Eigenvalue Statistics for Beta-Ensembles

by

Ioana Dumitriu

B.A., New York University, 1999

Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2003

© Ioana Dumitriu, MMIII. All rights reserved.

The author hereby grants to MIT permission to reproduce and
distribute publicly paper and electronic copies of this thesis document
in whole or in part.

Author .....
Department of Mathematics
April 29, 2003

Certified by .....
Alan Edelman
Professor of Applied Mathematics
Thesis Supervisor

Accepted by .....
Rodolfo R. Rosales
Chairman, Applied Mathematics Committee

Accepted by .....
Pavel I. Etingof
Chairman, Department Committee on Graduate Students

Eigenvalue Statistics for Beta-Ensembles

by

Ioana Dumitriu

Submitted to the Department of Mathematics
on April 29, 2003, in partial fulfillment of the
requirements for the degree of
DOCTOR OF PHILOSOPHY

Abstract

Random matrix theory is a maturing discipline with decades of research in multiple fields now beginning to converge. Experience has shown that many exact formulas are available for certain matrices with real, complex, or quaternion entries. In random matrix jargon, these are the cases $\beta = 1, 2$ and 4 respectively. This thesis explores the general $\beta > 0$ case mathematically and with symbolic software.

We focus on generalizations of the Hermite distributions originating in physics (the "Gaussian" ensembles) and the Laguerre distributions of statistics (the "Wishart" matrices). One of our main contributions is the construction of tridiagonal matrix models for the general ($\beta > 0$) β-Hermite and ($\beta > 0, a > \beta(m-1)/2$) β-Laguerre ensembles of parameter a and size m, and investigate applications of these new ensembles, particularly in the areas of eigenvalue statistics.

The new models are symmetric tridiagonal, and with entries from *real* distributions, regardless of the value of β. The entry distributions are either normal or χ, so "classical", and the independence pattern is maximal, in the sense that the only constraints arise from the symmetric/semi-definite condition.

The β-ensemble distributions have been studied for the particular $1, 2, 4$ values of β as joint eigenvalue densities for full random matrix ensembles (Gaussian, or Hermite, and Wishart, or Laguerre) with real, complex, and quaternion entries (for references, see [66] and [70]). In addition, general β-ensembles were considered and studied as theoretical distributions ([8, 51, 50, 55, 56]), with applications in lattice gas theory and statistical mechanics (the β parameter being interpreted as an arbitrary inverse temperature of a Coulomb gas with logarithmic potential). Certain eigenvalue statistics over these general β-ensembles, namely those expressible in terms of integrals of symmetric polynomials with corresponding Hermite or Laguerre weights, can be computed in terms of multivariate orthogonal polynomials (Hermite or Laguerre).

We have written a Maple Library (MOPs: *Multivariate Orthogonal Polynomials symbolically*) which implements some new and some known algorithms for computing the Jack, Hermite, Laguerre, and Jacobi multivariate polynomials for arbitrary β. This library can be used as a tool for conjecture-formulation and testing, for statistical computations, or simply for getting acquainted with the mathematical concepts.

Some of the figures in this thesis have been obtained using MOPs.

Using the new β-ensemble models, we have been able to provide a unified perspective of the previously isolated $1, 2$, and 4 cases, and prove generalizations for some of the known eigenvalue statistics to arbitrary β. We have rediscovered (in the Hermite case) a strong version of the Wigner Law (semi-circle), and proved (in the Laguerre case) a strong version of the similar law (generalized quarter-circle). We have obtained first-order perturbation theory for the β large case, and we have reason to believe that the tridiagonal models in the large n (ensemble size) limit will also provide a link between the largest eigenvalue distributions for both Hermite and Laguerre for arbitrary β (for $\beta = 1, 2$, this link was proved to exist by Johannson [52] and Johnstone [53]). We also believe that the tridiagonal Hermite models will provide a link between the largest eigenvalue distribution for different values of β (in particular, between the Tracy-Widom [91] distributions for $\beta = 1, 2, 4$).

Thesis Supervisor: Alan Edelman
Title: Professor of Applied Mathematics

Acknowledgments

First and foremost, I want to thank my advisor, Alan Edelman. There are SO many things to say here... but no matter how much I said, I wouldn't feel I have said enough. So I will only say this: Alan has made me understand the concept of owing someone so much, that the only way you can repay your debt is by emulating that person. Thank you, Alan.

To Joel Spencer, who has been a wise friend and a great mentor each and every time I needed either one: thank you, Joel. Thank you-thank you-thank you.

To Michael Overton, who likes to take credit for introducing me to Numerical Linear Algebra and for sending me to MIT to work with Alan: you are absolutely right, Mike, it is completely your fault. I hope you're happy!... =)

I would like to thank Peter Winkler for many things, but especially for the phrase "If Linear Algebra doesn't work out, you should think about Combinatorics. You *are* good enough to do Combinatorics. Keep it in mind." Thank you, Pete.

Thank you, David Jackson, for the time you have taken to read this thesis so carefully and for the many valuable comments. Thank you for going out of your way to attend my thesis defense.

Many thanks to Gil Strang and Dan Spielman for comments and suggestions that improved this thesis, and to Richard Stanley, for always taking the time to listen and to provide answers to my questions.

Etienne – thank you for all the help with the "Dyck path" identities. "Ici nous prouvons des choses étranges et merveilleuses..."

I would like to thank Joe Gallian, for a wonderful summer I spent in Duluth at his REU. "Hey! Be nice."

Thank you, Dan, for being the first person to tell me that not only *could* I do more in mathematics, but that I *absolutely should* do more. Guess you were right. =)

And...

Thank you, Marsha Berger, for teaching me more than **C**.

Thank you, Cathy O'Neil, for always making me feel good with a word.

Thank you, Carly, for your warmth, for your optimism, and for your shoulder.

Tamar and Anda — THANK YOU. Without you, this never would have happened.

And one more thank you, the last one, and one of the biggest: thank you, Mom, for never letting me set the bar below my reach.

This thesis is dedicated to all the women in my life.

Contents

List of Figures

List of Tables

Chapter 1

Foreword

For a long time, the random matrix community has seen the cases $\beta = 1, 2$ and sometimes 4 as special – quite often, as the only interesting cases – in the study of β-Hermite (Gaussian), β-Laguerre (Wishart), and β-Jacobi (MANOVA) ensembles. One reason for this discrimination is that interesting matrix models were known only for these cases. Another reason is that for these three particular cases, there is a large variety of available methods of approach (which are parameter-specific, even among the $\{1, 2, 4\}$ values). One may even argue that the existence of these three different sets of methods is what created the impression that we are dealing with a "discrete" set of ensembles.

We show that "continuous" $\beta > 0$ is not "exotic", indeed, that it is merely a general case for which methods of approach can be devised, based on the matrix models. While the cases $\beta = 1, 2$, and 4 have some properties which may indeed be seen as special, we shall advance the idea that there exists a unified, general-β theory from which these cases stem, a theory which we are only starting to discover.

Chapter 2

The β-ensembles

2.1 Introduction

For close to a century now, researchers in all branches of science have used randomization as a tool for modeling large-scale systems. From studying populations to studying atomic interactions, randomization has been used whenever the size and complexity of the problem made a systematic, deterministic approach virtually impossible.

The study of random matrices emerged in the late 1920's (with the publishing of Wishart's most important work in 1928) and 1930's (with Hsu's work [41]), and they are a very quickly growing field of research, with communities like nuclear physics [23, 33, 94], multivariate statistics [14, 47, 70], algebraic and enumerative combinatorics [20, 36, 39], algorithms [84] and random graph theory [65], numerical analysis [2, 17, 26], and of late, computational biology and genomics [29], and wireless communications [93, 98].

Sometimes, different threads in this intricate web come together to produce very surprising connections. Such a result is the numerical evidence concerning the zeros of the Riemann zeta function on the $x = 1/2$ line; it was conjectured that $\zeta(\frac{1}{2} + i\lambda)$ has real roots because they correspond to a Hermitian operator. Much progress has been made in investigating the spacings of the zeros of the ζ function by Montgomery [69], and relating them to the spacings of a random Hermitian matrix (GUE). A tremendous computational effort by Odlyzko [72] provided enough numerical evi-

dence to bolster Montgomery's conjecture; this phenomenon is now known as the Montgomery-Odlyzko law.

Another point of confluence for the different threads of random matrix theory is the Baik-Deift-Johansson [7] proof that one of the important problems in combinatorics, the distribution of the longest increasing subsequence of a random permutation of length n is the same as the distribution of the largest eigenvalue of an $n \times n$ GUE matrix. The asymptotic value of this distribution, as $n \to \infty$, is also known as the Tracy-Widom law for the GUE, since Tracy and Widom were the first to compute it [91], and has since been identified as a limiting distribution for other combinatorics problems.

Of the many possible "nice", "classical" distributions that the entries of a random matrix could have, one model is prevalent in the literature: the normal distribution. This model is preferred not only because of its simplicity, but also (or mostly) because of two very important properties that distinguish it: invariance with respect to orthogonal transformations, and the Central Limit Theorem (which makes it a "natural" choice).

Many of the matrix models have thus standard normal entries, which are independent up to a symmetry/positive definiteness condition (since the spectrum, whether it corresponds to the energy levels of a Schrödinger operator or to the components of a sample covariance matrix, is real). There are two approaches to symmetrizing the matrix of standard, independent normal entries: in physics, the symmetric part of the matrix $(A + A^T)/2$ is extracted, and the resulting symmetric random matrix is defined to be from a *Gaussian* ensemble. In statistics, the matrix is symmetrized by multiplying by the transpose, AA^T, and the resulting ensemble of positive definite matrices was named after the researcher who first computed the joint element density of (AA^T): the *Wishart* ensemble.

Later, statisticians defined another type of positive definite random matrix, which essentially encodes two independent Wishart matrices, A and B, and named it a MANOVA[1]) (a.k.a. Jacobi) matrix: $(C \equiv A^{1/2}BA^{1/2})$.

[1]from Multivariate ANalysis Of VAriance.

The first normal distribution to be considered for the entries of the random matrix was real; shortly after, the complex normal distribution was investigated, and a question arose: what other types of normal entries could be considered, in terms of the field over which they were distributed?

This lead Dyson [23] to the classification of matrix ensembles, and the *threefold way*. To summarize, Dyson argued that in order for the random matrix model to be consistent with the physical model, the entries had to be distributed over a division algebra over $\mathbb{R}$; and since the only three division algebras over the real numbers are $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$ (the quaternions[2]), the only three types of normal variables were real, complex and quaternion.

Later, Zirnbauer [99] and Ivanov [43] extended Dyson's classification by studying symmetric spaces. Their taxonomies include some Laguerre and Jacobi cases, and also Dyson's circular ensembles [22]; however, each and every one of these cases is not outside of the "threefold way" realm, since they do not consider any other β values (they always work with real, complex, or quaternion entries).

Ever since Dyson's paper was published in 1963, the physics community has investigated closely the three Gaussian ensembles: Gaussian *Orthogonal* (GOE), *unitary* (GUE), and *symplectic* (GSE), thus named for the type of "orthogonal" invariance. These correspond, as in (2.1) to a choice of parameter β in the joint eigenvalue distribution (β can be seen as counting the number of real Gaussians in each entry, that is, 1 for the reals, 2 for the complexes, and 4 for the quaternions). Many statistics were thoroughly investigated for this triad of ensembles, starting with 1- and 2-point correlation functions (also known as the level density, respectively, the nearest neighbor spacing), determinants, and extremal eigenvalues. The approaches were somewhat different (though also somewhat similar) for each of the three cases, which perhaps contributed to the idea of a discrete set of ensembles.

On the other side, statisticians, though very interested in the real and complex Wishart and MANOVA models, were less so in the quaternion ones. Similar statistics

[2]Dyson used $\mathbb{Q}$ instead of $\mathbb{H}$, but since $\mathbb{Q}$ is more often identified with the rationals, the notation was later changed to $\mathbb{H}$ (from Hamilton, the "father" of the quaternion division algebra).

as in the case of the Gaussian ensembles were studied, sometimes with better success, as it seems that the Wishart and MANOVA ensembles are more tractable. The methods used were virtually the same in every case, which permitted a more homogeneous approach, and faster progress in the study of the eigenvalue statistics. Connections were quickly established with other fields, like special functions (through connections with Schur and zonal polynomials, hypergeometric functions, and multivariate orthogonal polynomials).

Many studies of integrals over the general β ensembles focused on the connection with Jack polynomials; of these we note the ones inspired by Selberg's work, like Aomoto [4], Kaneko [56], Kadell [55], and others. The last two authors also worked with multivariate Jacobi polynomials.

Another source of general β came in connection with the theory of Schrödinger operators of arbitrary (inverse temperature) parameter β; in particular we refer to the eigenvalues and eigenfunctions of these operators.

Finally, there are now instances of arbitrary β in algebraic geometry, which is one of the fields whose connection with random matrices is perhaps the most surprising, as random matrix integrals arise in the context of counting discrete things, like maps on certain surfaces [36].

This work continues in the same (arbitrary β) pattern of thinking. The results we prove here work in the general setting, and we believe they constitute strong grounds for the existence of a unifying theory, which would explain the similarities and differences between the different values of β, provide new insights on seemingly discrete patterns of behavior, and thus help advance the study of random matrices beyond the confines of the threefold way.

2.2 The Hermite (Gaussian) ensembles

The Gaussian ensembles were introduced by Wigner in the 1950's. Though he started with a simpler model for a random matrix (entries from the uniform distribution on $\{-1, 1\}$, [94]), he noted without proving that the (limiting level density) theory applies

in a more general setting. Wigner also derived the joint eigenvalue distribution for the Gaussian ensembles.

Aside from the limiting level density, finite n exact level densities for each one of the three ensembles are known, due to Gaudin [33] and Mehta [67]; asymptotics for the level densities have also been explored by Basor, Tracy, and Widom [9]. For a more in-depth history of the limiting level densities, see Chapter 6. Finally, the extremal eigenvalue asymptotics were computed by Tracy and Widom [91].

The three Gaussian ensembles have joint eigenvalue p.d.f. (probability density function)

$$\text{HERMITE:} \qquad f_\beta(\lambda) = c_H^\beta \prod_{i<j} |\lambda_i - \lambda_j|^\beta e^{-\sum_{i=1}^n \lambda_i^2/2} \ , \tag{2.1}$$

where

$$c_H^\beta = (2\pi)^{-n/2} \prod_{j=1}^n \frac{\Gamma(1+\frac{\beta}{2})}{\Gamma(1+\frac{\beta}{2}j)} \ . \tag{2.2}$$

Until the '80s, this topic was mostly studied by (mathematical) physicists (except, arguably, for Tracy and Widom). The '80s have marked the beginnings of a large body of work on the subject by algebraic combinatorialists. A significant number of combinatorial problems was found to be related to random matrix theory (mostly to statistics of the GOE and GUE). These include the longest increasing subsequence in a random permutation (Baik, Deift, Johansson [7], Okounkov [74]), plane partitions (Borodin and Olshanski [12]), magic squares (Diaconis and Gamburd [32]), growth models (Gravner, Tracy and Widom [37]), and counting maps on conformal surfaces (Goulden and Jackson [36]).

Throughout this thesis, we will refer to the Gaussian ensembles (for general or particular β) as Hermite ensembles (this is a technically more accurate description, related to the type of weight function).

2.3 The Laguerre (Wishart) and Jacobi (MANOVA) ensembles

The Wishart ensembles were introduced by statisticians, and are older than the Gaussian ones. In a statistics paper from 1928 [96], Wishart proposed a matrix model which came to be known as the Wishart real model; he was also the first to compute the joint element distribution of this model.

The study of the Wishart models intensified in the late 1960's and 1970's, taking a different direction than the study of the Gaussian ensembles. The breakthrough for the limiting level densities came in the paper of Marcenko and Pastur [64], in an even larger context than the Wishart models.

Another breakthrough came through the work of James, who, in his study of the Wishart real distribution, defined and described zonal polynomials [47], and started the study of eigenvalue statistics in terms of special functions. Constantine [14] generalized the univariate hypergeometric functions, in terms of zonal polynomials. Eventually, this line of approach lead Jack to define what are now known as Jack polynomials [44].

A significant amount of work towards computing eigenvalue statistics in terms of special functions was done by Muirhead [71] and Chikuse [13]. Muirhead is also the author of a reference book for the study of real and complex Wishart ensembles [70].

Finally, from the study of extremal eigenvalues for the Wishart ensembles we mention the work of Krishnaiah and Chang [58], Silverstein [80], and Edelman [24].

Recently, Johnstone [53] and Johansson [52] have found a very interesting connection between the Gaussian ensembles and the Wishart ones: the distribution of the scaled largest eigenvalue in either case is given by the same Tracy-Widom law (F_1 for the real case, F_2 for the complex one, see [89, 90, 91]).

The Wishart (or Laguerre) $m \times n$ models have joint eigenvalue p.d.f.

$$\text{LAGUERRE:} \qquad f_\beta(\lambda) = c_L^{\beta,a} \prod_{i<j} |\lambda_i - \lambda_j|^\beta \prod_{i=1}^{m} \lambda_i^{a-p} e^{-\sum_{i=1}^n \lambda_i/2} \,, \tag{2.3}$$

with $a = \frac{\beta}{2} n$ and $p = 1 + \frac{\beta}{2}(m-1)$. Again, $\beta = 1$ for the reals, $\beta = 2$ for the

complexes, and $\beta = 4$ for the quaternions. The constant is

$$c_L^{\beta,a} = 2^{-ma} \prod_{j=1}^{m} \frac{\Gamma(1+\frac{\beta}{2})}{\Gamma(1+\frac{\beta}{2}j)\Gamma(a-\frac{\beta}{2}(m-j)))} \ . \tag{2.4}$$

Though this thesis does not use them explicitly, to complete the triad of classical orthogonal polynomials (Hermite-Laguerre-Jacobi), we mention here the β-MANOVA ensembles. They are better known in the literature as the Jacobi ensembles, with joint eigenvalue p.d.f.

$$\text{JACOBI:} \qquad f_\beta(\lambda) = c_J^{\beta,a_1,a_2} \prod_{i<j} |\lambda_i - \lambda_j|^\beta \prod_{j=1}^{m} \lambda_i^{a_1-p} \ (1-\lambda_i)^{a_2-p} \ , \tag{2.5}$$

with $a_1 = \frac{\beta}{2}n_1$, $a_2 = \frac{\beta}{2}n_2$, and $p = 1 + \frac{\beta}{2}(m-1)$. As usual, $\beta = 1$ for real and $\beta = 2$ for complex; also

$$c_J^{\beta,a_1,a_2} = \prod_{j=1}^{m} \frac{\Gamma(1+\frac{\beta}{2})\Gamma(a_1+a_2-\frac{\beta}{2}(m-j))}{\Gamma(1+\frac{\beta}{2}j)\Gamma(a_1-\frac{\beta}{2}(m-j))\Gamma(a_2-\frac{\beta}{2}(m-j))} \ . \tag{2.6}$$

The MANOVA real and complex cases ($\beta = 1$ and 2) have been studied by statisticians (for a good reference on the $\beta = 1$ and $\beta = 2$ cases, see [70]). For our purposes, they are relevant in the study of the general β context, as we will see in the next section.

For reasons similar to the ones described in Section 2.2, throughout this thesis, we will refer to the Wishart ensembles (for general or particular β) as Laguerre ensembles, and to the MANOVA ensembles as Jacobi ensembles.

2.4 General β-ensembles as theoretical distributions

One of the first (if not *the* first) researcher to consider general values as the power of the repulsion factor $\prod_{i\neq j} |\lambda_i - \lambda_j|$ was Selberg [79]. One of his most important contributions to random matrix theory was to compute the normalization constants for the general β-Jacobi distribution. From this single integral, one can obtain many

others (through changes of variables, and taking limits). For an extensive list, see [66].

In the early '80s, Askey and Richards [6] simplified the proof, and a few years later, Aomoto generalized the Selberg integral [4]. Related to this, there is a conjecture of Macdonald [62] involving Coxeter groups, which in certain special cases takes on the Selberg integral form.

The first to obtain multivariate Jacobi polynomials as polynomial eigenfunctions of (Jacobi forms of) the Schrödinger operator were Beerends and Opdam [10]. In particular, they showed orthogonality with respect to the multivariate (general β) Jacobi distribution.

Lasalle [59, 60, 61] pushed the study of multivariate orthogonal polynomials as eigenfunctions of Schrödinger operators even further, by defining and describing the multivariate Hermite and Laguerre polynomials as limiting cases of Jacobi.

Yan [97] generalized the Laguerre polynomials for every β and n (the number of variables) equal to 2.

The next important step was taken by Forrester and Baker [8]. They proved the orthogonality of the Hermite and Laguerre multivariate polynomials with respect to the general β-Hermite and β-Laguerre distributions, and computed certain arbitrary β eigenvalue statistics (such as the "ground state global density", i.e. the level density) in terms of them.

Finally, two important general β results were proved by Johansson; he showed that the eigenvalues of general β-Hermite ensembles [51] and β-Jacobi ensembles [50] obey a Central Limit Theorem (more about the former can be found in Chapter 6).

2.5 Contributions of this thesis

The most important contribution of this thesis is that it provides real tridiagonal matrix models for *any* β-Hermite ensemble with $\beta > 0$, and for *any* β-Laguerre ensemble of *any* parameter $a \geq (m-1)\beta/2$, where m is the size of the ensemble. Almost all the results in this thesis are applications of these models.

The Maple Library (MOPs: *Multivariate Orthogonal Polynomials (symbolically)*) was written for users in the many areas of application. We provide a complexity analysis of the main routines in Chapter 9. Even though the computational complexities are inherently superpolynomial, today's computers, because of the memory and speed available, make it possible to evaluate all the quantities mentioned in this thesis in a reasonable amount of time.

Below is the list of the main results of this thesis.

- We provide symmetric tridiagonal random matrix models for all β-Hermite ensembles. These models have classical entry distributions (standard normal and χ distributions) over the *real* numbers, and their entries are mutually independent up to the symmetry constraints. The joint eigenvalue p.d.f.s of these models are *exactly* given by (2.1), for any $\beta > 0$. For the explicit format see Table 2.1; for the proof, see Chapter 5, Section 5.2.

- We provide positive definite tridiagonal random matrix models for all β-Laguerre ensembles. These models have χ distributed entries over the *real* numbers, and their entries are mutually independent up to the positive definiteness constraints. The joint eigenvalue p.d.f.s of these models are *exactly* given by (2.3), for any $\beta > 0$, and for any value of the parameter $a \geq (m-1)\beta/2$. This is a natural restriction arising from the integrability of the p.d.f.Ṗreviously, the known matrix models restricted a to be of the form $n\beta/2$. For the explicit format see Table 2.1; for the proof, see Chapter 5, Section 5.3.

- We prove a strong semi-circle law for all β-Hermite ensembles (Strong Law of Large Numbers). While this result is subsumed by Johansson's Central Limit Theorem, our proof is combinatorial and uses the matrix models (it is similar to Wigner's proof and Arnold's proof). For statement and proof, see Chapter 6.

- We prove a strong type of semi-circle law for all β-Laguerre ensembles (Strong Law of Large Numbers). We believe that this result is new for the general β, general a case. For statement and proof, once again see Chapter 6.

- We obtain zero and first-order approximations for the large β eigenvalues of β-Hermite and β-Laguerre ensembles of constrained parameter a. The zero-order approximations (which are limiting values as $\beta \to \infty$) turn out to be roots of Hermite, respectively, Laguerre polynomials, while the first-order approximations (which are limiting distributions for the scaled eigenvalues) are Gaussian variables with mean 0 and explicitly computed variance (see Chapter 7).

- We obtain zero- and first-order approximations for the level densities of the β-Hermite and β-Laguerre ensembles of constrained a parameter, as weighted sums of Gaussian variables. These become exact as $\beta \to \infty$, but in fact β no larger than 6, 8, or 10 provide good approximations (see Chapter 7, Section 7.2).

- We obtain a duality principle (between β and $2/\beta$) for the (normalized) Jack polynomial average over the corresponding β-Hermite ensemble; we also obtain a simple proof for a conjecture of Goulden and Jackson [36], first proved by Okounkov [73]. We compute the second moment of the determinant of a general β-Hermite matrix, and provide 3− and 4−term recurrences for the third and fourth moments. All these results are in Chapter 8, Section 8.5.

- We present the algorithms used in our Maple Library MOPs, and analyze their complexities; we present both theoretical results and performance in practice; see Chapter 9.

- We present some immediate generalizations for β-Laguerre eigenvalue statistics.

We display our random matrix constructions in Table 2.1.

Along the way, we obtain various other results: we compute some (we believe) new Jacobians of matrix factorizations, including the Jacobian the symmetric tridiagonal eigenvalue decomposition, we rediscover the Selberg integral in Hermite and Laguerre form, and we prove that the expected value of a symmetric polynomial independent of β over either the β-Hermite and β-Laguerre distributions are polynomials in β.

Hermite Matrix $n \in \mathbb{N}$	$H_\beta \sim \frac{1}{\sqrt{2}} \begin{pmatrix} N(0,2) & \chi_{(n-1)\beta} & & & \\ \chi_{(n-1)\beta} & N(0,2) & \chi_{(n-2)\beta} & & \\ & \ddots & \ddots & \ddots & \\ & & \chi_{2\beta} & N(0,2) & \chi_\beta \\ & & & \chi_\beta & N(0,2) \end{pmatrix}$
Laguerre Matrix $m \in \mathbb{N}$ $a \in \mathbb{R}$ $a > \frac{\beta}{2}(m-1)$	$L_\beta = B_\beta B_\beta^T$, where $B_\beta \sim \begin{pmatrix} \chi_{2a} & & & \\ \chi_{\beta(m-1)} & \chi_{2a-\beta} & & \\ & \ddots & \ddots & \\ & & \chi_\beta & \chi_{2a-\beta(m-1)} \end{pmatrix}$

Table 2.1: Random Matrix Constructions.

We emphasize that in this thesis we systematically treat the $\beta = 1, 2, 4$ in a unified case, by providing formulas which are β-dependent, rather than listing three different cases.

The final contribution to the understanding of the unified, general β principle, even in the "threefold way" perspective, is given in the table of Figure 2-1.

Random Matrix Ensembles

Joint Eigenvalue Distributions: $c\,|\,\Delta\,|^{\beta}\prod_{i=1}^{n} V(\lambda_i)$

Technical name	Traditional name	β	Field	Property	Invariance	Connected Matrix Problem
Hermite ensembles	**Gaussian ensembles**					(EIG)
$V(\lambda) = e^{-\lambda^2/2}$	GOE	1	$\mathbb{R}$	Symmetric	$A \rightarrow Q^T A Q$	Symmetric
	GUE	2	$\mathbb{C}$	Hermitian	$A \rightarrow U^H A U$	Eigenvalue Problem
	GSE	4	$\mathbb{H}$	Self−Dual	$A \rightarrow S^D A S$	
Laguerre ensembles	**Wishart ensembles**					(SVD)
$V(\lambda) = \lambda^a e^{-\lambda/2}$	Wishart real	1	$\mathbb{R}$	Positive	$A \rightarrow Q^T A Q$	Singular Value
$a = \frac{\beta}{2}(n - m + 1) - 1$	Wishart complex	2	$\mathbb{C}$	Semi−Definite	$A \rightarrow U^H A U$	Decomposition
	Wishart quaternion	4	$\mathbb{H}$		$A \rightarrow S^D A S$	
Jacobi ensembles	**MANOVA ensembles**					(QZ)
$V(\lambda) = \lambda^a (1-\lambda)^b$	MANOVA real	1	$\mathbb{R}$	Positive	$X \rightarrow Q^T X Q_1$, $Y \rightarrow Q^T Y Q_1$	Generalized Symmetric
$a = \frac{\beta}{2}(n_1 - m + 1) - 1$	MANOVA complex	2	$\mathbb{C}$	Semi−Definite	$X \rightarrow U^H X U_1$, $Y \rightarrow U^H Y U_1$	Eigenvalue Problem
$b = \frac{\beta}{2}(n_2 - m + 1) - 1$	MANOVA quaternion	4	$\mathbb{H}$		$X \rightarrow S^D X S_1$, $Y \rightarrow S^D Y S_1$	

Type of ensemble	β	**MATLAB** code
Hermite	1	A = randn(n); A = (A+A') / 2;
	2	A = randn(n) + i * randn(n); A = (A+A') / 2;
	4	X = randn(n) + i * randn(n); Y = randn(n) + i * randn(n); A = [X Y; −conj(Y) conj(X)]; A = (A+A') / 2;
Laguerre	1	A = randn(m, n); A = A * A';
	2	A = randn(m, n) + i * randn(m, n); A = A * A';
	4	X = randn(m, n) + i * randn(m, n); Y = randn(m, n) + i * randn(m, n); A = [X Y; −conj(Y) conj(X)]; A = A * A';
Jacobi	1	X = randn(m, n_1); Y = randn(m, n_2); A = (X * X') / (X * X' + Y * Y');
	2	X = randn(m, n_1) + i * randn(m, n_1); Y = randn(m, n_2) + i * randn(m, n_2); A = (X * X') / (X * X' + Y * Y');
	4	X_1 = randn(m, n_1) + i * randn(m, n_1); X_2 = randn(m, n_1) + i * randn(m, n_1); Y_1 = randn(m, n_2) + i * randn(m, n_2); Y_2 = randn(m, n_2) + i * randn(m, n_2); X = [X_1 X_2; −conj(X_2) conj(X_1)]; Y = [Y_1 Y_2; −conj(Y_2) conj(Y_1)]; A = (X * X') / (X * X' + Y * Y');

Figure 2-1: Random Matrix Ensembles. As a guide to MATLAB notation, randn(m, n) produces an $m \times n$ matrix with i.i.d. standard normal entries, conj(X) produces the complex conjugate of the matrix X, and the $'$ operator produces the conjugate transpose of a matrix. Also $[X\ Y;\ Z\ W]$ produces a 2×2 block matrix.

Chapter 3

Random Matrix concepts, notation, and terminology

3.1 Basic distributions and random matrix models

One of the most important univariate (single variable) distributions in statistics and random matrix theory is the normal (or Gaussian) distribution $N(\mu, \sigma^2)$, of mean μ and variance σ^2. This distribution has a complex version which we denote by $N^2(\mu, \sigma^2)$, of the form $x + iy$, where x and y are i.i.d. real variables with distribution $N(\mu, \sigma^2)$. Similarly, one can define the quaternion complex version $N^4(\mu, \sigma^2)$, with form $x+iy+jz+kw$, where x, y, z, and w are all real i.i.d. variables with distribution $N(\mu, \sigma^2)$.

The normal (real, complex, quaternion) distribution easily generalizes to multivariate form, as a vector or matrix of normal (real, complex, quaternion) variables. Throughout this thesis, we will denote a random Gaussian matrix as described below. We will always use $m \leq n$.

$G(m, n)$ for an $m \times n$ matrix of real independent standard Gaussians (entries i.i.d. with distribution $N(0, 1)$).

$G^2(m, n)$ for an $m \times n$ matrix of complex independent standard Gaussians (entries i.i.d. with distribution $N^2(0, 1)$).

$G^4(m,n)$ for an $m \times n$ matrix of quaternion independent standard Gaussians (entries i.i.d. with distribution $N^4(0,1)$).

The most important property of the multivariate Gaussian, be it real, complex, or quaternion, is the *orthogonal invariance*. This makes the distribution impervious to multiplication by an orthogonal (unitary, symplectic) matrix, provided that the two are independent. We offer for example the following MATLAB experiment. A real matrix $A \sim G(m,n)$ is produced in MATLAB by making the assignment $A =$randn(m,n); if orthogonal matrices Q_1 $(m \times m)$ and Q_2 $(n \times n)$ are produced in whichever way the user might want, provided that A is not used at all in constructing them[1], no test can be devised that would differentiate between $Q_1 A$, A, and AQ_2.

This property, as we will later see, is the key in analyzing the random matrices which are constructed using the Gaussian distribution.

Many other important distributions are derived from the univariate Gaussian; of these, the most useful for our purposes is the χ_r distribution (also known as the square root of the χ_r^2 distribution). If the parameter r is a positive integer n, one definition of χ_n is given by $||G(n,1)||_2$, in other words, the 2-norm of a vector of independent standard normals. The probability density function of χ_n can then be extended to

$$f_r(x) = \frac{1}{2^{r/2-1}\,\Gamma\left(\frac{1}{2}n\right)} \quad x^{r-1} \quad e^{-x^2/2} \ ,$$

where r is any real number (the number of "degrees of freedom"). It is not hard to see that a variable x with distribution χ_r has mean

$$\mu = \sqrt{2} \ \ \frac{\Gamma\left(\frac{1}{2}(n+1)\right)}{\Gamma\left(\frac{1}{2}n\right)}$$

and variance

$$\sigma^2 = 2 \ \ \frac{\Gamma\left(\frac{1}{2}n\right)\Gamma\left(\frac{1}{2}n+1\right) - \Gamma^2\left(\frac{1}{2}(n+1)\right)}{\Gamma^2\left(\frac{1}{2}n\right)} \ .$$

Our *sparse* (bidiagonal, tridiagonal) matrix models are defined later in this thesis with the use of the χ distribution; we will also use the p.d.f. of the χ distribution in asymptotic approximations. However, we only need the Gaussian to define the

[1]One possible way is $[Q_1, R_1] =$qr(randn(n)), $[Q_2, R_2] =$qr(randn(n)).

following two lists which capture the two types of *full* random matrices we will use from now on. By way of construction, all matrix ensembles below are orthogonally invariant.

Gaussian Orthogonal Ensemble (GOE);

symmetric $m \times m$ matrix obtained as $(A + A^T)/2$ where A is $G(m, m)$. The diagonal entries are i.i.d. with distribution $N(0, 1)$, and the off-diagonal entries are i.i.d. (subject to the symmetry) with distribution $N(0, \frac{1}{2})$.

Gaussian Unitary Ensemble (GUE);

hermitian $m \times m$ matrix obtained as $(A + A^H)/2$ where A is $G^2(m, m)$ and H denotes the hermitian transpose of a complex matrix. The diagonal entries are i.i.d with distribution $N(0, 1)$, while the off-diagonal entries are i.i.d. (subject to being hermitian) with distribution $N^2(0, \frac{1}{2})$.

Gaussian Symplectic Ensemble (GSE);

self-dual $m \times m$ matrix obtained as $(A + A^D)/2$ where A is $G^4(m, m)$ and D denotes the dual transpose of a quaternion matrix. The diagonal entries are i.i.d with distribution $N(0, 1)$, while the off-diagonal entries are i.i.d. (subject to being self-dual) with distribution $N^4(0, \frac{1}{2})$.

Wishart real ensemble ($W(m, n)$, $m \leq n$);

symmetric $m \times m$ matrix which can be obtained as AA^T where A is $G(m, n)$.

Wishart complex ensemble ($W^2(m, n)$, $m \leq n$);

hermitian $m \times m$ matrix which can be obtained as AA^T where A is $G^2(m, n)$.

Wishart quaternion ensemble ($W^4(m, n)$, $m \leq n$);

self-dual $m \times m$ matrix which can be obtained as AA^T where A is $G^4(m, n)$.

As might have become apparent, we use the superscripts 2 and 4 to indicate the presence of 2 or 4 real Gaussians in each entry of the matrix. The normalizations have been chosen so as to agree with the distributions described Chapter 2; the literature contains these normalizations together with others.

To complete the list of random matrices we investigate in this thesis, we present below the sparse (bidiagonal, tridiagonal) matrix models we constructed for general β. To define these, we use the (perhaps familiar) MATLAB notation $[a : -t : a - kt]$ to indicate a descending arithmetic progression of length $k + 1$ starting at a and having step $-t$.

$G_{\beta,\ a}(m)$; a lower bidiagonal $m \times m$ matrix, with diagonal $\chi_{[2a\ :\ -\beta\ :\ 2a-(m-1)\beta]}$ and subdiagonal $\chi_{[(m-1)\beta\ :\ -\beta\ :\ \beta]}$; all entries are mutually independent.

β-Hermite ensemble ($\beta > 0$) ($H_\beta(n)$);

symmetric tridiagonal matrix with diagonal $G(n, 1)$ and subdiagonal $\frac{1}{\sqrt{2}}\ \chi_{[(n-1)\beta\ :\ -\beta\ :\ \beta]}$; all entries are mutually independent subject to the symmetry condition.

β-Laguerre ensemble of parameter a ($\beta > 0$, $a > \beta(m-1)/2$) ($L_{\beta,\ a}(n)$);

symmetric matrix which can be obtained as BB^T where B is $G_{\beta,\ a}(m)$.

3.2 Element and eigenvalue densities

In this section we list the joint element and joint eigenvalue densities of the six matrix models (three Gaussian and three Wishart) we enumerated in the previous section. These can be found in any book on random matrix theory (for good references, see [66], [70]).

We also include the joint element and joint eigenvalue densities of the two types of models (Hermite/Gaussian and Laguerre/Wishart) that we construct in Section 5 (we also refer to that section for the computation of the joint eigenvalue densities).

Before we proceed, we will need to define the multivariate Gamma function for arbitrary β; the real and complex versions are familiar from the literature, and the arbitrary β case represents an immediate extension:

$$\Gamma_m^\beta(a) \quad = \quad \pi^{m(m-1)\beta/4} \ \prod_{i=1}^{m} \Gamma\left(a + \frac{\beta}{2}(i-1)\right) \ . \tag{3.1}$$

Just as the univariate Gamma function generalizes to the multivariate one, the shifted factorial (Pochhammer symbol, rising factorial $(a)_k = \frac{\Gamma(a+k)}{\Gamma(a)}$) becomes the generalized multivariate shifted factorial.

We call

$$(a)_\kappa^\beta = \prod_{i=1}^{length(\kappa)} \left(a-(i-1)\frac{\beta}{2}\right)_{\kappa_i} = \prod_{i=1}^{length(\kappa)} \frac{\Gamma\left(a-(i-1)\frac{\beta}{2}+\kappa_i\right)}{\Gamma\left(a-(i-1)\frac{\beta}{2}\right)} \tag{3.2}$$

the *generalized shifted factorial*, or *generalized Pochhammer symbol*.

In Tables 3.1 and 3.2 we list the joint element and joint eigenvalue densities for the three Gaussian ensembles of the preceding section; the constants correspond to the normalization we have chosen.

We recall the orthogonal invariance property; this is easily seen in the joint element and eigenvalue density, since all matrices are symmetric. We use the notation $\mathrm{etr}(A)$ for the exponential of the trace of A.

Gaussian	orthogonal	$\beta = 1$	$\frac{1}{2^{n/2}} \frac{1}{\pi^{n/2+n(n-1)\beta/4}} \mathrm{etr}(-A^2/2)$
	unitary	$\beta = 2$	
	symplectic	$\beta = 4$	

Table 3.1: Joint element density of an $n \times n$ matrix A from a Gaussian ensemble.

Below we denote by Λ the diagonal matrix of eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. We use the notation $\Delta(\Lambda) = \prod_{1 \le i < j \le n} (\lambda_i - \lambda_j)$.

Gaussian	orthogonal	$\beta = 1$	$\frac{\pi^{n(n-1)\beta/4}}{(2\pi)^{n/2}} \frac{(\Gamma(1+\frac{\beta}{2}))^n}{\Gamma_n^\beta(1+\frac{\beta}{2})} \lvert\Delta(\Lambda)\rvert^\beta \mathrm{etr}(-\Lambda^2/2)$
	unitary	$\beta = 2$	
	symplectic	$\beta = 4$	

Table 3.2: Joint eigenvalue density (Λ) of the $n \times n$ Gaussian ensembles.

Tables 3.3 and 3.4 contain joint element and joint eigenvalue densities for the Wishart ensembles.

Finally, we present the joint element and joint eigenvalue densities for the two tridiagonal models (β-Hermite and β-Laguerre) described in Table 2.1. For an $n \times n$

Wishart	orthogonal unitary symplectic	$\beta = 1$ $\beta = 2$ $\beta = 4$	$\frac{1}{2^{mn\beta/2}} \frac{1}{\Gamma_m^\beta (n\beta/2)} \operatorname{etr}(-W/2) (\det W)^{\beta(n-m+1)/2-1}$

Table 3.3: Joint element density of an $m \times n$ matrix W from a Wishart ensemble.

Wishart	orthogonal unitary symplectic	$\beta = 1$ $\beta = 2$ $\beta = 4$	$\frac{\pi^{m(m-1)\beta/2}}{2^{mn\beta/2}} \frac{\Gamma(1+\frac{\beta}{2})^m}{\Gamma_m^\beta(1+\frac{\beta}{2})\Gamma_m^\beta((n-m+1)\frac{\beta}{2})}$ $\times \quad \lvert\Delta(\Lambda)\rvert^\beta \operatorname{etr}(-\Lambda/2) (\det \Lambda)^{\beta(n-m+1)/2-1}$

Table 3.4: Joint eigenvalue density (Λ) of the $m \times n$ Wishart ensembles.

symmetric tridiagonal matrix, we use the following notation:

$$T = \begin{pmatrix} a_n & b_{n-1} & & & & \\ b_{n-1} & a_{n-1} & b_{n-2} & & & \\ & b_{n-2} & a_{n-2} & \cdots & & \\ & & \vdots & \ddots & \cdots & \\ & & & \vdots & a_2 & b_1 \\ & & & & b_1 & a_1 \end{pmatrix} \tag{3.3}$$

and for a bidiagonal matrix we use the notation

$$B = \begin{pmatrix} x_n & & & & \\ y_{n-1} & x_{n-1} & & & \\ & \ddots & \ddots & & \\ & & y_2 & x_2 & \\ & & & y_1 & x_1 \end{pmatrix} . \tag{3.4}$$

Consequently, an $n \times n$ tridiagonal positive semidefinite matrix $L = BB^T$ will have

$$a_n = x_n^2 \ , \tag{3.5}$$

$$a_{n-i} = x_{n-i}^2 + y_{n-i}^2 \ , \quad \forall \ 1 \leq i \leq n-1 \ \ , \tag{3.6}$$

$$b_{n-i} = x_{n-i+1} y_{n-i} \ , \quad \forall \ 1 \leq i \leq n-1 \ \ . \tag{3.7}$$

Tables 3.5 and 3.6 summarize the joint element and joint eigenvalue densities for the β-Hermite ensembles and β-Laguerre ensembles.

$H_\beta(n)$	$\beta > 0$	$\frac{\pi^{(n-1)(n-2)\beta/4 - n/2}}{2^{1-n/2}} \frac{1}{\Gamma_{m-1}^{\beta}(\frac{\beta}{2})} \times \prod_{i=1}^{n-1} b_i^{i\beta-1} \ e^{-\sum_{i=1}^{n} a_i^2/2} \ e^{-\sum_{i=1}^{n-1} b_i^2}$
$L_{\beta,\ a}(m)$	$\beta > 0$ $a > \beta(m-1)/2$	$\frac{2^{2m-2} \ \pi^{(m-1)^2\beta/2}}{\Gamma_{m-1}^{\beta}(\frac{\beta}{2}) \ \Gamma_m^{\beta}(a-(m-1)\frac{\beta}{2})} \ x_1^{2(a-1)-\beta(m-1)} \ e^{-x_1^2/2} \times \prod_{i=2}^{m} x_i^{2a-(m-1)\beta-3} \ e^{-x_i^2/2} \ \prod_{i=1}^{m-1} y_i^{i\beta-1} e^{-y_i^2/2}$

Table 3.5: Joint element densities for the β-Hermite and β-Laguerre ensembles.

3.3 Level densities

Many statistics of the Gaussian and Wishart ensembles have been calculated exactly or asymptotically (as the size of the matrix models grows to infinity) in the half of century that followed their discovery. Of these, arguably the most important ones are level densities, spacing distributions, and extremal eigenvalue distributions.

$H_\beta(n)$	$\beta > 0$	$\frac{\pi^{n(n-1)\beta/4}}{(2\pi)^{n/2}} \frac{(\Gamma(1+\frac{\beta}{2}))^n}{\Gamma_n^\beta(1+\frac{\beta}{2})} \quad \lvert\Delta(\Lambda)\rvert^\beta \ \operatorname{etr}(-\Lambda^2/2)$
$L_{\beta,\ a}(m)$	$\beta > 0$ $a > \beta(m-1)/2$	$\frac{\pi^{m(m-1)\beta/2}}{2^{ma}} \frac{\Gamma(1+\frac{\beta}{2})^m}{\Gamma_m^\beta(1+\frac{\beta}{2})\Gamma_m^\beta(a-(m-1)\frac{\beta}{2})}$ $\times \quad \lvert\Delta(\Lambda)\rvert^\beta \operatorname{etr}(-\Lambda/2) \ (\det\Lambda)^{a-\beta(m-1)/2-1}$

Table 3.6: Joint eigenvalue (Λ) densities for the β-Hermite and β-Laguerre ensembles.

Once the β-Hermite and β-Laguerre ensembles were introduced as theoretical distributions for arbitrary β, there were many attempts at generalizing the known results for these three types of statistics for $\beta = 1, 2, 4$. Perhaps the most notable discoveries are those of Forrester and Baker [8], and of Johannson [50, 51]. Forrester and Baker provide a closed-form expression for the level densities of the β-Hermite ensembles, in terms of a multivariate Hermite polynomial. Johansson proves a very strong type of asymptotic convergence of the level densities of β-Hermite and β-Jacobi ensembles to given distributions (in the β-Hermite case, the semi-circle distribution).

Neither the spacings nor the extremal eigenvalues of the β-ensembles have been investigated much. We provide distributions for the extremal eigenvalues of β-Laguerre ensembles in Chapter 10.

The notion of level density or one-point correlation function [66, Chapter4] comes from nuclear physics, and the word "level" is meant to signify "energy level" of an electron.

Definition 3.3.1. *Let A be a matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$. The empirical distribution function for the eigenvalues of A is the distribution given by p.d.f. $\frac{1}{n}\sum_{i=1}^n \delta(x-\lambda_i)$.*

If a probability measure is placed on A (and thus, on the eigenvalues of A), we can talk about the distribution of a random eigenvalue of a random matrix of the ensemble; in other words, of a *random* empirical distribution.

If we average the random empirical density over the probability measure imposed on A, we obtain the distribution of a random eigenvalue *of the ensemble*; this is known as the *level density* of the ensemble.

Definition 3.3.2. *The level density of an $n \times n$ ensemble is the distribution of a random eigenvalue chosen from the ensemble. Equivalently, it is the average (over the ensemble) empirical density.*

There is another way to understand the level density. If one integrates out all but one of the variables in the joint (unordered) eigenvalue distribution of an ensemble, what is left is the level density.

For example, the level density (or one-point correlation function) of a Hermite ensemble is

$$\rho_{N,\beta}(\lambda_1) = c_H^\beta \int_{R^{n-1}} |\Delta(\Lambda)|^\beta e^{-\sum_{i=1}^{n} \lambda_i^2/2} d\lambda_2 \dots d\lambda_n \ .$$

The level densities have been computed exactly for $\beta = 1, 2, 4$ for the Gaussian and Wishart ensembles; for $\beta = 2$ they depend in a simple way on the (univariate) orthonormal Hermite (respectively Laguerre) polynomials, while for $\beta = 1$ and 4 they have slightly more complicated forms (see [66]).

Level densities in a more general setting (β an even integer) have been considered by Forrester and Baker [8].

Chapter 4

Jacobians and Perturbation Theory

4.1 Jacobians of matrix factorizations

One of the primary functions of Numerical Linear Algebra is that to provide direct and iterative algorithms for matrix operations; of these, a very important role is played by matrix factorizations.

Up to some sign-restricting conventions and orderings, a matrix factorization can be seen as a change of variables. Thus it has a Jacobian, which is the determinant of the linearization of this change of variables.

Many of the Jacobians of matrix factorizations were first computed by researchers of Random Matrix Theory, for the simple reason that they were needed in finding an eigenvalue or singular value distribution. For a survey of some of the matrix factorizations that occur in Random Matrix Theory, together with their Jacobians, see [76].

Wigner [95] was the first to compute the Jacobian of the $Q\Lambda Q^T$ factorization for a symmetric matrix. On the statistics part, the Jacobian for the $U\Sigma V^T$ transformation of a real rectangular matrix was needed in computing the joint eigenvalue density for the Wishart matrix; a derivation of this Jacobian can be found in [70].

Later, other matrix factorizations were considered, such as QR, LU, LL', etc. In what follows, we list some of the more important factorizations and their Jacobians, for all three cases (real $\beta = 1$, complex $\beta = 2$, quaternion $\beta = 4$). First we need

to define matrix differentials. Note that for a complex variable $x + iy$, if there is no dependence between x and y, we have $d(x + iy) = dx\ dy$; in other words, when we take differentials, we consider only the real parameters, as we will work over $\mathbb{R}^{mn}$, $\mathbb{R}^{2mn}$, $\mathbb{R}^{4mn}$. Similarly $d(x + iy + jz + kt) = dx\ dy\ dz\ dt$.

If S is an $n \times n$ symmetric matrix, it has a total of $n(n + 1)/2$ parameters, and its differential is defined as

$$dS = \prod_{i \le j} ds_{ij} \ .$$

Similarly, if H is $n \times n$ hermitian, then the strictly upper triangular part contains $n(n-1)$ parameters (the real and the imaginary parts), while the diagonal contributes an additional n (since the diagonal is real). Finally, if D is $n \times n$ quaternion self-dual, the off-diagonal part contributes $2n(n - 1)$ parameters, and the real diagonal adds another n, making the total of $2n^2 - n$.

If N is a non-symmetric rectangular matrix, it contains βmn parameters, and

$$dN = \prod_{i,j} dn_{ij} = \prod_{i,j} \prod_{k=1}^{\beta} dn_{ij}^k \ ,$$

with $\prod_{k=1}^{\beta} dn_{ij}^k$ the product of differentials of the β real parameters contained in n_{ij}.

If L is lower triangular, then it has a total of $\beta n(n + 1)/2$ parameters, and its differential is the product of differentials of all:

$$dL = \prod_{i \le j} dl_{ij} = \prod_{i \le j} \prod_{k=1}^{\beta} dl_{ij}^k \ ,$$

with the understanding that $\prod_{k=1}^{\beta} dl_{ij}^k$ is the product of differentials of the real parameters contained in l_{ij}.

If Λ is a diagonal matrix, it has βn parameters, and

$$d\Lambda = \prod_{i=1}^{n} \prod_{k=1}^{\beta} d\lambda_{ii}^k \ .$$

Differentials of orthogonal/unitary/symplectic matrices are slightly more complicated, because the matrix property includes a special (and intricate) interdependence among the matrix entries.

Let Q be an $m \times n$ orthogonal/unitary/symplectic matrix; the group of such matrices defines a submanifold $V_{m,n}^{\beta}$ of dimension $(\beta mn - \beta m(m-1)/2 - m)$ over $\mathbb{R}^{\beta mn}$. This is called a Stiefel manifold (see [28, 70]). The correct differential on this manifold is $Q'dQ$, where $'$ stands for conjugate transpose. This differential represents the surface element on this manifold. It can be shown that

$$\int_{V_{m,n}^{\beta}} (Q'dQ) \quad = \quad \frac{2^m \ \pi^{mn\beta/2}}{\Gamma_m^{\beta}(\frac{n\beta}{2})} \quad , \tag{4.1}$$

where Γ_m^{β} is defined as in 3.1. For $\beta = 1, 2$ we have multiple references, for example Edelman [24]; we have computed the $\beta = 4$ case ourselves, though we believe it probably appears in the literature.

The cases $m = 1$ and $m = n$ are special; $V_{1,n}^{\beta}$ represents the unit sphere in $\mathbb{R}^{\beta n}$ ($Q'dQ$ is in this case the surface element), and when $m = n$, $Q'dQ/\text{Volume}(V_{n,n}^{\beta})$ is called the (real, complex, or quaternion) *Haar measure*. The Haar measure is invariant under orthogonal/unitary/symplectic transformations.

We now have all the definitions that we need; first we list those factorizations which are well-known enough to be included in virtually any comprehensive book on Random Matrices. The originality of this approach comes from the fact that, to the best of our knowledge, this is the first attempt to provide a unified treatment for all three cases, using the β parameter, where appropriate. We have also included parameter counts and clear descriptions of the matrix properties. Most of the references we list are for the real and complex cases; most of the quaternion cases we have computed ourselves because, though similar to the real and complex, they do not appear in the literature.

Notation. *Throughout this chapter, we use the notation $'$ for conjugate transpose. This follows MATLAB notation.*

- **LU decomposition** ($A = LU$).

  Valid for all three cases ($\beta = 1, 2, 4$). All matrices are $n \times n$, L and U^T are

lower triangular, $l_{ii} = 1$, for all $1 \leq i \leq n$. Parameter count:

$$\beta n^2 = \beta \frac{n(n-1)}{2} + \beta \frac{n(n+1)}{2} \ .$$

Jacobian:

$$dA = \qquad \prod_{i=1}^{n} |u_{ii}|^{\beta(n-i)} \ dL \ dU.$$

References: real and complex cases, Muirhead [70].

- **LL′ (Cholesky) decomposition** ($A = LL'$).

  Valid for all three cases ($\beta = 1, 2, 4$). A must be square $n \times n$ and positive definite, and L is lower triangular with l_{ii} real for all $1 \leq i \leq n$. Parameter count:

$$\beta \frac{n(n-1)}{2} + n = \beta \frac{n(n-1)}{2} + n$$

  (the diagonals of A and L are real).

  Jacobian:

$$dA = \qquad 2^n \ \prod_{i=1}^{n} |l_{ii}|^{\beta(n-i)+1} \ dL.$$

  References: real and complex cases, Edelman [24].

- **QR (least squares) decomposition** ($A = QR$).

  Valid for all three cases ($\beta = 1, 2, 4$). Q is orthogonal/unitary/symplectic, R is upper triangular. A and Q are $m \times n$ (assume $m \geq n$), R is $n \times n$. Parameter count:

$$\beta mn = \beta mn - \beta \frac{n(n-1)}{2} - n + \beta \frac{n(n-1)}{2} + n$$

  The parameter count for the orthogonal matrix is the dimension of the Stiefel manifold $V_{n,m}$, since this time $m \geq n$.

  Jacobian:

$$dA = \prod_{i=1}^{n} r_{ii}^{\beta(n-i+1)-1} \ dR \ (Q'dQ) \ .$$

  References: real and complex cases, Edelman [24].

- **LQ (transpose least squares) decomposition** ($A = LQ$).

  Valid for all three cases ($\beta = 1, 2, 4$). Q is orthogonal/unitary/symplectic, L is lower triangular. A and Q are $m \times n$ (assume $m \leq n$), L is $m \times m$. Parameter count:

$$\beta mn = \beta mn - \beta\frac{m(m-1)}{2} - m + \beta\frac{m(m-1)}{2} + m \ .$$

  Jacobian:

$$dA = \prod_{i=1}^{m} l_{ii}^{\beta(m-i+1)-1} \ dL \ (Q'dQ) \ .$$

  Reference: real and complex cases, Edelman [24].

- **$\mathbf{Q\Lambda Q'}$ (symmetric eigenvalue) decomposition** ($A = Q\Lambda Q'$).

  Valid for all three cases ($\beta = 1, 2, 4$). A is $n \times n$ symmetric/hermitian/self-dual, Q is $n \times n$ and orthogonal/unitary/symplectic, Λ is $n \times n$ diagonal and real. To make the decomposition unique, we must fix the phases of the columns of Q (that eliminates $(\beta - 1)n$ parameters) and order the eigenvalues. Parameter count:

$$\beta\frac{n(n-1)}{2} + n = \beta\frac{n(n+1)}{2} - n - (\beta - 1)n + n \ .$$

  Jacobian:

$$dA = \prod_{i<j} (\lambda_i - \lambda_j)^{\beta} \ d\Lambda \ (Q'dQ) \ .$$

  References: real, complex, and quaternion cases, Mehta [66].

- **$\mathbf{U\Sigma V'}$ (singular value) decomposition** ($A = U\Sigma V'$).

  Valid for all three cases ($\beta = 1, 2, 4$). A is $m \times n$, U is $m \times n$ orthogonal/unitary/symplectic, V is $n \times n$ orthogonal/unitary/symplectic, Σ is $n \times n$ diagonal, positive, and real (suppose $m \geq n$). Again, to make the decomposition unique, we need to fix the phases on the columns of U (getting rid of $(\beta - 1)n$ parameters) and order the singular values. Parameter count:

$$\beta mn = \beta mn - \beta\frac{n(n-1)}{2} - n - (\beta - 1)n + n + \beta\frac{n(n+1)}{2} - n$$

Jacobian:

$$dA = \prod_{i<j}(\sigma_i^2 - \sigma_j^2)^\beta \prod_{i=1}^{n} \sigma_i^{\beta(m-n+1)-1} \ (U'dU)\ d\Sigma\ (V'dV)\ .$$

References: real and complex cases, Muirhead [70].

In addition to these Jacobians, we include a few that are less familiar from the literature. Though it stands to reason that these might have been computed before, we found no references for them, and recomputed them ourselves, and we briefly indicate how below.

- **LDL′ (generalized Cholesky) decomposition** ($A = LDL'$).

  Valid for all three cases ($\beta = 1, 2, 4$). A must be square $n \times n$ real/hermitian/self-dual, L is lower triangular with $l_{ii} = 1$ for all $1 \leq i \leq n$, and D is $n \times n$ real diagonal. Parameter count:

$$\beta\frac{n(n-1)}{2} + n = \beta\frac{n(n-1)}{2} + n$$

  (the diagonals of A and L are real). Note that A is not required to be positive definite.

  Jacobian:

$$dA = \prod_{i=1}^{n} |d_{ii}|^{\beta(n-i)}\ dL\ dD.$$

  The proof follows immediately from the LL' decomposition.

- **QS (polar) decomposition** ($A = QS$).

  Valid for all three cases ($\beta = 1, 2, 4$). A is $m \times n$ ($m \leq n$), Q is $m \times n$ orthogonal/unitary/symplectic, S is $n \times n$ positive definite. Parameter count

$$\beta mn = \beta mn - \beta\frac{n(n-1)}{2} - n + \beta\frac{n(n-1)}{2} + n\ .$$

  Jacobian:

$$dA = \prod_{i<j}(\sigma_i + \sigma_j)^\beta \prod_{i=1}^{m} \sigma_i^{\beta(n-m+1)-1} \ (Q'dQ)\ dS\ ,$$

where σ_i are the singular values of A and the eigenvalues of S.

The proof follows from the fact that, if $A = U\Sigma V'$ is the singular value decomposition for A, then $Q = UV'$ and $S = V\Sigma V'$, which also explains why the Jacobian is written in terms of the singular values σ_i of A.

- **CS decomposition** (see [35]).

  Valid for all three cases ($\beta = 1, 2, 4$). Q is $n \times n$ orthogonal/unitary/symplectic. Then for any $k + j = n$, $p = k - j \geq 0$, the decomposition is

$$Q = \begin{pmatrix} U_{11} & U_{12} & 0 \\ U_{21} & U_{22} & 0 \\ 0 & 0 & U_2 \end{pmatrix} \begin{pmatrix} I_p & 0 & 0 \\ 0 & C & S \\ 0 & S & -C \end{pmatrix} \begin{pmatrix} V_{11}' & V_{12}' & 0 \\ V_{21}' & V_{22}' & 0 \\ 0 & 0 & V_2' \end{pmatrix},$$

  such that U_2, V_2 are $j \times j$ orthogonal/unitary/symplectic, $\begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}$ and $\begin{pmatrix} V_{11}' & V_{12}' \\ V_{21}' & V_{22}' \end{pmatrix}$ are $k \times k$ orthogonal/unitary/symplectic, with U_{11} and V_{11} being $p \times p$, and C and S are $j \times j$ real, positive, and diagonal, and $C^2 + S^2 = I_j$.

  Parameter count:

$$\begin{aligned} \beta\frac{n(n+1)}{2} - n &= \Big(\beta j(j+1) - (\beta-1)j\Big) + j \; + \\ &\quad + \; \Big(\beta k(k+1) - k - \beta\frac{p(p+1)}{2} + p\Big) \end{aligned}$$

  This count is a little special, and so we will give it in detail. The first part in the right hand side counts the number of parameters in U_2 and V_2 (which are independent), and accounts for the choice of phases, since $U_2 C V_2'$ is the SVD of the bottom $j \times j$ part of Q. The second term accounts for the j parameters of C (which are the same in S).

  Finally, the last parenthesis counts the number of parameters in $\begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}$ and $\begin{pmatrix} V_{11}' & V_{12}' \\ V_{21}' & V_{22}' \end{pmatrix}$, and accounts for the fact that $\begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} \begin{pmatrix} V_{11}' & V_{12}' \end{pmatrix}$ is

determined by the other matrices in the decomposition and thus, one has to subtract the missing number of parameters. Since the number of parameters in either $\begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix}$ or $\begin{pmatrix} V'_{11} & V'_{12} \end{pmatrix}$, given the rest of the orthogonal matrices, is $\beta p(p+1)/2 - p$, the count follows.

Note that the choice of phases in U_2, V_2 determines all phases in the decomposition.

Now let $\theta_i \in (0, \frac{\pi}{2})$, $q \leq i \leq j$ be the angles such that $C = \text{diag}(\cos(\theta_1), \ldots, \cos(\theta_j))$, and $S = \text{diag}(\sin(\theta_1), \ldots, \sin(\theta_j))$. To ensure uniqueness of the decomposition we order the angles, $\theta_i \geq \theta_j$, for all $i \leq j$.

For notational purposes, let $U_1 = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}$, and $V'_1 = \begin{pmatrix} V'_{21} & V'_{22} \end{pmatrix}$.

Jacobian:

$$\begin{aligned} (Q'dQ) &= \prod_{i<j} \sin(\theta_i - \theta_j)^\beta \ \sin(\theta_i + \theta_j)^\beta \ \prod_{i=1}^{j} \cos(\theta_i)^{\beta-1} \ \sin(\theta_i) \ d\theta \quad \times \\ &\quad \times \quad (U'_1 dU_1) \ (U'_2 dU_2) \ (V'_1 dV_1) \ (V'_2 dV_2) \ . \end{aligned}$$

The proof follows from the decomposition itself. Note that $V_1 \in V_{j,k}^\beta$.

- **Tridiagonal $\mathbf{Q\Lambda Q'}$ (eigenvalue) decomposition** ($T = Q\Lambda Q'$).

  Valid for real matrices. T is an $n \times n$ tridiagonal symmetric matrix as in (3.3), Q is an orthogonal $n \times n$ matrix, and Λ is diagonal. To make the factorization unique, we impose the condition that the first row of Q is all positive. The number of independent parameters in Q is $n-1$ and the can be seen as being all in the first row q of Q. The rest of Q can be determined from the orthogonality constraints, the tridiagonal symmetric constraints on A, and from Λ. Parameter count:

$$2n - 1 = n - 1 + n \ .$$

  Jacobian:

$$dT = \frac{\prod_{i=1}^{n-1} b_i}{\prod_{i=1}^{n} q_i} \ (dq) \ d\Lambda.$$

Note that the Jacobian is written as a combination of parameters from T and q, the first row of Q, and (dq) is the surface element on the sphere.

For a proof, the reader is referred to Section 4.3.

- **Tridiagonal BB′ (Cholesky) decomposition** ($T = BB'$).

  Valid for real matrices. T is an $n \times n$ real positive definite matrix following notation (3.3), B is an $n \times n$ real bidiagonal matrix as in (3.4). Parameter count:

  $$n - 1 = n - 1 \ .$$

  Jacobian:

  $$dT = 2^m x_1 \prod_{i=2}^{n} x_i^2 \ dB \ . \tag{4.2}$$

  The proof follows from (3.5, 3.6, 3.7).

4.2 An expression for the Vandermonde

Given a tridiagonal matrix T defined by the diagonal $a = (a_n, \ldots, a_1)$ and subdiagonal $b = (b_{n-1}, \ldots, b_1)$, with each b_i positive, let $T = Q\Lambda Q'$ be the eigendecomposition of T. Let q be the first row of Q and $\lambda = \text{diag}\ (\Lambda)$.

Lemma 4.2.1. *Under the assumptions above, starting from q and λ, one can uniquely reconstruct Q and T.*

This is a special case of a more general result (Theorem 7.2.1, Parlett [77]).

Our next result establishes a formula for the Vandermonde determinant of the eigenvalues of a tridiagonal matrix (denoted by $\Delta(\Lambda)$, and equal to $\prod_{i<j}(\lambda_i - \lambda_j)$).

Lemma 4.2.2. *The Vandermonde determinant for the ordered eigenvalues of a symmetric tridiagonal matrix with positive sub-diagonal $b = (b_{n-1}, \ldots, b_1)$ is given by*

$$\Delta(\Lambda) = \prod_{i<j}(\lambda_i - \lambda_j) = \frac{\prod_{i=1}^{n-1} b_i^i}{\prod_{i=1}^{n} q_i} \ ,$$

where $(q_1, \ldots, q_n)$ is the first row of the eigenvector matrix.

Proof. Let $\lambda_i^{(k)}$, $i = 1 \ldots k$, be the eigenvalues of the $k \times k$ lower right-corner submatrix of T. Then $P_k(x) = \prod_{i=1}^{k}(x - \lambda_i^{(k)})$ is the associated characteristic polynomial of that submatrix.

For $k = 1, \ldots, n$ we have the three-term recurrence

$$P_k(x) = (x - a_k)P_{k-1}(x) - b_{k-1}^2 P_{k-2}(x) \ , \tag{4.3}$$

and the two-term identity

$$\prod_{\substack{1 \le i \le k \\ 1 \le j \le k-1}} |\lambda_i^{(k)} - \lambda_j^{(k-1)}| = \prod_{i=1}^{k} |P_{k-1}(\lambda_i^{(k)})| = \prod_{j=1}^{k-1} |P_k(\lambda_j^{(k-1)})| \ . \tag{4.4}$$

From (4.3) we get

$$|\prod_{i=1}^{k-1} P_k(\lambda_i^{(k-1)})| = b_{k-1}^{2(k-1)} |\prod_{i=1}^{k-1} P_{k-2}(\lambda_i^{(k-1)})| \ . \tag{4.5}$$

By repeatedly applying (4.3) and (4.4) we obtain

$$\begin{aligned}
\prod_{i=1}^{n-1} |P_n(\lambda_i^{(n-1)})| &= b_{n-1}^{2(n-1)} \prod_{i=1}^{n-2} |P_{n-1}(\lambda_i^{(n-2)})| \\
&= b_{n-1}^{2(n-1)} \ b_{n-2}^{2(n-2)} \ |\prod_{i=1}^{n-2} P_{n-3}(\lambda_i^{(n-2)})| \\
&= \ldots \\
&= \prod_{i=1}^{n-1} b_i^{2i} \ .
\end{aligned}$$

Finally, we use the following formula due to Paige, found in [77] as the more general Theorem 7.9.2:

$$q_i^2 = \left| \frac{P_{n-1}(\lambda_i)}{P_n'(\lambda_i)} \right| = \left| \frac{P_{n-1}(\lambda_i^{(n)})}{P_n'(\lambda_i^{(n)})} \right| \ .$$

It follows that

$$\prod_{i=1}^{n} q_i^2 = \frac{\prod_{i=1}^{n} |P_{n-1}(\lambda_i^{(n)})|}{\Delta(\Lambda)^2} = \frac{\prod_{i=1}^{n-1} b_i^{2i}}{\Delta(\Lambda)^2} ,$$

which proves the result. □

Remark 4.2.3. *The Vandermonde determinant formula of Lemma 4.2.2 can also be obtained from the Heine formula for the square of the Vandermonde determinant, integrated against a weight function, as presented in Szegö [78, page 27, (2.2.11)].*

4.3 Derivation of the tridiagonal $T = Q\Lambda Q'$ Jacobian

This section presents a Random Matrix derivation for the Jacobian of the transformation $T = Q\Lambda Q'$, where T is a symmetric tridiagonal matrix. This result is found in Lemma 4.3.1.

The derivation is done by relating the tridiagonal and diagonal forms of a GOE matrix, as illustrated in Figure 4-1.

In this section, we will use the notations given in (3.3).

Lemma 4.3.1. *The Jacobian J of the tridiagonal transformation $T = Q\Lambda Q'$ can be written as*

$$J = \frac{\prod_{i=1}^{n-1} b_i}{\prod_{i=1}^{n} q_i} .$$

Proof. To obtain the Jacobian, we shall study the transformation from GOE to 1-Hermite ensemble (see Figure 4-1).

Let T be a 1-Hermite matrix (distributed like (2.1) with $\beta = 1$). The eigenvalues of T are distributed as the eigenvalues of a symmetric GOE matrix A, from which T can be obtained by tridiagonal reduction (for an explicit proof, see [21]).

We recall the notation of (3.3): denote by $a = (a_n, a_{n-1}, \ldots, a_1)$ the diagonal of T, and by $b = (b_{n-1}, b_{n-2}, \ldots, b_1)$ the subdiagonal of T. The joint element density for

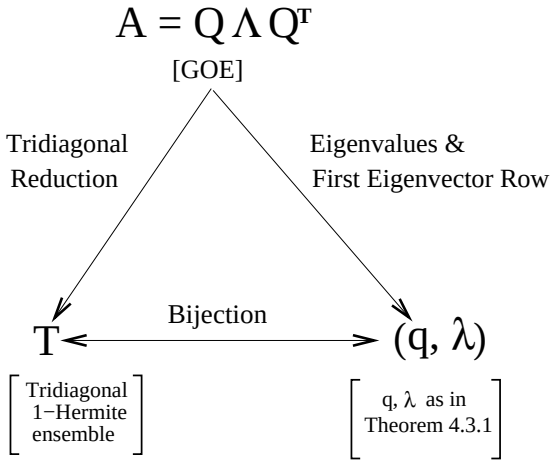

Figure 4-1: A dense symmetric matrix T can be tridiagonalized (left side) or diagonalized (right side). In brackets, we provide the distributions starting with that of A (GOE).

T is then

$$\mu(a,b) \;=\; c_{a,b}\; e^{-\frac{1}{2}\sum_{i=1}^{n} a_i^2} \prod_{i=1}^{n} b_i^{i-1} e^{-\sum_{i=1}^{n} b_i^2} \quad , \qquad \text{where} \qquad c_{a,b} = \frac{2^{n-1}}{(2\pi)^{n/2} \prod_{i=1}^{n-1} \Gamma(\frac{i}{2})} \; .$$

Let

$$da = \wedge_{i=1}^{n} da_i, \quad db = \wedge_{i=1}^{n-1} db_i, \quad d\lambda = \wedge_{i=1}^{n} \lambda_i, \tag{4.6}$$

and (dq) be the surface element of the n-dimensional sphere. Let $\mu(a(q,\lambda), b(q,\lambda))$ be the expression for $\mu(a,b)$ in the new variables q, λ. We have that

$$\mu\,(a,b) \quad da\; db \;=\; \mu\,(a(q,\lambda), b(q,\lambda)) \quad J\;(dq)\;d\lambda \quad \equiv\; \nu(q,\lambda) \;\; (dq)\;d\lambda. \tag{4.7}$$

We recall that matrices from the GOE have the following properties:

Property 1. The joint eigenvalue density is $c_H^1\; |\Delta(\Lambda)| e^{-\frac{1}{2}\sum_i \lambda_i^2}$ [66];

Property 2. The first row of the eigenvector matrix, q, is distributed uniformly on the sphere, and it is independent of the eigenvalues.

We combine Properties 1 and 2 to get the joint p.d.f. $\nu(q,\lambda)$ of the eigenvalues and first eigenvector row of a GOE matrix, and rewrite it as

$$\nu(q,\lambda)\; dq\; d\lambda \;=\; n!\; c_H^1\; \frac{2^{n-1}\Gamma(\frac{n}{2})}{\pi^{n/2}}\; \Delta(\Lambda)\; e^{-\frac{1}{2}\sum_i \lambda_i^2}\; dq\; d\lambda.$$

We have introduced the $n!$ and removed the absolute value from the Vandermonde determinant, because the eigenvalues are ordered. We have also included the distribution of q (as mentioned in Property 2, it is uniform, but only on the all-positive 2^{-n}th of the sphere because of the condition $q_i \geq 0$).

Since orthogonal transformations do not change the Frobenius norm $||A||_F = \sum_{i,j=1}^{n} a_{ij}^2$ of a matrix A, from (4.7), it follows that

$$J = \frac{\nu(q,\lambda)}{\mu(a,b)} = \frac{n!\ c_H^1 \frac{2^{n-1}\Gamma(\frac{n}{2})}{\pi^{n/2}}}{c_{a,b}} \frac{\Delta(\Lambda)}{\prod_{i=1}^{n} b_i^{i-1}} .$$

All constants cancel, and by Lemma 4.2.2 we obtain

$$J = \frac{\prod_{i=1}^{n-1} b_i}{\prod_{i=1}^{n} q_i} .$$

□

Note that we have not expressed $\mu(a,b)$ in terms of q and λ in the above, and have thus obtained the expression for the Jacobian neither in the variables q and λ, nor a and b, solely; but rather in a mixture of the two sets of variables. The reason for this is that of simplicity.

Remark 4.3.2. *Our derivation of the Jacobian is a true Random Matrix derivation. Alternate derivations of the Jacobian can be obtained either via symplectic maps or through direct calculation. We thank Percy Deift and Peter Forrester, respectively, for having shown them to us.*

4.4 Perturbation Theory

In Chapter 7, we make use of perturbation theory in order to compute the first-order asymptotics for the eigenvalues of the large β, fixed n, β-Hermite and β-Laguerre ensembles. We give here the lemma we need in order to do that. For a reference on perturbation theory, in particular to a more general form of the result below, see Demmel's book [18, Section 4.3].

Lemma 4.4.1. *Let A and B be $n \times n$ symmetric matrices, and let $\epsilon > 0$. Assume A has all distinct eigenvalues. Let $M = A + \epsilon B + o(\epsilon)$, where by $o(\epsilon)$ we mean a matrix in which every entry goes to 0 faster than ϵ. Let $\lambda_i(X)$ denote the ith eigenvalue of X, for every $1 \leq i \leq n$. Finally, let Q be an eigenvector matrix for A. Then*

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon}(\lambda_i(M) - \lambda_i(A)) \ = Q(:,i)^T B Q(:,i) \ ,$$

where, following MATLAB notation, $Q(:,i)$ represents the ith column of Q.

Remark 4.4.2. *Equivalently, for every $1 \leq i \leq n$,*

$$\lambda_i(M) = \lambda_i(A) + \epsilon Q(:,i)^T B Q(:,i) + o(\epsilon) \ .$$

Chapter 5

Matrix Models for general β-Hermite (Gaussian) and β-Laguerre (Wishart) ensembles

In this Chapter we present the two new random matrix models, and we prove that the eigenvalue distributions are those given by (2.1) and (2.3). The results in this chapter, together with some of the results of Chapter 4 were published in [21].

We make use of the lemmas we proved in Section 4. We begin with a motivation for the tridiagonal models.

5.1 Bidiagonalization and tridiagonalization

Most efficient algorithms for computing eigenvalues of symmetric matrices or singular values of general rectangular matrices consist of two steps. The first step is a "sparsification" of the matrix; in the symmetric case, it involves the reduction of the matrix to symmetric tridiagonal form ("tridiagonalization"), and in the case of the rectangular matrix, it involves the reduction of the matrix to bidiagonal form ("bidiagonalization"). This step is "finite", that is, in perfect arithmetic, it is achievable in finite time.

The second step is to use an iterative method to compute the eigenvalues/singular

values of the sparse model, and it is "infinite", i.e. in perfect arithmetic, in finite time, it can obtain infinitely good approximations, but no exact answer.

The bidiagonalization method is due to Golub and Kahan [34], while the tridiagonalization method is due to Householder [40].

Because of the orthogonality invariance property of the multivariate Gaussian, these two linear algebra algorithms can be applied stochastically, and it is not very hard to compute the distributions of the resulting tridiagonal/bidiagonal matrix. The following two theorems illustrate this.

Theorem 5.1.1. *If A is an $n \times n$ matrix from the GOE ($\beta = 1$), GUE ($\beta = 2$), or GSE ($\beta = 4$), then reduction of A to tridiagonal form shows that the matrix H_β from the β-Hermite ensemble (see below) has joint eigenvalue p.d.f. given by (2.1) with $\beta = 1, 2, 4$.*

We recall that the $1, 2$ and 4 β-Hermite ensembles have the joint element distribution illustrated below (tridiagonal, symmetric, and real):

$$H_\beta \sim \frac{1}{\sqrt{2}} \begin{pmatrix} N(0,2) & \chi_{(n-1)\beta} & & & & \\ \chi_{(n-1)\beta} & N(0,2) & \chi_{(n-2)\beta} & & & \\ & \ddots & \ddots & \ddots & & \\ & & \chi_{2\beta} & N(0,2) & \chi_\beta & \\ & & & \chi_\beta & N(0,2) \end{pmatrix}.$$

Since the tridiagonal reduction is done by symmetric orthogonal multiplication by a sequence of orthogonal Householder matrices, *each of which only depends on the current first column*, the subdiagonal elements have the distribution of the norms of the multivariate Gaussians (real, complex, quaternion).

Theorem 5.1.2. *If W is from $G(m,n)$ ($\beta = 1$), $G^2(m,n)$ ($\beta = 2$), or $G^4(m,n)$ ($\beta = 4$), then reduction of W to bidiagonal form shows that the matrix B_β from the $G_{\beta,a}(m)$ distribution of parameter $a = n\beta/s$ (see below) has joint singular value p.d.f. given by (2.3) with $\beta = 1, 2, 4$ and $a = n\beta/2$.*

Remark 5.1.3. *Implicitly, this shows that the Wishart real, complex, and quaternion ensembles have the same eigenvalue distributions as the β-Laguerre ensembles of parameter $a = n\beta/2$.*

References for the real, complex and quaternion cases are [80] and [92].

The reasons why bidiagonalization yields the χ structures on the diagonal and subdiagonal are again the orthogonal invariance of the multivariate Gaussian, and the distribution of the integer-valued χ.

5.2 β-Hermite (Gaussian) model

Let H_β be a random real symmetric, tridiagonal matrix whose distribution we schematically depict as

$$H_\beta \sim \frac{1}{\sqrt{2}} \begin{pmatrix} N(0,2) & \chi_{(n-1)\beta} & & & & \\ \chi_{(n-1)\beta} & N(0,2) & \chi_{(n-2)\beta} & & & \\ & \ddots & \ddots & \ddots & & \\ & & \chi_{2\beta} & N(0,2) & \chi_\beta & \\ & & & \chi_\beta & N(0,2) \end{pmatrix}$$

By this we mean that the n diagonal elements and the $n-1$ sub-diagonal elements are mutually independent, with standard normal variables on the diagonal, and $\frac{1}{\sqrt{2}}\chi_{k\beta}$ on the sub-diagonal.

Theorem 5.2.1. *Let $H_\beta = Q\Lambda Q^T$ be the eigendecomposition of H_β; fix the signs of the first row of Q to be non-negative and order the eigenvalues in increasing order on the diagonal of Λ. Let λ be the diagonal of Λ. Then λ and q, the first row of Q, are independent. Furthermore, the joint density of the eigenvalues is*

$$f_\beta(\lambda) = c_H^\beta \prod_{i<j} |\lambda_i - \lambda_j|^\beta \, e^{-\frac{1}{2}\sum_{i=1}^n \lambda_i^2} \quad = \quad c_H^\beta \, |\Delta(\Lambda)|^\beta \, e^{-\frac{1}{2}\sum_{i=1}^n \lambda_i^2} \, ,$$

and $q = (q_1, \ldots, q_n)$ is distributed as $(\chi_\beta, \ldots, \chi_\beta)$, normalized to unit length; c_h^β is given by (2.2).

Proof. Just as before, we denote by $a = (a_n, \ldots, a_1)$ the diagonal of H_β, and by $b = (b_{n-1}, \ldots, b_1)$ the sub-diagonal. The differentials $da, db, dq, d\lambda$ are the same as in statement (4.6). For general β, we have that

$$(dH_\beta) \equiv \mu(a,b)\, da\, db = c_{a,b} \prod_{k=1}^{n-1} b_k^{k\beta-1} e^{-\frac{1}{2}||T_1||_F}\, da\, db = c_{a,b}\, J \prod_{k=1}^{n-1} b_k^{k\beta-1} e^{-\frac{1}{2}||T_1||_F}\, dq\, d\lambda\,,$$

where

$$c_{a,b} = \frac{2^{n-1}}{(2\pi)^{n/2} \prod_{k=1}^{n-1} \Gamma\left(\frac{\beta}{2}k\right)}\,.$$

With the help of Lemmas 4.2.1 and 4.3.1 this identity becomes

$$\begin{aligned}(dH_\beta) &= c_{a,b}\, \frac{\prod_{k=1}^{n-1} b_k}{\prod_{k=1}^{n} q_k} \prod_{k=1}^{n-1} b_k^{k\beta-1} e^{-\frac{1}{2}||T_1||_F}\, dq\, d\lambda &(5.1)\\ &= c_{a,b}\, \frac{\prod_{k=1}^{n-1} b_k^{k\beta}}{\prod_{i=1}^{n} q_i^{\beta}} \prod_{i=1}^{n} q_i^{\beta-1} e^{-\frac{1}{2}\sum_i \lambda_i^2}\, dq\, d\lambda\,. &(5.2)\end{aligned}$$

Thus

$$(dH_\beta) = \left(c_q^\beta \prod_{i=1}^{n} q_i^{\beta-1}\, dq\right)\ \left(n!\ c_H^\beta\ \Delta(\Lambda)^\beta e^{-\frac{1}{2}\sum_i \lambda_i^2}\, d\lambda\right)\,.$$

Since the joint density function of q and λ separates, q and λ are independent. Moreover, once we drop the ordering imposed on the eigenvalues, it follows that the joint eigenvalue density of H_β is $c_H^\beta\ |\Delta(\Lambda)|^\beta e^{-\frac{1}{2}\sum_i \lambda_i^2}$, and q is distributed as $(\chi_\beta, \ldots, \chi_\beta)$, normalized to unit length. From (5.2), it also follows that

$$c_q^\beta = \frac{2^{n-1}\Gamma(\frac{\beta}{2}n)}{\left[\Gamma(\frac{\beta}{2})\right]^n}\,. \tag{5.3}$$

□

5.3 β-Laguerre (Wishart) model

Let

$$B_\beta \sim \begin{pmatrix} \chi_{2a} & & & \\ \chi_{\beta(m-1)} & \chi_{2a-\beta} & & \\ & \ddots & \ddots & \\ & & \chi_\beta & \chi_{2a-\beta(m-1)} \end{pmatrix}\,,$$

by this meaning that all of the $2m-1$ diagonal and subdiagonal elements are mutually independent with the corresponding χ distribution.

Let $L_\beta = B_\beta B_\beta^T$ be the corresponding tridiagonal matrix.

Theorem 5.3.1. *Let $L_\beta = Q\Lambda Q^T$ be the eigendecomposition of L_β; fix the signs of the first row of Q to be non-negative and order the eigenvalues increasingly on the diagonal of Λ. Let λ be the diagonal of Λ. Then λ and the first row q of Q are independent. Furthermore, the joint density of the eigenvalues is*

$$f_\beta(\lambda) = c_L^{\beta,a} \ |\Delta(\Lambda)|^\beta \ \prod_{i=1}^{m} \lambda_i^{a-p} e^{-\sum_{i=1}^n \lambda_i/2} \ ,$$

where $p = 1 + \frac{\beta}{2}(m-1)$, and q is distributed as $(\chi_\beta, \ldots, \chi_\beta)$ normalized to unit length; $c_L^{\beta,a}$ is given by (2.4).

Proof. We will use throughout the proof the results of Lemma 4.2.1, Lemma 4.3.1, and(4.2), which are true in the context of tridiagonal symmetric matrices with positive sub-diagonal entries. By definition, L_β is such a matrix.

We will again use the notations of Lemma 4.3.1, (4.2), and (4.6) for the differentials da, db, dq, $d\lambda$, dx, and dy.

We define (dB_β) to be the joint element distribution on B_β

$$(dB_\beta) \ \equiv \ \mu(x,y)\ dx\ dy \ = \ c_{x,y} \ \prod_{i=0}^{m-1} x_{m-i}^{a-\beta i-1} e^{-x_i^2/2} \ \prod_{i=1}^{m-1} y_i^{\beta i-1} e^{-y_i^2/2}\ dx\ dy \ .$$

By (4.2), the joint element distribution on L_β is

$$\begin{aligned} (dL_\beta) \ &\equiv \ J_{B\to T}^{-1} \mu(x,y)\ dx\ dy & (5.4)\\ &= \ 2^{-m} c_{x,y}\ x_1^{2a-\beta(m-1)-2} e^{-x_1^2/2} \prod_{i=0}^{m-2} x_{m-i}^{a-\beta i-3} e^{-x_i^2/2} \prod_{i=1}^{m-1} y_i^{\beta i-1} e^{-y_i^2/2}\ dx\ dy & (5.5) \end{aligned}$$

where

$$c_{x,y} = \frac{\prod_{i=1}^{m-1} \Gamma(i\frac{\beta}{2}) \prod_{i=1}^{m} \Gamma(a - \frac{\beta}{2}(i-1))}{2^{2m-1}} \ .$$

We rewrite (5.5) in terms of x, y, λ, and q:

$$\begin{aligned}(dL_\beta) &= 2^{-m} c_{x,y}\, e^{-\sum_{i=1}^{m} x_i^2/2} e^{-\sum_{i=1}^{m-1} y_i^2/2} \frac{\prod_{i=1}^{m-1}(x_{i+1}y_i)}{\prod_{i=1}^{m} q_i} x_1^{2a-\beta(m-1)-2} \times \\ &\quad \times \prod_{i=0}^{m-2} x_{m-i}^{2a-\beta(m-i)-3} \prod_{i=1}^{m-1} y_i^{\beta i-1}\; dq\; d\lambda \\ &= 2^{-m} c_{x,y}\, e^{-\sum_{i=1}^{m} x_i^2/2} e^{-\sum_{i=1}^{m-1} y_i^2/2} \frac{\prod_{i=0}^{m-1} x_{m-i}^{2a-\beta(m-i)-2} \prod_{i=1}^{m-1} y_i^{\beta i}}{\prod_{i=1}^{m} q_i}\; dq\; d\lambda.\end{aligned}$$

Since the Vandermonde with respect to b and q and the ordered eigenvalues λ can be written as

$$\Delta(\Lambda) = \frac{\prod_{i=1}^{m-1} b_i^i}{\prod_{i=1}^{m} q_i}\ ,$$

it follows that

$$\Delta(\Lambda) = \frac{\prod_{i=1}^{m-1} \Big(x_{i+1}y_i\Big)^i}{\prod_{i=1}^{m} q_i}\ .$$

Then

$$\begin{aligned}(dL_\beta) &= 2^{-m} c_{x,y}\, e^{-\sum_{i=0}^{m-1} x_{n-i}^2/2} e^{-\sum_{i=1}^{m-1} y_i^2/2} \frac{\prod_{i=1}^{m-1} \Big(x_{i+1}y_i\Big)^{\beta i}}{\prod_{i=1}^{m} q_i^{\beta}} \prod_{i=1}^{m-1} q_i^{\beta-1} \prod_{i=0}^{m-1} x_{m-i}^{2a-\beta(m-1)-2}\; dq\; d\lambda \\ &= 2^{-m} c_{x,y}\, e^{-\sum_{i=0}^{m-1} x_{n-i}^2/2} e^{-\sum_{i=1}^{m-1} y_i^2/2} \Delta(\Lambda)^\beta \prod_{i=1}^{m-1} q_i^{\beta-1} \left(\prod_{i=0}^{m-1} x_{m-i}\right)^{2a-\beta(m-1)-2} dq\; d\lambda\ .\end{aligned}$$

The trace and the determinant are invariant under orthogonal similarity transformations, so $\mathrm{tr}(L_\beta) = \mathrm{tr}(\Lambda)$, and $\det(L_\beta) = \det(\Lambda)$. This is equivalent to

$$\begin{aligned}\sum_{i=0}^{m-1} x_{m-i}^2 + \sum_{i=1}^{m-1} y_i^2 &= \sum_{i=1}^{m} \lambda_i\ , \\ \prod_{i=0}^{m-1} x_{m-i}^2 &= \prod_{i=1}^{m} \lambda_i\ .\end{aligned}$$

Using this, and substituting p for $1 + \frac{\beta}{2}(m-1)$, we obtain

$$(dL_\beta) = \left(c_q^\beta \prod_{i=1}^{m-1} q_i^{\beta-1}\; dq\right) \left(m!\; c_L^{\beta,a} e^{-\sum_{i=1}^{m} \lambda_i/2} \Delta(\Lambda)^\beta \prod_{i=1}^{m} \lambda_i^{a-p}\; d\lambda\right)\ ,$$

where c_q^β is given in (5.3).

From the above we see that q and λ are independent, and once we drop the ordering the joint eigenvalue density is given by the β-Laguerre ensemble of parameter a, while q is distributed like a normalized vector of χ_β's.

This concludes the proof of Theorem 5.3.1. □

5.4 A few immediate applications

5.4.1 A new proof for Hermite and Laguerre forms of the Selberg Integral

Here we present a quick proof for the Hermite and Laguerre forms of the Selberg Integral [66], using the β-Hermite, respectively, β-Laguerre ensembles.

The Hermite Selberg integral is

$$I_H(\beta, n) \equiv \int_{\mathbb{R}^n} |\Delta(\Lambda)|^\beta e^{-\sum_{i=1}^n \lambda_i^2/2}\, d\lambda$$

We have that

$$I_H(\beta, n) \;=\; n! \left(\int_{0 \le \lambda_1 \le \ldots \le \lambda_n < \infty} \Delta(\Lambda)^\beta e^{-\sum_{i=1}^n \lambda_i^2/2}\, d\lambda \right) \quad \left(c_q^\beta \int_{S_+^{n-1}} \prod_{i=1}^n q_i^{\beta-1}\, dq \right) ,$$

where c_q^β is as in (5.3). We introduce the $n!$ because in the first integral we have ordered the eigenvalues; S_+^{n-1} signifies that all q_i are positive.

Note that c_q^β can easily be computed independently of the β-Hermite ensembles.

Using the formula for the Vandermonde given by Lemma 4.2.2, the formula for the Jacobian J given in Lemma 4.3.1, and the fact that the Frobenius norm of a matrix in the tridiagonal 1-Hermite ensemble is the same as the Frobenius norm of its eigenvalue matrix, one obtains

$$\begin{aligned} I_H(\beta, n) &= n!\; c_q^\beta \int_{\mathbb{R}^n \times (0,\infty)^{n-1}} \frac{\prod_{i=1}^n q_i}{\prod_{i=1}^{n-1} b_i} \frac{\prod_{i=1}^{n-1} b_i^{\beta i}}{\prod_{i=1}^n q_i^\beta} \prod_{i=1}^n q_i^{\beta-1} e^{-\sum_{i=1}^n b_i^2 - \sum_{i=1}^n a_i^2/2}\, da\, db \\ &= n!\; c_q^\beta (2\pi)^{n/2} \prod_{i=1}^{n-1} \int_{(0,\infty)} b_i^{\beta i - 1} e^{-b_i^2}\, db_i \\ &= n!\; \frac{2^{n-1}\Gamma(\frac{\beta}{2}n)}{\left(\Gamma(\frac{\beta}{2})\right)^n} (2\pi)^{n/2} \prod_{i=1}^{n-1} \frac{\Gamma(\frac{\beta}{2}i)}{2} \;=\; \frac{1}{c_H^\beta} . \end{aligned}$$

The same reasoning yields the Laguerre Selberg Integral formula

$$I_L^{\beta,a,n} = \frac{1}{c_L^{\beta,a}} \ .$$

5.4.2 The expected characteristic polynomial

The result below might be seen as an extension of the classical Heine theorem (see Szegö [78, page 27, (2.2.11)] and Deift [15, Proposition 3.8]) which has $\beta = 2$. Note that for $\beta \neq 2$, $\Delta(\Lambda)^\beta$ can no longer be written as the determinant of a Vandermonde matrix times its transpose, and the proof cannot be adapted.

The same result is found in a slightly more general form in [24], and its Jacobi case was first derived by Aomoto [4].

Theorem 5.4.1. *The expected characteristic polynomial $P_n(y) = \det(yI_n - S)$ over S in the β-Hermite and β-Laguerre ensembles are proportional to*

$$H_n\left(\frac{y}{\sqrt{2\beta}}\right), \qquad \text{and} \qquad L_n^{\frac{2a}{\beta}-n}\left(\frac{y}{2\beta}\right) \ ,$$

respectively. Here H_n and $L_n^{\frac{2a}{\beta}-n}$ are the Hermite, respectively Laguerre, polynomials, and the constant of proportionality accounts for the fact that $P_n(y)$ is monic.

Proof. Both formulas follow immediately from the 3-term recurrence for the characteristic polynomial of a tridiagonal matrix (see formula (4.3)) and from the independence of the variables involved in the recurrence. □

5.4.3 Expected values of symmetric polynomials

Using the three-term recurrence for the characteristic polynomial of a tridiagonal matrix, we obtain Theorem 5.4.2.

Theorem 5.4.2. *Let p be any fixed (independent of β) multivariate symmetric polynomial on n variables. Then the expected value of p over the β-Hermite or β-Laguerre ensembles is a polynomial in β.*

We remark that it is difficult to see this from the eigenvalue density.

Proof. The elementary symmetric functions

$$e_i(x_1, x_2, \ldots, x_n) = \sum_{1 \le j_1 < \ldots < j_i \le n} x_{j_1} x_{j_2} \ldots x_{j_i} \qquad i = 0, 1, \ldots, n \ ,$$

can be used to generate any symmetric polynomial of degree n (in particular p).

The e_i evaluated at the eigenvalues of a matrix are the coefficients of its characteristic polynomial, and hence they can be written in terms of the matrix entries. Thus p can be written as a polynomial of the $n \times n$ tridiagonal matrix entries.

To obtain the expected value of p over the β-Hermite or β-Laguerre ensemble, one can write p in terms of the corresponding matrix entries, use the symmetry to condense the expression, then replace the powers of the matrix entries by their expected values.

The diagonal matrix entries are either normal random variables in the Hermite case or sums of χ^2 random variables in the Laguerre case. The subdiagonal entries appear only raised at even powers in the e_i and hence in p (this is an immediate consequence of the three-term recurrence for the characteristic polynomial, (4.3)). Since all even moments of the involved χ distributions are polynomials in $\beta/2$, it follows that the expectation of p is a polynomial in β. □

As an easy consequence we have the following corollary.

Corollary 5.4.3. *All moments of the determinant of a β-Hermite matrix are* ***integer-coefficient*** *polynomials in $\beta/2$.*

Proof. Note that even moments of the $\chi_{\beta i}$ distribution are integer-coefficient polynomials in $\beta/2$, and that the determinant is e_n. □

5.4.4 Moments of the discriminant

The discriminant of a polynomial equation of order m is the square of the Vandermonde determinant of the m zeroes of the equation. Thus, the discriminant of the characteristic polynomial of a β-Hermite or β-Laguerre ensemble matrix is simply $D(\Lambda) = \Delta(\Lambda)^2$.

A simple calculation shows that the kth moments of $D(\Lambda)$ in the two cases are, respectively,

$$\frac{c_H^{\beta}}{c_H^{\beta+2k}} = \prod_{j=1}^{n} \frac{(1+\frac{\beta}{2}j)_{kj}}{(1+\frac{\beta}{2})_k} ,$$

and

$$\frac{c_L^{\beta,a}}{c_L^{\beta+2k,\ a+k(m-1)}} = 2^{km(m-1)} \prod_{j=1}^{m} \frac{(1+\frac{\beta}{2}j)_{kj}\ (a-\frac{\beta}{2}(m-j))_{k(j-1)}}{(1+\frac{\beta}{2})_k} .$$

where n and m are the matrix sizes for the Hermite and Laguerre cases, respectively. Recall the Pochhammer symbolor shifted factorial $(x)_k \equiv \Gamma(x+k)/\Gamma(x)$.

Using the Jacobi form of the Selberg integral, one obtains that the moments of the discriminant for the β-Jacobi case are

$$\frac{c_J^{\beta,\ a_1,\ a_2}}{c_J^{\beta+2k,\ a_1+k(m-1),\ a2+k(m-1)}} = \prod_{j=1}^{m} \frac{(1+\frac{\beta}{2}j)_{kj}\ (a_1-\frac{\beta}{2}(m-j))_{k(j-1)}\ (a_2-\frac{\beta}{2}(m-j))_{k(j-1)}}{(1+\frac{\beta}{2})_k\ (a_1+a_2-\frac{\beta}{2}(m-j))_{k(m+j-2)}} .$$

Chapter 6

Limiting empirical distributions for β-Hermite and β-Laguerre ensembles

As mentioned in Section 3.3, exact level densities have been computed for the Gaussian ensembles for fixed n in terms of orthogonal polynomials [66]. Similar methods apply to the study of Wishart matrices.

The asymptotic (scaled, and with $n \to \infty$) limiting shape of the empirical distribution has been determined to be the semi-circle distribution of density (6.1) for the Gaussian ensembles and the compactly supported distribution corresponding to density (6.2) for the Wishart real and complex ensembles.

Most of the work covers the cases $\beta = 1, 2$, though $\beta = 4$ has also been considered in some cases.

The convergence was first determined (in a simpler format) by the method of moments in the Gaussian case by Wigner [94]. This result establishes that the level densities converge to the semi-circle distribution. Later the result was improved to convergence in probability of the underlying empirical distributions to the semi-circle distribution by Grenander [38], and then strengthened by Arnold [5] to almost sure convergence of the empirical distributions. Later, Trotter [92] found a different proof for the same type of convergence proved by Grenander, using the tridiagonal matrix

model in Chapter 5, for $\beta = 1$.

For the Wishart ensembles, convergence (again by the method of moments) was investigated in a larger context by Marcenko-Pastur [64]; later Jonsson [54] strengthened the result to convergence in probability, and Silverstein and Bai [82] proved almost sure convergence for a larger class of matrices.

Recall from Section 3.3 that the density of the empirical distribution function for the eigenvalues $\lambda_1, \ldots, \lambda_n$ of an $n \times n$ matrix A is $\frac{1}{n}\sum_{i=1}^{n} \delta(x - \lambda_i)$, and that the level density ρ for an $n \times n$ ensemble is the p.d.f. of a random eigenvalue of a random matrix from the ensemble. Equivalently, the level density is the the expected (over the ensemble) empirical density

$$\rho(x) = E_A \left[\frac{1}{n} \sum_{i=1}^{n} \delta(x - \lambda_i) \right] .$$

There are multiple possibilities for convergence as $n \to \infty$, if one considers all three layers: the random eigenvalue, the empirical distribution (or the uniform (discrete) measure on the set of eigenvalues), and the level density. To illustrate this, we propose the following thought experiment.

Suppose we have an infinite sequence of matrices, the first being of size 1, the second of size 2, the third of size 3, ..., which we denote by $(A_1, A_2, \ldots, A_n, \ldots)$. Suppose we choose a random eigenvalue (uniformly) from the eigenvalues of A_1, one eigenvalue (uniformly) from the eigenvalues of A_2, etc; we denote these eigenvalues $(\lambda_1(A_1), \lambda_2(A_2), \ldots, \lambda_n(A_n), \ldots)$. These are random variables with the empirical distributions of the eigenvalues of A_1, A_2, etc, which we denote by $(F_{A_1}, F_{A_2}, \ldots, F_{A_n}, \ldots)$.

In addition to this, we place a probability measure on the set of sequences $(A_1,\ A_2,\ \ldots,\ A_n,\ \ldots)$, which is the tensor product of a measure on 1×1 matrices with a measure on 2×2 matrices, etc.

The level densities are then the p.d.f.s of the expected distributions (over the probability measures we placed on $A_1,\ A_2,\ \ldots$) of $(\lambda_1(A_1), \lambda_2(A_2), \ldots, \lambda_n(A_n), \ldots)$. We denote by $(\lambda_1, \lambda_2, \ldots, \lambda_n, \ldots)$, a set of variables with c.d.f. corresponding to the level density.

Let F be a distribution uniquely determined by its set of moments.

To have convergence of moments of the level density to μ is to have that for every $k \in \mathbb{N}$, $\lim_{n\to\infty} E[\lambda_n^k] = E[X^k]$, where X is a variable distributed according to μ (see [11]). Note that, by the symmetry of the eigenvalues,

$$E_n[\lambda_n^k] = \frac{1}{n} E_{A_n}[\sum_{i=1}^{n} \lambda_i(A_n)^k] = \frac{1}{n} E[\mathrm{tr}(A_n^k)] \ .$$

Note that $\frac{1}{n}\mathrm{tr}(A_n^k)$ is the kth moment of the *empirical distribution function* of the matrix A_n.

Convergence of moments is equivalent to convergence in distribution (pointwise, see [11]), hence the above is equivalent to convergence of the expected empirical distribution (i.e. the c.d.f. of the level density) to F, in the limit as $n \to \infty$. In other words, if one averages over $(A_1, \ldots, A_n, \ldots)$ the empirical distributions $F_{A_1}, \ldots, F_{A_n}, \ldots$, the obtained sequence of distributions converges to F.

Convergence in probability is stronger; it says that not only the average empirical distribution converges, but that given ϵ, given any sequence of matrices $(A_1, A_2, \ldots, A_n, \ldots)$, with probability $1-\epsilon$, there is an $N = N_\epsilon$ such that the empirical distributions μ_{A_n} themselves, for all $n \geq N$, are "within ϵ" of μ.

This is equivalent to saying that for every $k \in \mathbb{N}$, the sequence of kth moments of the empirical distributions, that is, $(\frac{1}{1}\mathrm{tr}(A_1^k), \frac{1}{2}\mathrm{tr}(A_2^k), \ldots, \frac{1}{n}\mathrm{tr}(A_n^k), \ldots)$ converges in probability to the kth moment of μ, $E[X^k]$).

To prove this, it is enough to show that $\mathrm{Var}(\frac{1}{n}\mathrm{tr}(A_n^k)) \to 0$ for all k, as $n \to \infty$ (this is immediate; for more on this see [38]).

Finally, almost sure convergence is even stronger; it says that with probability 1, given any sequence of matrices $(A_1, A_2, \ldots, A_n, \ldots)$, the distributions μ_{A_n} converge to μ. This is equivalent to saying that for every $k \in \mathbb{N}$, the sequence $(\frac{1}{1}\mathrm{tr}(A_1^k), \frac{1}{2}\mathrm{tr}(A_2^k), \ldots, \frac{1}{n}\mathrm{tr}(A_n^k), \ldots)$ converges almost surely to the kth moment of μ, $E[X^k]$). Once again this is immediate [38].

The Borel-Cantelli Lemma (for a reference, see for example [11]) provides a technical condition for almost sure convergence; one needs to show that for every k and ϵ, $\sum_{n\geq 1} \Pr(|\frac{1}{n}\mathrm{tr}(A_n^k) - E[X^k]| > \epsilon) < \infty$. Note that since we already have that $E[\frac{1}{n}\mathrm{tr}(A_n^k)] \to E[X^k]$, it is enough to show that $\sum_{n\geq 1} Pr(|\frac{1}{n}\mathrm{tr}(A_n^k) - E[\frac{1}{n}\mathrm{tr}(A_n^k)]]| >$

$\epsilon) < \infty$.

But by Chebyshev's inequality, for this to happen for every ϵ it is enough to have that for every k, $\sum_{n\geq 1} \mathrm{Var}(\frac{1}{n}\mathrm{tr}(A_n^k)) < \infty$.

In Table 6 we illustrate these three types of convergence. We use the notation $\mu_{n,k}$ for $\frac{1}{n}\mathrm{tr}(A_n^k)$, σ_k for $E[X^k]$, F_n^β for the empirical distribution of the eigenvalues of A_n, and F for the limiting distribution. The technical and sufficient conditions must be fulfilled for all $k \in \mathbb{N}$.

Convergence	Notation	Technical Condition	Sufficient condition
Of moments	$E[\mu_{n,k}] \longrightarrow \sigma_k$		
In probability	$\begin{cases} \mu_{n,k} \xrightarrow{p} \sigma_k \\ F_n^\beta \xrightarrow{p} F \end{cases}$	$\Pr[\lvert\mu_{n,k} - E[\mu_{n,k}]\rvert > \epsilon] \to 0$	$\mathrm{Var}(\mu_n^k) \to 0$
Almost sure	$\begin{cases} \mu_{n,k} \xrightarrow{a.s.} \sigma_k \\ F_n^\beta \xrightarrow{a.s.} F \end{cases}$	$\sum_{n=1}^{\infty} \Pr[\lvert\mu_{n,k} - E[\mu_{n,k}]\rvert > \epsilon] \ < \infty$	$\sum_{n=1}^{\infty} \mathrm{Var}(X_n^k) < \infty$

Table 6.1: Three types of convergence for our distributions; convergence of moments refers to the *level density*, while convergence in probability and almost sure convergence refer to the *empirical distribution*.

The two theorems below gather the main results mentioned in the papers cited.

Theorem 6.0.4. *Let $\beta \in \{1, 2, 4\}$. Let A_n be a matrix from the β-Gaussian ensemble, scaled by $1/\sqrt{2n\beta}$. Let F_n^β be the empirical distribution function for the eigenvalues of A_n. Then, as $n \to \infty$, $F_n^\beta(x) \xrightarrow{a.s.} S(x)$, where $S(x)$ is the distribution (c.d.f.) of the semi-circle*

$$s(x) \equiv \begin{cases} \frac{2}{\pi}\sqrt{1-x^2}, & \text{for } x \in [-1, 1] \ , \\ 0, & \text{for } x \notin [-1, 1] \ . \end{cases} \tag{6.1}$$

Theorem 6.0.5. *Let $\beta \in \{1, 2, 4\}$. Let B_n be a matrix from the $W^\beta(m, n)$ Wishart ensemble, scaled by $\sqrt{\gamma/(m\beta)}$. Let $F^\beta_{m,n}$ be the empirical distribution function for the eigenvalues of A_n. Suppose that $\lim_{m \to \infty} \frac{m\beta}{n} = \gamma \leq 1$. Let $a = (\sqrt{\gamma} - 1)^2$, $b = (\sqrt{\gamma} + 1)^2$. Then, as $m \to \infty$, $F^\beta_{m,n}(x) \xrightarrow{a.s.} E_\gamma(x)$, where $E_\gamma(x)$ is the distribution (c.d.f.) corresponding to the density*

$$e_\gamma(x) \equiv \begin{cases} \frac{1}{2\pi\gamma} \frac{\sqrt{(x-a)(x-b)}}{x}, & \text{for } x \in [a, b] \ , \\ 0, & \text{for } x \notin [a, b] \ . \end{cases} \tag{6.2}$$

When $\gamma = 1$ in the above, by making the change of variables $x = y^2$, one obtains the well-known quarter-circle law for the singular values of a matrix from $G^\beta(n, n)$.

Note that in both cases, the β parameter only influences the scaling, not the limiting distribution law.

The two theorems above can be thought of as illustrating the Strong Law of Large Numbers for the eigenvalues of a β-Hermite (Gaussian) or β-Laguerre (Wishart) ensemble, with $\beta = 1, 2, 4$.

In 1998, Johansson [51] has proved that these two laws hold true in a more general setting, and in a stronger form, for all β. Roughly speaking, Johansson proved that a "Central Limit Theorem" is true for the eigenvalues of an arbitrary β-Hermite ensemble. We present a weaker version of his result in Theorem 6.0.6.

Theorem 6.0.6. *For any continuous function $h : \mathbb{R} \to \mathbb{R}$,*

$$\int_{\mathbb{R}} h(x) \ d(F^\beta_n(x)) \longrightarrow \int_{-1}^{1} h(x) s(x) dx \ ,$$

where $s(x)$ is given by (6.1). Moreover, there exists a distribution ν on $[-1, 1]$ (independent of β) such that

$$\log \left(\int_{\mathbb{R}} e^{h(x)} \ d(F^\beta_n(x)) \right) - n \int_{-1}^{1} h(x) s(x) dx) \longrightarrow \left(1 - \frac{2}{\beta}\right) \int_{-1}^{1} h(x) \nu(x) dx + \frac{2}{\beta} A(h) \ ,$$

where $A(h)$ is a quadratic functional independent of β.

In recent work, Silverstein and Bai [83] have proved a similar result for a more general class of Wishart-type matrices, for the real and complex case.

Unlike the proofs of Theorem 6.0.4, which consist mainly of combinatorial counting techniques (based on the full matrix models for $\beta = 1, 2, 4$), Johanssen's proof makes use of analysis tools (based only on the p.d.f. of the arbitrary β-Hermite ensembles).

In this chapter we use basic combinatorial methods, together with our new matrix models, to obtain a generalization of Theorem 6.0.4 (which is not as strong as the Johansson result for the β-Hermite ensembles) and a generalization of Theorem 6.0.5 for the general β case.

First we will need to show some combinatorial identities involving counting weighted paths; we do this in Section 6.1. These identities will be used in the following sections.

To prove convergence, we use the "classical" approach of examining traces of powers of the random matrix model in order to study the empirical distribution.

Remark 6.0.7. *Many of the results mentioned above were obtained by examining traces of powers of the random matrix model; it is also the method employed by Soshnikov and Sinai [85] for proving a central limit theorem for traces of high powers of the (full) matrix model.*

In Sections 6.2 and 6.3, we show, starting from the models, that after the scaling, the moments of the β-Hermite and β-Laguerre ensembles converge to those of the limiting distribution (independent of β).

In Section 6.4 and 6.5, we compute the asymptotics of the variance for the variables $\frac{1}{n}\mathrm{tr}(A_n^k)$, and use them to show that almost sure convergence to the limiting distribution occurs.

6.1 Dyck and alternating Motzkin paths, Catalan and Narayana numbers

In this section we develop some combinatorial identities involving Dyck and Motzkin paths, which will be used in the remaining sections of this chapter. To prove almost sure convergence, we will be looking at products of elements of the random matrix models, the indices of which form integer sequences corresponding to Dyck

and Motzkin paths. The identities we develop here will be used for counting the number of asymptotically relevant terms.

The relationship between Dyck paths, Catalan and Narayana numbers, and level densities of random matrix ensembles has been known to some researchers for quite a while, starting with Wigner [94]. We explore here this relationship to what we believe to be a new depth, by finding and proving new combinatorial identities which are crucial to our proofs of almost sure convergence for the level densities.

A good reference for Dyck paths, Catalan numbers, and Narayana numbers, is Stanley [87].

We start by giving a number of definitions.

Definition 6.1.1. *The nth Catalan number C_n is defined as*

$$C_n = \frac{1}{n+1}\binom{2n}{n} .$$

The Catalan numbers count many different combinatorial structures; in particular, they count Dyck paths.

Definition 6.1.2. *A Dyck path of length $2n$ is a lattice path consisting of "rise" steps or "rises" ($\nearrow$) and "fall" steps or "falls" ($\searrow$), which starts at $(0,0)$ and ends at $(2n,0)$, and stays above the x-axis (see Figure 6-1).*

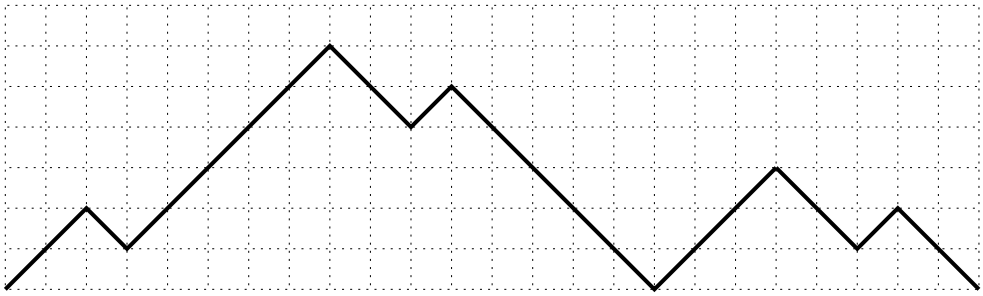

Figure 6-1: A Dyck path of length 24.

Lemma 6.1.3. *The number of Dyck paths of length $2n$ is C_n.*

Proof. Consider a Dyck path; add at the end a fall step (from $(2n, 0)$ to $(2n+1, -1)$). The new path has $n+1$ falls, and n rises. Choose one of the $n+1$ falls, label it, and cut the path into 2 pieces: before the labeled step and after (including the labeled step). Translate the second piece to $(0, 0)$, and glue the first piece at the end of it.

Now one has a path from $(0, 0)$ to $(2n+1, -1)$, on which the first step is a fall.

Consider any path from $(0, 0)$ to $(2n+1, -1)$, on which the first step is a fall. Find the leftmost minimum, cut the path in two (before the leftmost minimum and after). Translate the second part to $(0, 0)$, and glue the first part at the end of it – the result is a Dyck path with an extra fall step at the end.

The above constructs a bijection between the set of Dyck paths with an extra fall at the end and with a labeled fall, and the set of paths from $(0, 0)$ to $(2n+1, -1)$, on which the first step is a fall. It follows that the number N of Dyck paths satisfies

$$(n+1)N = \binom{2n}{n},$$

and so $N = C_n$. □

Motzkin paths are relatively well-known to combinatorialists; we will need a special type of Motzkin path, which we call an *alternating Motzkin path*. The definitions are below.

Definition 6.1.4. *A Motzkin path is a path consisting of "rise" steps or "rises" ($\nearrow$), "fall" steps or "falls" ($\searrow$), and "level" steps ($\rightarrow$), which starts at $(0, 0)$, ends at $(2k, 0)$, and stays above the x-axis.*

Definition 6.1.5. *An alternating Motzkin path of length $2k$ is a Motzkin path in which rises are allowed only on even numbered steps, and falls are only allowed on odd numbered steps. See Figure 6-2.*

Remark 6.1.6. *It follows from the definition that an alternating Motzkin path starts and ends with a level step.*

We count the alternating Motzkin paths with a given number r of rises, and prove two simple lemmas.

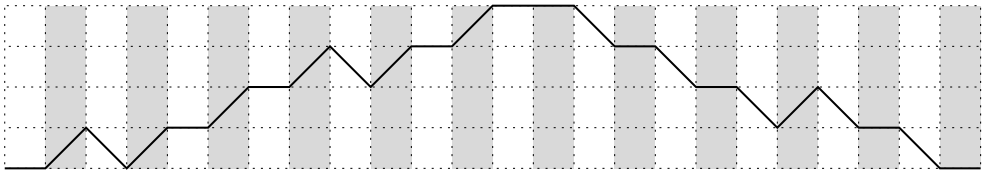

Figure 6-2: An alternating Motzkin path of length 24, with a total of 7 rises.

Lemma 6.1.7. *The number of alternating Motzkin paths of length $2k$ with r rises is the Narayana number $N(k,r) = \frac{1}{r+1}\binom{k}{r}\binom{k-1}{r}$.*

Proof. The proof follows very closely the proof of Lemma 6.1.3. Given an alternating Motzkin path of length $2k$ with r rises, we delete the initial level step, and add at the end of it a fall; we obtain a path from $(1,0)$ to $(2k+1,-1)$ which has $r+1$ falls and r rises; we label one of the falls and cut the path, translate the second part to $(0,0)$, and glue the first part at the end. We obtain a path from $(0,0)$ to $(2k,-1)$ which starts with a fall step; the initial path can be recovered by cutting this new path at the leftmost minimum and reassembling the two parts in inverse order.

Note that the "new" path from $(0,0)$ to $(2k,-1)$ which starts with a fall still has the alternating property, *because we have deleted the initial level step*: falls only on odd numbered steps, rises only on even numbered steps. Therefore the number of such paths is $\binom{k}{r}\binom{k-1}{r}$, since the first step is pre-determined.

Just as before, this operation constructs a bijection; it follows that the number of alternating Motzkin paths of length $2k$ which have exactly r rises is

$$N(k,r) = \frac{1}{r+1}\binom{k}{r}\binom{k-1}{r}\ .$$

□

Lemma 6.1.8. *The number of level steps taken at altitude i is even, and exactly half of them are on even-numbered steps.*

Proof. Let us examine a "maximal" sequence of level steps at altitude i; we use "maximal" to express the fact that the steps preceding and succeeding the sequence of level steps (if they exist) are rises or falls. For the benefit of the reader, we include Figure 6-3).

Assume $i > 0$, so that there are steps preceding and succeeding the sequence of level steps.

If the sequence of level steps has even length, then half of them are on even-numbered steps. Moreover, due to the alternating constraint, they have to either be preceded by a rise and succeeded by a fall, or the reverse (see regions B,D in Figure 6-3).

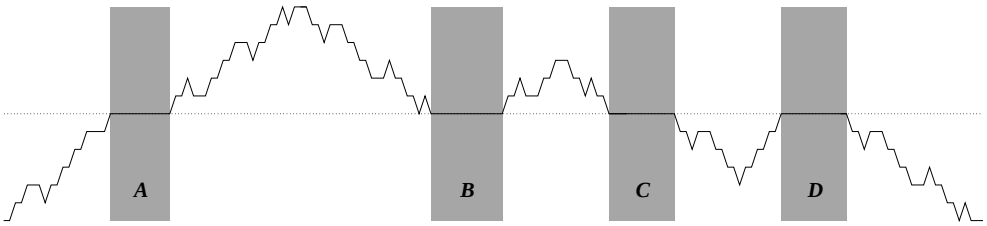

Figure 6-3: Looking at the four types of level sequences found at some level i in an alternating Motzkin path: even-length ones (B, D) and odd-length ones (A,C).

If the sequence has odd length, there are two possibilities: either both the preceding and the succeeding steps are rises (region A in Figure 6-3), or they are both falls (region C). It is enough to examine the first case (region A).

In the first case, the path climbs to a higher altitude, and since it ends at $(2k, 0)$, it will have to eventually go below altitude i. Look at the closest place where the path does that. The only way in which it can come back and leave level i is by a sequence "fall, level, level, ..., level, fall" (see region C). This sequence will also have odd length; moreover, because of the alternating property, this pair of maximal length level sequences will have exactly half of its steps on odd-numbered steps.

Note that the path cannot have two regions A without a region C between them (nor the converse), since a region A implies that a descent to altitude $i-1$ has already taken place and the only way in which this can happen is by passing through a region C. So the regions A and C alternate in the path, with a region A being first and a region C being last.

Thus, we can pair *all* the odd-length maximal level sequences at altitude i (each region A gets paired with the following region C), so that each pair has exactly half

of its steps on odd-numbered steps; this shows the claim for $i > 0$.

Assume now $i = 0$. If there are both preceding and succeeding steps, they can only be a fall and a rise (in this order); in this case the sequence of level steps has even length. Suppose that either the preceding or the succeeding step is missing (i.e. we are at one end of the path or at the other). In the first case, the succeeding step can only be a rise, so the path has odd length, and one more odd-numbered step than even-numbered steps. We thus know that any alternating Motzkin path starts with an odd-length sequence of level steps. Similarly, it ends with an odd-length sequence of level steps; this sequence has one more even-numbered step. Hence the pair formed by the first and last maximal sequences of level steps has exactly as many odd-numbered steps as even-numbered steps. This concludes the proof. □

Suppose we associate to each alternating Motzkin path of length $2k$ with r rises a weight γ^r for some $\gamma > 0$.

Definition 6.1.9. *We define the Narayana polynomial $N_k(\gamma)$ to be the total weight of the alternating Motzkin paths of length $2k$ (this is a generating series in γ). In other words,*

$$N_k(\gamma) = \sum_{r=0}^{k-1} \gamma^r \ \frac{1}{r+1} \binom{k}{r} \binom{k-1}{r} \ .$$

We have now the ingredients to prove the three main results of this section, which are technical results and will be used in Sections 6.4 and 6.5.

Lemma 6.1.10. *Let p and q be two Dyck paths of length $2k$. For $i \geq 0$, let $k_i(p)$ (respectively $k_i(q)$) be the number of rises p (respectively q) take from altitude i to altitude $i+1$. Then*

$$\sum_{p,q} \sum_{i \geq 0} k_i(p) k_i(q) = C_{2k} - C_k^2 \ .$$

Proof. Note that the right hand side of the equality counts the Dyck paths of length $4k$ which do not cross the x-axis in the middle; it is not hard to see that such paths are at an even altitude $2i$ in the middle.

We prove the lemma by constructing a bijection between the set

$$(p, q, i, \{1, \ldots, k_i(p)\}, \{1, \ldots, k_i(q)\}) \ ,$$

and the paths of length $4k$ which do not cross the x-axis in the middle. Since the size of the first set is given by the right-hand side of the equality, while the size of the second set is given by the left-hand side of the equality, constructing this bijection is enough to prove the lemma.

Note that $k_i(q)$ is also the number of falls from altitude i to altitude $i-1$.

Given p and q, choose an $i \geq 0$. Choose a rise x in p from altitude $i-1$ to altitude i, and one fall y in q from altitude i to altitude $i-1$ (Figure 6-4). From now on, every operation we perform on p is accompanied by a mirror-reversed operation in q ("left" becomes "right", "rise" becomes "fall"). We only describe the moves on p; however, we illustrate both the moves on p and on q.

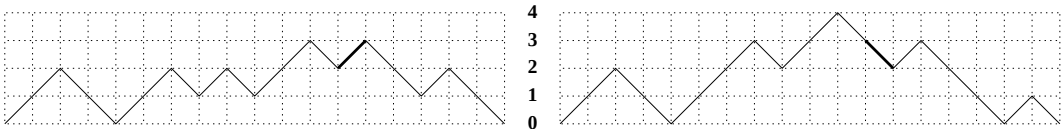

Figure 6-4: Choosing a rise from altitude 2 in p (left) and a fall from altitude 3 in q (right).

In the path p, start at x, mark it, and go left; mark the first rise from altitude $i-2$ to altitude $i-1$, then go left and mark the first rise from $i-3$ to $i-2$, etc. Each of these i marked edges has a "closing" fall on the right side of x (see Figure 6-5).

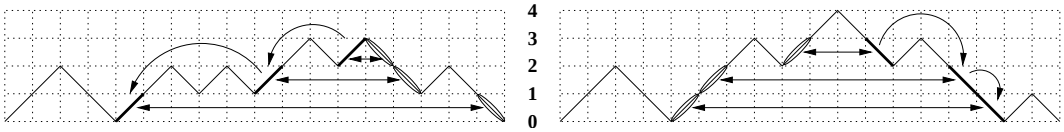

Figure 6-5: Finding the "first rise" steps from 0 to 2 in p (left), and the "first fall" steps from 2 to 0 in q (right); the curved arrows mark them, and the horizontal double arrows find their respective "closing" steps.

We will be interested in these i "closing" fall steps. Flip them all to rises; each flip increases the final altitude of the path by 2, for a total of $2i$ final altitude increase.

Note that the flipped edges correspond to the rightmost rise from altitude i, the rightmost rise from altitude $i+1$, the rightmost rise from altitude $i+2$, etc; hence, given a path of length $2k$ made of $k+i$ rises and $k-i$ falls, which stays above the x-axis, there is a simple transformation which flips the rightmost rises from altitude $i+1$, $i+2$, etc, to falls, to get a Dyck path. Thus this process is reversible (see Figure 6-6).

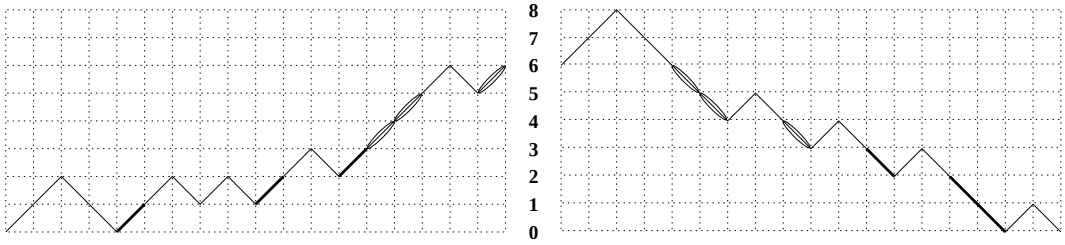

Figure 6-6: Flipping the rises in p and the falls in q. The flipped edges correspond to the rightmost rise from altitude i, the rightmost rise from altitude $i+1$, the rightmost rise from altitude $i+2$, etc in the new path; the converse is true for q.

We concatenate the two paths obtained from p and q to obtain a Dyck path of length $4k$ which is at altitude $2i$ in the middle (see Figure 6-7).

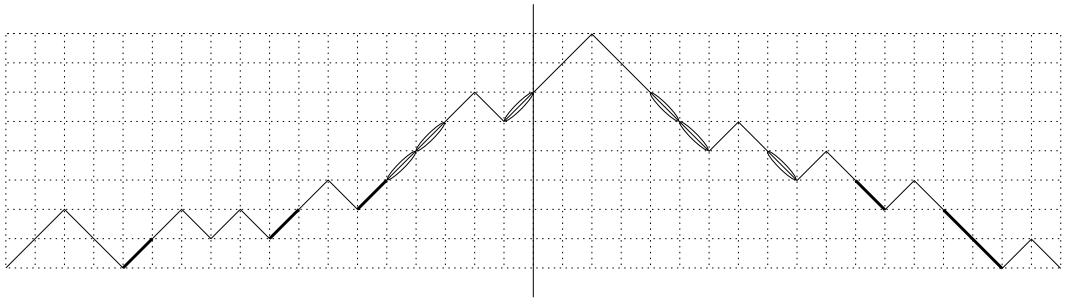

Figure 6-7: Concatenating the two paths from Figure 6-6; the resulting path is a Dyck path of double length and altitude $6 = 2 \times 3$ in the middle.

Everything we did is reversible, and hence what we have in fact constructed is a bijection from the set

$$(p, q, i, \{1, \ldots, k_i(p)\}, \{1, \ldots, k_i(q)\}) \ ,$$

and the paths of length $4k$ which do not cross the x-axis in the middle.

Thus the lemma is proved. □

Lemma 6.1.11. *Let p and q be two Dyck paths of length $2k$, and let $m_i(p)$ and $m_i(q)$ be the number of vertices at altitude i in p, respectively q. Then*

$$\sum_{p,q} \sum_{i \geq 0} m_i(p) m_i(q) \;=\; C_{2k+1} \;.$$

Proof. The proof is similar and involves constructing a bijection between the set $\{(p, q, i, x \in \{1, \ldots, m_i(p)\}, y \in \{1, \ldots, m_i(q)\})\}$ and the Dyck paths of length $4k+2$.

Note that concatenation is no longer enough, because the total length increases by 2. Also note that Dyck paths of length $4k+2$ are at an odd altitude in the middle.

We use the same method of construction as in the previous lemma; given p, q, i, we choose a vertex x at altitude i in p and a vertex y at altitude i in q. Then in p we look left and identify the closest rises from altitude $i-1$ to altitude i, from altitude $i-2$ to altitude $i-1$, etc; we mark their "closing" falls on the right, and then flip them. This increases the final altitude by $2i$; we insert an additional rise from i to $i+1$ starting from vertex x, so that the total increase in altitude is $2i+1$. We do the mirror-image operation for q, and glue the resulting paths together; we have a Dyck path of length $4k+2$ which is at altitude $2i+1$ in the middle.

To recover the paths we started with, cut the Dyck path of length $4k+2$ in the middle, and then identify (on the left) the rightmost rises from altitude i to $i+1$, $i+1$ to $i+2$, etc; delete the one from i to $i+1$, and flip the others to falls to get p; do the mirror-image transformation on the right to obtain q.

Since the size of the set $\{(p, q, i, x \in \{1, \ldots, m_i(p)\}, y \in \{1, \ldots, m_i(q)\})\}$ is the left hand side of the equation, the lemma is proved. □

The last lemma of this section is similar to the ones preceding it.

Lemma 6.1.12. *Let p and q be two alternating Motzkin paths of length $2k$, p with r_1 rises, and q with r_2 rises. For $i \geq 0$, let $k_i(p)$ (respectively $k_i(q)$) the number of rises p (respectively q) take from altitude i to altitude $i+1$. Also for $i \geq 0$, let $l_i(p)$*

(respectively $l_i(q)$) be the number of level steps p (respectively q) takes from altitude i to altitude i, on odd-numbered steps. Then

$$\sum_{p,q} \left(\gamma^{r_1+r_2} \sum_{i\geq 0} k_i(p)k_i(q) + \gamma^{r_1+r_2+1} \sum_{i\geq 0} l_i(p)l_i(q) \right) = N_{2k}(\gamma) - N_k(\gamma)^2 \ .$$

Proof. The proof is very similar to the one given for Lemma 6.1.10, and it once again consists of finding a bijection. The right-hand side represents the total weight of the alternating Motzkin paths of length $4k$ which do not have two level steps at altitude 0 in the middle. Such paths can be at both even or odd altitude in the middle.

On the left-hand side, we have two terms in the sum; the first term of which, $\sum_{p,q} \gamma^{r_1+r_2} \sum_{i\geq 0} k_i(p)k_i(q)$, we prove to correspond to the total weight of the alternating Motzkin paths of length $4k$ which are at an even altitude in the middle.

The construction is the same as in Lemma 6.1.10, with a simple change: we can not just simply flip falls to rises edges, because of the alternating property (see Figure 6-8)! Instead, for each fall z from altitude l to altitude $l-1$ we choose the first level step from altitude $l-1$ to altitude $l-1$ to the right of it (see Figure 6-9), switch them, and then flip the falls to rises (see Figure 6-10)).

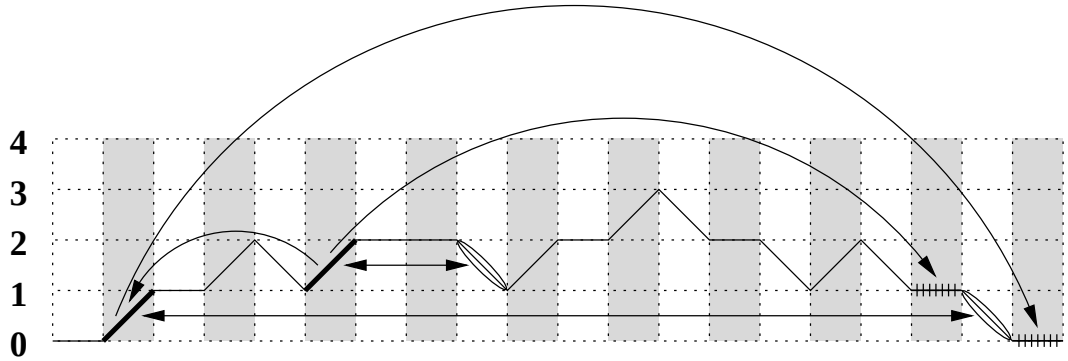

Figure 6-8: Choosing the rise; finding the corresponding ascending sequence, the "closing" one, and the next level steps. The thick lines represent the ascending sequence, the tripled lines – the "closing" one, and the "hatched" lines are the next level steps.

We claim that this move is allowable, since the first level step will necessarily be on an even-numbered step, because it must have been preceded by a fall.

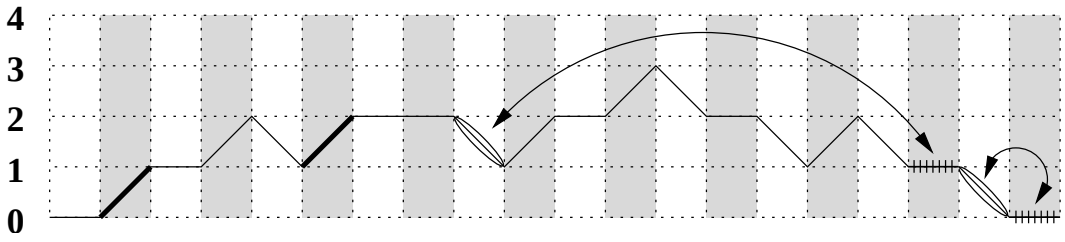

Figure 6-9: Putting each of the "closing" fall steps in correspondence with its next level step.

Indeed, by definition it can not have been preceded by a level step. If it had been preceded by a rise, we would have already visited altitude $l - 2$ again, so we must have descended there somehow after z, and any descent implies the existence of a sequence of the form "fall, an odd number of level steps, fall" – but this implies we have already had some level steps before our "first" level step.

Hence the level step must have been preceded by a fall, and thus it is on an even-numbered step.

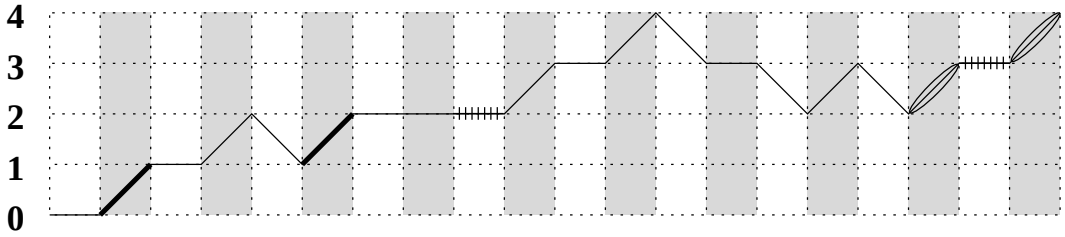

Figure 6-10: Switching each "closing" fall with the corresponding level step, and then flipping the falls to rises.

This is the construction for the left path; for the right path, we do the transformation in the mirror (instead of a rise, we choose a fall, because they are just as many; instead of looking left to find the closest rises, we look to the right, etc). Note that this preserves the alternating structure.

By the same reasoning as in Lemma 6.1.10, everything is reversible; also, the weight of the pair of paths (given by $\gamma^{r_1+r_2}$) is preserved (the total number of up steps is unchanged). Hence this transformation represents a weight-preserving bijection to

the set of alternating Motzkin paths of length $4k$ which are at an even altitude in the middle.

For the second term, we use a similar argument. Given two paths, p and q, we examine p and choose a level step at altitude i that is on an odd-numbered step. Then we perform the same "search-and-mark-then-flip" as before, except that we change one more step, namely, the level step we started with - we replace it by an rise (which is allowable). Just as before, we do the mirror-transformation to the right hand side path.

We obtain thus a bijection to the set of alternating Motzkin paths of length $4k$ which are at odd altitude in the middle; moreover, the bijection increases the weight of the two paths by exactly γ (since we get rid of two level steps, and create a rise on the left path, and a fall on the right one).

The statement of the lemma follows. □

6.2 Convergence of moments: β-Hermite

We prove that for arbitrary β the level density of the $n \times n$ β-Hermite ensemble (scaled by $1/\sqrt{2n\beta}$) converges to the semi-circle distribution S; we also find the first-order asymptotics.

First, we prove an easy lemma.

Lemma 6.2.1. *Let the vector $q = (q_1, \ldots, q_n)$ have distribution as in Theorems 5.2.1 and 5.3.1, that is, that of a vector of n independent χ_β variables, normalized to unit length. Then*

$$\begin{aligned} E[q_i^2] &= \frac{1}{n}, & \forall\ 1 \le i \le n \\ E[q_i^4] &= \frac{\beta+2}{n(n\beta+2)}, & \forall\ 1 \le i \le n \\ E[q_i^2 q_j^2] &= \frac{\beta}{n(n\beta+2)}, & \forall\ 1 \le i \ne j \le n \quad . \end{aligned}$$

Proof. Let $x = (x_1, \ldots, x_n)$ be a vector of n independent χ_β variables.

We make the change of variables $x = rq$, and note that $dx = r^{n-1}\, dq\, dr$, and that r and q separate. Moreover r is a $\chi_{n\beta}$ variable.

The calculations follow. □

Remark 6.2.2. *One can get any joint moments $\lambda = (\lambda_1, \ldots, \lambda_n)$ of the q_i variables by the method sketched above.*

Let X_n be a random eigenvalue of the matrix A from the scaled $n \times n$ β-Hermite ensemble (scaled by $1/\sqrt{2n\beta}$). Then the distribution of X_n is the distribution of $1/n$ $\mathrm{tr}(A) = 1/n \sum_{i=1}^{n} \lambda_i$, and the kth moment $\mu_{n,k}$ of X_n is the same as the expected value of $1/n$ $\mathrm{tr}(A^k) = 1/n \sum_{i=1}^{n} \lambda_i^k$.

Theorem 6.2.3. *Let β be arbitrary. Then $\mu_{n,k} \to \mu_k$, as $n \to \infty$, for all $k \in \mathbb{N}$. Here*

$$\mu_k = \begin{cases} \frac{1}{4^k} \frac{1}{k/2+1} \binom{k}{k/2} \ , & \text{if } k \text{ is even} \ , \\ 0, & \text{if } k \text{ is odd} \ , \end{cases}$$

are the moments of the semi-circle.

First we need to prove the following easy lemma.

Lemma 6.2.4.

$$E\left[\frac{1}{n} tr(A^k)\right] = E\left[(A^k)_{11}\right] \ ,$$

where $(A^k)_{11}$ is the $(1,1)$ element of the kth power of A.

Proof. Let $A = Q\Lambda Q'$ be the eigenvalue decomposition of A. The scaling on A translates into a scaling of the eigenvalues; Q is the same as in Theorem 5.2.1. We have

$$(A^k)_{11} = \sum_{i=1}^{n} \lambda_i^k q_i^2 \ .$$

By Theorem 5.2.1, by the symmetry of q and independence of q and Λ, and by Lemma 6.2.1,

$$E[(A^k)_{11}] = E[q_1^2] \ E[\sum_{i=1}^{n} \lambda_i^k] = \frac{1}{n} \mathrm{tr}(A^k) \ .$$

□

Now we are ready to prove Theorem 6.2.3.

Proof. It is enough to show that

$$E[(A^k)_{11}] \to \mu_k \ ,$$

as $n \to \infty$.

We denote the variables on the diagonal of A by $(a_n, \ldots, a_1)$, and those on the subdiagonal by $(b_{n-1}, \ldots, b_1)$. From the model, for $1 \leq i \leq \lceil k/2 \rceil$, $b_{n-i} \sim \frac{1}{\sqrt{2n\beta}} \frac{1}{\sqrt{2}} \chi_{(n-i)\beta} = \frac{1}{2} \frac{\chi_{(n-1)\beta}}{\sqrt{n\beta}}$. Hence

$$E[b_{n-i}^{2j}] \to \frac{1}{4^j} \ , \ \forall \ \ 1 \leq j \leq \lfloor \frac{k}{2} \rfloor \quad \text{and} \quad \forall \ \ 1 \leq i \leq \lceil \frac{k}{2} \rceil \ . \tag{6.3}$$

On the other hand, for all $1 \leq i \leq \lceil k/2 \rceil$, a_{n-i} is a Gaussian scaled by $\frac{1}{\sqrt{2n\beta}}$, and hence

$$E[a_{n-i}^{2j+1}] = 0 \ , \quad \forall \ \ 1 \leq j \leq \lfloor \frac{k-1}{2} \rfloor \ , \tag{6.4}$$

$$E[a_{n-i}^{2j}] = \frac{(2j-1)!!}{(2\beta n)^j} \to 0, \ \forall \ \ 1 \leq j \leq \lfloor \frac{k}{2} \rfloor \ . \tag{6.5}$$

Let us now examine $E[(A^k)_{11}]$. We find that

$$A_{11}^k = \sum A_{1i_1} A_{i_1 i_2} \ldots A_{i_{k-2} i_{k-1}} A_{i_{k-1} 1} \ ,$$

where the sum is taken over all sequences $i_0 = 1, i_1, i_2, \ldots, i_{k-1}, i_k = 1$. Note that a term in the sum is nonzero if and only if $|i_j - i_{j+1}| \in \{0, 1\}$, for all $0 \leq j \leq k-1$ (all other entries are 0).

Consequently

$$E[A_{11}^k] = \sum E[A_{1i_1} A_{i_1 i_2} \ldots A_{i_{k-2} i_{k-1}} A_{i_{k-1} 1}] \ ,$$

over all sequences $i_0 = 1, i_1, i_2, \ldots, i_{k-1}, i_k = 1$ with $|i_j - i_{j+1}| \in \{0, 1\}$, for all $0 \leq j \leq k-1$.

Note that for k fixed, the sum above has a *finite* number of terms, depending only on k. Also note that since the sequence starts with 1 and must return to 1, $i_j \leq \lfloor \frac{k}{2} \rfloor$.

Associate to each sequence a lattice path starting at $(0, 0)$ such that a rise occurs when $i_{j+1} - i_j = 1$, a fall occurs when $i_{j+1} - i_j = 1$, and a level step occurs when $i_j = i_{j+1}$.

Remark 6.2.5. *Note that the path always stays above the x-axis (the indices are always greater than or equal to* 1*). This is a* Motzkin path *(note the that we have no alternation requirement). Such paths have the same number of rises and falls.*

Consider a term $A_{i_1 i_2} A_{i_2 i_3} \dots A_{i_k i_1} = a_n^{l_1} b_{n-1}^{l_2} a_{n-1}^{l_3} \dots a_{(n-\lfloor \frac{k}{2} \rfloor)}^{l_{2\lfloor \frac{k}{2} \rfloor + 1}} b_{(n-\lfloor \frac{k}{2} \rfloor - 1)}^{l_{2\lfloor \frac{k}{2} \rfloor + 2}}$.

By (6.4), (6.5), and (6.3), if one of $l_1, l_3, \dots, l_{2\lfloor \frac{k}{2} \rfloor + 1}$ is non-zero (i.e. the corresponding path has at least one level step), the expectation of such a term goes to 0.

Hence the only terms whose expected value will survive, asymptotically, are the ones that contain *no* diagonal terms; those corresponding to Dyck paths (with only rises and falls). Such sequences must then have even length.

If $k = 2v + 1$ is odd, there are no such sequences. Moreover, by Remark 6.2.5, the total number of level steps is odd, and hence the total number of level steps at some altitude i is odd. But this implies that there is an i such that the sequence contains $a_{(n-i)}$ to some odd power; since $a_{(n-i)}$ is independent from all the other entries, it follows that the expectation of such a term is exactly 0. Thus $\mu_{n,2v+1} = 0$.

If $k = 2v$ is even, the number of such sequences is exactly the number of Dyck paths of size $2v$. Let us now examine a term corresponding to such a sequence. First, note that each entry b_{n-i} must appear an even number of terms (because each entry corresponds to either a rise or a fall between altitudes $i - 1$ and i, and the number of steps between i and $i - 1$ is even). Hence we can write such a term as $b_{n-1}^{2k_1} b_{n-2}^{2k_2} \dots b_{n-v}^{2k_v}$, with $\sum_{i=1}^{v} k_i = 2v$. Since the b_{n-i}'s are independent, by (6.3), it follows that each path contributes asymptotically $1/4^v$ to the expectation, and thus $\mu_{n,2v} \to \frac{1}{4^v} C_v = \frac{1}{4^v} \frac{1}{v+1} \binom{2v}{v}$.

Hence we obtain convergence of moments. □

We will now look at the first-order behavior of the expectation. To do this, we have to examine more carefully the moments of the variables involved.

First, note that if $b_{(n-i)} \sim \frac{1}{2\sqrt{n\beta}}\chi_{(n-i)\beta}$, then

$$E[b_{(n-i)}^{2j}] = \frac{1}{4^j}\,\frac{1}{(n\beta)^j}\,2^j\,\frac{\Gamma\left(\frac{(n-i)\beta+2j}{2}\right)}{\Gamma\left(\frac{(n-i)\beta}{2}\right)} \tag{6.6}$$

$$= \frac{1}{4^j}\left(1-\frac{i}{n}\right)\left(1-\frac{i}{n}+\frac{2}{\beta n}\right)\ldots\left(1-\frac{i}{n}+\frac{2(i-1)}{\beta n}\right) \tag{6.7}$$

$$= \frac{1}{4^j}\left(1-j\frac{i}{n}+\frac{2}{\beta n}\binom{j}{2}+O\left(\frac{1}{n^2}\right)\right)\ . \tag{6.8}$$

Let $k = 2v$ be even. There is a source of first-order terms from the highest-order terms, corresponding to Dyck paths (with the sequence $1, i_1, \ldots, i_{2v-1}, 1$ describing a Dyck path). Let $\mathcal{D}_v$ be the set of Dyck paths of length $2v$. This source can be obtained as

$$\sum_{p\in\mathcal{D}_v} E[\ b_{n-1}^{2k_1}b_{n-2}^{2k_2}\ldots b_{n-v}^{2k_v}\ ] = \sum_{p\in\mathcal{D}_v} E[b_{n-1}^{2k_1}]\ldots E[b_{n-v}^{2k_v}] \tag{6.9}$$

$$= \sum_{p\in\mathcal{D}_v}\frac{1}{4^v}\prod_{i=1}^{v}\left(1-\frac{ik_i}{n}+\frac{2}{\beta n}\binom{k_i}{2}+O(n^{-2})\right) \tag{6.10}$$

$$= \frac{1}{4^v}C_v - \frac{1}{4^v n}\sum_{p\in\mathcal{D}_v}\left(\sum_{0\le i\le v} i\ k_i + \frac{2}{\beta n}\sum_{0\le i\le v}\binom{k_i}{2}\right) \tag{6.11}$$

The variables k_i in the above count the number of even rises (or the number of falls) between altitudes $i-1$ and i.

There is another source of first-order terms. Let us consider paths that take level steps (corresponding to sequences which contain diagonal terms). Since $k = 2v$ and the paths have the same number of rises and falls, it follows that the total number of level steps is always even (and thus at least 2).

The presence of a power of a diagonal term is described by equations (6.4) and (6.5). Hence we have a source of first-order terms from the generalized paths with exactly 2 level steps, both at the same altitude (because they must correspond to the *same* diagonal term), as in Figure 6-11. One can think of these paths as Dyck paths of length $2v-2$, with two level steps inserted at the same altitude. In expectation, such a term will contribute asymptotically $\frac{1}{4^{v-1}}\frac{1}{2n\beta}$, since the moments of the χ variables are as given by (6.3).

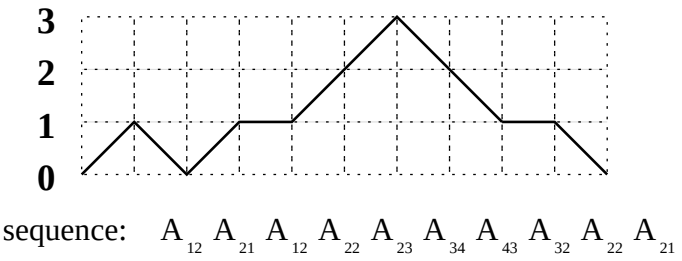

Figure 6-11: An example of path with exactly 2 level steps, both at the same altitude (altitude 1); such a path accounts for a first-order term in the expectation of the $2v$th moment.

To count these paths, we note that for each Dyck path p of length $2v-2$, there are $\sum_{i\geq 0}\binom{m_i(p)+1}{2}$ ways of choosing two places at the same altitude to insert level steps, where $m_i(p)$ is the number of times altitude i is visited (that is, the number of "vertices" on the path at altitude i).

For example, for the Dyck path of length 8 implicitly present in Figure 6-11, we have $\binom{4}{2}=6$ ways of inserting two level steps at altitude 0, $\binom{4}{2}=6$ ways of inserting two level steps at altitude 1, $\binom{3}{2}=3$ ways of inserting two level steps at altitude 2, and finally $\binom{2}{2}=1$ way of inserting two level steps at altitude 3, for a total of 16 possibilities.

Hence the total first-order contribution coming from such steps is

$$\frac{1}{2n\beta}\frac{1}{4^{v-1}}\sum_{p\in\mathcal{D}_{v-1}}\sum_{i\geq 0}\binom{m_i(p)+1}{2}.$$

Adding this to the first-order term in equation (6.11), we get that

$$\mu_{n,2v} = \frac{1}{4^v}C_v - \frac{D_v}{4^v n} + 2\frac{E_v}{4^v n\beta} + O(n^{-2}), \quad \text{with} \tag{6.12}$$

$$D_v = \sum_{p\in\mathcal{D}_v}\sum_{0\leq i\leq v} i\ k_i \quad , \quad \text{and} \tag{6.13}$$

$$E_v = \left(\sum_{p\in\mathcal{D}_v}\sum_{0\leq i\leq v}\binom{k_i}{2} + \sum_{p\in\mathcal{D}_{v-1}}\sum_{i\geq 0}\binom{m_i(p)+1}{2}\right) . \tag{6.14}$$

If $k=2v+1$, since $\mu_{n,2v+1}=0$, there are no first-order terms.

6.3 Convergence of moments: β-Laguerre

We prove that for fixed β and $\gamma_m = (m\beta)/2a \to \gamma \leq 1$, the level density of the $m \times m$ β-Laguerre ensemble of parameter a (scaled by $\sqrt{\gamma_m/(m\beta)}$) converges to the E_γ distribution in moments, and we compute the second-order terms.

We use some of the results from Section 6.2, like the Lemmas 6.2.1 and 6.2.4.

Let Y_m be a random eigenvalue of the matrix $L = BB^T$ from the scaled $m \times m$ β-Laguerre ensemble of parameter $a = \frac{m\beta}{2\gamma_m}$ (recall that the scaling is $\sqrt{\gamma_m/(m\beta)}$).

Then the distribution of Y_m is the distribution of $1/m \operatorname{tr}(L) = 1/m \sum_{i=1}^m \lambda_i$, and the kth moment $\mu_{m,k}$ of Y_m is the same as the expected value of $1/m \operatorname{tr}(L^k) = 1/m \sum_{i=1}^m \lambda_i^k$.

Theorem 6.3.1. *Let β be arbitrary, and γ_m, a, and γ as above. Then $\mu_{m,k,\gamma_m} \to \mu_{k,\gamma}$, as $m \to \infty$, for all $k \in \mathbb{N}$. Here*

$$\mu_{k,\gamma} = \sum_{r=0}^{k-1} \frac{1}{r+1} \binom{k}{r} \binom{k-1}{r} \gamma^r$$

is the kth moment of the distribution e_γ (see Theorem 6.0.5).

Proof. Let $L = Q\Lambda Q'$. Since by Theorem 5.3.1, the distribution of the first row q of Q has the same distribution as for the β-Hermite ensembles, it follows that Lemma 6.2.4 still applies, and we only need to compute the asymptotics for $E[(L^k)_{11}]$.

Since $L = BB^T$, with B diagonal, it follows that

$$E[(L^k)_{11}] = E[((BB^T)^k)_{11}] = E[(BB^TBB^T \dots BB^T)_{11}] = \sum B_{1i_1} B_{i_1 i_2} \dots B_{i_{2k-1}1} \ ,$$

where once again the sum is over all allowable sequences $(i_0 = 1, i_1, \dots, i_{2k-1}, i_{2k} = 1)$.

Since B is bidiagonal, it follows that

1. $|i_j - i_{j+1}| \in \{0, 1\}$, for all $0 \leq j \leq 2k-1$, and
2. $i_{2j} - i_{2j+1} \in \{0, 1\}$, for all $0 \leq j \leq k-1$, while
3. $i_{2j+1} - i_{2j+2} \in \{-1, 0\}$, for all $0 \leq j \leq k-1$.

Thus the sequences $(i_0 = 1, i_1, i_2, \ldots, i_{2k-1}, i_{2k} = 1)$ which produce non-zero terms in the above sum correspond to alternating Motzkin paths of length $2k$.

As before, we will use the notation $a = (a_m, a_{m-1}, \ldots, a_1)$ for the diagonal of B, and $b = (b_{m-1}, b_{m-2}, \ldots, b_1)$ for the subdiagonal of B.

Note that since k is fixed, and the matrix B is scaled by $\sqrt{\gamma_m/(m\beta)}$,

$$E[b_{(m-i)}^{2j}] \sim \gamma_m^j \to \gamma^j, \quad \forall 1 \leq \jmath \leq k \ , \tag{6.15}$$

while

$$E[a_{(m-i)}^{2j}] \to 1, \quad \forall 1 \leq \jmath \leq k \ . \tag{6.16}$$

If we think of the a's in term corresponding to the sequence $(i_0 = 1, i_1, \ldots, i_{2k-1}, i_{2k} = 1)$ as level steps, and of the b's as rises/falls steps (rises if they are on the odd-numbered places – since they come from B^T, falls if they are not – since they must come from B), then the summand corresponding to sequence $(i_0 = 1, i_1, i_2, \ldots, i_{2k-1}, i_{2k} = 1)$ has highest-order term γ_m^r, where r is the total number of rises.

Since we have as many summands with highest order term γ_m^r as the number of alternating Motzkin paths, it follows that

$$\mu_{m,k,\gamma_m} \sim N_k(\gamma_m) \to N_k(\gamma) = \sum_{r=0}^{k-1} \frac{1}{r+1} \binom{k}{r} \binom{k-1}{r} \gamma^r = \mu_{k,\gamma} \ .$$

This proves convergence of moments. □

We want to examine this in more detail, and find first-order terms in the above. As before, for an alternating Motzkin path p, let $u_i(p)$ be the number of rises between altitudes i and $i-1$, and let $l_i(p)$ be the number of level steps p takes from altitude i to altitude i, on odd-numbered steps. Let $\mathcal{AGD}_{k,r}$ be the set of alternating Motzkin paths of length $2k$ with r rises.

Then the expectation becomes

$$E[((BB^T)^k)_{11}] = \sum_{r=0}^{k-1} \sum_{p \in \mathcal{AGD}_{k,r}} E[b_{m-1}^{2u_1(p)}] \ \ldots \ E[b_{m-k}^{2u_k(p)}] \quad E[a_m^{2l_1(p)}] \ \ldots \ E[a_{m-k}^{2l_k(p)}] \ .$$

When we examine the b terms above more in detail we obtain that for all $1 \leq i \leq k$,

$$E[b_{(m-i)}^{2u_i(p)}] \sim \gamma_m^{u_i(p)} \left(1 - \frac{1}{m} i u_i \ + \ \frac{2}{m\beta} \binom{u_i}{2} + O(m^{-2})\right), \tag{6.17}$$

while for all $0 \leq i \leq k$,

$$E[a_{(m-i)}^{2l_i(p)}] \sim \left(1 - \frac{\gamma_m}{m} i l_i \ + \ 2\frac{\gamma_m}{m\beta}\binom{l_i}{2} + O(m^{-2})\right) \ . \tag{6.18}$$

It then follows that the first-order terms are given by

$$\mu_{m,k,\gamma} \ = \ \mu_{k,\gamma_m} - \frac{1}{m} D_k + \frac{2}{m\beta} E_k \ + \ O(m^{-2}), \quad \text{with} \tag{6.19}$$

$$D_k \ = \ \sum_{r=0}^{k-1} \sum_{p \in \mathcal{AGD}_{k,r}} \gamma_m^r \sum_{i \geq 0} i \left(u_i + \gamma_m l_i\right) \quad , \quad \text{and} \tag{6.20}$$

$$E_k \ = \ \sum_{r=0}^{k-1} \sum_{p \in \mathcal{AGD}_{k,r}} \gamma_m^r \sum_{i \geq 0} \left(\binom{u_i}{2} + \gamma_m \binom{l_i}{2}\right) \ . \tag{6.21}$$

6.4 Almost sure convergence: β-Hermite

In Section 6.2, we have examined the asymptotical behavior of the moments of the level density, obtaining a weak convergence of moments. Now we will look at the variance of those moments, and show that the first-order terms disappear – in other words, that the variance $\mathrm{Var}(X_n^k) = E[(X_n^k - \mu_{n,k})^2] = O(n^{-2})$ asymptotically, which implies that for every k, $\sum_{n=1}^{\infty} \mathrm{Var}(X_n^k) \ < \infty$.

We will thus prove the following theorem:

Theorem 6.4.1. *The set of moments $(\mu_{n,k})_{k=1,\ldots,\infty}$ converges almost surely to $(\mu_k)_{k=1,\ldots,\infty}$. Thus the distribution function F_n^β converges almost surely to the semi-circle distribution S.*

We need to examine $\mathrm{Var}(X_n^k) = E[(X_n^k - \mu_{n,k})^2] = E[(X_n^k)^2] - \mu_{n,k}^2$ for k fixed.

Since we already know the zero- and first-order expansion of $\mu_{n,k}$ for all k, we only nee to look at $E[(X_n^k)^2]$.

We need the following lemma.

Lemma 6.4.2.

$$E[(X_n^k)^2] = \frac{n\beta + 2}{n\beta} E[((A^k)_{11})^2] - \frac{2}{n\beta} E[(A^{2k})_{11}] \ .$$

Proof. Note that

$$E[(X_n^k)^2] = E\left[\frac{1}{n^2}(tr(A^k))^2\right] = \frac{1}{n^2}E[(\sum_{i=1}^n \lambda_i^k)^2] \tag{6.22}$$

$$= \frac{1}{n^2}\left(E[\sum_{i=1}^n \lambda_i^{2k}] + E[\sum_{i\neq j} \lambda_i^k \lambda_j^k]\right) \tag{6.23}$$

$$= \frac{1}{n}E[(A^{2k})_{11}] + \frac{1}{n^2}E[\sum_{i\neq j} \lambda_i^k \lambda_j^k] \ . \tag{6.24}$$

As before, let $A = Q\Lambda Q'$ and q be the first row of Q; write

$$E[((A^k)_{11})^2] = E[(\sum_{i=1}^n q_i^2 \lambda_i^k)^2] = E[\sum_{i=1}^n q_i^4 \lambda_i^{2k}] + E[\sum_{i\neq j} q_i^2 q_j^2 \lambda_i^k \lambda_j^k] \ .$$

By the independence of q and Λ, symmetry of q, and linearity of expectation, applying Lemma 6.2.1 yields

$$E[((A^k)_{11})^2] = E[q_1^4]\ E[\sum_{i=1}^n \lambda_i^{2k}] + E[q_i^2 q_2^2]\ E[\sum_{i\neq j} \lambda_i^k \lambda_j^k] \tag{6.25}$$

$$= \frac{\beta+2}{n(n\beta+2)}E[(A^{2k})_{11}] + \frac{\beta}{n(n\beta+2)}\ E[\sum_{i\neq j} \lambda_i^k \lambda_j^k] \ . \tag{6.26}$$

From (6.24) and (6.26) it follows that

$$E[(X_n^k)^2] = \frac{n\beta+2}{n\beta}E[((A^k)_{11})^2] - \frac{2}{n\beta}E[(A^{2k})_{11}] \ .$$

□

Remark 6.4.3. *It is not hard to see that using the same kind of reasoning, one could get all the moments of $\frac{1}{n}tr(A^k)$ in terms of moments of $(A^{jk})_{11}$ with $j \in \mathbb{N}$.*

Now we can prove Theorem 6.4.1.

Proof. Assume $k = 2v+1$. We need to show that

$$\frac{n\beta+2}{n\beta}E[((A^{2v+1})_{11})^2] - \frac{2}{n\beta}E[(A^{4v+2})_{11}] = O(n^{-2}) \ .$$

Note that

$$((A^{2v+1})_{11})^2 = (\sum A_{1i_1}A_{i_1i_2}\dots A_{i_{2v}1})^2 \ ,$$

with $i_0 = 1, i_1, \ldots, i_{2v}, i_{2v+1} = 1$ describing a Motzkin path.

So

$$((A^{2v+1})_{11})^2 = \sum A_{1i_1} A_{i_1 i_2} \ldots A_{i_{2v}1} A_{1j_1} \ldots A_{j_{2v}1} \ ,$$

with both $i_0 = 1, i_1, \ldots, i_{2v}, i_{2v+1} = 1$ and $j_0 = 1, j_1, \ldots, j_{2v}, j_{2v+1} = 1$ describing Motzkin paths.

Note that since the length of the paths is $2v+1$, each path has at least one level step, and in fact each path has an odd number of level steps. A level step, as before, corresponds to the presence of a diagonal term $A_{ii} = a_{n+1-i}$ in the product.

Since the powers of the diagonal terms have expectations described by (6.4) and (6.5), it follows that the highest-order term is given by exactly one level step in the left path ($i_0 = 1, i_1, \ldots, i_{2v}, i_{2v+1} = 1$) and one level step in the right path ($j_0 = 1, j_1, \ldots, j_{2v}, j_{2v+1} = 1$), both at the same altitude (i.e. corresponding to the same diagonal term). Indeed if the altitudes are different, then one has two different Gaussians in the sum, and the expectation of the summand will be 0 since the two Gaussians are independent from the rest of the terms in the summand.

Hence the terms that are asymptotically relevant are those terms which can be described by a pair of Dyck paths, an altitude, and two places at that altitude for the two level steps (one in the left path, on in the right path).

The expectation of such a term can be obtained from (6.4) and (6.5) as $\frac{1}{2n\beta}\frac{1}{4^{2v}}$. Hence

$$E[((A^{2v+1})_{11})^2] = \frac{1}{4^{2v}}\frac{1}{2n\beta} \sum_{p,q \in \mathcal{D}_v} \sum_{i \geq 0} m_i(p) m_i(q) + O(n^{-2}) \ . \tag{6.27}$$

We know by (6.12) that $E[(A^{4v+2})_{11}] = \frac{1}{4^{2v+1}} C_{2v+1} + O(n^{-1})$.

Hence

$$\begin{aligned}
E[(X_n^{2v+1})^2] &= \frac{n\beta+2}{n\beta} E[((A^{2v+1})_{11})^2] - \frac{2}{n\beta} E[(A^{4v+2})_{11}] && (6.28)\\
&= \frac{1}{4^{2v}}\frac{1}{2n\beta} \sum_{p,q \in \mathcal{D}_v} \sum_{i \geq 0} m_i(p) m_i(q) - \frac{2}{n\beta}\frac{1}{4^{2v+1}} C_{2v+1} + O(n^{-2}) && (6.29)\\
&= \frac{1}{24^{2v}} \left(\sum_{p,q \in \mathcal{D}_v} \sum_{i \geq 0} m_i(p) m_i(q) - C_{2v+1} \right) + O(n^{-2}) && (6.30)\\
&= O(n^{-2}) \ , && (6.31)
\end{aligned}$$

where the last equality uses Lemma 6.1.11 .

Assume now $k = 2v$.

We examine $E[((A^{2v})_{11})^2]$; as before,

$$((A^{2v})_{11})^2 = \sum A_{1i_1} A_{i_1 i_2} \dots A_{i_{2v-1}1} A_{1j_1} \dots A_{j_{2v-1}1} \ ,$$

with both $i_0 = 1, i_1, \dots, i_{2v-1}, i_{2v} = 1$ and $j_0 = 1, j_1, \dots, j_{2v-1}, j_{2v} = 1$ describing Motzkin paths.

Just as before, the presence of level steps in the Motzkin path (i.e. the presence of diagonal terms in the two sequences) determines the order of the summand; moreover, since k is even, the number of level steps in each path has to be even.

The zero-order terms will be given by the summands that contain no diagonal terms; these summand will also yield some first-order terms. An additional source of first-order terms will come from pairs of paths (p, q) such that exactly one has no level steps, and the other one has exactly two level steps, both at the same altitude. Assume p is a true Dyck path, and q a Dyck path to which we add two level steps, both at the same altitude. Note that from an expectation point of view, it makes no difference at what altitude we add the level steps, i.e. which Gaussian we insert in the sequence, because the Gaussians are i.i.d., and the contribution from the square of a Gaussian to the expectation of the summand is $1/(2n\beta)$ by (6.5).

As usual, let $k_i(p)$ and $k_i(q)$ be the number of rises at altitude i (corresponding to how many times we choose $b^2_{(n-i)}$ in our summand), and let $m_i(q)$ be the number of vertices in the path q, at altitude i. Then the number of ways in which one can insert two level steps at the same altitude in q is

$$\sum_{i \geq 0} \binom{m_i(q) + 1}{2} \ .$$

We obtain that the highest order terms are given by

$$E[\sum_{p,q \in \mathcal{D}_v} b_{n-1}^{2k_1(p)+2k_1(q)} b_{n-2}^{2k_1(p)+2k_2(q)} \dots b_{n-v}^{2k_v(p)+2k_v(q)}] \ ;$$

by (6.3), following the same line of thought that lead to the equation (6.12), the above

is equal to

$$\frac{1}{4^{2v}}C_v^2 - \frac{D_v}{4^{2v}n} + 2\frac{E_v}{4^{2v}n\beta} + O(n^{-2}) \ , \quad \text{where} \tag{6.32}$$

$$D_v = \sum_{p,q\in\mathcal{D}_v}\sum_{i\le 0} i\ (k_i(p)+k_i(q)) \quad , \quad \text{and} \tag{6.33}$$

$$E_v = \left(\sum_{p,q\in\mathcal{D}_v}\sum_{i\le 0}\binom{k_i(p)+k_i(q)}{2}\right) \ . \tag{6.34}$$

The second source of first-order terms, as mentioned, comes from summands that correspond to choosing a Dyck path and a Dyck path plus two level steps; the first-order contribution from those terms is

$$\frac{1}{2n\beta}2E[\sum_{p\in\mathcal{D}_v}\sum_{q\in\mathcal{D}_{v-1}}\left(\sum_{i\ge 0}\binom{m_i(q)+1}{2}\right)b_{n-1}^{2k_1(p)+2k_1(q)}\cdots b_{n-v}^{2k_v(p)+2k_v(q)}] \ , \tag{6.35}$$

the above is equal to

$$\frac{2}{2n\beta}\frac{1}{4^{2v-1}}\sum_{p\in\mathcal{D}_v}\sum_{q\in\mathcal{D}_{v-1}}\left(\sum_{i\ge 0}\binom{m_i(q)+1}{2}\right) \ .$$

Putting (6.32) and (6.35) together we obtain that

$$E[((A^{2v})_{11})^2] = \frac{1}{4^{2v}}C_v^2 - \frac{D_v}{4^{2v-1}n} + 2\frac{E_v}{4^{2v}n\beta} + \frac{F_v}{4^{2v}n\beta} + O(n^{-2}) \quad \text{where} \tag{6.36}$$

$$D_v = \sum_{p,q\in\mathcal{D}_v}\sum_{i\le 0} i\ (k_i(p)+k_i(q)) \quad , \quad \text{and} \tag{6.37}$$

$$E_v = \sum_{p,q\in\mathcal{D}_v}\sum_{i\le 0}\binom{k_i(p)+k_i(q)}{2} \quad , \quad \text{and} \tag{6.38}$$

$$F_v = \sum_{p\in\mathcal{D}_v}\sum_{q\in\mathcal{D}_{v-1}}\sum_{i\le 0}\binom{m_i(q)+1}{2} \ . \tag{6.39}$$

On the other hand, when we square $\mu_{n,k}$ (see 6.12), we obtain

$$\mu_{n,k}^2 = \frac{1}{4^{2v}}C_v^2 - \frac{\tilde{D}_v}{4^{2v}n} + \frac{\tilde{E}_v}{4^{2v}n\beta} + \frac{\tilde{F}_v}{4^{2v}n\beta} + O(n^{-2}) \quad \text{where} \tag{6.40}$$

$$\tilde{D}_v = 2C_v\sum_{p\in\mathcal{D}_v}\sum_{i\le 0} i\ k_i(p) \quad , \quad \text{and} \tag{6.41}$$

$$\tilde{E}_v = 2C_v\sum_{p\in\mathcal{D}_v}\sum_{i\le 0}\binom{k_i(p)}{2} \quad , \quad \text{and} \tag{6.42}$$

$$\tilde{F}_v = 4C_v\sum_{q\in\mathcal{D}_{v-1}}\sum_{i\ge 0}\binom{m_i(q)+1}{2} \ . \tag{6.43}$$

Note that $D_v = \tilde{D}_v$, and that $F_v = \tilde{F}_v$.

Hence, when we compute $\frac{n\beta+2}{n\beta} E[((A^{2v})_{11})^2] - \mu_{n,k}^2$, we get some cancellations; what remains is

$$\frac{n\beta+2}{n\beta} E[((A^{2v})_{11})^2] - \mu_{n,k}^2 = \frac{2}{n\beta}\frac{1}{4^{2v}} \left(C_v^2 + E_v - \tilde{E}_v \right) \ . \tag{6.44}$$

But

$$\begin{aligned} E_v - \tilde{E}_v &= \sum_{p,q \in \mathcal{D}_v} \sum_{i \geq 0} \binom{k_i(p) + k_i(p)}{2} - 2C_v \sum_{p \in \mathcal{D}_v} \sum_{i \leq 0} \binom{k_i(p)}{2} & (6.45) \\ &= \sum_{p,q \in \mathcal{D}_v} \sum_{i \geq 0} \left(\binom{k_i(p) + k_i(p)}{2} - \binom{k_i(p)}{2} - \binom{k_i(q)}{2} \right) & (6.46) \\ &= \sum_{p,q \in \mathcal{D}_v} \sum_{i \geq 0} k_i(p) k_i(q) \ . & (6.47) \end{aligned}$$

On the other hand,

$$E[(A^{4v})_{11}] = \frac{1}{4^{2v}} C_{2v} + O(n^{-1}) \ . \tag{6.48}$$

Putting (6.44), (6.47), (6.48) together, it follows that

$$\begin{aligned} E[(X_n^k)^2] - \mu_{n,k}^2 &= \frac{n\beta+2}{n\beta} E[((A^k)_{11})^2] - \frac{2}{n\beta} E[(A^{2k})_{11}] - \mu_{n,k}^2 & (6.49) \\ &= \frac{2}{n\beta}\frac{1}{4^{2v}} \left(C_v^2 + \sum_{p,q \in \mathcal{D}_v} \sum_{i \geq 0} k_i(p) k_i(q) - C_{2v} \right) + O(n^{-2}) & (6.50) \end{aligned}$$

By Lemma 6.1.10, the quantity between parentheses is 0, and hence

$$Var[X_n, k] = E[(X_n^k)^2] - \mu_{n,k}^2 = O(n^{-2}) \ .$$

□

6.5 Almost sure convergence: β-Laguerre

In Section 6.3, we have examined the asymptotical behavior of the moments of the level density, obtaining a weak convergence of moments. Now we will look at the variance of those moments, and show that the first-order terms disappear – in other words, that the variance $\mathrm{Var}(Y_m^k) = E[(Y_m^k - \mu_{m,k,\gamma_m})^2] = O(m^{-2})$ asymptotically, which implies that for every k, $\sum_{m=1}^{\infty} \mathrm{Var}(Y_m^k) < \infty$.

We will thus prove the following theorem:

Theorem 6.5.1. *The set of moments $(\mu_{m,k,\gamma_m})_{k=1,\ldots,\infty}$ converges almost surely to $(\mu_{k,\gamma})_{k=1,\ldots,\infty}$. Thus the distribution function F^{β}_{m,γ_m} converges almost surely to the distribution e_γ.*

Proof. We need to examine $\mathrm{Var}(Y_m^k) = E[(Y_m^k - \mu^2_{m,k,\gamma_m}] = E[(Y_m^k)^2] - \mu^2_{m,k,\gamma_m}$ for k fixed.

Note that Lemma 6.4.2 still applies (the distribution of q, the first row of the eigenvector matrix for L, is the same as in the Hermite case, and symmetry and independence of q and the eigenvalues of L still apply).

Thus

$$Var[(Y_m^k)] = \frac{m\beta+2}{m\beta}E[((L^k)_{11})^2] - \frac{2}{m\beta}E[(L^{2k})_{11}] - \mu^2_{m,k,\gamma_m} \ ;$$

hence we need to find zero- and first-order behavior for $E[((L^k)_{11})^2]$ and μ^2_{m,k,γ_m}, and zero-order behavior for $E[(L^{2k})_{11}]$. The latter two we obtain from (6.19).

As usual, $u_i(p)$ is the number of rises in a path $p \in \mathcal{AGD}$ at level i and $l_i(p)$ is the number of level steps at altitude i on odd-numbered steps.

We obtain

$$\begin{aligned}
E[(L^{2k})_{11}] &= \mu_{2k,\gamma_m} + O(m^{-1}) \ , && (6.51)\\
\mu^2_{m,k,\gamma_m} &= \mu^2_{k,\gamma_m} - \frac{D_v}{m} + \frac{E_v}{m\beta} + O(m^{-2}) \ , \text{ with} && (6.52)\\
D_v &= 2\mu_{k,\gamma_m}\sum_{r=0}^{k-1}\gamma_m^r \sum_{p\in\mathcal{AGD}_{k,r}}\sum_{i\geq 0} i \ (u_i(p) + \gamma_m l_i(p)) \ , \text{ and} && (6.53)\\
E_v &= 2\mu_{k,\gamma_m}\sum_{r=0}^{k-1}\sum_{p\in\mathcal{AGD}_{k,r}}\gamma_m^r \sum_{i\geq 0}\left(\binom{u_i(p)}{2} + \gamma_m\binom{l_i(p)}{2}\right) \ . && (6.54)
\end{aligned}$$

Let us now turn our attention to $E[((L^k)_{11})^2]$. We have that

$$E[((L^k)_{11})^2] = E[(((BB^T)^k)_{11})^2] = E[\sum(B_{1i_1}Bi_1i_2\ldots B_{i_{2k-1}1}B_{1j_1}B_{j_1j_2}\ldots B_{j_{2k-1}1})] \ ,$$

where $i_0 = 1, i_1, \ldots, i_{2k-1}, i_{2k} = 1$ and $j_0 = 1, j_1, \ldots, j_{2k-1}, j_{2k} = 1$ are sequences corresponding to alternating Motzkin paths. Hence, following the notational conventions

for u_i and l_i,

$$E[((L^k)_{11})^2] = \sum_{r_1,r_2=0}^{k-1} \sum_{\substack{p \in \mathcal{AGD}_{k,r_1} \\ q \in \mathcal{AGD}_{k,r_2}}} E\left[b_{n-1}^{2u_1(p)+2u_1(q)} b_{n-2}^{2u_2(p)+2u_2(q)} \dots b_{n-k}^{2u_k(p)+2u_k(q)}\right] \times (6.55)$$

$$\times \; E\left[a_n^{2l_1(p)+2l_1(q)} a_{n-2}^{2l_2(p)+2l_2(q)} \dots a_{n-k}^{2l_k(p)+2l_k(q)}\right] \; , \; (6.56)$$

and using (6.17) and (6.18) we obtain

$$E[((L^k)_{11})^2] = \mu_{k,\gamma_m}^2 - \frac{\tilde{D}_v}{m} + 2\frac{\tilde{E}_v}{m\beta} + O(m^{-2}) \; , \text{ with} \quad (6.57)$$

$$\tilde{D}_v = \sum_{r_1,r_2=0}^{k-1} \sum_{\substack{p \in \mathcal{AGD}_{k,r_1} \\ q \in \mathcal{AGD}_{k,r_2}}} \gamma_m^{r_1+r_2} \sum_{i \geq 0} (u_i(p) + u_i(q) + \gamma_m(l_i(p) + l_i(q))) \; , \quad (6.58)$$

$$\tilde{E}_v = \sum_{r_1,r_2=0}^{k-1} \sum_{\substack{p \in \mathcal{AGD}_{k,r_1} \\ q \in \mathcal{AGD}_{k,r_2}}} \gamma_m^{r_1+r_2} \sum_{i \geq 0} \binom{u_i(p) + u_i(q)}{2} + \gamma_m \binom{l_i(p) + l_i(q)}{2} \; . \quad (6.59)$$

Since, upon examination, $\tilde{D}_v = D_v$, using (6.51 - 6.54) and (6.57-6.59), we can write

$$\begin{aligned} E[(Y_m^k)^2] &= \frac{m\beta + 2}{m\beta} E[((L^k)_{11})^2] - \frac{2}{m\beta} E[(L^{2k})_{11}] - \mu_{m,k,\gamma_m}^2 \\ &= \frac{2}{m\beta}(\mu_{k,\gamma_m} + \tilde{E}_v - E_v - \mu_{2k,\gamma_m}) + O(m^{-2}) \; . \end{aligned}$$

But we can write

$$\begin{aligned} \tilde{E}_v - E_v &= \sum_{r_1,r_2=0}^{k-1} \sum_{\substack{p \in \mathcal{AGD}_{k,r_1} \\ q \in \mathcal{AGD}_{k,r_2}}} \gamma_m^{r_1+r_2} \left(\sum_{i \geq 0} \binom{u_i(p) + u_i(q)}{2} - \binom{u_i(p)}{2} - \binom{u_i(q)}{2} \; + \right. \\ &\qquad \left. + \; \gamma_m \left(\binom{l_i(p) + l_i(q)}{2} - \binom{l_i(p)}{2} - \binom{l_i(q)}{2} \right) \right) \\ &= \sum_{r_1,r_2=0}^{k-1} \sum_{\substack{p \in \mathcal{AGD}_{k,r_1} \\ q \in \mathcal{AGD}_{k,r_2}}} \gamma_m^{r_1+r_2} \sum_{i \geq 0} u_i(p)u_i(q) + \gamma_m l_i(p) l_i(q) \; ; \end{aligned}$$

by Lemma 6.1.12 this means that

$$\tilde{E}_v - E_v = N_{2k}(\gamma_m) - N_k(\gamma_m)^2 \; ;$$

and since $\mu_{k,\gamma_m} = N_k(\gamma_m)$, for all k, it follows that

$$E[(Y_m^k)^2] \;=\; O(m^{-2}) \;,$$

for all k. $\square$

Chapter 7

Eigenvalue distributions for large β

In Chapter 6, we study the asymptotics of the empirical distribution of the β-Hermite and β-Laguerre ensembles when β is fixed and $n \to \infty$. The approach we use in that chapter is combinatorial; we use path-counting methods to prove our results.

In this chapter, we switch the focus from n to β, and study the *individual* eigenvalue asymptotics of the β-Hermite and β-Laguerre ensembles at fixed n and $\beta \to \infty$. We use perturbation theory and probability results to prove these asymptotics. As a bonus, we also obtain level density asymptotics for the β-Hermite and β-Laguerre ensembles.

Remark 7.0.2. *It is worth noting that while the approaches we take in Chapters 6 and 7 are quite different, both approaches are based on the matrix models.*

As we mentioned in Chapter 2, the β parameter in either the Hermite case or the Laguerre case can be thought of as an inverse temperature, while the presence of the Vandermonde determinant $\Delta(\Lambda)$ functions as a repelling factor. As β goes to 0, the strength of this repelling factor decreases until it is annihilated; the interdependence disappears, and the randomness (in a certain sense) increases (as the temperature $1/\beta$ increases to ∞, the entropy reaches its maximum).

In the Hermite case, at $\beta = 0$, the eigenvalues become i.i.d. variables with standard normal distribution, whereas in the Laguerre case they become i.i.d variables with distribution χ^2_{2a}.

What about when β increases to ∞? As one might guess, the effect is opposite: the eigenvalues "freeze", randomness decreases. For n fixed, the eigenvalues approach the roots of the corresponding Hermite polynomial, respectively, the roots of the corresponding Laguerre polynomial. At $\beta = \infty$ $(1/\beta = 0)$ we are at a "freezing point"; the eigenvalues are no longer random, but fixed at the roots of the Hermite or Laguerre polynomial.

In this chapter, we use our matrix models to study the way in which this convergence occurs at large β. We find zero- and first-order approximations for any eigenvalue at large β, and we use these approximations to estimate the level densities at large β; we find that these approximations work surprisingly well for relatively small values of β.

7.1 Eigenvalue distributions for β large; zero- and first-order approximations

In this section we show how, for fixed n and β large, we can obtain zero- and first-order approximations for the eigenvalues.

Throughout this section, we will use the concept of variables on the same probability space; intuitively, and from a computational point of view, this is equivalent to random variable generation from the same group of random bits.

To provide an example, assume that we are able to generate a variable t from the uniform distribution on $[0,\ 1]$. Then we can generate any variable X with known c.d.f. F, by the following process: we generate $Y = t$, then solve $F(r) = t$, and return $X = r$.

Remark 7.1.1. *Note that with probability* 1*, the solution is unique (if it is not, it follows that there exists a smallest* r_{min} *such that* $F(r_{min}) = t$*, and a largest* r_{max} *such that* $F(r_{min}) = t$ *(because* F *is non-decreasing, and not identically constant). Since any value between* r_{min} *and* r_{max} *corresponds to the same* t*,* $Pr[X \in [r_{min},\ r_{max}]] = 0$*. But since* $F(r_{min}) = F(r_{max})$*, this is consistent with being a* 0*-probability event.*

The process is illustrated in Figure 7-1.

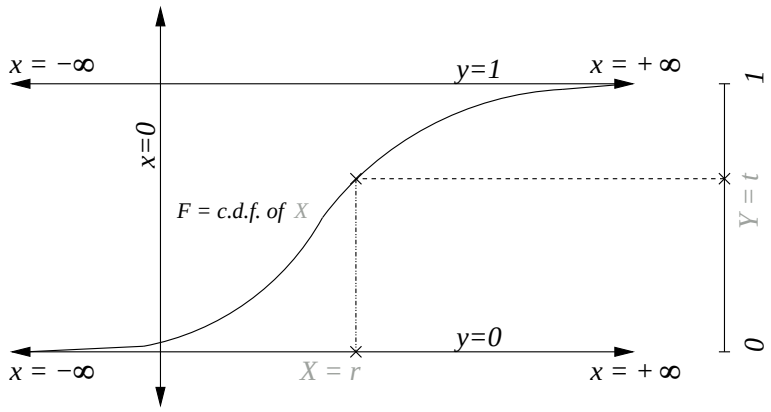

Figure 7-1: How to generate a random variable X from a known distribution F, assuming that you can generate a random variable Y from the uniform distribution on $[0,\ 1]$. X and Y (together with their values, r and t) are in lighter font.

Throughout this chapter, when we talk about a set of random variables on the same probability space, we think of them as being all generated by the same value t of the uniform distribution variable Y (thus dependent on the same random "bits").

We can now present the following crucial lemma.

Lemma 7.1.2. *For $n \geq \mathbb{N}$, let $r_n \in \mathbb{R}_+$, and let $(X_n)_{n\in\mathbb{N}}$ be a set of variables lying on the same probability space, such that X_n has distribution χ_{r_n}. If $\lim_{n\to\infty} r_n = \infty$, then $X_n - \sqrt{r_n}$ converges almost surely to a normal distribution of variance $\frac{1}{2}$ (which lies in the same probability space).*

Proof. We prove this lemma by looking at the density function of χ_{r_n} when $r_n \to \infty$.

First we will show that $E[X_n] - \sqrt{r_n} \to 0$ as $n \to \infty$. Since

$$E[X_n] = \frac{2^{1-r_n/2}}{\Gamma\left(\frac{r_n}{2}\right)} \frac{\Gamma\left(\frac{r_n+1}{2}\right)}{2^{1-(r_n+1)/2}} = \frac{1}{\sqrt{2}} \frac{\Gamma\left(\frac{r_n+1}{2}\right)}{\Gamma\left(\frac{r_n}{2}\right)} \ . \tag{7.1}$$

Using the Lanczos approximation formula

$$\Gamma(z+1) \ = \ z^{z+1/2} e^{-z} \sqrt{2\pi} \left(c_0 + \frac{c_1}{z} + O(\frac{1}{z^2}) \right) \ , \tag{7.2}$$

for r_n large, (7.1) becomes

$$\begin{aligned} E[X_n] &= \frac{1}{\sqrt{2}} \frac{\left(\frac{r_n-1}{2}\right)^{(r_n-1)/2} e^{-(r_n-1)/2} \sqrt{2\pi(r_n-1)/2}}{\left(\frac{r_n-2}{2}\right)^{(r_n-2)/2} e^{-(r_n-2)/2} \sqrt{2\pi(r_n-2)/2}} \left(1+O(r_n^{-1})\right) , \\ &= \sqrt{r_n} \, \frac{(r_n-1)^{(r_n-2)/2} e^{-1/2}}{(r_n-2)^{(r_n-2)/2}} \left(1+O(r_n^{-1})\right) , \\ &= \sqrt{r_n} \left(1+O(r_n^{-1})\right) . \end{aligned}$$

So as $r_n \to \infty$, the mean of $X_n - \sqrt{r_n}$ approaches 0. Now let us examine the p.d.f. of $X_n - \sqrt{r_n}$; denote by $Y_n = X_n - \sqrt{r_n}$. Then the p.d.f. of Y_n is

$$f_n(t) = \frac{2^{1-r_n/2}}{\Gamma\left(\frac{r_n}{2}\right)} \, (t+\sqrt{r_n})^{r_n-1} e^{-(t+\sqrt{r_n})^2/2} .$$

We examine this p.d.f. in a "small" neighborhood of 0, such that $t = o(r_n^{1/2})$. Again, we use the Lanczos approximation (7.2) for the Gamma function at infinity, to obtain

$$\begin{aligned} f_n(t) &= \frac{2^{1-r_n/2} \; r_n^{r_n/2-1} \; \left(1+\frac{t}{r_n}\right)^{r_n-1} \; e^{-t^2/2-\sqrt{r_n}t-r_n/2}}{\left(\frac{r_n}{2}-1\right)^{r_n/2-1} \; e^{-r_n/2+1} \; \sqrt{\pi r_n}} \; \left(1+O(\frac{1}{r_n})\right) \quad , \\ &= \frac{1}{\sqrt{\pi}} \; \left(1+\frac{t}{\sqrt{r_n}}\right)^{r_n-1} \; e^{-t^2/2-\sqrt{r_n}t} \; \left(1+O(\frac{1}{r_n})\right) , \end{aligned}$$

and by using the Taylor series for $(r_n-1)\ln(1+t/\sqrt{r_n})$ and the fact that $t = o(r_n^{1/2})$, we obtain

$$f_n(t) = \frac{1}{\sqrt{\pi}} e^{-t^2} \; \left(1+O\left(\frac{t}{\sqrt{r_n}}\right)\right) . \tag{7.3}$$

Thus, on any fixed interval, the p.d.f. of $X_n - \sqrt{r_n}$ converges to the p.d.f. of a centered normal of variance 1/2. This is enough to prove that the c.d.f. of $X_n - \sqrt{r_n}$ converges to the c.d.f. of a centered normal of variance 1/2.

We now use the fact that the variables $X_n - \sqrt{r_n}$ are on the same probability space with the variable Y which corresponds to the normal. Let $t_1 = -\ln r_n$ and $t_2 = \ln r_n$; the maximum amount by which $X_n - \sqrt{r_n}$ can differ from Y given that $X_n - \sqrt{r_n}$ is between t_1 and t_2 is then $O\left(\frac{(\ln r_n)^2}{\sqrt{r_n}}\right)$, from (7.3).

Taking derivatives, one obtains that the maximum of the function $x^{r_n-1}e^{-x^2/2}$ is obtained at $x = \sqrt{r_n-1} > \sqrt{r_n} - \ln r_n$ since $r_n \to \infty$; the function is increasing

before and decreasing after that point. Hence on the interval $[0,\ \sqrt{r_n} - \ln r_n]$ this allows us to bound from above the probability that X_n is smaller than $\sqrt{r_n} - \ln r_n$:

$$\begin{aligned}
\Pr[X_n \leq \sqrt{r_n} - \ln r_n] &= \frac{2^{1-r_n/2}}{\Gamma(\frac{r_n}{2})} \int_0^{\sqrt{r_n}-\ln r_n} t^{r_n-1} e^{-t^2/2}\, dt \\
&< \frac{2^{1-r_n/2}}{\Gamma(\frac{r_n}{2})} (\sqrt{r_n} - \ln r_n)^{r_n} e^{-(\sqrt{r_n}-\ln r_n)^2/2} \\
&= O(r_n e^{-(\ln r_n)^2/2} = O(r_n^{-\ln r_n/2-1}) \ .
\end{aligned}$$

Similarly, the probability that X_n is larger than $\sqrt{r_n} + \ln r_n$ is found to be $O(r_n^{-\ln r_n/2-1})$.

Thus, given a certain ϵ,

$$\begin{aligned}
\Pr[|X_n - \sqrt{r_n} - Y| > \epsilon] &= \Pr\big[|X_n - \sqrt{r_n} - Y| > \epsilon \mid X_n \in [t_1, t_2]\big]\ \Pr[X_n \in [t_1, t_2]] + \\
&\quad + \Pr\big[|X_n - \sqrt{r_n} - Y| > \epsilon \mid X_n | X_n \leq t_1\big]\ \Pr[X_n \leq t_1] + \\
&\quad + \Pr\big[|X_n - \sqrt{r_n} - Y| > \epsilon \mid X_n | X_n \geq t_1\big]\ \Pr[X_n \geq t_1] \\
&< O(r_n^{-\ln r_n/2-1}) \ ,
\end{aligned}$$

since $\Pr[|X_n - Y| > \epsilon \mid X_n \in [t_1, t_2]]$ becomes 0 as soon as n is large enough.

But then

$$\sum_{n=1}^{\infty} \Pr[|X_n - Y| > \epsilon] = O(1) + \sum_{n=1}^{\infty} O(r_n^{-\ln r_n/2-1}) < \infty \ ,$$

and hence we have almost sure convergence of $X_n - \sqrt{r_n}$ to Y. □

Lemma 7.1.2 will turn out to be crucial for the rest of the chapter.

7.1.1 The Hermite case

Let n be fixed, and $(\beta_m)_{m\geq 1}$ be a sequence of positive reals going to infinity.

Let $h_1^{(n)}, \ldots, h_n^{(n)}$ be the roots of the nth univariate Hermite polynomial H_n (where the Hermite polynomials $H_0(x), H_1(x), \ldots$ are orthonormal with respect to the weight e^{-x^2} on $(-\infty, \infty)$).

Let A_1, A_2, …, A_m, …, be random matrices from the β_1, β_2, …, β_m, …, -Hermite ensembles of size n (fixed), scaled by $1/\sqrt{2n\beta_1}$, $1/\sqrt{2n\beta_2}$, …, $1/\sqrt{2n\beta_m}$, ….

Furthermore, assume that for each pair (i,j) with $|i-j| \leq 1$, the variables $A_1(i,j), \ldots, A_m(i,j), \ldots$ are on the same probability space (in particular, if $i=j$, this means that $A_m(i,i) = A_{m'}(i,i)$ for all i, m, m', for each realization of the Gaussian).

We can now state and prove the following theorem.

Theorem 7.1.3. *Let $\lambda_i(A_m)$ be the ith largest eigenvalue of A_m, for any fixed $1 \leq i \leq n$. Then, as $m \to \infty$,*

$$\sqrt{\beta_m}\ \left(\lambda_i(A_m) - \frac{1}{\sqrt{2n}} h_i^{(n)}\right) \stackrel{a.s.}{\to} \frac{1}{\sqrt{2n}}\ G_i\ ,$$

where G_i is a Gaussian with mean 0 and variance

$$Var(G_i) \ = \ \frac{\sum_{j=0}^{n-1} H_j^4(h_i^{(n)}) + \sum_{j=0}^{n-2} H_{j+1}^2(h_i^{(n)}) H_j^2(h_i^{(n)})}{\left(\sum_{j=0}^{n-1} H_j^2(h_i^{(n)})\right)^2}\ .$$

Proof. Let H be the $n \times n$ symmetric tridiagonal matrix

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & \sqrt{n-1} & & & & \\ \sqrt{n-1} & 0 & \sqrt{n-2} & & & \\ & \sqrt{n-2} & 0 & & & \\ & & & \ddots & & \\ & & & & 0 & \sqrt{1} \\ & & & & \sqrt{1} & 0 \end{pmatrix}\ . \tag{7.4}$$

Orthogonal polynomial/tridiagonal symmetric matrix theory tell us that the first matrix has the following property: its eigenvalues are the roots of the nth Hermite polynomial $H_n(x)$ (recall that we denoted them by $h_1^{(n)}, \ldots, h_n^{(n)}$), and an eigenvector corresponding to the ith eigenvalue $h_i^{(n)}$ is

$$v_i = \begin{pmatrix} H_{n-1}(h_i^{(n)}) \\ H_{n-2}(h_i^{(n)}) \\ \vdots \\ H_1(h_i^{(n)}) \\ H_0(h_i^{(n)}) \end{pmatrix}\ .$$

We define $\phi_i \equiv v_i/||v_i||_2$ to be a length 1 eigenvector which corresponds to $h_i^{(n)}$.

Lemma 7.1.4. *Let A_1, ..., A_m ..., be as defined in the beginning of this section. Then almost surely*

$$\lim_{m\to\infty} \sqrt{2n\beta_m} A_m - \sqrt{\beta_m} H = Z \ ,$$

where Z is a tridiagonal matrix with standard normal variables on the diagonal and normal variables of mean 0 and variance 1/4 on the subdiagonal. All normal variables in Z are mutually independent, subject only to the symmetry.

Lemma 7.1.4 follows immediately from Lemma 7.1.2, since we are dealing with a finite number $(n-1)$ of χ variables on the sub-diagonal of A_m, each converging almost surely to a normal variable.

Hence we have

$$A_m = \frac{1}{\sqrt{2n}} H + \frac{1}{\sqrt{2n\beta_m}} Z + o\left(\frac{1}{\sqrt{\beta_m}}\right) \ ,$$

almost surely as $m \to \infty$.

Thus all zero- and first-order properties of A_m are the same as for the simpler model $\frac{1}{\sqrt{2n}} H + \frac{1}{\sqrt{2n\beta_m}} Z$, where Z as as above. In particular, for any $1 \leq i \leq n$,

$$\lambda_i(A_m) = \lambda_i\left(\frac{1}{\sqrt{2n}} H + \frac{1}{\sqrt{2n\beta_m}} Z\right) + o\left(\frac{1}{\sqrt{\beta_m}}\right) \ ,$$

almost surely as $m \to \infty$.

Finally, with the help of Perturbation Theory Lemma 4.4.1, we obtain that for any $1 \leq i \leq n$,

$$\lambda_i(A_m) = \frac{1}{\sqrt{2n}} h_i^{(n)} + \frac{1}{\sqrt{2n\beta_m}} \psi_i^T Z \psi_i + o\left(\frac{1}{\sqrt{\beta_m}}\right) \ ,$$

almost surely as $m \to \infty$.

Since $G_i \equiv \psi_i^T Z \psi_i$ is a Gaussian of variance

$$\mathrm{Var}(G_i) \ = \ \frac{\sum_{j=0}^{n-1} H_j^4(h_i^{(n)}) + \sum_{j=0}^{n-2} H_{j+1}^2(h_i^{(n)}) H_j^2(h_i^{(n)})}{\left(\sum_{j=0}^{n-1} H_j^2(h_i^{(n)})\right)^2} \ ,$$

the proof of Theorem 7.1.3 is complete. □

7.1.2 The Laguerre case

Let n be fixed, and $(\beta_m)_{m\geq 1}$ be a sequence of positive reals going to infinity.

Given $\gamma > 0$, let $l_1^{(n)}, \ldots, l_n^{(n)}$ be the roots of the nth univariate Laguerre polynomial of parameter $\gamma - 1$, $L_n^{\gamma-1}$ (for any $\gamma \geq 0$, the Laguerre polynomials $L_0^\gamma, L_1^\gamma, \ldots$ are orthonormal with respect to the weight $x^\gamma e^{-x}$ on $[0, \infty)$; for $\gamma \in (-1, 0)$ they admit formal definitions).

Let B_1, B_2, ..., B_m, ..., be random matrices from the β_1, β_2, ..., β_m, ..., -Laguerre ensembles of size n and parameters a_1, a_2, ..., a_m, ..., scaled by $1/n\beta_1$, $1/n\beta_2$, ..., $1/n\beta_m$, ... Suppose that

$$\lim_{m\to\infty} \frac{a_m}{\beta_m} = \frac{1}{2}(n + \gamma - 1) \ .$$

Note that the requirement $a_m > (n-1)\beta_m/2$ constrains γ to be positive.

Furthermore, assume that for each i and j with $|i - j| \leq 1$, all of the (i, j) entries of the matrices B_m are on the same probability space.

Theorem 7.1.5. *Let $\lambda_i(B_m)$ be the ith largest eigenvalue of B_m, for any fixed $1 \leq i \leq n$. Then, as $m \to \infty$,*

$$\sqrt{\beta_m} \ \left(\lambda_i(B_m) - \frac{1}{n} l_i^{(n)}\right) \stackrel{a.s.}{\to} \frac{1}{n} G_i \ ,$$

where G_i is a Gaussian with mean 0 and variance

$$Var(G_i) \ = \ 2 \ \frac{(\gamma + n - 1)(L_{n-1}^\gamma(l_i^{(n)}))^4 + A_n + B_n + 2C_n}{\left(\sum_{j=0}^{n-1} (L_j^\gamma(l_i^{(n)}))^2\right)^2} \ ,$$

where

$$\begin{aligned}
A_n &= \sum_{j=1}^{n-1} (\gamma + 2(n-j) - 1)(L_{n-j-1}^\gamma(l_i^{(n)}))^4 \ , \\
B_n &= \sum_{j=1}^{n-1} (\gamma + 2(n-j))(L_{n-j-1}^\gamma(l_i^{(n)}))^2 (L_{n-j}^\gamma(l_i^{(n)}))^2 \ , \quad \textit{and} \\
C_n &= \sum_{j=1}^{n-1} \sqrt{\gamma+n-j}\sqrt{n-j} \ \left((L_{n-j-1}^\gamma(l_i^{(n)}))^3 L_{n-j}^\gamma(l_i^{(n)}) \ + \ L_{n-j-1}^\gamma(l_i^{(n)})(L_{n-j}^\gamma(l_i^{(n)}))^3\right) \ .
\end{aligned}$$

Proof. The proof follows in the footsteps of that of Theorem 7.1.3.

Let L_γ be the $n \times n$ (symmetric) positive definite matrix

$$L_\gamma = \begin{pmatrix} \gamma+n-1 & \sqrt{\gamma+n-1}\sqrt{n-1} & & & & \\ \sqrt{\gamma+n-1}\sqrt{n-1} & 2(n-2)+\gamma+1 & \sqrt{\gamma+n-2}\sqrt{n-2} & & & \\ & \sqrt{\gamma+n-2}\sqrt{n-2} & 2(n-3)+\gamma+1 & & & \\ & & \ddots & & & \\ & & & \sqrt{\gamma+2}\sqrt{2} & & \\ & & \sqrt{\gamma+2}\sqrt{2} & 3+\gamma & \sqrt{\gamma+1}\sqrt{1} & \\ & & & \sqrt{\gamma+1}\sqrt{1} & 1+\gamma & \end{pmatrix} \tag{7.5}$$

We can write $L_\gamma = B_\gamma B_\gamma^T$, with

$$B_\gamma = \begin{pmatrix} \sqrt{\gamma-n+1} & & & & \\ \sqrt{n-1} & \sqrt{\gamma-n} & & & \\ & \ddots & \ddots & & \\ & & \sqrt{2} & \sqrt{\gamma+1} & \\ & & & \sqrt{1} & \sqrt{\gamma} \end{pmatrix} . \tag{7.6}$$

The matrix L_γ has as eigenvalues the roots of the nth Laguerre polynomial of parameter $\gamma - 1$, $L_n^{\gamma-1}(x)$ (recall that we have denoted them by $l_1^{(n)}, \ldots, l_n^{(n)}$), and an eigenvector corresponding to the ith eigenvalue $l_i^{(n)}$ is

$$w_i = \begin{pmatrix} L_{n-1}^\gamma(l_i^{(n)}) \\ L_{n-2}^\gamma(l_i^{(n)}) \\ \vdots \\ L_1^\gamma(l_i^{(n)}) \\ L_0^\gamma(l_i^{(n)}) \end{pmatrix} .$$

We define $\phi_i \equiv w_i/||w_i||_2$ to be a length 1 eigenvector corresponding to the ith eigenvalue l_i.

Lemma 7.1.6. *Let B_1, ..., B_m ..., be as in the statement of Theorem 7.1.5. Then almost surely*

$$\lim_{m\to\infty} n\beta_m B_m - \beta_m L^\gamma = \frac{1}{\sqrt{2}}(B_\gamma Z^T + Z B_\gamma^T) ,$$

where Z is a lower bidiagonal matrix with standard normal variables on the diagonal and normal variables of mean 0 and variance 1 on the subdiagonal. All normal variables in Z are mutually independent, subject only to the symmetry.

Once again, the proof for this lemma follows from the construction of the Laguerre matrices as a lower bidiagonal random matrix times its transpose, and from Lemma 7.1.2 applied to the χ entries on the bidiagonal random matrices (there is a finite number $2n-1$ of them).

Just as in the Hermite case, Lemma 7.1.6 allows us to write

$$B_m = \frac{1}{n} L_\gamma + \frac{1}{n\sqrt{2\beta_m}}(B_\gamma Z^T + Z B_\gamma^T) + o\left(\frac{1}{\sqrt{\beta_m}}\right) ,$$

almost surely, as $m \to \infty$. Thus once again,

$$\lambda_i(B_m) = \lambda_i(\frac{1}{n} L_\gamma + \frac{1}{n\sqrt{2\beta_m}}(B_\gamma Z^T + Z B_\gamma^T)) + o\left(\frac{1}{\sqrt{\beta_m}}\right) ,$$

almost surely, as $m \to \infty$.

Finally, Perturbation Theory Lemma 4.4.1 applies once again to yield that almost surely as $m \to \infty$,

$$\lambda_i(B_m) = \frac{1}{n} l_i^{(n)} + \frac{1}{n\sqrt{2\beta_m}} \phi_i^T (B_\gamma Z^T + Z B_\gamma^T)) \phi_i + o\left(\frac{1}{\sqrt{\beta_m}}\right) .$$

Note that $\phi_i^T B_\gamma Z^T \phi_i = \phi_i^T Z B_\gamma^T \phi_i$; thus $G_i \equiv \sqrt{2}\phi_i^T B_\gamma Z^T \phi_i$ is a Gaussian of mean 0 and variance

$$\mathrm{Var}(G_i) = 2\, \frac{(\gamma+n-1)(L_{n-1}^\gamma(l_i^{(n)}))^4 + A_n + B_n + 2C_n}{\left(\sum_{j=0}^{n-1} (L_j^\gamma(l_i^{(n)}))^2\right)^2} ,$$

with

$$\begin{aligned}
A_n &= \sum_{j=1}^{n-1} (\gamma + 2(n-j) - 1)(L_{n-j-1}^\gamma(l_i^{(n)}))^4 , \\
B_n &= \sum_{j=1}^{n-1} (\gamma + 2(n-j))(L_{n-j-1}^\gamma(l_i^{(n)}))^2 (L_{n-j}^\gamma(l_i^{(n)}))^2 , \quad \text{and} \\
C_n &= \sum_{j=1}^{n-1} \sqrt{\gamma+n-j}\sqrt{n-j}\, \left((L_{n-j-1}^\gamma(l_i^{(n)}))^3 L_{n-j}^\gamma(l_i^{(n)}) \;+\; L_{n-j-1}^\gamma(l_i^{(n)}) (L_{n-j}^\gamma(l_i^{(n)}))^3 \right) .
\end{aligned}$$

□

7.2 Asymptotics for β large; level densities

The following corollaries follow immediately from Theorems 7.1.3 and 7.1.5.

Corollary 7.2.1. *Let n be fixed, and $f_{n,\beta}$ be the level density of the scaled (by $1/\sqrt{2n\beta}$) $n \times n$ β-Hermite ensemble. Let $g_{n,\beta}$ be as below:*

$$g_{n,\beta}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \ ,$$

where $\mu_i = \frac{h_i^{(n)}}{\sqrt{2n}}$ and $\sigma_i = \frac{1}{\sqrt{2n\beta}}\sqrt{Var(G_i)}$, with h_i and $Var(G_i)$ as in Section 7.1.1. Then for any x,

$$\lim_{\beta \to \infty} \sqrt{\beta} \ (f_{n,\beta}(x) - g_{n,\beta}(x)) \ = \ 0 \ .$$

Corollary 7.2.2. *Let n and $\gamma > 0$ be fixed, and $f_{n,\beta,\gamma}$ be the level density of the scaled (by $1/(n\beta)$) $n \times n$ β-Laguerre ensemble of parameter $a = \frac{\beta}{2}(n - 1 + \gamma)$. Let $g_{n,\beta,\gamma}$ be as below:*

$$g_{n,\beta,\gamma}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \ ,$$

where $\mu_i = \frac{l_i^{(n)}}{n}$ and $\sigma_i = \frac{1}{n\sqrt{\beta}}\sqrt{Var(G_i)}$, with l_i and $Var(G_i)$ as in Section 7.1.2. Then for any x,

$$\lim_{\beta \to \infty} \sqrt{\beta} \ (f_{n,\beta,\gamma}(x) - g_{n,\beta,\gamma}(x)) \ = \ 0 \ .$$

While these approximations are simple enough (a sum of Gaussians is an easily recognizable shape that is also easy to work with), one may wonder how big β has to be in order for these approximations to become "accurate" (for example, in order to *appear* accurate in a plot, the approximations have to be accurate to about 2-3 digits). We have found that, in either of the two cases, the answer is surprisingly low, as the following set of plots demonstrate.

The first two sets of plots (Figures 7-2 and 7-3) are for the Hermite case. In Figure 7-2, we start with $n = 5$, and gradually increase β (from 2 to 10) to show how the plots become more and more similar. In the last one, for $\beta = 10$, the two plots appear to overlap.

If we do the same thing for $n = 7$ (as in Figure 7-3), $\beta = 6$ already provides a very good approximation.

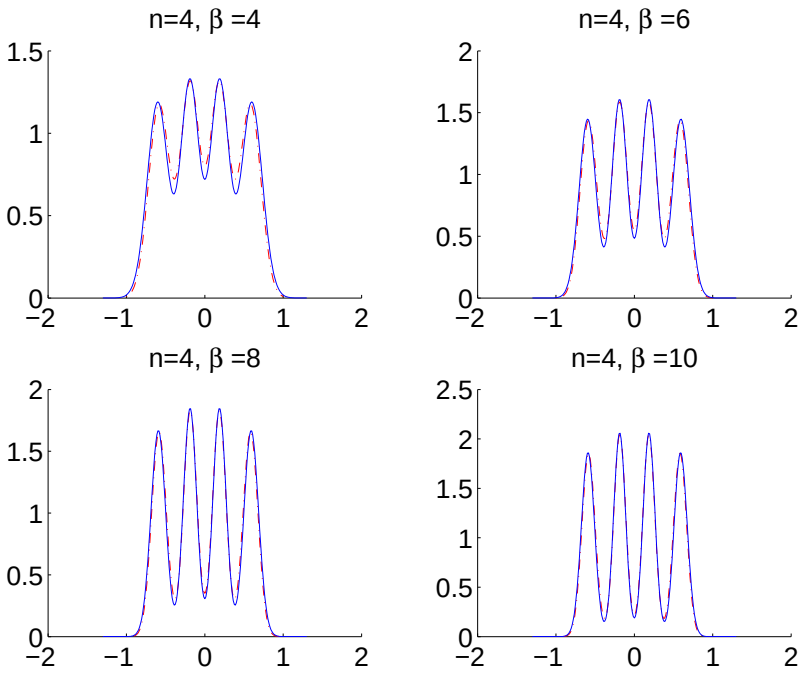

Figure 7-2: Hermite case: approximate level densities (in dash-dot lines) and exact level densities (in solid lines) for $n = 4$, and $\beta = 4, 6, 8, 10$.

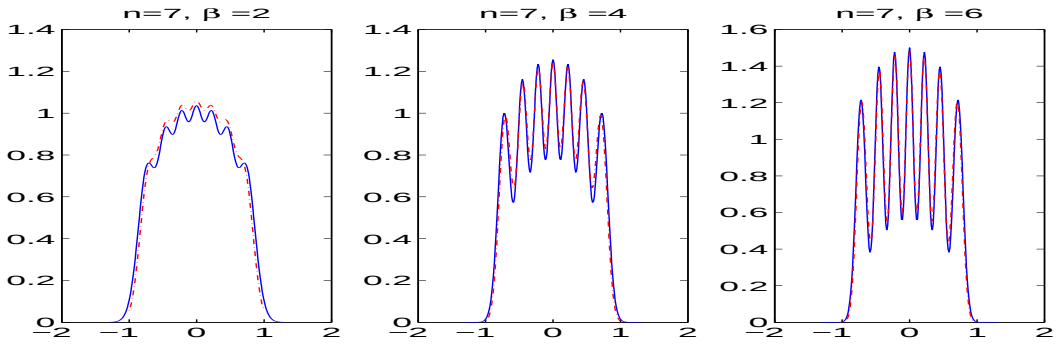

Figure 7-3: Hermite case: approximate level densities (in dash-dot lines) and exact level densities (in solid lines) for $n = 7$, and $\beta = 2, 4, 6$.

We can conclude that the approximation works very well for low values of β, in the Hermite case.

In the Laguerre case, there are two possible perspectives, which produce two possible types of tests. The first one stems from the theory: choose $\gamma \geq 0$ and $n \geq 0$,

and then see how big β and $a = \frac{\beta}{2}(n-1+\gamma)$ must be before the level density and the zero- and first-order approximation overlap.

Note that both the Laguerre parameter a and the power $p = a - \frac{(m-1)}{2}\beta - 1 = \gamma\frac{\beta}{2} - 1$ are increasing functions of β.

Moreover, by prescribing γ, as $\beta \to \infty$, in the limit, the plot should become a sum of delta functions at the roots of the Laguerre polynomial $L_n^{\gamma-1}$.

In Figure 7-4; we take $n = 4$, $\gamma = 1$, $\beta = 4, 6, 8, 10$, and $a = 8, 12, 16, 20$ (equivalently, $p = 1, 2, 3, 4$). Note that the approximation is very good for $\beta = 10$.

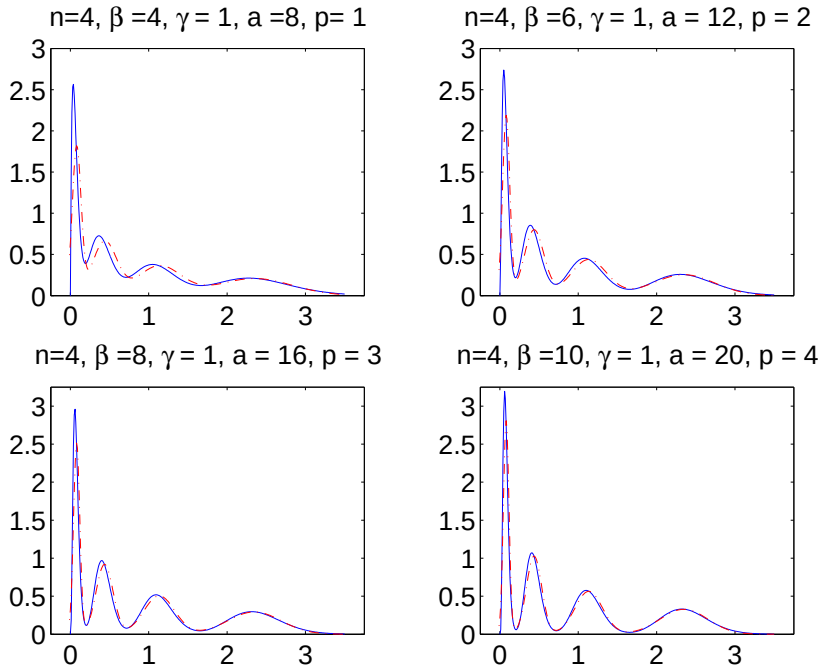

Figure 7-4: Laguerre case 1: approximate level densities (in dash-dot lines) and exact level densities (in solid lines) for $n = 4$, $\beta = 4, 6, 8, 10$, and $a = 8, 12, 16, 20$ (equivalently, $p = 1, 2, 3, 4$).

A second test arises from a practical question: suppose that one has a given Laguerre ensemble of fixed n, (relatively large) β, and given power p, and one would like to know how good the level density approximation is for that particular case. Note that to compute the approximation in this case we need to take $\gamma = (p+1)\frac{2}{\beta}$ and since β is relatively large and p "fixed", this yields a relatively small γ.

Moreover, in this case, as $\beta \to \infty$, $\gamma \to 0$, and the plot should become a sum of delta functions at the roots of the polynomial L_n^{-1}.

The approximation works, once again, surprisingly well, as demonstrated by Figure 7-5, where $n = 4$, $p = 1$, $\beta = 4, 6, 8, 10$, and $\gamma = 1, 2/3, 1/2, 2/5$ (or $a = 8, 11, 14, 17$).

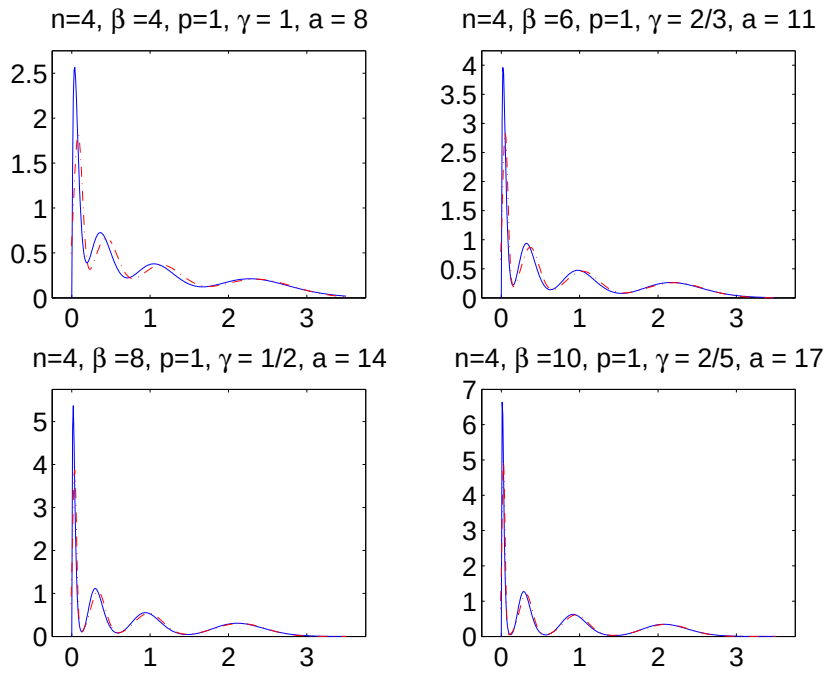

Figure 7-5: Laguerre case 2: approximate level densities (in dash-dot lines) and exact level densities (in solid lines) for $n = 4$, $p = 1$, $\beta = 4, 6, 8, 10$, and $\gamma = 1, 2/3, 1/2, 2/5$ (or $a = 8, 11, 14, 17$).

Thus we can conclude that in both cases, a good approximation is obtained even for β relatively small.y.

We have used only even integer values of β for our plots, because (in addition to $\beta = 1$) those are the only ones for which (to the best of our knowledge) there are exact formulas for the level densities. The plots were obtained with the help of our Maple Library, MOPs, which was used for computing the orthogonal and Jack polynomial quantities involved; these were translated into polynomials which were then plotted in MATLAB. For more on MOPs see Chapters 8 and 9.

Chapter 8

Jack Polynomials and Multivariate Orthogonal Polynomials

8.1 History and the connection with Random Matrix Theory

The Jack polynomials have a very rich history. They represent a family of orthogonal polynomials dependent on a positive parameter α, and some of them are more famous than others. Unsurprisingly, there are three values of α which have been studied independently, namely, $\alpha = 2, 1, 1/2$. The 1-Jack polynomials are better known as the Schur functions; the 2-Jack polynomials are better known as the zonal polynomials, whereas the 1/2-Jack polynomials are known as the quaternion zonal polynomials.

In an attempt to evaluate the integral (8.1) in connection with the non-central Wishart distribution, James [46] discovered the zonal polynomials in 1960.

$$\int_{O(n)} (\mathrm{tr}(AHBH'))^k \ \ (H'dH) = \sum_{\kappa \vdash k} c_\kappa Z_\kappa(A) Z_\kappa(B) \ . \tag{8.1}$$

Inspired by the work of James [46] and Hua [42], in his own attempt to evaluate (8.1), Jack was lead to define the polynomials eventually associated with his name [44]. More explicitly, he defined a new one-parameter (α) class of polynomials, which for $\alpha = 1$ he proved were the Schur functions, and for $\alpha = 2$ he conjectured to be the

zonal polynomials (and proved it in a very special case). He consequently generalized the α parameter to any real non-zero number, and noted that for $\alpha = -1$ he obtained yet another special class of functions, which he called the "augmented" monomial symmetric functions. Later it was noted that the orthogonalizing inner product was positive definite only if $\alpha > 0$.

During the next decade, the study of Jack polynomials intensified; Macdonald [63, page 387] points out that in 1974, H.O.Foulkes [31] raised the question of finding combinatorial interpretations for the Jack polynomials. This question was satisfactorily answered in 1997 by Knop and Sahi [57].

In the late '80s, the Jack polynomials were the subject of investigation in Macdonald's book [63] and Stanley's paper [86]; these two authors generalized many of the known properties of the Schur functions and zonal polynomials to Jack polynomials.

Below we mention a few of the researchers who have studied the connection between Jack polynomials and Random Matrix Theory.

James [47] was one of the first to make the connection between the zonal polynomials ($\alpha = 2$ Jack polynomials) and the 1-ensembles, when he calculated statistical averages of zonal polynomials over the 1-Laguerre ensemble (Wishart central and non-central distributions).

At about the same time, Constantine and Muirhead provided a generalization of the hypergeometric series, using the zonal polynomials, and studied the multivariate Laguerre polynomials for $\beta = 1$ (for a reference, see [70]).

In a survey paper, James defines and describes multivariate 1-Laguerre, Hermite and Jacobi polynomials [49]. Chikuse [13] studied more extensively the multivariate 1-Hermite polynomials.

In the early '90s, Kaneko [56] studied the general α binomial coefficients, and used them in connection with the study of hypergeometric series and multivariate Jacobi polynomials. He was also one of the first to study Selberg-type integrals, and establish the connection with generalized Jacobi polynomials. A few years later, Okounkov and Olshanski [75] considered shifted Jack polynomials for all α, and proved that they

were the same as the generalized binomial coefficients.

Kadell [55] was perhaps the first to consider averages of many valued Jack polynomials, with his study of the average of the Jack polynomial of parameter $1/k$ (with k an integer) over the corresponding $2k$-Jacobi distribution. Later it was noticed that this constraint is unnecessary.

Lasalle [59, 60, 61], considered all three types of multivariate polynomials, and among many other things computed generating functions for them.

These are just a few of the researchers who have made significant contributions in this area; we have mentioned a few more in the introduction. The last results that we mention here are those of Forrester and Baker [8], who studied in detail the generalized α-Hermite and α-Laguerre polynomials, in connection with the $2/\alpha$-Hermite and Laguerre ensembles (some of their work built on Lasalle [59, 61]). For a good reference on multivariate generalizations of many of the univariate properties of the Hermite and Laguerre ensembles, see [30].

In this chapter, we define the Jack, Hermite, Laguerre, and Jacobi multivariate orthogonal polynomials, and add a few facts to the growing body of knowledge. In Chapter 9, we show how these polynomials can be computed.

8.2 Partitions and Symmetric Functions

Definition 8.2.1. *A partition λ is a finite, ordered, non-increasing sequence of positive integers $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_l$.*

Throughout this chapter, we will refer to $l = l(\lambda)$ as the length of λ, and to $k = |\lambda| = \sum_{i=1}^{l} \lambda_i$ as the sum of λ.

Remark 8.2.2. *One can remove the constraint "finite" from the definition of the partition, and replace it with "of finite sum", since one can always "pad" a partition with 0s at the end; in this context l becomes the index of the smallest non-zero component of the partition λ.*

We will work with two orderings of the partitions. The first one is the *lexicographic* one, denoted by $\leq$.

Definition 8.2.3. *We say that $\lambda \leq \kappa$ in lexicographical ordering if for the largest integer m such that $\lambda_i = \kappa_i$ for all $i < m$, we have $\lambda_m \leq \kappa_m$. If $\lambda_m < \kappa_m$, we say that $\lambda < \kappa$.*

Remark 8.2.4. *This is a total ordering of the partitions.*

The second ordering is the *dominance* ordering, sometimes also called the *natural* ordering.

Definition 8.2.5. *We say that $\lambda \preceq \kappa$ (or, equivalently, that κ "dominates" λ) if, given $m = \max\{length(\kappa), length(\lambda)\}$,*

$$\begin{aligned} \sum_{i=1}^{j} \lambda_i &\leq \sum_{i=1}^{j} \kappa_i, \quad \forall\ j < m\ , \quad \text{and} \\ \sum_{i=1}^{m} \lambda_i &= \sum_{i=1}^{m} \kappa_i\ . \end{aligned}$$

If one of the inequalities above is strict, we say that $\lambda \prec \kappa$.

Remark 8.2.6. *Note that we compare two partitions only if they sum to the same integer. Also note that even with this constraint, $\preceq$ is only a partial ordering of the set of partitions of a given number: for example, $[4,1,1]$ and $[3,3]$ are incomparable.*

Definition 8.2.7. *A symmetric polynomial of m variables, $x_1, \ldots, x_m$, is a polynomial which is invariant under every permutation of $x_1, \ldots, x_m$.*

Remark 8.2.8. *The symmetric polynomials form a vector space over $\mathbb{R}$.*

Over the course of time, combinatorialists have defined a variety of *homogeneous* bases for this vector space; each such basis is indexed by partitions (which correspond to the terms of highest order in lexicographical ordering of the polynomial). By homogeneity we mean that all terms of the polynomial have the same total degree.

Some of these homogeneous bases are displayed in the table below:

Name	Definition for $l = 1$	Definition for $l > 1$
power-sum functions	$p_{\lambda_1} = \sum_{j=1}^{m} x_j^{\lambda_1}$	$p_\lambda = \prod_{i=1}^{l} p_{\lambda_i}$
elementary functions	$e_{\lambda_1} = \sum_{j_1<j_2<\ldots<j_{\lambda_1}} x_{j_1} \ldots x_{j_{\lambda_1}}$	$e_\lambda = \prod_{i=1}^{l} e_{\lambda_i}$
Complete homogeneous functions	$h_{\lambda_1} = \sum_{j_1\leq j_2\leq\ldots\leq j_{\lambda_1}} x_{j_1} \ldots x_{j_{\lambda_1}}$	$h_\lambda = \prod_{i=1}^{l} h_{\lambda_i}$

Another important basis is given by the monomial functions m,

$$m_\lambda = \sum_{\sigma \in S_\lambda} x_{\sigma(1)}^{\lambda_1} x_{\sigma(2)}^{\lambda_2} \ldots x_{\sigma(m)}^{\lambda_m} \ ;$$

here S_λ is the set of permutations giving distinct terms in the sum; λ is considered as infinite.

The last basis we mentioned distinguishes itself from the other ones in two ways; the advantage is that it is very easy to visualize, and proving that it is indeed a basis is immediate. The disadvantage is that it is not multiplicative[1].

Monomials seem to be the basis of choice for most people working in statistics or engineering. Combinatorialists often prefer to express series in the power-sum basis, because of connections with character theory.

8.3 Multivariate Orthogonal Polynomials

8.3.1 Jack Polynomials

The Jack polynomials C_λ^α constitute a far more complex class of homogeneous bases (depending on the parameter α) than any of the previously mentioned ones. They allow for several equivalent definitions (up to certain normalization constraints). We present here two (Definitions 8.3.1 and 8.3.2). Definition 8.3.1 arose in combinatorics, whereas Definition 8.3.2 arose in statistics. We will mainly work with Definition 8.3.2.

Definition 8.3.1. *(following Macdonald [63]) The Jack polynomials P_λ^α are orthogonal with respect to the inner product defined below on power-sum functions*

$$\langle p_\lambda, p_\mu \rangle_\alpha = \alpha^{l(\lambda)} z_\lambda \delta_{\lambda\mu},$$

[1]While $x_\lambda x_\mu = x_{\lambda+\mu}$, for $x \in \{p, e, h\}$, $m_\lambda m_\mu$ in general is not a monomial

where $z_\lambda = \prod_{i=1}^{l(\lambda)} a_i! i^{a_i}$, a_i *being the number of occurrences of* i *in* λ. *In addition,*

$$P_\lambda^\alpha = m_\lambda + \sum_{\mu \preceq \lambda} u_{\lambda,\mu}^\alpha m_\mu \ .$$

There are two main normalizations of the Jack polynomials used in combinatorics, the "J" normalization (which makes the coefficient of the lowest-order monomial, $[1^n]$, be exactly $n!$) and the "P" normalization (which makes the coefficient of the highest-order monomial be 1, and is given in Definition 8.3.1). To convert between these normalizations, see Tables 8.1 and 8.2. In Table 8.1, $I_m = (1, 1, 1, \ldots, 1)$, where the number of variables is m.

We use the notation $\kappa \vdash k$ for κ a partition of k, and ρ_κ^α for $\sum_{i=1}^m k_i(k_i - 1 - \frac{2}{\alpha}(i-1))$.

Definition 8.3.2. *(following Muirhead, [70]) The Jack polynomial* C_κ^α *is the only homogeneous polynomial eigenfunction of the following Laplace-Beltrami-type operator*

$$D^* = \sum_{i=1}^m x_i^2 \frac{d^2}{dx_i^2} + \frac{2}{\alpha} \sum_{1 \le i \ne j \le m} \frac{x_i^2}{x_i - x_j} \frac{d}{dx_i} \ ,$$

with eigenvalue $\rho_\kappa^\alpha + k(m-1)$, *having highest-order term corresponding to* κ. *In addition,*

$$\sum_{\kappa \vdash k, \ \ l(\kappa) \le m} C_\kappa^\alpha(x_1, x_2, \ldots, x_m) = (x_1 + x_2 + \ldots x_m)^k \ .$$

Remark 8.3.3. *The Jack "C" polynomials are the normalization which allows for defining scalar hypergeometric functions of multivariate (or matrix) argument. These supply Selberg-type integrals which appear in various fields, from the theory of random walks to multivariate statistics and quantum many-body problems.*

Remark 8.3.4. *David M. Jackson [45] pointed out that the* D^* *operator also appears in algebraic geometry, for example in the context of ramified covers.*

Definition 8.3.5. *Given the diagram of a partition* κ *(see Figure 1), define* $a_\kappa(s)$ *(the "arm-length") as the number of squares at the right of* s; $l_\kappa(s)$ *(the "leg-length") as the number of squares below* s; $h_\kappa^*(s) = l_\kappa(s) + \alpha(1 + a_\kappa(s))$ *(the "upper hook length") and* $h_*^\kappa(s) = l_\kappa(s) + 1 + \alpha a_\kappa(s)$ *(the "lower hook length").*

Figure 8-1: The Arm-length and the Leg-length.

Finally, a further definition is needed in order to present the conversion table.

Definition 8.3.6. *Let*

$$\begin{aligned} c(\kappa,\alpha) &= \prod_{s\in\kappa} h^*_\kappa(s)\ , \\ c'(\alpha,\kappa) &= \prod_{s\in\kappa} h^\kappa_*(s)\ , \\ j_\kappa &= c(\alpha,\kappa)\ c'(\alpha,\kappa)\ , \end{aligned}$$

where h^*_κ and h^κ_* have been defined above.

To explain the conversions between "J", "P", and "C", we recall the definition of the generalized Gamma function and generalized shifted factorial from Chapter 3, (3.1) and (3.2); instead of β, we will work now with $\alpha = 2/\beta$, as in the rest of the chapter.

We can now present Tables 8.1 and 8.2; the entries have been filled out using James [48], Forrester and Baker [8], and Stanley [86].

Normalization	Value at $I_m = (1,1,1,\ldots,1)$
C	$C^\alpha_\kappa(I_m) = \frac{\alpha^{2k}k!}{j_\kappa}\left(\frac{m}{\alpha}\right)_\kappa$
J	$J^\alpha_\kappa(I_m) = \alpha^k\left(\frac{m}{\alpha}\right)_\kappa$
P	$P^\alpha_\kappa(I_m) = \frac{\alpha^k}{c(\alpha,\kappa)}\left(\frac{m}{\alpha}\right)_\kappa$

Table 8.1: Values of the different normalizations of Jack polynomials of partition κ and parameter α at I_m.

	C	J	P
C		$\frac{\alpha^k\ k!}{j_k}$	$\frac{\alpha^k\ k!}{c'(\kappa,\alpha)}$
J	$\frac{j_k}{\alpha^k\ k!}$		$c(\kappa,\alpha)$
P	$\frac{c'(\kappa,\alpha)}{\alpha^k\ k!}$	$\frac{1}{c(\kappa,\alpha)}$	

Table 8.2: Conversions between the three normalizations for the Jack polynomials; the $a(I,J)$ entry above is defined as $I^\alpha_\kappa(x_1,\ldots,x_m) = a(I,J)J^\alpha_\kappa(x_1,\ldots,x_m)$.

8.3.2 Generalized binomial coefficients

Many algebraic quantities (and the identities they satisfy) can be extended from the univariate case to the multivariate case through Jack polynomials. One such example is the generalized binomial coefficient.

Definition 8.3.7. *We define the generalized binomial coefficients* $\binom{\kappa}{\sigma}$ *as*

$$\frac{C^\alpha_\kappa(x_1+1,x_2+1,\ldots,x_n+1)}{C^\alpha_\kappa(1,1,\ldots,1)} = \sum_{s=0}^{k}\sum_{\sigma\vdash s,\ \sigma\subseteq\kappa}\binom{\kappa}{\sigma}\frac{C^\alpha_\sigma(x_1,x_2,\ldots,x_n)}{C^\alpha_\sigma(1,1,\ldots,1)}\ ,$$

where $\sigma\subset\kappa$ *means that* $\sigma_i\le\kappa_i$ *for all* i.

The generalized binomial coefficients depend on α, but are independent of both the number of variables m and the normalization of the Jacks (the latter independence is easily seen from the definition).

The multivariate binomial coefficients generalize the univariate ones; some simple properties of the former are straightforward generalizations of properties of the latter. For example,

$$\begin{aligned}
\tbinom{\kappa}{(0)} &= 1\ ,\\
\tbinom{\kappa}{(1)} &= |\kappa|\ ,\\
\tbinom{\kappa}{\sigma} &= 0 \text{ if } \sigma\not\subseteq\kappa\ ,\\
\tbinom{\kappa}{\sigma} &= \delta_\kappa \text{ if } |\kappa|=|\sigma|\ ,\\
\tbinom{\kappa}{\sigma} &\neq 0 \text{ if } |\kappa|=|\sigma|+1,\ \text{iff } \sigma=\kappa_{(i)}\ ,
\end{aligned}$$

where $\kappa_{(i)}=(k_1,\ldots,k_i-1,\ldots,k_m)$. The above are true for all κ and α, and σ subject to the constraints.

8.3.3 Jacobi Polynomials

The Jacobi polynomials can be obtained by the Gram-Schmidt orthogonalization of the Jack polynomials[2] C_λ^α with respect to the weight/density function proportional to the $2/\alpha$-Jacobi density

$$\begin{aligned} d\mu_J^\alpha(x_1, x_2, \ldots, x_m) &= \frac{\Gamma_m\big(g_1 + g_2 + \frac{2}{\alpha}(m-1) + 2\big)}{\Gamma_m(g_1 + \frac{m-1}{\alpha} + 1)} \quad \times \\ &\quad \times \quad \prod_{i=1}^m \Big[x_i^{g_1}(1-x_i)^{g_2}\Big] \prod_{i<j} |x_i - x_j|^{2/\alpha} dx_1 \ldots dx_m \ ; \end{aligned}$$

for the purpose of well-definiteness we assume

$$g_1, g_2 > -1 \ . \tag{8.2}$$

Define

$$\begin{aligned} \delta^* &= \sum_i x_i \frac{d^2}{dx_i^2} + \frac{2}{\alpha} \sum_{i \neq j} \frac{x_i}{x_i - x_j} \frac{d}{dx_i} \\ E &= \sum_i x_i \frac{d}{dx_i} \\ \epsilon &= \sum_i \frac{d}{dx_i} \ ; \end{aligned}$$

then the Jacobi polynomials are eigenfunctions of the following Laplace-Beltrami-type operator:

$$D^* + (g_1 + g_2 + 2)E - \delta^* - (g_1 + 1)\epsilon \ ,$$

with eigenvalue $\rho_\kappa^\alpha + |\kappa|(g_1 + g_2 + \frac{2}{\alpha}(m-1) + 2)$.

8.3.4 Laguerre Polynomials

The multivariate Laguerre polynomials are orthogonal with respect to the weight/density proportional to the Wishart (or Laguerre) density:

$$\begin{aligned} d\mu_L^\alpha(x_1, x_2, \ldots, x_m) &= \frac{\pi^{-m(m-1)/\alpha} \ (\Gamma(1 + \frac{1}{\alpha}))^m}{\Gamma_m^\alpha(1 + \frac{m}{\alpha}) \ \Gamma_m^\alpha(\gamma + \frac{m-1}{\alpha} + 1)} \quad \times \\ &\quad \times \quad e^{-\sum_i x_i} \prod_i x_i^\gamma \prod_{i \neq j} |x_i - x_j|^{2/\alpha} dx_1 \ldots dx_m. \end{aligned}$$

[2]Or any other set of homogeneous polynomials with leading degree corresponding to κ.

Note that for the purpose of well-definiteness, we must have $\gamma > -1$.

This density can be obtain from the Jacobi density of the previous subsection by substituting $(g_1 + g_2 + \frac{2}{\alpha}(m-1) + 2)^{-1}(x_1, \ldots, x_m)$ for $(x_1, \ldots, x_m)$ and then taking the limit as $g_2 \to \infty$. The same limiting process applied to the Jacobi polynomials yields the Laguerre polynomials.

Under the transformation mentioned above, the Jacobi differential operator becomes

$$\delta^* - E + (\gamma + 1)\epsilon \ ,$$

and the Laguerre polynomials are simply eigenfunctions of this operator, with eigenvalue $|\kappa|$.

8.3.5 Hermite Polynomials

The multivariate Hermite polynomials are orthogonal with respect to the Gaussian (Hermite) weight/density proportional to

$$\begin{aligned} d\mu_H^\alpha(x_1, x_2, \ldots, x_m) &= 2^{-n/2}\ \pi^{n(n-1)/\alpha - n/2}\ \frac{(\Gamma(1+\frac{1}{\alpha}))^n}{\Gamma_n^\alpha(1+\frac{n}{\alpha})} \quad \times \\ &\times \quad e^{-\sum_i x_i^2/2} \prod_{i \neq j} |x_i - x_j|^{2/\alpha} dx_1 \ldots dx_n \ . \end{aligned}$$

This density can be obtained by taking $(\gamma + \sqrt{\gamma}x_1, \gamma + \sqrt{\gamma}x_2, \ldots, \gamma + \sqrt{\gamma}x_m)$ in the density of the Laguerre polynomial, and then letting γ go to infinity; note that this way the only non-matrix parameter remaining is α.

Under this limiting process, the differential operator becomes

$$\delta^{**} - E \ ,$$

where

$$\delta^{**} = \sum_i \frac{d^2}{dx_i^2} + \frac{2}{\alpha} \sum_{i \neq j} \frac{1}{x_i - x_j} \frac{d}{dx_i} \ .$$

The Hermite polynomials are eigenfunctions of this operator with eigenvalue $|\kappa|$.

Remark 8.3.8. *Similarly,*

$$\lim_{\gamma \to \infty} \gamma^{-k/2} L_\kappa^{\alpha,\gamma}(\gamma + \sqrt{\gamma}x_1, \gamma + \sqrt{\gamma}x_2, \ldots, \gamma + \sqrt{\gamma}x_m) = (-1)^k H_\kappa^\alpha(x_1, \ldots, x_m) \ . \quad (8.3)$$

8.3.6 Hypergeometric functions

The hypergeometric functions are perhaps the easiest to generalize from univariate to multivariate. For the multivariate versions, a good reference is Forrester's unpublished book [30].

Definition 8.3.9. *We define the hypergeometric function* ${}_pF_q^\alpha$ *of parameters* $a_1, \ldots, a_p$, *respectively* $b_1, \ldots, b_q$ *and of variables* $(x_1, \ldots, x_m)$ *by*

$$ {}_pF_q^\alpha(a_1, \ldots, a_p; b_1, \ldots, b_q; x_1, \ldots, x_m) = \sum_{k=0}^{\infty} \sum_{\kappa \vdash k} \frac{(a_1)_\kappa \ldots (a_p)_\kappa}{k!\ (b_1)_\kappa \ldots (b_q)_\kappa}\ C_\kappa^\alpha(x_1, \ldots, x_m)\ . $$

Note that this is a formal definition; additional conditions on the variables a_i and b_j are needed in order for the above series to converge. Similarly one can extend this to hypergeometric functions of 2, 3, ... matrix arguments.

Hypergeometric functions provide answers to many statistics and statistics-related questions; below are two examples.

1. Krishnaiah and Chang [58] have proved in 1971 that the density of the smallest root of a real ($\alpha = 2$) Wishart matrix with m variables and n degrees of freedom such that $p = \frac{n-m-1}{2}$ is an integer is proportional to

   $$ \rho(x) = x^{pm}\ e^{-xm/2}\ {}_2F_0(-p, \frac{m+2}{2}; -2I_{m-1}/x)\ . $$

   Note that the eigenvalue density of the matrix described above is $d\mu_L^\alpha$ with $\alpha = 2$ and $a = n/2$.

   In Chapter 10 we extend this to any α and any a such that $p = a - \frac{m-1}{\alpha} - 1$ is an integer. We obtain that for this case the density of the smallest eigenvalue is proportional to

   $$ \rho(x) = x^{pm}\ e^{-xm/2}\ {}_2F_0^\alpha(-p, \frac{m}{\alpha} + 1; -2I_{m-1}/x)\ . $$

2. The largest eigenvalue (l_1) distribution for a Wishart real matrix with m variates and n degrees of freedom ($\alpha = 2$, $a = n/2$) can be expressed as

   $$ P[l_1 < x] = \frac{\Gamma_m\big[\frac{1}{2}(m+1)\big]}{\Gamma_m\big[\frac{1}{2}(n+m+1)\big]} \det\big(\frac{1}{2}xI_m\big)^{n/2}\, {}_1F_1\big(\frac{1}{2}n, \frac{1}{2}(n+m+1); -\frac{1}{2}xI_m\big)\ . $$

The above is a corollary of a stronger theorem proved by Constantine [14], and it can also be found in Muirhead [70, page 421].

This result generalizes to any α and a (as noted in Chapter 10) to

$$P[l_1 < x] = \frac{\Gamma_m\left[\frac{1}{\alpha}(m-1)+1\right]}{\Gamma_m\left[a+\frac{1}{\alpha}(m-1)+1\right]} \det\left(\frac{1}{2}xI_m\right)^{am} {}_1F_1\left(a, a+\frac{1}{\alpha}(m-1)+1; -\frac{1}{2}xI_m\right)$$

8.4 Computing Integrals over the β-ensembles

In the univariate case we have the following moments for the Hermite, Laguerre, and Jacobi weight functions:

$$\begin{aligned}
\int_{\mathbb{R}} x^k e^{-x^2/2} dx &= (2k-1)!! = (-1)^{k/2} H_k(0) \ , \\
\int_{[0,\infty)} x^k x^\gamma e^{-x} dx &= (\gamma+1)_k = L_k^\gamma(0) \ , \quad \text{and} \\
\int_{[0,1]} x^k x^a (1-x)^b dx &= \frac{(a+1)_k \Gamma(a+b+2)}{\Gamma(a+1)\Gamma(a+b+k+2)} = P_k^{a,b}(0) \ .
\end{aligned}$$

In the above, $k \geq 0$.

These three identities provide in the univariate case, formulas for integrating (averaging) any polynomial over the univariate Hermite, Laguerre, and Jacobi distributions.

A similar triad of formulas is can be established for the multivariate case. In the Laguerre and Jacobi cases, the closed-form formulas extend to the multivariate case:

$$\int_{[0,\infty)^m} C_\kappa^\alpha(x_1, \ldots, x_m) d\mu_L^\alpha = (\gamma + \frac{m-1}{\alpha} + 1)_\kappa C_\kappa^\alpha(I_m) = L_\kappa^{\alpha,\gamma}(0) \ , \tag{8.4}$$

$$\int_{[0,1]^m} C_\kappa^\alpha(x_1, \ldots, x_m) d\mu_J^\alpha = \frac{(g_1 + \frac{m-1}{\alpha} + 1)_\kappa}{(g_1 + g_2 + \frac{2}{\alpha}(m-1) + 2)_\kappa} C_\kappa^\alpha(I_m) = P_{\alpha,\kappa}^{g1,g2}(0) \tag{8.5}$$

For a good reference for the first formula, see Forrester and Baker [8]; the second one was obtained by Kadell [55].

For the Hermite case,

$$\int_{\mathbb{R}^n} C_\kappa^\alpha(x_1, \ldots, x_m) d\mu_H^\alpha = (-1)^{k/2} H_\kappa^\alpha(0) \ , \tag{8.6}$$

but to the best of our knowledge, no simpler closed-form formula is known.

Below is the most explicit formula we were able to obtain for $H_\kappa^\alpha(0)$. Here $\Big[\dots\Big]$ is the notation for a coefficient in a polynomial/power series. To the best of our knowledge, this formula is new. We use I_t for a vector of t ones.

Lemma 8.4.1.

$$[C_\sigma^\alpha(x_1,\dots,x_m)]\ H_\kappa^\alpha(x_1,\dots,x_m) \;=\; (-1)^k \frac{C_\kappa^\alpha(I_m)}{C_\sigma^\alpha(I_m)} \sum_{j=s}^{\frac{k+s}{2}} (-1)^{k-j} \sum_{\sigma\subseteq\mu\subseteq\kappa;\mu\vdash j} \binom{\kappa}{\mu}\binom{\mu}{\sigma} \times \\ \times \Big[r^{\frac{k+s}{2}-j}\Big] F(r,\alpha,m,\kappa,\sigma)\ ,$$

where

$$F(r,\alpha,m,\kappa,\sigma) \;=\; \frac{(r+\frac{1}{\alpha}(m+\alpha-1))_\kappa}{(r+\frac{1}{\alpha}(m+\alpha-1))_\sigma}\ .$$

In particular, for the empty partition $[]$*,*

$$H_\kappa^\alpha(0) \;=\; [C_{[]}^\alpha(x_1,\dots,x_m)]\ H_\kappa^\alpha(x_1,\dots,x_m) \tag{8.7}$$

$$\;=\; (-1)^k C_\kappa^\alpha(I_m) \sum_{j=0}^{\frac{k}{2}} (-1)^{k-j} \sum_{\mu\subseteq\kappa;\mu\vdash j} \binom{\kappa}{\mu} \Big[r^{\frac{k}{2}-j}\Big] F(r,\alpha,m,\kappa,[])\ . \tag{8.8}$$

Remark 8.4.2. *Note that if* κ *sums to an odd integer, or if* $|\kappa|$ *and* $|\sigma|$ *have different parities, the above is trivially* 0*.*

Proof. Starting with (8.3), one writes the Laguerre polynomial in closed-form, then expands the Jack polynomials by the generalized binomial formula, re-groups the terms, and takes the limit. □

8.5 Some integrals over β-Hermite ensembles

8.5.1 A new proof for a conjecture of Goulden and Jackson

In this section, we take a closer look at formula (8.6): the expected value of a Jack polynomial over the β-Hermite ensembles. We provide an alternate proof for a conjecture by Goulden and Jackson [36], which was proved by Okounkov [73]. The proof is based on the material presented in this chapter.

We give the conjecture as a theorem below.

Theorem 8.5.1. *Let k be an even integer. Then*

$$\int_{\mathbb{R}^n} P_\kappa^\alpha(x_1, x_2, \ldots, x_n) d\mu_H^\alpha(x_1, x_2, \ldots, x_n) \;=\; \left(\frac{n}{\alpha}\right)_\kappa \frac{\langle P_\kappa^\alpha,\; p_{[2^{k/2}]}\rangle_\alpha}{2^{k/2}\;(k/2)!}\;,$$

where $|\kappa| = k$ and $[2^{k/2}] = [2, 2, \ldots, 2]$. Note that if $k = 0$ or k is odd, the above becomes $0 = 0$.

Proof. For reasons of space, we denote by $X = (x_1, x_2, \ldots, x_n)$, by $T = (t_1, t_2, \ldots, t_n)$, and by $I_n = (1, 1, \ldots, 1)$.

First we rewrite the equation with the help of (8.6) to

$$H_\kappa^\alpha(0) = \frac{(-1)^{k/2}\alpha^k\; k!}{c'(\alpha,\kappa)}\;\left(\frac{n}{\alpha}\right)_\kappa \frac{\langle p_\kappa^\alpha,\; p_{[2^{k/2}]}\rangle_\alpha}{2^{k/2}\;(k/2)!}. \tag{8.9}$$

We use the generating function for the Hermite polynomials. From [8], Proposition 3.1 (which the authors attribute to Lasalle), rewritten in accordance with our normalization of the Hermite polynomials, we have that

$$\sum_{k=0}^{\infty}\sum_{\kappa\vdash k} \frac{H_\kappa^\alpha(X)\; C_\kappa^\alpha(T)}{k!\; C_\kappa^\alpha(I_n)} \;=\; {}_0F_0^\alpha(X, T)\;\mathrm{etr}(-T^2/2)\;.$$

Because the left-hand side has the normalization constant for the Jack polynomial built-in, we can rewrite it as

$$\sum_{k=0}^{\infty}\sum_{\kappa\vdash k} \frac{H_\kappa^\alpha(X)\; P_\kappa^\alpha(T)}{k!\; P_\kappa^\alpha(I_n)} \;=\; {}_0F_0^\alpha(X, T)\;\mathrm{etr}(-T^2/2)\;.$$

Let $X = 0$ in the above; by the homogeneity of $P_\kappa^\alpha(T)$ it follows that, for every $k \geq 0$,

$$\sum_{\kappa\vdash k} \frac{H_\kappa^\alpha(0)\; P_\kappa^\alpha(T)}{k!\; P_\kappa^\alpha(I_n)} \;=\; \frac{1}{(k/2)!}\mathrm{tr}(-T^2/2)^{k/2} \;=\; \frac{(-1)^{k/2}}{2^{k/2}\;(k/2)!} p_{[2^{k/2}]}(T)\;.$$

Thus

$$H_\kappa^\alpha(0) = \frac{k!\; P_\kappa^\alpha(I_n)}{2^{k/2}\;(k/2)!}\,[P_\kappa^\alpha(T)]\;\; p_{[2^{k/2}]}(T)\;,$$

which can be rewritten as

$$H_\kappa^\alpha(0) = \frac{k!\; P_\kappa^\alpha(I_n)}{2^{k/2}\;(k/2)!}\frac{\langle p_{[2^{k/2}]}, P_\kappa^\alpha\rangle_\alpha}{\langle P_\kappa^\alpha, P_\kappa^\alpha\rangle_\alpha}\;, \tag{8.10}$$

and since $\langle P_\kappa^\alpha, P_\kappa^\alpha\rangle_\alpha = \frac{c'(\alpha,\kappa)}{c(\alpha,\kappa)}$, and $P_\kappa^\alpha(I_n) = \frac{\alpha^k}{c(\alpha,\kappa)}\left(\frac{n}{\alpha}\right)_\kappa$, the statement of the theorem follows. □

Remark 8.5.2. *The Goulden-Jackson conjecture was meant to provide a way of connecting two different perspectives on a map counting problem. Okounkov was interested in finding a way to compute the expected value of a Jack polynomial over a β-Hermite ensemble. Formula (8.8) also provides such a means, and it has been implemented in our Maple package, MOPs.*

8.5.2 A duality principle

This section explores formula (8.6) to a greater extent, and obtains an interesting duality formula by using the $\mathbb{Q}(\alpha)$-algebra family of automorphisms defined in [63] and [86].

Theorem 8.5.3.

$$E_\alpha\left[\frac{C_\kappa^\alpha(x_1,x_2,\ldots,x_n)}{C_\kappa^\alpha(1,1,\ldots,1)}\right] = (-\alpha)^{-k/2}\ E_{1/\alpha}\left[\frac{C_{\kappa'}^{1/\alpha}(y_1,y_2,\ldots,y_m)}{C_{\kappa'}^{1/\alpha}(1,1,\ldots,1)}\right]\ ,$$

where the first expectation is taken over the $2/\alpha$-Hermite ensemble of size n, while the second is taken over the 2α-Hermite ensemble of size m (n not necessarily equal to m), and κ' is the conjugate of κ.

Proof. Let $\mathbb{Q}(\alpha)$ be the field of all rational functions of α with rational coefficients.

Let $\Lambda \times \mathbb{Q}(\alpha)$ be the vector space of all symmetric polynomials of bounded degree with coefficients in $\mathbb{Q}(\alpha)$.

For every $0 \neq \beta \in \mathbb{Q}(\alpha)$, define a $\mathbb{Q}(\alpha)$-algebra automorphism $\omega_\beta : \Lambda \times \mathbb{Q}(\alpha) \to \Lambda \times \mathbb{Q}(\alpha)$ by the condition $\omega_\theta(p_k) = (-1)^{k-1}\theta p_k$, for all $k \geq 1$. This family of automorphisms is defined in [63, Chapter 10], and similarly in [86]. In particular, $\omega = \omega_1$ is the Macdonald involution [63, Chapter 1].

We will use the following three formulas, to be found as formulas (10.8), (10.9) and (10.24) in [63]; the first two follow easily from the definition of ω_θ. The third one is due to Stanley [86].

$$\langle \omega_\theta f,\ g\rangle_\alpha = \langle f,\ \omega_\theta g\rangle_\alpha\ , \quad \text{(self-adjointness)} \tag{8.11}$$

$$\langle \omega_{1/\alpha} f,\ g\rangle_\alpha = \langle \omega f,\ g\rangle_1\ , \quad \text{(duality with respect to } \langle , \rangle_\alpha) \tag{8.12}$$

$$\omega_\alpha J_\kappa^\alpha = \alpha^{|\kappa|}\ J_{\kappa'}^{1/\alpha}\ . \tag{8.13}$$

We rewrite 8.9 as follows:

$$\frac{H^\alpha_\kappa(0)}{C^\alpha_\kappa(I_n)} = \frac{(-1)^{k/2}}{2^{k/2}\ \alpha^k\ (k/2)!}\langle p_{[2^{k/2}]},\ J^\alpha_\kappa\rangle_\alpha\ , \tag{8.14}$$

where the last equality can be obtained by re-normalizing (see Table 6), and from Proposition (3.6) from [86]. Once again, I_n denotes a vector of n ones.

We now rewrite $\langle p_{[2^{k/2}]},\ J^\alpha_\kappa\rangle_\alpha$ as follows, making use of (8.11)-(8.13):

$$\begin{aligned}
\langle p_{[2^{k/2}]},\ J^\alpha_\kappa\rangle_\alpha &= \alpha^k\ \langle p_{[2^{k/2}]},\ \omega_{1/\alpha}J^{1/\alpha}_{\kappa'}\rangle_\alpha \\
&= \alpha^k\ \langle p_{[2^{k/2}]},\ \omega J^{1/\alpha}_{\kappa'}\rangle_1 \\
&= \alpha^k\ \langle \omega p_{[2^{k/2}]},\ J^{1/\alpha}_{\kappa'}\rangle_1 \\
&= \alpha^k\ \langle \omega_\alpha p_{[2^{k/2}]},\ J^{1/\alpha}_{\kappa'}\rangle_{1/\alpha} \\
&= (-1)^{k/2}\alpha^{3k/2}\langle p_{[2^{k/2}]},\ J^{1/\alpha}_{\kappa'}\rangle_{1/\alpha}\ .
\end{aligned}$$

Hence

$$\begin{aligned}
\frac{H^\alpha_\kappa(0)}{C^\alpha_\kappa(I_n)} &= \frac{(-\alpha)^{k/2}}{2^{k/2}\ (k/2)!}\ \langle p_{[2^{k/2}]},\ J^{1/\alpha}_{\kappa'}\rangle_{1/\alpha}\ , \\
&= (-\alpha)^{-k/2}\ \frac{H^{1/\alpha}_{\kappa'}(0)}{C^{1/\alpha}_{\kappa'}(I_m)}\ ,
\end{aligned}$$

and by (8.6), the statement of the theorem follows. □

8.5.3 Moments of determinants

In this section we examine the problem of computing moments of the determinant of the $2/\alpha$-Hermite distribution. The reason why the Hermite case is interesting is because it is the only non-trivial one (in the Laguerre (or even Jacobi) cases, the power of the determinant can be absorbed into the Laguerre (or Jacobi) power, and the answer follows from Selberg integral formulas.

This problem was our entry point into the theory of random matrices, and attempting to solve it lead us to many of the computations and theorems in this thesis. Our interest has since shifted to other areas, and the original question of a nice general formula for the moments of the determinant of the $2/\alpha$-Hermite distribution is

not fully solved. We present here some partial results, and mention the fact that a routine for computing moments of the determinant is available as part of our Maple Library, MOPs.

As with most of the problems approached in this thesis, the cases $\beta = 1, 2, 4$ (or, in the notation of this chapter, $\alpha = 2, 1, 1/2$) are the only ones which were studied. The first to examine the problem and provide a complete answer for $\beta = 2$ ($\alpha = 1$) were Mehta and Normand [68]. Mehta also discovered a duality between the cases $\beta = 1$ ($\alpha = 2$) and $\beta = 4$ ($\alpha = 1/2$).

This duality is simply a particular case of Theorem 8.5.3. Indeed, for a $n \times n$ matrix A with eigenvalues $x_1, \ldots, x_n$,

$$\det(A)^k = \frac{C^{\alpha}_{[k^n]}(x_1, x_2, \ldots, x_n)}{C^{\alpha}_{[k^n]}(I_n)} ,$$

for every α. Moreover, the set of partitions $B = \{[k^n] \mid k \in \mathbb{N}, n \in \mathbb{N}\}$ is closed with respect to conjugation.

Remark 8.5.4. *This means that instead of computing the kth moment for all α and n (size of the matrix) we can transform the problem to computing all moments n for any α for a matrix of fixed size k.*

A few years after [68], Delannay and Le Caër [16] computed the closed-form answer for $\beta = 1$ ($\alpha = 2$), and noted that the duality also provides the answer for $\beta = 4$ ($\alpha = 1/2$).

Jackson, Goulden, and Andrews [3] rediscovered these results for $\alpha = 1$ and 2, and linked them to a particular coefficient of the Jack polynomial corresponding to rectangular partitions. Their proof greatly simplifies the Delannay and Le Caër proof, and is entirely self-contained.

To the best of our knowledge, there are no references in the literature for other values of α. In this chapter, we give a closed-form answer for the second moment (the square of the determinant), for any α, and we provide 3-term and 4-term recurrences for the third and fourth moments.

Theorem 8.5.5. *Let n and α be fixed; then*

$$E_\alpha[\det(A_n)^2] \;=\; (2n-1)!!\;(-\alpha)^{-n}\;\;{}_2F_1(-n,-n-\alpha;-n+\frac{1}{2};\frac{1}{2})\;,$$

where the matrix A_n is from the $2/\alpha$-Hermite ensemble of size n, and the expectation is taken over the ensemble.

Proof. The proof is based on the duality. Based on Theorem 8.5.3, it is enough to show that given a matrix A_2 from the 2×2 $2/\alpha$-Hermite ensemble,

$$E_\alpha[\det(A_2)^n] = (2n-1)!!\;\;{}_2F_1(-n,-n-\frac{1}{\alpha};-n+\frac{1}{2};\frac{1}{2})\;.$$

To this extent, we start with the actual distribution of the eigenvalues (x_1,x_2), which we order $(x_1\geq x_2)$. The distribution is

$$f_{\alpha,2} = \frac{1}{\pi}\frac{\Gamma\left(1+\frac{1}{\alpha}\right)}{\Gamma\left(1+\frac{2}{\alpha}\right)}\;(x_1-x_2)^{2/\alpha}\;e^{-x_1^2/2-x_2^2/2}\;.$$

We will compute the exponential generating function for the moments of the determinant, which is (formally) $E_\alpha[e^{t\det(A_2)}]$.

We denote by $T=x_1+x_2$, and by $D=(x_1-x_2)^2$, the trace and the discriminant of the matrix; we change variables (the Jacobian of the transformation is $1/(4\sqrt{D})$), and obtain

$$E_\alpha[e^{t\det(A_2)}] = \frac{1}{\pi}\frac{\Gamma\left(1+\frac{1}{\alpha}\right)}{\Gamma\left(1+\frac{2}{\alpha}\right)}\;\int_{\mathbb{R}} e^{-T^2(1-t)/4}dT\;\int_{[0,\infty)} D^{1/\alpha-1/2}\;e^{-(1+t)D/4}\;dD\;. \quad (8.15)$$

With the help of the Legendre duplication formula

$$\Gamma(2z) = \frac{1}{2\pi}2^{2z-1/2}\Gamma(z+\frac{1}{2})\Gamma(z)\;,$$

(8.15) becomes

$$E_\alpha[e^{t\det(A_2)}] = (1-t)^{-1/2}\;(1+t)^{-1/\alpha-1/2}\;.$$

In other words, we obtain that

$$E_\alpha[\det(A_2)^n] = \frac{\mathrm{d}^n}{\mathrm{d}t^n}(1-t)^{-1/2}\;(1+t)^{-1/\alpha-1/2}\;\Big|_{t=0}\;.$$

Though the Jacobi polynomials are usually defined only for parameters $a,b>-1$ (as being orthogonal with respect to the weight function $(1-t)^a(1+t)^b$ on $[-1,1]$),

formally, the Rodrigues formula definition (8.16, see Abramowitz and Stegun [1]) below extends to and a, b, away from 1 and -1:

$$P_n^{a,b}(t) \;=\; \frac{(-1)^n}{2^n\ n!}\ (1-x)^{-a}(1+x)^{-b}\ \frac{\mathrm{d}^n}{\mathrm{d}t^n}(1-t)^{a+n}\ (1+t)^{b+n}\ . \tag{8.16}$$

Thus

$$E_\alpha[\det(A_2)^n] = (-1)^n\ 2^n\ n!\ P_n^{a,b}(0)\ . \tag{8.17}$$

Similarly, away from $t = \pm 1$, the following identity extends for any a, b (see Abramowitz and Stegun [1]):

$$P_n^{a,b}(t) \;=\; \frac{(a+1)_n}{n!}\ \ {}_2F_1(-n, n+a+b+1; a+1; \frac{1}{2}(1-t))\ .$$

By letting $a = -n-1/2$ and $b = -n-1/\alpha-1/2$ in the above, we can thus rewrite (8.17) as

$$E_\alpha[\det(A_2)^n] \;=\; (2n-1)!!\ \ {}_2F_1(-n, -n-\frac{1}{\alpha}; -n+\frac{1}{2}; \frac{1}{2})\ .$$

□

For the third and fourth moments of the determinant, we will have to apply a different strategy, based on the 3-term recurrence for the determinant.

Let A_n be a matrix from the size n $2/\alpha$-Hermite ensemble, and let A_{n-k} be the lower right $(n-k)\times(n-k)$ submatrix, for all $k \le n$; clearly A_{n-k} is a matrix from the size $n-k$ $2/\alpha$-Hermite ensemble. Let $T_n = E_\alpha[\det(A_n)^3]$ for all n.

Theorem 8.5.6.

$$\begin{aligned} T_n \quad = \quad & - \ \left(\frac{n-1}{\alpha}\right)_1 \left(3 + \left(\frac{n-2}{\alpha}\right)_2 + \left(\frac{n-1}{\alpha}+1\right)_2\right)\ \ T_{n-2}\ - \\ & - \ \left(\frac{n-1}{\alpha}\right)_1 \left(\frac{n-2}{\alpha}\right)_2 \left(\frac{n-3}{\alpha}\right)_3\ T_{n-4}\ , \end{aligned}$$

with $T_0 = 1$, $T_1 = 0$, $T_2 = -\frac{5}{\alpha} - \frac{3}{\alpha^2} - \frac{1}{\alpha^3}$, *and* $T_3 = 0$. *Note that* $T_{2k+1} = 0$ *for any* k.

Remark 8.5.7. *The above is a linear recurrence.*

Remark 8.5.8. *Technically speaking, this appears to be a 5-term recurrence. However, since the odd-indexed terms are always 0, and for the even-indexed terms the recurrence depends only on the previous two (T_n depends only on T_{n-2} and T_{n-4}), one can safely call this a 3-term recurrence.*

The proof is based on the 3-term recurrence for the determinant of a symmetric tridiagonal matrix. Here and for the rest of this section we use the notation $a_n, \ldots, a_1$ for the diagonal of the matrix, $b_{n-1}, \ldots, b_1$ for the subdiagonal. The distribution of the elements can be found in Chapter 5.

Proof. We start with

$$\det(A_n) = a_n \det(A_{n-1}) - b_{n-1}^2 \det(A_{n-2}) \ , \tag{8.18}$$

once we take the third power on each side and take expectations, we obtain

$$\begin{aligned} T_n &= E_\alpha[\ a_n^3 \det(A_{n-1})^3 \ - \ 3a_n^2 b_{n-1}^4 \det(A_{n-1})^2 \det(A_{n-2}) \ + \\ &\qquad + \ 3a_n b_{n-1}^6 \det(A_{n-1}) \det(A_{n-2})^2 \ - \ b_{n-1}^8 \det(A_{n-2})^3] \ , \end{aligned}$$

whence

$$T_n = -3 \left(\frac{n-1}{\alpha} \right)_1 E_\alpha[\det(A_{n-1})^2 \det(A_{n-2})] - \left(\frac{n-1}{\alpha} \right)_3 T_{n-2} \ , \tag{8.19}$$

where the last equality comes form the mutual independence of a_n and b_{n-1}, and their individual independence from either one of A_{n-1} and A_{n-2}.

We now write, with the help of (8.18) for $n-1$,

$$\begin{aligned} E_\alpha[\det(A_{n-1})^2 \det(A_{n-2})] &= E_\alpha[\det(A_{n-2})(a_{n-1} \det(A_{n-2}) - b_{n-2}^2 \det(A_{n-3}))^2] \\ &= E_\alpha[a_{n-1}^2 \det(A_{n-2})^3 \ - \ 2a_{n-1} b_{n-2}^2 \det(A_{n-2})^2 \det(A_{n-3}) \ + \\ &\qquad + \ b_{n-2}^4 \det(A_{n-2}) \det(A_{n-3})^2] \\ &= T_{n-2} + \left(\frac{n-2}{\alpha} \right)_2 E_\alpha[\det(A_{n-2}) \det(A_{n-3})^2] \ , \end{aligned}$$

once again by using the mutual independence of b_{n-2} and a_{n-1}, and their individual independence from A_{n-2} and A_{n-3}.

We now write

$$\begin{aligned} E_\alpha[\det(A_{n-2})\det(A_{n-3})^2] &= E_\alpha[(a_{n-2}\det(A_{n-3}) - b_{n-3}^2\det(A_{n-4}))\det(A_{n-3})^2] \\ &= -\left(\frac{n-3}{\alpha}\right)_1 E_\alpha[\det(A_{n-3})\det(A_{n-4})^2] \ , \end{aligned}$$

since a_{n-2} and A_{n-3} are mutually independent, and b_{n-3} is independent from either A_{n-3} and A_{n-4}.

Finally, we use a rewrite of (8.19) for $n-2$ to get

$$E_\alpha[\det(A_{n-3})\det(A_{n-4})^2] = \frac{T_{n-2} + \left(\frac{n-3}{\alpha}\right)_3 T_{n-4}}{-3\left(\frac{n-3}{\alpha}\right)_1} \ ,$$

and working backward to (8.19), we obtain Theorem 8.5.6. □

The last recurrence we work out in this section is the one for the fourth moment. Let $Y_n = E_\alpha[\det(A_n)^4]$.

Theorem 8.5.9.

$$\begin{aligned} Y_n &= \left(3 + \left(\frac{n-1}{\alpha}\right)\right) Y_{n-1} + \left(\frac{n-1}{\alpha}\right)_2 \left(3 + \left(\frac{n-1}{\alpha} + 2\right)_2\right) Y_{n-2} - \\ &\quad - \left(\frac{n-1}{\alpha}\right)_2 \left(\frac{n-2}{\alpha}\right)_4 Y_{n-3} \ , \end{aligned}$$

with $Y_0 = 1$, $Y_1 = 3$. $Y_2 = \frac{1}{\alpha^4} + \frac{6}{\alpha^3} + \frac{17}{\alpha^2} + \frac{12}{\alpha} + 9$.

Remark 8.5.10. *The above is a linear recurrence.*

Proof. Using the same notation and independence properties as in the proof of Theorem 8.5.6, we obtain that

$$\begin{aligned} Y_n &= E_\alpha[\det(A_n)^4] = E_\alpha[(a_n\det(A_{n-1}) - b_{n-1}^2\det(A_{n-2}))^4] \\ &= E_\alpha[a_n^4\det(A_{n-1})^4 - 4a_n^3 b_{n-1}^2\det(A_{n-1})^3\det(A_{n-2}) + \\ &\qquad + 6a_n^2 b_{n-1}^4\det(A_{n-1})^2\det(A_{n-2})^2 - \\ &\qquad - 4a_n b_{n-1}^6\det(A_{n-1})\det(A_{n-2})^3 + b_{n-1}^8\det(A_{n-2})^4] \ , \end{aligned}$$

whence

$$Y_n = 3Y_{n-1} + 6\left(\frac{n-1}{\alpha}\right)_2 E_\alpha[\det(A_{n-1})^2\det(A_{n-2})^2] + \left(\frac{n-1}{\alpha}\right)_4 Y_{n-2} \ . \quad (8.20)$$

The next step is to write

$$\begin{aligned} E_\alpha[\det(A_{n-1})^2 \det(A_{n-2})^2] &= E_\alpha[\det(A_{n-2})^2 (a_{n-1}\det(A_{n-2}) - b_{n-2}^2 \det(A_{n-3}))^2] \\ &= E_\alpha[a_{n-1}^2 \det(A_{n-2})^4 - 2a_{n-1}b_{n-2}^2 \det(A_{n-2})^3 \det(A_{n-3}) + \\ &\qquad + b_{n-2}^4 \det(A_{n-2})^2 \det(A_{n-3})^2] \\ &= Y_{n-2} + \left(\frac{n-2}{\alpha}\right)_2 E_\alpha[\det(A_{n-2})^2 \det(A_{n-3})^2] \ ; \end{aligned}$$

finally, rewriting (8.20) for $n-1$, we obtain

$$E_\alpha[\det(A_{n-2})^2 \det(A_{n-3})^2] = \frac{Y_{n-1} - 3Y_{n-2} - \left(\frac{n-2}{\alpha}\right)_4 Y_{n-3}}{6\left(\frac{n-2}{\alpha}\right)_2} \ ,$$

working backwards to (8.20) we obtain the 4-term recurrence for Y_n. □

Remark 8.5.11. *The method we employed in finding recurrences for the third and fourth moments extends to any other power k.*

Consider the $(k+1) \times n$ table which has in its (i,j) entry the expected value $E_\alpha[\det(A_j)^{i-1} \det(A_{j-1})^{k-i+1}]$. Note that $A_0 = 1$ (non-randomly), and we can fill the first column with the moments of the Gaussian.

It is not hard to see that the entries in the jth column depend exclusively on the $(k+1)$ entries of the previous column (an immediate consequence of the 3-term recurrence for the determinant). Moreover, the coefficients involved essentially encode the 0 through $2k$ moments of a_j and b_j; hence for k fixed, they take $O(1)$ to compute.

Since we are interested in the kth row of the matrix, it follows that the work done in filling this matrix is linear in n. Thus with the help of the 3-term recurrence for the determinant, one can compute $E_\alpha[\det(A_n)^k]$ in time linear in n for any k fixed.

Chapter 9

MOPs: A Maple Library for Multivariate Orthogonal Polynomials (symbolically)

As mentioned in Section 2.5, our Maple Library *MOPs* implements some new and some known algorithms for computing the Jack, Hermite, Laguerre, and Jacobi multivariate polynomials for arbitrary $\alpha = 2/\beta$. This library can be used as a tool for conjecture-formulation and testing, for statistical computations, or simply for getting acquainted with the mathematical concepts.

As an example of a calculation that can pe performed with *MOPs* is the probability that a random Gaussian matrix has exactly k real eigenvalues. This probability has a closed-form answer as an integral, which was computed by Edelman [27, Theorem 7.1]. The main term in the integral is given by an expectation over the 1-Hermite ensemble or GOE of a product of determinants. Using *MOPs*, we can evaluate that expectation and perform the rest of the computation, to get an exact answer.

All the algorithms presented here keep the variables in symbolic format; whether we expand the polynomials in the monomial or in the Jack polynomial basis, we ultimately compute *coefficients*, which can either be numerical values, or (in the symbolic evaluations) rational functions of m, the number of variables, and/or α, the Jack parameter.

Aside from the routines analyzed here, *MOPs* includes routines for numerical evaluation of the Jack, Hermite, Laguerre, and Jacobi polynomials. These algorithms take numerical values for the *variables*; they compute quantities such as $C^3_{[3,2]}(2.5, 1.09, -10)$. The algorithms we use for these routines have been developed and analyzed by Koev and Demmel [19].

9.1 Computing Jack and Multivariate Orthogonal Polynomials

In this section we present the recurrences and formulas which can be used in order to compute Jack, Hermite, Laguerre, and Jacobi polynomials, and generalized binomial coefficients. These recurrences appear in the literature (see James [49]) for $\alpha = 2$; we have used the methods employed by James to compute the general α versions, and obtained similar results, the derivations of which we choose not to present. Since we implemented these recurrences, we were forced to consider certain details that might have been otherwise overlooked. We present them here, in Lemmas 9.1.2 and 9.1.4.

9.1.1 Computing Jack Polynomials

From the Laplace-Beltrami equation, one can find an expansion for the Jack polynomials of the type

$$C^\alpha_\kappa(x_1, x_2, \ldots, x_m) = \sum_{\lambda \le \kappa} c^\alpha_{\kappa,\mu} m_\lambda(x_1, x_2, \ldots, x_m) \ ,$$

where λ and κ are both partitions of the same integer $|\kappa|$, and the order imposed on partitions is the lexicographic one. The coefficients $c^\alpha_{\kappa,\lambda}$ depend on all three parameters; $m_\lambda(x_1, x_2, \ldots, x_m)$ is the monomial function corresponding to λ.

Note that as a consequence of the above, if $l(\kappa) > m$, $C^\alpha_\kappa(x_1, x_2, \ldots, x_m) = 0$ ("there is no highest-order term").

Using the eigenfunction equation

$$D^* C^\alpha_\kappa = (\rho^\alpha_\kappa + k(m-1)) C^\alpha_\kappa \ , \tag{9.1}$$

where

$$\rho_\kappa^\alpha = \sum_{i=1}^{m} k_i(k_i - 1 - \frac{2}{\alpha}(i-1))$$

one can obtain a recurrence for $c_{\kappa,\lambda}^\alpha$ from which the Jack polynomials can be explicitly calculated. This recurrence is

$$c_{\kappa,\lambda}^\alpha = \frac{\frac{2}{\alpha}}{\rho_\kappa^\alpha - \rho_\lambda^\alpha} \sum_{\lambda \langle \mu \leq \kappa} \Big((l_i + t) - (l_j - t)\Big) c_{\kappa,\mu}^\alpha \ , \tag{9.2}$$

where $\lambda = (l_1, \ldots, l_i, \ldots, l_j, \ldots, l_m)$, $\mu = (l_1, \ldots, l_i + t, \ldots, l_j - t, \ldots, l_m)$, and μ has the property that, when properly reordered, it is between λ (strictly) and κ in lexicographic order.

In fact we can do better, using two propositions found in Macdonald's book [63, (10.13), (10.15)]. Roughly the content of the two propositions is that the Jack polynomials, in "P" normalization, can be written as

$$P_\kappa^\alpha = m_\kappa + \sum_{\lambda \prec \kappa} u_{\kappa,\lambda}^\alpha m_\lambda \ ,$$

with $u_{\kappa,\lambda}^\alpha > 0$ whenever $\kappa \succ \lambda$ (the order imposed on partitions here being the domination order).

Thus it follows that the recurrence can be improved to

$$c_{\kappa,\lambda}^\alpha = \frac{\frac{2}{\alpha}}{\rho_\kappa^\alpha - \rho_\lambda^\alpha} \sum_{\lambda \prec \mu \preceq \kappa} \Big((l_i + t) - (l_j - t)\Big) c_{\kappa,\mu}^\alpha \ , \tag{9.3}$$

where $\lambda = (l_1, \ldots, l_i, \ldots, l_j, \ldots, l_m)$, $\mu = (l_1, \ldots, l_i + t, \ldots, l_j - t, \ldots, l_m)$, and μ has the property that, when properly reordered, it is between λ (strictly) and κ in domination order.

This recurrence, at first glance, seems to be enough to compute all coefficients $c_{\kappa,\lambda}^\alpha$, once $c_{\kappa,\kappa}^\alpha$ is found. However, one has to account for the possibility that $\rho_\kappa^\alpha = \rho_\lambda^\alpha$ for some λ different from κ; what can one do in that case?

Fortunately, this never happens. We first need the following well known Proposition.

Proposition 9.1.1. *The dominance ordering is a lattice on the set of partitions of a given number. In particular, between any partitions κ and λ such that $\kappa \succ \lambda$, there*

exists a "path" on this lattice, $\sigma^0 = \kappa \succ \sigma^1 \succ \ldots \succ \sigma^t = \lambda$*, such that* σ^{i+1} *differs from* σ^i *in the following way: there exists* $i_1 < i_2$ *such that* σ^{i+1} *and* σ^i *agree in all places but* i_1 *and* i_2*,* $(\sigma^{i+1})_{i_1} = (\sigma^i)_{i_1} - 1$*, and* $(\sigma^{i+1})_{i_2} = (\sigma^i)_{i_2} + 1$*.*

Now we can prove that we never divide by 0 in computing Recurrence 9.3.

Lemma 9.1.2. *If* $\lambda \prec \kappa$*, then* $\rho^\alpha_\lambda \neq \rho^\alpha_\kappa$*, for all* $\alpha > 0$*.*

Proof. Let $\lambda \prec \kappa$ be two partitions, let $m = \max\{length(\kappa), length(\lambda)\}$, and assume that there is some $\alpha > 0$ such that

$$\rho^\alpha_\lambda = \rho^\alpha_\kappa \ .$$

Since the two partitions sum to the same number, the above is equivalent to

$$\sum_{i=1}^{m} k_i^2 - \lambda_i^2 = \frac{2}{\alpha} \sum_{i=1}^{m} (k_i - \lambda_i)(i-1) \ .$$

The right-hand side is non-negative (as an immediate consequence of the strict ordering).

We show that the left-hand side is positive by induction. For that we will use Proposition 9.1.1, which shows that it is enough to prove that

$$\sum_{i=1}^{m} k_i^2 - \lambda_i^2 \geq 0$$

in the case when κ and λ differ only in two places, $i_1 < i_2$. Note that if $\kappa_{i_1} = \lambda_{i_1} + 1$ and $\kappa_{i_2} = \lambda_{i_2} - 1$, this implies that $\kappa_{i_1} \geq \kappa_{i_2} + 2$. Hence

$$\sum_{i=1}^{m} k_i^2 - \lambda_i^2 = k_{i_1}^2 - \lambda_{i_1}^2 + k_{i_2}^2 - \lambda_{i_2}^2 = 2k_{i_1} - 1 - 2k_{i_2} - 1 \geq 2 > 0 \ ,$$

and we are done. □

Proposition 9.1.1 ensures thus that once $c^\alpha_{\kappa\kappa}$ is determined, every other non-zero coefficient is uniquely determined.

Finally, for $c^\alpha_{\kappa\kappa}$ we use the following formula (deduced on the basis of Table 8.1 and the fact that P^α_κ has highest-order coefficient 1):

$$c^\alpha_{\kappa\kappa} = \frac{\alpha^k k!}{c'(\kappa, \alpha)} \ .$$

9.1.2 Computing Generalized Binomial Coefficients

One can prove, using the eigenfunction equation (9.1) and the definition of the generalized binomial coefficients, that

$$\sum_i \binom{\sigma^{(i)}}{\sigma} \binom{\kappa}{\sigma^{(i)}} = (k-s)\binom{\kappa}{\sigma} , \tag{9.4}$$

where $|\sigma| = s$, $|\kappa| = k$, $\sigma^{(i)} = (\sigma_1 \ldots, \sigma_i + 1, \ldots, \sigma_m)$. All generalized binomial coefficients can be found by recursion, once one has a way to compute the so-called "contiguous" coefficients $\binom{\sigma^{(i)}}{\sigma}$.

To compute the contiguous coefficients, we use Proposition 2 from [56], applied to $\kappa = \sigma^{(i)}$, and simplified slightly:

$$\binom{\sigma^{(i)}}{\sigma} = j_\sigma^{-1} g_{\sigma\ 1}^{\sigma^{(i)}} , \tag{9.5}$$

where $g_{\sigma\ 1}^{\sigma^{(i)}}$ is

$$g_{\sigma\ 1}^{\sigma^{(i)}} = \left(\prod_{s\in\sigma} A_{\sigma^{(i)}}\right)\left(\prod_{s\in\sigma} B_{\sigma^{(i)}}\right) .$$

Here

$$A_{\sigma^{(i)}} = \begin{cases} h_*^\sigma(s), & \text{if } s \text{ is not in the } i\text{th column of } \sigma , \\ h_\sigma^*(s), & \text{otherwise .} \end{cases}$$

$$B_{\sigma^{(i)}} = \begin{cases} h^*_{\sigma^{(i)}}(s), & \text{if } s \text{ is not in the } i\text{th column of } \sigma , \\ h_*^{\sigma^{(i)}}(s), & \text{otherwise .} \end{cases}$$

Knowing the contiguous coefficients allows for computing all the generalized binomial coefficients.

Remark 9.1.3. *The generalized binomial coefficients are independent of the number of variables. They are rational functions of α.*

9.1.3 Computing Jacobi Polynomials

From the differential equation, one obtains that the Jacobi polynomials can be written in the corresponding Jack polynomial basis as

$$P_\kappa^{\alpha,g_1,g_2}(x_1,\ldots,x_m) = (g_1 + \frac{m-1}{\alpha} + 1)_\kappa C_\kappa^\alpha(I_m) \sum_{\sigma\subseteq\kappa} \frac{(-1)^s c_{\kappa\sigma}}{(g_1+\frac{m-1}{\alpha}+1)_\sigma} \frac{C_\sigma^\alpha(x_1,\ldots,x_m)}{C_\sigma^\alpha(I_m)} \tag{9.6}$$

where the coefficients $c^\alpha_{\kappa\sigma}$ satisfy the recursion

$$c^\alpha_{\kappa\sigma} = \frac{1}{\left((g_2+g_1+\frac{2}{\alpha}(m-1)+2)(k-s)+\rho^\alpha_\kappa-\rho^\alpha_\sigma\right)} \sum_{i \text{ allowable}} \binom{\kappa}{\sigma^{(i)}}\binom{\sigma^{(i)}}{\sigma} c^\alpha_{\kappa\sigma^{(i)}} \quad (9.7)$$

with the previous notation for ρ^α_κ, I_m, and $\sigma^{(i)}$. The question is again whether we are always allowed to make the division.

Lemma 9.1.4. *Under the assumptions that* $g_1, g_2 > -1$, $(g_2+g_1+\frac{2}{\alpha}(m-1)+2)(k-s)+\rho^\alpha_\kappa-\rho^\alpha_\sigma$ *is never* 0.

Proof. The proof is very similar with the proof of Lemma 9.1.2; the two crucial facts here are that one needs one show it for the case $\kappa = \sigma^{(i)}$, and that g_1 and g_2 are both larger than -1 (due to (8.2)). □

Letting $c^\alpha_{\kappa,\kappa} = 1$ for all κ and α allows all the coefficients to be uniquely determined.

9.1.4 Computing Laguerre Polynomials

The Laguerre polynomials have the simplest expansion in terms of the Jack polynomials, which generalizes straightaway from the expansion of the univariate Laguerre polynomials in terms of powers.

The multivariate expansion can be written in closed form as

$$L^{\alpha,\gamma}_\kappa(x_1,\ldots,x_m) = (\gamma+\tfrac{m-1}{\alpha}+1)_\kappa C^\alpha_\kappa(I_m) \sum_{\sigma\subseteq\kappa} \frac{(-1)^s\binom{\kappa}{\sigma}}{(\gamma+\frac{m-1}{\alpha}+1)_\sigma} \frac{C^\alpha_\sigma(x_1,\ldots,x_m)}{C^\alpha_\sigma(I_m)} \ . \quad (9.8)$$

Note that the coefficient of $C^\alpha_\kappa(x_1,\ldots,x_m)$ in $L^{\alpha,\gamma}_\kappa(x_1,\ldots,x_m)$ is $(-1)^k$.

9.1.5 Hermite Polynomials

We present here a recurrence used to compute the Hermite polynomials, rather than the formula of Lemma 8.4.1, because the latter is more computationally expensive.

Once again using the corresponding differential operator, we obtain a recurrence for the coefficients of the polynomial. Let

$$H^\alpha_\kappa(x_1,\ldots,x_n) = \sum_{\sigma\subseteq\kappa} c^\alpha_{\kappa,\sigma} \frac{C^\alpha_\sigma(x_1,\ldots,x_n)}{C_\sigma(I_n)} \ , \quad (9.9)$$

and

$$c^{\alpha}_{\kappa,\sigma} = \frac{1}{k-s}\left(\sum_i \binom{\sigma^{(i)(i)}}{\sigma^{(i)}}\binom{\sigma^{(i)}}{\sigma} c^{\alpha}_{\kappa,\sigma^{(i)(i)}} + \right. \tag{9.10}$$

$$\left. + \sum_{i<j}(\sigma_i - \sigma_j - \frac{1}{\alpha}(i-j))\left(\binom{\sigma^{(i)(j)}}{\sigma^{(j)}}\binom{\sigma^{(j)}}{\sigma}\right) c^{\alpha}_{\kappa,\sigma^{(i)(j)}}\right) . \tag{9.11}$$

In the above, $i < j$ take on all admissible values.

Note that if $\sigma \not\subseteq \kappa$ or $k \neq s \pmod 2$, then the above is 0 (this is a nice generalization of the univariate case).

For consistency with the explicit formula of Lemma 8.4.1, we choose $c^{\alpha}_{\kappa,\kappa} = C^{\alpha}_{\kappa}(I_n)$.

9.2 Algorithms

In this section we analyze the complexity of the algorithms to compute the Jack, Hermite, Laguerre, and Jacobi polynomials. We also present and analyze the algorithm for generalized binomial coefficients, as it is a component of the analysis of algorithms for the aforementioned polynomials.

Our complexity bounds are upper bounds, but we believe them to be asymptotically correct. They work well for the numerical evaluation; symbolic evaluation of the polynomials is considerably slower. We are not aware of the existence of a good symbolic performance model for Maple, and hence it would be difficult to predict how much slower symbolic evaluation is than numerical evaluation. Since the coefficients in the polynomials we compute are rational functions of m (the number of variables) and α, of degrees that can go up to $|\kappa|$ (the partition size), storage is another issue in symbolic evaluations, and hence one would expect that the running times for symbolic evaluation would be orders of magnitude slower than for numerical evaluation (and hence, the partition size that would be "maximally reasonable" to attempt evaluation on should be considerably lower than for numerical evaluation).

For each algorithm, we provide a complexity analysis, and we illustrate the performance in practice by providing running times for different tests (both numerical and symbolic); then we examine the running times and draw a set of conclusions.

Each time we used N/A for an entry in a running time table, we have done so because that particular computation has exhausted the memory available to Maple, and hence (regardless of the time it took up to that point) the computation was not completed.

The computer we have performed all tests is a Pentium 4 by Dell, 1.8 Ghz, 512 MB, and we have used Maple 8.

9.3 Complexity bounds: theory and practice

The algorithms we use to compute the polynomials are implementations of the recurrences in Section 9.1; thus we do not see the need to present them in pseudo-code format. We also note that the U_κ or in other cases $U_{\kappa,\sigma}$ (see Table 9.1) are natural lower bounds for any of the quantities we compute.

One thing worth mentioning is that Maple allows for the storage (and recall) of each quantity previously computed, and our library uses this option.

Throughout this section, we will follow the notations given below.

$k = \|\kappa\|$	size of partition κ
$s = \|\sigma\|$	size of partition σ
l =length(κ)	length of partition κ
P_κ	number of partitions of k dominated by κ
$P_{[k]}$	number of partitions of the number k (each partition of k is dominated by $[k]$)
U_κ	number of subpartitions of κ
$U_{\kappa,\sigma}$	number of subpartitions of κ which are superpartitions for σ (this implies σ is a subpartition of κ)
A_κ	number of subpartitions of κ which sum to a number with the same parity with k

Table 9.1: Notations to be used throughout Chapter 9.

Note that by Ramanujan's formula

$$P_{[k]} \sim \frac{1}{4k\sqrt{3}}\, e^{\pi\sqrt{2k/3}} \quad . \tag{9.12}$$

9.3.1 Jack Polynomials

The algorithm uses recurrence 9.2, together with the "boundary" conditions $c_{\kappa,\lambda} = 0$ if $\kappa \not\geq \lambda$ in dominance ordering, and $c_{\kappa,\kappa} = \frac{\alpha^k k!}{c'(\kappa,\alpha)}$. The length of the recurrence is at most $O(k_1\binom{k+1}{2})$, with k_1 being the first entry in the partition, and the algorithm will check each of the possible partitions μ (at most $k_1\binom{k+1}{2}$) to see if they are dominated by κ and dominating λ (this involves l additions and l comparisons). The rest of the computation has complexity $O(k)$. Thus the complexity of the algorithm is $O(k_1 k^3 P_\kappa)$. Note that P_κ, at worst, is $P_{[k]}$, so the algorithm has super-exponential complexity.

Below we illustrate the running times for both numerical and symbolic computations. For numerical computations, we have chosen to make $\alpha = 1$, (so that the Jack polynomials are the Schur functions). Note that we do not test the partition $[k]$; for that particular partition we have a closed-form formula for the Jack polynomial, due to Stanley [86], which has complexity only $O(kP_k)$.

Remark 9.3.1. *Note that the ratio of the running times increases when the partition size increases. At $k = 30$, the number of partitions is 5604, and each of the monomial coefficients is a rational function of α. Issues like storage and memory access become important, and influence negatively the running times. Another important factor is that in order to make things easier to store and access, not to mention easier to read and interpret, we use the procedures "simplify" and "factor", which are relatively costly.*

Extrapolation. Suppose the speed/memory of a top-of-the-line computer goes up by a factor of 10^3 every 10 years. Then within the next decade, using MOPS, computing $J^\alpha_{(59,1)}$ will take about 30 minutes.

k	κ	Running time, $\alpha = 1$	Running time, α symbolic	Ratio
15	$\kappa = [14, 1]$	2.48	4.54	1.83
	$\kappa = [8, 7]$	1.79	3.17	1.77
	$\kappa = [3, 3, 3, 3, 3]$	0.39	0.50	1.28
20	$\kappa = [19, 1]$	16.97	30.45	1.79
	$\kappa = [10, 10]$	11.53	20.32	1.76
	$\kappa = [4, 4, 4, 4, 4]$	2.91	4.02	1.38
25	$\kappa = [24, 1]$	93.42	189.66	2.03
	$\kappa = [9, 8, 8]$	46.85	79.85	1.70
	$\kappa = [5, 5, 5, 5, 5]$	16.08	24.18	1.50
30	$\kappa = [29, 1]$	634.32	1819.65	2.86
	$\kappa = [10, 10, 10]$	214.10	418.19	1.95
	$\kappa = [6, 6, 6, 6, 6]$	73.54	113.55	1.54

Table 9.2: Running times (in seconds) for the Jack polynomial computation.

9.3.2 Generalized Binomial Coefficients

We use 9.4, together with the boundary conditions listed in Section 8.3.2 and with the contiguous binomial formula 9.5. Computing each contiguous binomial coefficient has complexity $O(k)$, and one needs to compute no more than l such coefficients per subpartition σ of κ. Thus one immediately obtains the bound $O(klU_{\kappa,\sigma})$ for the complexity of computing $\binom{\kappa}{\sigma}$.

Note that by computing $\binom{\kappa}{\sigma}$, one also obtains $\binom{\kappa}{\mu}$, for each $\sigma \subseteq \mu \subset \kappa$. So we have chosen for our tests to compute $\binom{\kappa}{[1,1]}$ for different κ, as this yields all the binomial coefficients having κ as top partition (except $\binom{\kappa}{2}$, but that requires only an additional kl complexity).

Remark 9.3.2. *Once again, size and length of the partition increase the symbolic running times; however, note that the running times are relatively small, even for partitions of* 30*. We believe that the generalized binomial coefficients are rational functions of α which can always be factored in small-degree factors, so that they are*

k	κ	Running time, $\alpha = 1$	Running time, α symbolic	$U_{\kappa,\ [1^2]}$
15	$[6,4,2,2,1]$	0.22	1.12	139
	$[3,3,3,3,3]$	0.05	0.18	56
	$[10,5]$	0.03	0.15	51
20	$[6,4,3,2,2,1,1,1]$	1.01	6.68	418
	$[4,4,4,4,4]$	0.17	0.6	126
	$[12,8]$	0.07	0.28	81
25	$[7,5,4,3,2,2,1,1]$	3.41	23.37	1077
	$[5,5,5,5,5]$	0.41	1.67	252
	$[16,9]$	0.15	0.62	125
30	$[8,6,4,3,2,2,1,1,1,1,1]$	11.87	89.61	2619
	$[6,6,6,6,6]$	0.91	3.95	462
	$[20,10]$	0.24	1.20	176

Table 9.3: Running times (in seconds) for the generalized binomial coefficient computation.

easy to store and operate with.

9.3.3 Jacobi Polynomials

To compute the Jacobi polynomials, we use Format 9.6 and Recurrence 9.7. One can easily see that at each step, one needs to compute at most l contiguous binomial coefficients, each of which has complexity $O(k)$; in addition, one needs to compute another at most l binomial coefficients; each of these takes only $O(l)$, as the contiguous coefficients needed *have already been computed* at the previous step. Hence the total complexity is $O(kl)$ (since $l \leq k$) at each step, for a total of $O(klU_{\kappa,\ [1^2]})$.

Hence computing numerically the Jacobi polynomials is comparable to computing the generalized binomial coefficients $\binom{\kappa}{[1,1]}$; however, the constant for the Jacobi polynomial complexity is considerably larger (we are guessing it is around 8).

The Jacobi parameters we chose for each of the computations below are 0 and 1.

k	κ	Running time, $\alpha = 1,\ m = l$	Running time, m symbolic	Running time, α, m symbolic	U_κ
10	$[4,2,2,1,1]$	0.27	0.74	22.12	42
	$[4,3,3]$	0.11	0.35	1.88	30
	$[7,3]$	0.10	0.30	1.57	26
15	$[6,4,2,2,1]$	1.05	11.08	N/A	139
	$[3,3,3,3,3]$	0.39	0.87	63.07	56
	$[10,5]$	0.19	1.01	27.98	51
20	$[6,4,3,2,2,1,1,1]$	5.94	N/A	N/A	418
	$[4,4,4,4,4]$	0.63	8.24	N/A	126
	$[12,8]$	0.26	3.51	N/A	81
25	$[7,5,4,3,2,2,1,1]$	18.61	N/A	N/A	1077
	$[5,5,5,5,5]$	1.23	N/A	N/A	252
	$[16,9]$	0.45	N/A	N/A	125

Table 9.4: Running times (in seconds) for the Jacobi polynomial computation.

Remark 9.3.3. *While the running times for numerical evaluation are reasonable, they explode when a symbolic parameter is introduced. The coefficients of the polynomial are rational functions of that parameter or combination of parameters, of order up to $k(k-1)/2$. We recall that there are $U_{\kappa,\ [1^2]}$ of them, a potentially superpolynomial number, which explains the tremendous increase in the running time.*

9.3.4 Laguerre Polynomials

We use Format 9.8; it is easily established that the complexity of computing the Laguerre polynomial is dominated by the cost of computing the binomial coefficients, that is $O(klU_{\kappa,\ [1^2]})$.

The Laguerre parameter we chose for each of the computations below is 1.

k	κ	Running time, $\alpha = 1,\ m = l$	Running time, m symbolic	Running time, α, m symbolic	U_κ
10	$[4,2,2,1,1]$	0.12	0.23	0.54	42
	$[4,3,3]$	0.07	0.14	0.31	30
	$[7,3]$	0.07	0.10	0.28	26
15	$[6,4,2,2,1]$	0.49	0.82	2.95	139
	$[3,3,3,3,3]$	0.18	0.27	0.84	56
	$[10,5]$	0.11	0.22	0.81	51
20	$[6,4,3,2,2,1,1,1]$	2.26	3.37	16.08	418
	$[4,4,4,4,4]$	0.44	0.69	2.74	126
	$[12,8]$	0.20	0.37	1.79	81
25	$[7,5,4,3,2,2,1,1]$	7.23	11.06	67.92	1077
	$[5,5,5,5,5]$	0.96	1.53	8.06	252
	$[16,9]$	0.32	0.69	4.21	125

Table 9.5: Running times (in seconds) for the Laguerre polynomial computation.

Remark 9.3.4. *For the Laguerre polynomials, even in the all-symbolic case, the computation is very easy, and the storage required is relatively small. This explains why it is possible to obtain them without much effort, in any one of the cases.*

9.3.5 Hermite Polynomials

We use Format 9.9 and Recurrence 9.10. We only do work for those coefficients that correspond to subpartitions σ of κ such that $|\sigma| \equiv k \pmod 2$. There are A_κ of them. For each, we compute at most $\binom{l}{2}$ contiguous coefficients, each done with work $O(k)$; all of the other work is $O(k)$. Hence the total work is $O(kl^2 A_\kappa)$.

Remark 9.3.5. $A_\kappa = O(U_\kappa)$; $A_\kappa \sim U_\kappa/2$.

Remark 9.3.6. *Note that when m is parametrized, but $\alpha = 1$, the computation is almost as fast as in the all-numerical case. That happens because the dependence*

k	κ	Running time, $\alpha = 1,\ m = l$	Running time, m symbolic	Running time, α, m symbolic	A_κ
10	$[4,2,2,1,1]$	0.21	0.24	0.75	22
	$[4,3,3]$	0.09	0.11	0.33	16
	$[7,3]$	0.05	0.06	0.24	14
15	$[6,4,2,2,1]$	0.41	2.83	42.92	88
	$[3,3,3,3,3]$	0.13	0.17	1.83	38
	$[10,5]$	0.10	0.12	1.10	30
20	$[6,4,3,2,2,1,1,1]$	1.93	2.39	N/A	211
	$[4,4,4,4,4]$	0.35	0.51	N/A	66
	$[12,8]$	0.18	0.25	13.49	43
25	$[7,5,4,3,2,2,1,1]$	6.23	7.53	N/A	1077
	$[5,5,5,5,5]$	0.90	1.20	N/A	252
	$[16,9]$	0.29	0.50	106.56	125

Table 9.6: Running times (in seconds) for the Hermite polynomial computation.

on m is very simple, and it only involves Pochhammer symbols, which do not get expanded (so that the storage required is minimal). However, the dependence on α is more complicated, and the rational functions obtained as coefficients are complex and hard to store. Hence the running time for the all-symbolic computation increases dramatically.

9.3.6 Computing Integrals

To evaluate integral s over the β-Hermite, β-Laguerre, and β-Jacobi ensembles, we are using formulas (8.6), (8.4), and (8.5) as building blocks; they provide the answer for the integrals of Jack polynomials.

To compute the integral of a symmetric polynomial over the ensembles, one needs to write the polynomial either in the Jack basis or in the monomial basis. We have conversion algorithms that take expressions involving sums of products of Jack/monomial

symmetric functions, and rewrite them as simple sums of Jack polynomials.

Once the symmetric polynomial is expressed in terms of Jack symmetric functions, the integral is computed as the sum of the integral s of the Jack polynomials involved.

We provide an example below.

Example. Suppose we want to compute the expected value of

$$z(\alpha, x_1, x_2, x_3) := J^{\alpha}_{[2,1]}(x_1, x_2, x_3) C^{\alpha}_{[1,1,1]}(x_1, x_2, x_3)$$

over the Hermite distribution. First we have to express z as a linear combination of Jack "C" Polynomials. Note that the number of variables, as well as α, must be the same in the two terms of z.

First, we express the two terms in the monomial basis (this can be obtained with a call to the "jack" routine):

$$\begin{aligned} J^{\alpha}_{[2,1]}(x_1, x_2, x_3) &= (2+\alpha)\ m_{[2,1]}(x_1, x_2, x_3) + 6\ m_{[1,1,1]}(x_1, x_2, x_3)\ , \\ C^{\alpha}_{[1,1,1]}(x_1, x_2, x_3) &= \frac{6\alpha^2}{(1+\alpha)(2+\alpha)}\ m_{[1,1,1]}(x_1, x_2, x_3). \end{aligned}$$

Their product thus becomes a linear combination of sums of products of two monomials, which are in turn converted to linear combinations of monomials (this can be obtained with a call to the "m2m" routine):

$$\begin{aligned} m_{[2,1]}(x_1, x_2, x_3)\ m_{[1,1,1]}(x_1, x_2, x_3) &= m_{[3,2,1]}(x_1, x_2, x_3)\ , \quad \text{while} \\ m_{[1,1,1]}(x_1, x_2, x_3)^2 &= m_{[2,2,2]}(x_1, x_2, x_3)\ . \end{aligned}$$

Putting it all together, in the monomial basis,

$$\begin{aligned} z(\alpha, x_1, x_2, x_3) &= \frac{6\alpha^2}{1+\alpha}\ m_{[3,2,1]}(x_1, x_2, x_3) \quad + \\ &\quad + \quad \frac{36\alpha^2}{(1+\alpha)(2+\alpha)}\ m_{[2,2,2]}(x_1, x_2, x_3)\ . \end{aligned}$$

All that is left now is to convert from the monomial basis back to the Jack Polynomial basis, and to obtain that

$$z(\alpha, x_1, x_2, x_3) = \frac{1}{120} \frac{(2+3\alpha)(1+2\alpha)^2}{\alpha(1+\alpha)}\ C^{\alpha}_{[3,2,1]}(x_1, x_2, x_3)$$

The above involves a call to the "m2jack" routine.

We are now able to finish the work:

$$E_H\big[z(\alpha, x_1, x_2, x_3)\big] = -\frac{36(\alpha^2 + 2\alpha - 3)}{(1+\alpha)(2+\alpha)} \ .$$

All of the above can be realized with a simple call to "expHjacks", which runs in less than 0.1 seconds.

The implementations of the (8.4) and (8.5) formulas have both complexity $O(k)$; (8.6) is slightly more complicated. We have chosen to use (8.8) in order to compute the value at $(0, 0, \ldots, 0)$ of the Hermite polynomial. The complexity of this formula is comparable to the complexity of computing the generalized binomial coefficient $\binom{\kappa}{[1,1]}$; it is $O(klA_\kappa)$.

Below we present some sample running times.

k	κ	Running time, $\alpha = 1,\ m = l$	Running time, α symbolic	Running time, α, m symbolic	A_κ
10	$[4,2,2,1,1]$	0.08	0.39	0.47	22
	$[4,3,3]$	0.05	0.26	0.36	16
	$[7,3]$	0.04	0.24	0.29	14
16	$[6,4,3,2,1]$	0.42	10.47	12.77	88
	$[4,4,4,4]$	0.14	8.96	9.92	38
	$[10,6]$	0.08	7.11	8.34	30
20	$[6,4,3,2,2,1,1,1]$	1.36	N/A	N/A	211
	$[4,4,4,4,4]$	0.32	N/A	N/A	66
	$[12,8]$	0.17	106.47	109.66	43
24	$[7,5,4,3,2,2,1]$	3.21	N/A	N/A	425
	$[4,4,4,4,4,4]$	0.61	N/A	N/A	110
	$[16,8]$	0.27	N/A	N/A	61

Table 9.7: Running times (in seconds) for the integral of the Jack polynomial with respect to the $2/\alpha$-Hermite distribution.

Since the algorithms for evaluating the integral s of the Jack polynomial over the $2/\alpha$-Laguerre and $2/\alpha$-Jacobi distributions have complexity $O(k)$, they are incredibly fast, literally taking fractions of a second to compute when all the parameters are numerical, and taking seconds in the worst cases, when all parameters are symbolic.

Chapter 10

Other eigenvalue statistics

In this chapter we list some of the Laguerre/Wishart $\beta = 1, 2$ eigenvalue statistics that are found in the literature, and present their generalizations to general β. There is an extensive literature on this subject, starting with the work of James [47], Constantine [14], continuing with Muirhead [70], which is probably the best reference for $\beta = 1$, Krishnaiah and Chang [58], Sugiyama [88], Silverstein [81], Edelman [24, 25, 26], and many others.

It is our strong belief that each and every one of the Laguerre/Wishart eigenvalue statistics that have been proved for $\beta = 1$ and 2 admits a simple and similar generalization to $\beta > 0$. Moreover, we believe that most eigenvalue statistics that have been proved for the Gaussian/Hermite ensembles also admits a (perhaps less obvious) generalization to $\beta > 0$.

10.1 Smallest eigenvalue statistics

We present here two theorems which are immediate generalizations from the $\beta = 1$ case to all β. Except for the changes in the parameters, the proofs are identical to the original ones. In the first case, all the tools involving $\beta = 1$ hypergeometric functions have been generalized to all β (for a reference, see Forrester [30]). In the second case, no new tools are necessary; the conditions that are imposed arise from the proof in the same way they arose for $\beta = 1$.

The $\beta = 1$ version of the theorem below belongs to Krishnaiah and Chang [58]. It can also be found in Muirhead [70, page 423], as a finite sum of zonal polynomials.

Theorem 10.1.1. *Let L be a $n \times n$ β-Laguerre matrix of parameter a, and assume that $k = a - \frac{\beta}{2}(n-1) - 1$ is an integer. Then the p.d.f. of the smallest eigenvalue of L is proportional to*

$$f_{n,\beta}(x) = x^{kn}\ e^{-xn/2}\,{}_2F_0^{\beta}(-k, \frac{\beta}{2}n + 1; -\frac{2}{x}I_{n-1})\ ,$$

where ${}_2F_0^{\beta}$ is the multivariate hypergeometric function of parameter β, and $I_{n-1} = (1, 1, \ldots, 1)$.

The second theorem in this section has a $\beta = 1$ version which belongs to Edelman [24, 25].

Theorem 10.1.2. *Let L be a $n \times n$ β-Laguerre matrix of parameter a, and assume that $a - \frac{\beta}{2}(n-2) = 1$. Then the p.d.f. of the smallest eigenvalue of L is given by*

$$f_{n,\beta}(x) = \frac{n\ 2^{\beta/2-1}}{\Gamma(1-\frac{\beta}{2})}\ x^{-\beta/2}\ e^{-xn/2}\ U(\frac{\beta}{2}(n-1), -\frac{\beta}{2}, \frac{x}{2})\ ,$$

where $U(a, b, z)$ is the Tricomi function, i.e. the unique solution of the Kummer equation

$$z^2\frac{d^2w}{dz^2} + (b - z)\frac{dw}{dz} - aw = 0$$

with $U(a, b, 0) = \frac{\Gamma(1-b)}{\Gamma(1+a-b)}$ and $U(a, b, \infty) = 0$.

Remark 10.1.3. *Note that the requirement $a - \frac{\beta}{2}(n-2) = 1$, together with the fact that $a > \frac{\beta}{2}(n-1)$, implies that $\beta < 2$. Moreover, for each such β, for each n, there exists precisely one a which satisfies the constraints.*

10.2 Largest eigenvalue statistics

We present here a theorem giving the c.d.f. of the largest eigenvalue of a β-Laguerre matrix of size n and parameter a. The $\beta = 1$ version of this theorem belongs to Constantine [14] (it is a particular case of Theorem 7) and it is found in numerous

places, including Muirhead [70, page 421]. The proof for the general case ($\beta > 0$, $a > (n-1)\beta/2$) is once again identical to the original one, and it is based on the formula for the average of a Jack polynomial of the β-Jacobi distribution (see Kadell [55]), and on the generalization of the β multivariate hypergeometric functions (for a reference, see [8]).

Theorem 10.2.1. *Let L be an $n \times n$ β-Laguerre matrix of parameter a. Then the c.d.f. of the largest eigenvalue l of L is given below:*

$$Pr[\ l \leq x\ ] = \frac{\Gamma_n(\frac{\beta}{2}(n-1)+1)}{\Gamma_n(a+\frac{\beta}{2}(n-1)+1)}\ \left(\frac{x}{2}\right)^{an}\ {}_1F_1^{\beta}(a; a+\frac{\beta}{2}(n-1)+1; -\frac{x}{2}I_{n-1})\ ,$$

where ${}_1F_1^{\beta}$ is the multivariate hypergeometric function of parameter β, and $I_{n-1} = (1, 1, \ldots, 1)$.

The second result of this section is a generalization of a result obtained by Silverstein [81].

Theorem 10.2.2. *Let L be an $n \times n$ β-Laguerre matrix of parameter a, and let n and a go to infinity in such a way that $(n\beta/2a) \to y \in [0,\ 1]$. Then the largest eigenvalue λ_{max} of L satisfies*

$$\lim_{n\to\infty} n\lambda_{max} = \beta(1+\sqrt{y})^2\ .$$

10.3 Condition number

In this section, we present a theorem whose $\beta = 1$ version belongs to Edelman [24, 26]; once again, the original proof applies with minor modifications and we choose not to present it.

Theorem 10.3.1. *Let L be a $n \times n$ β-Laguerre matrix of parameter a, and assume that $a - \frac{\beta}{2}(n-2) = 1$. Let $\kappa = \lambda_{\max}/\lambda_{\min}$ be the condition number of L, and let $f_{n,\beta}$ be the p.d.f. of the scaled condition number κ/n^2. Then for any x,*

$$\lim_{n\to\infty} f_{n,\beta}(x) = \frac{8\beta^{3/2-\beta/4}}{\Gamma(1-\frac{\beta}{2})}\ t^{-4+\beta/2}\ e^{-2/t^2}\ K_{-\beta/2-1}\left(\frac{2}{t}\right)\ ,$$

where K_ν, the modified Bessel function, is the unique solution of the equation

$$z^2 \frac{d^2 w}{dz^2} + z \frac{dw}{dz} - (z^2 + \nu^2) w = 0$$

with $K_\nu(z) \to 0$ as $z \to \infty$ and $|arg(z)| < \pi/2$.

This generalization, due to the condition that $\beta < 2$, fails to provide an answer to the question of finding the condition number *for any* β, when $a = n\beta/2$; this question remains open.

Index

Bibliography

[1] M. Abramowitz and I.A. Stegun, editors. *Handbook of Mathematical Functions*. Dover Publications, New York, 1970.

[2] Alan L. Andrew. Choosing test matrices for numerical software. In W.L. Hogarth and B.J. Nege, editors, *Computational Techniques and Applications Conference: CTAC-89*, pages 561–567, 1990.

[3] G.E. Andrews, David M. Jackson, and I.P. Goulden. Determinants of random matrices and Jack polynomials of rectangular shape. *Studies Applied Math.*, 2003. To appear.

[4] K. Aomoto. Jacobi polynomials associated with Selberg integrals. *SIAM J. Math. Anal.*, 18:545–549, 1987.

[5] Ludwig Arnold. On Wigner's semicircle law for the eigenvalues of random matrices. *Z. Wahrsch. Verw. Gebiete*, 19:191–198, 1971.

[6] Richard Askey and Donald Richards. Selberg's second beta integral and an integral of Mehta. In T.W. Anderson et al., editor, *Probability, Statistics, and Mathematics: Papers in Honor of Samuel Karlin*, pages 27–39. Academic Press, New York, 1989.

[7] Jinho Baik, Percy A. Deift, and Kurt Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.*, 12(4):1119–1178, 1999.

[8] T. Baker and Peter Forrester. The Calogero-Sutherland model and generalized classical polynomials. *Commun.Math.Phys.*, 188:175–216, 1997.

[9] Estelle L. Basor, Craig A. Tracy, and Harold Widom. Asymptotics of level-spacing distributions for random matrices. *Phys. Rev. Letters*, 69:5–8, 1992.

[10] R.J. Beerends and M.Opdam. Certain hypergeometric series related to the root system BC. *Trans. Amer. Math. Soc*, 339:581–609, 1993.

[11] P. Billingsley. *Probability and Measure*. John Wiley and Sons, New York, 1979.

[12] Alexei Borodin and Grigori Olshanski. Z-measures on partitions and their scaling limits. Technical Report math-ph/0210048.

[13] Yasuko Chikuse. Properties of Hermite and Laguerre polynomials in matrix argument and their applications. *Lin. Alg. Appl.*, 176:237–260, 1992.

[14] A.G. Constantine. Some noncentral distribution problems in multivariate analysis. *Ann. Math. Statist.*, 34:1270–1285, 1963.

[15] Percy Deift. *Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert Approach*. American Mathematical Society, Providence, 1998.

[16] R. Delannay and G. Le Caër. Distribution of the determinant of a random real-symmetric matrix from the Gaussian orthogonal ensemble. *Phys. Rev. E*, 62:1526–1536, 2000.

[17] James Demmel. The probability that a numerical analysis problem is difficult. *Math. of Comp.*, 50:449–480, 1988.

[18] James Demmel. *Applied Numerical Linear Algebra*. SIAM, Philadelphia, 1997.

[19] James Demmel and Plamen Koev. Efficient and accurate evaluation of Schur and Jack polynomials. 2003. Preprint.

[20] Persi Diaconis. *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, Hayward, California, 1998.

[21] Ioana Dumitriu and Alan Edelman. Matrix models for beta-ensembles. *J. Math. Phys.*, 43:5830–5847, 2002.

[22] Freeman J. Dyson. Statistical theory of energy levels of complex systems, i, ii, and iii. *J. Math. Phys.*, 3:140–156, 157–165, 166–175, 1962.

[23] Freeman J. Dyson. The threefold way. Algebraic structures of symmetry groups and ensembles in Quantum Mechanics. *J. Math. Phys.*, 3:1199–1215, 1963.

[24] Alan Edelman. *Eigenvalues and Condition Numbers of Random Matrices*. PhD thesis, Massachusetts Institute of Technology, 1989.

[25] Alan Edelman. The distribution and moments of the smallest eigenvalue of a random matrix of Wishart type. *Lin. Alg. Appl.*, 159:55–80, 1991.

[26] Alan Edelman. On the distribution of a scaled condition number. *Mathematics of Computation*, 58:185–190, 1992.

[27] Alan Edelman. The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law. *Journal of Multivariate Analysis*, 60:203–232, 1997.

[28] Alan Edelman. *Random Eigenvalues: Notes for Spring 2003, MIT Course*. 2003. Preprint.

[29] M.B. Eisen, P.T. Spellman, P.O Brown, and D. Botstein. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci.*, 95:14863–14868, 1998.

[30] Peter Forrester. *Random Matrices*. 2001. Preprint.

[31] H. O. Foulkes. A survey of some combinatorial aspects of symmetric functions. In *Permutations*. Gauthier-Villars, Paris, 1974.

[32] Alex Gamburd and Persi Diaconis. Magic squares, 2003. Preprint.

[33] Michel Gaudin. Sur la loi limite de l'éspacement des valeurs propres d'une matrice aléatoire. *Nucl. Phys.*, 25:447–458, 1961.

[34] Gene Golub and William Kahan. Calculating the singular values and pseudo-inverse of a matrix. *SIAM J. Num. Anal.*, 2:205–224, 1965.

[35] Gene Golub and Charles Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore and London, third edition, 1996.

[36] I. Goulden and David M. Jackson. Maps in locally orientable surfaces and integrals over real symmetric matrices. *Canadian J. Math.*, 49:865–882, 1997.

[37] J. Gravner, Craig A. Tracy, and Harold Widom. A growth model in a random environment. *Ann. Probab.*, 30:1340–1368, 2002.

[38] Ulf Grenander. *Probabilities on Algebraic Structures*. John Wiley and Sons, New York, 1963.

[39] Philip J. Hanlon, Richard P. Stanley, and John R. Stembridge. Some combinatorial aspects of the spectra of normally distributed random matrices. *Contemp. Math.*, 138:151–174, 1992.

[40] A. S. Householder. Unitary triangularization of a nonsymmetric matrix. *Journal of the Association of Computing Machinery*, 5:339–342, 1958.

[41] P.L. Hsu. On the distribution of the roots of certain determinantal questions. *Ann. Eugenics*, 9:250–258, 1939.

[42] L.K. Hua. Harmonic Analysis of functions of several complex variables in the classical domains. *Transl. Math. Monogr. Am. Math. Soc.*, 6, 1963.

[43] Dmitri A. Ivanov. Random-matrix ensembles in p-wave vortices. In *Proceedings of the Dresden workshop "Vortices in unconventional superconductors and superfluids"*. Springer, Heidelberg, 2002. [e-print cond-mat/0103089].

[44] Henry Jack. A class of symmetric polynomials with a parameter. *Proc. R. Soc. Edinburgh*, 69:1–18, 1970.

[45] David M. Jackson. Personal communication, April 2003.

[46] Alan T. James. The distribution of the latent roots of the covariance matrix. *Ann. Math. Stat.*, 31:151–158, 1960.

[47] Alan T. James. Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Stat.*, 35:475–501, 1964.

[48] Alan T. James. Calculation of the zonal polynomial coefficients by use of the Laplace Beltrami operator. *Ann. Math. Stat.*, 39:1711–1718, 1968.

[49] Alan T. James. Special functions of matrix and single argument in Statistics. In Richard A. Askey, editor, *Theory and Application of Special Functions*, pages 497–520. Academic Press, New York, 1975.

[50] Kurt Johansson. On random matrices from the compact classical groups. *Ann. of Math.*, 145(3):519–545, 1997.

[51] Kurt Johansson. On fluctuations of random hermitian matrices. *Duke Math. J.*, 91:151–203, 1998.

[52] Kurt Johansson. Shape fluctuations and random matrices. *Commun. Math. Phys.*, 209:437–476, 2000.

[53] Iain M. Johnstone. On the distribution of the largest principal component. *Ann. of Stat.*, 29(2):295–327, 2001.

[54] Dag Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate. Anal.*, 12:1–38, 1982.

[55] K. Kadell. The Selberg-Jack polynomials. *Advances in Mathematics*, 130:33–102, 1997.

[56] Joichi Kaneko. Selberg integrals and hypergeometric functions associated with Jack polynomials. *SIAM J. Math. Anal.*, 24:1086–1110, 1993.

[57] F. Knop and S. Sahi. A recursion and a combinatorial formula for the Jack polynomials. *Invent. Math.*, 128:9–22, 1997.

[58] P.R. Krishnaiah and T.C. Chang. On the exact distribution of the smallest root of the Wishart matrix using zonal polynomials. *Ann. I. Math. Stat.*, 23:293–295, 1971.

[59] M. Lasalle. Polynômes de hermite généralisés. *C.R. Acad. Sci. Paris, Séries I*, 313:579–582, 1991.

[60] M. Lasalle. Polynômes de jacobi généralisés. *C.R. Acad. Sci. Paris, Séries I*, 312:425–428, 1991.

[61] M. Lasalle. Polynômes de laguerre généralisés. *C.R. Acad. Sci. Paris, Séries I*, 312:725–728, 1991.

[62] I.G. Macdonald. Some conjectures for root systems. *SIAM J. Math. Anal.*, 13:998–1004, 1982.

[63] I.G. Macdonald. *Symmetric Functions and Hall Polynomials*. Oxford University Press Inc, New York, 1995.

[64] V.A. Marčenko and L.A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Math USSR Sbornik*, 1:457–483, 1967.

[65] Brendan D. McKay. The expected eigenvalue distribution of a large regular graph. *Lin. Alg. Appl.*, 40:203–216, 1981.

[66] Madan Lal Mehta. *Random Matrices*. Academic Press, Boston, second edition, 1991.

[67] Madan Lal Mehta and Michel Gaudin. On the density of eigenvalues of a random matrix. *Nucl. Phys.*, 4:420–427, 1960.

[68] Madan Lal Mehta and Jean-Marie Normand. Probability density of the determinant of a random hermitian matrix. *J. Phys. A*, 31:5377–5391, 1998.

[69] H. L. Montgomery. The pair correlation of zeros of the zeta function. In *Analytic Number Theory, Proceedings of the 1972 St. Louis Symposium*, pages 181–193. American Mathematical Society, 1973.

[70] Robb J. Muirhead. *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, New York, 1982.

[71] Robb J. Muirhead and Yasuko Chikuse. Asymptotic expansions for the joint and marginal distributions of the latent roots of the covariance matrix. *Ann. of Stat.*, 3:1011–1017, 1975.

[72] A.M. Odlyzko. On the distribution of spacings between zeros of the zeta function. In M. van Frankenhuysen and M. L. Lapidus, editors, *Dynamical, Spectral and Arithmetic Zeta-Functions, Contemporary Mathematics Series*, volume 290, pages 139–144. American Mathematical Society, Providence, Rhode Island, 2001.

[73] Andrei Okounkov. Proof of a conjecture of Goulden and Jackson. *Canadian J. Math.*, 49:883–886, 1997.

[74] Andrei Okounkov. Random matrices and random permutations. *Internat. Math. Res. Notices*, 20:1043–1095, 2000.

[75] Andrei Okounkov and Grigori Olshanski. Shifted Jack polynomials, binomial formula, and applications. *Mathematical Research Letters*, 4:69–78, 1997.

[76] Ingram Olkin. The 70th anniversary of the distribution of random matrices: A survey. *Lin. Alg. Appl.*, 354:231–243, 2002.

[77] Beresford N. Parlett. *The Symmetric Eigenvalue Problem*. SIAM Classics in Applied Mathematics, 1998.

[78] Gábor Szegö. *Orthogonal Polynomials*. American Mathematical Society, Providence, 1975. 4th edition.

[79] A. Selberg. Bemerkninger om et multiplet integral. *Norsk Matematisk Tiddskrift*, 26:71–78, 1944.

[80] Jack W. Silverstein. The smallest eigenvalue of a large dimensional Wishart matrix. *Ann. of Probab.*, 13:1364–1368, 1985.

[81] Jack W. Silverstein. On the weak limit of the largest eigenvalue of a large dimensional sample covariance matrix. *J. Multivariate Anal.*, 30:307–311, 1989.

[82] Jack W. Silverstein and Z.D. Bai. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *J. Multivariate Anal.*, 54(2):175–192, 1995.

[83] Jack W. Silverstein and Z.D. Bai. CLT of linear spectral statistics of large dimensional sample covariance matrices. preprint, 2003. Accepted for publication in *Annals. of Probab.*

[84] Steve Smale. On the efficiency of algorithms of analysis. *Bulletin of the American Math. Soc.*, 13:87–121, 1985.

[85] Alexander Soshnikov and Yakov Sinai. Central limit theorem for traces of large symmetric matrices with independent matrix elements. *Bol. Soc. Bras. Mat., Nova Sr.*, 29:1–24, 1998.

[86] Richard P. Stanley. Some combinatorial properties of Jack symmetric functions. *Adv. in Math.*, 77:76–115, 1989.

[87] Richard P. Stanley. *Enumerative Combinatorics, vol. 2.* Cambridge University Press, 1999.

[88] T. Sugiyama. On the distribution of the largest latent root of the covariance matrix. *Ann. Math. Stat.*, 38:1148–1151, 1967.

[89] Craig A. Tracy and Harold Widom. On orthogonal and symplectic matrix ensembles. *J. Stat. Phys.*, 92:809–835, 1996.

[90] Craig A. Tracy and Harold Widom. Universality of the distribution functions of random matrix theory. In *Statistical Physics on the Eve of the 21st Century: In*

Honour of J.B.McGuire on the Occasion of His 65th Birthday, pages 230–239. World Scientific Pub, 1999.

[91] Craig A. Tracy and Harold Widom. The distribution of the largest eigenvalue in the Gaussian ensembles. In *Calogero-Moser-Sutherland Models, CRM Series in Mathematical Physics*, volume 4, pages 461–472. Springer-Verlag, 2000.

[92] Hale F. Trotter. Eigenvalue distributions of large hermitian matrices; Wigner's semi-circle law and a theorem of Kac, Murdock, and Szegö. *Advances in Mathematics*, 54:67–82, 1984.

[93] P. Viswanath, D. Tse, and V. Anantharam. Asymptotically optimal waterfilling in vector multiple access channels. *IEEE Transactions on Information Theory*, 47(1):241–267, 2001.

[94] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math.*, 62:548–564, 1955.

[95] Eugene P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math.*, 67:325–327, 1958.

[96] J. Wishart. The generalized product moment distribution in samples from a normal multivariate population. *Biometrika A*, 20:32–43, 1928.

[97] Z. Yan. Generalized hypergeometric functions and Laguerre polynomials in two variables. *Contemp. Math.*, 138:239–259, 1992.

[98] L. Zheng and D. Tse. Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels. *IEEE Trans. Inf. Th.*, 2002.

[99] M. Zirnbauer. Riemannian symmetric superspaces and their origin in random matrix theory. *J. Math. Phys.*, 37:4986–5018, 1996.