Acta Numerica (2005), pp. 233–297
DOI: 10.1017/S0962492904000236

Random matrix theory

Alan Edelman
Department of Mathematics,
Massachusetts Institute of Technology,
Cambridge, MA 02139, USA
E-mail: edelman@math.mit.edu

N. Raj Rao
Department of Electrical Engineering and Computer Science,
Massachusetts Institute of Technology,
Cambridge, MA 02139, USA
E-mail: raj@mit.edu

Random matrix theory is now a big subject with applications in many disciplines of science, engineering and finance. This article is a survey specifically oriented towards the needs and interests of a numerical analyst. This survey includes some original material not found anywhere else. We include the important mathematics which is a very modern development, as well as the computational software that is transforming the theory into useful practice.

CONTENTS

1 Introduction 234
2 Linear systems 234
3 Matrix calculus 235
4 Classical random matrix ensembles 243
5 Numerical algorithms stochastically 254
6 Classical orthogonal polynomials 257
7 Multivariate orthogonal polynomials 262
8 Hypergeometric functions of matrix argument 264
9 Painlevé equations 265
10 Eigenvalues of a billion by billion matrix 275
11 Stochastic operators 278
12 Free probability and infinite random matrices 283
13 A random matrix calculator 285
14 Non-Hermitian and structured random matrices 288
15 A segue 290
References 291

1. Introduction

Texts on 'numerical methods' teach the computation of solutions to non-random equations. Typically we see integration, differential equations, and linear algebra among the topics. We find 'random' there too, but only in the context of random number generation.

The modern world has taught us to study stochastic problems. Already many articles exist on stochastic differential equations. This article covers topics in stochastic linear algebra (and operators). Here, the equations themselves are random. Like the physics student who has mastered the lectures and now must face the sources of randomness in the laboratory, numerical analysis is heading in this direction as well. The irony to newcomers is that often randomness imposes more structure, not less.

2. Linear systems

The limitations on solving large systems of equations are computer memory and speed. The speed of computation, however, is not only measured by clocking hardware; it also depends on numerical stability, and for iterative methods, on convergence rates. At this time, the fastest supercomputer performs Gaussian elimination, *i.e.*, solves $Ax = b$ on an n by n matrix A for $n \approx 10^6$. We can easily imagine $n \approx 10^9$ on the horizon. The standard benchmark HPL ('high-performance LINPACK') chooses A to be a random matrix with elements from a uniform distribution on $[-1/2, 1/2]$. For such large n, a question to ask would be whether a double precision computation would give a single precision answer.

Turning back the clock, in 1946 von Neumann and his associates saw $n = 100$ as the large number on the horizon. How do we pick a good test matrix A? This is where von Neumann and his colleagues first introduced the assumption of random test matrices distributed with elements from independent normals. Any analysis of this problem necessarily begins with an attempt to characterize the condition number $\kappa = \sigma_1/\sigma_n$ of the $n \times n$ matrix A. They give various 'rules of thumb' for κ when the matrices are so distributed. Sometimes these estimates are referred to as an expectation and sometimes as a bound that holds with high, though unspecified, probability. It is interesting to compare their 'rules of thumb' with what we now know about the condition numbers of such random matrices as $n \to \infty$ from Edelman (1989).

Quote. *For a 'random matrix' of order n the expectation value has been shown to be about n.*
(von Neumann 1963, p. 14)

Fact. $E[\kappa] = \infty$.

Quote. *... we choose two different values of* κ*, namely* n *and* $\sqrt{10}n$*.*
(von Neumann 1963, p. 477)

Fact. $\Pr(\kappa < n) \approx 0.02, \Pr(\kappa < \sqrt{10}\, n) \approx 0.44$.

Quote. *With probability* $\approx$ *1,* $\kappa < 10n$
(von Neumann and Goldstine 1947, p. 555)

Fact. $\Pr(\kappa < 10\, n) \approx 0.80$.

Results on the condition number have been extended recently by Edelman and Sutton (2004), and Azaïs and Wschebor (2004). Related results include the work of Viswanath and Trefethen (1998).

Analysis of Gaussian elimination of random matrices[1] began with the work of Trefethen and Schreiber (1990), and later Yeung and Chan (1997). Of specific interest is the behaviour of the 'growth factor' which influences numerical accuracy. More recently, Sankar, Spielman and Teng (2004) analysed the performance of Gaussian elimination using smoothed analysis, whose basic premise is that bad problems tend to be more like isolated spikes. Additional details can be found in Sankar (2003).

Algorithmic developers in need of guinea pigs nearly always take random matrices with standard normal entries, or perhaps close cousins, such as the uniform distribution of $[-1, 1]$. The choice is highly reasonable: these matrices are generated effortlessly and might very well catch programming errors. But are they really 'test matrices' in the sense that they can catch every type of error? It really depends on what is being tested; random matrices are not as random as the name might lead one to believe. Our suggestion to library testers is to include a carefully chosen range of matrices rather than rely on randomness. When using random matrices as test matrices, it can be of value to know the theory.

We want to convey is that random matrices are very *special* matrices. It is a mistake to link psychologically a random matrix with the intuitive notion of a 'typical' matrix or the vague concept of 'any old matrix'. In fact, the larger the size of the matrix the more predictable it becomes. This is partly because of the central limit theorem.

3. Matrix calculus

We have checked a few references on 'matrix calculus', yet somehow none were quite right for a numerical audience. Our motivation is twofold. Firstly, we believe that condition number information has not traditionally been widely available for matrix-to-matrix functions. Secondly, matrix

[1] On a personal note, the first author started down the path of random matrices because his adviser was studying Gaussian elimination on random matrices.

calculus allows us to compute Jacobians of familiar matrix functions and transformations.

Let $x \in \mathbb{R}^n$ and $y = f(x) \in \mathbb{R}^n$ be a differentiable vector-valued function of x. In this case, it is well known that the Jacobian matrix

$$J = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial f_n}{\partial x_1} & \cdots & \dfrac{\partial f_n}{\partial x_n} \end{pmatrix} = \left(\frac{\partial f_i}{\partial x_j} \right)_{i,j=1,2,\ldots,n} \tag{3.1}$$

evaluated at a point x approximates $f(x)$ by a linear function. Intuitively $f(x + \delta x) \approx f(x) + J\delta x$, *i.e.*, J is the matrix that allows us to invoke first-order perturbation theory. The function f may be viewed as performing a change of variables. Often the matrix J is denoted $\mathrm{d}f$ and 'Jacobian' refers to $\det J$. In the complex case, the Jacobian matrix is real $2n \times 2n$ in the natural way.

3.1. Condition numbers of matrix transformations

A matrix function/transformation (with no breakdown) can be viewed as a local linear change of variables. Let f be a (differentiable) function defined in the neighbourhood of a (square or rectangular) matrix A.

We think of functions such as $f(A) = A^3$ or $f(A) = \mathrm{lu}(A)$, the LU factorization, or even the SVD or QR factorizations. The linearization of f is $\mathrm{d}f$ which (like Kronecker products) is a linear operator on matrix space. For general A, $\mathrm{d}f$ is $n^2 \times n^2$, but it is rarely helpful to write $\mathrm{d}f$ explicitly in this form.

We recall the notion of condition number which we put into the context of matrix functions and factorizations. The condition number for $A \to f(A)$ is defined as

$$\begin{aligned} \kappa &= \frac{\text{relative change in } f(A)}{\text{relative change in } A} \\ &= \lim_{\epsilon \to 0} \sup_{\|E\| = \epsilon} \frac{\|f(A+E) - f(A)\| / \|f(A)\|}{\|E\| / \|A\|} \\ &= \|\mathrm{d}f\| \left(\frac{\|A\|}{\|f(A)\|} \right). \end{aligned}$$

Figure 3.1 illustrates the condition number to show that the key factor in the two-norm κ is related to the largest axis of an ellipsoid in the matrix factorization space, *i.e.*, the largest singular value of $\mathrm{d}f$. The product of the semi-axis lengths is related to the volume of the ellipsoid and is the Jacobian determinant of f.

In summary

$$\kappa = \sigma_{\max}(\mathrm{d}f)\,\frac{\|A\|}{\|f(A)\|}, \tag{3.2}$$

$$J = \prod_i \sigma_i(\mathrm{d}f) = \det(\mathrm{d}f). \tag{3.3}$$

Example 1. Let $f(A) = A^2$ so that $\mathrm{d}f(E) = AE + EA$. This can be rewritten in terms of the Kronecker (or tensor) product operator $\otimes$ as $\mathrm{d}f = I \otimes A + A^T \otimes I$. Therefore

$$\kappa = \sigma_{\max}(I \otimes A + A^T \otimes I)\frac{\|A\|}{\|A^2\|}.$$

Recall that $A \otimes B : X \to BXA^T$ is the linear map from X to BXA^T. The Kronecker product has many wonderful properties, as described in the article by Van Loan (2000).

Example 2. Let $f(A) = A^{-1}$, so that $\mathrm{d}f(E) = -A^{-1}EA^{-1}$, or in terms of the Kronecker product operator as $\mathrm{d}f = -A^{-T} \otimes A^{-1}$.

This implies that the singular values of $\mathrm{d}f$ are $(\sigma_i(A)\sigma_j(A))^{-1}$, for $1 \le i, j \le n$.

The largest singular value $\sigma_{\max}(\mathrm{d}f)$ is thus equal to $1/\sigma_n(A)^2 = \|A^{-1}\|^2$ so that κ as defined in (3.2) is simply the familiar matrix condition number

$$\kappa = \|A\|\,\|A^{-1}\| = \frac{\sigma_1}{\sigma_n},$$

while in contrast, the Jacobian given by (3.3) is

$$\text{Jacobian} = \prod_{i,j} \frac{1}{\sigma_i(A)\sigma_j(A)} = (\det A)^{-2n}.$$

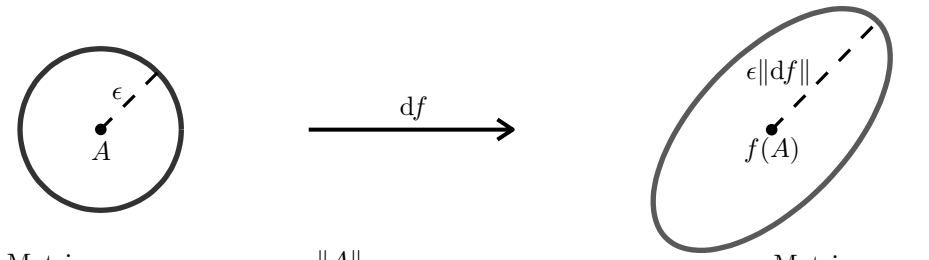

Figure 3.1. The condition number of a matrix factorization is related to the largest axis of an ellipsoid in matrix factorization space.

Without dwelling on the point, κ has a worst case built into the 'lim sup', while J contains information about an average behaviour under perturbations.

3.2. Matrix Jacobians numerically computed with finite differences

Consider the symmetric eigenvalue decomposition $A = Q\Lambda Q'$, where A is an $n \times n$ symmetric matrix. The Jacobian for this factorization is the term $\prod_{i<j} |\lambda_i - \lambda_j|$ in

$$(\mathrm{d}A) = \prod_{i<j} |\lambda_i - \lambda_j| \; (\mathrm{d}\Lambda) \; (Q' \, \mathrm{d}Q). \tag{3.4}$$

This equation is derived from the first-order perturbation $\mathrm{d}A$ in A due to perturbations $\mathrm{d}\Lambda$ and $Q' \, \mathrm{d}Q$ in the eigenvalues Λ and the eigenvectors Q. Note that since Q is orthogonal, $Q'Q = I$ so that $Q' \, \mathrm{d}Q + \mathrm{d}Q'Q = 0$ or that $Q' \, \mathrm{d}Q$ is antisymmetric with zeros along the diagonal. Restricting $Q' \, \mathrm{d}Q$ to be antisymmetric ensures that $A + \mathrm{d}A$ remains symmetric.

Numerically, we compute the perturbations in Λ and Q due to perturbations in A. As numerical analysts we always think of A as the input and Q and Λ as output, so it is natural to ask for the answer in this direction. Assuming the eigenvalue decomposition is unique after fixing the phase of the columns of Q, the first-order perturbation in Λ and Q due to perturbations in A is given by

$$\frac{(\mathrm{d}\Lambda)(Q' \, \mathrm{d}Q)}{(\mathrm{d}A)} = \frac{1}{\prod_{i<j} |\lambda_i - \lambda_j|} = \frac{1}{\Delta(\Lambda)}, \tag{3.5}$$

where $\Delta(\lambda) = \prod_{i<j} |\lambda_i - \lambda_j|$ is the absolute value of the Vandermonde determinant.

We can create an $n \times n$ symmetric matrix A by, for example, creating an $n \times n$ matrix X with independent Gaussian entries and then symmetrizing it as $A = (X + X')/n$. This can be conveniently done in MATLAB as

```
n=15;
X=randn(n);
A=(X+X')/n;
```

The fact that X is a random matrix is incidental, *i.e.*, we do not exploit the fact that it is a random matrix.

We can compute the decomposition $A = Q\Lambda Q'$ in MATLAB as

```
[Q,L]=eig(A);
L=diag(L);
```

Since A is an $n \times n$ symmetric matrix, the Jacobian matrix as in (3.1) resides in an $(n(n+1)/2)^2$-dimensional space:

```
JacMatrix=zeros(n*(n+1)/2);
```

Table 3.1. Jacobians computed numerically with finite differences.

```
n=15;                              % Size of the matrix
X=randn(n);
A=(X+X')/n;                        % Generate a symmetric matrix
[Q,L]=eig(A);                      % Compute its eigenvalues/eigenvectors
L=diag(L);
JacMatrix=zeros(n*(n+1)/2);        % Initialize Jacobian matrix
epsilon=1e-7; idx=1;
mask=triu(ones(n),1); mask=logical(mask(:)); % Upper triangular mask
for i=1:n
    for j=i:n

        %%% Perturbation Matrix
        Eij=zeros(n);              % Initialize perturbation
        Eij(i,j)=1; Eij(j,i) = 1; % Perturbation matrix
        Ap=A+epsilon*Eij;          % Perturbed matrix

        %%% Eigenvalues and Eigenvectors

        [Qp,Lp] = eig(Ap);
        dL= (diag(Lp)-L)/epsilon; % Eigenvalue perturbation
        QdQ = Q'*(Qp-Q)/epsilon;  % Eigenvector perturbation

        %%% The Jacobian Matrix
        JacMatrix(1:n,idx)=dL;                 % Eigenvalue part of Jacobian
        JacMatrix((n+1):end,idx) = QdQ(mask); % Eigenvector part of Jacobian
        idx=idx+1;                             % Increment column counter
    end
end
```

Let ϵ be any small positive number, such as

```
epsilon=1e-7;
```

Generate the symmetric perturbation matrix E_{ij} for $1 \le i \le j < n$ whose entries are equal to zero except in the (i, j) and (j, i) entries, where they are equal to 1. Construct the Jacobian matrix by computing the eigenvalues and eigenvectors of the perturbed matrix $A + \epsilon\, E_{ij}$, and the quantities $\mathrm{d}\Lambda$ and $Q'\,\mathrm{d}Q$. This can be done in MATLAB using the code in Table 3.1. We note that this simple forward difference scheme can be replaced by a central difference scheme for added accuracy.

We can compare the numerical answer obtained by taking the determinant of the Jacobian matrix with the theoretical answer expressed in terms of the Vandermonde determinant as in (3.5). For a particular choice of A we can

run the MATLAB code in Table 3.1 to get the answer:

```
format long
disp([abs(det(JacMatrix))  1/abs(det(vander(L)))]);
>> ans = 1.0e+49 *
         3.32069877679394   3.32069639128242
```

This is, in short, the 'proof by MATLAB' to show how Jacobians can be computed numerically with finite differences.

3.3. Jacobians of matrix factorizations

The computations of matrix Jacobians can be significantly more complicated than the scalar derivatives familiar in elementary calculus. Many Jacobians have been rediscovered in various communities. We recommend Olkin (1953, 2002), and the books by Muirhead (1982) and Mathai (1997). When computing Jacobians of matrix transformations or factorizations, it is important to identify the dimension of the underlying space occupied by the matrix perturbations.

Wedge products and the accompanying notation are used to facilitate the computation of matrix Jacobians. The notation also comes in handy for expressing the concept of volume on curved surfaces as in differential geometry. Mathai (1997) and Muirhead (1982) are excellent references for readers who truly wish to understand wedge products as a tool for computing the Jacobians of commonly used matrix factorizations such as those listed below.

While we expect our readers to be familiar with real and complex matrices, it is reasonable to consider quaternion matrices as well. The parameter β has been traditionally used to count the dimension of the underlying algebra as in Table 3.2. In other branches of mathematics, the parameter $\alpha = 2/\beta$ is used.

We provide, without proof, the formulas containing the Jacobians of familiar matrix factorizations. We encourage readers to notice that the vanishing

Table 3.2. Notation used to denote whether the elements of a matrix are real, complex or quaternion ($\beta = 2/\alpha$).

β	α	Division algebra
1	2	real ($\mathbb{R}$)
2	1	complex ($\mathbb{C}$)
4	1/2	quaternion ($\mathbb{H}$)

of the Jacobian is connected to difficult numerical problems. The parameter count is only valid where the Jacobian does not vanish.

QR (Gram–Schmidt) 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 $n \times n$. The parameter count for the orthogonal matrix is the dimension of the Stiefel manifold $V_{m,n}$.

Parameter count:

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

Jacobian:

$$(\mathrm{d}A) = \prod_{i=1}^{n} r_{ii}^{\beta(m-i+1)-1}\ (\mathrm{d}R)\ (Q'\,\mathrm{d}Q). \tag{3.6}$$

Notation: $(\mathrm{d}A)$, $(\mathrm{d}R)$ are volumes of little boxes around A and R, while $(Q'\,\mathrm{d}Q)$ denotes the volume of a little box around the strictly upper triangular part of the antisymmetric matrix $Q'\,\mathrm{d}Q$ (see a numerical illustration in Section 3.2).

LU (Gaussian elimination) decomposition ($A = LU$). Valid for all three cases ($\beta = 1, 2, 4$). All matrices are $n \times n$, L and U are lower and upper triangular respectively, $l_{ii} = 1$ for all $1 \leq i \leq n$. Assume there is no pivoting.

Parameter count:

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

Jacobian:

$$(\mathrm{d}A) = \prod_{i=1}^{n} |u_{ii}|^{\beta(n-i)}\ (\mathrm{d}L)\ (\mathrm{d}U). \tag{3.7}$$

$\mathbf{Q\Lambda Q'}$ (symmetric eigenvalue) decomposition ($A = Q\Lambda Q'$). Valid for all three cases ($\beta = 1, 2, 4$). Here 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:

$$(\mathrm{d}A) = \prod_{i<j} (\lambda_i - \lambda_j)^{\beta}\ (\mathrm{d}\Lambda)\ (Q'\,\mathrm{d}Q). \tag{3.8}$$

UΣV′ (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 (removing $(\beta - 1)n$ 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:

$$(\mathrm{d}A) = \prod_{i<j}(\sigma_i^2 - \sigma_j^2)^\beta \prod_{i=1}^{n} \sigma_i^{\beta(m-n+1)-1} \, (U'\,\mathrm{d}U) \, (\mathrm{d}\Sigma) \, (V'\,\mathrm{d}V). \tag{3.9}$$

References: real, Muirhead (1982), Dumitriu (2003), Shen (2001).

CS (Cosine–sine) decomposition. 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,

$$\begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} V_{11}' & V_{12}' \\ V_{21}' & V_{22}' \end{pmatrix}$$

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

This parameter count is a little special since we have to account for the choice of the cases in the decomposition.

Parameter count:

$$\beta\frac{n(n+1)}{2} - n = \big(\beta j(j+1) - (\beta - 1)j\big) + j + \left(\beta k(k+1) - k - \beta\frac{p(p+1)}{2} + p\right).$$

Jacobian:

$$(Q'\,\mathrm{d}Q) = \prod_{i<j} \sin(\theta_i - \theta_j)^\beta \, \sin(\theta_i + \theta_j)^\beta \prod_{i=1}^{j} \cos(\theta_i)^{\beta-1} \, \sin(\theta_i)\,\mathrm{d}\theta \times (U_1'\,\mathrm{d}U_1)\,(U_2'\,\mathrm{d}U_2)\,(V_1'\,\mathrm{d}V_1)\,(V_2'\,\mathrm{d}V_2).$$

Tridiagonal QΛQ′ (eigenvalue) decomposition ($T = Q\Lambda Q'$). Valid for real matrices. T is an $n \times n$ tridiagonal symmetric matrix, 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 they 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:

$$(\mathrm{d}T) = \frac{\prod_{i=1}^{n-1} T_{i+1,i}}{\prod_{i=1}^{n} q_i}\ \mu(\mathrm{d}q)\ (\mathrm{d}\Lambda). \tag{3.10}$$

Note that the Jacobian is written as a combination of parameters from T and q, the first row of Q, and $\mu(\mathrm{d}q)$ is the surface area on the sphere.

Tridiagonal BB′ (Cholesky) decomposition ($T = BB'$). Valid for real matrices. T is an $n \times n$ real positive definite tridiagonal matrix, B is an $n \times n$ real bidiagonal matrix.

Parameter count:

$$2n - 1 = 2n - 1.$$

Jacobian:

$$\mathrm{d}T = 2^n b_{11} \prod_{i=2}^{n} b_{ii}^2\ (\mathrm{d}B). \tag{3.11}$$

4. Classical random matrix ensembles

We now turn to some of the most well-studied random matrices. They have names such as Gaussian, Wishart, MANOVA, and circular. We prefer Hermite, Laguerre, Jacobi, and perhaps Fourier. In a sense, they are to random matrix theory as Poisson's equation is to numerical methods. Of course, we are thinking in the sense of the problems that are well tested, well analysed, and well studied because of nice fundamental analytic properties.

These matrices play a prominent role because of their deep mathematical structure. They have arisen in a number of fields, often independently. The tables that follow are all keyed by the first column to the titles Hermite, Laguerre, Jacobi, and Fourier.

4.1. Classical ensembles by discipline

We connect classical random matrices to problems roughly by discipline. In each table, we list the 'buzz words' for the problems in the field; where a

Table 4.1. Matrix factorizations associated with the classical random matrix ensembles.

Ensemble	Numerical procedure	MATLAB
Hermite	symmetric eigenvalue decomposition	`eig`
Laguerre	singular value decomposition	`svd`
Jacobi	generalized singular value decomposition	`gsvd`
Fourier	unitary eigenvalue decomposition	`eig`

Table 4.2. Equilibrium measure for classical random matrix ensembles.

Ensemble	Weight function	Equilibrium measure
Hermite	$\mathrm{e}^{-x^2/2}$	semi-circular law (Wigner 1958)
Laguerre	$x^a\mathrm{e}^{-x}$	Marčenko and Pastur (1967)
Jacobi	$(1-x)^a(1+x)^b$	generalized McKay law
Fourier	$\mathrm{e}^{j\theta}$	uniform

classical random matrix has not yet appeared or, as we would rather believe, is yet to be discovered, we indicate with a blank. The actual definition of the random matrices may be found in Section 4.3. Note that for every problem there is an option of considering random matrices over the reals ($\beta = 1$), complexes ($\beta = 2$), quaternions ($\beta = 4$), or there is the general β approach.

We hope the reader will begin to see a fascinating story already emerging in Tables 4.1 and 4.2, where the symmetric eigenvalue problem is connected to the Hermite weight factor e^{-x^2}, the SVD to the Laguerre weight factor $x^a\mathrm{e}^{-x}$ and so forth.

In multivariate statistics, the problems of interest are random covariance matrices (known as Wishart matrices) and ratios of Wishart matrices that

Table 4.3. Multivariate statistics.

Ensemble	Problem solved	Univariate distribution
Hermite	–	normal
Laguerre	Wishart	chi-squared
Jacobi	MANOVA	beta
Fourier	–	–

Table 4.4. Graph theory.

Ensemble	Type of graph	Author
Hermite	undirected	Wigner (1955)
Laguerre	bipartite	Jonsson (1982)
Jacobi	d-regular	McKay (1981)
Fourier	–	–

Table 4.5. Free probability and operator algebras.

Ensemble	Terminology
Hermite	semi-circle
Laguerre	free Poisson
Jacobi	free product of projections
Fourier	–

arise in the multivariate analysis of variance (MANOVA). This is a central theme of texts such as Muirhead (1982).

The same matrices also arise elsewhere, especially in the modern physics of super-symmetry. This is evident in the works of Zirnbauer (1996), Ivanov (2002) and Caselle and Magnea (2004). More classically Dyson and Wigner worked on the Hermite and Fourier cases, known, respectively, as the Gaussian and circular ensembles. (See Mehta (1991).)

The cases that correspond to symmetric spaces are quantized perhaps unnecessarily. In mathematics, a symmetric space is a geometric object such as a sphere that can be identified as the quotient space of two Lie groups, with an involution that preserves geodesics. The Grassmann manifold is a symmetric space, while the Stiefel manifold of $m \times n$ orthogonal matrices is not, unless $m = 1$ or $m = n$, *i.e.*, the sphere and the orthogonal group respectively.

Many of the classical techniques for computing the eigenvalue distributions are ultimately related to interconnectivity of the matrix. For each case Table 4.4 shows a graph structure underlying the matrix.

'Free probability' is an important branch of operator algebra developed in 1985 by Voiculescu that has deep connections to random matrix theory. Table 4.5 uses the names found in that literature. From the random matrix viewpoint, free probability describes the eigenvalues of such operations as $A+B$ or AB in a language similar to that of the distribution of independent

random variables $a + b$ or ab, respectively. There will be more on this in Section 12.

The authors would be delighted if the reader is awed by the above set of tables. Anything that manifests itself in so many ways in so many fields must be deep in the foundations of the problem. We indicate the four channels of structure lurking underneath computation (Table 4.1), multivariate statistics (Table 4.3), graph theory (Table 4.4) and operator algebras (Table 4.5).

There is a deep structure begging the dense matrix expert to forget the SVD for a moment, or the sparse expert to forget bipartite graphs, if only briefly, or the statistician to forget the chi-squared distribution and sample covariance matrices. Something ties these experts together. Probably random matrix theory is not the only way to reveal the hidden message, but it is the theory that has compelled us to see what is truly there.

A few words for the numerical analyst. The symmetric and unitary eigenvalue problems, the SVD, and the GSVD have important mathematical roles because of certain symmetries not enjoyed by LU or the asymmetric eigenvalue problem. More can be said, but this may not be the place. We plant the seed and we hope it will be picked up by many.

In the remainder of this chapter we will explore these random matrix ensembles in depth. We begin with the basic Gaussian matrices and briefly consider the joint element density and invariance properties. We then construct the classical ensembles, derive their joint element densities, and their joint eigenvalue densities, all in the context of the natural numerical procedures listed in Table 4.1.

4.2. Gaussian random matrices

$G_1(m, n)$ is an $m \times n$ matrix of independent and identically distributed (i.i.d.) standard real random normals. More simply, in MATLAB notation:

```
G1=randn(m,n);
```

Table 4.6 lists MATLAB commands that can be used to generate $G_\beta(m, n)$ for general β. Note that since quaternions do not exist in MATLAB they are 'faked' using 2×2 complex matrices.

If A is an $m \times n$ Gaussian random matrix $G_\beta(m, n)$ then its joint element density is given by

$$\frac{1}{(2\pi)^{\beta mn/2}} \exp\left(-\frac{1}{2}\|A\|_F^2\right). \tag{4.1}$$

Some authors also use the notation $\mathrm{etr}(A)$ for the exponential of the trace of a matrix.

The most important property of G_β, be it real, complex, or quaternion, is its *orthogonal invariance*. This makes the distribution impervious to multiplication by an orthogonal (unitary, symplectic) matrix, provided that

Table 4.6. Generating the Gaussian random matrix $G_\beta(m,n)$ in MATLAB.

β	MATLAB command
1	`G = randn(m,n)`
2	`G = randn(m,n) + i*randn(m,n)`
4	`X = randn(m,n) + i*randn(m,n); Y = randn(m,n) + i*randn(m,n);` `G = [X Y; - conj(Y) conj(X)]`

the two are independent. This can be inferred from the joint element density in (4.1) since its Frobenius norm, $\|A\|_F$, is unchanged when A is multiplied by an orthogonal (unitary, symplectic) matrix. The orthogonal invariance implies that no test can be devised that would differentiate between Q_1A, A, and AQ_2, where Q_1 and Q_2 are non-random orthogonal and A is Gaussian.

4.3. Construction of the classical random matrix ensembles

The classical ensembles are constructed from G_β as follows. Since they are constructed from multivariate Gaussians, they inherit the orthogonality property as well, *i.e.*, they remain invariant under orthogonal transformations.

Gaussian orthogonal ensemble (GOE): symmetric $n \times n$ matrix obtained as $(A + A^T)/2$ where A is $G_1(n,n)$. 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 $n \times n$ matrix obtained as $(A + A^H)/2$, where A is $G_2(n,n)$ 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 $n \times n$ matrix obtained as $(A + A^D)/2$, where A is $G_4(n,n)$ 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 $N_4(0, \frac{1}{2})$.

Similarly, the Wishart and MANOVA ensembles can be defined as follows.

Wishart ensemble ($W_\beta(m,n)$, $m \geq n$): symmetric/Hermitian/self-dual $n \times n$ matrix which can be obtained as $A'A$, where A is $G_\beta(m,n)$ and A' denotes A^T, A^H and A^D, depending on whether A is real, complex, or quaternion, respectively.

MANOVA ensemble ($J_\beta(m_1, m_2, n)$, $m_1, m_2 \geq n$): symmetric/Hermitian/self-dual $n \times n$ matrix which can be obtained as $A/(A+B)$, where A and B are $W_\beta(m_1, n)$ and $W_\beta(m_2, n)$, respectively. See Sutton (2005) for details on a construction using the CS decomposition.

Circular ensembles: constructed as $U^T U$ and U for $\beta = 1, 2$ respectively, where U is a uniformly distributed unitary matrix (see Section 4.6). For $\beta = 4$, it is defined analogously as in Mehta (1991).

The β-Gaussian ensembles arise in physics, and were first identified by Dyson (1963) by the group over which they are invariant: Gaussian orthogonal or, for short, GOE (with real entries, $\beta = 1$), Gaussian unitary or GUE (with complex entries, $\beta = 2$), and Gaussian symplectic or GSE (with quaternion entries $\beta = 4$).

The Wishart ensembles owe their name to Wishart (1928), who studied them in the context of statistics applications as sample covariance matrices. The β-Wishart models for $\beta = 1, 2, 4$ could be named Wishart real, Wishart complex, and Wishart quaternion respectively, though the β notation is not as prevalent in the statistical community.

The MANOVA ensembles arise in statistics in the Multivariate Analysis of Variance, hence the name. They are in general more complicated to characterize, so less is known about them than the Gaussian and Wishart ensembles.

4.4. Computing the joint element densities

The joint eigenvalue densities of the classical random matrix ensembles have been computed in many different ways by different authors. Invariably, the basic prescription is as follows.

We begin with the probability distribution on the matrix elements. The next step is to pick an appropriate matrix factorization whose Jacobians are used to derive the joint densities of the elements in the matrix factorization space. The relevant variables in this joint density are then appropriately transformed and 'integrated out' to yield the joint eigenvalue densities.

This prescription is easy enough to describe, though in practice the normal distribution seems to be the best choice to allow us to continue and get analytical expressions. Almost any other distribution would stop us in our tracks, at least if our goal is some kind of exact formula.

Example. Let A be an $n \times n$ matrix from the Gaussian orthogonal ensemble ($\beta = 1$). As described earlier, this is an $n \times n$ random matrix with elements distributed as $N(0,1)$ on the diagonal and $N(0, 1/2)$ off the diagonal, that is,

$$a_{ij} \sim \begin{cases} N(0,1) & i = j, \\ N(0,1/2) & i > j. \end{cases}$$

Table 4.7. Joint element densities of an $n \times n$ matrix A from a Gaussian ensemble.

Gaussian	orthogonal	$\beta = 1$	$\frac{1}{2^{n/2}} \; \frac{1}{\pi^{n/2+n(n-1)\beta/4}} \; \exp\Big(-\frac{1}{2}\|A\|_F^2\Big)$
	unitary	$\beta = 2$	
	symplectic	$\beta = 4$	

Recall that the normal distribution with mean μ and variance σ^2, *i.e.*, $N(\mu, \sigma^2)$, has a density given by

$$\frac{1}{\sqrt{2\pi\,\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

from which it is fairly straightforward to verify that the joint element density of A written as

$$\frac{1}{2^{n/2}} \frac{1}{\pi^{n(n+1)/4}} \exp\big(-\|A\|_F^2/2\big) \tag{4.2}$$

can be obtained by taking products of the n normals along the diagonal having density $N(0,1)$ and $n(n-1)/2$ normals in the off-diagonals having density $N(0,1/2)$.

Table 4.7 lists the joint element density for the three Gaussian ensembles parametrized by β.

Now that we have obtained the joint element densities, we simply have to follow the prescription discussed earlier to obtain the joint eigenvalue densities.

In the case of the Gaussian ensembles, the matrix factorization $A = Q\Lambda Q'$ directly yields the eigenvalues and the eigenvectors. Hence, applying the Jacobian for this transformation given by (3.8) allows us to derive the joint densities for the eigenvalues and the eigenvectors of A. We obtain the joint eigenvalue densities by 'integrating' out the eigenvectors.

We like to think of the notion of the 'most natural' matrix factorization that allows us to compute the joint eigenvalue densities in the easiest manner. For the Gaussian ensembles, the symmetric eigenvalue decomposition $A = Q\Lambda Q'$ is the most obvious choice. This not the case for the Wishart and the MANOVA ensembles. In this context, what makes a matrix factorization 'natural'? Allow us to elaborate.

Consider the Wishart matrix ensemble $W_\beta(m,n) = A'A$, where $A = G_\beta(m,n)$ is a multivariate Gaussian. Its joint element density can be computed rather laboriously in a two-step manner whose first step involves writing $W = QR$ and then integrating out the Q, leaving the R. The second step is the transformation $W = R'R$ which is the Cholesky factorization of a matrix in numerical analysis. The conclusion is that although we may obtain the joint density of the elements of W as listed in Table 4.8, this procedure is much more involved than it needs to be.

Table 4.8. Joint element density of the Wishart ensemble $W_\beta(m,n)$ $(m \geq n)$.

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

This is where the notion of a 'natural' matrix factorization comes in. Though it seems statistically obvious to think of Wishart matrices as covariance matrices and compute the joint density of the eigenvalues of $A'A$ directly, it is more natural to derive the joint density of the singular values of A instead. Since A is a multivariate Gaussian, the Jacobian of the factorization $A = U\Sigma V'$ given by (3.9) can be used to directly determine the joint density of the singular values and the singular vectors of W from the joint element density of A in (4.1). We can then integrate out the singular vectors to obtain the joint density of the singular values of A and hence the eigenvalues of $W = A'A$. The technicalities of this may be found in Edelman (1989).

Similarly, the corresponding 'natural' factorization for the MANOVA ensembles is the generalized singular value decomposition. Note that the square of the generalized singular values of two matrices A and B is the same as the eigenvalues of $(BB')^{-1}(AA')$, so that the eigenvalues of the MANOVA matrix $J_\beta(m_1, m_2, n) = (I + W(m_1, n)^{-1}W(m_2, n))^{-1}$ can be obtained by a suitable transformation.

Table 4.1 summarizes the matrix factorizations associated with the classical random matrix ensembles that allow us to compute the joint eigenvalue densities in the most natural manner. Later we will discuss additional connections between these matrix factorizations, and classical orthogonal polynomials.

4.5. Joint eigenvalue densities of the classical ensembles

The three Gaussian ensembles have joint eigenvalue probability density function

$$\text{Gaussian:} \quad f_\beta(\lambda) = c_H^\beta \prod_{i<j} |\lambda_i - \lambda_j|^\beta \mathrm{e}^{-\sum_{i=1}^n \lambda_i^2/2}, \tag{4.3}$$

with $\beta = 1$ corresponding to the reals, $\beta = 2$ to the complexes, $\beta = 4$ to the quaternion, and with

$$c_H^\beta = (2\pi)^{-n/2} \prod_{j=1}^n \frac{\Gamma(1 + \frac{\beta}{2})}{\Gamma(1 + \frac{\beta}{2}j)}. \tag{4.4}$$

The best references are Mehta (1991) and the original paper by Dyson (1963).

Similarly, the Wishart (or Laguerre) models have joint eigenvalue PDF

$$\text{Wishart:} \quad f_\beta(\lambda) = c_L^{\beta,a} \prod_{i<j} |\lambda_i - \lambda_j|^\beta \prod_i \lambda_i^{a-p} \mathrm{e}^{-\sum_{i=1}^n \lambda_i/2}, \tag{4.5}$$

with $a = \frac{\beta}{2}m$ and $p = 1 + \frac{\beta}{2}(n-1)$. Again, $\beta = 1$ for the reals, $\beta = 2$ for the complexes, and $\beta = 4$ for the quaternion. The constant

$$c_L^{\beta,a} = 2^{-na} \prod_{j=1}^n \frac{\Gamma(1+\frac{\beta}{2})}{\Gamma(1+\frac{\beta}{2}j)\Gamma(a-\frac{\beta}{2}(n-j)))}. \tag{4.6}$$

Good references are Muirhead (1982), Edelman (1989), and James (1964), and for $\beta = 4$, Macdonald (1998).

To complete the triad of classical orthogonal polynomials, we will mention the β-MANOVA ensembles, which are associated to the multivariate analysis of variance (MANOVA) model. They are better known in the literature as the Jacobi ensembles, with joint eigenvalue PDF, that is,

$$\text{MANOVA:} \quad f_\beta(\lambda) = c_J^{\beta,a_1,a_2} \prod_{i<j} |\lambda_i - \lambda_j|^\beta \prod_{j=1}^n \lambda_i^{a_1-p} \, (1-\lambda_i)^{a_2-p}, \tag{4.7}$$

with $a_1 = \frac{\beta}{2}m_1$, $a_2 = \frac{\beta}{2}m_2$, and $p = 1 + \frac{\beta}{2}(n-1)$. As usual, $\beta = 1$ for real and $\beta = 2$ for complex; also

$$c_J^{\beta,a_1,a_2} = \prod_{j=1}^n \frac{\Gamma(1+\frac{\beta}{2})\Gamma(a_1+a_2-\frac{\beta}{2}(n-j))}{\Gamma(1+\frac{\beta}{2}j)\Gamma(a_1-\frac{\beta}{2}(n-j))\Gamma(a_2-\frac{\beta}{2}(n-j))}. \tag{4.8}$$

Good references are the original paper by Constantine (1963), and Muirhead (1982) for $\beta = 1, 2$.

4.6. Haar-distributed orthogonal, unitary and symplectic eigenvectors

The eigenvectors of the classical random matrix ensembles are distributed with Haar measure. This is the uniform measure on orthogonal/unitary/symplectic matrices; see Chapter 1 of Milman and Schechtman (1986) for a derivation.

A measure $\mu(E)$ is a generalized volume defined on E. A measure μ, defined on a group G, is a Haar measure if $\mu(gE) = \mu(E)$, for every $g \in G$. For the example $O(n)$ of orthogonal $n \times n$ matrices, the condition that our measure be Haar is, for any continuous function f, that

$$\int_{Q \in O(n)} f(Q) \, \mathrm{d}\mu(Q) = \int_{Q \in O(n)} f(Q_o \, Q) \, \mathrm{d}\mu(Q), \qquad \text{for any } Q_o \in O(n).$$

In other words, Haar measure is symmetric: no matter how we rotate our

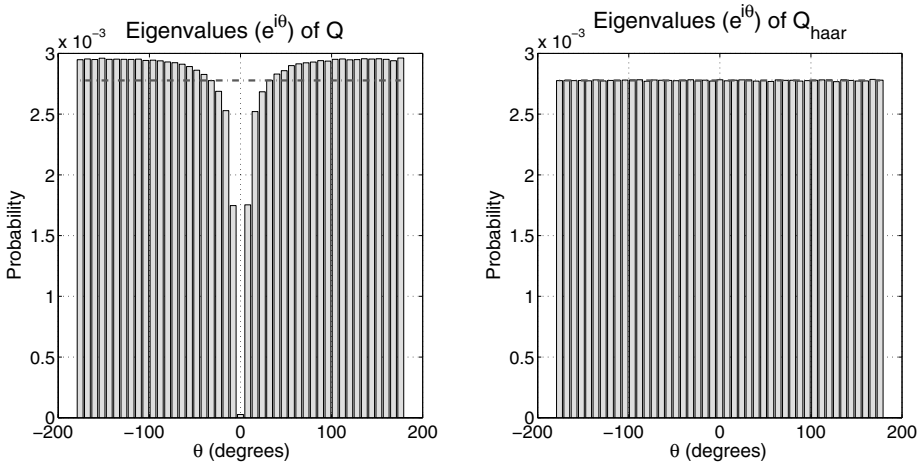

Figure 4.1. QR (Gram–Schmidt) factorization of `randn(n)`; no correction in the left panel, phase correction in the right panel.

sets, we get the same answer. In numerical terms, we can devise the following experiment to get some intuition on whether or not randomly generated unitary matrices are Haar-distributed.

Suppose we started with an $n \times n$ complex random matrix A constructed in MATLAB as

```
% Pick n
A=randn(n)+i*randn(n);
```

Compute its QR decomposition to generate a random unitary matrix Q:

```
[Q,R]=qr(A);
```

The eigenvalues of Q will be complex with a magnitude of 1, *i.e.*, they will be distributed on the unit circle in the complex plane. Compute the phases associated with these complex eigenvalues:

```
Qphase=angle(eig(Q));
```

Now, perform this experiment several times and collect the phase information in the variable `Qphase`. Plot the histogram of the phases (in degrees) normalized to have area 1. The left-hand panel of Figure 4.1 shows this histogram for $n = 50$ and $100,000$ trials. The dotted line indicates a uniform density between $[-180, 180]$. From this we conclude that since the phases of Q are not uniformly distributed, Q as constructed in this experiment is not distributed with Haar measure.

It is interesting to recognize why the experiment described above does not produce a Haar-distributed unitary matrix. This is because the QR

algorithm in MATLAB does not guarantee nonnegative diagonal entries in R. A simple correction by randomly perturbing the phases as:

```
Q=Q*diag(exp(i*2*pi*rand(n,1)));
```

or even by the sign of the diagonal entries of R:

```
Q=Q*diag(sign(diag(R)));
```

would correct this problem and produce the histogram in the right-hand panel of Figure 4.1 for the experiment described above. Note that the MATLAB command `rand` generates a random number uniformly distributed between 0 and 1. While this method of generating random unitary matrices with Haar measure is useful for simplicity, it is not the most efficient. For information on the efficient numerical generation of random orthogonal matrices distributed with Haar measure, see Stewart (1980).

4.7. The longest increasing subsequence

There is an interesting link between the moments of the eigenvalues of Q and the number of permutations of length n with longest increasing subsequence k. For example, the permutation $(3\,1\,8\,4\,5\,7\,2\,6\,9\,10)$ has $(1\,4\,5\,7\,9\,10)$ or $(1\,4\,5\,6\,9\,10)$ as the longest increasing subsequences of length 6.

This problem may be recast for the numerical analyst as the parallel complexity of solving an upper triangular system whose sparsity is given by a permutation π:

$$U_{ij}(\pi)\begin{cases}\neq 0 & \text{if } \pi(i)\leq\pi(j) \text{ and } i\leq j,\\ =0 & \text{if } \pi(i)>\pi(j) \text{ or } i>j.\end{cases}$$

The result from random matrix theory is that the number of permutations of length n with longest increasing subsequence less than or equal to length k is given by

$$E_{Q_k}\big(|\mathrm{tr}(Q_k)|^{2n}\big).$$

We can verify this numerically using what we know about generating Haar unitary random matrices from Section 4.6. We can construct a function in MATLAB that generates a Haar unitary matrix, computes the quantity $|\mathrm{tr}Q_k|^{2n}$ and averages it over many trials:

```
function L = longestsubsq(n,k,trials);
expt=[];
for idx=1:trials,
   % Generate Haar unitary matrix
   [Q,R]=qr(randn(k)+i*randn(k));
   Q=Q*diag(exp(i*2*pi*rand(k,1)));
   expt=[expt;abs(trace(Q))^(2*n)];
end

mean(exp)
```

Table 4.9. Permutations for $n = 4$.

1 2 3 4	2 1 3 4	3 1 2 4	4 1 2 3
1 2 4 3	2 1 4 3	3 1 4 2	4 1 3 2
1 3 2 4	2 3 1 4	3 2 1 4	4 2 1 3
1 3 4 2	2 3 4 1	3 2 4 1	4 2 3 1
1 4 2 3	2 4 1 3	3 4 1 2	4 3 1 2
1 4 3 2	2 4 3 1	3 4 2 1	4 3 2 1

For $n = 4$, there are 24 possible permutations listed in Table 4.9. We underline the fourteen permutations with longest increasing subsequence of length ≤ 2. Of these, one permutation (4 3 2 1) has length 1 and the other thirteen have length 2.

If we were to run the MATLAB code for $n = 4$ and $k = 2$ and 30000 trials we would get:

```
>> longestsubsq(4,2,30000)
ans = 14.1120
```

which is approximately equal to the number of permutations of length less than or equal to 2. It can be readily verified that the code gives the right answer for other combinations of n and k as well. We note that for this numerical verification, it was critically important that a Haar unitary matrix was generated. If we were to use a matrix without Haar measure, for example simply using the command `[Q,R]=qr(randn(n)+i*randn(n))` without randomly perturbing the phases, as described in Section 4.6, we would not get the right answer.

The authors still find it remarkable that the answer to a question this simple (at least in terms of formulation) involves integrals over Haar unitary matrices. There is, of course, a deep mathematical reason for this that is related to the correspondence between, on the one hand, permutations and combinatorial objects known as Young tableaux, via the Schensted correspondence, and, on the other hand, representations of the symmetric group and the unitary group. The reader may wish to consult Rains (1998), Aldous and Diaconis (1999) and Odlyzko and Rains (1998) for additional details. Related works include Borodin (1999), Tracy and Widom (2001) and Borodin and Forrester (2003).

5. Numerical algorithms stochastically

Matrix factorization algorithms may be performed stochastically given Gaussian inputs. What this means is that instead of performing the matrix reductions on a computer, they can be done by mathematics. The three

that are well known, though we will focus on the latter two, are:

(1) Gram–Schmidt (the `qr` decomposition)
(2) symmetric tridiagonalization (standard first step of `eig`), and
(3) bidiagonalization (standard first step of `svd`).

The bidiagonalization method is due to Golub and Kahan (1965), while the tridiagonalization method is due to Householder (1958).

These two linear algebra algorithms can be applied stochastically, and it is not very hard to compute the distributions of the resulting matrix.

The two key ideas are:

(1) the χ_r distribution, and
(2) the orthogonal invariance of Gaussians.

The χ_r is the χ-distribution with r degrees of freedom where r is any real number. It can be derived from the univariate Gaussian and is also the square root of the χ^2_r-distribution. Hence it may be generated using the MATLAB Statistics Toolbox using the command `sqrt(chi2rnd(r))`. 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 an $n \times 1$ vector of independent standard normals (or `norm(randn(n,1))` in MATLAB). The probability density function of χ_n can then be extended to any real number r so that the probability density function of χ_r is given by

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

The orthogonal invariance of Gaussians is mentioned in Section 4.3. In this form it means that

$$H \begin{pmatrix} G_1 \\ G_1 \\ \vdots \\ \vdots \\ G_1 \end{pmatrix} \stackrel{\mathcal{D}}{=} \begin{pmatrix} G \\ G \\ \vdots \\ \vdots \\ G \end{pmatrix},$$

if each G denotes an independent standard Gaussian and H any independent orthogonal matrix (such as a reflector).

Thus, for example, the first two steps of Gram–Schmidt applied to an $n \times n$ real Gaussian matrix ($\beta = 1$) are:

$$\begin{pmatrix} G & G & \cdots & G \\ G & G & \cdots & G \\ \vdots & \vdots & \cdots & \vdots \\ G & G & \cdots & G \end{pmatrix} \to \begin{pmatrix} \chi_n & G & \cdots & G \\ & G & \cdots & G \\ & \vdots & \cdots & \vdots \\ & G & \cdots & G \end{pmatrix} \to \begin{pmatrix} \chi_n & G & \cdots & G \\ & \chi_{n-1} & \cdots & G \\ & & \cdots & \vdots \\ & & G & G \end{pmatrix}.$$

Table 5.1. Tri- and bidiagonal models for the Gaussian and Wishart ensembles.

Gaussian Ensemble $n \in \mathbb{N}$	$H_n^\beta \sim \frac{1}{\sqrt{2}} \begin{pmatrix} N(0,2) & \chi_{(n-1)\beta} & & & & \\ \chi_{(n-1)\beta} & N(0,2) & \chi_{(n-2)\beta} & & & \\ & \ddots & \ddots & \ddots & & \\ & & \chi_{2\beta} & N(0,2) & \chi_\beta & \\ & & & \chi_\beta & N(0,2) & \end{pmatrix}$
Wishart ensemble $n \in \mathbb{N}$, $a \in \mathbb{R}$, $a > \frac{\beta}{2}(n-1)$	$L_n^\beta = B_n^\beta B_n^{\beta\prime}$, where $B_n^\beta \sim \begin{pmatrix} \chi_{2a} & & & \\ \chi_{\beta(n-1)} & \chi_{2a-\beta} & & \\ & \ddots & \ddots & \\ & & \chi_\beta & \chi_{2a-\beta(n-1)} \end{pmatrix}$

Applying the same ideas for tridiagonal or bidiagonal reduction gives the answer listed in Table 5.1, where the real case corresponds to $\beta = 1$, complex $\beta = 2$ and quaternion $\beta = 4$. For the Gaussian ensembles, before scaling the diagonal elements are i.i.d. normals with mean 0 and variance 2. The subdiagonal has independent elements that are χ variables as indicated. The superdiagonal is copied to create a symmetric tridiagonal matrix. The diagonal and the subdiagonals for the bidiagonal Wishart ensembles are independent elements that are χ-distributed with degrees of freedom having arithmetic progressions of step size β.

There is a tridiagonal matrix model for the β-Jacobi ensemble also, as described in Killip and Nenciu (2004); the correspondence between the CS decomposition and the Jacobi model is spelled out in Sutton (2005). Other models for the β-Jacobi ensemble include Lippert (2003).

There is both an important computational and theoretical implication of applying these matrix factorizations stochastically. Computationally speaking, often much of the time goes into performing these reductions for a given realization of the ensemble. Having them available analytically means that the constructions in Section 4.3 are highly inefficient for numerical simulations of the Hermite and Laguerre ensembles. Instead, we can generate then much more efficiently using MATLAB code and the Statistics Toolbox as listed in Table 5.2. The tangible savings in storage $O(n^2)$ to $O(n)$ is reflected in similar savings in computational complexity when computing their eigenvalues too. Not surprisingly, these constructions have been rediscovered independently by several authors in different contexts. Trotter (1984) used it in his alternate derivation of Wigner's semi-circular law.

Table 5.2. Generating the β-Hermite and β-Laguerre ensembles efficiently.

Ensemble	MATLAB commands

β-Hermite	`% Pick n, beta`
	`d=sqrt(chi2rnd(beta*[n:-1:1]))';`
	`H=spdiags(d,1,n,n)+spdiags(randn(n,1),0,n,n);`
	`H=(H+H')/sqrt(2);`

β-Laguerre	`% Pick m, n, beta`
	`% Pick a > beta*(n-1)/2;`
	`d=sqrt(chi2rnd(2*a-beta*[0:1:n-1]))';`
	`s=sqrt(chi2rnd(beta*[n:-1:1]))';`
	`B=spdiags(s,-1,n,n)+spdiags(d,0,n,n)`
	`L=B*B';`

Similarly, Silverstein (1985) and, more recently, Baxter and Iserles (2003) have rederived this result; probably many others have as well.

Theoretically the parameter β plays a new important role. The answers show that insisting on $\beta = 1, 2$ and 4 is no longer necessary. While these three values will always play something of a special role, like the mathematician who invents the Gamma function and forgets about counting permutations, we now have a whole continuum of possible betas available to us. While clearly simplifying the 'book-keeping' in terms of whether the elements are real, complex or quaternion, this formalism can be used to re-interpret and rederive familiar results as in Dumitriu (2003).

The general β version requires a generalization of $G_\beta(1, 1)$. We have not seen any literature but formally it seems clear how to work with such an object (rules of algebra are standard, rules of addition come from the normal distribution, and the absolute value must be a χ_β distribution). From there, we might formally derive a general Q for each β.

6. Classical orthogonal polynomials

We have already seen in Section 4 that the weight function associated with classical orthogonal polynomials plays an important role in random matrix theory.

Given a weight function $w(x)$ and an interval $[a, b]$ the orthogonal polynomials satisfy the relationship

$$\int_a^b p_j(x)p_k(x)w(x)\,\mathrm{d}x = \delta_{jk}.$$

In random matrix theory there is interest in matrices with joint eigen-

Table 6.1. The classical orthogonal polynomials.

Polynomial	Interval $[a, b]$	$w(x)$
Hermite	$(-\infty, \infty)$	$\mathrm{e}^{-x^2/2}$
Laguerre	$[0, \infty)$	$x^k \mathrm{e}^{-x}$
Jacobi	$(-1, 1)$	$(1-x)^a(1+x)^b$

value density proportional to $\prod w(\lambda_i)|\Delta(\lambda)|^\beta$ where $\Delta(x) = \prod_{i<j}(x_i - x_j)$. Table 6.1 lists the weight functions and the interval of definition for the classical Hermite, Laguerre and Jacobi orthogonal polynomials as found in classical references such as Abramowitz and Stegun (1970).

Note that the Jacobi polynomial reduces to the Legendre polynomial when $\alpha = \beta = 0$, and to the Chebyshev polynomials when $\alpha, \beta = \pm 1/2$.

Classical mathematics suggests that a procedure such as Gram–Schmidt orthonormalization can be used to generate these polynomials. Numerically, however, other procedures are available, as detailed in Gautschi (1996).

There are deep connections between these classical orthogonal polynomials and three of the classical random matrix ensembles as alluded to in Section 4.

The most obvious link is between the form of the joint eigenvalue densities for these matrix ensembles and the weight functions $w(x)$ of the associated orthogonal polynomial. Specifically, the joint eigenvalue densities of the Gaussian (Hermite), Wishart (Laguerre) and MANOVA (Jacobi) ensembles given by (4.3), (4.5), and (4.7) can be written in terms of the weight functions where $\Delta(\Lambda) = \prod_{i<j}|\lambda_i - \lambda_j|$ is the absolute value of the Vandermonde determinant.

6.1. Equilibrium measure and the Lanczos method

In orthogonal polynomial theory, given a weight function $w(x)$, with integral 1, we obtain Gaussian quadrature formulas for computing

$$\int f(x)w(x)\,\mathrm{d}x \approx \sum_{i=1}^{n} f(x_i)q_i^2.$$

In other words, we have the approximation

$$w(x) \approx \sum \delta(x - x_i)q_i^2.$$

Here the x_i are the roots of the nth orthogonal polynomial, and the q_i^2 are the related Christoffel numbers also obtainable from the nth orthogonal polynomial.

The Lanczos algorithm run on the continuous operator $w(x)$ gives a tridiagonal matrix whose eigenvalues are the x_i and the first component of the ith eigenvector is q_i.

As the q_i^2 weigh each x_i differently, the distribution of the roots

$$e_n(x) = \sum_{i=1}^{n} \delta(x - x_i)$$

converges to a different answer from $w(x)$. For example, if $w(x)$ corresponds to a Gaussian, then the limiting $e_n(x)$ is semi-circular. Other examples are listed in Table 4.2.

These limiting measures, $e(x) = \lim_{n\to\infty} e_n(x)$, are known as the equilibrium measure for $w(x)$. They are characterized by a solution to a two-dimensional force equilibrium problem on a line segment. These equilibrium measures become the characteristic densities of random matrix theory as listed in Table 4.2. They have the property that, under the right conditions,

$$\operatorname{Re} m(x) = \frac{w'(x)}{w(x)},$$

where $m(x)$ is the Cauchy transform of the equilibrium measure.

In recent work, Kuijlaars (2000) has made the connection between the equilibrium measure and how Lanczos finds eigenvalues. Under reasonable assumptions, if we start with a large matrix, and take a relatively smaller number of Lanczos steps, then Lanczos follows the equilibrium measure. This is more or less intuitively clear. What he discovered was how one interpolates between the equilibrium measure and the original measure as the algorithm proceeds. There is a beautiful combination of a cut-off equilibrium measure and the original weight that applies during the transition.

For additional details on the connection see Kuijlaars and McLaughlin (2000). For a good reference on equilibrium measure, see Deift (1999, Chapter 6).

6.2. Matrix integrals

A strand of random matrix theory that is connected to the classical orthogonal polynomials is the evaluation of matrix integrals involving the joint eigenvalue densities. One can see this in works such as Mehta (1991).

Definition. Let A be a matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$. The empirical distribution function for the eigenvalues of A is the probability measure

$$\frac{1}{n} \sum_{i=1}^{n} \delta(x - \lambda_i).$$

Definition. The level density of an $n \times n$ ensemble A with real eigenvalues is the distribution of a random eigenvalue chosen from the ensemble. Equivalently, it is the average (over the ensemble) empirical density. It is denoted by ρ_n^A.

There is another way to understand the level density in terms of a matrix integral. 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.

Specifically, the level density can be written in terms of the joint eigenvalue density $f_A(\lambda_1, \ldots, \lambda_n)$ as

$$\rho_{n,\beta}^A(\lambda_1) = \int_{\mathbb{R}^{n-1}} f_A(\lambda_1, \ldots, \lambda_n)\, \mathrm{d}\lambda_2 \cdots \mathrm{d}\lambda_n.$$

For the case of the β-Hermite ensemble, this integral can be written in terms of its joint eigenvalue density in (4.3) as

$$\rho_{n,\beta}^H(\lambda_1) = c_H^\beta \int_{\mathbb{R}^{n-1}} |\Delta(\Lambda)|^\beta \mathrm{e}^{-\sum_{i=1}^n \lambda_i^2/2}\, \mathrm{d}\lambda_2 \cdots \mathrm{d}\lambda_n. \tag{6.1}$$

The integral in (6.1) certainly looks daunting. Surprisingly, it turns out that closed form expressions are available in many cases.

6.3. Matrix integrals for complex random matrices

When the underlying random matrix is complex ($\beta = 2$), some matrix integrals become particularly easy. They are an application of the Cauchy–Binet theorem that is sometimes familiar to linear algebraists from texts such as Gantmacher and Krein (2002).

Theorem 6.1. (Cauchy–Binet) Let $C = AB$ be a matrix product of any kind. Let $M\binom{i_1 \cdots i_p}{j_1 \cdots j_p}$ denote the $p \times p$ minor

$$\det(M_{i_k j_l})_{1 \le k \le p, 1 \le l \le p}.$$

In other words, it is the determinant of the submatrix of M formed from rows $i_1, \ldots, i_p$ and columns $j_1, \ldots, j_p$.

The Cauchy–Binet theorem states that

$$C\begin{pmatrix} i_1, \ldots, i_p \\ k_1, \ldots, k_p \end{pmatrix} = \sum_{j_1 < j_2 < \cdots < j_p} A\begin{pmatrix} i_1, \ldots, i_p \\ j_1, \ldots, j_p \end{pmatrix} B\begin{pmatrix} j_1, \ldots, j_p \\ k_1, \ldots, k_p \end{pmatrix}.$$

Notice that when $p = 1$ this is the familiar formula for matrix multiplication. When all matrices are $p \times p$, then the formula states that

$$\det C = \det A \det B.$$

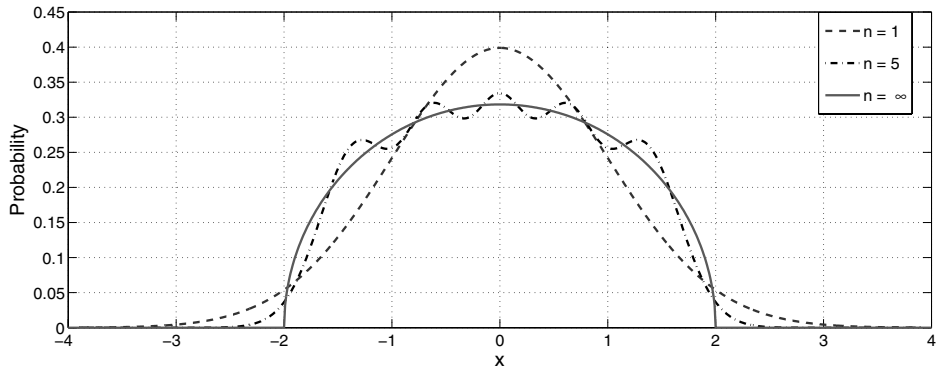

Figure 6.1. Level density of the GUE ensemble ($\beta = 2$) for different values of n. The limiting result when $n \to \infty$ is Wigner's famous semi-circular law.

Cauchy–Binet extends in the limit to matrices with infinitely many columns. If the columns are indexed by a continuous variable, we now have a vector of functions.

Replacing A_{ij} with $\varphi_i(x_j)$ and B_{jk} with $\psi_k(x_j)$, we see that Cauchy–Binet becomes

$$\det C = \int \cdots \int \det(\varphi_i(x_j))_{i,j=1,\ldots,n} \det(\psi_k(x_j))_{k,j=1,\ldots,n} \, \mathrm{d}x_1 \, \mathrm{d}x_2 \cdots \mathrm{d}x_n.$$

where $C_{ik} = \int \varphi_i(x)\psi_k(x) \, \mathrm{d}x, \ i,k = 1,\ldots,n$.

This continuous version of Cauchy–Binet may be traced back to Andréief (1883).

We assume that $\beta = 2$ so that $w_n(x) = \Delta(x)^2 \prod_{i=1}^n w(x_i)$. For classical weight function $\omega(x)$, Hermitian matrix models have been constructed. We have already seen the GUE corresponding to Hermite matrix models, and complex Wishart matrices for Laguerre. We also get the complex MANOVA matrices corresponding to Jacobi.

Notation: we define $\phi_n(x) = p_n(x)w(x)^{1/2}$. Thus the $\phi_i(x)$ are not polynomials, but they do form an orthonormal set of functions on the support of the weight function, $w(x)$.

It is a general fact that the level density of an $n \times n$ complex ($\beta = 2$) classical random matrix ensemble

$$f_w(x) = \frac{1}{n} \sum_{i=0}^{n-1} \phi_i(x)^2.$$

Figure 6.1 compares the normalized level density of the GUE for different values of n using $w(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$. When $n = 1$, it is simply the normal distribution. The asymptotic result is the celebrated Wigner's semi-circular law (Wigner 1958).

Analogously to the computation of the level density, given any function $f(x)$ one can ask for

$$E(f) \equiv E_{\omega_n}\Big(\prod(f(x_i)\Big).$$

When we have a matrix model, this is $E(\det(f(X))$.

It is a simple result that $E(f) = \int(\det(\phi_i(x)\phi_j(x)f(x))_{i,j=0,\dots,n-1}\,\mathrm{d}x$. This implies, by the continuous version of the Cauchy–Binet theorem, that

$$E(f) = \det C_n,$$

where $(C_n)_{ij} = \int \phi_i(x)\phi_j(x)f(x)\,\mathrm{d}x$.

Some important functions to use are $f(x) = 1 + \sum z_i(\delta(x - y_i))$. The coefficients of the resulting polynomial are then the marginal density of k eigenvalues. See Tracy and Widom (1998) for more details.

Another important function is $f(x) = 1-\chi_{[a,b]}$, where $\chi_{[a,b]}$ is the indicator function on $[a, b]$. Then we obtain the probability that no eigenvalue is in the interval $[a, b]$. If b is infinite, we obtain the probability that all eigenvalues are less than a, that is, the distribution function for the largest eigenvalue.

Research on integrable systems is a very active area within random matrix theory in conjunction with applications in statistical physics, and statistical growth processes. Some good references on this subject are van Moerbeke (2001), Tracy and Widom (2000*b*), Its, Tracy and Widom (2001), Deift, Its and Zhou (1997) and Deift (2000). The connection with the Riemann–Hilbert problem is explored in Deift (1999), Kuijlaars (2003) and Bleher and Its (1999).

7. Multivariate orthogonal polynomials

We feel it is safe to say that classical orthogonal polynomial theory and the theory of special functions reached prominence in numerical computation just around or before computers were becoming commonplace. The knowledge has been embodied in such handbooks as Abramowitz and Stegun (1970), Erdélyi, Magnus, Oberhettinger and Tricomi (1981*a*), Erdélyi, Magnus, Oberhettinger and Tricomi (1981*b*), Erdélyi, Magnus, Oberhettinger and Tricomi (1955), Spanier and Oldham (1987) and Weisstein (2005).

Very exciting developments linked to random matrix theory are the orthogonal polynomials and special functions of a matrix argument. These are scalar functions of a matrix argument that depend on the eigenvalues of the matrix, but in highly nontrivial ways. They are not mere trivial generalizations of the univariate objects. They are also linked to the other set of special functions that arise in random matrix theory: the Painlevé equations (see Section 9).

We refer readers to works by James (1964), Muirhead (1982), and Forrester (2005) for statistical and random matrix applications, and Hanlon,

Stanley and Stembridge (1992) for combinatorial aspects. Stanley (1989) is another good reference on the subject.

The research terrain is wide open to study fully the general multivariate orthogonal polynomial theory. Generalizations of Lanczos and other applications seem like low-hanging fruit for anyone to pick. Also, the numerical computation of these functions were long considered out of reach. As we describe in Section 8, applications of dynamic programming have suddenly now made these functions computable.

Our goal is to generalize orthogonal polynomials $p_k(x)$ with respect to a weight function $w(x)$ on $[a, b]$. The objects will be denoted $p_\kappa(X)$, where $\kappa \equiv (k_1, k_2, \ldots)$ is a partition of K, *i.e.*, $k_1 \geq k_2 \geq \cdots$ and $K = k_1 + k_2 + \cdots$. The partition κ is the multivariate degree in the sense that the leading term of $p_\kappa(X)$ is

$$\sum_{\text{sym terms}} \lambda_1^{k_1} \lambda_2^{k_2} \cdots,$$

where the $\lambda_1 \leq \cdots \leq \lambda_n$ are the eigenvalues of X.

We define $W(X) = \det(w(X)) = \prod_i w(\lambda_i)$ for X such that $\lambda_1 \geq a$ and $\lambda_n \leq b$. The multivariate orthogonality property is then

$$\int_{aI \leq X \leq bI} p_\kappa(X) p_\mu(X) W(X)\, \mathrm{d}X = \delta_{\kappa\mu}.$$

The multivariate orthogonal polynomials may also be defined as polynomials in n variables:

$$\int_{\substack{a \leq x_i \leq b, \\ i=1,2,\ldots,n}} p_\kappa(x_1, \ldots, x_n) p_\mu(x_1, \ldots, x_n) \prod_{i<j} |x_i - x_j|^\beta \prod_{i=1}^n w(x_i)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n = \delta_{\kappa\mu},$$

where $\beta = 1, 2, 4$, according to Table 3.2, or may be arbitrary.

The simplest univariate polynomials are the monomials $p_n(x) = x^n$. They are orthogonal on the unit circle. This is Fourier analysis. Formally we take $w(x) = 1$ if $|x| = 1$ for $x \in \mathbb{C}$. The multivariate version is the famous Jack polynomial $C_\kappa^{2/\beta}(X)$ introduced in 1970 by Henry Jack as a one-parameter family of polynomials that include the Schur functions ($\beta = 2, \alpha = 1$) and (as conjectured by Jack (1970) and later proved by Macdonald (1982)) the zonal polynomials ($\beta = 1, \alpha = 2$). The Schur polynomials are well known in combinatorics, representation theory and linear algebra in their role as the determinant of a generalized Vandermonde matrix: see Koev (2002). One may also define the Jack polynomials by performing the QR factorization on the matrix that expresses the power symmetric functions $p_\kappa(X) = \prod \mathrm{tr}(X^{k_i})$ in terms of the monomial symmetric function $m_\kappa(X) = \sum x_i^{\kappa_i}$. The Q in the QR decomposition becomes a generalized character table while R defines

the Jack polynomials. Additional details may be found in Knop and Sahi (1997).

Dumitriu has built a symbolic package (MOPs) written in Maple, for the evaluation of multivariate polynomials symbolically. This package allows the user to write down and compute the Hermite, Laguerre, Jacobi and Jack multivariate polynomials.

This package has been invaluable in the computation of matrix integrals and multivariate statistics for general β or a specific $\beta \neq 2$ for which traditional techniques fall short. For additional details see Dumitriu and Edelman (2004).

8. Hypergeometric functions of matrix argument

The classical univariate hypergeometric function is well known:

$$ {}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; x) \equiv \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_p)_k}{k!(b_1)_k \cdots (b_q)_k} \cdot x^k, $$

where $(a)_k = a(a+1)\cdots(a+k-1)$.

The multivariate version is

$$ {}_pF_q^{\alpha}(a_1, .., a_p; b_1, .., b_q; x_1, .., x_n) \equiv \sum_{k=0}^{\infty} \sum_{\kappa \vdash k} \frac{(a_1)_\kappa \cdot \cdot (a_p)_\kappa}{k!(b_1)_\kappa \cdot \cdot (b_q)_\kappa} C_\kappa^{\alpha}(x_1, .., x_n), $$

where

$$ (a)_\kappa \equiv \sum_{(i,j) \in \kappa} \left(a - \frac{i-1}{\alpha} + j - 1 \right) $$

is the Pochhammer symbol and $C_\kappa^{\alpha}(x_1, x_2, \ldots, x_n)$ is the Jack polynomial.

Some random matrix statistics of the multivariate hypergeometric functions are the largest and smallest eigenvalue of a Wishart matrix. As in Section 5, the Wishart matrix can be written as $L = BB^T$, where

$$ B = \begin{bmatrix} \chi_{2a} & & & \\ \chi_{\beta(n-1)} & \chi_{2a-\beta} & & \\ & \ddots & \ddots & \\ & & \chi_{\beta} & \chi_{2a-\beta(n-1)} \end{bmatrix}, $$

where $a = m\frac{\beta}{2}$. The probability density function of the smallest eigenvalue of the Wishart matrix is

$$ f(x) = x^{kn} \cdot \mathrm{e}^{-\frac{nx}{2}} \cdot {}_2F_0^{2/\beta}\left(-k, \beta\frac{n}{2} + 1; ; -\frac{2}{x} I_{n-1} \right), $$

where $k = a - (n-1)\frac{\beta}{2} - 1$ is a nonnegative integer. Figure 8.1 shows

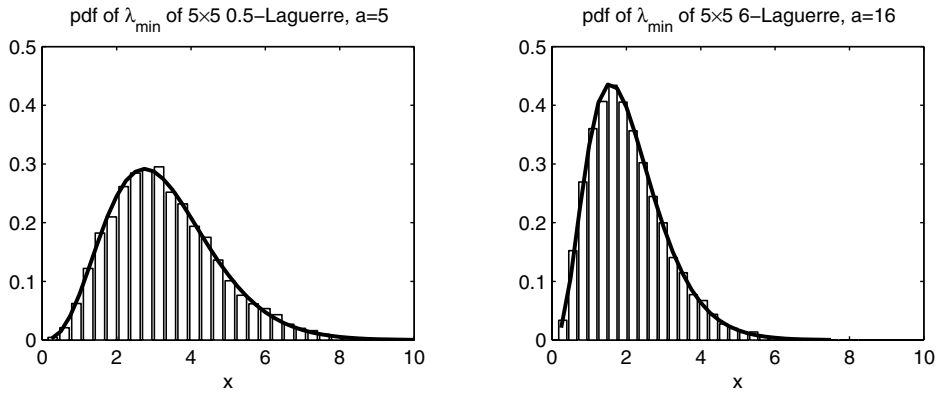

Figure 8.1. The probability density function of $\lambda_{\min}$ of the β-Laguerre ensemble.

this distribution against a Monte Carlo simulation for 5×5 matrices with $\beta = 0.5$ and $a = 5$ and $\beta = 6$ and $a = 16$.

Hypergeometrics of matrix argument also solve the random hyperplane angle problem. One formulation picks two random p-hyperplanes through the origin in n dimensions and asks for the distribution of the angle between them. For numerical applications and the formulae see Absil, Edelman and Koev (2004).

A word on the computation of these multivariate objects. The numerical computation of the classical function is itself difficult if the user desires accuracy over a large range of parameters. Many articles and books on multivariate statistics consider the multivariate function difficult.

In recent work Koev has found an algorithm for computing matrix hypergeometrics based on exploiting the combinatorial properties of the Pochhammer symbol, dynamic programming, and the algorithm for computing the Jack function. For a specific computation, this replaces an algorithm in 2000 that took 8 days to one that requires 0.01 seconds. See Koev and Edelman (2004) for more details.

9. Painlevé equations

The Painlevé equations, already favourites of those who numerically study solitons, now appear in random matrix theory and in the statistics of zeros of the Riemann zeta function. In this section we introduce the equations, show the connection to random matrix theory, and consider numerical solutions matched against theory and random matrix simulations.

We think of the Painlevé equations as the most famous equations not found in the current standard handbooks. This will change rapidly. They are often introduced in connection to the problem of identifying second-

order differential equations whose singularities may be poles depending on the initial conditions ('movable poles') and other singularities that are not movable. For example, the first-order equation

$$y' + y^2 = 0, \quad y(0) = \alpha$$

has solution

$$y(x) = \frac{\alpha}{\alpha x + 1},$$

which has a movable pole at $x = -1/\alpha$. (To repeat, the pole moves with the initial condition.) The equation

$$y'' + (y')^2 = 0, \quad y(0) = \alpha, \quad y'(0) = \beta$$

has solution

$$y(x) = \log(1 + x\beta) + \alpha.$$

This function has a movable log singularity ($x = -1/\beta$) and hence would not be of the type considered by Painlevé.

Precisely, Painlevé allowed equations of the form $y'' = R(x, y, y')$, where R is analytic in x and rational in y and y'. He proved that the equations whose only movable singularities are poles can be transformed into either a linear equation, an elliptic equation, a Riccati equation or one of the six families of equations below:

(I) $y'' = 6y^2 + t,$

(II) $y'' = 2y^3 + ty + \alpha$

(III) $y'' = \dfrac{1}{y}y'^2 - \dfrac{y'}{t} + \dfrac{\alpha y^2 + \beta}{t} + \gamma y^3 + \dfrac{\delta}{y},$

(IV) $y'' = \dfrac{1}{2y}y'^2 + \dfrac{3}{2}y^3 + 4ty^2 + 2(t^2 - \alpha)y + \dfrac{\beta}{y},$

(V) $y'' = \left(\dfrac{1}{2y} + \dfrac{1}{y-1}\right)y'^2 - \dfrac{1}{t}y' + \dfrac{(y-1)^2}{t}\left(\alpha y + \dfrac{\beta}{y}\right) + \gamma\dfrac{y}{t} + \delta\dfrac{y(y+1)}{y-1},$

(VI) $y'' = \dfrac{1}{2}\left(\dfrac{1}{y} + \dfrac{1}{y-1} + \dfrac{1}{y-t}\right)y'^2 - \left(\dfrac{1}{t} + \dfrac{1}{t-1} + \dfrac{1}{y-t}\right)y'$
$+ \dfrac{y(y-1)(y-t)}{t^2(t-1)^2}\left[\alpha - \beta\dfrac{t}{y^2} + \gamma\dfrac{t-1}{(y-1)^2} + \left(\dfrac{1}{2} - \delta\right)\dfrac{t(t-1)}{(y-t)^2}\right].$

A nice history of the Painlevé equation may be found in Takasaki (2000). Deift (2000) has a good exposition on this as well, where the connection to Riemann–Hilbert problems, explored in greater detail in Deift *et al.* (1997), is explained nicely. (A Riemann–Hilbert problem prescribes the jump condition across a contour and asks which problems satisfy this condition.)

In random matrix theory, distributions of some statistics related to the eigenvalues of the classical random matrix ensembles are obtainable from solutions to a Painlevé equation. The Painlevé II, III, V equations have been well studied, but others arise as well. More specifically, it turns out that integral operator discriminants related to the eigenvalue distributions satisfy differential equations, which involve the Painlevé equations in the large n limit. Connections between Painlevé theory and the multivariate hypergeometric theory of Section 7 are discussed in Forrester and Witte (2004) though more remains to be explored.

9.1. Eigenvalue distributions for large random matrices

In the study of eigenvalue distributions, two general areas can be distinguished. These are, respectively, the *bulk*, which refers to the properties of all of the eigenvalues and the *edges*, which (generally) addresses the largest and smallest eigenvalues.

A kernel $K(x, y)$ defines an operator K on functions f via

$$K[f](x) = \int K(x, y) f(y)\, \mathrm{d}y. \tag{9.1}$$

With appropriate integration limits, this operator is well defined if $K(x, y)$ is chosen as in Table 9.1. Discretized versions of these operators are famous 'test matrices' in numerical analysis as in the case of the sine-kernel which discretizes to the prolate matrix (Varah 1993).

The determinant becomes a 'Fredholm determinant' in the limit of large random matrices. This is the first step in the connection to Painlevé theory. The full story may be found in the Tracy–Widom papers (Tracy and Widom 1993, 1994*a*, 1994*b*) and in the paper by Forrester (2000). The term 'soft

Table 9.1. Operator kernels associated with the different eigenvalue distributions.

Painlevé	Statistic	Interval ($s > 0$)	Kernel	$K(x, y)$
V	'bulk'	$[-s, s]$	sine	$\dfrac{\sin(\pi(x-y))}{\pi(x-y)}$
III	'hard edge'	$(0, s]$	Bessel	$\dfrac{\sqrt{y}J_\alpha(\sqrt{x})J'_\alpha(\sqrt{y}) - \sqrt{x}J_\alpha(\sqrt{y})J'_\alpha(\sqrt{x})}{2(x-y)}$
II	'soft edge'	$[s, \infty)$	Airy	$\dfrac{\mathrm{Ai}(x)\,\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\,\mathrm{Ai}(y)}{x-y}$

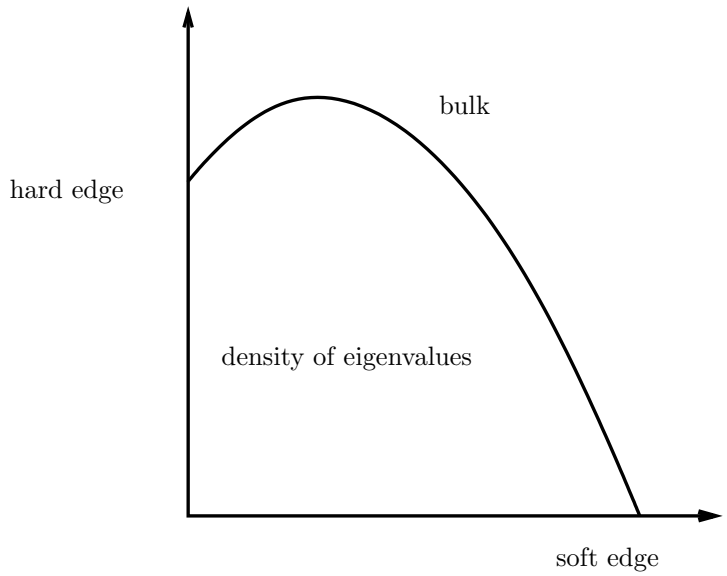

Figure 9.1. Regions corresponding to eigenvalue distributions that are of interest in random matrix theory.

edge' applies (because there is still 'wiggle room') when the density hits the horizontal axis, while the 'hard edge' applies when the density hits the vertical axis (no further room on the left because of positivity constraints on the eigenvalues, for example as is the case for the smallest eigenvalue of the Laguerre and Jacobian ensembles). This is illustrated in Figure 9.1 and is reflected in the choice of the integration intervals in Table 9.1 as well.

The distributions arising here are becoming increasingly important as they are showing up in many places. Authors have imagined a world (perhaps in the past) where the normal distribution might be found experimentally or mathematically but without the central limit theorem to explain why. This is happening here with these distributions as in the connection to the zeros of the Riemann zeta function (discussed in Section 9.3), combinatorial problems (Deift 2000), and growth processes (Johansson 2000*a*). The relevance of β in this context has not been fully explored.

9.2. The largest eigenvalue distribution and Painlevé II

The distribution of the appropriately normalized largest eigenvalues of the Hermite ($\beta = 1, 2, 4$) and Laguerre ($\beta = 1, 2$) ensembles can be computed from the solution of the Painlevé II equation:

$$q'' = sq + 2q^3 \tag{9.2}$$

with the boundary condition

$$q(s) \sim \mathrm{Ai}(s), \qquad \text{as } s \to \infty. \tag{9.3}$$

The probability distributions thus obtained are the famous Tracy–Widom distributions.

The probability distribution $f_2(s)$, corresponding to $\beta = 2$, is given by

$$f_2(s) = \frac{\mathrm{d}}{\mathrm{d}s} F_2(s), \tag{9.4}$$

where

$$F_2(s) = \exp\left(- \int_s^\infty (x - s) q(x)^2 \, \mathrm{d}x \right). \tag{9.5}$$

The distributions $f_1(s)$ and $f_4(s)$ for $\beta = 1$ and $\beta = 4$ are the derivatives of $F_1(s)$ and $F_4(s)$ respectively, which are given by

$$F_1(s)^2 = F_2(s) \exp\left(- \int_s^\infty q(x) \, \mathrm{d}x \right) \tag{9.6}$$

and

$$F_4\left(\frac{s}{2^{\frac{2}{3}}} \right)^2 = F_2(s) \left(\cosh\left(\int_s^\infty q(x) \, \mathrm{d}x \right) \right)^2. \tag{9.7}$$

These distributions can be readily computed numerically. To solve using MATLAB, first rewrite (9.2) as a first-order system:

$$\frac{\mathrm{d}}{\mathrm{d}s} \begin{pmatrix} q \\ q' \end{pmatrix} = \begin{pmatrix} q' \\ sq + 2q^3 \end{pmatrix}. \tag{9.8}$$

This can be solved as an initial value problem starting at $s = s_0 =$ sufficiently large positive number, and integrating backwards along the s-axis. The boundary condition (9.3) then becomes the initial values

$$\begin{cases} q(s_0) &= \mathrm{Ai}(s_0), \\ q'(s_0) &= \mathrm{Ai}'(s_0). \end{cases} \tag{9.9}$$

This problem can be solved in just a few lines of MATLAB using the built-in Runge–Kutta-based ODE solver `ode45`. First define the system of equations as an inline function

```
deq=inline('[y(2); s*y(1)+2*y(1)^3]','s','y');
```

Next specify the integration interval and the desired output times:

```
s0=5;
sn=-8;
sspan=linspace(s0,sn,1000);
```

The initial values can be computed as

```
y0=[airy(s0); airy(1,s0)]
```

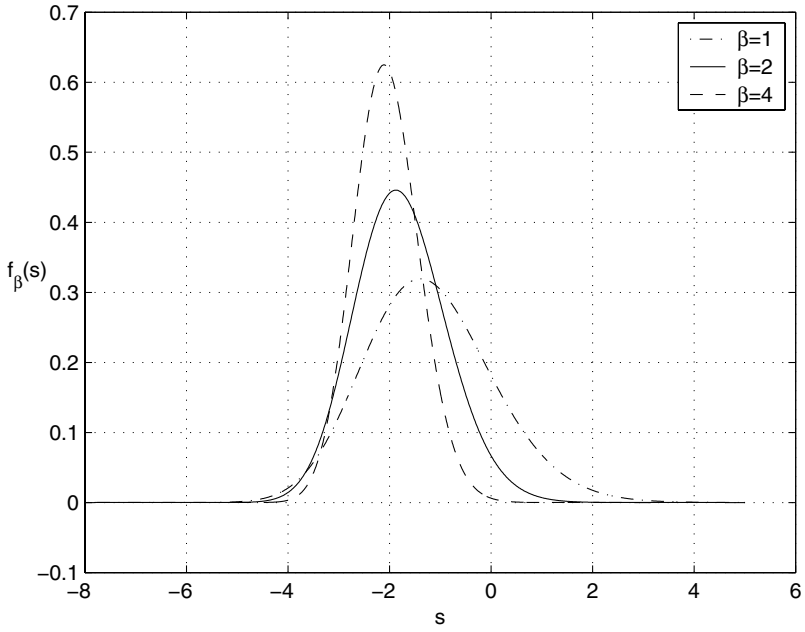

Figure 9.2. The Tracy–Widom distributions for $\beta = 1, 2, 4$.

Now, the integration tolerances can be set and the system integrated:

```
opts=odeset('reltol',1e-13,'abstol',1e-15);
[s,y]=ode45(deq,sspan,y0,opts);
q=y(:,1);
```

The first entry of the MATLAB variable `y` is the function $q(s)$. The distributions $F_2(s)$, $F_1(s)$ and $F_4(s)$ can be obtained from $q(s)$ by first setting the initial values:

```
dI0=I0=J0-0;
```

then numerically integrating to obtain:

```
dI=-[0;cumsum((q(1:end-1).^2+q(2:end).^2)/2.*diff(s))]+dI0;
I=-[0;cumsum((dI(1:end-1)+dI(2:end))/2.*diff(s))]+I0;
J=-[0;cumsum((q(1:end-1)+q(2:end))/2.*diff(s))]+J0;
```

Finally, using equations (9.5), (9.6), and (9.7) we obtain the desired distributions as:

```
F2=exp(-I);
F1=sqrt(F2.*exp(-J));
F4=sqrt(F2).*(exp(J/2)+exp(-J/2))/2;
s4=s/2^(2/3);
```

Note that the trapezoidal rule (`cumsum` function in MATLAB) is used to approximate numerically the integrals in (9.5), (9.6) and (9.7) respectively.

The probability distributions $f_2(s), f_1(s)$, and $f_4(s)$ can then computed by numerical differentiation:

```
f2=gradient(F2,s);
f1=gradient(F1,s);
f4=gradient(F4,s4);
```

The result is shown in Figure 9.2. Note that more accurate techniques for computing the Tracy–Widom distributions are known and have been implemented as in Edelman and Persson (2002). Dieng (2004) discusses the numerics of another such implementation.

These distributions are connected to random matrix theory by the following theorems.

Theorem 9.1. (Tracy and Widom 2000*a*) Let $\lambda_{\max}$ be the largest eigenvalue of $G_\beta(n,n)$, the β-Hermite ensemble, for $\beta = 1, 2, 4$. The normalized largest eigenvalue $\lambda'_{\max}$ is calculated as

$$\lambda'_{\max} = n^{\frac{1}{6}}(\lambda_{\max} - 2\sqrt{n}).$$

Then, as $n \to \infty$,

$$\lambda'_{\max} \xrightarrow{\mathcal{D}} F_\beta(s).$$

Theorem 9.2. (Johnstone 2001) Let $\lambda_{\max}$ be the largest eigenvalue of $W_1(m,n)$, the real Laguerre ensemble ($\beta = 1$). The normalized largest eigenvalue $\lambda'_{\max}$ is calculated as

$$\lambda'_{\max} = \frac{\lambda_{\max} - \mu_{mn}}{\sigma_{mn}},$$

where μ_{mn} and σ_{mn} are given by

$$\mu_{mn} = (\sqrt{m-1} + \sqrt{n})^2, \sigma_{mn} = (\sqrt{m-1} + \sqrt{n})\left(\frac{1}{\sqrt{m-1}} + \frac{1}{n}\right)^{\frac{1}{3}}.$$

Then, if $m/n \to \gamma \geq 1$ as $n \to \infty$,

$$\lambda'_{\max} \xrightarrow{\mathcal{D}} F_1(s).$$

Theorem 9.3. (Johansson 2000*b*) Let $\lambda_{\max}$ be the largest eigenvalue of $W_2(m,n)$, the complex Laguerre ensemble ($\beta = 2$). The normalized largest eigenvalue $\lambda'_{\max}$ is calculated as

$$\lambda'_{\max} = \frac{\lambda_{\max} - \mu_{mn}}{\sigma_{mn}},$$

where μ_{mn} and σ_{mn} are given by

$$\mu_{mn} = (\sqrt{m} + \sqrt{n})^2, \sigma_{mn} = (\sqrt{m} + \sqrt{n})\left(\frac{1}{\sqrt{m}} + \frac{1}{n}\right)^{\frac{1}{3}}.$$

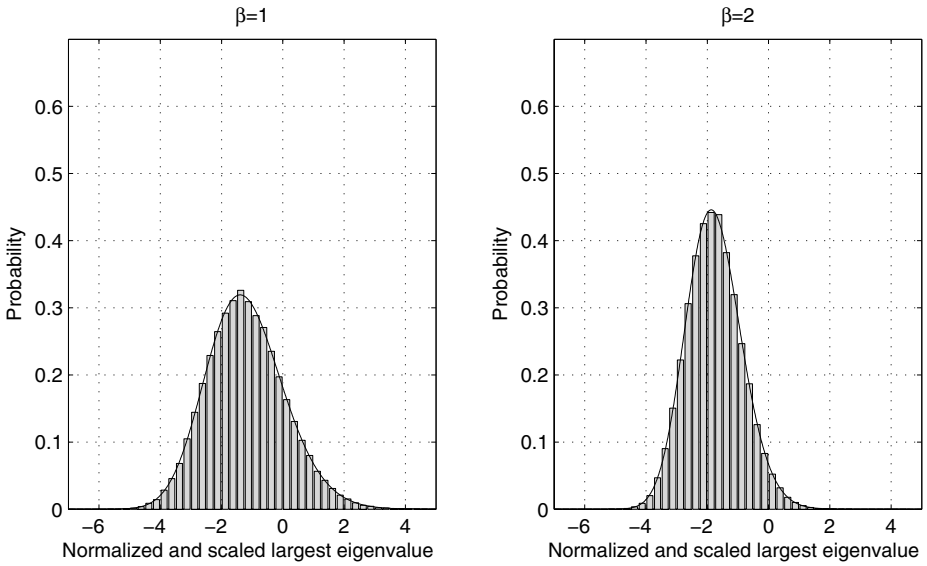

Figure 9.3. Probability distribution of scaled largest eigenvalue of the Hermite ensembles (10^5 repetitions, $n = 10^9$).

Then, if $m/n \to \gamma \geq 1$ as $n \to \infty$,

$$\lambda'_{\max} \xrightarrow{\mathcal{D}} F_2(s).$$

Figure 9.3 compares the probability distribution of the scaled large eigenvalue of the GOE, and GUE with the numerical results for a billion by billion matrix over 10^5 trials. We talk about how we generate data points for a billion by billion matrix later in this article. Related results include Soshnikov (1999). Dieng (2004) derives Painlevé-type expressions for the distribution of the kth-largest eigenvalue in the GOE and GSE in the edge scaling limit.

9.3. The GUE level spacings distribution and Painlevé V

Another quantity with an interesting probability distribution is the spacings of the eigenvalues of the Gaussian unitary ensemble, $G_2(n, n)$. The normalized spacings of the eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m$ are computed according to

$$\delta'_k = \frac{\lambda_{k+1} - \lambda_k}{\pi\beta}\sqrt{2\beta n - \lambda_k^2}, \qquad k \approx n/2. \tag{9.10}$$

The distribution of the eigenvalues is almost uniform, with a slight deviation at the two ends of the spectrum. Therefore, only half of the eigenvalues are used, and one quarter of the eigenvalues at each end is discarded.

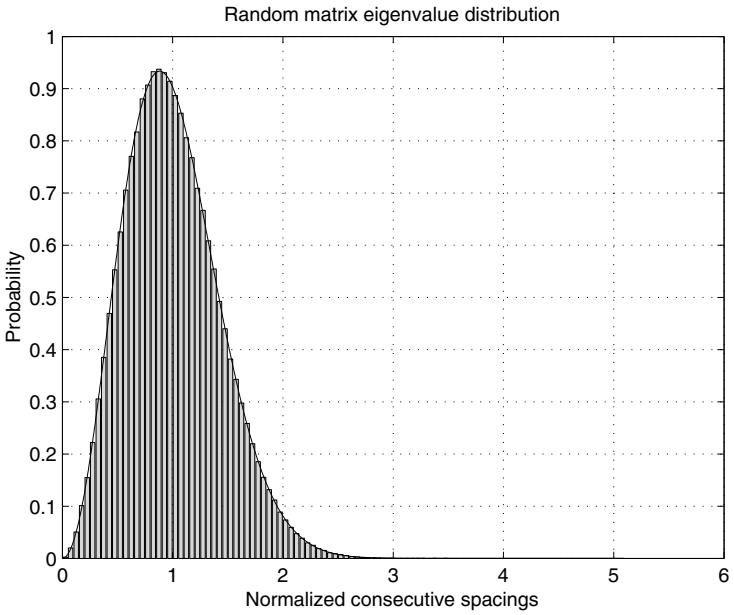

Figure 9.4. Probability distribution of consecutive spacings of the eigenvalues of a GUE ensemble (1000 repetitions, $n = 1000$).

The probability distribution $p(s)$ for the eigenvalue spacings when $\beta = 2$ can be computed with the solution to the Painlevé V nonlinear differential equation:

$$(t\sigma'')^2 + 4(t\sigma' - \sigma)\big(t\sigma' - \sigma + (\sigma')^2\big) = 0 \tag{9.11}$$

with the boundary condition

$$\sigma(t) \approx -\frac{t}{\pi} - \left(\frac{t}{\pi}\right)^2, \qquad \text{as } t \to 0^+. \tag{9.12}$$

Then $p(s)$ is given by

$$p(s) = \frac{\mathrm{d}^2}{\mathrm{d}s^2} E(s), \tag{9.13}$$

where

$$E(s) = \exp\left(\int_0^{\pi s} \frac{\sigma(t)}{t}\,\mathrm{d}t\right). \tag{9.14}$$

Explicit differentiation gives

$$p(s) = \frac{1}{s^2}\big(\pi s\sigma'(\pi s) - \sigma(\pi s) + \sigma(\pi s)^2\big)E(s). \tag{9.15}$$

The second-order differential equation (9.11) can be written as a first-order

system of differential equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\sigma\\ \sigma'\end{pmatrix} = \begin{pmatrix}\sigma'\\ -\frac{2}{t}\sqrt{(\sigma - t\sigma')(t\sigma' - \sigma + (\sigma')^2)}\end{pmatrix}. \tag{9.16}$$

This is solved as an initial value problem starting at $t = t_0 =$ very small positive number. The value $t = 0$ has to be avoided because of the division by t in the system of equations. This is not a problem, since the boundary condition (9.12) provides an accurate value for $\sigma(t_0)$ (as well as $E(t_0/\pi)$). The boundary conditions for the system (9.16) then become

$$\begin{cases}\sigma(t_0) &= -\frac{t_0}{\pi} - (\frac{t_0}{\pi})^2,\\ \sigma'(t_0) &= -\frac{1}{\pi} - \frac{2t_0}{\pi}.\end{cases} \tag{9.17}$$

This system can be solved numerically using MATLAB.

9.4. The GUE level spacings distribution and the Riemann zeta zeros

It has been observed that the zeros of the Riemann zeta function along the critical line $\mathrm{Re}(z) = \frac{1}{2}$ (for z large) behave similarly to the eigenvalues of random matrices in the GUE. Here, the distribution of the scaled spacings of the zeros is compared to the corresponding level spacing distribution computed using the Painlevé V equation.

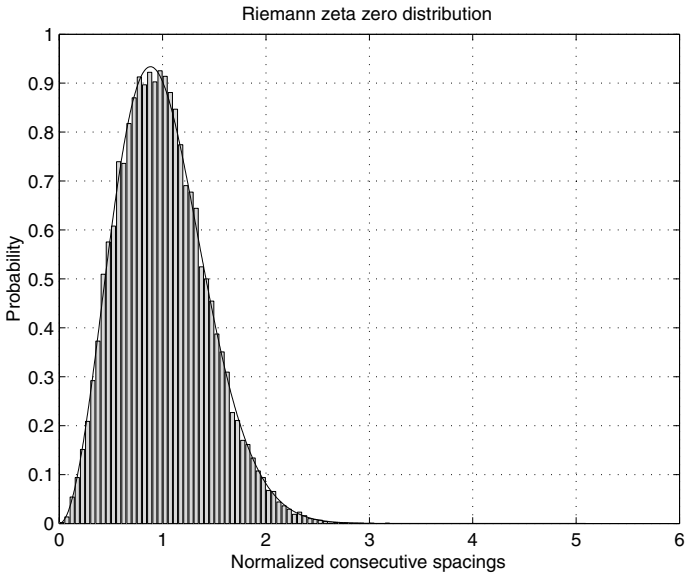

Figure 9.5. Probability distribution of consecutive spacings of Riemann zeta zeros (30,000 zeros, $n \approx 10^{12}, 10^{21}, 10^{22}$).

Define the nth zero γ_n as

$$\zeta\left(\frac{1}{2}+i\gamma_n\right)=0, \qquad 0<\gamma_1<\gamma_2<\cdots. \tag{9.18}$$

Compute a normalized spacing:

$$\tilde{\gamma}_n=\frac{\gamma_n}{\text{av spacing near } \gamma_n}=\gamma_n\cdot\left[\frac{\log\gamma_n/2\pi}{2\pi}\right]. \tag{9.19}$$

Zeros of the Riemann zeta function can be downloaded from Odlyzko (2001). Assuming that the MATLAB variable `gamma` contains the zeros, and the variable `offset` the offset, these two lines compute the consecutive spacings $\tilde{\gamma}_{n+1}-\tilde{\gamma}_n$ and plot the histogram:

```
delta=diff(gamma)/2/pi.*log((gamma(1:end-1)+offset)/2/pi);

% Normalize and plot the histogram of the spacings
```

The result can be found in Figure 9.5, along with the Painlevé V distribution.

10. Eigenvalues of a billion by billion matrix

We discuss how knowledge of numerical algorithms and software allow us to perform random matrix simulations very efficiently. In this case study, we illustrate an improvement rarely seen in computation. We succeeded in going from $n=100$ to $n=10^9$, *i.e.*, we can compute the largest eigenvalue of a billion by billion matrix in the time required by naive methods for a hundred by hundred matrix. Pushing to $n=10^{12}$ is within reach.

We devise a numerical experiment to verify that the distribution of the appropriately normalized and scaled largest eigenvalue of the GOE ensemble is given by the Tracy–Widom distribution $F_2(s)$ in (9.5).

Recall that an instance of the GOE ensemble ($\beta=1$) can be created conveniently in MATLAB as:

```
A=randn(n);
A=(A+A')/2;
```

It is now straightforward to compute the distribution for $\lambda'_{\max}$ by simulation:

```
for idx=1:trials
  A=randn(n);
  A=(A+A')/2;
  lmax=max(eig(A));
  lmaxscaled=n^(1/6)*(lmax-2*sqrt(n));
  % Store lmax
end

% Create and plot histogram
```

The problem with this technique is that the computational requirements and the memory requirements grow rapidly with n. Storing the matrix A requires n^2 double-precision numbers, so on most computers today n has to be less than 10^4. Furthermore, computing all the eigenvalues of a full Hermitian matrix requires computing time proportional to n^3. This means that it will take many days to create a smooth histogram by simulation, even for relatively small values of n.

To improve upon this situation, we can instead study the β-Hermite tridiagonal ensemble as in Table 5.1:

$$H_n^\beta \sim \frac{1}{\sqrt{2}} \begin{pmatrix} N(0,2) & \chi_{(n-1)\beta} & & & & \\ \chi_{(n-1)\beta} & N(0,2) & \chi_{(n-2)\beta} & & & \\ & \ddots & \ddots & \ddots & & \\ & & \chi_{2\beta} & N(0,2) & \chi_\beta & \\ & & & \chi_\beta & N(0,2) \end{pmatrix}. \tag{10.1}$$

Recall that $N(0,2)$ is a zero-mean Gaussian with variance 2, and χ_r is the square-root of a χ^2-distributed number with r degrees of freedom. Note that the matrix is symmetric, so the subdiagonal and the superdiagonal are always equal.

This matrix has a tridiagonal sparsity structure, and only $2n-1$ double-precision numbers are required to store an instance of it. The time for computing the largest eigenvalue is proportional to n, either using Krylov subspace-based methods or the method of bisection (Trefethen and Bau 1997). This is certainly an improvement, though not substantial enough to do a simulation of a billion by billion GOE as in Figure 9.3.

The following code can, however, be used to compute the largest eigenvalue of a billion by billion GOE ($\beta = 1$):

```
beta=1; n=1e9; opts.disp=0; opts;issym=1;
alpha=10;k=round(alpha*n^(1/3)); % cutoff parameters
d=sqrt(chi2rnd(beta*n:-1:(n-k-1)))';
H=spdiags(d,1,k,k)+spdiags(randn(k,1),0,k,k);
H=(H+H')/sqrt(4*n*beta); % Scale so largest eigenvalue is near 1
eigs(H,1,1,opts);
```

The technology underlying this code is remarkable and deserves to be widely known. A number of interesting tricks are combined together.

- The observation that if $k = 10n^{1/3}$, then the largest eigenvalue is determined numerically by the top $k \times k$ segment of n. (This is related to the decay of the Airy function that arises in the kernel whose eigenvalues determine the largest eigenvalue distribution. The 'magic number' 10 here is not meant to be precise. It approximates the index k such that $\frac{v(k)}{v(1)} \approx \epsilon$, where $\epsilon = 2^{-52}$ for double precision arithmetic, and v is

the eigenvector corresponding to the largest eigenvalue. For small β, it may be necessary to crank up the number 10 to a larger value.)

- Sparse matrix storage. (Only $O(n)$ storage is used.)
- Tridiagonal ensemble formulas. (Any beta is available due to the tridiagonal ensemble.)
- The Lanczos algorithm for eigenvalue computation. (This allows the computation of the largest eigenvalue faster than typical general purpose eigensolvers.)
- The shift-and-invert accelerator to Lanczos and Arnoldi. (Since we know the eigenvalues are near 1, we can accelerate the convergence of the largest eigenvalue.)
- The ARPACK software package as made available seamlessly in MATLAB. (The Arnoldi package contains state of the art data structures and numerical choices.)

Two of these tricks are mathematical. The first one is the ability to use tridiagonal ensembles to generate matrices whose eigenvalues match the GOE distribution. This allows us to avoid using dense matrices again for random matrix experiments. The second mathematical trick is the ability to cut off the top segment of the matrix to obtain accurately the largest eigenvalue.

It would be all too easy to take for granted the technology available for the remaining tricks. It was not so many years ago that the user would have to code up the sparse matrix storage made available by the '`spdiags`' command or the ability to peel off the largest eigenvalue and give a starting guess that is made available in '`eigs`'. Though numerical analysts are well versed in such numerical techniques, we would probably still not have bothered to implement the shift-and-invert Arnoldi-style algorithms ourselves. It has been said that technology advances by the number of operations that we do *not* have to think about. This is a great example.

Incidentally, for users interested in all of the eigenvalues of the tridiagonal matrix (Hermite ensembles such as the GOE, GUE, GSE) or all the singular values of a bidiagonal matrix (Laguerre ensembles such as Wishart matrices), then the LAPACK routines DSTEQR and DBDSQR can compute the eigenvalues with linear storage and in quadratic time. Users who simply use MATLAB's `eig`, Mathematica's `Eigenvalues`, or Maple's `linalg[eigenvalues]` are at a severe disadvantage.

We remark that further improvements are possible (and have been implemented!) if we use the approximation $\chi_n \approx \sqrt{n} + \frac{1}{\sqrt{2}} G$. This approximation forms the basis of the ideas in the next section. There are further tricks available, such as using the method of bisection (Trefethen and Bau 1997)

and approximating χ_n with simply $\sqrt{n}$. See Edelman and Persson (2002) for more details.

11. Stochastic operators

For years, the first author was mystified by the notation $\sqrt{\mathrm{d}t}$ often found in integrals connected with the Black–Scholes model of options pricing in finance. The simple fact that he was missing is that, if one has Gaussian random variables, the natural quantity that adds (thus, the linear function) is the variance, which is connected to the square of the variable.

There is some mathematics to be completed to understand fully how well-defined is the notion of the eigenvalues of a stochastic operator. Careful analysis will tell whether different discretizations give the same limiting eigenvalue distributions. Nonetheless, as we will outline, there is an idea here that we feel is sufficiently important that we can not afford to wait for this sort of analysis.

We define a Wiener process differentially as

$$\mathrm{d}W = (\text{standard normal}) \cdot \sqrt{\mathrm{d}t}.$$

The integral of such a process $W(t)$ (Brownian motion) is

$$W(t) = \int \mathrm{d}W.$$

This is probably best graphed in MATLAB with the command:

```
t = [dt:dt:1]';
W = cumsum(randn(length(t),1))*sqrt(dt);
plot([0;t],[0;W])
```

where $\mathrm{d}t$ = very small number not equal to zero and $W(0) = 0$. A good reference on Brownian motion is Karatzas and Shreve (1991).

Every time we 'roll the dice' we get a new W, but it is always the case that $W(t)$ is a Gaussian with variance t.

We are interested in operators exactly or closely related to the form

$$\begin{array}{lccccc} & \dfrac{\mathrm{d}^2}{\mathrm{d}x^2} & + & V(x) & + & \sigma\,\mathrm{d}W. \\ & \uparrow & & \uparrow & & \uparrow \\ \text{Discretization:} & \text{tridiagonal} & & \text{diagonal} & & \text{diagonal or tridiagonal} \end{array}$$

When discretized each term can be thought of as a tridiagonal or diagonal matrix. The last part requires Gaussians.

11.1. From random matrices to stochastic operators

Consider the β-Hermite ensemble. The eigenvalue distribution of this ensemble is shared by a tridiagonal matrix with real elements that could be constructed as

$$H_n^\beta = \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{2}G & \chi_{\beta(n-1)} & & & \\ \chi_{\beta(n-1)} & \sqrt{2}G & \chi_{\beta(n-2)} & & \\ & \ddots & \ddots & \ddots & \\ & & \chi_{\beta\cdot 2} & \sqrt{2}G & \chi_\beta \\ & & & \chi_\beta & \sqrt{2}G \end{bmatrix}.$$

This matrix is symmetric with independent entries in the upper triangular part. G represents an element taken from the standard Gaussian distribution, and χ_r represents an element taken from the χ-distribution with r degrees of freedom.

We are interested in the distribution of the largest eigenvalue, which is related to the solution of the Painlevé II transcendent.

Consider the β-Hermite ensemble from a numerical linear algebra point of view. The tridiagonal form suggests that H_n^β may be a finite difference approximation to some differential operator. We proposed that the β-Hermite ensemble is a finite difference approximation of the *stochastic Airy operator*:

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} - x + \sigma\,\mathrm{d}W, \tag{11.1}$$

in which $\mathrm{d}W$ represents a Wiener process. Recall that the Airy kernel in Table 9.1 plays an important role.

Hence, the random matrix model *itself* has a large n limit, and the eigenvalues should converge in distribution to the eigenvalues of the stochastic Airy operator as $n \to \infty$.

When $\sigma = 0$, the stochastic Airy operator in (11.1) specializes to the well-known, non-noisy, Airy operator on $[0, \infty)$ with boundary condition $u(0) = 0$. It has a particularly simple eigendecomposition in terms of the Airy special function,

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}x^2} - x\right) u_i(x) = u_i''(x) - x u_i(x) = \lambda_i u_i(x),$$

which has solutions

$$u_i(x) = \frac{1}{\mathrm{Ai}'(\lambda_i)^2}\,\mathrm{Ai}(x + \lambda_i),$$

where λ_i is the ith root of the Airy function, $\mathrm{Ai}(x)$.

We can discretize the non-noisy Airy operator using finite differences. Taking some mesh size $h = h(n) \to 0$ and defining $x_i = hi$, the matrix

$$A_n = \frac{1}{h^2}\begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -2 \end{bmatrix} - \begin{bmatrix} x_1 & & & & \\ & x_2 & & & \\ & & \ddots & & \\ & & & x_{n-1} & \\ & & & & x_n \end{bmatrix}$$

$$= \frac{1}{h^2} D_n^2 - h\,\mathrm{diag}(1, 2, \ldots, n)$$

is a natural finite difference approximation to the non-noisy Airy operator, *i.e.*, the stochastic Airy operator in (11.1) with $\sigma = 0$. We expect the eigenvalues nearest 0 and the corresponding eigenvectors to converge to the eigenvalues and eigenfunctions of the Airy operator as $n \to \infty$.

The β-Hermite ensemble H_n^β, which is clearly 'noisy', admits a similar representation. There are some manipulations that need to be done to get to that form.

The first step is to obtain the right scaling, focusing on the largest eigenvalue. From Tracy and Widom's result on the distribution of the largest eigenvalue, we know that the largest eigenvalue of

$$\tilde{H}_n^\beta = \frac{\sqrt{2}}{\sqrt{\beta}} n^{1/6} (H_m^\beta - \sqrt{2\beta n} I)$$

converges in distribution as $n \to \infty$ for $\beta = 1, 2, 4$.

Using the approximation $\chi_r \approx \sqrt{r} + \frac{1}{\sqrt{2}} G$, valid for large r, and breaking the matrix into a sum of a non-random part and a random part, it follows that

$$\tilde{H}_n^\beta \approx \begin{bmatrix} -2n^{2/3} & n^{1/6}\sqrt{n-1} & & & \\ n^{1/6}\sqrt{n-1} & -2n^{2/3} & n^{1/6}\sqrt{n-2} & & \\ & \ddots & \ddots & \ddots & \\ & & n^{1/6}\sqrt{2} & -2n^{2/3} & n^{1/6}\sqrt{1} \\ & & & n^{1/6}\sqrt{1} & -2n^{2/3} \end{bmatrix}$$

$$+ \frac{1}{\sqrt{2\beta}} n^{1/6} \begin{bmatrix} 2G & G & & & \\ G & 2G & G & & \\ & \ddots & \ddots & \ddots & \\ & & G & 2G & G \\ & & & G & 2G \end{bmatrix}.$$

Next, replacing $\sqrt{n-i}$ with the first-order Taylor series expansion $\sqrt{n} - \frac{1}{2}n^{-1/2}i$, the following approximation is obtained:

$$\tilde{H}_n^\beta \approx n^{2/3} \begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -2 \end{bmatrix} + \frac{1}{\sqrt{2\beta}} n^{1/6} \begin{bmatrix} 2G & G & & & \\ G & 2G & G & & \\ & \ddots & \ddots & \ddots & \\ & & G & 2G & G \\ & & & G & 2G \end{bmatrix}$$

$$- \frac{1}{2} n^{-1/3} \begin{bmatrix} & 1 & & & \\ 1 & & 2 & & \\ & \ddots & \ddots & \ddots & \\ & & n-2 & & n-1 \\ & & & n-1 & \end{bmatrix}.$$

The first term is a second difference operator, the second term injects noise, and the third term resembles a diagonal multiplication. Introducing $h = n^{-1/3}$ and replacing the second and third terms with analogous diagonal matrices, preserving total variance, the final approximation obtained is:

$$\begin{aligned} \tilde{H}_n^\beta &\approx \frac{1}{h^2} D_n^2 - h \operatorname{diag}(1, 2, \ldots, n) + \frac{2}{\sqrt{\beta}} \frac{1}{\sqrt{h}} \operatorname{diag}(G, G, \ldots, G) \\ &\approx A_n + \frac{2}{\sqrt{\beta}} \frac{1}{\sqrt{h}} \operatorname{diag}(G, G, \ldots, G), \\ h &= n^{-1/3}. \end{aligned}$$

This final approximation appears to be a reasonable discretization of the stochastic Airy operator

$$L^\beta = \frac{\mathrm{d}^2}{\mathrm{d}x^2} - x + \frac{2}{\sqrt{\beta}} \mathrm{d}W, \tag{11.2}$$

with the boundary conditions $f(0) = f(+\infty) = 0$, in which W is Gaussian white noise.

Therefore, the largest eigenvalue distribution of L^β should follow the Tracy–Widom distribution in the cases $\beta = 1, 2, 4$. Figure 11.1 plots the distribution for $\beta = 1, 2, 4$ and compares it to simulation results for $\beta = 1$.

The stochastic operator approach is also advantageous when dealing with 'general β'. The traditional approaches are limited to the cases $\beta = 1, 2, 4$. In the stochastic operator approach, β is related to the variance of the noise; specifically, $\sigma = 2/\sqrt{\beta}$ in the case of the stochastic Airy operator as in (11.2). Instead of working with three discrete values of β, the stochastic operators vary continuously with β. Numerical simulations, as in Figure 11.1, indicate

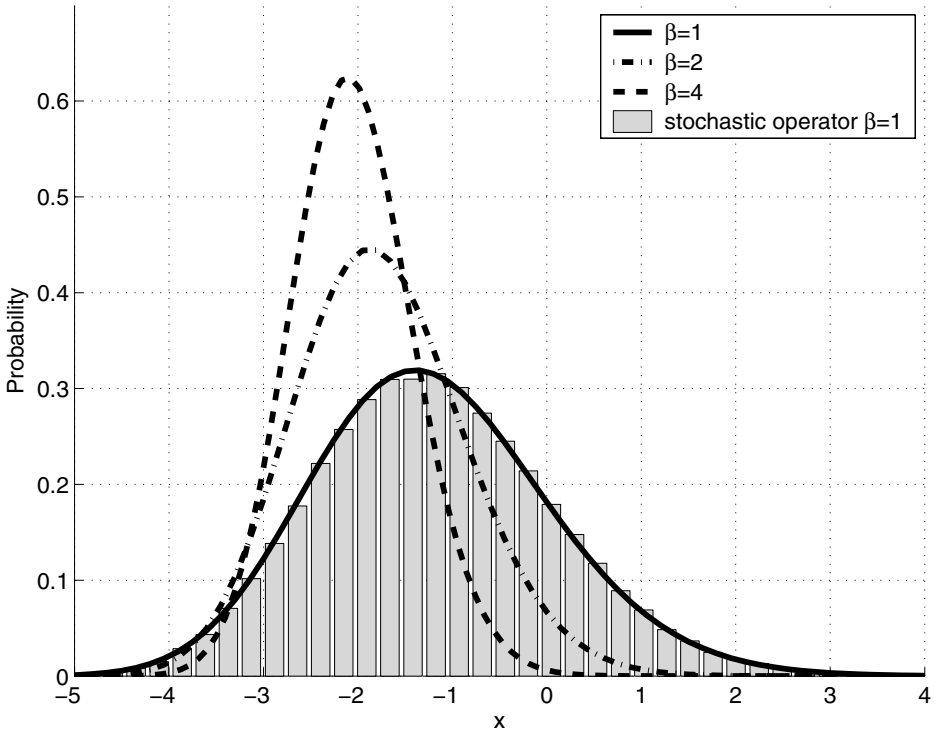

Figure 11.1. The largest eigenvalue distribution: comparison of discretized stochastic Airy operator with the Tracy–Widom law ($\beta = 1$). Monte Carlo simulations involved 10^5 trials of 500-by-500 matrices.

some sort of convection-diffusion process that can be explained in general terms.

The diffusion comes from the high noise associated with small β. Increase the volatility (decrease β) and we increase the range. The convection comes from the repulsion of eigenvalues seen by any perturbation.

The reader can play with a simple experiment to observe the same phenomenon. Consider the 2×2 symmetric random matrix

$$\begin{bmatrix} x & z \\ z & y \end{bmatrix} + \frac{2}{\sqrt{\beta}} \begin{bmatrix} G & 0 \\ 0 & G \end{bmatrix},$$

where the G are independent standard normals. As $\beta \to 0$ the largest eigenvalue will have a larger mean and a larger variance no matter what matrix you start with, *i.e.*, for any choice of x, y, and z.

Similar stochastic operators corresponding to the discretization of the sine and Bessel kernels in Table 9.1 can also be readily derived, as detailed in Sutton (2005).

12. Free probability and infinite random matrices

There is a new mathematical field of 'free probability' emerging as a counterpart to classical probability. Some good references are Voiculescu, Dykema and Nica (1992), Hiai and Petz (2000) and Biane (2003). These references and even the name 'free probability' are worthy of some introduction. The forthcoming book by Speicher and Nica (2005) promises to serve as invaluable resource for making this subject more accessible.

We begin with a viewpoint on classical probability. If we are given probability densities f and g for random variables X and Y respectively, and if we know that X and Y are independent, we can compute the moments of $X+Y$, and XY, for example, from the moments of X and Y.

Our viewpoint on free probability is similar. Given two random matrices, A and B with eigenvalue density f and g, we would like to compute the eigenvalue densities for $A+B$ and AB in terms of the moments of f and g. Of course, A and B do not commute so we are in the realm of noncommutative algebra. Since all possible products of A and B are allowed, we have the 'free' product, *i.e.*, all words in A and B are allowed. (We recall that this is precisely the definition of the free product in algebra.) The theory of free probability allows us to compute the moments of these products in the limit of large matrices, as long as at least one of A or B has what amounts to eigenvectors that are essentially uniformly distributed with Haar measure. Speicher (2003) places these moment computations in an elegant combinatorial context.

We like to think of the difference between classical and free probability as being illustrated by the following maxim:

sum of the eigenvalues of random matrices	(*classical probability*)
versus	
eigenvalues of the sum of random matrices	(*free probability*)

We take a closer look with an example.

Suppose A_i is an $m \times m$ matrix from the Gaussian orthogonal ensemble (GOE). Let λ_i be a random eigenvalue chosen uniformly from the m eigenvalues of A_i.

The classical central limit theorem states that if we form

$$\lambda = \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_n}{\sqrt{n}},$$

no matter what m is, for large n, we obtain a normal distribution. The central limit theorem does not care at all that these λ_is were eigenvalues of random matrices.

However, if rather λ is a random eigenvalue of $A_1 + \cdots + A_n$ (eigenvalue of the sum), then λ is no longer normal. Free probability tells us that as $m, n \to \infty$, the λ follows Wigner's semi-circular density. This is the analogous 'free' central limit theorem for asymptotically large random matrices.

In a broader sense, free probability is studied in the context of non-commutative operator algebras. The synergy between random matrices and free probability arises because matrices are a natural model for a non-commutative algebra. The general theory of free probability is, however, more than just infinite random matrix theory.

In this sense, we find it remarkable that free probabilists were able to derive many of the well-known results in infinite random matrix theory by abstracting away the matrix in question. In special cases, techniques first used by Marčenko and Pastur (1967) and later perfected by Silverstein (1986) and Girko (1998) yield the same results as well. More details on these techniques can be found in Bai (1999) and the references therein.

12.1. Finite free probability

We propose that there is a finite counterpart, which we might call finite free probability. This is an area that is yet to be fully explored but some of the formulas for the moments of AB may be computed using the Jack polynomial theory mentioned in Section 7. There would be a beta dependence that is not necessary when $n = 1$ or $n = \infty$, but otherwise the theory is sensible.

In Figure 12.1, we illustrate (what we call) the finite free central limit theorem for a case when $n = 5$ and $\beta = 2$ (complex random matrices). The answer is neither a semi-circle as in standard free probability or a normal distribution as in classical probability. Here we took 5×5 complex Wishart

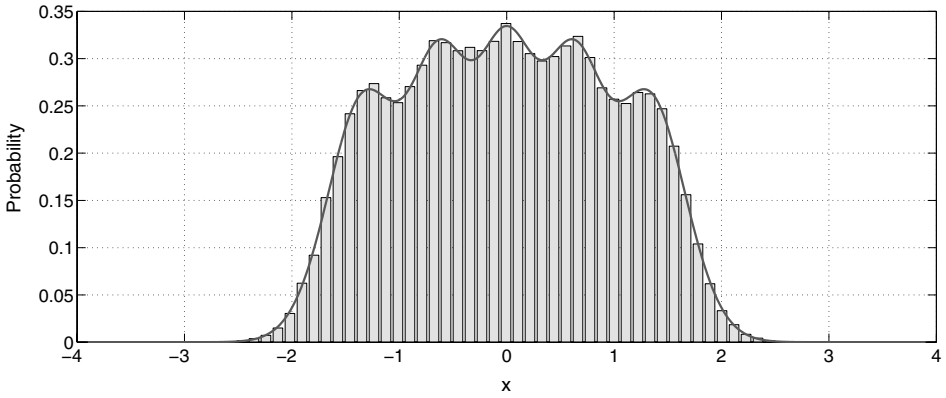

Figure 12.1. Finite free probability: the level density of the $\beta = 2$, $n = 5$ Hermite ensemble obtained by summing a large number of independent realizations of the $\beta = 2$, $n = 5$ Laguerre ensemble.

matrices, subtracted the mean and added them. There is a sensible notion of finite free probability, though it is not clear if finite free cumulants can or do exist. The details have yet to be worked out, though efforts by many authors on different fronts are underway. We invite readers to work in this area.

13. A random matrix calculator

In principle, the formulas from free probability allow us to combine very general combinations of random matrices and still compute the eigenvalue densities. In practice, however, researchers have been constrained from doing so because the relevant theorems are expressed explicitly in terms of transforms that are difficult to compute beyond some simple 'toy examples'.

It turns out that these theorems can be described implicitly as well. The key object is not the transform itself but the algebraic equation that the transform satisfies. The practical implication of this is that we can actually compute the limiting level density and moments for an infinitely large class of random matrices. We label such random matrices as 'characterizable'. Figure 13.1 uses a calculator analogy to describe how one characterizable matrix can be constructed from another.

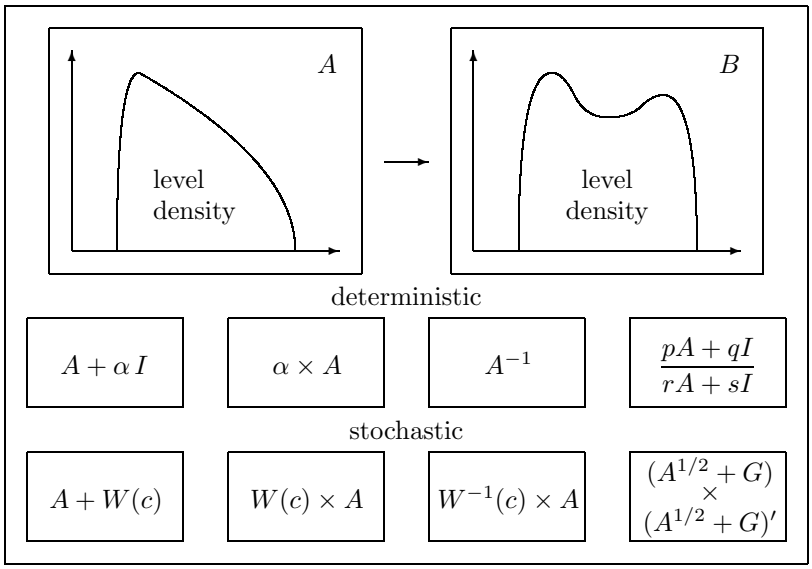

Figure 13.1. A random matrix calculator where a sequence of deterministic and stochastic operations performed on a 'characterizable' matrix A produces a 'characterizable' matrix B. The level density and moments of a 'characterizable' matrix can be computed analytically.

The 'buttons' in the top row of Figure 13.1 represent deterministic operations that can be performed on it (α, p, q, r, s are scalars). The 'buttons' in the bottom row are stochastic operations where additional randomness is injected.

The G matrix is an $m \times n$ matrix with independent, identically distributed (i.i.d.) zero mean elements with a variance of $1/n$ and bounded higher-order moments. We could generate G of this form in MATLAB as

```
G=randn(m,n)/sqrt(n);
```

or

```
G=sign(randn(m,n))/sqrt(n);
```

as examples. The $W(c)$ matrix is a Wishart-like matrix constructed as $W(c) = GG'$ where $m/n \to c > 0$ as $m, n \to \infty$.

The idea behind the calculator is that if we start off with a characterizable matrix A and if we were to generate the matrix B by pressing any of the buttons of the calculator we generate another characterizable matrix B. We can repeat this process forever, and by virtue of it being characterizable we can compute the limiting level density and limiting moments, often in closed form.

We can extend this idea even further by using the theorems of free probability. If we are given two characterizable random matrices, A_1 and A_2, then we can make them 'free' relative to each other by letting $A_2 = QA_2Q'$, where Q is an independent Haar orthogonal/unitary matrix. Then the matrices $A_1 + A_2$, and A_1A_2 are characterizable as well. Other transformations such as $i(A_1A_2 - A_2A_1)$ (the matrix commutator in Lie algebra) are possible as well. The mathematical principles behind this method and the computational realization that makes all of this possible may be found in Rao and Edelman (2005) and Rao (2005). We illustrate this with an example.

Suppose we start off with $A_1 = I$. In MATLAB we perform a sequence of simple transformations corresponding to buttons on our calculator:

```
% Pick n, N1, N2
c1=n/N1; c2=n/N2;
A1=eye(n,n);
```

Then, we generate $A_2 = W_1(c_1) \times A_1$:

```
G1=randn(n,N1)/sqrt(N1);
W1=G1*G1';
A2=A1*W1;
```

Let $A_3 = A_2^{-1}$ and $A_4 = W_2(c_2) \times A_3$:

```
A3=inv(A2);
G2=randn(n,N2)/sqrt(N2);
W2=G2*G2';
A4=A3*W2
```

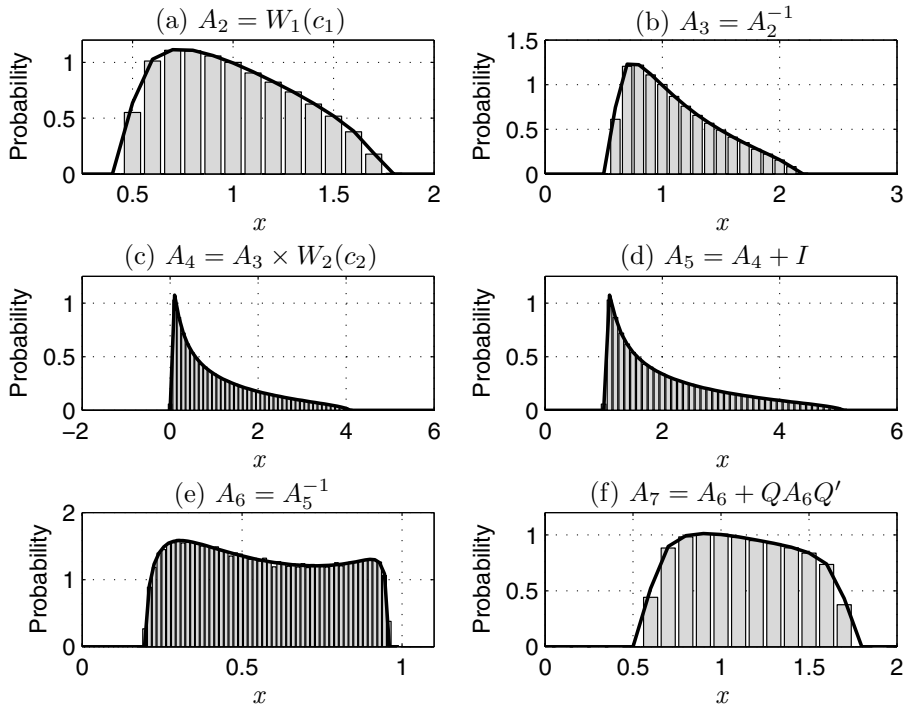

Figure 13.2. Comparison of the theoretical limiting level density (solid line) with the experimental level density for 1000 random matrix ensemble realizations with $c_1 = 0.1$, $c_2 = 0.625$, with $n = 100$, $N_1 = n/c_1 = 1000$ and $N_2 = n/c_2 = 160$.

Now, $A_5 = A_4 + I$ and $A_6 = A_5^{-1}$:

```
A5=A4+eye(n,n);
A6=inv(A5);
```

Generate a Haar unitary matrix and let $A_7 = A_6 + Q\,A_6\,Q'$:

```
[Q,R]=qr(randn(n)+i*randn(n));
Q=Q*diag(exp(2*pi*i*rand(n,1)));
A7=A6+Q*A6*Q';

% Collect eigenvalues
% Repeat over several trials
% Histogram eigenvalues
```

Since we constructed the matrices A_2 to A_7 using the 'buttons' of the random matrix calculator, they turn out to be characterizable. Figure 13.2 shows the limiting level density of these matrix ensembles compared with the experimental version. It is clear that although the predictions were asymptotic in nature (with respect to large n, N_1, N_2) the agreement with

experimental data is excellent. Empirical evidence suggests that a 10×10 matrix is often 'good enough' to corroborate the limiting predictions of free probability.

14. Non-Hermitian and structured random matrices

Our understanding of non-Hermitian and structured random matrices is very limited at present. Relatively recent results on non-Hermitian random matrices include the works by Goldsheid and Khoruzhenko (2000), Fyodorov, Khoruzhenko and Sommers (1997), and Feinberg and Zee (1997).

The most celebrated theorem, Girko's circular law (Girko 1994) states that under reasonable conditions, the eigenvalues of an $n \times n$ matrix with independent entries of mean 0 and variance $1/n$ fall uniformly on a circular disk of radius 1 as $n \to \infty$. Figure 14.1 illustrates this numerically. The theorem is correct whether the matrix is real or complex. When the matrix is real there is a larger attraction of eigenvalues on the real axis and a small repulsion just off the axis. This disappears as $n \to \infty$.

Mehlig and Chalker (2000) study the eigenvectors of such non-Hermitian random matrices. General questions regarding eigenvectors or spacings remain open at this time, as do studies of the β-arbitrary generalization.

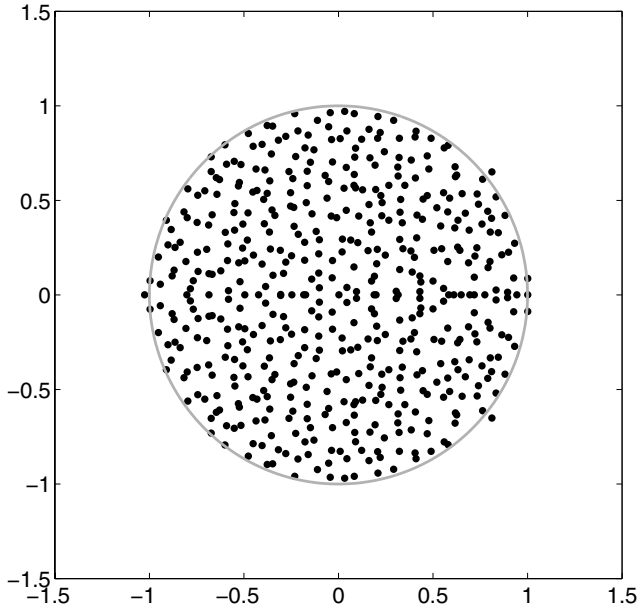

Figure 14.1. The eigenvalues of a 500×500 Gaussian random matrix (`randn(500)/sqrt(500)` in MATLAB) in the complex plane.

The theory of pseudospectra is a rich area that allows for the study of non-Hermitian matrices, specifically those that are highly non-normal. Many tools for drawing pseudospectra are available, such as EigTool by Wright (2000). Figure 14.2 shows the pseudospectra for the same random matrix whose eigenvalues were plotted in Figure 14.1. The Pseudospectra Gateway compiled by Embree and Trefethen (2000) and their well-researched book, Trefethen and Embree (2005), contain discussions of nonsymmetric random matrices.

Random matrix theory is relevant in two distinct ways. An instance of a random matrix itself becomes a valuable object to study, as in Girko's circular law or in the Hatano and Nelson's non-Hermitian Anderson model as described by Trefethen, Contedini and Embree (2001). Also, perturbing a matrix randomly allows for insights into the pseudospectra and has been elevated to the level of a tool, as in the book by Chaitin-Chatelin and Frayssé (1996), where, for example, the Jordan structure is studied.

Another interesting result concerns the probability $p_{n,k}$ that $G_1(n,n)$ has k real eigenvalues. A formula for this may be found in Edelman (1997). Numerical analysts might be interested in the use of the real Schur decomposition in the computation $p_{n,k}$. This is the decomposition used in standard

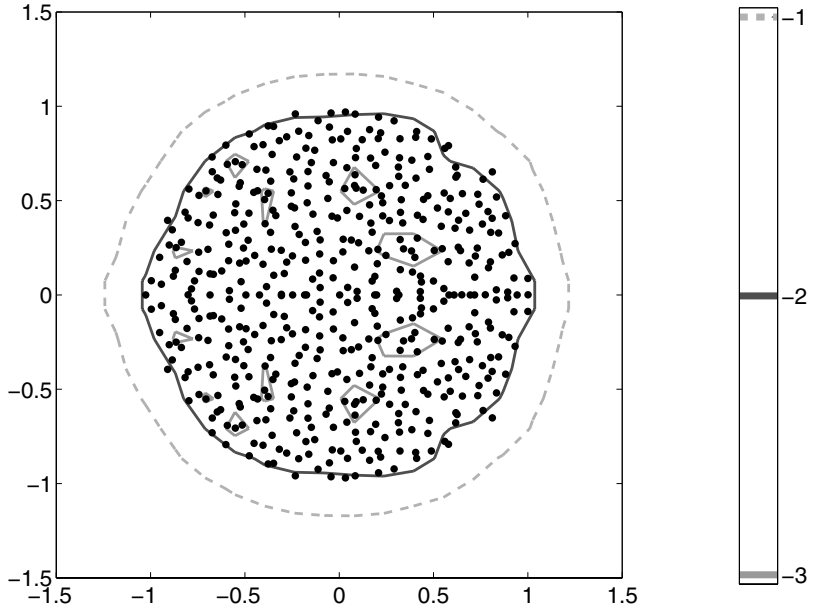

Figure 14.2. The pseudospectra of a 500×500 Gaussian random matrix (`randn(500)/sqrt(500)` in MATLAB). The illustration was produced with the `eigtool` pseudospectra plotter from Oxford University. The values on the colour bar show $10 \log_{10} \epsilon$.

eigenvalue computations. For example, to compute the probability that a matrix has all real eigenvalues, one integrates the measure on $G_1(n, n)$ over matrices of the form $A = QRQ^T$, where Q is orthogonal and R is upper triangular with ordered diagonal elements. This is the Schur form for real matrices with all real eigenvalues.

For random sparse matrices we refer the reader to Rodgers and Bray (1988) and Semerjian and Cugliandolo (2002), and the general theory of random graphs (Bollobás 1985). In Spiridonov (2005) one finds an interesting fractal pattern in the histogram of the eigenvalues of a sparse random matrix depending on the degree of sparsity.

The classical reference on deterministic Toeplitz matrices is Grenander and Szegő (1958). Recent work by Byrc, Dembo and Jiang (2005) provides a free probability-like characterization of the limiting spectral measure of Toeplitz, Hankel and Markov random matrices. Anderson and Zeitouni (2005) discuss central limit theorems related to generalized 'banded' random matrix models.

15. A segue

We make some final predictions about the application of random matrix theory: the pattern will follow that of numerical analysis in general. Most disciplines of science and engineering will find random matrix theory a valuable tool. Random matrix history started in the physics of heavy atoms and multivariate statistics. It has found its way into wireless communications and combinatorial mathematics. The latest field is financial analysis. More will follow; the word has to spread. Hopefully, this is a start.

Acknowledgements

We would like to acknowledge the contributions of Ioana Dumitriu, Brian Sutton, Per-Olof Persson and Plamen Koev. We feel privileged to have them as colleagues and have often been inspired by interacting with them, including while writing this article. There have been many instances when, in the midst of our discussion, we realized that they had said something better than we could ever have; this article thus reflects their words and thoughts on this fascinating subject along with our own. In particular, portions of Ioana's (Dumitriu 2003) and Brian's (Sutton 2005) dissertation work formed the basis for Section 4 and Section 11 respectively. We borrowed the MATLAB code in Section 9 from Persson's beautifully written documentation of the same in Edelman and Persson (2002). We thank Pierre-Antoine Absil, Peter Forrester, Matthew Harding, Nick Higham, Eric Kostlan, Julius Kusuma, Gil Strang, Craig Tracy, and Nick Trefethen for their valuable feedback. Especially, we thank Arieh Iserles and Brad Baxter for their comments and

encouragement, Glennis Starling for being incredibly patient with us in the face of severe time constraints, and Brett Coonley for his typesetting wizardry. We thank the National Science Foundation (DMS-0411962) and the Singapore-MIT Alliance for supporting our work.

REFERENCES

M. Abramowitz and I. Stegun, eds (1970), *Handbook of Mathematical Functions*, Dover Publications, New York.

P.-A. Absil, A. Edelman and P. Koev (2004), 'On the largest principal angle between random subspaces'. Submitted.

D. Aldous and P. Diaconis (1999), 'Longest increasing subsequences: from patience sorting to the Baik–Deift–Johansson theorem', *Bull. Amer. Math. Soc.* **36**, 413–432.

G. W. Anderson and O. Zeitouni (2005), A CLT for a band matrix model: `arXiV.org/math.PR/0412040`.

C. Andréief (1883), 'Note sur une relation les intégrales définies des produits des fonctions', *Mém. de la Soc. Sci. Bordeaux* **2**, 1–14.

J.-M. Azaïs and M. Wschebor (2004), Upper and lower bounds for the tails of the distribution of the condition number of a Gaussian matrix: `www.lsp.ups-tlse.fr/Azais/publi/upper.ps`.

Z. D. Bai (1999), 'Methodologies in spectral analysis of large-dimensional random matrices: a review', *Statist. Sinica* **9**(3), 611–677. With comments by G. J. Rodgers and J. W. Silverstein, and a rejoinder by the author.

B. J. C. Baxter and A. Iserles (2003), On the foundations of computational mathematics, in *Handbook of Numerical Analysis*, Vol. XI, North-Holland, Amsterdam, pp. 3–34.

P. Biane (2003), Free probability for probabilists, in *Quantum probability communications, Vol. XI: Grenoble, 1998*, World Scientific, River Edge, NJ, pp. 55–71.

P. Bleher and A. Its (1999), 'Semiclassical asymptotics of orthogonal polynomials, Riemann–Hilbert problem, and the universality in the matrix model', *Ann. Math.* **150**, 185–266.

B. Bollobás (1985), *Random Graphs*, Academic Press, London.

A. Borodin (1999), 'Longest increasing subsequences of random colored permutations', *Electron. J. Combin.* **6**, # 13 (electronic).

A. Borodin and P. J. Forrester (2003), 'Increasing subsequences and the hard-to-soft edge transition in matrix ensembles', *J. Phys. A* **36**(12), 2963–2981.

W. Byrc, A. Dembo and T. Jiang (2005), Spectral measure of large random Hankel, Markov, and Toeplitz matrices: `arXiV.org/abs/math.PR/0307330`.

M. Caselle and U. Magnea (2004), 'Random matrix theory and symmetric spaces', *Phys. Rep.* **394**(2–3), 41–156.

F. Chaitin-Chatelin and V. Frayssé (1996), *Lectures on Finite Precision Computations: Software, Environments, and Tools*, SIAM, Philadelphia, PA.

A. Constantine (1963), 'Some noncentral distribution problems in multivariate analysis', *Ann. Math. Statist.* **34**, 1270–1285.

P. A. Deift (1999), *Orthogonal Polynomials and Random Matrices: A Riemann–Hilbert Approach*, Vol. 3 of *Courant Lecture Notes in Mathematics*, New York University Courant Institute of Mathematical Sciences, New York.

P. Deift (2000), 'Integrable systems and combinatorial theory', *Notices Amer. Math. Soc.* **47**(6), 631–640.

P. Deift, A. Its and X. Zhou (1997), 'A Riemann–Hilbert problem approach to asymptotic problems arising in the theory of random matrix models, and also in the theory of integrable statistical mechanics', *Ann. Math.* **146**(1), 149–235.

M. Dieng (2004), 'Distribution functions for edge eigenvalues in orthogonal and symplectic ensembles: Painlevé representations':
`arXiV.org/math.PR/0411421`.

I. Dumitriu (2003), Eigenvalue statistics for beta-ensembles, PhD thesis, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA.

I. Dumitriu and A. Edelman (2004), MOPS: Multivariate Orthogonal Polynomials (symbolically): `arXiV.org/abs/math-ph/0409066`.

F. J. Dyson (1963), 'The threefold way: algebraic structures of symmetry groups and ensembles in quantum mechanics', *J. Math. Phys.* **3**, 1199–1215.

A. Edelman (1989), Eigenvalues and condition numbers of random matrices, PhD thesis, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA.

A. Edelman (1997), 'The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law', *J. Multivar. Anal.* **60**, 203–232.

A. Edelman and P.-O. Persson (2002), Numerical methods for random matrices. Technical report, Massachusetts Institute of Technology:
`arXiV.org/abs/math-ph/0501068`.

A. Edelman and B. Sutton (2004), 'Tails of condition number distributions'. Submitted.

M. Embree and L. N. Trefethen (2000), 'Pseudospectra gateway':
`www.comlab.ox.ac.uk/pseudospectra`.

A. Erdélyi, W. Magnus, F. Oberhettinger and F. G. Tricomi (1955), *Higher Transcendental Functions*, Vol. III, McGraw-Hill, New York/Toronto/London.

A. Erdélyi, W. Magnus, F. Oberhettinger and F. G. Tricomi (1981*a*), *Higher Transcendental Functions*, Vol. I, Robert E. Krieger, Melbourne, FL. Reprint of the 1953 original.

A. Erdélyi, W. Magnus, F. Oberhettinger and F. G. Tricomi (1981*b*), *Higher Transcendental Functions*, Vol. II, Robert E. Krieger, Melbourne, FL. Reprint of the 1953 original.

J. Feinberg and A. Zee (1997), 'Non-Hermitian random matrix theory: method of Hermitian reduction', *Nuclear Phys. B* **504**(3), 579–608.

P. J. Forrester (2000), Painlevé transcendent evaluation of the scaled distribution of the smallest eigenvalue in the Laguerre orthogonal and symplectic ensembles. Technical report: `www.lanl.gov`, `arXiV.org/nlin.SI/0005064`.

P. J. Forrester (2005), *Log-Gases and Random Matrices*, book in progress.

P. J. Forrester and N. S. Witte (2004), 'Application of the τ-function theory of Painlevé equations to random matrices: P_{VI}, the JUE, CyUE, cJUE and scaled limits', *Nagoya Math. J.* **174**, 29–114.

Y. Fyodorov, B. A. Khoruzhenko and H.-J. Sommers (1997), 'Almost-Hermitian random matrices: Crossover from Wigner–Dyson to Ginibre eigenvalue statistics', *Phys. Rev. Lett.* **79**, 557–560.

F. P. Gantmacher and M. G. Krein (2002), *Oscillation Matrices and Kernels and Small Vibrations of Mechanical Systems*, revised edn, AMS Chelsea Publishing, Providence, RI. Translation based on the 1941 Russian original.

W. Gautschi (1996), Orthogonal polynomials: Applications and computation, in *Acta Numerica*, Vol. 5, Cambridge University Press, pp. 45–119.

V. L. Girko (1994), 'The circular law: ten years later', *Random Oper. Stoch. Equations* **2**, 235–276, 377–398.

V. L. Girko (1998), *An Introduction to Statistical Analysis of Random Arrays*, VSP, Utrecht. Translated from the Russian.

I. Y. Goldsheid and B. A. Khoruzhenko (2000), 'Eigenvalue curves of asymmetric tridiagonal random matrices', *Electron. J. Probab.* **5**, # 16 (electronic).

G. Golub and W. Kahan (1965), 'Calculating the singular values and pseudo-inverse of a matrix', *SIAM J. Numer. Anal.* **2**, 205–224.

U. Grenander and G. Szegő (1958), *Toeplitz Forms and their Applications*, California Monographs in Mathematical Sciences, University of California Press, Berkeley.

P. J. Hanlon, R. P. Stanley and J. R. Stembridge (1992), Some combinatorial aspects of the spectra of normally distributed random matrices, in *Hypergeometric Functions on Domains of Positivity, Jack Polynomials, and Applications: Tampa, FL, 1991*, Vol. 138 of *Contemp. Math.*, AMS, Providence, RI, pp. 151–174.

F. Hiai and D. Petz (2000), *The Semicircle Law, Free Random Variables and Entropy*, Mathematical Surveys and Monographs, AMS.

A. S. Householder (1958), 'Unitary triangularization of a nonsymmetric matrix', *J. Assoc. Comput. Mach.* **5**, 339–342.

A. Its, C. A. Tracy and H. Widom (2001), 'Random words, Toeplitz determinants and integrable systems II', *Physica* **152–153D**, 199–224.

D. A. Ivanov (2002), Random-matrix ensembles in p-wave vortices, in *Proc. Dresden Workshop 'Vortices in Unconventional Superconductors and Superfluids'*, Springer, Heidelberg: `arXiV.org/abs/cond-mat/0103089`.

H. Jack (1970), 'A class of symmetric polynomials with a parameter', *Proc. R. Soc. Edinburgh* **69**, 1–18.

A. T. James (1964), 'Distributions of matrix variates and latent roots derived from normal samples', *Ann. Math. Stat.* **35**, 475–501.

K. Johansson (2000*a*), 'Random growth and random matrices', *European Congress of Mathematics* **I**, 445–456.

K. Johansson (2000*b*), 'Shape fluctuations and random matrices', *Comm. Math. Phys.* **209**, 437–476.

I. M. Johnstone (2001), 'On the distribution of the largest eigenvalue in principal components analysis', *Ann. Statist.* **29**(2), 295–327.

D. Jonsson (1982), 'Some limit theorems for the eigenvalues of a sample covariance matrix', *J. Multivar. Anal.* **12**, 1–38.
I. Karatzas and S. E. Shreve (1991), *Brownian Motion and Stochastic Calculus*, Vol. 113 of *Graduate Texts in Mathematics*, second edn, Springer, New York.
R. Killip and I. Nenciu (2004), 'Matrix models for circular ensembles', *Int. Math. Research Notices* **50**, 2665–2701.
F. Knop and S. Sahi (1997), 'A recursion and a combinatorial formula for Jack polynomials', *Invent. Math.* **128**(1), 9–22.
P. Koev (2002), Accurate and efficient matrix computation with structured matrices, PhD thesis, University of California, Berkeley.
P. Koev and A. Edelman (2004), The efficient evaluation of the hypergeometric function of a matrix argument. Preprint.
A. B. J. Kuijlaars (2000), 'Which eigenvalues are found by the Lanczos method?', *SIAM J. Matrix Anal. Appl.* **22**(1), 306–321 (electronic).
A. B. J. Kuijlaars (2003), Riemann–Hilbert analysis for orthogonal polynomials, in *Orthogonal Polynomials and Special Functions: Leuven, 2002*, Vol. 1817 of *Lecture Notes in Mathematics*, Springer, Berlin, pp. 167–210.
A. Kuijlaars and K. T.-R. McLaughlin (2000), 'Generic behavior of the density of states in random matrix theory and equilibrium problems in the presence of real analytic external fields', *Comm. Pure Appl. Math.* **53**, 736–785.
R. A. Lippert (2003), 'A matrix model for the β-Jacobi ensemble', *J. Math. Phys.* **44**(10), 4807–4816.
I. Macdonald (1982), 'Some conjectures for root systems', *SIAM J. Math. Anal.* **13**, 998–1004.
I. Macdonald (1998), *Symmetric Functions and Hall Polynomials*, Oxford Mathematical Monographs, 2nd edn, Oxford University Press.
B. D. McKay (1981), 'The expected eigenvalue distribution of a large regular graph', *Linear Algebra Appl.* **40**, 203–216.
V. Marčenko and L. Pastur (1967), 'Distribution of eigenvalues for some sets of random matrices', *Math USSR Sbornik* **1**, 457–483.
A. Mathai (1997), *Jacobians of Matrix Transformations and Functions of Matrix Arguments*, World Scientific, Singapore.
B. Mehlig and J. T. Chalker (2000), 'Statistical properties of eigenvectors in non-Hermitian Gaussian random matrix ensembles', *J. Math. Phys.* **41**(5), 3233–3256.
M. L. Mehta (1991), *Random Matrices*, second edn, Academic Press, Boston.
V. D. Milman and G. Schechtman (1986), *Asymptotic Theory of Finite Dimensional Normed Spaces*, Vol. 1200 of *Lecture Notes in Mathematics*, Springer.
R. J. Muirhead (1982), *Aspects of Multivariate Statistical Theory*, Wiley, New York.
A. M. Odlyzko (2001), 'Tables of zeros of the Riemann zeta function': `www.dtc.umn.edu/~odlyzko/zeta_tables/index.html`.
A. Odlyzko and E. Rains (1998), On longest increasing subsequences in random permutations. Technical report, AT&T Laboratories.
I. Olkin (1953), 'Note on "The Jacobians of certain matrix transformations useful in multivariate analysis"', *Biometrika* **40**, 43–46.
I. Olkin (2002), 'The 70th anniversary of the distribution of random matrices: a survey', *Linear Algebra Appl.* **354**, 231–243.

E. M. Rains (1998), 'Increasing subsequences and the classical groups', *Electron. J. Combin.* **5**, # 12 (electronic).
N. R. Rao (2005), Infinite random matrix theory for multi-channel signal processing, PhD thesis, Massachusetts Institute of Technology.
N. R. Rao and A. Edelman (2005), The polynomial method for random matrices. Preprint.
G. J. Rodgers and A. J. Bray (1988), 'Density of states of a sparse random matrix', *Phys. Rev. B* (3) **37**(7), 3557–3562.
A. Sankar (2003), Smoothed analysis of Gaussian elimination, PhD thesis, Massachusetts Institute of Technology.
A. Sankar, D. A. Spielman and S.-H. Teng (2004), Smoothed analysis of the condition numbers and growth factors of matrices:
`arXiV.org/abs/cs.NA/0310022`.
G. Semerjian and L. F. Cugliandolo (2002), 'Sparse random matrices: the eigenvalue spectrum revisited', *J. Phys. A* **35**(23), 4837–4851.
J. Shen (2001), 'On the singular values of Gaussian random matrices', *Linear Algebra Appl.* **326**(1–3), 1–14.
J. W. Silverstein (1985), 'The smallest eigenvalue of a large-dimensional Wishart matrix', *Ann. Probab.* **13**(4), 1364–1368.
J. W. Silverstein (1986), Eigenvalues and eigenvectors of large dimensional sample covariance matrices, chapter in *Random Matrices and their Applications*, AMS, Providence, RI, pp. 153–159.
A. Soshnikov (1999), 'Universality at the edge of the spectrum in Wigner random matrices', *Comm. Math. Phys.* **207**, 697–733.
J. Spanier and K. B. Oldham (1987), *An Atlas of Functions*, Taylor & Francis/ Hemisphere.
R. Speicher (2003), Free probability theory and random matrices, in *Asymptotic Combinatorics with Applications to Mathematical Physics: St. Petersburg, 2001*, Vol. 1815 of *Lecture Notes in Mathematics*, Springer, Berlin, pp. 53–73.
R. Speicher and A. Nica (2005), Combinatorics of free probability. In preparation.
A. N. Spiridonov (2005), Spectra of sparse graphs and matrices. Preprint.
R. P. Stanley (1989), 'Some combinatorial properties of Jack symmetric functions', *Adv. Math.* **77**, 76–115.
G. W. Stewart (1980), 'The efficient generation of random orthogonal matrices with an application to condition estimators', *SIAM J. Numer. Anal.* **17**(3), 403–409 (loose microfiche suppl).
B. D. Sutton (2005), The stochastic operator approach to random matrix theory, PhD thesis, Massachusetts Institute of Technology, Department of Mathematics.
K. Takasaki (2000), 'Painlevé equations', Soliton Laboratory, Chronology of Mathematics:
`www.math.h.kyoto-u.ac.jp/~takasaki/soliton-lab/chron/painleve.html`.
C. A. Tracy and H. Widom (1993), Introduction to random matrices, in *Geometric and Quantum Aspects of Integrable Systems* (G. Helminck, ed.), Vol. 424 of *Lecture Notes in Physics*, Springer, Berlin, pp. 103–130.
C. A. Tracy and H. Widom (1994*a*), 'Fredholm determinants, differential equations and matrix models', *Comm. Math. Phys.* **163**, 33–72.

C. A. Tracy and H. Widom (1994*b*), 'Level-spacing distributions and the Bessel kernel', *Comm. Math. Phys.* **161**, 289–310.

C. A. Tracy and H. Widom (1998), 'Correlation functions, cluster functions and spacing distributions for random matrices', *J. Statist. Phys.* **92**, 809–835.

C. A. Tracy and H. Widom (2000*a*), The distribution of the largest eigenvalue in the Gaussian ensembles, in *Calogero–Moser–Sutherland Models*, Vol. 4 of *CRM Series in Mathematical Physics* (J. van Diejen and L. Vinet, eds), Springer, Berlin, pp. 461–472.

C. A. Tracy and H. Widom (2000*b*), Universality of the distribution functions of random matrix theory II, in *Integrable Systems: From Classical to Quantum* (J. Harnad, G. Sabidussi and P. Winternitz, eds), Vol. 26 of *CRM Proceedings & Lecture Notes*, AMS, Providence, pp. 251–264.

C. A. Tracy and H. Widom (2001), 'On the distributions of the lengths of the longest monotone subsequences in random words', *Probab. Theory Rel. Fields* **119**, 350–380.

L. N. Trefethen and D. Bau, III (1997), *Numerical Linear Algebra*, SIAM, Philadelphia, PA.

L. N. Trefethen and A. Embree (2005), *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*, Princeton.

L. N. Trefethen and R. S. Schreiber (1990), 'Average-case stability of Gaussian elimination', *SIAM J. Matrix Anal. Appl.* **11**(3), 335–360.

L. N. Trefethen, M. Contedini and M. Embree (2001), 'Spectra, pseudospectra, and localization for random bidiagonal matrices', *Comm. Pure Appl. Math.* **54**(5), 595–623.

H. F. Trotter (1984), 'Eigenvalue distributions of large Hermitian matrices; Wigner's semi-circle law and a theorem of Kac, Murdock, and Szegő', *Adv. Math.* **54**, 67–82.

C. F. Van Loan (2000), 'The ubiquitous Kronecker product', *J. Comput. Appl. Math.* **123**(1–2), 85–100.

P. van Moerbeke (2001), Integrable lattices: random matrices and random permutations, in *Random Matrices and their Applications*, MSRI Publications, Cambridge University Press, Cambridge.

J. M. Varah (1993), 'The prolate matrix', *Linear Algebra Appl.* **187**, 269–278.

D. Viswanath and L. N. Trefethen (1998), 'Condition numbers of random triangular matrices', *SIAM J. Matrix Anal. Appl.* **19**(2), 564–581 (electronic).

D. V. Voiculescu, K. J. Dykema and A. Nica (1992), *Free Random Variables*, AMS, Providence, RI.

J. von Neumann (1963), *Collected Works*, Vol. V: *Design of Computers, Theory of Automata and Numerical Analysis* (A. H. Taub, ed.), Pergamon Press, Macmillan, New York.

J. von Neumann and H. H. Goldstine (1947), 'Numerical inverting of matrices of high order', *Bull. Amer. Math. Soc.* **53**, 1021–1099.

E. W. Weisstein (2005), 'Special function', From MathWorld, A Wolfram Web Resource: `http://mathworld.wolfram.com/SpecialFunction.html`

E. P. Wigner (1955), 'Characteristic vectors of bordered matrices with infinite dimensions', *Ann. Math.* **62**, 548–564.

E. P. Wigner (1958), 'On the distribution of the roots of certain symmetric matrices', *Ann. Math.* **67**, 325–327.

J. Wishart (1928), 'The generalized product moment distribution in samples from a normal multivariate population', *Biometrika* **20A**, 32–52.

T. G. Wright (2000), 'EigTool':
`www.comlab.ox.ac.uk/projects/pseudospectra/eigtool`.

M.-C. Yeung and T. F. Chan (1997), 'Probabilistic analysis of Gaussian elimination without pivoting', *SIAM J. Matrix Anal. Appl.* **18** (2), 499–517.

M. R. Zirnbauer (1996), 'Riemannian symmetric superspaces and their origin in random-matrix theory', *J. Math. Phys.* **37**(10), 4986–5018.