Random Matrix Theory, Numerical Computation and Applications

Alan Edelman, Brian D. Sutton, and Yuyang Wang

Abstract. This paper serves to prove the thesis that a computational trick can open entirely new approaches to theory. We illustrate by describing such random matrix techniques as the stochastic operator approach, the method of ghosts and shadows, and the method of "Riccatti Diffusion/Sturm Sequences," giving new insights into the deeper mathematics underneath random matrix theory.

Contents

1. Introduction: A Computational Trick Can Also Be a Theoretical Trick

We advise mathematicians not to dismiss an efficient computation as mere "implementation details", it may be where the next theory comes from. This note will supply examples (real case outlined in Table 1). (Throughout the notes, MATLAB codes are in typewriter font. In Table 1, `trideig` and `maxeig` can be downloaded from [**Per**].)

We start with the famous semicircle distribution $f(x) = \frac{1}{2\pi}\sqrt{4-x^2}$; illustrated at the bottom (a) of Algorithm 1. This distribution depicts the histogram of the

Key words and phrases. Random Matrix Theory, Numerical Linear Algebra, Stochastic Operator, Ghosts and Shadows.

The first author was supported in part by DMS 1035400 and DMS 1016125. Note to our readers: These notes are in large part a precursor to a book on Random Matrix Theory that will be forthcoming. We reserve the right to reuse materials in the book. All codes were tested in MATLAB2012a.

	RMT Laws	**Naive Computation**	**Clever Computational Tricks**	
All eigs	*Semicircle Law*	`A=randn(n)` `v=eig((A+A')/sqrt(2*n))`	`A=sqrt(chi2rnd((n-1):-1:1))` `v=trideig(randn(n,1),A)`	*Tridiagonal models (2.3)*
		Space: $\mathcal{O}(n^2)$	$\mathcal{O}(\boldsymbol{n})$	
		Time: $\mathcal{O}(n^3)$	$\mathcal{O}(\boldsymbol{n^2})$	

Max eig	*Tracy-Widom Law*	`A=randn(n)`
		`vs=eig((A+A')/sqrt(2*n))`
		`v=max(vs)`
			`k=round(n-10*n^(1/3)-1)`
			`A=sqrt(chi2rnd((n-1):-1:k))`
			`v=maxeig(randn(n-k+1,1),A)`
				Truncated Storage, Bisection, Sturm Sequence, Sparse Eigensolver

		Space: $\mathcal{O}(n^2)$	$\mathcal{O}(\boldsymbol{10\,n^{1/3}})$	
		Time: $\mathcal{O}(n^3)$	$\mathcal{O}(\boldsymbol{(10\,n^{1/3})^2})$	
Theories Inspired by Computation		Tridiagonal and Bidiagonal models (Section 2) Stochastic Operators (Section 3) Sturm sequence and Ricatti difussion (Section 4) Method of Ghosts and Shadows (Section 5)		

TABLE 1. A Computational Trick Can Also Be a Theoretical Trick.

n eigenvalues of a symmetric random $n \times n$ matrix $S = (A^T + A)/\sqrt{2n}$ obtained by symmetrizing a matrix whose elements follow the standard normal distribution, i.e., in MATLAB notation: `A=randn(n)`.

The complex version starts with `A = randn(n) + sqrt(-1)*randn(n)` and forms $(A^H + A)/2\sqrt{n}$ to get the semicircle law. The *Tracy-Widom* distribution (illustrated in Algorithm 1 bottom (b)) describes the normalized largest eigenvalue, which, in the complex case, is

$$f(x) = \frac{\mathrm{d}}{\mathrm{d}x} \exp\left(- \int_x^\infty (t-x) q(t)^2 \mathrm{d}t \right), \tag{1.1}$$

where $q(t)$ is the solution of a so-called Painlevé II differential equation $\ddot{q}(t) = tq(t) + 2q(t)^3$, with the boundary condition that as $t \to \infty$, $q(t)$ is asymptotic to the Airy function $Ai(t)$. Algorithm 1 shows Monte Carlo experiments for the semicircle law and the Tracy-Widom law.

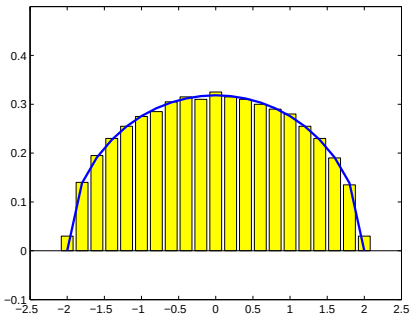
We recommend Bornemann's code as the current best practice for computing the Tracy-Widom density $f(x)$ [**Bor10**]. Alternatively, we present a simpler method in Algorithm 2, showing that even the formidable is but a few lines of MATLAB.

The semicircle and Tracy-Widom laws are theorems as $n \to \infty$ but computations for small n suffice for illustration. The real S is known as the Gaussian Orthogonal Ensemble (GOE) and the "complex S" the Gaussian Unitary Ensemble (GUE). In general, they are instances of β-Hermite ensemble where $\beta = 1, 2$ correspond to the real and complex cases respectively.

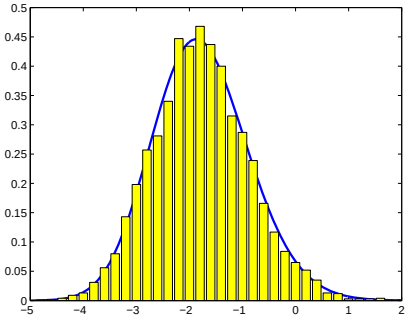
As we can see in Algorithm 1, direct random matrix experiments usually involve calculating the eigenvalues of random matrices, i.e. `eig(s)`. Since many linear algebra computations require $\mathcal{O}(n^3)$ operations, it seems more feasible to take n relatively small, and take a large number of Monte Carlo instances. This is our strategy in Algorithm 1.

Algorithm 1 Semicircle Law ($\beta = 1$) and the Tracy-Widom distribution ($\beta = 2$)

```
%Experiment:   Demonstration of Semicircle and Tracy-Widom distribution
%Plot:         Histogram of the eigenvalues and the largest eigenvalue
%Theory:       Semicircle and Tracy-Widom as n->infinity;
%% Parameters
n=100;          % matrix size
t=5000;         % trials
v=[];           % eigenvalue samples
vl=[];          % largest eigenvalue samples
dx=.2;          % binsize
%% Experiment
for i=1:t
    %% Sample GOE and collect their eigenvalues
    a=randn(n);    % n by n matrix of random Gaussians
    s=(a+a')/2;    % symmetrize matrix
    v=[v;eig(s)]; % eigenvalues
    %% Sample GUE and collect their largest eigenvalues
    a=randn(n)+sqrt(-1)*randn(n);     % random nxn complex matrix
    s=(a+a')/2;           % Hermitian matrix
    vl=[vl;max(eig(s))];  % Largest Eigenvalue
end
%% Semicircle law
v=v/sqrt(n/2);     % normalize eigenvalues
% Plot
[count,x]=hist(v,-2:dx:2);
bar(x, count/(t*n*dx),'y')
% Theory
hold on
plot(x,sqrt(4-x.^2)/(2*pi),'LineWidth',2)
axis([-2.5 2.5 -.1 .5])
hold off
%% Tracy-Widom distribution
vl=n^(1/6)*(vl-2*sqrt(n));    % normalized eigenvalues
% Plot
figure; [count,x]=hist(vl,-5:dx:2);
bar(x, count/(t*dx),'y')
% Theory
hold on
tracywidom
```



(a) Semicircle Law



(b) Tracy-Widom distribution

Algorithm 2 Tracy-Widom distribution ($\beta = 2$)

```
%Theory:       Compute and plot the Tracy-Widom distribution
%%Parameters
t0=5;          % right endpoint
tn=-8;         % left  endpoint
dx=.005;       % discretization
%%Theory: The differential equation solver
deq=@(t,y) [y(2); t*y(1)+2*y(1)^3; y(4); y(1)^2];
opts=odeset('reltol',1e-12,'abstol',1e-15);
y0=[airy(t0);airy(1,t0);0;airy(t0)^2]; % boundary conditions
[t,y]=ode45(deq,t0:-dx:tn,y0,opts);    % solve
F2=exp(-y(:,3));                       % the distribution
f2=gradient(F2,t);                     % the density
%% Plot
plot(t,f2,'LineWidth',2)
axis([-5 2 0 .5])
```

In fact, sophisticated matrix computations involve a series of reductions. With normally distributed matrices, the most expensive reduction steps can be avoided on the computer as they can be done with mathematics! All of a sudden $\mathcal{O}(n^3)$ computations become $\mathcal{O}(n^2)$ or even better. The resulting matrix requires less storage either using sparse formulas or data structures with even less overhead.

The story gets better. Random matrix experiments involving complex numbers or even over the quaternions reduce to real matrices even before they need to be stored on a computer.

The story gets even better yet. On one side, for finite n, the reduced form leads to the notion of a "ghost" random matrix quantity that exists for every β (not only real, complex and quaternions), and a "shadow" quantity which may be real or complex which allows for computation. On the other hand, the reduced forms connect random matrices to the continuous limit, *stochastic operators*, which in some ways represent a truer view of why random matrices behave as they do.

The rest of the notes is organized as follows. In Chapter 2, we prepare our readers with preliminaries of matrix factorization for random matrices. In Chapter 3, stochastic operator is introduced with applications and we discuss Sturm sequences and Ricatti diffusion in Chapter 4. We introduce "ghost" and "shadow" techniques for random matrices in Chapter 5. The final chapter is devoted to the smallest singular value of `randn(n)`.

Note: It has now been eight years since the first author has written a large survey for Acta Numerica [**ER05**], and two years since the applications survey [**EW13**]. This survey is meant to be different as we mean to demonstrate the thesis in the very name of this section.

2. Random Matrix Factorization

In this section, we will provide the details of matrix reductions that do not require a computer. Then, we derive the reduced forms of β-Hermite and β-Laguerre ensembles, which is summarized in Table 2 and Table 3 shows how to generate them in sparse formula. Later this section, we give an overview of how these reductions lead to various computational and theoretical impact.

Ensemble	Matrices	Numeric	Models	MATLAB ($\beta = 1$)
Hermite	Wigner	`eig`	Tridiagonal (2.3)	`g = randn(n,n); H=(g+g')/2;`
Laguerre	Wishart	`svd`	Bidiagonal (2.4)	`g = randn(m,n); L=(g'*g)/m;`

TABLE 2. Hermite and Laguerre ensembles.

Ensemble	MATLAB commands (Statistics Toolbox required)

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);`

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';`

TABLE 3. Generating the Hermite and Laguerre ensembles as sparse matrices.

2.1. The Chi-distribution and orthogonal invariance. There are two key facts to know about a vector of independent standard normals. Let $\boldsymbol{v}_n$ denote such a vector. In MATLAB this would be `randn(n,1)`. Mathematically, we say that the n elements are iid standard normals (i.e., mean 0, variance 1).

- **Chi distribution**: the Euclidean length $\|\boldsymbol{v}_n\|$, which is the square root of the sum of the n squares of Gaussians, has what is known as the χ_n distribution.
- **Orthogonal invariance**: for any fixed orthogonal matrix Q, or if Q is random and independent of $\boldsymbol{v}_n$, the distribution of $Q\boldsymbol{v}_n$ is identical to that of $\boldsymbol{v}_n$. In other words, it is impossible to tell the difference between a computer generated $\boldsymbol{v}_n$ or $Q\boldsymbol{v}_n$ upon inspecting only the output. It is easy to see that the density of $\boldsymbol{v}_n$ is $(2\pi)^{-\frac{n}{2}} e^{-\frac{\|\boldsymbol{v}_n\|^2}{2}}$ which only depends on the length of $\boldsymbol{v}_n$.

We shall see that these two facts allow us to transform matrices involving standard normals to simpler forms.

For reference, we mention that the χ_n distribution has the probability density

$$f(x) = \frac{x^{n-1}e^{-x^2/2}}{2^{n/2-1}\Gamma(n/2)}.$$

Notice that there is no specific requirement that n be an integer, despite our original motivation as the length of a Gaussian vector. The square of χ_n is the distribution that underlies the well-known *Chi-squared test*. It can be seen that the mean of χ_n^2 is n. For integers, it is the sum of the n standard normal variables. We have

that $\boldsymbol{v}_n$ is the product of the random scalar χ_n, which serves as the length, and an independent vector that is uniform on the sphere, which serves as the direction.

2.2. The QR decomposition of `randn(n)`. Given a vector $\boldsymbol{v}_n$, we can readily construct an orthogonal reflection or rotation H_n such that $H_n\boldsymbol{v}_n = \pm\|\boldsymbol{v}_n\|\boldsymbol{e}_1$, where $\boldsymbol{e}_1$ denotes the first column of the identity. We do this using the standard technique of *Householder transformations* [**TB97**] (see Lec. 10) in numerical linear algebra, which is a reflection across the external angle bisector of these two vectors. In this case, $H_n = \mathbb{I} - 2\frac{\boldsymbol{w}\boldsymbol{w}^T}{\boldsymbol{w}^T\boldsymbol{w}}$ where $\boldsymbol{w} = \boldsymbol{v_n} \pm \|\boldsymbol{v}_n\|\boldsymbol{e}_1$.

Therefore, if $\boldsymbol{v}_n$ follows a multivariate standard normal distribution, $H_n\boldsymbol{v}_n$ yields a Chi distribution for the first element and 0 otherwise. Furthermore, let `randn(n)` be an $n \times n$ matrix of iid standard normals. It is easy to see now that through successive Householder reflections of size $n, n-1, \ldots, 1$ we can orthogonally transform `randn(n)` into the upper triangular matrix

$$H_1H_2\cdots H_{n-1}H_n \times \texttt{randn(n)} = R_n = \begin{pmatrix} \chi_n & \mathrm{G} & \cdots & \mathrm{G} & \mathrm{G} \\ & \chi_{n-1} & \cdots & \mathrm{G} & \mathrm{G} \\ & & \ddots & \vdots & \vdots \\ & & & \chi_2 & \mathrm{G} \\ & & & & \chi_1 \end{pmatrix}.$$

Here all elements are independent and represent a distribution and each G is an iid standard normal. It is helpful to watch a 3×3 matrix turn into R_3:

$$\begin{pmatrix} \mathrm{G} & \mathrm{G} & \mathrm{G} \\ \mathrm{G} & \mathrm{G} & \mathrm{G} \\ \mathrm{G} & \mathrm{G} & \mathrm{G} \end{pmatrix} \to \begin{pmatrix} \chi_3 & \mathrm{G} & \mathrm{G} \\ 0 & \mathrm{G} & \mathrm{G} \\ 0 & \mathrm{G} & \mathrm{G} \end{pmatrix} \to \begin{pmatrix} \chi_3 & \mathrm{G} & \mathrm{G} \\ 0 & \chi_2 & \mathrm{G} \\ 0 & 0 & \mathrm{G} \end{pmatrix} \to \begin{pmatrix} \chi_3 & \mathrm{G} & \mathrm{G} \\ 0 & \chi_2 & \mathrm{G} \\ 0 & 0 & \chi_1 \end{pmatrix}.$$

The Gs as the computation progresses are not the same numbers, but merely indicating that the distributions remain unchanged. With a bit of care we can say that

$$\texttt{randn(n)} = (\text{orthogonal uniform with Haar measure}) \cdot R_n$$

is the QR decomposition of `randn(n)`. Notice that in earlier versions of LAPACK and MATLAB `[Q, R]=qr(randn(n))` did not always yield Q with Haar measure. Random matrix theory provided the impetus to fix this!

One immediate consequence is the following interesting fact

$$\mathbb{E}[\mathbf{det}[\texttt{randn(n)}]^2] = n!. \tag{2.1}$$

2.3. The tridiagonal reduction of the GOE. Eigenvalues are usually introduced for the first time as the roots of the characteristic polynomial. Many people just assume that this is the definition that is used during a computation, but it is well-established that this is not a good method for computing eigenvalues. Rather, a matrix factorization is used. In the case that S is symmetric, an orthogonal matrix Q is found such that $Q^TSQ = \Lambda$ is diagonal. The columns of Q are the eigenvectors and the diagonal of Λ are the eigenvalues.

Mathematically, the construction of Q is an iterative procedure, requiring infinitely many steps to converge. In practice, S is first tridiagonalized through a finite process which usually takes the bulk of the time. The tridiagonal is then iteratively diagonalized. Usually, this tridiagonal to diagonal step takes a negligible amount of time to converge in finite precision.

Suppose $A =$ `randn(n)` and $S = (A + A^T)/\sqrt{2}$, we can tridiagonalize S with the finite Householder procedure (see [**TB97**] for general algorithms). The result [**DE02**] is

$$T_n = \begin{pmatrix} \mathrm{G}\sqrt{2} & \chi_{n-1} & & & \\ \chi_{n-1} & \mathrm{G}\sqrt{2} & \chi_{n-2} & & \\ & \ddots & \ddots & \ddots & \\ & & \chi_2 & \mathrm{G}\sqrt{2} & \chi_1 \\ & & & \chi_1 & \mathrm{G}\sqrt{2} \end{pmatrix}, \tag{2.2}$$

where $\mathrm{G}\sqrt{2}$ refers to a Gaussian with mean 0 and variance 2. The superdiagonal and diagonal are independent, as the matrix is symmetric. The matrix T_n has the same eigenvalue distribution as S, but numerical computation of the eigenvalues is considerably faster when the right software is used, for example, LAPACK's DSTEQR or DSTEBZ (bisection). The largest eigenvalue benefits further as we only need to build around a $10n^{1/3} \times 10n^{1/3}$ matrix ad we can input an estimate for the largest eigenvalues such as $\lambda_{\max} = 2$. See Section 2.5 for details.

A dense eigensolver requires $\mathcal{O}(n^3)$ operations and will spend nearly all of its time constructing T_n. Given that we know the distribution for T_n a priori, this is wasteful. The eigenvalues of T_n require $\mathcal{O}(n^2)$ time or better. In addition, a dense matrix requires $\mathcal{O}(n^2)$ storage while the tridiagonal matrix only needs $\mathcal{O}(n)$.

2.4. Bidiagonal reduction of Real Wishart Matrices. Suppose `A=randn(m,n)`, $W = A^T A/m$ is called the Wishart matrix or Laguerre ensemble ($\beta = 1$). Computing its eigenvalues amounts to calculating the singular values of A. For that purpose, we need to reduce A to lower bidiagonal form [**TB97**] (Lec. 31) (shown here for $n > m$),

$$B_n = \begin{pmatrix} \chi_{n-1} & & & & & \\ \chi_{m-1} & \chi_{n-1} & & & & \\ & \ddots & \ddots & & & \\ & & \chi_2 & \chi_{n-m+2} & & \\ & & & \chi_1 & \chi_{n-m+1} \end{pmatrix}.$$

See [**Sil85**] and [**Tro84**] for details. Computation of singular values is greatly accelerated in bidiagonal form when using, for example, LAPACK's DBDSQR.

2.5. Superfast computation. Most of earlier numerical experiments computed the eigenvalues of random matrices and then histogrammed them. Can we histogram without histogramming? The answer is Yes! *Sturm sequences* can be used with T_n for the computation of histograms [**ACE08**]. This is particularly valuable when there is interest in a relatively small number of histogram intervals (say 20 or 30) and n is very large. This is an interesting idea, particularly because most people think that histogramming eigenvalues first requires that they compute the eigenvalues, then sort them into bins. The Sturm sequence [**TB97**] idea gives a count without computing the eigenvalues at all. This is a fine example of not computing more than is needed: if you only need a count, why should one compute the eigenvalues at all? We will further discuss Sturm sequence in Section 4.

For the largest eigenvalue, the best trick for very large n is to only generate the upper left $10\,n^{1/3} \times 10\,n^{1/3}$ of the matrix. Because of what is known as the "Airy

decay" in the corresponding eigenvector, the largest eigenvalue, which technically depends on every element in the tridiagonal matrix — numerically depends significantly only on the upper left part. This is a huge savings in Monte Carlo sampling. Further savings can be obtained by using the Lanczos "shift and invert" strategy given an estimate for the largest eigenvalue. Similar ideas may be used for singular values. We refer interested reads to Section 10 of [**ER05**]. Algorithm 3 provides an example of how we succeed to compute the largest eigenvalue of a billion by billion matrix in the time required by naive methods for a hundred by hundred matrix.

2.6. Generalizations to complex and quaternion. We can consider extending the same matrix algorithms to random complex (GUE) and quaternion (GSE) matrices. For the complex case, we take `randn(n)+i*randn(n)`. Quaternions may be less familiar. Not available in MATLAB (without special programming) but easily imagined is `randn(n)+i*randn(n)+j*randn(n)+k*randn(n)`, where $ij = k$, $jk = i$, $ki = j$, $ji = -k$, $kj = -i$, $ik = -j$, $ijk = -1$. One can complete to an entire algebraic system obtaining the third division ring. Remember that a division ring is an algebra where $ab = 0$ implies at least one of a or b is 0. Matrices are not a division ring even though they are an algebra.

In MATLAB, one can simulate scalar quaternions $a+bi+cj+dk$ with the matrix `[a+bi  c+di;-c+di  a-bi]`. Similarly, the quaternion matrix $A + Bi + Cj + Dk$ can be simulated with the MATLAB matrix `[A+Bi  C+Di;-C+Di  A-Bi]`.

The generalizations to $\beta = 2, 4$ are as follows. Let β count the number of independent real Gaussians, and let G_β be a complex ($\beta = 2$) or quaternion ($\beta = 4$) Gaussian respectively. G denotes G_1 by default.

Therefore, the upper triangular R_n, tridiagonal T_n (β-Hermite ensemble) and bidiagonal B_n (β-Laguerre ensemble) reductions have the following form

$$R_n = \begin{pmatrix} \chi_{n\beta} & \mathrm{G}_\beta & \dots & \mathrm{G}_\beta & \mathrm{G}_\beta \\ & \chi_{(n-1)\beta} & \dots & \mathrm{G}_\beta & \mathrm{G}_\beta \\ & & \ddots & \vdots & \vdots \\ & & & \chi_{2\beta} & \mathrm{G}_\beta \\ & & & & \chi_\beta \end{pmatrix},$$

$$T_n = \begin{pmatrix} \mathrm{G}\sqrt{2} & \chi_{(n-1)\beta} & & & \\ \chi_{(n-1)\beta} & \mathrm{G}\sqrt{2} & \chi_{(n-2)\beta} & & \\ & \ddots & \ddots & \ddots & \\ & & \chi_{2\beta} & \mathrm{G}\sqrt{2} & \chi_\beta \\ & & & \chi_\beta & \mathrm{G}\sqrt{2} \end{pmatrix}, \quad \text{and} \tag{2.3}$$

$$B_n = \begin{pmatrix} \chi_{(n-1)\beta} & & & & & \\ \chi_{(m-1)\beta} & \chi_{(n-1)\beta} & & & & \\ & & \ddots & \ddots & & \\ & & & \chi_{2\beta} & \chi_{(n-m+2)\beta} & \\ & & & & \chi_\beta & \chi_{(n-m+1)\beta} \end{pmatrix}. \tag{2.4}$$

Of interest is that T_n and B_n are real matrices whose eigenvalue and singular value distributions are exactly the same as the original complex and quaternion matrices. This leads to even greater computational savings because only real numbers need to be stored or computed with.

Algorithm 3 Compute the largest eigenvalues of a billion by billion matrix.

```
%% This code requires statistics toolbox
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 eigenvlaue is near 1
eigs(H, 1, 1, opts);
```

2.7. Generalization Beyond. We follow out there that a computation trick lead to deep theoretical results. We summarize two generalizations and will survey recent results in Section 3 and Section 5 correspondingly.

- **Stochastic Operator**: that tridiagonals tend to a stochastic operator was first announced by Edelman [**Ede03**] in 2003 and subsequently developed in [**Sut05, ES07**] with a formal argument that was perhaps satisfactory at the physics or applied math level. Pure mathematics treatment was rigorously investigated in [**RR09, RRV11, Blo11, BV11**].
- **Ghosts and Shadows of Random Matrices**: there is little reason other than history and psychology to restrict β to only the values corresponding to the reals, complexes, and quaternions $\beta = 1, 2, 4$. The matrices given by T_n and B_n are well defined for any β, and are deeply related to generalizations of the Schur polynomials knows as the Jack Polynomials of parameter $\alpha = 2/\beta$. [**Ede10**] proposed in his method of "*Ghosts and Shadows*" that even G_β exists and has a meaning upon which algebra might be doable.

3. Stochastic Operators

Classically, many important distributions of random matrix theory were accessed through what now seems like an indirect procedure: first formulate an n-by-n random matrix, then compute an eigenvalue distribution, and finally let n approach infinity. The limiting distribution was reasonably called an eigenvalue distribution, but it did not describe the eigenvalue of any specific operator, since the matrices were left behind in the $n \to \infty$ limit.

All of that has changed with the stochastic operator approach to random matrix theory. The new framework is this:

- Select a *stochastic differential operator* such as the stochastic Airy operator
$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} - x + \frac{2}{\sqrt{\beta}} W'(x),$$
where $W(x)$ is the Wiener process.
- Compute an eigenvalue distribution.

That's it. This approach produces the same eigenvalue statistics that have been studied by the random matrix theory community for decades but in a more direct fashion. The reason: the stochastic differential operators of interest are the $n \to \infty$ continuum limits of the most-studied random matrix models, as we shall see.

The stochastic operator approach was introduced by Edelman [**Ede03**] in 2003 and developed by Edelman and Sutton [**Sut05, ES07**].

3.1. Brownian motion and white noise. We begin by discussing simple Brownian motion and its derivative, "white noise." Right away we would like to demystify ideas that almost fit the usual calculus framework, but with some differences. Readers familiar with the Dirac delta function (an infinitesimal spike) have been in this situation before.

The following simple MATLAB code produces a figure of the sort that resembles logarithmic stock market prices. Every time we execute this code we get a different random picture (shown in Figure 1 Left).

```
x = [0:h:1];   %Think of h as "Delta x"
dW = randn(length(x),1)*sqrt(h); %Think sqrt(Delta x)
W= cumsum(dW);
plot(x,W)
```

Intuitively, we break $[0, x]$ into intervals each having length Δx. For each interval, we sample ΔW which is a zero mean normal with variance equal to Δx and then sum them up. Thus, if we look at one point x, we have

$$W(x) \stackrel{d}{=} \sum_{i=1}^{\left[\frac{x}{\Delta x}\right]} \Delta W = \sum_{i=1}^{\left[\frac{x}{\Delta x}\right]} \mathrm{G} \cdot \sqrt{\Delta x}.$$

$W(x)$ is a normal with mean 0 and variance $\frac{x}{\Delta x} \times \Delta x = x$, i.e. $W(x) \sim N(0, x)$ (shown in Figure 1 Center and Right). We can write this as $W(x) = \sqrt{x} \cdot \mathrm{G}$, G denoting a standard normal. In particular, $W(1)$ is a standard normal, and $W(x) - W(y)$ has mean 0 and variance $(x - y)$, i.e. $W(x) - W(y) = \sqrt{x - y} \cdot \mathrm{G}$. $W(x)$ is known as the *Wiener process* or *standard Brownian motion*. It has the property that $W(x) - W(y)$ has the distribution $N(0, x - y)$.

A suggestive notation is

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

and the corresponding Wiener process is

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

The $\sqrt{\mathrm{d}x}$ seems troubling as notation, until one realizes that the `cumsum` then has quantities that do not depend on h (or Δx) at all. Like the standard integral, mathematics prefers quantities that at least in the limit do not depend on the discretization size or method. Random quantities are the same. The $\sqrt{\mathrm{d}x}$ captures the idea that variances add when adding normals. If each increment depends on $\mathrm{d}x$ instead of $\sqrt{dx}$, then there will be no movement at all because the variance of $W(x)$ will be $\frac{x}{\Delta x} \times (\Delta x)^2 = x \times \Delta x$ which will be 0 when $\Delta x \to 0$.

The derivative $W'(x) = \frac{\mathrm{d}W}{\mathrm{d}x}$ at first seems strange. The discretization would be `dW/h` in the MATLAB code above, which is a discrete-time white noise process. At every point, it is a normal with mean 0 and variance $1/h$, and the covariance matrix is $\frac{1}{h}\mathbb{I}$.In the continuous limit, the differential form $\mathrm{d}W$ denotes a *white noise process* formally satisfying

$$\int f(x, W)W'(x)\mathrm{d}x = \int f(x, W)\mathrm{d}W.$$

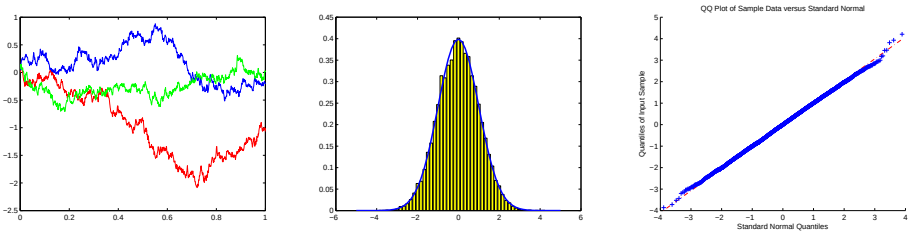


FIGURE 1. Left: Sample paths for standard Brownian motion; Center: histogram of $W(1)$ vs. the pdf of the standard normal; Right: quantile-quantile plot of $W(1)$.

Its covariance function is the Dirac delta $\mathrm{d}W_x \mathrm{d}W_y = \delta(x-y)$. We might say that $W'(x)$ has a "variance density" of 1, referring to the variance divided by the step size of the discretization.

In general we can consider integrals of the form

$$\int_0^x f(t)\mathrm{d}W = \lim_{\Delta x \to 0} \sum_{i=1}^{\left[\frac{x}{\Delta x}\right]} f(i\,[\Delta x])\Delta W,$$

which discretizes to `cumsum(f(t).*dW)`. We can think of $\mathrm{d}W$ as an operator such that $f\,\mathrm{d}W$ is a distribution—not a function in the classical sense, but able to serve as the differential in a stochastic integral. Multiplication by $\mathrm{d}W$ is called the *white noise transformation* [**Ros09**].

3.2. Three local eigenvalue behaviors; three stochastic differential operators. The most commonly studied random matrix models over the years have been the Gaussian, Wishart, and MANOVA ensembles, also known as the Hermite, Laguerre, and Jacobi ensembles. We are primarily concerned with local eigenvalue behavior (that is, a single eigenvalue or a small number of eigenvalues rather than the entire spectrum), which depends on the location in the spectrum as well as the random matrix distribution. Remarkably, though, we see only three different local behaviors among the classical ensembles:

	Region of spectrum		
Ensemble	Left edge	Interior	Right edge
Hermite	soft edge	bulk	soft edge
Laguerre	hard edge	bulk	soft edge
Jacobi	hard edge	bulk	hard edge

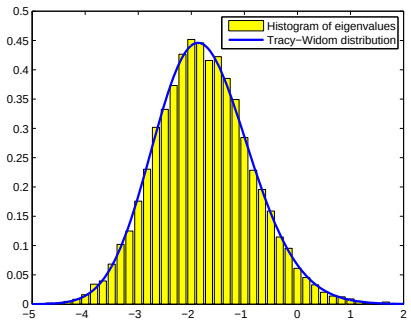
In the next section, we will explore how the operator

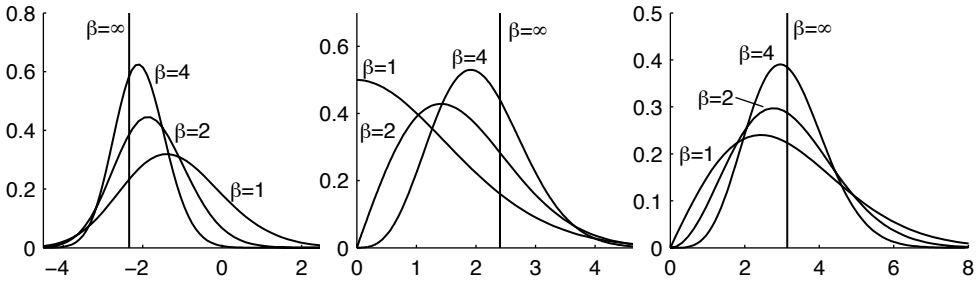
$$\mathcal{A}_\beta = \frac{\mathrm{d}^2}{\mathrm{d}x^2} - x + \frac{2}{\sqrt{\beta}} W'(x)$$

might reasonably be considered to have a random largest eigenvalue that follows the limiting largest eigenvalue law of random matrices. Before proceeding to rigorous mathematical treatment, the authors hope readers will be convinced after running the following numeric experiment.

Algorithm 4 Distribution of the largest eigenvalue of the stochastic Airy operator.

```
% Experiment: Largest eigenvalue of a Stochastic Airy Operator
% Plot:       Histogram of the largest eigenvalues
% Theory:     The Tracy–Widom law
%% Parameters
t=10000;        % number of trials
v=zeros(t,1);   % samples
n=1e9;          % level of discretization
beta=2;
h=n^(-1/3);     % h serves as dx
x=[0:h:10];     % discretization of x
N=length(x);
%% Experiment
% generate the off diagonal elements
b=(1/h^2)*ones(1,N-1);
for i=1:t
    %% discretize stochastic airy operator
    % discretize airy operator
    a=-(2/h^2)*ones(1,N); % differential operator: d^2/dx^2
    a=a-x;                % d^2/dx^2 - x
    % add the stochastic part
    dW=randn(1,N)*sqrt(h);
    a=a+(2/sqrt(beta))*dW/h;
    %% calculate the largest eigenvalue of tridiagonal matrix T
    % diagonal of T: a
    % subdiagonal of T: b
    v(i) = maxeig(a,b);
    % maxeig(a,b): eigenvalue solver for tridiagonal matrices
    % downloadable at http://persson.berkeley.edu/mltrid/index.html
end
%% Plot
binsize = 1/6;
[count, x] = hist(v,-6:binsize:6);
bar(x, count/(t*binsize),'y');
%% Theory
hold on;
tracywidom
```





(a) Soft edge / Stochastic Airy (b) Hard edge / Stochastic Bessel (c) Bulk / Stochastic sine

FIGURE 2. Local eigenvalue behavior

For example, after appropriate recentering and rescaling, the largest eigenvalues of the Hermite and Laguerre ensembles are indistinguishable in the $n \to \infty$ limit, because the limiting distributions are identical—both show "soft edge" behavior. In contrast, the limiting behavior of the smallest eigenvalues of the Laguerre ensemble—those at the "hard edge"—follow a very different law. Near a point in the interior of the spectrum support—in the "bulk"— a pair of eigenvalues is more interesting than a single eigenvalue, and the spacing between consecutive eigenvalues is the most commonly studied distribution. Figure 2 contains plots for the three scaling regimes. Plot (a) often describes a largest eigenvalue; plot (b) often describes a smallest singular value; and plot (c) often describes the spacing between two consecutive eigenvalues in the interior of a spectrum.

The stochastic differential operators mentioned above are associated with the three local eigenvalue behaviors:

Local eigenvalue behavior	Stochastic differential operator
Soft edge	Stochastic Airy operator
Hard edge	Stochastic Bessel operator
Bulk	Stochastic sine operator

They are simple to state:

Stochastic Airy operator:

$$\mathcal{A}^\beta = \frac{\mathrm{d}^2}{\mathrm{d}x^2} - x + \frac{2}{\sqrt{\beta}} W'(x),$$

b.c.'s: $f : [0, +\infty) \to \mathbb{R},\ f(0) = 0,\ \lim_{x \to +\infty} f(x) = 0;$

Stochastic Bessel operator:

$$\mathcal{J}_a^\beta = -2\sqrt{x}\frac{\mathrm{d}}{\mathrm{d}x} + \frac{a}{\sqrt{x}} + \frac{2}{\sqrt{\beta}} W'(x),$$

b.c.'s: $f : [0, 1] \to \mathbb{R},\ f(1) = 0,\ (\mathcal{J}_a^\beta f)(0) = 0;$

Stochastic sine operator:

$$\mathcal{S}^\beta = \begin{bmatrix} & \mathcal{J}^\infty_{-1/2} \\ (\mathcal{J}^\infty_{-1/2})^* & \end{bmatrix} + \frac{2}{\sqrt{\beta}} \begin{bmatrix} W'_{11}(x) & \frac{1}{\sqrt{2}} W'_{12}(x) \\ \frac{1}{\sqrt{\beta}} W'_{12}(x) & W'_{22}(x) \end{bmatrix},$$

b.c.'s: $\mathcal{S}^\beta$ acts on $\left[\begin{smallmatrix} f \\ g \end{smallmatrix}\right]$, b.c.'s of $\mathcal{J}^\infty_{-1/2}$ and $(\mathcal{J}^\infty_{-1/2})^*$ apply.

The eigenvalues of the stochastic Airy operator show soft edge behavior; the singular values of the stochastic Bessel operator show hard edge behavior; and the spacing between consecutive eigenvalues of the stochastic sine operator show bulk behavior. These operators allow us to study classical eigenvalue distributions directly, rather than finding the eigenvalue of a finite random matrix and then taking an $n \to \infty$ limit.

3.3. Justification: from random matrices to stochastic operators. The stochastic operators were discovered by interpreting the tridiagonal (2.3) and bidiagonal beta models (2.4) as finite difference schemes. We have three classical ensembles, and each has three spectrum regions, as discussed in the previous section. Continuum limits for eight of the $3 \times 3 = 9$ combinations have been found [**Sut05, ES07**]. We shall review one derivation.

Consider the largest eigenvalues of the β-Hermite matrix model $H = [h_{ij}]$. These lie at a soft edge, and therefore we hope to find the stochastic Airy operator as $n \to \infty$.

First, a similarity transform produces a nonsymmetric matrix whose entries are totally independent and which is easier to interpret as a finite difference scheme. Define D to be the diagonal matrix whose ith diagonal entry equals

$$(n/2)^{-(i-1)/2} \prod_{k=1}^{i-1} h_{k,k+1}.$$

Then DHD^{-1} equals

$$\frac{1}{\sqrt{2\beta}} \begin{bmatrix} \mathrm{G}\sqrt{2} & \sqrt{\beta n} & & & & \\ \frac{1}{\sqrt{\beta n}}\chi^2_{(n-1)\beta} & \mathrm{G}\sqrt{2} & \sqrt{\beta n} & & & \\ & \ddots & \ddots & \ddots & & \\ & & \frac{1}{\sqrt{\beta n}}\chi^2_{2\beta} & \mathrm{G}\sqrt{2} & \sqrt{\beta n} & \\ & & & \frac{1}{\sqrt{\beta n}}\chi^2_{\beta} & \mathrm{G}\sqrt{2} & \end{bmatrix},$$

and all entries are independent.

To see the largest eigenvalues more clearly, the matrix is recentered and rescaled. We consider $\sqrt{2}n^{1/6}(DHD^{-1} - \sqrt{2n}I)$. The distribution of the algebraically largest eigenvalue of this matrix, when $\beta \in 1, 2, 4$, converges to one of the curves in Figure 2(a) as $n \to \infty$.

The recentered and rescaled matrix has a natural interpretation as a finite difference scheme on the grid $x_i = hi$, $i = 1, \ldots, n$, with $h = n^{-1/3}$. First, the

tridiagonal matrix is expressed as a sum of three simpler matrices:

$$\begin{aligned}
&\sqrt{2}n^{1/6}(DHD^{-1} - \sqrt{2n}I) \\
&\quad = \frac{1}{h^2}\begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -2 \end{bmatrix} \\
&\qquad - \begin{bmatrix} 0 & 0 & & & \\ x_1 & 0 & 0 & & \\ & \ddots & \ddots & \ddots & \\ & & x_{n-2} & 0 & 0 \\ & & & x_{n-1} & 0 \end{bmatrix} \\
&\qquad + \frac{2}{\sqrt{\beta}} \cdot \frac{1}{\sqrt{2h}} \begin{bmatrix} \mathrm{G} & 0 & & & & \\ \tilde{\chi}^2_{(n-1)\beta} & \mathrm{G} & 0 & & & \\ & \tilde{\chi}^2_{(n-2)\beta} & \mathrm{G} & 0 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \tilde{\chi}^2_{2\beta} & \mathrm{G} & 0 \\ & & & & \tilde{\chi}^2_{\beta} & \mathrm{G} \end{bmatrix},
\end{aligned}$$

with $\tilde{\chi}^2_r$ shorthand for $\frac{1}{\sqrt{2\beta n}}(\chi^2_r - r)$. More briefly,

$$\sqrt{2}n^{1/6}(DHD^{-1} - \sqrt{2n}I) = \frac{1}{h^2}\Delta - \mathrm{diag}_{-1}(x_1, x_2, \ldots, x_n - 1) + \frac{2}{\sqrt{\beta}}N.$$

The random variable $\tilde{\chi}^2_{(n-j)\beta}$ has mean zero and standard deviation $1 + O(h^2)$ uniformly for j satisfying $x_j \le M$ for fixed M. Hence, the total standard deviation on row i is asymptotic to $\frac{1}{\sqrt{2h}}\sqrt{1^2 + 1^2} = h^{-1/2}$. The recentered and rescaled matrix model encodes a finite difference scheme for

$$\mathcal{A}^\beta = \frac{\mathrm{d}^2}{\mathrm{d}x^2} - x + \frac{2}{\sqrt{\beta}}W'(x).$$

4. Sturm sequences and Riccati Diffusion

Probabilists and engineers seem to approach eigenvalues in different ways. When a probabilist considers a cumulative distribution function $F(\lambda) = \Pr[\Lambda < \lambda]$, he or she conducts a test: Is the random eigenvalue less than a fixed cutoff? In contrast, when an engineer types `eig(A)` into MATLAB, he or she expects to receive the locations of the eigenvalues directly.

If one looks under the hood, however, the distinction may disappear. A competitive numerical method for eigenvalues is *bisection iteration* with *Sturm sequences*. The method is easiest to describe for the largest eigenvalue. Starting from an initial guess λ_0, the method determines if there are any eigenvalues greater than λ_0. If so, the guess is increased; if not, it is decreased. In time, the largest eigenvalue is captured within an interval, and then the interval is halved with each step. This is *linear convergence*, because the number of correct bits increases by one with each step. At the end, the numerical location is found from a sequence of tests.

Linear convergence is fast, but it is not as fast as, say, Newton's iteration, which converges quadratically. What makes the overall method competitive is the sheer speed with which the test against λ_k can be conducted. The key tool is the Sturm sequence for tridiagonal matrices, which has a close connection to the *Sturm-Liouville theory* of ordinary differential equations. Below, we show how these theories inspire two seemingly different approaches to computing random eigenvalue distributions. The Sturm sequence approach was introduced by Albrecht, Chan, and Edelman and applied to computing eigenvalue distributions of the β-Hermite ensemble [**ACE08**]. The continuous Riccati diffusion was introduced by Ramírez, Rider, and Virág [**RRV11**], by applying a change of variables to the stochastic differential operators of the previous section [**RRV11**].

4.1. Sturm sequences for numerical methods. A Sturm sequence can reveal the inertia of a matrix, i.e., the number of positive, negative, and zero eigenvalues.

For an n-by-n matrix $A = [a_{ij}]$, define A_k to be the k-by-k principal submatrix in the lower-right corner, e.g., $A_1 = [a_{n,n}]$ and $A_2 = \left[\begin{smallmatrix} a_{n-1,n-1} & a_{n-1,n} \\ a_{n,n-1} & a_{n,n} \end{smallmatrix}\right]$. Because the eigenvalues of A_k interlace those of A_{k+1}, the *Sturm sequence*

$$(\det A_0, \det A_1, \det A_2, \ldots, \det A_n)$$

reveals the inertia. Specifically, assuming that no zeros occur in the sequence, the number of sign changes equals the number of negative eigenvalues. (Because a zero determinant occurs with zero probability in our random matrices of interest, we will maintain the assumption of a zero-free Sturm sequence.) Alternatively, the *Sturm ratio sequence* can be used. If $r_i = (\det A_i)/(\det A_{i-1})$, then the number of negative values in $(r_1, r_2, \ldots, r_n)$ equals the number of negative eigenvalues.

The Sturm ratio sequence can be computed extremely quickly when A is tridiagonal. Labeling the diagonal entries $a_n, a_{n-1}, \ldots, a_1$ and the subdiagonal entries $b_{n-1}, b_{n-2}, \ldots, b_1$ from top-left to bottom-right, the ith Sturm ratio is

$$r_i = \begin{cases} a_1, & i = 1; \\ a_i - \frac{b_{i-1}^2}{r_{i-1}}, & i > 1. \end{cases}$$

This reveals in quick order the number of negative eigenvalues of A, or the number of eigenvalues less than λ if $A - \lambda\mathbb{I}$ is substituted for A.

Computing eigenvalues is one of those "impossible" problems that follow from the insolubility of the quintic. It is remarkable that counting eigenvalues is so quick and easy.

4.2. Sturm sequences in random matrix theory. The Sturm sequence enables the computation of various random eigenvalue distributions. Let us consider the largest eigenvalue of the β-Hermite ensemble.

The tridiagonal β-Hermite matrix model H_n^β has diagonal entries $a_i = G_i$ and subdiagonal entries $b_i = \chi_{(i-1)\beta}/\sqrt{2}$. Hence, the Sturm ratio sequence of $H_n^\beta - \lambda I$ is

$$r_i = \begin{cases} G_1 - \lambda, & i = 1; \\ G_i - \lambda - \frac{\chi_{(i-1)\beta}^2}{2r_{i-1}}, & i > 1. \end{cases}$$

The distribution of the largest eigenvalue is

$$\begin{aligned}\Pr[\Lambda_{\max} < \lambda] &= \Pr[H_n^\beta - \lambda I \text{ has all negative eigenvalues}] \\ &= \Pr[\text{all Sturm ratios } r_i \text{ are negative}] \\ &= \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} f_{r_1,\ldots,r_n}(s_1,\ldots,s_n)\mathrm{d}s_1 \cdots \mathrm{d}s_n,\end{aligned}$$

in which $f_{r_1,\ldots,r_n}$ is the joint density of all Sturm ratios. Albrecht, Chan, and Edelman compute this joint density from (4.2) and find

$$f_{r_1,\ldots,r_n}(s_1,\ldots,s_n) = \frac{1}{\sqrt{2\pi}} e^{-(s-\lambda)^2/2} \prod_{i=2}^{n} f_{r_i|r_{i-1}}(s_i|s_i - 1),$$

$$f_{r_i|r_{i-1}}(s_i|s_{i-1}) = \frac{|s_{i-1}|^{p_i}}{\sqrt{2\pi}} e^{-\frac{1}{2}(s_i+\lambda)^2 + z_i^2/4} D_{-p_i}(\mathrm{sign}(s_{i-1})(s_i + \lambda + s_{i-1})),$$

with D_p denoting a parabolic cylinder function.

The level density, i.e., the distribution of a randomly chosen eigenvalue from the spectrum, has also been computed using Sturm sequences[**ACE08**].

4.3. Sturm-Liouville theory. The presentation of Sturm sequences focused on finite matrices. However, there is a deep connection with the continuous world, through Sturm-Liouville theory.

Recall that the eigenvalues of A that are less than λ are equal in number to the negative Sturm ratios of $A - \lambda\mathbb{I}$. This is not all. A similar relationship exists with the solution vector $x = (x_n, x_{n-1}, \ldots, x_1)$ of $(A - \lambda\mathbb{I})x = 0$. Of course, if λ is not an eigenvalue, then no nontrivial solution exists. However, the underdetermined matrix

$$\tilde{T} = \begin{bmatrix} e_1 & (A - \lambda\mathbb{I}) \end{bmatrix},$$

with e_1 denoting the first standard basis vector, always has a nontrivial solution $\tilde{x} = (x_{n+1}, x_n, x_{n-1}, \ldots, x_1)$, and λ is an eigenvalue of A if and only if $\tilde{T}x = 0$ has a nontrivial solution with $x_{n+1} = 0$. This recalls the shooting method for boundary value problems—the solution space is expanded by relaxing a boundary condition, a solution is found, and then the boundary condition is reasserted.

As just mentioned, the test $x_{n+1} \stackrel{?}{=} 0$ highlights the eigenvalues of A. The other solution vector entries $x_n, x_{n-1}, \ldots, x_1$ provide useful information as well. Letting $s_i = x_i/x_{i-1}$, $i = 2, \ldots, n+1$, the reader can check that

$$s_i = \begin{cases} -\frac{r_{i-1}}{b_{i-1}}, & i = 2, \ldots, n; \\ -r_n, & i = n+1. \end{cases}$$

If the subdiagonal entries b_{i-1} are all positive—as they are for the β-Hermite matrix model with probability one—then the "shooting vector ratios" $s_{n+1}, \ldots, s_2$ and the Sturm ratios $r_n, \ldots, r_1$ have opposite signs.

In particular, A has no eigenvalues greater than λ if and only if the shooting vector ratios $s_{n+1}, \ldots, s_2$ are all positive.

This result may sound familiar to a student of differential equations. One of the important results of Sturm-Liouville theory is this: the nth eigenfunction of a regular Sturm-Liouville operator $\mathcal{L}$ has exactly n zeros. In particular, the lowest-energy eigenfunction never crosses 0. This leads to the already-mentioned shooting method: From a guess λ for the lowest eigenvalue, relax a boundary condition and

solve $\mathcal{L} - \lambda f = 0$. If the solution has no zeros, then the guess was too low; if the solution has zeros, then the guess was too high.

4.4. Riccati diffusion. The stochastic Airy operator is a regular Sturm-Liouville problem [**ES07, Blo11**]. It can be analyzed by the shooting method and a Riccati transform, following Ramírez, Rider, and Virág [**RR09**], and the largest eigenvalue distribution can be computed with the help of Kolmogorov's backward equation, as shown by Bloemendal and Virág [**BV10, BV11**]. Bloemendal and Sutton have developed an effective numerical method based on this approach [**BS12**].

First, the Riccati transform. Consider the stochastic Airy operator $\mathcal{A}^\beta$ acting on a function $f(x)$. Define $w(x) = f'(x)/f(x)$. Then

$$w'(x) = \frac{f''(x)}{f(x)} - \left(\frac{f'(x)}{f(x)}\right)^2 = \frac{f''(x)}{f(x)} - w(x)^2.$$

If $f(x)$ is an eigenfunction of $\mathcal{A}^\beta$, then it passes two tests: the differential equation

$$f''(x) - xf(x) + \frac{2}{\sqrt{\beta}} W'(x) f(x) = \Lambda f(x)$$

and the boundary conditions $f(0) = 0$ and $\lim_{x\to+\infty} f(x) = 0$. In fact, the boundary condition at $+\infty$ forces $f(x)$ to decay at the same rate as $\mathrm{Ai}(x)$. After the change of variables, these conditions become

$$w'(x) = x + \Lambda + \frac{2}{\sqrt{\beta}} W'(x) - w(x)^2.$$

and

$$\lim_{x\to 0^+} w(x) = +\infty,$$
$$w(x) \sim \mathrm{Ai}'(x)/\,\mathrm{Ai}(x) \sim -\sqrt{x} \quad (x \to +\infty).$$

Conversely, if $w(x)$ satisfies the first-order differential equation and satisfies the boundary conditions, then $f(x) = \exp(\int w(x)\,\mathrm{d}x)$ is an eigenfunction with eigenvalue Λ. Sturm-Liouville theory leads to the following three equivalent statements concerning the largest eigenvalue $\Lambda_{\max}$ of the stochastic Airy operator:

(1) $\Lambda_{\max} < \lambda$.
(2) Suppressing the right boundary condition, the solution to $(\mathcal{A}^\beta - \lambda)f(x) = 0$ has no zeros on the nonnegative half-line.
(3) Suppressing the right boundary condition, the solution $w(x)$ to the first-order ODE $w'(x) = x + \lambda - w(x)^2 + \frac{2}{\sqrt{\beta}} W'(x)$ has no poles on the nonnegative half-line.

Computing the probability of any of these events gives the desired distribution, the generalization of the Tracy-Widom distribution to arbitrary $\beta > 0$.

One final trick is in order before moving to computation. The test value λ can be removed from the diffusion equation with the change of variables $t = x + \lambda$. The resulting equation is equivalent in distribution to

$$w'(t) = t - w(t)^2 + \frac{2}{\sqrt{\beta}} W'(t), \tag{4.1}$$

and the left boundary condition becomes $lim_{t\to\lambda^+} w(t) = +\infty$. We have $\Lambda_{\max} < \lambda$ if and only if $w(t)$ has no poles in $[\lambda, +\infty)$.

4.5. Kolmogorov's backward equation. The probability of a pole in the solution to the stochastic Riccati diffusion (4.1) turns out to be a tractable computation; Kolmogorov's backward equation is designed for this sort of problem.

We ultimately need to enforce the left boundary condition $\lim_{t\to 0^+} w(t) = +\infty$. The trick is to broaden the problem, analyzing all boundary conditions before finding the original one as a special case [**BV10**]. That is, we compute $\Pr_{(t_0,w_0)}[\text{no poles}]$ for all initial conditions $w(t_0) = w_0$.

For initial conditions with large t_0, it is rather easy to predict whether a pole appears in the solution of the Riccati equation. When noise is removed, the equation has two fundamental solutions: $\operatorname{Ai}'(t)/\operatorname{Ai}(t) \sim -\sqrt{t}$, which is like an unstable equilibrium in that it repels nearby solutions, and $\operatorname{Bi}'(t)/\operatorname{Bi}(t) \sim \sqrt{t}$, which is like a stable equilibrium. Solutions with $w(t_0) < \operatorname{Ai}'(t_0)/\operatorname{Ai}(t_0)$ hit $w = -\infty$ in finite time when run forward, and solutions with $w(t_0) > \operatorname{Ai}'(t_0)/\operatorname{Ai}(t_0)$ become asymptotic to $\sqrt{t}$ when run forward. White noise has no effect on this behavior in the $t_0 \to +\infty$ limit. Hence, we know a slice: $\lim_{t_0\to+\infty} \Pr_{(t_0,w_0)}[\text{no poles}] = \mathbf{1}_{w_0 \geq -\sqrt{t_0}}$.

Kolmogorov's backward equation specifies how this probability evolves as the initial condition moves backward. Let $F(t,w) = \Pr_{(t,w)}[\text{no poles}]$. (Notice that we have dropped subscripts from t_0 and w_0, but these are still initial values.) The backward equation is

$$\frac{\partial F}{\partial t} + (t - w^2)\frac{\partial F}{\partial w} + \frac{2}{\beta}\frac{\partial^2 F}{\partial w^2} = 0.$$

With our initial condition and the boundary condition $\lim_{w\to-\infty} F(t,w) = 0$, this has a unique solution. The desired quantity is a horizontal slice:

$$\begin{aligned}\Pr[\Lambda_{\max} < \lambda] &= \Pr[\text{Riccati diffusion started at } w(\lambda) = +\infty \text{ has no poles with } t > \lambda] \\ &= F(\lambda, +\infty).\end{aligned}$$

Bloemendal and Sutton have developed a numerical routine for solving the PDE numerically. Some challenges arise, particularly when β becomes large. Then the PDE is dominated by convection, and its solution develops a jump discontinuity (a butte, so to speak). The solution can be smoothed out by an additional change of variables [**BS12**].

5. Ghosts and Shadows

We propose to abandon the notion that a random matrix exists only if it can be sampled. Much of today's applied finite random matrix theory concerns real or complex random matrices ($\beta = 1, 2$). The "threefold way" so named by Dyson in 1962 adds quaternions ($\beta = 4$). While it is true there are only three real division algebras (β="dimension over the reals"), this mathematical fact while critical in some ways, in other ways is irrelevant and perhaps has been over interpreted over the decades.

We introduce the notion of a "ghost" random matrix quantity that exists for every beta, and a "shadow" quantity which may be real or complex which allows for computation. Any number of computations have successfully given reasonable answers to date though difficulties remain in some cases.

Though it may seem absurd to have a "three and a quarter" dimensional or "π" dimensional algebra, that is exactly what we propose and what we compute with. In the end β becomes a noisiness parameter rather than a dimension.

This section contains an "idea" which has become a "technique." Perhaps it might be labeled "a conjecture," but we think "idea" is the better label right now. Soon, we hopefully predict, this idea will be embedded in a rigorous theory.

The idea was discussed informally to a number of researchers and students at MIT for years now, probably dating back to 2003 or so. It was also presented at a number of conferences [**Ede03**] and in a paper [**Ede10**].

Mathematics has many precedents, the number 0 was invented when we let go of the notion that a count requires objects to exist. Similarly negative numbers are more than the absence of existing objects, imaginary numbers can be squared to obtain negative numbers, and infinitesimals act like the "ghosts of departed quantities." Without belaboring the point, mathematics makes great strides by letting go of what at first seems so dear.

What we will obtain here is a rich algebra that acts in every way that we care about as a β-dimensional real algebra for random matrix theory. Decades of random matrix theory have focused on reals, complexes, and quaternions or $\beta = 1, 2, 4$. Statisticians would say the real theory is more than enough and those who study wireless antenna networks would say that the complexes are valuable, while physicists are an applied community that also find the quaternions of value. Many random matrix papers allow for general betas formally, perhaps in a formula with factor $\prod |x_i - x_j|^\beta$; we wish to go beyond the formal.

Though it may seem absurd to have a "three and a quarter" dimensional algebra, as long as $\alpha = 2/\beta$ is associated with "randomness" rather than dimension, there is little mathematical difficulty. Thus we throw out two notions that are held very dear: 1) a random object has to be capable of being sampled to exist and 2) the three division algebras so important to non-random matrix theory must take an absolute role in random matrix theory. One reference that captures some of this philosophy is [**Par03**].

The entire field of *free probability* introduced by Dan Voiculescu around 1986 is testament to the power of the first idea, that a random object need not be sampled in order to exist. Some good references are [**VDN92**] or [**NS06**]. In free probability, the entire theory is based on moments and generating functions rather than on sampling. To be sure $\beta = 1, 2, 4$ will always be special, perhaps in the same way that as the factorial function melts away into the gamma function: permutations are no longer counted but analysis goes so very far.

We introduce the notion of a "ghost" in a straightforward manner in the next section. We propose that one can compute with "ghosts" through "shadow" quantities thereby making the notions concrete. Some of the goals that we wish to see are

(1) The definition of a continuum of Haar measures on matrices that generalize the orthogonals, unitaries and symplectics;
(2) A mechanism to compute arbitrary moments of the above quantities;
(3) A mechanism to compute Jacobians of matrix factorizations over beta-dimensional objects;
(4) Various new definitions of the Jack polynomials that generalize the Zonal and Schur Polynomials;
(5) New proofs and insights on any number of aspects of random matrix theory.

In section 3 we showed that the large n limit of random matrix theory corresponds to a stochastic integral operator where β inversely measures the amount of randomness. We believe that finite random matrix theory deserves an equal footing.

5.1. Ghost Random Variables. In [**Ede10**] we provided the beginnings of a formal theory. In these notes we prefer the informal approach. Let x_1 be real standard normal, x_2 be a complex number with independent real and imaginary part that are iid standard normals, and x_4 be a quaternion composed of four independent standard normals. We observe that

- $|x_\beta| \sim \chi_\beta$ (the absolute value has a real Chi distribution)
- $\Re(x_\beta) \sim G$ (the real part is a standard normal)
- $\|\boldsymbol{v}_\beta\| \sim \chi_{n\beta}$ if $\boldsymbol{v}_\beta$ is an n-dimensional vector whose entries are independent and distributed as x_β
- $Q_\beta \boldsymbol{v}_\beta \sim \boldsymbol{v}_\beta$ if $\boldsymbol{v}_\beta$ is defined as above and Q_β is (orthogonal/unitary/symplectic).

We pretend that the above objects (and others) make sense not only for $\beta = 1, 2, 4$ for any $\beta > 0$. We call x_β a ghost Gaussian, $\boldsymbol{v}_\beta$ a vector of ghost Gaussians, and Q_β, a ghost unitary matrix.

Definition 1. **(Shadows)** A shadow is a real (or complex) quantity derived from a ghost that we can sample and compute with.

We therefore have that the norm $\|x_\beta\| \sim \chi_{beta}$ is a shadow. So is $\Re(x_\beta)$.

5.2. Ghost Orthogonals ("The Beta Haar Distribution"). We reason by analogy with $\beta = 1, 2, 4$ and imagine a notion of orthogonals that generalizes the orthogonal, unitary, and symplectic groups. A matrix Q of ghosts may be said to be orthogonal if $Q^T Q = \mathbb{I}$. The elements of course will not be independent.

We sketch an understanding based on the QR decomposition on general matrices of independent ghost Gaussians. We imagine using Householder transformations as is standard in numerical linear algebra software. We obtain immediately

Proposition 2. *Let A be an $n \times n$ matrix of standard β ghost Gaussians We may perform the QR decomposition into ghost orthogonal times ghost upper triangular. The matrix R has independent entries in the upper triangle. Its entries are standard ghost Gaussians above the diagonal, and the non-negative real quantity $R_{ii} = \chi_{\beta(n+1-i)}$ on the diagonal. resulting Q may be thought of as a β analogue of Haar measure. It is the product of Householder matrices H_k obtained by reflecting on the uniform k-dimensional "β sphere."*

The Householder procedure may be thought of as an analog for the $O(n^2)$ algorithm for representing random real orthogonal matrices as described by Stewart [**Ste80**].

We illustrate the procedure when $n = 3$. We use G_β to denote independent standard ghost Gaussians as distributions. They are not meant in any way to indicate common values or that even there is a meaning to having values at all.

$$\begin{pmatrix} \mathrm{G}_\beta & \mathrm{G}_\beta & \mathrm{G}_\beta \\ \mathrm{G}_\beta & \mathrm{G}_\beta & \mathrm{G}_\beta \\ \mathrm{G}_\beta & \mathrm{G}_\beta & \mathrm{G}_\beta \end{pmatrix} = H_3^T \begin{pmatrix} \chi_{3\beta} & \mathrm{G}_\beta & \mathrm{G}_\beta \\ 0 & \mathrm{G}_\beta & \mathrm{G}_\beta \\ 0 & \mathrm{G}_\beta & \mathrm{G}_\beta \end{pmatrix}$$
$$= H_2^T H_3^T \begin{pmatrix} \chi_{3\beta} & \mathrm{G}_\beta & \mathrm{G}_\beta \\ 0 & \chi_{2\beta} & \mathrm{G}_\beta \\ 0 & 0 & \mathrm{G}_\beta \end{pmatrix} \quad .$$
$$= H_1^T H_2^T H_3^T \begin{pmatrix} \chi_{3\beta} & \mathrm{G}_\beta & \mathrm{G}_\beta \\ 0 & \chi_{2\beta} & \mathrm{G}_\beta \\ 0 & 0 & \chi_\beta \end{pmatrix}$$

The H_i are reflectors that do nothing on the first $n-i$ elements and reflect uniformly on the remaining i elements. The absolute values of the elements on the sphere behave like i independent χ_β random variables divided by their root mean square. The Q is the product of the Householder reflectors.

We remark that the β-Haar are different from the circular β ensembles for $\beta \neq 2$.

5.3. Ghost Gaussian Ensembles and Ghost Wishart Matrices. It is very interesting that if we tridiagonalize a complex Hermitian matrix (or a quaternion self-dual matrix), as is done with software for computing eigenvalues, the result is a real tridiagonal matrix. Equally interesting, and perhaps even easier to say, is that the bidigonalization procedure for computing singular values takes general rectangular complex (or quaternion) matrices into real bidiagonal matrices.

The point of view is that the Hermite and Laguerre models introduced in Section 2 are not artificial constructions, but they are shadows of symmetric or general rectangular ghost matrices respectively. If we perform the traditional Householder reductions on the ghosts the answers are the tridiagonal and bidiagonal models. The tridiagonal reduction of a normalized symmetric Gaussian ("The Gaussian β-orthogonal Ensemble") is

$$H_n^\beta \sim \frac{1}{2\sqrt{n\beta}} \begin{pmatrix} \mathrm{G}\sqrt{2} & \chi_{(n-1)\beta} & & & \\ \chi_{(n-1)\beta} & \mathrm{G}\sqrt{2} & \chi_{(n-2)\beta} & & \\ & \ddots & \ddots & \ddots & \\ & & \chi_{2\beta} & \mathrm{G}\sqrt{2} & \chi_\beta \\ & & & \chi_\beta & \mathrm{G}\sqrt{2} \end{pmatrix},$$

where the elements on the diagonal are each independent Gaussians with mean 0 and variance 2. The $\chi's$ on the super and subdiagonal are equal giving a symmetric tridiagonal.

The bidiagonal for a the singular values of a general ghost is similar with chi's running on the diagonal and off-diagonal respectively. See [**DE02**] for details.

We repeat the key point that these "shadow" matrices are real and can therefore be used to compute the eigenvalues or the singular values very efficiently. The notion is that they are not artificial constructions, but what we must get when we apply the ghost Householder transformations.

5.4. Jack Polynomials and Ghosts. Around 1970, Henry Jack, a Scottish mathematician, obtained a sequence of symmetric polynomials $J_\kappa^\alpha(x)$ that are closely connected to our ghosts. The parameter $\alpha = 2/\beta$ for our purposes, and κ

is a partition of an integer k. The argument x can be a finite vector or a matrix. It can also be a formal infinite sequence.

With MOPS [**DES07**], we can press a few buttons before understanding the polynomials just to see what they look like for the partition [2,1,1] of 4:

$$J^1_{2,1,1}(x,y,z) = 3xyz(x+y+z)$$

$$J^\alpha_{2,1,1}(x,y,z) = \frac{12\alpha^2}{(1+\alpha)^2}xyz(x+y+z)$$

$$J^\alpha_{2,1,1} = 2(3+\alpha)m_{2,1,1} + 24m_{1,1,1,1},$$

where $m_{2,1,1}$ and $m_{1,1,1,1}$ denote the monomial symmetric functions.

When $\beta = 2$, the Jack polynomials are the Schur polynomials that are widely used in combinatorics and representation theory. When $\beta = 1$, the Jack polynomials are the zonal polynomials. A wonderful reference for $\beta = 1$, is [**Mui82**]. In general see [**Sta89, Mac98**].

We will not define the Jack polynomials here. Numerical and symbolic routines for their computation may be found in [**DES07, KE06**] respectively.

We expect that the Jack Polynomial formula gives consistent moments for Q through what might be seen as a generating function. Let A and B be diagonal matrices of indeterminates. The formula

$$E_Q J_\kappa(AQBQ') = J_\kappa(A)J_\kappa(B)/J_\kappa(I),$$

provides expressions for moments in Q. Here the J_κ are the Jack Polynomials with parameter $\alpha = 2/\beta$ [**Jac70, Sta89**]. This formula is an analog of Theorem 7.2.5 of page 243 of [**Mui82**]. It must be understood that the formula is a generating function involving the moments of Q and Q'. This is formally true whether or not one thinks that Q exists, or whether the formula is consistent or complete. For square Ghost Gaussian matrices, we expect an analog such that

$$E_G J_\kappa(AGBG') = c^{(\beta)}_\kappa J_\kappa(A)J_\kappa(B).$$

5.5. Ghost Jacobian Computations. We propose a β-dimensional volume in what in retrospect must seem a straightforward manner. The volume element $(\mathrm{d}\boldsymbol{x})^\wedge$ satisfies the key scaling relationship. This makes us want to look a little into "fractal theory," but at the moment we are suspecting this is not really the key direction. Nonetheless we keep an open mind. The important relationship must be

$$\int_{a<\|\boldsymbol{x}\|<b} (\mathrm{d}\boldsymbol{x})^\wedge = \int_a^b S_{\beta-1} r^{\beta-1}\mathrm{d}r = S_{\beta-1}(b^\beta - a^\beta)/\beta,$$

where S_β is the surface area of the sphere (in β dimensions) for any positive real β (integer or not!), i.e.,

$$S_\beta = \frac{2\pi^{\beta/2}}{\Gamma(\beta/2)}.$$

We use the wedge notation to indicate the wedge product of the independent quantities in the vector or matrix of differentials.

This allows the computations of Jacobians of matrix factorizations for general β. As an example relevant to the tridiagonalization above, we can compute the

usual Jacobian for the symmetric eigenvalue problem obtaining

$$(\mathrm{d}A)^\wedge = \prod_{i<j}(|\lambda_i - \lambda_j|^\beta)(Q^T\mathrm{d}Q)^\wedge(\mathrm{d}\Lambda)^\wedge.$$

The derivation feels almost straightforward from the differential of $A = Q\Lambda Q^T$ or $Q^T\mathrm{d}AQ = (Q^T\mathrm{d}Q)\Lambda - \Lambda(Q^T\mathrm{d}Q) + \mathrm{d}\Lambda$. The reason it is straightforward is that the quantity in the (i,j) position that multiplies $(q_i^T\,\mathrm{d}q_j)$ is exactly $\lambda_i - \lambda_j$. In a β-dimensional space this must be scaled with a power of β, respecting the dimensionality scaling $(r\mathrm{d}x)^\wedge = r^\beta(\mathrm{d}x)^\wedge$.

5.6. Application: Numerical Generation of Samples of Singular Values from $G\Sigma^{-1/2}$. To sample from the singular values of an $m \times n$ matrix of standard ghosts, all that is necessary is to compute the singular values of the real bidiagonal form (2.4). LAPACK contains algorithms that very efficiently compute the singular values. Standard procedures in MATLAB, Mathematica, or MAPLE are less efficient, as they presume dense or general sparse formats, not taking advantage of the bidiagonal matrix.

Sampling the singular values of an $m\times n$ matrix of standard ghosts with columns scaled by a diagonal matrix $\Sigma^{-1/2}$ is harder. In an upcoming work [**DE12**], we will show how the method of ghosts allowed for the computation of a practical algorithm.

When $\Sigma = \mathbb{I}$, the joint density of the squared singular values has the Laguerre density

$$c_{m,n,\beta}\prod_{i=1}^{n}\lambda_i^{\frac{m-n+1}{2}\beta-1}\prod_{j<k}|\lambda_j - \lambda_k|^\beta e^{-\frac{\beta}{2}\sum_{i=1}^n\lambda_i}.$$

The more general form when $\Sigma \neq I$, which is sometimes associated with Harish-Chandra-Itzykson-Zuber integral when $\beta = 2$ is more complicated:

$$c_{m,n,\beta,\Sigma}\prod_{i=1}^{n}\lambda_i^{\frac{m-n+1}{2}\beta-1}\prod_{j<k}|\lambda_j - \lambda_k|^\beta {}_0F_0^{(\beta)}(-\frac{\beta}{2}\Lambda, \Sigma^{-1}),$$

where ${}_0F_0$ denotes the hypergeometric function of matrix arguments, which is itself defined in terms of Jack Polynomials [**KE06**].

We outline the idea of our algorithm here. See [**DE12**] for the details. If $n = 1$, i.e., we only have one column, then we can replace each ghost with a real χ_{beta} without changing the singular value. By induction we assume that we can get the SVD of a matrix with $n-1$ columns and thus get the SVD with n columns. We therefore assume that we can put the SVD problem in the form

$$Z = \begin{pmatrix} & & d_n g_{1,n} \\ & UTV' & \vdots \\ & & d_n g_{m,n} \end{pmatrix},$$

where U and V are ghost-unitary matrices, the g's are standard ghost Gaussians and d_n is real as is T which is diagonal. We can multiply the right $n-1$ columns by V and on the left by U taking advantage of the invariance of ghost Gaussian

vectors. This yields a new matrix with the same singular values:

$$\begin{pmatrix} t_1 & & & & d_n\chi_\beta \\ & t_2 & & & \\ & & \ddots & & \vdots \\ & & & t_{n-1} & \\ & & & & d_n\chi_\beta \end{pmatrix}.$$

We can then proceed to replace the ghosts with independent χ_β without changing the singular values. We think of this as moving the phases into the t_i and then removing them on the left. The resulting real matrix is

$$\begin{pmatrix} t_1 & & & & d_n g_{1,n} \\ & t_2 & & & \\ & & \ddots & & \vdots \\ & & & t_{n-1} & \\ & & & & d_n g_{m,n} \end{pmatrix}.$$

As evidence of the effectiveness of this algorithm, see Algorithm 5 and Figure 3.

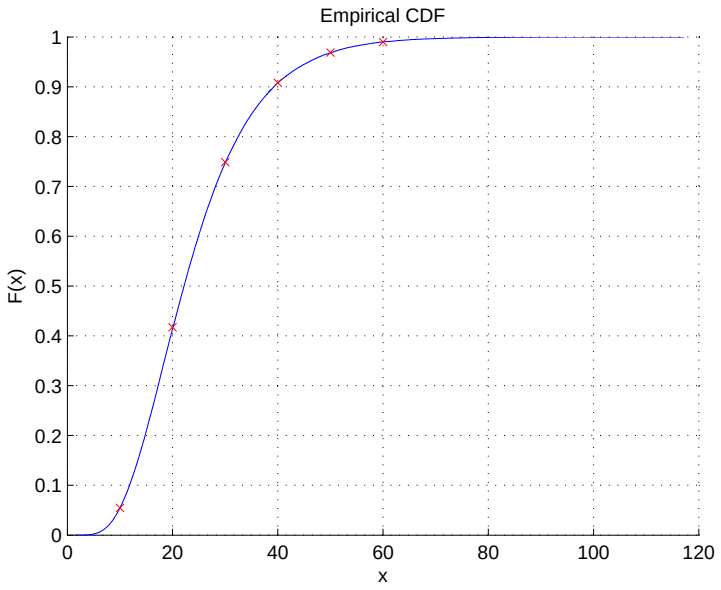


FIGURE 3. The empirical (blue) and analytic (red) cdf's of the largest eigenvalue of the general β-Wishart Ensemble.

Algorithm 5 Singular values of General β-Wishart Ensemble.

```
%Experiment: Singular Values of the General Beta-Wishart Ensemble.
%Plot:       The empirical (blue) and analytic (red) cdf's of the largest eig.
%Techniques: Method of Ghost and Shadows
%Theory:     The General Beta-Wishart Ensemble.
function ghost_exp
%% Parameters
m = 5; n = 4;                   % matrix size
t = 10^5;                       % trials
beta = .75;                     % dimension
Sigma = diag([1,2,3,4]);        % Wishart Covariance
M = 120;                        % terms in hypergeometric sum
%% Experiment
sv_list = zeros(t,1);
for j = 1:t
    sv2 = mxn(m,n,beta,Sigma).^2;
    sv_list(j) = max(sv2);
end
%% Experiment Plot
hold on, cdfplot(sv_list);
%% Theory Plot
xrange = 10:10:60;
for j = 1:6
    yrange(j) = cdfBetaWishart(xrange(j),n,m,beta,Sigma,M);
end
plot(xrange,yrange,'rx');
hold off
end
function sv = mxn(m,n,beta,Sigma)
%% Computes the General Beta-Wishart Singular Values
if m < n
    sv = [];
elseif n == 1
    sv = sqrt(Sigma(1,1)*chi2rnd(m*beta));
else
    sv_0 = mxn(m,n-1,beta,Sigma(1:n-1,1:n-1));
    sv = svd([[diag(sv_0); zeros(m-n+1,n-1)], ...
              sqrt(Sigma(n,n)*chi2rnd(beta,m,1))]);
end
end
function y = cdfBetaWishart(x,m,n,beta,Sigma,M)
%% The Theoretical General Beta-Wishart Max-Eig Distribution
alpha = 2/beta;
y = multigamma((m-1)*beta/2+1,m,beta)/multigamma((m+n-1)*beta/2+1,m,beta);
y = y * det(.5*x*inv(Sigma))^(n*beta/2);
y = y*mhg(M,alpha,(m+n-1)*beta/2+1-n*beta/2,(m+n-1)*beta/2+1, ...
          -eig(-.5*x*inv(Sigma)))*exp(trace(-.5*x*inv(Sigma)));
% mgh: downloadable from: http://www-math.mit.edu/~plamen/software/mhgref.html
end
function y = multigamma(c,m,beta)
y = pi^(m*(m-1)*beta/4)*prod(gamma(c-(beta/2)*(0:1:(m-1))));
end
```

6. Universality of the Smallest Singular Value

The Central Limit Theorem (CLT) is a theorem in pure mathematics and a way of life in applied mathematics, science and engineering. Outside of pure mathematics, it is a methodology that says that if quantities are independent enough, and perhaps do misbehave in some egregious manner, one can pretend that random variables "mixed up" enough behave as if they are normally distributed.

One can make many variations, but for this section, let us imagine that we have M_n, an $n \times n$ random matrix with independent elements all of mean 0 and variance 1. We would imagine that some suitably mixed up quantity would behave as if the entries were normal, i.e., the distribution of the smallest singular value of M_n should be close to that of `randn(n)`, which asymptotically, equals to [**Ede89**]

$$\Pr[n\lambda_{\min} \leq t] = \int_0^t \frac{1+\sqrt{x}}{2\sqrt{x}} e^{-(x/2+\sqrt{x})} \mathrm{d}x.$$

Twenty four years ago, the first author performed a kind of "parallel MATLAB" experiment that convinced him beyond a shadow of a doubt that two things were true:

- One can replace `randn(n)` with other nice distributions with mean 0 and variance 1 and the answer would hardly care;
- This was already true even for small n.

The experiments were performed in "the computer room" on the third floor of the Math department at MIT. Every Sun workstation in the room had a MATLAB session running with a polite handwritten note explaining why the machines were being used, and permission to kill the computation if the machine was needed, but to please leave the machine running otherwise, because then MATLAB could not run in the background conveniently.

For many years, the first author thought the right probabilist would pull a theorem from a textbook, maybe modify it a little, and prove a big theorem that explained when random variables can be replaced with normals. The authors knew some variations of CLT, learned of Lindeberg and Berry-Esseen style results, but remain to this day disappointed, as well as hopeful. Recently, Tao and Vu [**TV10**] showed the universality of the smallest singular value and renewed hope that the right theorem would come along to make the applied mathematician's life easier.

Perhaps a personal satisfaction and disappointment is that with careful experiments one can usually know what is going on with random matrices, but the theory remains short. Even the celebrated Tao-Vu results are statements as $n \to \infty$; the proof details pessimistically require that n be huge before the phenomenon kick in. In this section, we provide a reformulation of the Tao-Vu intuition in numerical linear algebra language.

Consider computing a block 2×2 QR decomposition of an $n \times n$ matrix M.

$$\underset{n-s}{M} = \begin{pmatrix} \underset{n-s}{M_1} & \underset{s}{M_2} \end{pmatrix} = QR = \begin{pmatrix} \underset{n-s}{Q_1} & \underset{s}{Q_1^\perp} \end{pmatrix} \times \begin{pmatrix} \underset{n-s}{R_{11}} & \underset{s}{R_{12}} \\ & \underset{s}{R_{22}} \end{pmatrix}^{n-s}_{s}.$$

Thus, we have

$$(Q_1^\perp)^T M_2 = R_{22}.$$

The smallest singular value of the lower right triangular block R_{22} (of size $s \times s$) scaled by $\sqrt{n/s}$ is a good estimate for the smallest singular value of M.

- If M were exactly singular, and in the most generic way (the last column or one of the last columns) is a linear combination of the previous columns that are independent. Then R_{22} is exactly singular, hence its smallest singular value is exactly 0. The factor $\sqrt{n/s}$ is explained by Tao and Vu as the factor seen from recent results on sampling theory. We encourage readers to try the following code.

```
n=1000;v=[];
[q,r]=qr(randn(n));
ss=min(svd(r));
for i=1:400,
    k=(n+1-i):n;
    v(i)=min(svd(r(k,k)));
end
%scale the singular values
v=v.*sqrt(1:400)/sqrt(n);
plot(v/ss,'-*')
```

- The CLT would give a standard normal if a random vector on the unit sphere is dotted into a random vector of independent elements of mean 0 and variance 1. This is a kind of stirring up the non-gaussian to smooth out its "rough edges." Indeed the vector on the unit sphere needs not be random, it just has to be not to concentrated on any of the coordinate. For example, a vector with one 1 and $n-1$ 0's would work, but anything that mixes things up is fine.

Combine these two ideas together, we can see that multiplying the $s \times n$ matrix $Q_1^\perp$ by the $n \times s$ matrix M_2 gives an $s \times s$ matrix that behaves like `randn(s)` as $n \to \infty$. It is particularly clean that $\mathrm{span}\{(Q_1^\perp)^T\}$ depends only on M_1, and hence independent of M_2. There is no guarantee that $Q_1^\perp$ is not concentrated, but with high probability it will "mix things up". We also wish to mention some recent progresses [**ESYY12, EY12**].

Acknowledgement

We would like to thank Alexander Dubbs, Plamen Koev, Oren Mangoubi for valuable comments and suggestions. We also acknowledge Alexander Dubbs for providing his code for Algorithm 5.

References

[ACE08] James Albrecht, Cy P Chan, and Alan Edelman, *Sturm sequences and random eigenvalue distributions*, Foundations of Computational Mathematics **9** (2008), no. 4, 461–483.

[Blo11] Alex Bloemendal, *Finite rank perturbations of random matrices and their continuum limits*, Ph.D. thesis, University of Toronto, 2011.

[Bor10] Folkmar Bornemann, *On the numerical evaluation of distributions in random matrix theory: A review*, Markov Processes and Related Fields **16** (2010), 803–866, arXiv:0904.1581.

[BS12] Alex Bloemendal and Brian D. Sutton, *General-beta computation at the soft edge*, in preparation (2012).

[BV10] Alex Bloemendal and Bálint Virág, *Limits of spiked random matrices i*, arXiv preprint arXiv:1011.1877 (2010).

[BV11] ———, *Limits of spiked random matrices ii*, arXiv preprint arXiv:1109.3704 (2011).

[DE02] Ioana Dumitriu and Alan Edelman, *Matrix Models for Beta Ensembles*, Journal of Mathematical Physics (2002), no. 11, 5830–5847.

[DE12] Alexander Dubbs and Alan Edelman, *A ghost and shadow approach to the singular values of general β-wishart matrices*, in preparation (2012).

[DES07] Ioana Dumitriu, Alan Edelman, and Gene Shuman, *Mops: Multivariate orthogonal polynomials (symbolically)*, Journal of Symbolic Computation **42** (2007), 587–620.

[Ede89] Alan Edelman, *Eigenvalues and condition numbers of random matrices*, Ph.D. thesis, Massachusetts Institute of Technology, 1989.

[Ede03] Alan. Edelman, *SIAM conference on applied linear algebra*, The College of William and Mary, Williamsburg, VA (2003).

[Ede10] Alan Edelman, *The random matrix technique of ghosts and shadows*, Markov Processes and Related Fields **16** (2010), no. 4, 783–790.

[ER05] Alan Edelman and N. Raj Rao, *Random matrix theory*, Acta Numerica **14** (2005), no. 233-297, 139.

[ES07] Alan Edelman and Brian D. Sutton, *From random matrices to stochastic operators*, Journal of Statistical Physics **127** (2007), no. 6, 1121–1165.

[ESYY12] L. Erdős, B. Schlein, H.T. Yau, and J. Yin, *The local relaxation flow approach to universality of the local statistics for random matrices*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, vol. 48, Institut Henri Poincaré, 2012, pp. 1–46.

[EW13] Alan Edelman and Yuyang Wang, *Random matrix theory and its innovative applications*, Advances in Applied Mathematics, Modeling, and Computational Science **66** (2013), 91–116.

[EY12] L. Erdős and H.T. Yau, *Universality of local spectral statistics of random matrices*, Bull. Amer. Math. Soc **49** (2012), 377–414.

[Jac70] Henry Jack, *A class of symmetric polynomials with a parameter*, Proc. R. Soc. Edinburgh **69** (1970), 1–18.

[KE06] Plamen Koev and Alan Edelman, *The efficient evaluation of the hypergeometric function of a matrix argument*, Mathematics of Computation **75** (2006), 833–846.

[Mac98] I.G. Macdonald, *Symmetric functions and hall polynomials*, 2nd ed., Oxford Mathematical Monographs, Oxford University Press, 1998.

[Mui82] R. J. Muirhead, *Aspects of multivariate statistical theory*, John Wiley & Sons Inc., New York, 1982, Wiley Series in Probability and Mathematical Statistics. MR MR652932 (84c:62073)

[NS06] A. Nica and R. Speicher, *Lectures on the combinatorics of free probability*, vol. 335, Cambridge University Press, 2006.

[Par03] Giorgio Parisi, *Two spaces looking for a geometer*, Bulletin of Symbolic Logic **9** (2003), no. 2, 181–196.

[Per] Per-Olof Persson, *Eigenvalues of tridiagonal matrices in matlab*, `http://persson.berkeley.edu/mltrid/index.html`.

[Ros09] Sheldon Ross, *Introduction to probability models*, Academic press, 2009.

[RR09] José A. Ramírez and Brian Rider, *Diffusion at the random matrix hard edge*, Communications in Mathematical Physics **288** (2009), no. 3, 887–906.

[RRV11] José A. Ramírez, Brian Rider, and Bálint Virág, *Beta ensembles, stochastic Airy spectrum, and a diffusion*, Journal of the American Mathematical Society **24** (2011), no. 4, 919–944, arXiv:math/0607331v3.

[Sil85] Jack W. Silverstein, *The smallest eigenvalue of a large-dimensional Wishart matrix*, Ann. Probab. **13** (1985), no. 4, 1364–1368. MR MR806232 (87b:60050)

[Sta89] Richard P. Stanley, *Some combinatorial properties of jack symmetric functions*, Adv. Math. **77** (1989), no. 1, 76–115.

[Ste80] G. W. Stewart, *The efficient generation of random orthogonal matrices with an application to condition estimators*, SIAM Journal on Numerical Analysis **17** (1980), no. 3, 403–409.

[Sut05] Brian D. Sutton, *The stochastic operator approach to random matrix theory*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, June 2005, Department of Mathematics.

[TB97] Lloyd N. Trefethen and David Bau, III, *Numerical linear algebra*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997. MR MR1444820 (98k:65002)

[Tro84] Hale F. Trotter, *Eigenvalue distributions of large Hermitian matrices; Wigner's semicircle law and a theorem of Kac, Murdock, and Szegö*, Adv. in Math. **54** (1984), 67–82.

[TV10] Terence Tao and Van Vu, *Random matrices: The distribution of the smallest singular values*, Geometric And Functional Analysis **20** (2010), no. 1, 260–297.

[VDN92] D.V. Voiculescu, K.J. Dykema, and A. Nica, *Free random variables*, vol. 1, Amer Mathematical Society, 1992.

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139

Department of Mathematics, Randolph-Macon College, Ashland, VA 23005

Department of Computer Science, Tufts University, Medford, MA 02155