to show that the contribution of the heavy couples cannot be too large.

Lemma 31. *If the* **b.d.c.**, **m.d.c** *and* **discrepancy** *properties all hold then*

$$\sum_{(u,v)\in\mathcal{H}} |x_u M_{uv} y_v| \leq C\tfrac{f(d)}{\sqrt{d}},$$

for some constant $C = C(c_*)$.

Finally, by combining these individual bounds, we are able to prove that the spectral gap is large with high probability.

Theorem 11. *For all $k \geq 0$ and $1/4 \geq \epsilon > 0$ there exists $C = C(k,\epsilon)$ so that for all $\eta > 0$,*

$$p \geq \frac{(k+1)\log n + C(\log n)^{1-2\epsilon}\log\log n}{n}$$

then

$$\mathbb{P}\left[\tilde{\lambda}_2 \vee |\tilde{\lambda}_n| \geq \eta d^{-1/4+\epsilon}\right] = o(n^{-k}).$$

Proof. Recall that $f(d) = c_3 d^{1/4+\epsilon}$, where c_3 has yet to be chosen. Fix some $C > 0$ to be determined later. By Lemma 27 if

$$\tilde{\lambda}_2 \vee |\tilde{\lambda}_n| \geq \frac{4C}{(1-\delta)^2}\tfrac{f(d)}{\sqrt{d}}$$

then either there exists $x \in \mathcal{T}$ and $y \in \mathcal{U}$ such that either

$$\left|\sum_{(u,v)\in\mathcal{L}} x_u M_{uv} y_v\right| \geq C\tfrac{f(d)}{\sqrt{d}} \qquad \text{or} \qquad \left|\sum_{(u,v)\in\mathcal{H}} x_u M_{uv} y_v\right| \geq C\tfrac{f(d)}{\sqrt{d}}. \tag{4.6}$$

Let $E_{x,y}$ be the event that

$$\left|\sum_{(u,v)\in\mathcal{L}} x_u M_{uv} y_v\right| > C\frac{f(d)}{\sqrt{d}} \text{ and } \mathbf{b.d.c.}$$

By Lemma 30 the left hand side of (4.6) can happen only if $E_{x,y}$ happens for some $(x,y) \in \mathcal{T} \times \mathcal{U}$ or if the bounded degree condition fails. By Lemma 31 the right hand side of (4.6) can happen only if at least one of the bounded degree condition, the minimal degree condition or the discrepancy conditions fail. Thus

$$\begin{aligned} \mathbb{P}\left(\tilde{\lambda}_2 \vee |\tilde{\lambda}_n| \geq \tfrac{4C}{(1-\delta)^2}\cdot\tfrac{f(d)}{\sqrt{d}}\right) &\leq \sum_{x,y\in\mathcal{T}} \mathbb{P}(E_{x,y}) \\ &\quad +\mathbb{P}(\ \mathbf{m.d.c.}, \mathbf{b.d.c.} \text{ or } \mathbf{discrepancy} \text{ fails}) \end{aligned}$$

By Lemma 30 if $C > t(1+c_*)$ then

$$\mathbb{P}(E_{x,y}) \leq C_0 \exp\left(\frac{-nt^2}{C_1 + C_2\frac{t}{\sqrt{d}}}\right).$$

We can choose t such that

$$\exp\left(\frac{t^2}{C_1 + C_2\frac{t}{\sqrt{d}}}\right) > 1000,$$

so that provided that $C > t(1+c_*)$,

$$\mathbb{P}(E_{x,y}) \leq 1000^{-n}.$$

Applying the regularity Lemmas 28 and 29 with $m = k$,

$$\begin{aligned} \mathbb{P}\left(\tilde{\lambda}_2 \vee |\tilde{\lambda}_n| \geq \tfrac{4C}{(1-\delta)^2}\cdot\tfrac{f(d)}{\sqrt{d}}\right) &\leq \sum_{x,y\in\mathcal{T}} \mathbb{P}(E_{x,y}) \\ &\quad +\mathbb{P}(\mathbf{m.d.c.}, \mathbf{b.d.c.} \text{ or } \mathbf{discrepancy} \text{ fails}) \\ &\leq |\mathcal{T}||\mathcal{U}|1000^{-n} + o(n^{-k}) \\ &\leq \left(\tfrac{9\cdot 2e\pi}{4\delta^2}\right)^n 1000^{-n} + o(n^{-k}) \\ &\leq o(n^{-k}), \end{aligned}$$

where we pick δ sufficiently large. Thus, choosing c_3 sufficiently small that

$$\frac{4C}{(1-\delta)^2}c_3 < \eta,$$

it has been shown that

$$\mathbb{P}\left(\tilde{\lambda}_2 \vee |\tilde{\lambda}_n| \geq \eta d^{-1/4+\epsilon}\right) = o(n^{-k}).$$

□

From this point, it is straightforward to transfer the results about the spectrum of M to statements about the spectral gap of L.

Proof of Theorem 9. Apply Theorem 11 with $\eta = c$, and it follows that there is a constant $C = C(\epsilon, k)$ so that

$$\mathbb{P}\left(\tilde{\lambda}_2 \vee |\tilde{\lambda}_n| \geq cd^{-1/4+\epsilon}\right) = o(n^{-k})$$

for

$$p \geq \frac{(k+1)\log n + C(\log n)^{1-2\epsilon}\log\log n}{n}.$$

As the **m.d.c.** held, the graph has an edge except for with probability $o(n^{-k})$, and so the largest eigenvalue of M is exactly 1. The spectrum of L is just one minus the spectrum of M, and so

$$\max_{i \neq 1} |1-\lambda_i| \leq \tilde{\lambda}_2 \vee |\tilde{\lambda}_n| \leq cd^{-1/4+\epsilon}$$

except for with probability $o(n^{-k})$.

□

4.3.2 Bounds for the failure probabilities of the regularity conditions

Proof of Lemma 28. For any vertex v, $\deg(v)$ is a binomial random variable with mean $d > \log(n)$. By Lemma 35

$$\begin{aligned} \mathbb{P}(\deg(v) > c_* d) &\leq \exp\left(-\tfrac{d c_* \log c_*}{3}\right) \\ &\leq \exp\left(-\log(n)(1+m)\log c_*\right) \\ &= o(n^{-1-m}). \end{aligned}$$

Taking the union bound over all vertices, the graph has the bounded degree property with probability $1 - o(n^{-m})$.

We will now turn to showing the discrepancy property.

Let D be the event that the discrepancy condition fails and let $D(A, B)$ be the event that the discrepancy condition fails for sets A and B. Then by the union bound

$$\begin{aligned} \mathbb{P}(D) &\leq \mathbb{P}(\exists A, B \text{ with } |A| \wedge |B| \geq n/e) : D(A,B) \text{ occurs}) \\ &+ \mathbb{P}(\exists A, B \text{ with } |A| \vee |B| \geq n/e \geq |A| \wedge |B|) : D(A,B) \text{ occurs}) \\ &+ \sum_{A,B:\ |A| \vee |B| < n/e} \mathbb{P}(D(A,B)) \end{aligned}$$

If $|A| \wedge |B| \geq \frac{n}{e}$, and

$$e(A, B) > c_* \mu(A, B) > c_*(n/e)^2 d/n > nd$$

and there are at least nd edges in the graph. The distribution of the number of edges is binomial with mean $n(n-1)p/2 = nd/2$. The probability of this is going to zero exponentially in $n(n-1)p/2 > n$ and

$$\mathbb{P}(\exists A, B \text{ with } |A| \wedge |B| \geq n/e) : D(A,B) \text{ occurs}) = o(n^{-m}) \tag{4.7}$$

If $|A| \vee |B| \geq \frac{n}{e} > |A| \wedge |B|$ and the bounded degree condition fails for some vertex in $A \cup B$ or else $e(A, B) \leq (|A| \vee |B|) c_* d$, so that

$$\frac{e(A,B)}{\mu(A,B)} \leq \frac{c_* nd(|A| \vee |B|)}{|A||B|d} = \frac{c_* n}{|A| \wedge |B|} \leq c_* e.$$

Thus

$$\begin{aligned}\mathbb{P}(\exists A, B \text{ with } |A| \vee |B| \geq n/e \geq |A| \wedge |B|) : D(A,B)\text{occurs}) &\leq \mathbb{P}(\text{b.d.c. fails}) \\ &= o(n^{-m}). \qquad (4.8)\end{aligned}$$

Now we need to deal with the case that both A and B are less than $\frac{n}{e}$. Choose $r = r(A, B, n) = c_* \vee r_1$ where r_1 is the solution to

$$\mu(A, B, n) r_1 \log(r_1) = c_*(|A| \vee |B|) \log \tfrac{n}{|A| \vee |B|}.$$

For any A, B and n we must have either

- $e(A, B) \leq r\mu(A, B, n)$ and $r = c_*$
- $e(A, B) \leq r\mu(A, B, n)$ and $r = r_1$ or
- $e(A, B) > r\mu(A, B, n)$

Thus if $D(A, B)$ occurs then at least one of the following three events occur.

- $D_1 = D_1(A, B) = \Big\{ e(A, B) \leq r\mu(A, B, n),\ r = c_* \text{ and } e(A, B) > c_*\mu(|A|, |B|, n) \Big\}$
- $D_2 = D_2(A, B) = \Big\{ e(A, B) \leq r\mu(A, B, n), r = r_1 \text{ and } e(A, B) \log \frac{e(A,B)}{\mu(A,B,n)} > c_*(|A| \vee |B|) \log \frac{n}{|A| \vee |B|} \Big\}$
- $D_3 = D_3(A, B) = \{e(A, B) > r\mu(A, B, n)\}$

For D_1 the conditions are mutually exclusive as $e(A, B)$ cannot be simultaneously greater than and less than or equal to $c_*\mu(|A|, |B|, n)$. Thus $D_1(A, B)$ is empty. For

D_2 we get similar contradiction after a little work.

$$\begin{aligned} e(A,B)\log\tfrac{e(A,B)}{\mu(A,B,n)} &> c_*(|A|\vee|B|)\log\tfrac{n}{|A|\vee|B|} \\ e(A,B)\log\tfrac{e(A,B)}{\mu(A,B,n)} &> \mu(A,B,n)r_1\log r_1 \\ \tfrac{e(A,B)}{\mu(A,B,n)}\log\tfrac{e(A,B)}{\mu(A,B,n)} &> r_1\log r_1 \\ \tfrac{e(A,B)}{\mu(A,B,n)} &> r_1 \\ e(A,B) &> r_1\mu(A,B,n) \\ e(A,B) &> r\mu(A,B,n). \end{aligned}$$

This is a contradiction so $D_2(A,B)$ is also empty.

Now we bound $\mathbb{P}(D_3(A,B))$. We have that $e(A,B)$ is a binomial random variable with mean $\mu = |A|\cdot|B|d/n$. Thus by Lemma 35

$$\mathbb{P}(e(A,B) > r\mu) \le \exp\left(-\tfrac{\mu r\log r}{3}\right)$$

for any $r \ge 4$.

$$\mathbb{P}(D_3(A,B)) \le \exp\left(-\tfrac{\mu(|A|,|B|,n)r\log r}{3}\right)$$

For all A,B we have $D \subset D_1\cup D_2\cup D_3$ and $\mathbb{P}(D_1(A,B)) = \mathbb{P}(D_2(A,B)) = 0$. Combining this with (4.7) and (4.8) we get

$$\begin{aligned} \mathbb{P}(D) &\le \mathbb{P}(\exists A,B:\ D(A,B)\text{ occurs}) \\ &\le \mathbb{P}(\exists A,B:\ |A|,|B|<n/e\text{ and }D(A,B)\text{ occurs}) + o(n^{-m}) \\ &\le \mathbb{P}(\exists A,B:\ |A|,|B|<n/e\text{ and }D_3(A,B)\text{ occurs}) + o(n^{-m}) \\ &\le \sum_{|A|,|B|\le n/e}\mathbb{P}(D_3(A,B)) + o(n^{-m}) \\ &\le \sum_{1\le a,b\le n/e}\ \sum_{|A|=a,|B|=b}\exp\left(-\tfrac{\mu r\log r}{3}\right) + o(n^{-m}) \\ &\le \sum_{1\le a,b\le n/e}\binom{n}{a}\binom{n}{b}\exp\left(-\tfrac{\mu(a,b,n)r\log r}{3}\right) + o(n^{-m}) \end{aligned}$$

To evaluate the last term we get

$$\begin{aligned}
\tfrac{\mu r \log r}{3} &\geq \Big(6+m\Big)\Big((|A| \vee |B|) \log \tfrac{n}{|A| \vee |B|}\Big) \\
&> \Big(2+2+(2+m)\Big)\Big)\Big((|A| \vee |B|) \log \tfrac{n}{|A| \vee |B|}\Big) \\
&> 2|A|(\log \tfrac{n}{|A|}) + 2|B|(\log \tfrac{n}{|B|}) + (2+m) \log n \\
&> |A|(1+\log \tfrac{n}{|A|}) + |B|(1+\log \tfrac{n}{|B|}) + (2+m) \log n.
\end{aligned}$$

The first line is due to the definitions of r and c_*. In the third line we use the monotonicity of $x \log \frac{n}{x}$ on $[1, n/e]$ by substituting in $|A|$, $|B|$ and 1 for x. In the fourth line we use that $|A| \vee |B| \leq \frac{n}{e}$ so $\log \frac{n}{|A|}, \log \frac{n}{|B|} > 1$

Exponentiating we get

$$\exp\left[\tfrac{\mu r \log r}{3}\right] \geq \left(\tfrac{en}{|A|}\right)^n \left(\tfrac{en}{|B|}\right)^n n^{2+m+\alpha},$$

for some $\alpha > 0$. It follows that

$$\begin{aligned}
\binom{n}{a}\binom{n}{b} \exp\left(-\tfrac{\mu(a,b,n) r \log r}{3}\right) &\leq \binom{n}{a}\binom{n}{b} \left(\tfrac{en}{a}\right)^{-n} \left(\tfrac{en}{b}\right)^{-n} n^{-2-m-\alpha} \\
&\leq n^{-2-m-\alpha}.
\end{aligned}$$

Putting this together we get

$$\begin{aligned}
\mathbb{P}(D) &\leq \sum_{1 \leq a,b \leq n/e} \binom{n}{a}\binom{n}{b} \exp\left(-\tfrac{\mu(a,b,n) r \log r}{3}\right) + o(n^{-m}) \\
&\leq n^2 n^{-2-m-\alpha} + o(n^{-m}) \\
&= o(n^{-m})
\end{aligned}$$

Thus the lemma is satisfied. □

Proof of Lemma 29. First we set

$$t = \frac{d^{3/2}}{f(d)^2} = \frac{d^{1-2\epsilon}}{(c_3)^2}.$$

Then, we apply the union bound, so

$$\mathbb{P}\left[\ \textbf{m.d.c.}\ \text{fails}\ \right] \leq n\ \mathbb{P}\left[\deg(v) \leq t\right].$$

The distribution of $\deg(v)$ is a binomial distribution with mean d. As $\epsilon > 0$ and hence $t < d$ for n sufficiently large, it is possible to apply Lemma 34

$$\mathbb{P}\left[\deg(v) \leq t\right] \leq \exp\left(-d + t(1 + \log \tfrac{d}{t})\right).$$

Applying this inequality and taking logs we get

$$\begin{aligned} \log \mathbb{P}\left[\ \textbf{m.d.c.}\ \text{fails}\ \right] &\leq \log n - d + t\left[1 + \log \tfrac{d}{t}\right] \\ &\leq \log n - d + \frac{d^{1-2\epsilon}}{(c_3)^2}[1 + \log((c_3)^2 d^{2\epsilon})]. \end{aligned}$$

Note that for $d > 1$, this bound is monotone decreasing in d. Thus we get an upper bound on this quantity by assuming that d is as small as possible, which means

$$d = (k+1)\log n + C(\log n)^{1-2\epsilon} \log\log n.$$

As $\epsilon > 0$ this implies that

$$d < C' \log n \qquad \text{and} \qquad \frac{1}{c_3^2}\left[1 + \log((c_3)^2 d^{2\epsilon})\right] \leq (C')^{2\epsilon} \log\log n$$

for $C' = k + 2$ and all n sufficiently large.

$$\begin{aligned} \log \mathbb{P}\left[\ \textbf{m.d.c.}\ \text{fails}\ \right] &\leq -k\log n - C(\log n)^{1-2\epsilon}\log\log n + [C'\log n]^{1-2\epsilon}((C')^{2\epsilon}\log\log n) \\ &\leq -k\log n - C(\log n)^{1-2\epsilon}\log\log n + C'(\log n)^{1-2\epsilon}(\log\log n) \\ &\leq -k\log n - (C - C')(\log n)^{1-2\epsilon}\log\log n \\ &\leq -k\log n - \omega(1), \end{aligned}$$

where the last line is true for all $C > C'$. Thus

$$\mathbb{P}\left[\ \textbf{m.d.c.}\ \text{fails}\ \right] = o(n^{-k}).$$

□

4.3.3 Light Couple Bounds

To get tail bounds on the contributions of the light couples we first bound their expectation, and then we prove a large deviation result.

Lemma 32. *The expectation over the light couples is at most* $1/f(d)(1+o(1))$*. More precisely,*

$$\left|\mathbb{E}\sum_{(u,v)\in\mathcal{L}} x_u M_{uv} y_v\right| \leq \tfrac{1}{n-1} + \left[1+\tfrac{1}{n-1}\right]\frac{1}{f(d)}.$$

Proof. By symmetry, $\mathbb{E}M_{uv}$ is independent of u and v provided that $u \neq v$. When $u = v$, $M_{uv} \equiv 0$. Define $E = \mathbb{E}M_{uv}$ and define $\mathcal{D} = \{(u,u) \ : \ u \in 1,\ldots,n\}$ so that

$$\mathbb{E}\sum_{(u,v)\in\mathcal{L}} x_u M_{uv} y_v = E \sum_{(u,v)\in\mathcal{L}\setminus\mathcal{D}} x_u y_v.$$

Because x_u sums to 0, the sum over all light couple contributions $x_u y_v$ can be related to the sum over all heavy couple contributions. Specifically,

$$0 = \sum_{(u,v)\in\{1,\ldots,n\}^2} x_u y_v = \sum_{(u,v)\in\mathcal{L}\setminus\mathcal{D}} x_u y_v + \sum_{(u,v)\in\mathcal{H}\setminus\mathcal{D}} x_u y_v + \sum_{(u,v)\in\mathcal{D}} x_u y_v,$$

and thus

$$\left|\sum_{(u,v)\in\mathcal{L}\setminus\mathcal{D}} x_u y_v\right| \leq \left|\sum_{(u,v)\in\mathcal{H}\setminus\mathcal{D}} x_u y_v\right| + \left|\sum_{(u,v)\in\mathcal{D}} x_u y_v\right|.$$

The sum over the diagonal $\mathcal{D}$ is simply the dot product of these vectors and is thus bounded as

$$\left|\sum_{(u,v)\in\mathcal{D}} x_u y_v\right| = |x\cdot y| \leq \|x\|\|y\| \leq 1, \tag{4.9}$$

where we recall that both $\|x\|, \|y\| \leq 1$. The contribution of the heavy couples can be bounded as well

$$\left|\sum_{\{u,v\}\in\mathcal{H}\setminus\mathcal{D}} x_u y_v\right| \le \sum_{\{u,v\}\in\mathcal{H}} |x_u y_v|$$
$$= \sum_{\{u,v\}\in\mathcal{H}} \frac{x_u^2 y_v^2}{|x_u y_v|}$$
$$\le \frac{n}{f(d)} \sum_{\{u,v\}\in\mathcal{H}} x_u^2 y_v^2,$$

where we have used the defining property of heavy couples, and thus we have shown

$$\left|\sum_{\{u,v\}\in\mathcal{H}\setminus\mathcal{D}} x_u y_v\right| \le \frac{n}{f(d)}. \tag{4.10}$$

Combine both bounds (4.9) and (4.10), and it is possible to bound the contribution of the off-diagonal light couples by

$$\left|\sum_{(u,v)\in\mathcal{L}\setminus\mathcal{D}} x_u y_v\right| \le 1 + \frac{n}{f(d)}. \tag{4.11}$$

It remains to establish an upper bound for E to complete the proof. Recall that E can be given by

$$E = \mathbb{E}\left[\frac{\mathbf{1}_{u\leftrightarrow v}}{\sqrt{\deg u}\sqrt{\deg v}}\right]$$

for any fixed distinct vertices u and v. Let $\deg \hat{u} = \deg u - \mathbf{1}_{u\leftrightarrow v}$ and define $\deg \hat{v}$ analogously. Then $\deg \hat{u}$ and $\deg \hat{v}$ and ($\frac{1}{\sqrt{\deg \hat{u}+1}}$ and $\frac{1}{\sqrt{\deg \hat{v}+1}}$) are independent. Computing E,

$$E = \mathbb{P}\left(\mathbf{1}_{u\leftrightarrow v} = 1\right)\mathbb{E}\left[\frac{\mathbf{1}_{u\leftrightarrow v}}{\sqrt{\deg u}\sqrt{\deg v}} \;\middle|\; \mathbf{1}_{u\leftrightarrow v} = 1\right]$$
$$= p\mathbb{E}\left[\frac{1}{\sqrt{\deg \tilde{u}+1}\sqrt{\deg \tilde{v}+1}}\right]$$
$$= p\left(\mathbb{E}\left[\frac{1}{\sqrt{\deg \tilde{u}+1}}\right]\right)^2$$
$$\le p\frac{1}{p(n-1)} = \frac{p}{d},$$

where in the last step, Lemma 33 has been applied, noting $\deg \tilde{u} \sim \text{Binomial}(n-1, p)$. Combining this with (4.11) we get

$$\left| \mathbb{E} \sum_{\{u,v\} \in \mathcal{L}} x_u M_{uv} y_v \right| \leq \frac{p}{d} \left[1 + \frac{n}{f(d)} \right] \leq \frac{1}{n-1} + \left(1 + \frac{1}{n-1} \right) \frac{1}{f(d)}.$$

Note that $f(d) \leq \sqrt{d} \leq \sqrt{n-1}$, and thus the additive error term is of a much smaller order than the $\frac{1}{f(d)}$ contribution. □

Proof of Theorem 30. In executing the concentration of measure inequality, it is helpful if instead of proving the inequality for the sum

$$\sum_{(u,v) \in \mathcal{L}} x_u M_{uv} y_v,$$

we partition this sum into smaller pieces. It is possible to partition the vertices into two ways, $\{U_i\}_{i=1}^4$ and $\{V_i\}_{i=1}^4$, so that each block has cardinality at most $\lceil n/3 \rceil$ and so that x_u has constant sign over each block of $\{U_i\}$ (and likewise y_v has constant sign over each block of $\{V_i\}$).

The large deviation inequality we will establish will bound

$$\mathbb{P}\left[\left| \sum_{(u,v) \in \mathcal{L} \cap U_i \times V_j} x_u M_{uv} y_v - \mathbb{E} x_u M_{uv} y_v \right| > \tfrac{t(2+c_*) f(d)}{\sqrt{d}} \text{ and b.d.c.} \right]$$

for all $n > n_0(t)$. Having established the inequality for every pair of $\{U_i\}$ and $\{V_j\}$, a similar bound can be established for the full sum by applying the union bound and adjusting the constants in the bound.

Note that because $|U_i|, |V_j| \leq \lceil \frac{n}{3} \rceil = r$, there are at least $\frac{n}{4}$ vertices not included in these lists (for $n \geq n_0$ universal). Call this set of vertices $\mathcal{A}$. Assume for the following that all x_u and all y_v are non-negative. Other possibilities can be handled in an identical manner. Enumerate the distinct edges between U_i and V_j as $e_1, e_2, \ldots,$ and note there are at most r^2 such edges. Form the edge-revelation filtration, so that

$\mathscr{F}_k = \mathscr{F}_{k-1} \vee \sigma(e_k)$. According to this filtration, let

$$S_k = \mathbb{E}\left[\sum_{(u,v)\in\tilde{\mathcal{L}}} x_u M_{uv} y_v \middle| \mathscr{F}_k\right]$$

define a martingale, where $\tilde{\mathcal{L}} = \mathcal{L} \cap U_i \times V_j$ and let X_k be the associated martingale increments. Let τ be the stopping time defined by

$$\tau = \inf\left\{k \ : \ \max_{1\le i\le n} \mathbb{E}\left[\deg u_i \ \middle|\mathscr{F}_k\right] \ge (1+c_*)d\right\}.$$

If the bounded degree condition is satisfied, then $\deg u_i \le c_* d$ for all i, and so

$$\text{b.d.c} \implies \mathbb{E}\left[\deg u_i \ \middle|\mathscr{F}_k\right] \le (1+c_*)d \quad \forall\ 0 \le k \le n.$$

Especially, the bounded degree condition implies $\tau = \infty$.

The desired large deviation bound can now be cast in terms of S_k and τ as

$$\begin{aligned}
\mathbb{P}\left[\left|\sum_{\{u,v\}\in\mathcal{L}\cap U_i\times V_j} x_u M_{uv} y_v - \mathbb{E} x_u M_{uv} y_v\right| > t \text{ and b.d.c.}\right]& \\
\le \mathbb{P}\left[\exists\ k \le \tau \text{ so that } |S_k - S_0| \ge t \text{ and } T_n \ge b\right]& \\
\le 2\exp\left(\frac{-t^2}{2(Rt+b)}\right),&
\end{aligned}$$

provided that b satisfies

$$\sum_{k=1}^{\tau} \mathbb{E}\left[X_k^2 \mid \mathscr{F}_{k-1}\right] \le b$$

and R satisfies $|X_k| \le R$ for $k \le \tau$.

This reduces the proof to finding suitable choices of R and b. The starting point for computing either value is the expression for X_k, with $e_k = \{u, v\}$, given by

$$X_k = \mathbf{1}_{(u,v)\in\tilde{\mathcal{L}}}\left[D(u,v) + L(u,v) + U(u,v)\right] + \mathbf{1}_{(v,u)\in\tilde{\mathcal{L}}}\left[D(v,u) + L(v,u) + U(v,u)\right] \tag{4.12}$$

where D, L and U are each handled separately.

Bounding $D(u,v)$

$$D(u,v) = \mathbb{E}\left[\frac{x_u \mathbf{1}_{u\leftrightarrow v} y_v}{\sqrt{\deg u}\sqrt{\deg v}} \;\middle|\; \mathscr{F}_k\right] - \mathbb{E}\left[\frac{x_u \mathbf{1}_{u\leftrightarrow v} y_v}{\sqrt{\deg u}\sqrt{\deg v}} \;\middle|\; \mathscr{F}_{k-1}\right].$$

This can be simplified substantially. Let $\deg \hat{u} = \deg u - \mathbf{1}_{u\leftrightarrow v}$ and let $\deg \hat{v} = \deg v - \mathbf{1}_{u\leftrightarrow v}$. The second term vanishes unless $\mathbf{1}_{u\leftrightarrow v} = 1$, so that

$$\mathbb{E}\left[\frac{x_u \mathbf{1}_{u\leftrightarrow v} y_v}{\sqrt{\deg u}\sqrt{\deg v}} \;\middle|\; \mathscr{F}_{k-1}\right] = \mathbb{E}\left[\frac{p x_u y_v}{\sqrt{\deg \hat{u}+1}\sqrt{\deg \hat{v}+1}} \;\middle|\; \mathscr{F}_k\right].$$

Substituting this into the definition of $D(u,v)$, it is seen that

$$|D(u,v)| \le x_u y_v \left|\mathbf{1}_{u\leftrightarrow v} - p\right| \mathbb{E}\left[\frac{1}{\sqrt{\deg \hat{u}+1}\sqrt{\deg \hat{v}+1}} \;\middle|\mathscr{F}_k\right].$$

These two degrees, $\deg \hat{u}$ and $\deg \hat{v}$ are conditionally independent. Moreover, they can be bounded below by $\deg \hat{u} \ge \deg(u,\mathcal{A})$ and $\deg \hat{v} \ge \deg(v,\mathcal{A})$, both of which are mutually independent of one another and of $\mathscr{F}_k$. From this it follows that

$$|D(u,v)| \le x_u y_v \left|\mathbf{1}_{u\leftrightarrow v} - p\right| \frac{4}{d}, \tag{4.13}$$

where Lemma 33 and $|\mathcal{A}| \le \frac{n}{4}$ has been used.

Bounding $L(u,v)$

$$L(u,v) = \sum_{\substack{w\ne v \\ (u,w)\in\tilde{\mathcal{L}}}} x_u y_w \left[\mathbb{E}\left[\frac{\mathbf{1}_{u\leftrightarrow w}}{\sqrt{\deg u}\sqrt{\deg w}} \;\middle|\; \mathscr{F}_k\right] - \mathbb{E}\left[\frac{\mathbf{1}_{u\leftrightarrow w}}{\sqrt{\deg u}\sqrt{\deg w}} \;\middle|\; \mathscr{F}_{k-1}\right]\right].$$

As before, let $\deg \hat{u} = \deg u - \mathbf{1}_{u\leftrightarrow v}$. Let $\chi \sim \text{Bernoulli}(p)$ be independent of the entire graph. The second conditional expectation may be rewritten as

$$\mathbb{E}\left[\frac{\mathbf{1}_{u\leftrightarrow w}}{\sqrt{\deg u}\sqrt{\deg w}} \;\middle|\; \mathscr{F}_{k-1}\right] = \mathbb{E}\left[\frac{\mathbf{1}_{u\leftrightarrow w}}{\sqrt{\deg \hat{u}+\chi}\sqrt{\deg w}} \;\middle|\; \mathscr{F}_k\right].$$

In this alternate form, the two conditional expectations may be joined and a common denominator formed to yield a single expression

$$\mathbb{E}\left[\frac{\mathbf{1}_{u\leftrightarrow w}(\chi - \mathbf{1}_{u\leftrightarrow v})}{\sqrt{\deg u}\sqrt{\deg w}\sqrt{\deg \hat{u}+\chi}(\sqrt{\deg u}+\sqrt{\deg \hat{u}+\chi})} \;\middle|\; \mathscr{F}_k\right].$$

The bound for this quantity depends entirely on whether $\mathbf{1}_{u\leftrightarrow w}$ is $\mathscr{F}_k$ measurable or not. If it is not measurable, however, then we will condition on it to advance the calculation anyhow. Consider bounding

$$\mathbb{E}\left[\frac{\mathbf{1}_{u\leftrightarrow w}(\chi - \mathbf{1}_{u\leftrightarrow v})}{\sqrt{\deg u}\sqrt{\deg w}\sqrt{\deg \hat{u}+\chi}(\sqrt{\deg u}+\sqrt{\deg \hat{u}+\chi})} \;\middle|\; \mathscr{F}_k \vee \sigma(\mathbf{1}_{u\leftrightarrow w})\right]. \tag{4.14}$$

To bound this from above, bound each of the degree expressions in the denominator from below by either $X_u = \deg(u, \mathcal{A})$ or $X_w = \deg(w, \mathcal{A})$. Using these bounds, expression (4.14) can be bounded from above by

$$\mathbb{E}\left[\frac{\mathbf{1}_{u\leftrightarrow w}\chi}{2\sqrt{1+X_u}^3\sqrt{1+X_w}} \;\middle|\; \mathscr{F}_k \vee \sigma(\mathbf{1}_{u\leftrightarrow w})\right] \leq \frac{8p\mathbf{1}_{u\leftrightarrow w}}{d^2}.$$

To bound (4.14) from below, drop the χ from the numerator. This makes the expression nonpositive, so the analogous bound

$$\mathbb{E}\left[\frac{-\mathbf{1}_{u\leftrightarrow w}\mathbf{1}_{u\leftrightarrow v}}{2\sqrt{1+X_u}^3\sqrt{1+X_w}} \;\middle|\; \mathscr{F}_k \vee \sigma(\mathbf{1}_{u\leftrightarrow w})\right] \geq \frac{-8\mathbf{1}_{u\leftrightarrow v}\mathbf{1}_{u\leftrightarrow w}}{d^2}$$

holds. Combining each of these bounds, it follows that $L(u,v)$ can be bounded by

$$\sum_{\substack{w\neq v\\(u,w)\in\tilde{\mathcal{L}}}} -8\mathbf{1}_{u\leftrightarrow v}x_u y_w \mathbb{E}\left[\mathbf{1}_{u\leftrightarrow w} \;\middle|\; \mathscr{F}_k\right] \leq L(u,v)d^2 \leq \sum_{\substack{w\neq v\\(u,w)\in\tilde{\mathcal{L}}}} 8p x_u y_w \mathbb{E}\left[\mathbf{1}_{u\leftrightarrow w} \;\middle|\; \mathscr{F}_k\right]. \tag{4.15}$$

Bounding $U(u,v)$

$$U(u,v) = \sum_{\substack{w\neq u\\(w,v)\in\tilde{\mathcal{L}}}} x_w y_v \left[\mathbb{E}\left[\frac{\mathbf{1}_{w\leftrightarrow v}}{\sqrt{\deg w}\sqrt{\deg v}} \;\middle|\; \mathscr{F}_k\right] - \mathbb{E}\left[\frac{\mathbf{1}_{w\leftrightarrow v}}{\sqrt{\deg w}\sqrt{\deg v}} \;\middle|\; \mathscr{F}_{k-1}\right]\right].$$

Note the symmetry with $L(u,v)$. The same proof technique shows that

$$\sum_{\substack{w\neq u\\(w,u)\in\tilde{\mathcal{L}}}} -8\mathbf{1}_{w\leftrightarrow v}x_w y_v \mathbb{E}\left[\mathbf{1}_{w\leftrightarrow v} \;\middle|\; \mathscr{F}_k\right] \leq U(u,v)d^2 \leq \sum_{\substack{w\neq v\\(w,v)\in\tilde{\mathcal{L}}}} 8p x_w y_v \mathbb{E}\left[\mathbf{1}_{w\leftrightarrow v} \;\middle|\; \mathscr{F}_k\right].$$

Bounding $|X_k|$, $k \le \tau$

Recall that X_k is given by (4.12). The triangle inequality will be applied and each term bounded separately. Recall that only pairs $(u,v) \in \tilde{\mathcal{L}}$ contribute to X_k, and thus $x_u y_v \le \frac{f(d)}{n}$. With these preliminaries in mind, the simplest bound is of D, for which

$$|D(u,v)| \le x_u y_v \left|\mathbf{1}_{u\leftrightarrow v} - p\right| \frac{4}{d} \le \frac{4f(d)}{nd}.$$

For the bound on L, recall the bound in (4.15), use $x_u y_v \le \frac{f(d)}{n}$ and use $k \le \tau$:

$$|L(u,v)| \le \sum_{\substack{w\neq v \\ (u,w)\in\tilde{\mathcal{L}}}} \frac{8x_u y_w \mathbb{E}\left[\mathbf{1}_{u\leftrightarrow w} \;\middle|\; \mathscr{F}_k\right]}{d^2} \le \frac{8f(d)(1+c_*)d}{nd^2} = \frac{8(1+c_*)f(d)}{nd}.$$

The bound on U is identical, and all bounds may be combined to give

$$|X_k| \le \frac{(40+32c_*)f(d)}{nd}, \qquad\qquad k \le \tau.$$

Bounding $\sum_{k=1}^{\tau} \mathbb{E}\left[X_k^2 \;\middle|\mathscr{F}_{k-1}\right]$

Again, each of D, L and U will be handled separately, and the bound applied to $\mathbb{E}\left[X_k^2 \;\middle|\mathscr{F}_{k-1}\right]$ by virtue of convexity, i.e.

$$\mathbb{E}\left[X_k^2 \;\middle|\mathscr{F}_{k-1}\right] \le 6\left[\mathbb{E}\left[D(u,v)^2 \;\middle|\mathscr{F}_{k-1}\right] + \mathbb{E}\left[L(u,v)^2 \;\middle|\mathscr{F}_{k-1}\right] + \cdots\right].$$

The contribution of $\mathbb{E}\left[D(u,v)^2 \;\middle|\mathscr{F}_{k-1}\right]$ is bounded using (4.13) as

$$\mathbb{E}\left[D(u,v)^2 \;\middle|\mathscr{F}_{k-1}\right] \le \mathbb{E}\left[\frac{16x_u^2 y_v^2}{d^2}\left|\mathbf{1}_{u\leftrightarrow v} - p\right|^2 \;\middle|\mathscr{F}_{k-1}\right] \le \frac{32px_u^2 y_v^2}{d^2}.$$

Summing over all of these bounds,

$$\sum_{k=1}^{\tau} \mathbb{E}\left[D(u,v)^2 + D(v,u)^2 \;\middle|\mathscr{F}_{k-1}\right] \le \sum_{u,v} \frac{32px_u^2 y_v^2}{d^2} \le \frac{32}{nd},$$

where $\{u,v\} = e_k$ and where it has been used that $\|x\|, \|y\| \le 1$.

The contribution of L requires the bounds (4.15) to be simplified somewhat. Symmetrizing the bound (4.15),

$$|L(u,v)| \leq \sum_{\substack{w \neq v \\ (u,w) \in \tilde{\mathcal{L}}}} \frac{8x_u(\mathbf{1}_{u\leftrightarrow v} \vee p) y_w \mathbb{E}\left[\mathbf{1}_{u\leftrightarrow w} \;\middle|\; \mathscr{F}_k\right]}{d^2}.$$

Let $\mathcal{S}$ be the random, $\mathscr{F}_k$-measurable set of light edges emanating from u that are in the history of the process, i.e.

$$\mathcal{S} = \left\{ w \neq v \;\middle|\; 1 = \mathbf{1}_{u\leftrightarrow w} \in \mathscr{F}_k, (u,w) \in \tilde{\mathcal{L}} \right\}.$$

In terms of this random set,

$$|L(u,v)| \leq \frac{8(\mathbf{1}_{u\leftrightarrow v} \vee p)}{d^2} \left[\sum_{w \in \mathcal{S}} x_u y_w + \sum_{w \notin \mathcal{S}, w \neq v} x_u p y_w \right].$$

Thus provided τ has not happened yet, $|\mathcal{S}| \leq (1+c_*)d$, which produces the bound

$$|L(u,v)| \leq \frac{8(\mathbf{1}_{u\leftrightarrow v} \vee p)}{d^2} \left[\frac{f(d)}{n}(1+c_*)d + \frac{f(d)}{n}(n-1)p \right] \leq \frac{8(\mathbf{1}_{u\leftrightarrow v} \vee p)(2+c_*)f(d)}{nd}.$$

Thus, the contribution of $\mathbb{E}\left[L(u,v)^2 \;\middle|\mathscr{F}_{k-1}\right]$ to the sum of variances can be bounded by

$$\sum_{k=1}^{\tau} \mathbb{E}\left[L(u,v)^2 \;\middle|\mathscr{F}_{k-1}\right] \leq \sum_{k=1}^{n^2} \frac{128p(2+c_*)^2 f(d)^2}{n^2 d^2} \leq \frac{128(2+c_*)^2 f(d)^2}{nd}.$$

The bounds for the contributions of $L(v,u), U(u,v)$, and $U(v,u)$ are all identical, so that

$$\sum_{k=1}^{\tau} \mathbb{E}\left[X_k^2 \;\middle|\mathscr{F}_{k-1}\right] \leq \frac{192}{nd} + \frac{3072(2+c_*)^2 f(d)^2}{nd} \leq \frac{3200(2+c_*)^2 f(d)^2}{nd}.$$

Assembling the tail bound.

Having determined bounds for R and b in the Freedman bound:

$$\mathbb{P}\left[\exists\ k \leq \tau \text{ so that } |S_k - S_0| \geq t \text{ and } T_n \geq b\right] \leq 2\exp\left(\frac{-t^2}{2(Rt+b)}\right).$$

We make the scale change $t = \lambda(2 + c_*)\frac{f(d)}{\sqrt{d}}$ so that this bound becomes

$$2 \exp\left(\frac{-n\lambda^2}{6400 + 64\frac{\lambda}{\sqrt{d}}}\right).$$

It remains to establish the bound for the entire contribution of the light couples. Letting $Q_1, \ldots, Q_{16}$ enumerate each of

$$\left|\sum_{(u,v)\in\mathcal{L}\cap U_i\times V_j} x_u M_{uv} y_v - \mathbb{E} x_u M_{uv} y_v\right|$$

over all possible pairs (i, j). Then

$$\begin{aligned}\mathbb{P}\left[Q_1 + \cdots + Q_{16} > \tfrac{16t(2+c_*)f(d)}{\sqrt{d}} \text{ and } b.d.c\right] &\leq \mathbb{P}\left[\max Q_i > \tfrac{t(2+c_*)f(d)}{\sqrt{d}}\right] \\ &\leq 32 \exp\left(\frac{-nt^2}{6400 + 64\frac{t}{\sqrt{d}}}\right).\end{aligned}$$

It remains to include the contribution of the expectation into this inequality. With $f(d) \to \infty$, the expectation is of smaller order than the bound, and so can be ignored almost completely. To be absolutely correct, though, we have that by Lemma 32

$$\left|\mathbb{E}\sum_{(u,v)\in\mathcal{L}} x_u M_{uv} y_v\right| \leq \tfrac{1}{n-1} + \left[1 + \tfrac{1}{n-1}\right]\tfrac{1}{f(d)} \leq 2\tfrac{f(d)}{\sqrt{d}},$$

for n sufficiently large, so that

$$\left|\sum_{(u,v)\in\mathcal{L}} x_u[M_{uv} - \mathbb{E}M_{u,v}]y_v\right| > \tfrac{t(2+c_*)f(d)}{\sqrt{d}}$$

implies

$$\left|\sum_{(u,v)\in\mathcal{L}} x_u M_{uv} y_v\right| > \tfrac{t(1+c_*)f(d)}{\sqrt{d}},$$

for $t \geq 2$. □

4.3.4 Controlling the contribution of the heavy couples.

Proof of Lemma 31. The method of proof here is similar to [42] and [49]. As we have modified the setup slightly from their proofs, we provide a proof of this lemma for completeness. We start by reducing the claim to a claim about the adjacency matrix, by employing the **m.d.c.** Specifically, provided that the **m.d.c.** holds,

$$\sum_{(u,v)\in\mathcal{H}} |x_u M_{u,v} y_v| \le \frac{1}{\min_u \deg(u)} \sum_{(u,v)\in\mathcal{H}} |x_u \mathbf{1}_{u\leftrightarrow v} y_v| \le \frac{f(d)^2}{d^{3/2}} \sum_{(u,v)\in\mathcal{H}} |x_u \mathbf{1}_{u\leftrightarrow v} y_v| \,.$$

Thus, it suffices to show that there is a constant $C = C(c_*)$ so that

$$\sum_{(u,v)\in\mathcal{H}} |x_u \mathbf{1}_{u\leftrightarrow v} y_v| \le C\frac{d}{f(d)}.$$

In the remainder of the proof, we assume that **b.d.c.**, **m.d.c.** and **discrepancy** all hold, and we show how to produce this constant.

We will partition the summands into blocks where each term x_u or y_v has approximately the same magnitude. Thus let $\gamma_i = 2^i\delta$, and put

$$A_i = \left\{ u \;\middle|\; \tfrac{\gamma_{i-1}}{\sqrt{n}} \le |x_u| < \tfrac{\gamma_i}{\sqrt{n}} \right\}, \qquad 1 \le i \le \lceil \log_2 \sqrt{n}/\delta \rceil.$$

$$B_i = \left\{ u \;\middle|\; \tfrac{\gamma_{i-1}}{\sqrt{n}} \le |y_u| < \tfrac{\gamma_i}{\sqrt{n}} \right\}, \qquad 1 \le i \le \lceil \log_2 \sqrt{n}/\delta \rceil.$$

Let $\hat{\mathcal{H}}$ denote those pairs (i,j) so that $\gamma_i\gamma_j \ge f(d)$. The contribution of the absolute sum can, in these terms, be bounded by

$$\sum_{(u,v)\in\mathcal{H}} |x_u \mathbf{1}_{u\leftrightarrow v} y_v| \le \sum_{(i,j)\in\hat{\mathcal{H}}} \frac{\gamma_i\gamma_j}{n} e(A_i, B_j).$$

Let $\lambda_{i,j} = \frac{e(A_i,B_j)}{\mu(A_i,B_j)}$ denote the discrepancy, which can be controlled using Lemma 28. In terms of this quantity, the bound becomes

$$\sum_{(i,j)\in\hat{\mathcal{H}}} \frac{\gamma_i\gamma_j}{n} \lambda_{i,j} |A_i||B_j| \tfrac{d}{n}.$$

In this form, the magnitudes of each of the quantities are somewhat opaque. Consider the sum $\sum_i |A_i|\frac{\gamma_i^2}{n}$; it is at most $4\|x\|^2$. In particular, it is of constant order. Thus let $\alpha_i = |A_i|\frac{\gamma_i^2}{n}$ and $\beta_j = |B_j|\frac{\gamma_j^2}{n}$. This allows the bound to be rewritten as

$$d \sum_{(i,j)\in\hat{\mathcal{H}}} \frac{\gamma_i^2|A_i|}{n}\frac{\gamma_j^2|B_j|}{n}\frac{\lambda_{i,j}}{\gamma_i\gamma_j} = \frac{d}{f(d)} \sum_{(i,j)\in\hat{\mathcal{H}}} \alpha_i\beta_j\frac{\lambda_{i,j}f(d)}{\gamma_i\gamma_j}.$$

This exposes the quantity $\sigma_{i,j} = \frac{\lambda_{i,j}f(d)}{\gamma_i\gamma_j}$ as having some special importance. In effect, either for fixed j, $\sum_j \sigma_{i,j}\beta_j$ has constant order or for fixed i, $\sum_i \sigma_{i,j}\alpha_i$ has constant order.

In what follows, we will bound the contribution of the summands where $|A_i| \geq |B_j|$. By symmetry, the contribution of the other summands will have the same bound. The heavy couples will now be partitioned into 6 classes $\{\hat{\mathcal{H}}_i\}_{i=1}^6$ where their contribution is bounded in a different way. Let $\hat{\mathcal{H}}_i \subseteq \hat{\mathcal{H}}$ be those pairs (i,j) which satisfy the i^{th} property from the following list but none of the prior properties:

1. $\sigma_{i,j} \leq 1$.

2. $\lambda_{i,j} \leq ec_*$.

3. $\gamma_j > f(d)\gamma_i$.

4. $\log\lambda_{i,j} > \frac{1}{4}\left[2\log\gamma_i + \log\frac{1}{\alpha_i}\right]$.

5. $2\log\gamma_i \geq \log\frac{1}{\alpha_i}$.

6. $2\log\gamma_i < \log\frac{1}{\alpha_i}$.

The last properties are better understood when the second case of the discrepancy property is expressed in present notation. In its original form, it states

$$e(A_i, B_j)\log\lambda_{i,j} \leq c_*|A_i|\log\frac{n}{|A_i|}.$$

Substituting γ_i^2/α_i for $n/|A_i|$ and multiplying both sides of this equation through by $\frac{\gamma_i f(d)}{|B_j|\gamma_j d \log \lambda_{i,j}}$ produces the equivalent form

$$\sigma_{i,j}\beta_j \leq c_* \frac{f(d)\gamma_j}{d\gamma_i} \frac{\left[2\log\gamma_i + \log\frac{1}{\alpha_i}\right]}{\log\lambda_{i,j}}.$$

Thus, the last 3 cases cover each of the possible dominant log terms in this bound.

Bounding the contribution of $\hat{\mathcal{H}}_1$.

Quite trivially,

$$\sum_{(i,j)\in\hat{\mathcal{H}}_1} \alpha_i\beta_j\sigma_{i,j} \leq \left(\sum_i \alpha_i\right)\left(\sum_j \beta_j\right) \leq 16.$$

Bounding the contribution of $\hat{\mathcal{H}}_2$.

All the discrepancies in this sum are uniformly bounded. As $\gamma_i\gamma_j \geq f(d)$, it follows that $\sigma_{i,j} \leq ec_*$ and

$$\sum_{(i,j)\in\hat{\mathcal{H}}_2} \alpha_i\beta_j\sigma_{i,j} \leq 16ec_*.$$

Bounding the contribution of $\hat{\mathcal{H}}_3$.

In this case, the magnitudes of y_v dominate those of x_u. However, by bounded degree, there cannot be very many edges connecting to y_v that actualize this contribution, i.e. $e(A_i, B_j) \leq |B_j|c_*d$, so that the discrepancy $\lambda_{i,j}$ is at most $\frac{c_*n}{|A_i|} = \frac{c_*\gamma_i^2}{\alpha_i}$. Fixing a j,

$$\sum_{i\,:\,(i,j)\in\hat{\mathcal{H}}_3} \alpha_i\beta_j\sigma_{i,j} = \sum_{i\,:\,(i,j)\in\hat{\mathcal{H}}_3} \alpha_i\beta_j\frac{\lambda_{i,j}f(d)}{\gamma_i\gamma_j} \leq \sum_{i\,:\,(i,j)\in\hat{\mathcal{H}}_3} \beta_j\frac{c_*\gamma_i f(d)}{\gamma_j} \leq 2c_*\beta_j,$$

where in the last step it has been used that the sum is geometric with leading term less than $\gamma_j/f(d)$. Summing over all β_j shows that

$$\sum_{(i,j)\in\hat{\mathcal{H}}_3} \alpha_i\beta_j\sigma_{i,j} \leq \sum_{j\,:\,(i,j)\in\hat{\mathcal{H}}_3} 2c_*\beta_j \leq 8c_*.$$

Bounding the contribution of $\hat{\mathcal{H}}_4$.

We are not in case (2), and it follows that the second case of the discrepancy property holds. In present notation

$$\sigma_{i,j}\beta_j \leq c_* \frac{f(d)\gamma_j}{d\gamma_i} \frac{\left[2\log\gamma_i + \log\frac{1}{\alpha_i}\right]}{\log\lambda_{i,j}} \leq \frac{4c_*\gamma_j}{\gamma_i f(d)},$$

where $f(d) \leq \sqrt{d}$ and the hypothesis has been used. As we are not in case (3), the sum of these terms is bounded as

$$\sum_{j\,:\,(i,j)\in\hat{\mathcal{H}}_4} \alpha_i\beta_j\sigma_{i,j} \leq \alpha_i 8c_*,$$

where it has been used that the sum has a geometric dominator with leading term at most $\gamma_i f(d)$. Summing over all the i,

$$\sum_{(i,j)\in\hat{\mathcal{H}}_4} \alpha_i\beta_j\sigma_{i,j} \leq \sum_{i\,:\,(i,j)\in\hat{\mathcal{H}}_4} \alpha_i 8c_* \leq 32c_*.$$

Bounding the contribution of $\hat{\mathcal{H}}_5$.

Again, the second case of the discrepancy property holds. Now, in addition,

$$\log\lambda_{i,j} \leq \tfrac{1}{4}\left[2\log\gamma_i + \log\tfrac{1}{\alpha_i}\right] \leq \log\gamma_i,$$

i.e. that $\lambda_{i,j} \leq \gamma_i$. Furthermore, we are not in case (1) so $1 \leq \sigma_{i,j} = \frac{\lambda_{i,j}f(d)}{\gamma_i\gamma_j} \leq \frac{f(d)}{\gamma_j}$. Thus the second discrepancy bound becomes

$$\sigma_{i,j}\beta_j \leq c_* \frac{f(d)\gamma_j}{d\gamma_i} \frac{\left[2\log\gamma_i + \log\frac{1}{\alpha_i}\right]}{\log\lambda_{i,j}} \leq c_* \frac{\gamma_j 4\log\gamma_i}{f(d)\gamma_i(1+\log c_*)} \leq \frac{4c_*}{1+\log c_*}\frac{\gamma_j}{f(d)},$$

where it has been used that $\lambda_{i,j} \geq ec_*$. Fixing an i and summing over j,

$$\sum_{j\,:\,(i,j)\in\hat{\mathcal{H}}_5} \alpha_i\beta_j\sigma_{i,j} \leq \alpha_i \sum_{j\,:\,(i,j)\in\hat{\mathcal{H}}_5} \frac{4c_*}{1+\log c_*}\frac{\gamma_j}{f(d)} \leq \alpha_i\frac{8c_*}{1+\log c_*},$$

where it has been used that the sum is geometric with largest term $f(d)$. Summing over all i,

$$\sum_{(i,j)\in\hat{\mathcal{H}}_5} \alpha_i\beta_j\sigma_{i,j} \leq \sum_{i\,:\,(i,j)\in\hat{\mathcal{H}}_5} \alpha_i\frac{8c_*}{1+\log c_*} \leq \frac{32c_*}{(1+\log c_*)}.$$

Bounding the contribution of $\hat{\mathcal{H}}_6$.

The second case of the discrepancy property holds and in addition,

$$\log \lambda_{i,j} \leq \tfrac{1}{4}\left[2\log\gamma_i + \log\tfrac{1}{\alpha_i}\right] \leq \tfrac{1}{2}\log\tfrac{1}{\alpha_i}.$$

Note that we can drop the $\frac{1}{2}$ because $\lambda_{i,j} > ec_* \geq 1$ and so the logs are positive. This implies that σ satisfies the asymmetric bound $\sigma_{i,j} \leq \frac{1}{\alpha_i}\frac{f(d)}{\gamma_i\gamma_j}$. Thus fixing a j,

$$\sum_{i\,:\,(i,j)\in\hat{\mathcal{H}}_6} \alpha_i\beta_j\sigma_{i,j} \leq \sum_{i\,:\,(i,j)\in\hat{\mathcal{H}}_6} \beta_j\tfrac{f(d)}{\gamma_i\gamma_j} \leq 2\beta_j,$$

where it has been used that the sum is geometric with leading term $\frac{1}{f(d)}$ (which follows as $\gamma_i\gamma_j \geq f(d)$). At last, summing over all the i,

$$\sum_{(i,j)\in\hat{\mathcal{H}}_6} \alpha_i\beta_j\sigma_{i,j} \leq \sum_{j\,:\,(i,j)\in\hat{\mathcal{H}}_6} 2\beta_j \leq 8.$$

Combining the bounds

By summing all of the bounds, we can take C to be

$$C = 16 + 16ec_* + 8c_* + 32c_* + \frac{32c_*}{1+\log c_*}.$$

□

4.4 Estimates of Binomial Random Variables

Lemma 33. *Let* $X \sim \text{Binomial}(n,p)$*, then it follows that*

$$\frac{1}{\sqrt{np+1}} \leq \mathbb{E}\left[\frac{1}{\sqrt{1+X}}\right] \leq \frac{1}{\sqrt{p(n+1)}}.$$

Proof of Lemma 33. The lower bound follows as an immediate consequence of Jensen's inequality. For the upper bound, applying Jensen gives that

$$\mathbb{E}\left[\frac{1}{\sqrt{1+X}}\right] \leq \sqrt{\mathbb{E}\left[\frac{1}{1+X}\right]}.$$

The proof will be completed by verifying the identity

$$\mathbb{E}\left[\frac{1}{X+1}\right] = \frac{1}{p(n+1)}(1-(1-p)^{n+1}),$$

from which the claim follows immediately.

Expand the expectation as a sum

$$\mathbb{E}\left[\frac{1}{X+1}\right] = \sum_{i=0}^{n} \frac{1}{1+i}\binom{n}{i} p^i (1-p)^{n-i}$$

Apply the binomial coefficient identity $\frac{1}{i+1}\binom{n}{i} = \frac{1}{n+1}\binom{n+1}{i+1}$,

$$\begin{aligned}
&= \sum_{i=0}^{n} \frac{1}{n+1}\binom{n+1}{i+1} p^i (1-p)^{n-i} \\
&= \frac{1}{p(n+1)} \sum_{i=0}^{n} \binom{n+1}{i+1} p^{i+1} (1-p)^{(n+1)-(i+1)} \\
&= \frac{1}{p(n+1)} \left(1-(1-p)^{n+1}\right),
\end{aligned}$$

where the last step follows from comparison with the binomial theorem.

□

Lemma 34. *Let X be a binomial random variable with mean μ. Then for any $t \le \mu$*

$$\mathbb{P}\left[X \le t\right] \le \exp\left[-\mu + t(1+\log\tfrac{\mu}{t})\right],$$

Proof of Lemma 34. The proof follows from a standard estimate on the Laplace transform combined with Markov's inequality. For any $\lambda \in \mathbb{R}$, the Laplace transform of $X \sim \text{Binomial}(n,p)$ can be bounded by

$$\begin{aligned}
\mathbb{E}e^{\lambda X} &= \left(pe^{\lambda} + (1-p)\right)^n \\
&= \left(1 + p(e^{\lambda}-1)\right)^n \\
&\le \exp\left[\mu(e^{\lambda}-1)\right].
\end{aligned}$$

Provided that $\lambda < 0$, the tail bound now can be bounded by Markov's inequality by

$$\begin{aligned}\mathbb{P}\left[X \leq t\right] &= \mathbb{P}\left[e^{\lambda X} \geq e^{\lambda t}\right] \\ &\leq \left[\mathbb{E}e^{\lambda X}\right] e^{-\lambda t} \\ &\leq \exp\left[\mu(e^{\lambda} - 1) - \lambda t\right].\end{aligned}$$

Assuming that $t < \mu$, this bound holds with $\lambda = \log(t/\mu)$, which upon evaluation gives

$$\mathbb{P}\left[X \leq t\right] \leq \exp\left[\mu(e^{\log(t/\mu)} - 1) - \log(t/\mu)t\right] = \exp\left[-\mu + t(1 + \log\tfrac{\mu}{t})\right].$$

□

Lemma 35. *Let X be a binomial random variable with mean μ. Then for any $t > 4$*

$$\mathbb{P}\left[X \geq t\mu\right] \leq \exp\left[-\frac{t\mu\log(t)}{3}\right],$$

Proof of Lemma 35. The proof here is identical in approach to the proof of Lemma 34. As there, it is possible to bound the Laplace transform of X as

$$\mathbb{E}e^{\lambda X} \leq \exp\left[\mu(e^{\lambda} - 1)\right],$$

for any real λ. For $\lambda > 0$, the tail bound follows from Markov's inequality by

$$\begin{aligned}\mathbb{P}\left[X \geq t\mu\right] &= \mathbb{P}\left[e^{\lambda X} \geq e^{\lambda t\mu}\right] \\ &\leq \left[\mathbb{E}e^{\lambda X}\right] e^{-\lambda t\mu} \\ &\leq \exp\left[\mu(e^{\lambda} - 1) - \lambda t\mu\right].\end{aligned}$$

For $t > 1$, it is possible to take $\lambda = \log t$. This gives the bound on the tail probability

$$\mathbb{P}\left[X \geq t\mu\right] \leq \exp\left[\mu\left(t - 1 - t\log t\right)\right].$$

To complete the proof, it remains to show that $t - 1 \leq \frac{2}{3}t\log t$ when $t \geq 4$. The function $\frac{t}{t-1}\log t$ is monotonically increasing for $t > 1$, and thus it suffices to show that $\frac{4}{3}\log 4 \geq \frac{3}{2}$, or equivalently that $\log 4 \geq \frac{9}{8}$. This follows from $\log 4 = \int_1^4 \frac{1}{x}dx$ and bounding the integral from below by a right Riemann sum. □

4.5 Simulations for behavior of $\lambda_2(\mathscr{L})$ very near the connectivity threshold

It is tempting to conjecture that the threshold window is arbitrarily small. This is to say that for any $\omega \to \infty$ as $n \to \infty$, we have for all $\epsilon > 0$,

$$p \geq \frac{\log n + \omega}{n} \implies \mathbb{P}\left[\tilde{\lambda}_2 \vee |\tilde{\lambda}_n| > \epsilon\right] \to 0.$$

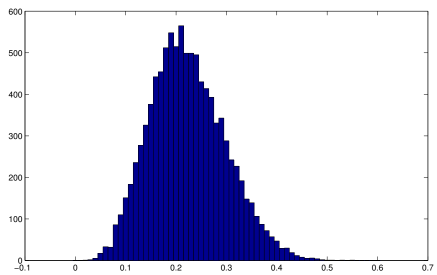
In particular, one wonders if the $\sqrt{\log n}\log\log n$ correction term is needed. To help elucidate this question, we have run some numerical simulations to test this effect.

Unfortunately, the correction is of such a small order that producing an effective simulation by varying the value of p in $G(n,p)$ is unreasonable. For this reason, we run the simulation with the *random graph process* instead. We define this as a discrete time Markov chain $G(t)$ on the set of simple graphs with n vertices, which at time t selects an edge e in $G(t)^c$ uniformly at random and defines $G(t+1) = G(t) \cup \{e\}$. When $np \to \infty$, we heuristically expect that $G([np]) \approx G(n,p)$. There are some exact, general transfer theorems by which statements for one model can be transferred to the other (see [65]).

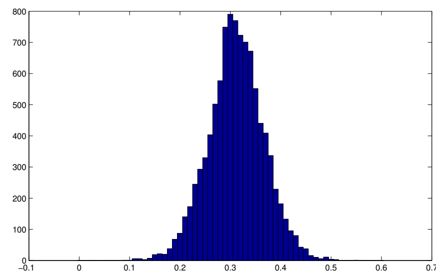
When working with $G(t)$, we can now consider stopping times that correspond to the p thresholds in $G(n,p)$. In particular, the $p \sim \log n/n$ threshold is really a proxy for the connection time of the graph. Thus, consider $\tau^c = \min\{t \mid G(t) \text{ connected }\}$. On the other hand, the $p \sim (\log n + k\log\log n)/n$ threshold corresponds to the first time when all the degrees of the graph are at least k. Thus let $\tau_k^d = \min\{t \mid \deg v \geq k\ \forall v \in \mathcal{V}(G(t))\}$. Note that with high probability $\tau^c \leq \tau_k^d$ for any $k \geq 1$, and in fact $\tau^c = \tau_d^1$ with high probability.

To better understand the n dependence of these variables, we consider sampling $\lambda_2(G(\tau^c))$ once for each value of n between 1000 and 50000 to create series data. The thousand sample moving average and thousand sample standard deviation are presented in Figure 4.5. For comparison, we run the same trend analysis for $\lambda_2(G(\tau_2^d))$.

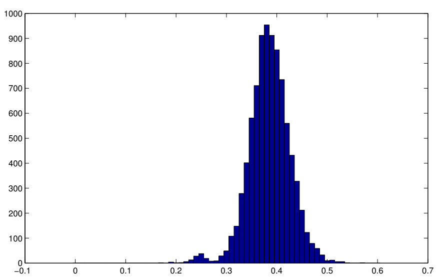
All simulations were run using SciPy and Starcluster on the Amazon EC2 servers.



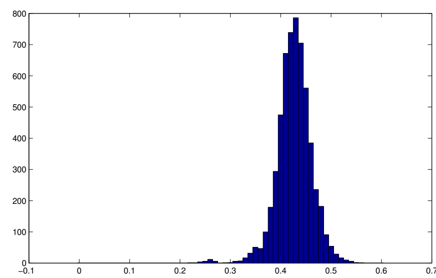
(a) $n = 100$, 10000 samples



(b) $n = 1000$, 10000 samples



(c) $n = 10000$, 8500 samples



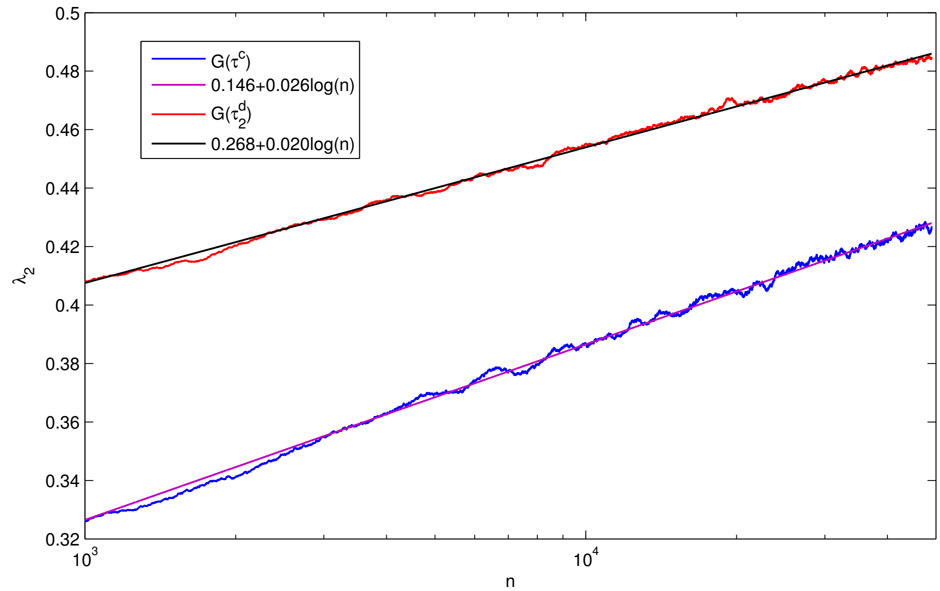
(d) $n = 50000$, 5000 samples

Figure 4.1: Histograms for samplings of λ_2 of $G(\tau^c)$ for different values of n.

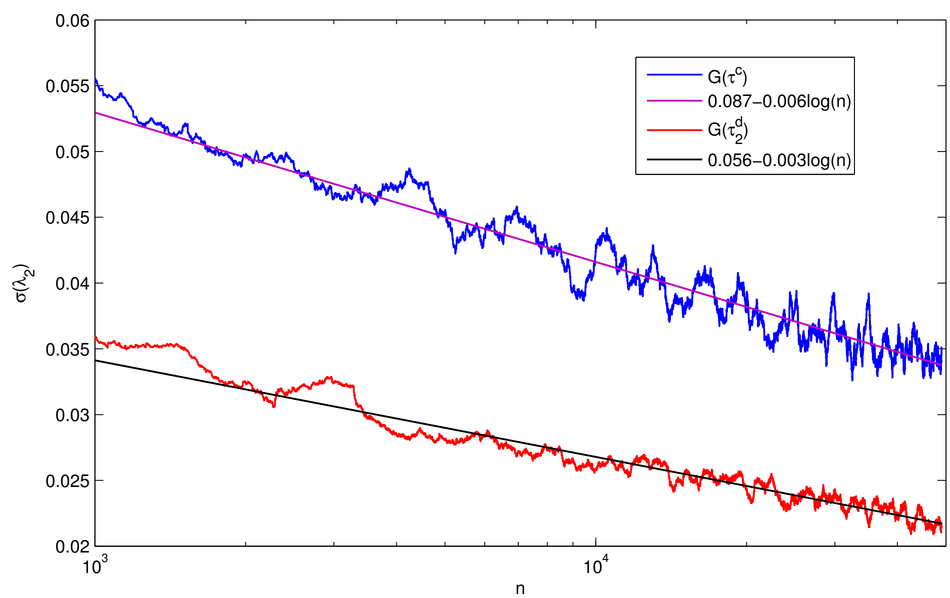
4.5.1 Discussion

From Figure 4.5 and Figure 4.5, one notices that the median value of λ_2 has not yet stabilized by the time n is 50000. Likewise, the standard deviation continues to shrink, leaving open the possibility that λ_2 converges to a constant in almost every realization of the graph process. The experiments show a very strong logarithmic trend in the n dependence of means and standard deviations of all the measured parameters, and they do not reveal the value of any horizontal asymptote. That said, one expects λ_2 to be statistically monotonic in p, as higher connectivity should yield bigger gap. Therefore, it is unlikely that $\mathbb{E}\lambda_2(G(\tau^c))$ is larger than $\mathbb{E}\lambda_2(G(\tau_2^d))$. However, extrapolating these lines, they must either cross, or the trends must change by around $n \approx 10^{20}$.

Preliminary evidence suggests that the trends could flatten sooner (see Figure 4.5.1). A small sample of $n = 500000$ provides some statistical evidence (with $p \approx 0.91$) that the trend overestimates $\lambda_2(G(\tau^c))$.



(a) Moving average of λ_2 as n grows.



(b) Moving average of $\sigma(\lambda_2)$ as n grows.

Figure 4.2: Independent samplings of λ_2 of $G(\tau^c)$ and $G(\tau_2^d)$ for different values of n as n varies between 1000 and 50000. These data were then averaged over a window of 1000 contiguous samples to produce the moving average λ_2 and the moving average $\sigma(\lambda_2)$.

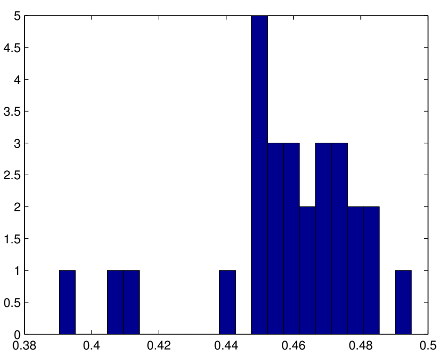


Figure 4.3: Histogram for 28 samplings of λ_2 of $G(\tau^c)$ with $n = 500000$. The sample mean is 0.4575, compared with the trend of 0.4886.

BIBLIOGRAPHY

[1] Milton Abramowitz and Irene A. Stegun, editors. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. A Wiley-Interscience Publication. John Wiley & Sons Inc., New York, 1984. Reprint of the 1972 edition, Selected Government Publications.

[2] Gernot Akemann, Jinho Baik, and Philippe Di Francesco, editors. *The Oxford handbook of random matrix theory*. Oxford University Press, Oxford, 2011.

[3] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[4] Greg W. Anderson and Ofer Zeitouni. A CLT for a band matrix model. *Probab. Theory Related Fields*, 134(2):283–338, 2006.

[5] Nicola Arcozzi, Richard Rochberg, Eric T. Sawyer, and Brett D. Wick. The Dirichlet space: a survey. *New York J. Math.*, 17A:45–86, 2011.

[6] Gérard Ben Arous and Kim Dang. On fluctuations of eigenvalues of random permutation matrices. 2011.

[7] Z. D. Bai and Jack W. Silverstein. CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.*, 32(1A):553–605, 2004.

[8] Zhidong Bai and Jack W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010.

[9] Dominique Bakry. Functional inequalities for Markov semigroups. In *Probability measures on groups: recent directions and trends*, pages 91–147. Tata Inst. Fund. Res., Mumbai, 2006.

[10] W. Ballmann and J. Świątkowski. On L^2-cohomology and property (T) for automorphism groups of polyhedral cell complexes. *Geom. Funct. Anal.*, 7(4):615–645, 1997.

[11] Ya. M. Barzdin and A. N. Kolmogorov. *Selected works of A. N. Kolmogorov. Vol. III*, volume 27 of *Mathematics and its Applications (Soviet Series)*, pages 194–202. Kluwer Academic Publishers Group, Dordrecht, 1993. English translation.

[12] Bachir Bekka, Pierre de la Harpe, and Alain Valette. *Kazhdan's property (T)*, volume 11 of *New Mathematical Monographs*. Cambridge University Press, Cambridge, 2008.

[13] Gordon Blower. *Random matrices: high dimensional phenomena*, volume 367 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2009.

[14] S. G. Bobkov. Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab.*, 27(4):1903–1921, 1999.

[15] S. G. Bobkov and M. Ledoux. From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities. *Geom. Funct. Anal.*, 10(5):1028–1052, 2000.

[16] Oriol Bohigas, Rizwan U. Haq, and Akhilesh Pandey. Higher-order correlations in spectra of complex systems. *Phys. Rev. Lett.*, 54:1645–1648, Apr 1985.

[17] B. Bollobás. *Random Graphs, Second Edition*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2001.

[18] A. Borodin. CLT for spectra of submatrices of Wigner random matrices. *ArXiv e-prints*, October 2010.

[19] A. A. Borovkov and S. A. Utev. An inequality and a characterization of the normal distribution connected with it. *Teor. Veroyatnost. i Primenen.*, 28(2):209–218, 1983.

[20] P. Bourgade. Bulk universality for one-dimensional log-gases. In A. Jensen, editor, *Xviith International Congress on Mathematical Physics*. WORLD SCIENTIFIC Publishing Company, 2013.

[21] A. Broder and E. Shamir. On the second eigenvalue of random regular graphs. In *28th Annual Symposium on Foundations of Computer Science (Los Angeles, 1987)*, pages 286–294. IEEE Comput. Soc. Press, Washington, D.C., 1987.

[22] Andrei Z. Broder, Alan M. Frieze, Stephen Suen, and Eli Upfal. Optimal construction of edge-disjoint paths in random graphs. *SIAM J. Comput.*, 28(2):541–573 (electronic), 1999.

[23] Thierry Cabanal-Duvillard. Fluctuations de la loi empirique de grandes matrices aléatoires. *Ann. Inst. H. Poincaré Probab. Statist.*, 37(3):373–402, 2001.

[24] M. Capitaine and M. Casalis. Asymptotic freeness by generalized moments for Gaussian and Wishart matrices. Application to beta random matrices. *Indiana Univ. Math. J.*, 53(2):397–431, 2004.

[25] S.-Y. A. Chang and D. E. Marshall. On a sharp inequality concerning the Dirichlet integral. *Amer. J. Math.*, 107(5):1015–1033, 1985.

[26] Sourav Chatterjee. Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields*, 143(1-2):1–40, 2009.

[27] Fan Chung, Linyuan Lu, and Van Vu. The spectra of random graphs with given expected degrees. *Internet Math.*, 1(3):257–275, 2004.

[28] Fan R. K. Chung. *Spectral graph theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1997.

[29] Amin Coja-Oghlan. On the Laplacian eigenvalues of $G_{n,p}$. *Combin. Probab. Comput.*, 16(6):923–946, 2007.

[30] Benoît Collins. Product of random projections, Jacobi ensembles and universality problems arising from free probability. *Probab. Theory Related Fields*, 133(3):315–344, 2005.

[31] Ovidiu Costin and Joel L. Lebowitz. Gaussian fluctuation in random matrices. *Phys. Rev. Lett.*, 75:69–72, Jul 1995.

[32] Michael Davis and Matthew Kahle. Random graph products of finite groups are rational duality groups. submitted, 1210.4577), 2012.

[33] Ioana Dumitriu. *Eigenvalue Statistics for Beta-Ensembles*. PhD thesis, MIT, 2003.

[34] Ioana Dumitriu and Alan Edelman. Global spectrum fluctuations for the β-Hermite and β-Laguerre ensembles via matrix models. *J. Math. Phys.*, 47(6):063302, 36, 2006.

[35] Ioana Dumitriu, Tobias Johnson, Soumik Pal, and Elliot Paquette. Functional limit theorems for random regular graphs. *Probability Theory and Related Fields*, 156(3-4):921–975, 2013.

[36] Freeman J. Dyson. Statistical theory of the energy levels of complex systems. I. *J. Mathematical Phys.*, 3:140–156, 1962.

[37] E. N. Economou and C. M. Soukoulis. Static conductance and scaling theory of localization in one dimension. *Phys. Rev. Lett.*, 46:618–621, Mar 1981.

[38] Alan Edelman and Brian D. Sutton. The beta-Jacobi matrix model, the CS decomposition, and generalized singular value problems. *Found. Comput. Math.*, 8(2):259–285, 2008.

[39] L. Erdos. Universality of Wigner random matrices: a Survey of Recent Results. *ArXiv e-prints*, April 2010.

[40] L. Erdos, A. Knowles, H.-T. Yau, and J. Yin. Spectral Statistics of Erd\H{o}s-R\'enyi Graphs I: Local Semicircle Law. *ArXiv e-prints*, March 2011.

[41] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[42] U. Feige and E. Ofek. Spectral techniques applied to sparse random graphs. *Random Structures Algorithms*, 27(2):251–275, 2005.

[43] P. J. Forrester. Quantum conductance problems and the Jacobi ensemble. *J. Phys. A*, 39(22):6861–6870, 2006.

[44] P. J. Forrester. *Log-gases and random matrices*, volume 34 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2010.

[45] Jacob Fox, Mikhail Gromov, Vincent Lafforgue, Assaf Naor, and János Pach. Overlap properties of geometric expanders. *J. Reine Angew. Math.*, 671:49–83, 2012.

[46] D. A. Freedman. On tail probabilities for martingales. *Ann. Probab.*, 3:100–118, 1975.

[47] Ehud Friedgut and Gil Kalai. Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.*, 124(10):2993–3002, 1996.

[48] J. Friedman. On the second eigenvalue and random walks in random d-regular graphs. *Combinatorica*, 11:331–362, 1991.

[49] J. Friedman, J. Kahn, and E. Szemerédi. On the second eigenvalue of random regular graphs. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, STOC '89, pages 587–598, New York, NY, USA, 1989. ACM.

[50] Joel Friedman. A proof of Alon's second eigenvalue conjecture. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, pages 720–724 (electronic), New York, 2003. ACM.

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

[52] Howard Garland. p-adic curvature and the cohomology of discrete subgroups of p-adic groups. *Ann. of Math. (2)*, 97:375–423, 1973.

[53] Nicolas Garrel. La propriété (T) de Kazhdan et les immeubles triangulaires. Master's thesis, École normale supérieure, Paris, 2012.

[54] Michel Gaudin. Sur la loi limite de l'espacement des valeurs propres d'une matrice aleÂť atoire. *Nuclear Physics*, 25(0):447 – 458, 1961.

[55] Eli Glasner. *Ergodic theory via joinings*, volume 101 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2003.

[56] Loukas Grafakos. *Classical Fourier analysis*, volume 249 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2008.

[57] C. Greenhill, S. Janson, J. H. Kim, and N. C. Wormald. Permutation pseudographs and contiguity. *Combin. Probab. Comput.*, 11(3):273–298, 2002.

[58] Mikhail Gromov. Singularities, expanders and topology of maps. Part 2: From combinatorics to topology via algebraic isoperimetry. *Geom. Funct. Anal.*, 20(2):416–526, 2010.

[59] Alice Guionnet. Large deviations upper bounds and central limit theorems for non-commutative functionals of Gaussian large random matrices. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(3):341–384, 2002.

[60] Alice Guionnet and Ofer Zeitouni. Addendum to: "Large deviations asymptotics for spherical integrals". *J. Funct. Anal.*, 216(1):230–241, 2004.

[61] Jonas Gustavsson. Gaussian fluctuations of eigenvalues in the GUE. *Ann. Inst. H. Poincaré Probab. Statist.*, 41(2):151–178, 2005.

[62] Allen Hatcher. *Algebraic topology.* Cambridge University Press, Cambridge, 2002.

[63] C. Hoffman, M. Kahle, and E. Paquette. A sharp threshold for Kazhdan's property (T). *ArXiv e-prints*, January 2012.

[64] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561 (electronic), 2006.

[65] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[66] Tiefeng Jiang. Approximation of Haar distributed matrices and limiting distributions of eigenvalues of Jacobi ensembles. *Probab. Theory Related Fields*, 144(1-2):221–246, 2009.

[67] Kurt Johansson. On random matrices from the compact classical groups. *Ann. of Math. (2)*, 145(3):519–545, 1997.

[68] Kurt Johansson. On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.*, 91(1):151–204, 1998.

[69] Dag Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.*, 12(1):1–38, 1982.

[70] Kevin W. J. Kadell. The Selberg-Jack symmetric functions. *Adv. Math.*, 130(1):33–102, 1997.

[71] Matthew Kahle. Sharp vanishing thresholds for cohomology of random flag complexes. submitted, arXiv:1207.0149, 2012.

[72] Matthew Kahle and Boris Pittel. Inside the critical window for cohomology of random k-complexes. submitted, arXiv:1301.1324, 2013.

[73] Rowan Killip. Gaussian fluctuations for β ensembles. *Int. Math. Res. Not. IMRN*, (8):Art. ID rnn007, 19, 2008.

[74] Rowan Killip and Irina Nenciu. Matrix models for circular ensembles. *Int. Math. Res. Not.*, (50):2665–2701, 2004.

[75] P.E. Koehler, F. BeÄŊvÃąÅŹ, M. KrtiÄŊka, K.H. Guber, and J.L. Ullmann. Neutron resonance data exclude random matrix theory. *Fortschritte der Physik*, 61(2-3):80–94, 2013.

[76] Timothy Kusalik, James A. Mingo, and Roland Speicher. Orthogonal polynomials and fluctuations of random matrices. *J. Reine Angew. Math.*, 604:1–46, 2007.

[77] Michel Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[78] W.-C. W. Li. Ramanujan hypergraphs. *Geom. Funct. Anal.*, 14(2):380–399, 2004.

[79] Nathan Linial and Roy Meshulam. Homological connectivity of random 2-complexes. *Combinatorica*, 26(4):475–487, 2006.

[80] H. I. Liou, H. S. Camarda, S. Wynchank, M. Slagowitz, G. Hacken, F. Rahn, and J. Rainwater. Neutron-resonance spectroscopy. viii. the separated isotopes of erbium: Evidence for dyson‘s theory concerning level spacings. *Physical Review, Part C, Nuclear Physics*, 5:974, 1972.

[81] E. Lubetzky, B. Sudakov, and V. Vu. Spectra of lifted Ramanujan graphs. *Adv. Math.*, 227(4):1612–1645, 2011.

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

[83] A. Lubotzky, B. Samuels, and U. Vishne. Explicit constructions of Ramanujan complexes of type. *European Journal of Combinatorics*, 26(6):965–993, 2005.

[84] A. Lubotzky, B. Samuels, and U. Vishne. Ramanujan complexes of typeÃ d. *Israel Journal of Mathematics*, 149(1):267–299, 2005.

[85] I. G. Macdonald. *Symmetric functions and Hall polynomials*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, second edition, 1995. With contributions by A. Zelevinsky, Oxford Science Publications.

[86] R. L. Macklin and C. W. Alexander. Neutron absorption cross section of ^{236}u. *Nuclear Science and Engineering*, 104:258, 1990.

[87] G. A. Margulis. Explicit constructions of expanders. *Problemy Peredači Informacii*, 9(4):71–80, 1973.

[88] Donald E. Marshall. A new proof of a sharp inequality concerning the Dirichlet integral. *Ark. Mat.*, 27(1):131–137, 1989.

[89] B. D. McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra Appl.*, 40:203–216, 1981.

[90] B. D. McKay, N. C. Wormald, and B. Wysocka. Short cycles in random regular graphs. *Electron. J. Combin.*, 11, 2004.

[91] M. L. Mehta. On the statistical properties of the level-spacings in nuclear spectra. *Nuclear Phys.*, 18:395–419, 1960.

[92] Madan Lal Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.

[93] R. Meshulam and N. Wallach. Homological connectivity of random k-dimensional complexes. *Random Structures Algorithms*, 34(3):408–417, 2009.

[94] James A. Mingo and Alexandru Nica. Annular noncrossing permutations and partitions, and second-order asymptotics for random matrices. *Int. Math. Res. Not.*, (28):1413–1460, 2004.

[95] Robb J. Muirhead. *Aspects of multivariate statistical theory*. John Wiley & Sons Inc., New York, 1982. Wiley Series in Probability and Mathematical Statistics.

[96] D. K. Olsen. Resolved resonance parameters for ^{238}u from 1 to 10 kev. *Nuclear Science and Engineering*, 94:102, 1986.

[97] Sean O'Rourke. Gaussian fluctuations of eigenvalues in Wigner random matrices. *J. Stat. Phys.*, 138(6):1045–1066, 2010.

[98] ELLIOT PAQUETTE and IOANA DUMITRIU. Global fluctuations for linear statistics of SS-jacobi ensembles. *Random Matrices: Theory and Applications*, 01(04):1250013, 2012.

[99] L. Pastur. Limiting laws of linear eigenvalue statistics for Hermitian matrix models. *J. Math. Phys.*, 47(10):103303, 22, 2006.

[100] Sandrine Péché and Alexander Soshnikov. Wigner random matrices with non-symmetrically distributed entries. *J. Stat. Phys.*, 129(5-6):857–884, 2007.

[101] M. Pinsker. On the complexity of a concentrator. *7th International Teletraffic Conference, Stockholm*, pages 318/1–318/4, 1973.

[102] Nathan Ross. Fundamentals of Stein's method. *Probab. Surv.*, 8:210–293, 2011.

[103] William T. Ross. The classical Dirichlet space. In *Recent advances in operator-related function theory*, volume 393 of *Contemp. Math.*, pages 171–197. Amer. Math. Soc., Providence, RI, 2006.

[104] Atle Selberg. Remarks on a multiple integral. *Norsk Mat. Tidsskr.*, 26:71–78, 1944.

[105] M. Shcherbina. Central limit theorem for linear eigenvalue statistics of orthogonally invariant matrix models. *Zh. Mat. Fiz. Anal. Geom.*, 4(1):171–195, 204, 2008.

[106] M. Shcherbina. Central limit theorem for linear eigenvalue statistics of the Wigner and sample covariance random matrices. *Zh. Mat. Fiz. Anal. Geom.*, 7(2):176–192, 197, 199, 2011.

[107] J. F. Shriner, Jr, H. A. Weidenmüller, and G. E. Mitchell. Data on Neutron Widths do not refute Random–Matrix Theory. *ArXiv e-prints*, September 2012.

[108] Ya. Sinai and A. Soshnikov. Central limit theorem for traces of large random symmetric matrices with independent matrix elements. *Bol. Soc. Brasil. Mat. (N.S.)*, 29(1):1–24, 1998.

[109] Alexander Soshnikov. Universality at the edge of the spectrum in Wigner random matrices. *Comm. Math. Phys.*, 207(3):697–733, 1999.

[110] Richard P. Stanley. Some combinatorial properties of Jack symmetric functions. *Adv. Math.*, 77(1):76–115, 1989.

[111] Elias M. Stein. *Singular integrals and differentiability properties of functions*. Princeton Mathematical Series, No. 30. Princeton University Press, Princeton, N.J., 1970.

[112] V. E. Stepanov. Combinatorial algebra and random graphs. *Teor. Verojatnost. i Primenen*, 14:393–420, 1969.

[113] Brian Sutton. *The Stochastic Operator Approach to Random Matrix Theory.* PhD thesis, MIT, 2005.

[114] Gábor Szegő. *Orthogonal polynomials.* American Mathematical Society, Providence, R.I., fourth edition, 1975. American Mathematical Society, Colloquium Publications, Vol. XXIII.

[115] M. Talagrand. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, 6(3):587–600, 1996.

[116] T. Tao and V. Vu. The Wigner-Dyson-Mehta bulk universality conjecture for Wigner matrices. *ArXiv e-prints*, January 2011.

[117] L. Tran, V. Vu, and K. Wang. Sparse random graphs: Eigenvalues and eigenvectors. 2010. Preprint. Available at arXiv:1011.6646v1.

[118] Hale F. Trotter. Eigenvalue distributions of large Hermitian matrices; Wigner's semicircle law and a theorem of Kac, Murdock, and Szegő. *Adv. in Math.*, 54(1):67–82, 1984.

[119] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[120] Van H. Vu. Spectral norm of random matrices. *Combinatorica*, 27(6):721–736, 2007.

[121] Sean Washburn and Richard A. Webb. Aharonov-bohm effect in normal metal quantum coherence and transport. *Advances in Physics*, 35(4):375–422, 1986.

[122] K. Wieand. Eigenvalue distributions of random permutation matrices. *The Annals of Probability*, 28(4):1563–1587, 2000.

[123] E. Wigner. *The probability of the existence of a self-reproducing unit.* Free Press, 1961.

[124] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)*, 62:548–564, 1955.

[125] Eugene P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)*, 67:325–327, 1958.

[126] John Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A(1/2):pp. 32–52, 1928.

[127] N. C. Wormald. Models of random regular graphs. In *Surveys in combinatorics, 1999 (Canterbury)*, volume 267 of *London Math. Soc. Lecture Note Ser.*, pages 239–298. Cambridge Univ. Press, Cambridge, 1999.

[128] A. Żuk. Property (T) and Kazhdan constants for discrete groups. *Geom. Funct. Anal.*, 13(3):643–670, 2003.

[129] Andrzej Żuk. La propriété (T) de Kazhdan pour les groupes agissant sur les polyèdres. *C. R. Acad. Sci. Paris Sér. I Math.*, 323(5):453–458, 1996.