Random Matrix Theory And Its Applications

Alan Julian Izenman

Department of Statistical Science,
Temple University,
1810 Liacouras Walk, Philadelphia, PA 19122

June 22, 2020

Abbreviated Title

Random Matrix Theory

Author's Footnote

Alan J. Izenman is Professor, Department of Statistical Science, Temple University, 1810 Liacouras Walk, Philadelphia, PA 19122 (e-mail: alan@temple.edu).

AMS 2010 subject classifications. Primary 15B52; Secondary 62H10.

Key words and phrases: Eigenvalue density, Gaussian ensembles, Jacobi ensembles, Marčenko-Pastur's quarter-circle law, Spiked covariance model, Tracy-Widom laws, Wigner matrix, Wigner's semicircle law, Wishart matrix, Wishart-Laguerre ensembles.

Abstract. This article reviews the important ideas behind random matrix theory (RMT), which has become a major tool in a variety of disciplines, including mathematical physics, number theory, combinatorics, and multivariate statistical analysis. Much of the theory involves ensembles of random matrices that are governed by some probability distribution. Examples include Gaussian ensembles and Wishart–Laguerre ensembles. Interest has centered on studying the spectrum of random matrices, especially the extreme eigenvalues, suitably normalized, for a single Wishart matrix and for two Wishart matrices, for finite and infinite sample sizes in the real and complex cases. The Tracy–Widom Laws for the probability distribution of a normalized largest eigenvalue of a random matrix have become very prominent in RMT. Limiting probability distributions of eigenvalues of a certain random matrix lead to Wigner's semicircle Law and Marčenko–Pastur's Quarter-Circle Law. Several applications of these results in RMT are described in this article.

Key words and phrases: Eigenvalue density, Gaussian ensembles, Jacobi ensembles, Marčenko-Pastur's quarter-circle law, Spiked covariance model, Tracy-Widom laws, Wigner matrix, Wigner's semicircle law, Wishart matrix, Wishart-Laguerre ensembles.

1. INTRODUCTION

Random-matrix theory (RMT) gained attention during the 1950s due to work by Eugene Wigner in mathematical physics. Specifically, Wigner wished to describe the general properties of the energy levels (or of their spacings) of highly excited states of heavy nuclei as measured in nuclear reactions (Wigner, 1957). Such a complex nuclear system is represented by an Hermitian operator $\mathbf{H}$ (called the *Hamiltonian*) living in an infinite-dimensional Hilbert space governed by physical laws. Unfortunately, in any specific case, $\mathbf{H}$ is unknown. Moreover, even if it were known, it would be much too complicated to write down and, even if we could write it down, no computer would be able to solve its eigenequation $\mathbf{Hv} = \lambda\mathbf{v}$ (the so-called Schrödinger equation of the physical system), where λ and $\mathbf{v}$ are an eigenvalue-eigenvector pair corresponding to $\mathbf{H}$.

Wigner argued that we should instead regard a specific Hamiltonian $\mathbf{H}$ as behaving like a large random matrix that is a member of a large class (or ensemble) of Hamiltonians, all of which would have similar general properties as the $\mathbf{H}$ in question (Wigner, 1955). The energy levels (represented by the eigenvalues of $\mathbf{H}$) of the physical system could then be approximated by the eigenvalues of a large random matrix. Furthermore, the spacings between energy levels of heavy nuclei could be more easily modeled by the spacings between successive eigenvalues of a random $(n \times n)$-matrix as $n \to \infty$.

Since the 1960s, Wigner and his colleagues, including Freeman Dyson and Madan Lal Mehta, worked on RMT and developed it to the point that it became a very powerful tool in mathematical physics (see Mehta, 2004). Dyson, in a series of papers in 1962, introduced a classification of three types of random matrix ensembles based upon the property of time-reversal invariance. The matrices corresponding to these three types of random matrix ensembles have elements that are complex (not time-reversal invariant), real (time-reversal invariant), or self-dual quaternion (time-reversal invariant, but with a restriction).

During the last few decades, we have seen more interest paid to RMT. One of the most important early discoveries in RMT was its connection to quantum chaos (Bohigas, Giannoni, and Schmit, 1984), which led to an RMT of quantum transport (Beenakker, 1997). RMT has since become a major tool in many fields, including number theory and combinatorics, electrical engineering, geophysical sciences, and wireless communications (Tulino and Verdú, 2004).

More recently, RMT has been used as an important tool for probabilistic modeling. Examples include:

- *Determinantal point processes* (DPPs), which arise in models for fermions in quantum physics (Kulesza and Taskar, 2012). A DPP is a stochastic point process whose probability density is characterized as the determinant of a real, symmetric, positive-semidefinite, kernel matrix $\mathbf{K}$, whose eigenvalues are in $[0, 1]$. Choice of $\mathbf{K}$ depends upon the application. In particular, the eigenvalues of a random matrix whose entries are drawn from a complex Gaussian distribution (GUE; see Section 3) are distributed as a DPP on $\mathbb{R}$. A good reference is Anderson, Guionnet, and Zeitouni (2009, Section 4.2).
- *Free probability theory*, which was introduced in an operator-algebraic context by Dan-Virgil Voiculescu in a series of papers during the 1980s. Later, he discovered its connection with RMT (Voiculescu, 1991, 1995) when he realized that his asymptotic results had already appeared in Wigner's semicircle law (see Section 5.1.2). The fact that operator algebras and random matrices are strongly related to each other proved to be a big breakthrough in the development of free probability. The two fundamental concepts in free probability theory are (1) *a noncommutative probability space* and (2) *free independence* (or *freeness*), which is a noncommutative analogue of classical independence. Voiculescu showed that certain random $(n \times n)$-matrices (e.g., independent Gaussian random matrices) have the property referred to as *asymptotic freeness*, as $n \to \infty$, which makes it easier to derive the asymptotic eigenvalue distribution of random matrices. Fan and Johnstone (2019), for example, apply free probability theory to proving certain results for the spectra of MANOVA estimates, motivated by issues in evolutionary genetics, and give an asymptotic freeness result for certain types of random matrices. See Mingo and Speicher (2017) for a survey.

A common element in these types of situations is that RMT has been used as an indirect method for solving complicated problems arising from physical or mathematical systems. Some examples of the use of RMT in physical or mathematical systems include the following: The positions at any given intermediate time of N independent one-dimensional Brownian motion paths with time in $[0, 1]$ that start and end at certain prescribed points but do not intersect at any intermediate time (Delvaux and Kuijlaars, 2010); last-passage time of a certain last-passage percolation model (Baik, 2003); height fluctuations of a certain random growth model (Johansson, 2000); and fluctuation properties of the energy levels of chaotic quantum systems (Bohigas, Giannoni, and Schmit, 1984). Each of these quantities behaves statistically like either the eigenvalues or the spacings between consecutive eigenvalues of a random $(n \times n)$-matrix as $n \to \infty$. Other fascinating examples of RMT include: The distances between parked cars in London (Šeba, 2009); waiting times for buses in Cuernavaca, Mexico (Baik, Borodin, Deift, and Suidan, 2006); and an airline boarding problem (Bachmat, Berend, Sapir, Skiena, and Stolyarov, 2006). These problems and others are discussed by Deift (2007). See also Krbalek and Šeba (2000).

Much of RMT has been directed towards dimensionality reduction techniques in multivariate statistical analysis, such as principal components analysis, canonical variate analysis, and classification, discriminant analysis, and cluster analysis (Johnstone, 2001, 2008). With the availability of high-speed computers and extensive data storage facilities, researchers have had to adapt to high-dimensional, large sample size situations. As a result, many of the classical statistical techniques, such as covariance estimation, hypothesis testing, multivariate analysis of variance (MANOVA), and multivariate regression analysis, have been reinvented to accommodate such "big data."

Recent work on random matrices has tried to establish the so-called *universality conjecture*. The Hermitian case of this conjecture was referred to by Tao and Vu (2011c) as the *Wigner-Dyson-Mehta conjecture*. The conjecture states that local behavior (i.e., fluctuation properties) of the eigenvalues of large random matrices have asymptotic limits that are independent of the probability distribution on the matrix ensembles. We refer the reader to Erdős (2010), who gives an historical account of the development of the various proofs of the universality conjecture (including those of Sinai and Soshnikov, 1998; Soshnikov, 1999, 2002; Erdős, Péché, Ramírez, Schlein, and Yau, 2010; Tao and Vu, 2010, 2011a, 2011b, 2012) for a wide class of (real and complex) Wigner random matrices.

In this article, we review the current state of RMT and provide some of the very interesting applications of that theory. Section 2 describes the basic terminology, including the idea of "ensembles" of random matrices. Section 3 deals with the three different types of Gaussian ensembles, orthogonal, unitary, or symplectic, according as the entries in a random matrix are real, complex, or real-quarternian, respectively. Section 4 introduces the spectrum of random matrices, which is divided into the bulk and extremes (or edges) of the eigenvalues of those matrices. Section 5 studies the bulk of the eigenvalues of an $(n \times n)$ Wishart matrix $\mathbf{S} = \mathcal{X}\mathcal{X}^{\tau}$, which is constructed from an $(n \times r)$-matrix $\mathcal{X}$ whose entries are iid standard Gaussian deviates, where r is dimensionality and n is sample size. Of interest are results for finite n (exact distribution) and large n (asymptotic distribution, including Wigner's semicircle law). Then, we describe the joint distribution of the eigenvalues of a single Wishart matrix in multivariate analysis. The cases considered are a mixture of finite and large n and fixed and large r. Also discussed in this Section is Iain Johnstone's spiked covariance model, which has become a very popular research topic in RMT. Next, the case of two independent Wishart matrices is described for fixed r and finite sample sizes, with applications to multivariate regression and MANOVA. In Section 6, we look at the edges of the spectrum for a single Wishart matrix and two Wishart matrices, and especially at the distribution of the largest eigenvalue of a Gaussian ensemble and a Wishart–Laguerre ensemble. When both r and n are large, we have the Tracy–Widom Laws. Discussion of the availability of computational packages for computing the distributional results of this article together with a brief survey of the RMT literature is given in Section 7.

Notation. We will use the following notation: $\stackrel{a.s.}{\to}$ means almost sure convergence and $\stackrel{\mathcal{D}}{\Rightarrow}$ means convergence in distribution. The abbreviation iid means independent and identically distributed, rhs means right-hand side, and iff means if and only if. If $\mathbf{X}$ is a real-valued random r-vector that has a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, we write $\mathbf{X} \sim \mathcal{N}_r(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. If $\mathbf{X}$ is a complex-valued random r-vector, then the random $2r$-vector $(\mathrm{Re}(\mathbf{X}), \mathrm{Im}(\mathbf{X}))^{\tau}$ with real components has the distribution (Brillinger, 1975, Section 4.2)

$$(1) \qquad \mathcal{N}_{2r}\left(\begin{pmatrix} \mathrm{Re}(\boldsymbol{\mu}) \\ \mathrm{Im}(\boldsymbol{\mu}) \end{pmatrix}, \frac{1}{2}\begin{pmatrix} \mathrm{Re}(\boldsymbol{\Sigma}) & -\mathrm{Im}(\boldsymbol{\Sigma}) \\ \mathrm{Im}(\boldsymbol{\Sigma}) & \mathrm{Re}(\boldsymbol{\Sigma}) \end{pmatrix}\right).$$

We write $\mathbf{X} \sim \mathcal{N}_r^C(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and say that $\mathbf{X}$ has the *complex multivariate Gaussian distribution* with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is an $(r \times r)$ Hermitian nonnegative-definite matrix. Note that the RMT literature has not been consistent in its notation, including the number of variables (p, n, N, k), the number of observations (M, N, n), population eigenvalues (λ_j, ℓ_j, α_j, σ_j), and sample eigenvalues (x_j, ℓ_j, λ_j, s_j, μ_j). In this article, we follow the notation and terminology of Izenman (2013).

2. MATRIX ENSEMBLES

We will need the following terminology. An *ensemble* of matrices is a family (or collection) of matrices together with a probability density p that shows how likely it is that any member of the family can be observed.

Wigner and Dyson were most interested in approximating the infinite-dimensional $\mathbf{H}$ by an ensemble of finite, large, $(n \times n)$ Hermitian matrices $\mathbf{H}_n$ whose probability density has the following form,

$$p(\mathbf{H}_n) \propto e^{-\beta \mathrm{tr}[V(\mathbf{H}_n)]}, \tag{2}$$

where $\mathrm{tr}(\mathbf{M})$ is the trace of the square and symmetric matrix $\mathbf{M}$, V is some function of $\mathbf{H}_n$, such as a finite polynomial function of $\mathbf{H}_n$, where the highest power is even and its coefficient positive, and where the constant of proportionality depends only on n. For example, a possible choice of V could be

$$V(\mathbf{H}_n) = a\mathbf{H}_n^2 + b\mathbf{H}_n + c, \tag{3}$$

where a, b, and c are real numbers and $a > 0$. The entries of $\mathbf{H}_n = (H_{ij})$ can be real ($\beta = 1$), complex ($\beta = 2$), or real-quaternion ($\beta = 4$). If $V(\mathbf{H}_n) \propto \mathbf{H}_n^2$, then $\mathrm{tr}(\mathbf{H}_n^2) = \sum_i \sum_j H_{ij}^2$, and (1) reduces to a Gaussian ensemble. For example, in the case of a real, symmetric (2×2) random matrix,

$$\mathbf{H}_n = \begin{pmatrix} H_{11} & H_{12} \\ H_{12} & H_{22} \end{pmatrix}, \tag{4}$$

with independent elements, we have that $\mathrm{tr}(\mathbf{H}_n^2) = H_{11}^2 + H_{22}^2 + 2H_{12}^2$, so that each element is an independent Gaussian random variate and the variance of each off-diagonal element is one-half that of the diagonal elements.

We define a "time-reversal" transformation as

$$\mathbf{H}_n \to \mathbf{U}\mathbf{H}_n\mathbf{U}^{-1}, \tag{5}$$

where $\mathbf{U}$ is orthogonal ($\beta = 1$), unitary ($\beta = 2$), or symplectic ($\beta = 4$). (A symplectic matrix is a unitary matrix with real-quaternion elements.) *Time-reversal invariance* means that the time-reversal transformation leaves $p(\mathbf{H}_n)$ invariant (Porter and Rosenzweig, 1960).

3. GAUSSIAN ENSEMBLES

Building upon (1) Wigner's group representations, including the property of "time-reversal" invariance, (2) Hermann Weyl's general theory of matrix algebras, which were of three possible types, orthogonal, unitary, and symplectic, and (3) Wigner's study of ensembles of random matrices, Dyson (1962) recognized a common theme that would unify these topics. He noted that there are exactly three division algebras that contain the real numbers, namely, the real numbers $\mathbb{R}$, the complex numbers $\mathbb{C}$, and the quaternians $\mathbb{Q}$. With that in mind, he introduced three categories of Gaussian random matrix ensembles: *Gaussian orthogonal ensemble* of matrices with entries from $\mathbb{R}$, *Gaussian unitary ensemble* of matrices with entries from $\mathbb{C}$, and *Gaussian symplectic ensemble* of matrices with entries from $\mathbb{Q}$. We define these ensembles as follows:

- *Gaussian Orthogonal Ensemble (GOE)*: Fill an $(n \times n)$-matrix $\mathbf{A}$ with real entries that are each iid as $\mathcal{N}(0, 1)$, so that $\mathbf{A}$ is not symmetric. Then a real *symmetric* $(n \times n)$-matrix $\mathbf{H}_n$ is formed[1] by

[1]Some authors (e.g., Edelman and Rao, 2005) replace $\sqrt{2}$ by 2 for all three types of ensembles.

setting $\mathbf{H}_n = (\mathbf{A} + \mathbf{A}^\tau)/\sqrt{2}$, where $\mathbf{A}^\tau$ is the transpose of the matrix $\mathbf{A}$. The diagonal entries of $\mathbf{H}_n = (H_{jk})$ are distributed as $H_{jj} \stackrel{iid}{\sim} \mathcal{N}(0,2)$ and, subject to being symmetric, the off-diagonal entries are distributed as $H_{jk} \stackrel{iid}{\sim} \mathcal{N}(0,1)$, $j \neq k$.

- *Gaussian Unitary Ensemble (GUE)*: Fill an $(n \times n)$-matrix $\mathbf{A}$ with complex-valued entries, each of which is iid as complex Gaussian, $\mathcal{N}^C(0,1)$, but not Hermitian. Then an *Hermitian* $(n \times n)$-matrix $\mathbf{H}_n$ is formed by setting $\mathbf{H}_n = (\mathbf{A} + \mathbf{A}^*)/\sqrt{2}$, where $\mathbf{A}^*$ is the conjugate transpose of the complex matrix $\mathbf{A}$. The diagonal entries of $\mathbf{H}_n = (H_{jk})$ are real and distributed as $H_{jj} \stackrel{iid}{\sim} \mathcal{N}(0,2)$ and, subject to being Hermitian (i.e., $H_{ij} = H^*_{ji}$), the off-diagonal entries are $H_{jk} = U_{jk} + iV_{jk}$, where $U_{jk}, V_{jk} \stackrel{iid}{\sim} \mathcal{N}(0, \frac{1}{2})$, $1 \leq j < k \leq n$. Thus, $\mathrm{E}\{H_{jk}\} = 0$, $\mathrm{E}\{H^2_{jk}\} = 0$, and $\mathrm{E}\{|H_{jk}|^2\} = \mathrm{E}\{H_{jk}H^*_{jk}\} = 1$, so that $H_{jk} \stackrel{iid}{\sim} \mathcal{N}^C(0,1)$.
- *Gaussian Symplectic Ensemble (GSE)*: Fill an $(n \times n)$-matrix $\mathbf{A}$ with entries that are each real-quarternian and iid as $\mathcal{N}^Q(0,1)$. Than a *self-dual* $(n \times n)$-matrix is formed by setting $\mathbf{H}_n = (\mathbf{A} + \mathbf{A}^D)/\sqrt{2}$, where $\mathbf{A}^D$ denotes the dual transpose of the quarternian matrix $\mathbf{A}$. The diagonal entries of $\mathbf{H}_n = (H_{\ell m})$ are distributed as $H_{mm} \stackrel{iid}{\sim} \mathcal{N}(0,1)$ and, subject to being self-dual, the off-diagonal entries are $H_{\ell m} = U_{\ell m} + iV_{\ell m} + jW_{\ell m} + kZ_{\ell m}$, where $U_{\ell m}, V_{\ell m}, W_{\ell m}, Z_{\ell m} \sim \mathcal{N}(0, \frac{1}{2})$, for $1 \leq \ell < m \leq n$.

For the GOE, $\mathbf{U}$ in (5) is orthogonal with real entries; for the GUE, $\mathbf{U}$ is unitary with complex entries; and for the GSE, $\mathbf{U}$ is symplectic with self-dual quaternion entries. See Table 1. It can be shown that the GOE is for systems that are time-reversal invariant, the GUE for systems that are not time-reversal invariant, and GSE for systems that are time-reversal invariant, but do not have spin-rotational symmetry. From a quantum-mechanical view, time-reversal invariance is the most realistic property, and, hence, the GOE is the most natural ensemble. If all three types of ensembles are time-reversal invariant and all elements of $\mathbf{H}_n$ are statistically independent (subject to the required symmetry constraint), then the form of $p(\mathbf{H}_n)$ is automatically restricted to have $V(\mathbf{H}_n) = a\mathbf{H}_n^2 + b\mathbf{H}_n + c$, with a, b, and c real and $a > 0$ (Porter and Rosenzweig, 1960). In this article, we will not be dealing with GSE matrices.

4. SPECTRUM OF RANDOM MATRICES

Of particular interest is the stochastic behavior of the *bulk* and the *extremes* (or *edges*) of the spectrum of large random matrices. The bulk deals with most of the eigenvalues of a given matrix and the extremes refer to the largest and smallest of those eigenvalues. Note that regardless of which ensemble we study (GOE, GUE, or GSE), the eigenvalues of $\mathbf{H}_n$ are all real and can, therefore, be rank-ordered (Bai and Silverstein, 2010). The extremes (and especially the smallest eigenvalue) of the spectrum are important in determining the stability and invertability of a square matrix. Recent work has shown that there are differences between the statistics of the bulk of the spectrum and those of the extreme eigenvalues at the edge of the spectrum.

One of the main features of all three random matrix ensembles is the idea of *repulsion*, that any two (correlated) eigenvalues obtained from a GOE (or GUE or GSE) matrix are unlikely to be close together; that is, the probability that adjacent eigenvalues are close together is small, and the probability quickly goes to zero as a power of the distance between them. Although this is also true for iid points on an interval, the repulsion for eigenvalues enjoys a faster probability convergence rate to zero than it does for iid points. Hence, the spacings distribution precludes near-zero spacings. This property is related to certain aspects of quantum chaos (Kiecherbauer, Marklof and Soshnikov, 2001).

5. BULK OF THE SPECTRUM

5.1 Gaussian Case: The Real Wigner Matrix

Wigner originally studied a real symmetric $(n \times n)$-matrix $\mathbf{H}_n = (H_{ij})$ where the diagonal entries were each 0 and the off-diagonal entries (subject to the symmetry constraint) were independently ± 1 with probability $\frac{1}{2}$. He later realized that his results for this matrix would continue to hold more generally.

TABLE 1

Dyson's classification of Gaussian ensembles. The Hermitian matrix $\mathbf{H}_n = (H_{ij})$ and its matrix of eigenvectors $\mathbf{U}$ are classified by the parameter $\beta \in \{1, 2, 4\}$, depending upon the presence or absence of time-reversal invariance (TRI) and spin-rotational symmetry (SRS). NA means 'not appropriate.' (Adapted from Table 1 in Beenakker, 1996.)

β	Ensemble	TRI	SRS	H_{ij}	$\mathbf{U}$
1	GOE	Yes	Yes	real	orthogonal
2	GUE	No	NA	complex	unitary
4	GSE	Yes	No	real-quaternion	symplectic

Several definitions of a *real Wigner matrix* have since appeared in the literature. When we say that "$\mathbf{H}_n$ is a Wigner matrix," what we mean is that $\mathbf{H}_n$ is a member of a family of random symmetric matrices that form a *Wigner ensemble*. This family generally consists of symmetric matrices with independent entries, where the off-diagonal entries are iid and the diagonal entries are iid, and the diagonal and off-diagonal entries may be drawn from different distributions. A special type of real Wigner matrix (Wigner, 1955) occurs when it is a member of the GOE (see Section 3), in which case, it is defined as a symmetric (i.e., $\mathbf{H}_n = \mathbf{H}_n^\tau$), random $(n \times n)$-matrix, where the ijth entry is real-valued with distribution $H_{ij} \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_{ij}^2\right), \sigma_{ij}^2 = 1 + \delta_{ij}$, where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. An analogous definition (see Section 3) has been given for a *complex Wigner matrix* if it is a member of the GUE.

REMARK 1. If we assume that $\mathbf{H}_n$ is a *real* symmetric Wigner matrix with independent entries, iid diagonal entries and iid off-diagonal entries, both types having mean zero, then a Gaussian assumption for the Wigner matrix $\mathbf{H}_n$ is not necessary, unless $\mathbf{H}_n$ is required to be a member of the GOE. In general, it could be replaced by a distribution with finite variance σ^2 for entries above the diagonal, and a distribution with no specific moment requirement along the diagonal (Wigner, 1955). See also Paul and Aue (2014). Some early articles did not distinguish between diagonal and off-diagonal distributions. For example, distributional assumptions are stated as having all entries iid with finite moments of all orders (Grenander, 1963), or all entries having finite 6th moment (Arnold, 1967). In the case of a *complex* Wigner matrix, similar relaxations of the Gaussian requirements can be found in the literature. Bai (1999), for example, sets up the Hermitian matrix to have iid off-diagonal entries with variance σ^2 and iid diagonal entries without any moment condition, while Tao (2012, Section 2.4) requires the diagonal entries of the Hermitian matrix to have bounded mean and variance.

Studying the behavior of real or complex Wigner matrices, which play important roles in nuclear and mathematical physics, finance, and communication theory, forms a large part of RMT.

5.1.1 *Finite n: Exact Distribution.* Let $\lambda_1 > \lambda_2 > \cdots > \lambda_n$ be the ordered eigenvalues of a real Wigner matrix $\mathbf{H}_n$ and let $\mathbf{U}$ be the matrix of associated eigenvectors. Then, $\mathbf{H}_n = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\tau$, where $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \ldots, \lambda_n\}$. Now, from (2), we see that $\text{tr}[V(\mathbf{H}_n)] = \sum_{j=1}^n V(\lambda_j)$ depends only on the eigenvalues. Thus, the distribution $p(\mathbf{H}_n)$ in (2) is independent of the eigenvectors, which can be viewed as being uniformly distributed over the members of each matrix ensemble. The exact joint probability distribution of the eigenvalues of an $(n \times n)$ Wigner matrix is then found by multiplying $p(\mathbf{H}_n)$ by the Jacobian of the transformation from the matrix to its eigenvalues and eigenvectors.

The exact distribution of the eigenvalues, therefore, has the form,

$$p(\lambda_1, \ldots, \lambda_n) = c_n e^{-\frac{1}{4}\text{tr}\{\mathbf{H}_n\}} \prod_{1 \le j < k \le n} |\lambda_j - \lambda_k|, \tag{6}$$

where

$$\text{tr}\{\mathbf{H}_n^2\} = \sum_{j=1}^n \lambda_j^2, \tag{7}$$

and the normalizing constant, c_n, is dependent upon n.

For general β-Gaussian-Hermite ensembles, where $\beta = 1$ (GOE), 2 (GUE), or 4 (GSE) form the *Dyson index*, the joint probability density of the eigenvalues of $\mathbf{H}_n$ is given by

$$p_\beta(\lambda_1, \cdots, \lambda_n) = c_{n,\beta} e^{-\frac{\beta}{4}\mathrm{tr}\{\mathbf{H}_n^2\}} \prod_{1 \le j < k \le n} |\lambda_j - \lambda_k|^\beta, \tag{8}$$

where the normalizing constant,

$$c_{n,\beta} = \frac{1}{(2\pi)^{n/2}\beta^{n/2+\beta n(n-1)/4}} \prod_{i=1}^{n} \frac{\Gamma(1+\frac{\beta}{2})}{\Gamma(1+\frac{\beta i}{2})}, \tag{9}$$

is dependent upon n and β (Mehta, 2004, p. 58). The distribution (8) is often referred to as *Dyson's β-ensemble*. In some treatments of this topic, β is replaced by $\alpha = 2/\beta$, so that for GOE, $\alpha = 2$; for GUE, $\alpha = 1$; and for GSE, $\alpha = 1/2$ (Edelman and Rao, 2005).

5.1.2 *Large n: Wigner's Semicircle Law.* Next, consider the empirical distribution of the eigenvalues of a real $(n \times n)$ Wigner matrix $n^{-1/2}\mathbf{H}_n$ (usually referred to as the *empirical spectral distribution* or ESD of $n^{-1/2}\mathbf{H}_n$) as n tends to infinity. Let $\#\{\cdot\}$ denote the number of elements in the set indicated, and let I_A denote the indicator function of the event A ($I_A = 1$ if A is true and 0 otherwise).

Wigner's result says that $G_n(\lambda)$, the ESD of $n^{-1/2}\mathbf{H}_n$, converges a.s. to a nonrandom limiting distribution $G(\lambda)$,

$$G_n(\lambda) = \frac{1}{n}\#\{i: \ \lambda_i \le \lambda\} \ = \frac{1}{n}\sum_{i=1}^{n} I_{[\lambda_i \le \lambda]} \ \stackrel{a.s.}{\to} \ G(\lambda), \quad n \to \infty, \tag{10}$$

where $G(k)$ has density

$$g(\lambda) = \frac{1}{2\pi}\sqrt{4-\lambda^2}, \quad |\lambda| \le 2, \tag{11}$$

and zero for $|\lambda| > 2$. This limiting density is a semicircle with radius 2 (Wigner, 1955, 1958) and is referred to as *Wigner's semicircle law*. In free probability theory, Wigner's semicircle law can be viewed as a matrix analogue of the standard Gaussian distribution (see, e.g., Anderson, Guionnet, and Zeitouni, 2009).

The $2k$-th moment of $g(\lambda)$ is the *kth Catalan number*,

$$C_k = \frac{1}{k+1}\binom{2k}{k}, \tag{12}$$

and, by symmetry, the $2k+1$-st moment is zero, $k = 0, 1, 2, \ldots$, where $C_0 = 1$ by convention. The Catalan numbers, which are 1, 1, 2, 5, 14, 42, 132, ..., satisfy the recursion formula, $C_{k+1} = \sum_{i=1}^{k} C_{i-1}C_{k-i}$. It turns out that Wigner's semicircle law is the unique distribution for which the even moments are the Catalan numbers.

There are several ways of proving results (10)-(11), including:

- The *Method of Moments*: This was the original proof by Wigner (1955). Let $g_n(\lambda)$ denote the empirical spectral density function of $n^{-1/2}\mathbf{H}_n$. This method uses the fact that the kth moment of $g_n(\lambda)$ can be written as
$$\int_{-2}^{2} \lambda^k g_n(\lambda)d\lambda = \frac{1}{n}\mathrm{tr}\{(n^{-1/2}\mathbf{H}_n)^k\} = \frac{1}{n^{1+k/2}}\mathrm{tr}\{\mathbf{H}_n^k\}.$$
As $n \to \infty$, the moments of $g_n(\lambda)$, expressed above as the normalized trace of powers of $\mathbf{H}_n$, converge to the Catalan numbers (12), which are the even moments of the semicircle law $g(\lambda)$.

- The *Stieltjes Transform Method* (also known as the *resolvent method*): The finite-n Stieltjes transform of a single realization of $n^{-1/2}\mathbf{H}_n$ can be expressed as
$$s_n(z) = \int_{-\infty}^{\infty} \frac{g_n(\lambda)}{\lambda - z}d\lambda = \frac{1}{n}\mathrm{tr}\{(n^{-1/2}\mathbf{H}_n - z\mathbf{I}_n)^{-1}\},$$

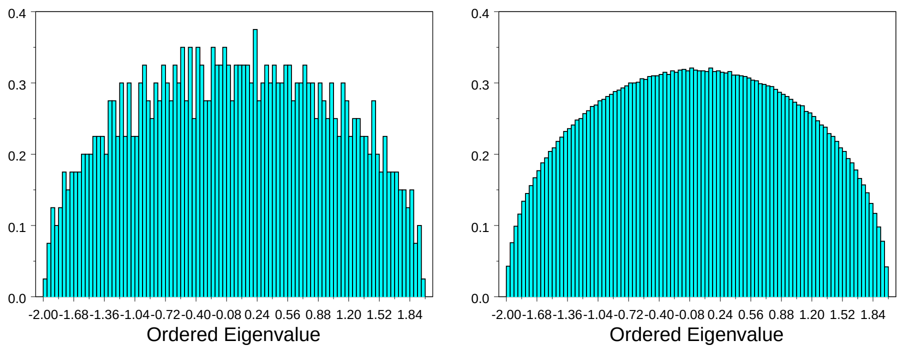

FIG. 1. *Illustration of the convergence to Wigner's Semicircle Law. Normalized histograms of the eigenvalues from a single* $(n \times n)$ *Wigner matrix. Left panel:* $n = 1,000$*. Right panel:* $n = 25,000$*. For each* n*, there are 100 bins.*

where z is not in the support of $g_n(\lambda)$ (i.e., $z \notin [-2, 2]$). The quantity $(n^{-1/2}\mathbf{H}_n - z\mathbf{I}_n)^{-1}$ is the normalized *resolvent* of $\mathbf{H}_n$, which is singular at the eigenvalues of $n^{-1/2}\mathbf{H}_n$. As $n \to \infty$, for each z, $\mathrm{E}\{s_n(z)\}$ converges almost surely to the function $s(z) = \frac{1}{2}\{-z + \sqrt{z^2 - 4}\}$, which is the Stieltjes transform of the semi-circular law $g(\lambda)$. Then, $g(\lambda)$ can be recovered by the inverse Stieltjes transform.

See, e.g., Tao (2012, Section 2.4), Anderson, Guionnet, and Zeitouni (2009, Chapter 2), and Bai (1999) for details.

REMARK 2. For (10) and (11) to hold, we only need the existence of second moments for the off-diagonal entries; we do not need such a moment requirement for the diagonal entries (Wigner, 1955).

REMARK 3. Wigner's semicircle law also holds for a much bigger class of random matrices. This follows because the limiting distribution (11) is independent of the distribution of the entries of $\mathbf{H}_n$. If the Gaussian assumption in the definition of the real Wigner's matrix $\mathbf{H}_n$ is replaced by any distribution (discrete or continuous) with mean zero, finite variance σ^2, and finite higher moments, then Wigner's semicircle law still holds. The limiting density (11) is then rescaled as

$$g_\sigma(\lambda) = \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - \lambda^2}, \quad |\lambda| \leq 2\sigma, \tag{13}$$

and is zero otherwise (see, e.g., Bai, 1999, Theorem 2.1). The limiting distribution (11) obtains by setting $\sigma = 1$. There are two other versions of the semicircle law that have appeared in the RMT literature, and which correspond to different values of σ. If $\sigma = \frac{1}{2}$, then the limiting distribution is $g_{1/2}(\lambda) = \frac{2}{\pi}\sqrt{1 - \lambda^2}$, $|\lambda| \leq 1$, and if $\sigma = \frac{1}{\sqrt{2}}$, then the limiting distribution is $g_{1/\sqrt{2}}(\lambda) = \frac{1}{\pi}\sqrt{2 - \lambda^2}$, $|\lambda| \leq \sqrt{2}$. See Erdős and Yau (2017) for further discussion.

Figure 1 shows an illustration of the convergence to Wigner's semicircle law for a single $(n \times n)$ Wigner matrix. We sampled $n = 1,000$ (left panel) and $n = 25,000$ (right panel) iid standard Gaussian deviates, computed $\mathbf{A}$ and then $\mathbf{H}_n$, and found the eigenvalues of $\mathbf{H}_n$. The histograms of the eigenvalues of $\mathbf{H}_n$ for both cases are given in Figure 1.

5.1.3 *Large n: Largest Eigenvalue.* Bai and Yin (1988) gave necessary and sufficient conditions for a.s. convergence, as $n \to \infty$, of the largest eigenvalue, λ_1, of the normalized Wigner matrix, $n^{-1/2}\mathbf{H}_n$, to a finite constant. The conditions were that diagonal entries have finite second moment and off-diagonal entries have mean (at most) zero and finite fourth moment, The necessity part of the proof used a novel approach that involved graph theory to obtain intermediate results.

5.2 Single Wishart Matrix

In multivariate statistical analysis, we are often interested in a random r-vector $\mathbf{X}$ that is distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, where

$$\boldsymbol{\mu} = \mathrm{E}\{\mathbf{X}\}, \quad \boldsymbol{\Sigma} = \mathrm{E}\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\tau\}. \tag{14}$$

We may sometimes need to assume that $\mathbf{X}$ is also Gaussian. Many topics in multivariate analysis, (e.g., principal component analysis, factor analysis, and multidimensional scaling) deal with dimensionality reduction and the study of functions of $\mathbf{\Sigma}$, such as its eigenvalues and associated eigenvectors (see, e.g., Izenman, 2013).

Typically, $\mathbf{\Sigma}$ is unknown, and so has to be estimated using a sample of data. Given a set of independent random r-vectors, $\mathbf{X}_i, i = 1, 2, \ldots, n$, drawn from the same underlying distribution as $\mathbf{X}$, the usual estimate of $\mathbf{\Sigma}$ is given by

$$\widehat{\mathbf{\Sigma}} = n^{-1} \sum_{i=1}^{n} (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^\tau = n^{-1} \mathcal{X}_c \mathcal{X}_c^\tau, \tag{15}$$

where the sample mean vector, $\bar{\mathbf{X}} = n^{-1} \sum_{i=1}^{n} \mathbf{X}_i$, is an estimator of the population mean vector $\boldsymbol{\mu}$. In (15), $\mathcal{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_n)$ and $\mathcal{X}_c = \mathcal{X}(\mathbf{I}_n - n^{-1}\mathbf{J}_n)$, where $\mathbf{J}_n = \mathbf{1}_n \mathbf{1}_n^\tau$ and $\mathbf{1}_n$ is an n-vector of 1s. We can make $\widehat{\mathbf{\Sigma}}$ an unbiased estimator of $\mathbf{\Sigma}$ by replacing $1/n$ by $1/(n-1)$ in the averaging operation in (15). We are interested in studying the behavior of the eigenvalues of $n\widehat{\mathbf{\Sigma}}$ under different assumptions on the number of variables r and the number of observations n.

5.2.1 *A Distributional Result.* We will need the following result below. Let $\mathbf{A}$ be an $(r \times r)$ positive-definite matrix with density function $p(\mathbf{A})$. The joint density of the eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_r$ of $\mathbf{A}$ is given by (Muirhead, 1982, Theorem 3.2.17)

$$p(\lambda_1, \ldots, \lambda_r) = \frac{\pi^{r^2/2}}{\Gamma_r(r/2)} \prod_{1 \le j < k \le r} |\lambda_j - \lambda_k| \int_{\mathcal{O}(r)} p(\mathbf{QLQ}^\tau)(d\mathbf{Q}), \tag{16}$$

where $\mathbf{L} = \text{diag}\{\lambda_1, \ldots, \lambda_r\}$ is a diagonal matrix and $(d\mathbf{Q})$ is the Haar invariant measure on the set $\mathcal{O}(r)$ of $(r \times r)$ orthogonal matrices, normalized so that $\int_{\mathcal{O}(r)} (d\mathbf{Q}) = 1$. In (16), the function Γ_r is a multivariate gamma function defined by (Muirhead, 1982, Section 2.1.2)

$$\Gamma_r(x) = \pi^{r(r-1)/4} \prod_{j=1}^{r} \Gamma\left(x - \frac{j-1}{2}\right), \quad \text{Re}(x) > \frac{r-1}{2}. \tag{17}$$

A sketch of the proof of (16) involves, first, obtaining the spectral decomposition of $\mathbf{A}$ as $\mathbf{A} = \mathbf{QLQ}^\tau$, where the ith column of $\mathbf{Q}$ is the normalized eigenvector of $\mathbf{A}$ corresponding to the eigenvalue λ_i in $\mathbf{L}$, suitably adjusted to make it 1–1, and then finding the Jacobian of that transformation. The derivative of $\mathbf{A}$ is given by

$$d\mathbf{A} = (d\mathbf{Q})\mathbf{LQ}^\tau + \mathbf{H}(d\mathbf{L})\mathbf{Q}^\tau + \mathbf{QL}(d\mathbf{Q}^\tau), \tag{18}$$

and the Jacobian is found by determining expressions for each of the three terms in (18). The product in (16) involving the pairwise differences of eigenvalues is the Jacobian term, and is the determinant of the *Vandermonde matrix*,

$$\mathbf{V}_r = \begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{r-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{r-1} \\ 1 & \lambda_3 & \lambda_3^2 & \cdots & \lambda_3^{r-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^{r-1} \end{pmatrix}; \tag{19}$$

the determinant of (19) is also known as the *Vandermonde determinant* (see, e.g., Bellman, 1960, p. 193).

5.2.2 *Fixed r, Finite n, and r < n: Exact Distributions.* Prior to the public availability of high-speed computation and large data storage facilities, the number of variables r was kept reasonably small and the number of observations n, though larger than r, was still small by modern standards. Distribution theory was either exact (fixed r and finite n) or asymptotic with a fixed r and $n \to \infty$. These are the two cases we deal with first.

Without loss of generality, suppose $\boldsymbol{\mu} = \mathbf{0}$. Suppose also that $\mathbf{X}_i \overset{iid}{\sim} \mathcal{N}_r(\mathbf{0}, \boldsymbol{\Sigma})$, $i = 1, 2, \ldots, n$, and set $\mathcal{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_n)$. Then,

$$\mathbf{S} = \mathcal{X}\mathcal{X}^\tau \sim \mathcal{W}_r(n, \boldsymbol{\Sigma}) \tag{20}$$

(Wishart, 1928). The distribution (20) is known as the *Wishart distribution* (or *Wishart ensemble*) with n degrees of freedom and covariance matrix $\boldsymbol{\Sigma}$. If $\boldsymbol{\Sigma} = \mathbf{I}_r$, then this is the so-called real *white Wishart* distribution.

When $\mathbf{S} \sim \mathcal{W}_r(n, \boldsymbol{\Sigma})$, we substitute into (16) the form of the Wishart density,

$$p(\mathbf{S}|n, \boldsymbol{\Sigma}) = c_{r,n} |\boldsymbol{\Sigma}|^{-n/2} |\mathbf{S}|^{(n-r-1)/2} e^{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S})}, \tag{21}$$

where

$$c_{r,n}^{-1} = 2^{nr/2} \pi^{r(r-1)/4} \prod_{i=1}^{r} \Gamma\left(\frac{n-i+1}{2}\right), \tag{22}$$

to obtain the exact joint distribution of the eigenvalues, $\lambda_1 > \lambda_2 > \cdots > \lambda_r$, of $\mathbf{S}$ (James, 1964),

$$p(\lambda_1, \ldots, \lambda_r) = c_{r,n} \prod_{j=1}^{r} \lambda_j^{(n-r-1)/2} \prod_{1 \le j < k \le r} |\lambda_j - \lambda_k| \int_{\mathcal{O}(r)} e^{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{Q}\mathbf{L}\mathbf{Q}^\tau)} (d\mathbf{Q}), \tag{23}$$

where $\mathbf{L} = \mathrm{diag}\{\lambda_1, \cdots, \lambda_r\}$, $|\mathbf{L}| = \prod_{j=1}^{r} \lambda_j$, $(d\mathbf{Q})$ is the Haar invariant measure on the set $\mathcal{O}(r)$ of $(r \times r)$ orthogonal matrices, normalized so that $\int_{\mathcal{O}(r)} (d\mathbf{Q}) = 1$, and $c_{r,n}$ is the normalization constant,

$$c_{r,n} = \frac{\pi^{r^2/2}}{2^{nr/2} |\boldsymbol{\Sigma}|^{n/2} \Gamma_r(r/2) \Gamma_r(n/2)}. \tag{24}$$

The integral in (23) over the orthogonal group $\mathcal{O}(r)$ is difficult to evaluate in the case of general $\boldsymbol{\Sigma}$. Some efforts in this direction have been made using infinite series expansions in zonal polynomials, but these have not yielded practical results.

If $\boldsymbol{\mu} \neq \mathbf{0}$, then $\mathcal{X}_c \mathcal{X}_c^\tau$ has the Wishart distribution $\mathcal{W}_r(n-1, n^{-1}\boldsymbol{\Sigma})$. In this case, the previous results, (23) and (24), for the eigenvalue density can be modified by substituting $n-1$ for n and $n\lambda_i$ for λ_i.

In the white Wishart case (i.e., $\boldsymbol{\Sigma} = \mathbf{I}_r$), the integral over the orthogonal group $\mathcal{O}(r)$ is easily evaluated to be $e^{-\frac{1}{2}\sum_j \lambda_j}$. The resulting density (23) reduces to

$$p(\lambda_1, \ldots, \lambda_r) = c_{r,n} \prod_{j=1}^{r} [w_{n,r}(\lambda_j)]^{1/2} \prod_{1 \le j < k \le r} |\lambda_j - \lambda_k|, \tag{25}$$

where

$$w_{n,r}(\lambda) = \lambda^{n-r-1} e^{-\lambda}, \quad \lambda \in [0, \infty), \quad n > r, \tag{26}$$

is the weight function for a generalized Laguerre family of orthogonal polynomials (Abromowitz and Stegun, 1970, Table 22.2), and $c_{r,n}$ is a normalizing constant dependent upon r and n. For a proof, see Anderson (1984, Section 13.3). The second product in (25) involving the pairwise differences of eigenvalues is, as before, the Jacobian term, and is the determinant of the *Vandermonde matrix* (19). The eigenvalue density (25) was found independently and simultaneously by Fisher, Girshick, Hsu, and Roy in 1939, and in 1951 independently by Mood.

As a result, the Wishart ensemble is often known as the Wishart-Laguerre ensemble, and a real parameter $\beta > 0$ is incorporated into its formulation to account for the spectral properties of different types of ensembles.

For β-Wishart-Laguerre ensembles, where $\beta = 1$, 2, or 4 for real, complex, or quaternian Gaussian entries of $\mathcal{X}$, respectively, the joint probability density of the eigenvalues of $\mathbf{S}$ is given by

$$p_{\beta}(\lambda_1, \cdots, \lambda_r) = c'_{n,r,\beta} \prod_{j=1}^{r} [w_{n,r,\beta}(\lambda_j)]^{1/2} \prod_{1 \leq j < k \leq r} |\lambda_j - \lambda_k|^{\beta}, \tag{27}$$

where

$$w_{n,r,\beta}(\lambda) = \lambda^{\beta(n-r+1)-2} e^{-\beta\lambda}, \quad \lambda \in [0, \infty), \tag{28}$$

and $c'_{n,r,\beta}$ is a normalizing constant,

$$c'_{n,r,\beta} = 2^{-\beta nr/2} \prod_{i=1}^{n} \frac{\Gamma(1 + \frac{\beta}{2})}{\Gamma(1 + \frac{\beta i}{2})\Gamma(\frac{\beta}{2}(r - n + i))}, \tag{29}$$

which is dependent upon n, r, and β. Proofs and discussions of these results may be found in Anderson, Guionnet, and Zeitouni (2009, Chapter 2). Note that if any two eigenvalues, λ_j and λ_k, $j \neq k$, say, in (27) (and (25)) are such that $\lambda_j = \lambda_k$, the density vanishes, indicating that the eigenvalues repel each other, ensuring that they are, a.s., distinct.

In the case of general $\mathbf{\Sigma}$, when the population eigenvalues are not all equal, the exact joint distribution of the sample eigenvalues is known (James, 1960, 1961) but is extremely complicated, involving zonal polynomials (i.e., power-series expansions in hypergeometric functions). For large n, the zonal polynomial series converges very slowly (Muirhead, 1982, Section 9.5), and so the results have "very limited value" in that a large number of terms of the series would be needed to be of use (James, 1964). More recently, Mo (2011) used zonal polynomials for an algebraic proof of the distribution of a rank-1 real Wishart spiked model, but made no comments as to convergence properties.

5.2.3 *Fixed r, Large n.* For fixed r and large n, and $\mathbf{X}_i \sim \mathcal{N}_r(\mathbf{0}, \mathbf{\Sigma})$, $i = 1, 2, \ldots, n$, the sample eigenvalues, $\widehat{\lambda}_j$, $j = 1, 2, \ldots, r$, of $n^{-1}\mathbf{S}$ are jointly asymptotically independently distributed according to

$$\sqrt{n}(\widehat{\lambda}_j - \lambda_j) \overset{\mathcal{D}}{\to} \mathcal{N}(0, 2\lambda_j^2), \quad \text{as } n \to \infty, \; j = 1, 2, \ldots, r, \tag{30}$$

where the $\{\lambda_j\}$ are the distinct eigenvalues of $\mathbf{\Sigma}$ (Anderson, 1963). This result shows that the jth sample eigenvalue, $\widehat{\lambda}_j$ is a consistent estimator of the jth population eigenvalue, λ_j, $j = 1, 2, \ldots, r$.

5.2.4 *Large r, Large n: The Marčenko–Pastur's Quarter-Circle Law.* One of the most important results of RMT for use in multivariate analysis is the Marčenko–Pastur Law, which is an analogue of Wigner's semicircle Law. The Marčenko–Pastur Law gives the limiting distribution of the eigenvalues of a sample covariance matrix (as the size of the matrix grows without bound) in the null case when $\mathbf{S} \sim \mathcal{W}_r(n, \mathbf{I}_r)$. In the null case, all the eigenvalues of the population covariance matrix $\mathbf{I}_r$ are equal to one. Although $n^{-1}\mathbf{S}$ is a good approximation to $\mathbf{\Sigma}$ for fixed r and large n, that does not hold when r and n are both large.

If we let $r \to \infty$ and $n \to \infty$ in such a way that the *matrix aspect ratio* converges to a non-zero constant, i.e., $r/n \to \gamma \in (0, \infty)$, then the empirical distribution of the eigenvalues, $\widehat{\lambda}_i, i = 1, 2, \ldots, r$, of $n^{-1}\mathbf{S}$ follows the Marčenko–Pastur Law:

$$\frac{1}{r}\#\{i : \; \widehat{\lambda}_i \leq x\} \overset{a.s.}{\to} G(x), \tag{31}$$

where the limiting distribution $G(x)$ has density $g(x) = G'(x)$ and

$$g(x) = \frac{1}{2\pi\gamma x} \sqrt{(b_+ - x)(x - b_-)} \; I_{[b_-,b_+]}(x), \quad b_{\pm} = (1 \pm \sqrt{\gamma})^2, \tag{32}$$

where $I_{[b_-,b_+]}(x)$ is the indicator function that is equal to 1 for $b_- < x < b_+$ and 0 otherwise (Marčenko and Pastur, 1967). This is the so-called *Quarter-Circle Law.* Note that the limiting density only depends upon γ. If $\gamma \in (0, 1)$, then $r < n$; in this case, the spectra of $\mathcal{X}\mathcal{X}^{\tau}$ and $\mathcal{X}^{\tau}\mathcal{X}$ differ by $n - r$ zero eigenvalues, and so there is an additional point mass at the origin $x = 0$ with weight $1 - \gamma$. Except for this point mass at the origin, the results given here for the nonzero eigenvalues hold regardless of whether r or n is larger.

A visual representation of the Marčenko–Pastur Law is given in Figure 2, where we have separated the values of γ by $\gamma \leq 1$ (left panel) and $\gamma \geq 1$ (right panel). We see that even though all the population eigenvalues are equal to 1, the spread of the sample eigenvalues varies directly with the ratio $\gamma = r/n$: the larger the ratio (i.e., the bigger r is relative to n), the more spread out are the sample eigenvalues. For example, when $\gamma = 1/4$, the density is supported on the interval $[\frac{1}{4}, \frac{9}{4}]$, when $\gamma = 1$ (i.e., $n = r$), the density is supported on $[0, 4]$, and when $\gamma = 4$, the density is supported on $[1, 9]$.

The a.s. result is fascinating, but for statistical inference purposes we would also like to have some insight into second-order information (i.e., variability) about the bulk of the sample eigenvalues, especially the largest eigenvalue, which is of importance in principal component analysis.

5.2.5 *Large r, Large n: Extreme Eigenvalues.* Much of the research work in RMT has focused on the convergence of extreme eigenvalues (i.e., largest and smallest eigenvalues) of sample covariance matrices.

Geman (1980) was the first to prove that the largest eigenvalue, $\widehat{\lambda}_1$, of a sample covariance matrix converges a.s. to the right-hand support point $b_+ = (1 + \sqrt{\gamma})^2$, where he assumed a certain growth condition on all the moments of the underlying distribution. Yin, Bai, and Krishnaiah (1988) then proved the same convergence result, where they assumed that the entries of $\mathcal{X}$ had zero mean and finite fourth moment, which Bai, Silverstein, and Yin (1988) showed was also necessary for the existence of the limit.

So, $\widehat{\lambda}_1$ is not a consistent estimator of the largest eigenvalue, λ_1, of $\mathbf{\Sigma}$, and if both r and n are large, $\widehat{\lambda}_1$ can be severely biased when $\mathbf{\Sigma} = \mathbf{I}_r$. For example, in the case when $r = n$ (i.e., $\gamma = 1$) and $\mathbf{\Sigma} = \mathbf{I}_r$, where all eigenvalues equal 1, the largest population eigenvalue $\lambda_1 = 1$, while the largest sample eigenvalue, $\widehat{\lambda}_1$, of $\widehat{\mathbf{\Sigma}}$ converges to the value $b_+ = (1 + \sqrt{\gamma})^2 = 4$ (see El Karoui, 2008, for further discussion). If $r > n$, then $\widehat{\lambda}_{n+1} = \cdots = \widehat{\lambda}_r = 0$. The results (31) and (32), unfortunately, remained obscure for a while (see, e.g., Wachter, 1978, who derived similar results apparently unaware of the Marčenko-Pastur paper).

We next turn to the smallest eigenvalue, $\widehat{\lambda}_r$, of the sample covariance matrix, $n^{-1}\mathbf{S}$. Silverstein (1985) showed that if the diagonal entries of $\mathcal{X}$ were distributed as $\mathcal{N}(0, 1)$, then $\widehat{\lambda}_r$ converged a.s. to the left-hand support point $b_- = (1 - \sqrt{\gamma})^2$. His proof, however, is too dependent upon the underlying Gaussian assumption and is difficult to extend. Tikhomirov (2015) proved the a.s. convergence of the smallest singular value of the $(r \times n)$ matrix $\mathcal{X}$ (square-root of the smallest eigenvalue of $\mathbf{S} = \mathcal{X}\mathcal{X}^\tau$) assuming only that the entries of $\mathcal{X}$ are iid with finite second moment, while higher moments could be infinite.

Then, in a unified approach, Bai and Yin (1993) were able to show that, assuming all entries in $\mathcal{X}$ are iid with a finite fourth moment and $r/n \to \gamma \in (0, 1)$ as $r, n \to \infty$, both $\widehat{\lambda}_1$ and $\widehat{\lambda}_r$ converge simultaneously and a.s. to their respective limits, b_+ and b_-. Their result for $\widehat{\lambda}_r$ was proved by, first, truncating the variables so that von Neumann's inequality (von Neumann, 1937) can be used to show that the difference between the square-root of the smallest eigenvalue of the truncated sample covariance matrix and that of the truncated and then centralized version converges to 0 as $n \to \infty$; next, the entries of $\mathcal{X}$ are identified with the edges in a graph (as used in Bai and Yin, 1988, for the Wigner matrix), and then graph theory is used to prove the convergence result. The result for $\widehat{\lambda}_1$ was found as a by-product. If the finite fourth moment condition is weakened, they also showed that convergence to the same limits holds, but instead convergence is in probability.

In related work, Livshyts, Tikhomirov, and Vershynin (2019) studied the invertability of an inhomogeneous square random matrix $\mathbf{A}$, say, by deriving a new probability inequality that yields a lower bound on the smallest singular value of $\mathbf{A}$ (i.e., square-root of the smallest eigenvalue of $\mathbf{A}^\tau\mathbf{A}$).

5.3 A Spiked Covariance Matrix

As a specific alternative to the "null" model of a white Wishart distribution, where $\mathbf{\Sigma} = \mathbf{I}_r$, Johnstone (2001) introduced the "non-null" concept of a "spiked" covariance model, where the covariance matrix $\mathbf{\Sigma}$ has most of the eigenvalues equal to 1 (which we refer to as "unit" eigenvalues) and a fixed number, $m \geq 1$, of eigenvalues greater than 1 (i.e., "non-unit" eigenvalues). The spiked covariance model has been proposed

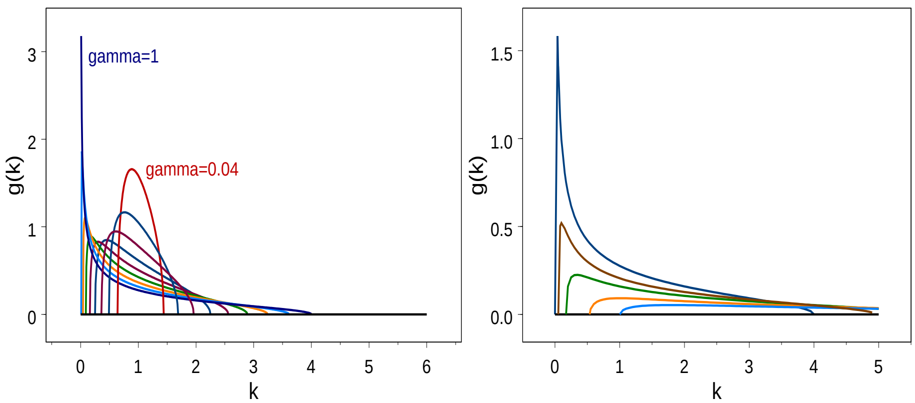

FIG. 2. *Density of eigenvalues from the Marčenko-Pastur Law. Left panel:* $\gamma = 0.04, 0.09, 0.16, 0.25, 0.36, 0.49, 0.64, 0.81, 1$ *(i.e., $r \leq n$). Right panel:* $\gamma = 1, 1.5, 2, 3, 4$ *(i.e., $r \geq n$).*

for a variety of applications, including speech recognition, wireless communication, signal processing, and mathematical finance.

The spiked covariance matrix is given by

$$\boldsymbol{\Sigma}_m = \text{diag}\{\lambda_1, \ldots, \lambda_m, 1 \ldots, 1\}, \quad \lambda_1 > \lambda_2 > \cdots > \lambda_m > 1, \tag{33}$$

where we assume the first m eigenvalues are distinct and that m is known, often being referred to as the number of spikes. Let $J_m = \{\lambda_1, \cdots, \lambda_m\}$. In the null model, $\lambda_1 = \lambda_2 = \cdots = \lambda_m = 1$. In the non-null model, the population covariance matrix $\boldsymbol{\Sigma}_m$ is a finite-rank perturbation of the identity matrix. In a different specification of the non-null model, a rescaling of the spiked covariance model has been proposed (Donoho, Gavish, and Johnstone, 2018, Birnbaum, Johnstone, Nadler, and Paul, 2013) in which the set of smallest eigenvalues, $\lambda_{m+1}, \ldots, \lambda_r$, take the common (known or unknown) value of σ^2 instead of the value 1.

A major discovery by Baik, Ben Arous, and Péché (2005) is that a phase transition exists for the spiked covariance model. The main features of this phase transition for the real and complex cases are as follows.

Real case. In the following discussion, we assume that $\mathbf{X}_i$, $i = 1, 2, \ldots, n$, represent a collection of n real random r-vectors, each iid as $\mathcal{N}_r(\mathbf{0}, \boldsymbol{\Sigma}_m)$, and $\mathcal{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_n)$ denotes the $(r \times n)$ random matrix whose columns are formed from the independent random vectors. We assume that $r/n \to \gamma$, where $\gamma > 0$ is a positive constant, although we often take $\gamma \in (0, 1)$. From (20), we are interested in the distribution of the largest few eigenvalues of the $(r \times r)$ sample covariance matrix $n^{-1}\mathbf{S}$ when $n \to \infty$, where $\mathbf{S} = \mathcal{X}\mathcal{X}^\tau \sim \mathcal{W}_r(n, \boldsymbol{\Sigma}_m)$.

It has been shown that the asymptotic behavior of the sample eigenvalues, which we denote by $\{\widehat{\lambda}_j\}$, of $n^{-1}\mathbf{S}$ depends upon the values of those population eigenvalues (i.e., the $\{\lambda_j\}$) that are larger than one. For $\lambda_j \in J_m$, under real Gaussian assumption on the $\{\mathbf{X}_i\}$, Paul (2004) proved for real Gaussian samples the following three results:

1. If the population eigenvalues that are greater than 1 are actually near 1, then the sample eigenvalues will behave as if $\boldsymbol{\Sigma} = \mathbf{I}_r$.
2. If $\lambda_j \leq 1 + \sqrt{\gamma}$ and $r/n \to \gamma \in (0, 1)$, then $\widehat{\lambda}_j \stackrel{a.s.}{\to} (1 + \sqrt{\gamma})^2$ as $n \to \infty$.
3. If $\lambda_j > 1 + \sqrt{\gamma}$ and $r/n \to \gamma \in (0, 1)$, then $\widehat{\lambda}_j \stackrel{a.s.}{\to} \rho_j = \lambda_j + \frac{\gamma\lambda_j}{\lambda_j - 1}$ as $n \to \infty$.

See also Baik and Silverstein (2006) and Paul (2007). From these results, as Donoho, Gavish, and Johnstone (2018) point out, for large r and n, the sample eigenvalues $\widehat{\lambda}_j$ are biased estimates of their corresponding population eigenvalues λ_j, and, in the third scenario, the amount of bias, $\gamma \cdot \lambda_j/(\lambda_j - 1)$, is asymptotically larger than γ. If $\gamma = 1$, then, in the second scenario, $\widehat{\lambda}_j \stackrel{a.s.}{\to} 4$ as $n \to \infty$, and in the third scenario, $\widehat{\lambda}_j \stackrel{a.s.}{\to} \lambda_j + \frac{\lambda_j}{\lambda_j - 1} = \frac{\lambda_j^2}{\lambda_j - 1}$ as $n \to \infty$.

The asymptotic distributions of the largest sample eigenvalue $\widehat{\lambda}_1$, suitably centered and scaled, are given by (Soshnikov, 2002; Baik and Silverstein, 2006): in the second scenario, when $\lambda_1 < 1+\sqrt{\gamma}$, the asymptotic distribution of $\widehat{\lambda}_1$ is the Tracy–Widom distribution, and when $\lambda_1 = 1+\sqrt{\gamma}$, the asymptotic distribution of $\widehat{\lambda}_1$ is a certain generalization of the Tracy–Widom distribution; and in the third scenario, $\widehat{\lambda}_j$ (assuming multiplicity 1) is asymptotically Gaussian distributed (Paul, 2007); that is,

$$\sqrt{n}(\widehat{\lambda}_j - \rho_j) \overset{\mathcal{D}}{\Rightarrow} \mathcal{N}(0, \sigma^2(\lambda_j)), \quad \sigma^2(\lambda_j) = 2\lambda_j^2\left(1 - \frac{\gamma}{(\lambda_j - 1)^2}\right). \tag{34}$$

For the second scenario, Soshnikov showed that when $\gamma = 1$, the limiting Tracy–Widom distribution does not depend upon the Gaussian assumption. Furthermore, although the Marčenko–Pastur Law (32) holds also for the spiked covariance model, it does not necessarily follow for that model that the largest sample eigenvalue, $\widehat{\lambda}_1$, and the smallest sample eigenvalue, $\widehat{\lambda}_r$, converge to the right and left support points $b_+ = (1+\sqrt{\gamma})^2$ and $b_- = (1-\sqrt{\gamma})^2$, respectively (see Section 5.2.4).

Complex case. In the case of samples from the complex Gaussian distribution $\mathcal{N}_r^C(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, Péché (2003) showed that the largest eigenvalue, $\widehat{\lambda}_1$, of the sample covariance matrix $\widehat{\boldsymbol{\Sigma}} = n^{-1}\mathcal{X}_c\mathcal{X}_c^*$ where $\mathcal{X}_c\mathcal{X}_c^*$ is Hermitian and $\mathcal{X}_c^*$ is the conjugate transpose of the centered $\mathcal{X}_c$, converged to the Tracy–Widom distribution, provided the spiked eigenvalues were not too large. Baik, Ben Arous, and Péché (2005), who studied the limiting distribution of $\widehat{\lambda}_1$ from $\widehat{\boldsymbol{\Sigma}}$, were particularly interested in the case when both $r, n \to \infty$, while $r/n \to \gamma \in (0, 1)$, and the population covariance matrix is the nonnull spiked version $\boldsymbol{\Sigma}_m$, where a small number, m, of the population eigenvalues, $\lambda_1, \ldots, \lambda_m$, are larger than 1, and the remaining eigenvalues each equal 1. They showed that, like the real case, there exists a *phase transition phenomenon* for the complex Gaussian case, where the *phase transition threshold* is $1+\sqrt{\gamma}$:

1. $\widehat{\lambda}_1$ will asymptotically separate from the remaining sample eigenvalues iff at least one of the non-unit eigenvalues of $\boldsymbol{\Sigma}_m$ is greater than $1+\sqrt{\gamma}$. If $\widehat{\lambda}_1$ is separated from the remaining eigenvalues, $\widehat{\lambda}_1$ exhibits a fluctuation of order $n^{1/2}$ rather than $n^{2/3}$. If none of the non-unit eigenvalues is greater than $1+\sqrt{\gamma}$, then $\widehat{\lambda}_1$ will not be separated from the other sample eigenvalues.

2. If $\lambda_1 > 1+\sqrt{\gamma}$, and has multiplicity k, then the distribution of $\widehat{\lambda}_1$, after centering and scaling by $n^{1/2}$, converges in distribution to the distribution of the largest eigenvalue of a $k \times k$ GUE. If $k = 1$, then $\widehat{\lambda}_1$, appropriately centered and scaled by $n^{2/3}$, converges in distribution to the Tracy–Widom F_2 law.

3. If a non-unit eigenvalue $\lambda_j < 1+\sqrt{\gamma}$, the asymptotic distribution of $\widehat{\lambda}_1$ exhibits an $n^{2/3}$ scaling. If one or more non-unit eigenvalues are greater than 2, then the scaling becomes $n^{1/2}$.

Further treatments of the phase transition phenomenon can be found in Nadler (2008) and Wang and Fan (2017).

The study of spiked covariance matrices and the estimation of spiked eigenvalues has become a popular research topic in recent years. See, for example, Dobriban (2017), Wang and Fan (2017), Wang and Yao (2017), Bloemendal, Knowles, Yau, and Yin (2015), Birnbaum, Johnstone, Nadler, and Paul (2013), Bai and Ding (2012), and Capitaine, Donati-Martin, and Féral (2009).

The spiked covariance model has proved to be helpful in understanding the behavior of multivariate statistical techniques such as principal component analysis (PCA) and related methods when both dimensionality r and sample size n are large. When both r and n are large and of the same order of magitude, the sample covariance matrix $n^{-1}\mathbf{S}$ fails to be a good estimator of $\boldsymbol{\Sigma}$. This complements the work of Charles Stein, who observed that the sample covariance matrix from a multivariate Gaussian sample turns out to be a poor estimate of $\boldsymbol{\Sigma}$. In the large r, large n case, Donoho, Gavish, and Johnstone (2018), building upon a large existing statistical literature on shrinkage estimation, sought to improve estimation of $\boldsymbol{\Sigma}$ in the spiked covariance model by optimal shrinkage of the entire spectrum of $\boldsymbol{\Sigma}$, where optimality was shown to depend upon the type of loss function used.

When r is very large and $r/n \to \gamma > 0$, the sample principal components derived from the standard version of PCA turn out to be inconsistent estimates of the population principal components. Although one can argue from this result that PCA cannot be a reliable dimensionality reduction technique (Johnstone and Lu, 2009, Paul, 2007), it has been applied successfully to many high-dimensional situations. This can be explained by the fact that the size of the error reflects eigenvalue size: in practical situations, the eigenvalues may be very large, in which case PCA becomes more precise. In PCA, m is the number of significant principal components and is usually unknown. To estimate m, Bai, Choi, and Fujikoshi (2018a) develop an approach based upon the AIC (Akaike, 1973) and BIC (Schwartz, 1978) estimation criteria when both r and n are large, and they give conditions under which these criteria produce consistent estimates without assuming an underlying Gaussian distribution.

Other approaches. There has been a lot of interest by statistical physicists and statisticians (and others) that the spiked covariance model could be used to investigate the presence of low-dimensional structure in data. Classical approaches to this dimensionality reduction problem include the use of PCA. Early treatments of the spiked model include Deshpande and Montanari (2014), who studied PCA for the spiked Wigner and Wishart models when sparsity constraints are added to the dimensionality reduction problem, and Montanari and Richard (2016), who studied PCA when it is known that the entries of the vector $\mathbf{v}$ (see (35)) are nonnegative (with applications to gene expression data, neural signal processing, and approximate message passing). Both of these articles showed the existence of a phase transition for the eigenvector corresponding to the largest eigenvalue of their specific models that occurred at a certain critical value of the SNR. Miolane (2019) gives a survey of Bayesian approaches to the problem of phase transitions in the spiked model.

Onatski, Moreira, and Hallin (2013) introduced a spiked model in the form of signal-plus-noise that enabled them to apply an hypothesis-testing approach to PCA and related methods, such as sparse PCA and nonnegative PCA. They specified an alternative hypothesis (H_1) as a "rank-m perturbation of the null" in which the population covariance matrix is proportional to a sum of the identity matrix (the null hypothesis H_0) and a matrix of rank m. In other words, if the data are $\mathbf{X}_j$, $j = 1, 2, \ldots, n$, and if $m = 1$, then, conditional on an r-vector $\mathbf{v}$ with norm one, the *rank-1 spiked covariance model* is

$$\mathbf{X}_j|\mathbf{v} \overset{iid}{\sim} \mathcal{N}_r(\mathbf{0}, \sigma^2(\mathbf{I}_r + \theta\mathbf{v}\mathbf{v}^\tau)), \quad j = 1, 2, \ldots, n, \tag{35}$$

where σ and θ are constants ($\sigma = 1$ is assumed), and θ is a *signal-to-noise ratio* (SNR). Sample size n and dimensionality r are assumed to satisfy the condition that $r/n \to \gamma$ for large r and n. This model is a *sparse spiked covariance model*. It is also referred to as the *detection problem*. The question is how would one test for the presence of a signal spike in the data. In this formulation, the competing hypotheses are $H_0 : \theta = 0$ (the pure noise case) versus $H_1 : \theta > 0$ (there is a spike). This leads, in turn, to the study of the likelihood ratio, which is the ratio of the density with $\theta \neq 0$ to that when $\theta = 0$. The likelihood ratio is represented as a contour integral, which is approximated using Laplace's method, and then used to ensure convergence of the log-likelihood ratio to a Gaussian process. These results were then used to establish the asymptotic power of various sphericity tests.

Building on Onatski et al.'s rank-1 spiked covariance model, Perry, Wein, Bandeira, and Moitra (2018) considered a Bayesian approach to the problem of whether PCA can detect low-rank structure in data when noise is also present. They defined a "spike" as the r-vector $\mathbf{v}$ in (35) (whereas Miolane (2019) calls $\mathbf{v}\mathbf{v}^\tau$ the spike and Johnstone and Onatski (2019) calls θ the spike). For a spiked (Gaussian) Wishart, the columns of the $(r \times n)$-matrix $\mathcal{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_n)$ are assumed to have been randomly sampled from the distribution (35) with $\sigma^2 = 1$. The vector $\mathbf{v}$ is assumed to have been drawn from some arbitrary, but known, prior distribution. The priors considered are the *spherical prior* ($\mathbf{v}$ is iid uniform on the r-dimensional unit sphere), the *Rademacher prior* ($\mathbf{v}$ is iid uniform on $\{\pm 1/\sqrt{r}\}$), and the *sparse Rademacher prior* ($\mathbf{v}$ is iid, where each entry is 0 with probability $1 - \rho$ and otherwise uniform on $\{\pm 1/\sqrt{\rho r}\}$). Given the prior and the SNR (i.e., the *Bayes optimal case*), the posterior distribution of the signal spike given the observations was studied. Contiguity was used (following Onatski et al.) as a means of comparing the joint distributions of the eigenvalues under the unspiked (H_0) distribution P_r and the spiked (H_1) distribution Q_r.

5.4 Two Wishart Matrices

5.4.1 *Fixed r, Finite n: Exact Distribution.*

Real case. Let $\mathbf{X}_i \stackrel{iid}{\sim} \mathcal{N}_r(\mathbf{0}, \mathbf{\Sigma})$, $i = 1, 2, \ldots, n$, and let $\mathcal{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_n)$ be an $(r \times n)$-matrix. Let $\mathbf{A} = \mathcal{X}\mathcal{X}^\tau \sim \mathcal{W}_r(n, \mathbf{\Sigma})$. Suppose we have another $(r \times r)$-matrix $\mathbf{B} \sim \mathcal{W}_r(m, \mathbf{\Sigma})$ that is independent of $\mathbf{A}$. Because we are interested in the eigenvalues of $\mathbf{A}^{-1}\mathbf{B}$ and because the distribution of those eigenvalues does not depend upon $\mathbf{\Sigma}$, without loss of generality, we can take $\mathbf{\Sigma} = \mathbf{I}_r$.

So, suppose we have two independent white Wishart matrices, $\mathbf{A} \sim \mathcal{W}_r(n, \mathbf{I}_r)$ and $\mathbf{B} \sim \mathcal{W}_r(m, \mathbf{I}_r)$, If $m, n \geq r$, then both $\mathbf{A}$ and $\mathbf{B}$ are invertable as is also their sum $\mathbf{A} + \mathbf{B}$. We are interested in solving (for λ) the following generalized eigenvalue problem,

$$|\mathbf{B} - \lambda(\mathbf{A} + \mathbf{B})| = 0. \tag{36}$$

That is, we are interested in the eigenvalues of $(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B}$. Because $\mathbf{A}$ is positive definite, it follows that $0 < \lambda < 1$ in (36). The eigenequation (36) can be reexpressed as

$$|\mathbf{B} - \theta\mathbf{A}| = 0, \tag{37}$$

and, in this form, we are interested in the eigenvalues of $\mathbf{A}^{-1}\mathbf{B}$ (or of the symmetric version $\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}$). The eigenvalues λ and θ are related by $\lambda = \theta/(1 + \theta)$ or $\theta = \lambda/(1 - \lambda)$.

The exact joint distribution of the eigenvalues of the generalized eigenequation (36) is given by

$$p(\lambda_1, \ldots, \lambda_r) = c_{m,n,r} \prod_i [w_{a,b}(\lambda_i)]^{1/2} \prod_{i<j} |\lambda_i - \lambda_j|, \tag{38}$$

where

$$w_{a,b}(\lambda) = \lambda^a (1 - \lambda)^b, \quad a = n - r - 1, \; b = m - r - 1, \tag{39}$$

is a weight function for the Jacobi family of orthogonal polynomials (Abromowitz and Stegun, 1970, Table 22.2), and c is a normalizing constant that depends upon m, n, and r.

The general β-form of (38) is given by

$$p_\beta(\lambda_1, \ldots, \lambda_r) = c_{m,n,r,\beta} \prod_i [w_{r,m,n,\beta}(\lambda_i)]^{1/2} \prod_{i<j} |\lambda_i - \lambda_j|^\beta, \tag{40}$$

where

$$w_{r,m,n,\beta}(\lambda) = \lambda^a (1 - \lambda)^b, \quad a = \beta(n - r + 1) - 2, \; b = \beta(m - r + 1) - 2, \tag{41}$$

and $c_{m,n,r,\beta}$ is a normalizing constant that depends upon m, n, r, and β. As before, $\beta = 1$ for the real case, $\beta = 2$ for the complex case, and $\beta = 4$ for the quaternion case.

Setting $\lambda_i = \theta_i/(1 + \theta_i)$, the joint distribution of the eigenvalues of the generalized eigenequation (37) is given by

$$p(\theta_1, \ldots, \theta_r) = c_{m,n,r} \prod_i [w_{a,b}(\theta_i)]^{1/2} \prod_{i<j} |\theta_i - \theta_j|, \tag{42}$$

where

$$w_{a,b}(\theta) = \theta^a (1 + \theta)^b, \quad a = n - r - 1, \quad b = m + n. \tag{43}$$

Proofs of these classical results can be found in Anderson (1984, Section 13.2) or Muirhead (1982, Section 3.3).

If we carry out a change of variable in (42) and (43) by setting $\theta = (1 + x)/2$, we obtain the *Jacobi orthogonal ensemble*,

$$p(x_1, \cdots, x_r) = c_{m,n,r} \prod_i [w_{a,b}(x_i)]^{1/2} \prod_{i<j} |x_i - x_j|, \tag{44}$$

TABLE 2

Families of orthogonal polynomials and their weight functions $w(x)$.

Case	$w(x)$	Interval	OrthoPoly
Gaussian	e^{-x^2}	$(-\infty, \infty)$	Hermite
Wishart	$x^a e^{-x}$	$[0, \infty)$	Laguerre
Two Wisharts	$x^a(1-x)^b$	$(0, 1)$	Jacobi

where

$$w_{a,b}(x) = (1-x)^a(1+x)^b, \quad a = n-r-1, \quad b = m-r-1, \tag{45}$$

is the weight function for the Jacobi family of orthogonal polynomials (Abromowitz and Stegun, 1970, Table 22.2).

Complex case. We write $\mathbf{X} \sim \mathcal{N}_r^C(\mathbf{0}, \mathbf{\Sigma})$, for a complex-valued random r-vector having a multivariate Gaussian distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$. If $\mathbf{X}_i \overset{iid}{\sim} \mathcal{N}_r^C(\mathbf{0}, \mathbf{\Sigma})$, $i = 1, 2, \ldots, n$, then

$$\mathcal{X}\mathcal{X}^* = \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^* \sim \mathcal{W}_r^C(n, \mathbf{\Sigma}), \tag{46}$$

where $\mathcal{X}^*$ denotes the complex-conjugate transpose of $\mathcal{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_n)$.

Suppose now that we have two independent complex-Wishart $(r \times r)$ random matrices, $\mathbf{A} \sim \mathcal{W}_r^C(n, \mathbf{I}_r)$ and $\mathbf{B} \sim \mathcal{W}_r^C(m, \mathbf{I}_r)$. Then, the exact joint density of the eigenvalues $1 \geq \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq 0$ of $(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B}$ is given by (Khatri, 1964; James, 1964)

$$p(\lambda_1, \cdots, \lambda_r) = c_{m,n,r} \prod_i w(\lambda_i) \prod_{i<j} |\lambda_i - \lambda_j|^2. \tag{47}$$

where

$$w_{a,b}(\lambda) = \lambda^a(1-\lambda)^b, \quad a = n-r, \; b = m-r, \tag{48}$$

and

$$c_{m,n,r} = \prod_{i=1}^{r} \frac{\Gamma(m+n+r+i)}{\Gamma(i)\Gamma(i+m)\Gamma(i+n)}. \tag{49}$$

See also Johnstone (2008).

5.4.2 *Application to Multivariate Analysis.* The eigenproblem (37) is of interest in multivariate reduced-rank regression (Izenman, 2013, Chapter 6), which includes as special cases canonical variate and correlation analysis, and linear discriminant analysis. See Johnstone (2008) and Kargin (2015). Related work on estimation and hypothesis-testing situations in multivariate analysis of variance (MANOVA) has also appeared recently.

Multivariate Regression. Suppose $\mathcal{X}_c$ is an $(r \times n)$-matrix and $\mathcal{Y}_c$ is an $(s \times n)$-matrix, where the subscript c indicates that both $\mathcal{X}$ and $\mathcal{Y}$ are centered (by subtracting out row means from each row), and where we assume $s \leq r$. Set $\mathbf{S}_{XX} = \mathcal{X}_c\mathcal{X}_c^\tau$, $\mathbf{S}_{YY} = \mathcal{Y}_c\mathcal{Y}_c^\tau$, and $\mathbf{S}_{XY} = \mathcal{X}_c\mathcal{Y}_c^\tau = \mathbf{S}_{YX}^\tau$. Then set $\mathbf{B} = \mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY}$ to be the variation due to the multivariate regression and $\mathbf{A} = \mathbf{S}_{YY} - \mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY}$ to be the residual variation, so that $\mathbf{A} + \mathbf{B} = \mathbf{S}_{YY}$. We have that $\mathbf{B} \sim \mathcal{W}_s(r, \mathbf{\Sigma}_{YY})$ and $\mathbf{A} \sim \mathcal{W}_s(n-r-1, \mathbf{\Sigma}_{YY})$. The eigenequation (36), thus, boils down to the following:

$$|\mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY} - \lambda\mathbf{S}_{YY}| = 0, \tag{50}$$

so that we are interested in the eigenvalues of $\mathbf{S}_{YY}^{-1}\mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY}$ (or of $\mathbf{S}_{YY}^{-1/2}\mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1/2}$, its symmetric version,). The joint distribution of the eigenvalues of the generalized eigenequation (36) is given by

$$p(\lambda_1, \ldots, \lambda_s) = c_{n,r,s} \prod_{i=1}^{s} [w_{a,b}(\lambda_i)]^{1/2} \prod_{i<j}^{s} |\lambda_i - \lambda_j|, \tag{51}$$

where

$$w_{a,b}(\lambda) = \lambda^a (1-\lambda)^b, \tag{52}$$

$$a = r - s - 1, \quad b = n - r - s - 2, \tag{53}$$

and c is a normalizing constant that depends upon n, r, and s (Anderson, 1984, Section 13.4).

Multivariate Analysis of Variance. In Section 5.2.4, we saw the Marčenko-Pastur result for a single Wishart matrix that the sample eigenvalues tend to be more spread out than their corresponding population eigenvalues in high-dimensional situations, which indicates that the largest sample eigenvalues are biased upwards. This phenomenon has also been observed in a highly structured genetics application by Fan and Johnstone (2019), who studied MANOVA estimators of certain variance component covariance matrices in multivariate random (and mixed) effects models. They show that, in an asymptotic sense, the spectra of these estimators are closely approximated by a certain generalization of the Marčenko-Pastur result, from which the bias effect can be removed from the sample eigenvalues.

Bai, Choi, and Fujikoshi (2018b) also studied the one-way MANOVA model. The r-dimensional observations form q independent samples or groups, the ith group having n_i observations and the underlying distribution for the ith group is Gaussian with mean vector $\boldsymbol{\mu}_i$ and common covariance matrix $\boldsymbol{\Sigma}$. Let $\mathbf{S}_b$ denote the between-groups covariance matrix and $\mathbf{S}_e$ denote the within-groups covariance matrix. The total-variation covariance matrix is $\mathbf{S}_t = \mathbf{S}_b + \mathbf{S}_e$. Under the null hypothesis of equal group means (i.e., $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_{q+1}$), $\mathbf{S}_b$ and $\mathbf{S}_e$ are independent Wishart matrices. Bai et al. derived the asymptotic joint distributions of the (normalized) eigenvalues of $\mathbf{S}_b\mathbf{S}_e^{-1}$ under null and nonnull formulations, when both $r, n \to \infty$ and $r/n \to \gamma \in (0,1)$, without assuming an underlying Gaussian distribution. They then applied those results to obtain the high-dimensional asymptotic distributions of some classical MANOVA test criteria that are defined in terms of functions of the eigenvalues of either $\mathbf{S}_b\mathbf{S}_e^{-1}$ or $\mathbf{S}_b\mathbf{S}_t^{-1}$. They showed that each of the suitably normalized versions of those test statistics converges in distribution to $\mathcal{N}(0, 2q/(1-\gamma))$. See also Han, Pan, and Zhang (2016), who determined Tracy–Widom-type universality for the largest eigenvalue (suitably centered and scaled) in a MANOVA setting.

6. EDGES OF THE SPECTRUM

Perhaps the most exciting results to have been derived from RMT are the Tracy–Widom laws for the distribution of the appropriately-normalized largest eigenvalue of a random matrix.

6.1 Largest Eigenvalue: Gaussian Ensembles

Complex case. In the GUE case, the distributional results turn out to be quite simple. Suppose that an $(n \times n)$ complex Gaussian Wigner matrix $\mathbf{X}$ is Hermitian (i.e., $\mathbf{X} = \mathbf{X}^*$) with diagonal elements $X_{jj} \overset{iid}{\sim} \mathcal{N}(0,2)$, and the real and imaginary parts of the off-diagonal entries X_{ij}, $i < j$, each iid $\mathcal{N}(0, \frac{1}{2})$. Tracy–Widom (1994) showed that the largest eigenvalue of $\mathbf{X}$ has the limiting distribution,

$$\lim_{n\to\infty} \mathrm{P}\left\{\sqrt{2}n^{1/6}\left(\widehat{\lambda}_1 - \sqrt{2n}\right) \leq t\right\} = F_2(t), \tag{54}$$

where F_2 is the *Tracy–Widom law of order 2* with distribution function,

$$F_2(t) = \exp\left\{-\int_t^\infty (x-t)[q(x)]^2 dx\right\}, \tag{55}$$

Table 3

Percentage points of the Tracy–Widom distributions for $\beta = 1, 2, 4$. Tabulated is the value of x such that $F_\beta(x) = P\{W_\beta < x\} = p$. (Adapted from Table 1 of Bejan, 2005.)

β	0.005	0.025	0.05	0.95	0.975	0.99	0.995	0.999
1	–4.1505	–3.5166	–3.1808	0.9703	1.4538	2.0234	2.4224	3.2724
2	–3.9139	–3.4428	–3.1945	–0.2325	0.0915	0.4776	0.7462	1.3141
4	–4.0531	–3.6608	–3.4556	–1.0904	–0.8405	–0.5447	–0.3400	0.0906

and q uniquely solves the Painlevé II ordinary differential equation,

$$q''(x) = xq(x) + 2[q(x)]^3, \quad q(x) \sim \text{Ai}(x) \text{ as } x \to \infty \tag{56}$$

for all x. The Airy function $\text{Ai}(x)$ satisfies

$$Ai''(x) = x \cdot Ai(x), \tag{57}$$

where $Ai''(x) = d^2 Ai(x)/dx^2$, with the boundary condition,

$$Ai(x) \sim \frac{e^{-\frac{2}{3}x^{3/2}}}{2\sqrt{\pi}x^{1/4}} \quad \text{as } x \to \infty. \tag{58}$$

In (56), $q(x) \sim \text{Ai}(x)$ as $x \to \infty$ means that $\lim_{x\to\infty} \frac{q(x)}{\text{Ai}(x)} = 1$ (Tracy and Widom, 1996). See also Johnstone and Ma (2012, Theorem 1). Table 3 gives the percentage points of the Tracy–Widom distributions for $\beta = 1, 2, 4$.

Let $K(x, y)$ denote the Airy kernel, which is defined as

$$K(x, y) = \frac{Ai(x)Ai'(y) - Ai'(x)Ai(y)}{x - y}, \qquad Ai'(x) = \frac{dAi(x)}{dx}. \tag{59}$$

Then, Tracy and Widom (1994) showed that the F_2 distribution can be expressed as a determinant:

$$F_2(t) = \det(\mathbf{I}_r - \mathbf{K}_t), \tag{60}$$

where $\mathbf{K}_t$ denotes the operator acting on $L^2(t, \infty)$ with kernel $K(x, y)$. See also Johnstone (2001).

Real case. For the GOE case, an $((n + 1) \times (n + 1))$ real Gaussian Wigner matrix $\mathbf{X}$ is symmetric (i.e., $\mathbf{X}^\tau = \mathbf{X}$), where, following Johnstone and Ma, $n+1$ is taken to be even, and its elements are distributed as $X_{ij} \sim \mathcal{N}(0, 1 + \delta_{ij})$, $i \leq j$, where $\delta_{ij} = 1$ if $i = j$, and 0 otherwise. Using results from the GUE case, Tracy and Widom (1996) established the asymptotic distribution of the largest eigenvalue for the GOE case. They showed that for the GOE case that

$$\lim_{n\to\infty} \text{P}\left\{\sqrt{2}n^{1/6}\left(\widehat{\lambda}_1 - \sqrt{2n+1}\right) \leq t\right\} = F_1(t), \tag{61}$$

where

$$[F_1(t)]^2 = F_2(t) \cdot \exp\left\{-\int_t^\infty q(x)dx\right\}, \tag{62}$$

$F_2(t)$ is given by (55), and $q(x)$ satisfies the Painlevé II differential equation that appears in (56). Johnstone and Ma (2012, Theorem 2) then determined centering and scaling constants that achieve $\mathcal{O}(n^{-2/3})$ rates of convergence to the Tracy–Widom limiting distributions of the largest eigenvalues for both GOE and GUE cases.

6.2 Largest Eigenvalue: Wishart–Laguerre Ensembles

Fill an $(r \times n)$-matrix $\mathcal{X}$ with iid $\mathcal{N}(0,1)$ deviates. Then, $\mathbf{S} = \mathcal{X}\mathcal{X}^\tau$ has the white Wishart distribution $\mathcal{W}_r(n, \mathbf{I}_r)$ and represents the null case (i.e., $\mathbf{\Sigma} = \mathbf{I}_r$). The eigenvalues of $\mathbf{S}$ are real and nonnegative. Denote the largest eigenvalue of $\mathbf{S}$ by $\widehat{\lambda}_1$.

6.2.1 *Fixed r, Finite n: Exact Distribution.* The exact distribution of $\widehat{\lambda}_1$ in the null case was found by Constantine (1963) and is expressed as an infinite expansion in zonal polynomials (i.e., hypergeometric functions of two matrix arguments). See Muirhead (1982, Chapter 7) for a detailed exposition of zonal polynomials. Unfortunately, such a series representation, which converges very slowly, is impractical for numerical computation and statistical usage.

6.2.2 *Large r, Large n: The Tracy–Widom Laws.* The development of RMT has provided us with the following useful results concerning the limiting distribution of the largest eigenvalue when the dimensions n and r of the matrix $\mathcal{X}$ are both very large.

Real case. Let

$$\mu_{nr} = (\sqrt{n-1} + \sqrt{r})^2, \tag{63}$$

and

$$\sigma_{nr} = (\sqrt{n-1} + \sqrt{r}) \left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{r}} \right)^{1/3} \tag{64}$$

be centering and scaling factors, respectively. Suppose both r and n are large and that $r/n \to \gamma \in (0, \infty)$. Johnstone (2001, 2007) showed that under the null model,

$$\frac{\widehat{\lambda}_1 - \mu_{nr}}{\sigma_{nr}} \overset{\mathcal{D}}{\Rightarrow} W_1 \sim F_1, \tag{65}$$

where F_1 is the *Tracy–Widom law of order 1* and has distribution function,

$$\begin{aligned} F_1(t) &= \exp\left(-\frac{1}{2} \int_t^\infty [q(x) + (x-t)(q(x))^2] dx \right) \\ &= [F_2(t)]^{1/2} \exp\left(-\frac{1}{2} \int_t^\infty q(x) dx \right). \end{aligned} \tag{66}$$

From (65), some authors write $TW_1(n,r)$ for the Tracy–Widom F_1 distribution of $\mu_{nr} + \sigma_{nr} W_1$, which can be used to approximate the distribution of $\widehat{\lambda}_1$. Johnstone showed that the asymptotic distribution result (65) is still useful for n and r as small as 10.

Remark 4. El Karoui (2003) extended the assumption $r/n \to \gamma \in (0, \infty)$ to include $\gamma = 0$ and $\gamma = \infty$. The extension to $\gamma = \infty$ is important in that it allows applications where $r \gg n$.

Remark 5. It turns out that the Gaussian assumption is not necessary. Soshnikov (2002) showed that if the Gaussian assumption on the elements of $\mathcal{X}$ is replaced by an assumption that $n - r = O(r^{1/3})$ as $r \to \infty$, and an assumption that the elements of the matrix $\mathcal{X}$ are symmetrically distributed with finite even moments and sufficiently light tails (i.e., a subGaussian distribution), then the Tracy–Widom Law (65) still holds. Under those conditions, the asymptotic distribution of the largest eigenvalue (suitably centered and scaled) is

$$\mathrm{P}\left\{ \left(\widehat{\lambda}_1 - (1 + \sqrt{\gamma})^2 \right) \frac{\gamma^{1/6} n^{2/3}}{(1+\sqrt{\gamma})^{4/3}} \leq x \right\} \to F_1(x), \quad x \in \mathbb{R}, \tag{67}$$

as $r \to \infty$. See also Baik, Ben Arous, and Péché (2005), Péché (2008), and Tao and Vu (2011a, 2011b).

Tracy and Widom (2000) and Johnstone (2001) show that F_1 has the following properties:

1. The F_1 distribution does not depend upon any parameters and, therefore, its role can be viewed in the same light as the role of the standard Gaussian distribution in the central limit theorem.
2. The F_1 density function is not symmetric, but is unimodal with mean approximately -1.21, standard deviation approximately 1.27, and different decay rates depending upon whether $x \to -\infty$ or $x \to +\infty$.
3. The standard deviation, σ_{nr}, increases with n as $n^{1/2}$.
4. Approximately 83% of the distribution lies below μ_{nr}.
5. Approximately 95% of the distribution lies below $\mu_{nr} + \sigma_{nr}$.
6. Approximately 99% of the distribution lies below $\mu_{nr} + 2\sigma_{nr}$.

The limiting distribution F_1 was discovered by Tracy and Widom to be one of a family of distributions, F_β, where $\beta = 1$ (real case), 2 (complex case), and 4 (real-quaternion case),

Complex case. The asymptotic distribution of the largest eigenvalue of an $(r \times r)$ complex Wishart matrix was actually found (Johansson, 2000) before the real-case result given by Johnstone (2001). Johansson showed that, for an $(r \times r)$ Hermitian (e.g., complex Wishart) matrix, as the size, r, of the matrix increases to ∞,

$$\frac{\widehat{\lambda}_1 - \mu_{nr}}{\sigma_{nr}} \overset{\mathcal{D}}{\Rightarrow} W_2 \sim F_2, \tag{68}$$

where F_2 is the *Tracy-Widom law of order 2* given by (55),

$$\mu_{nr} = (\sqrt{n} + \sqrt{r})^2, \tag{69}$$

and

$$\sigma_{nr} = (\sqrt{n} + \sqrt{r}) \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{r}} \right)^{1/3}. \tag{70}$$

In other words,

$$P\left\{ \frac{\widehat{\lambda}_1 - \mu_{nr}}{\sigma_{nr}} \leq x \right\} \to F_2(x), \quad \text{as } r \to \infty,\ x \in \mathbb{R}. \tag{71}$$

The real case with limiting distribution F_1 was found by Johnstone (2001) by using an independent approach with a different construction than was used in the complex case.

Remark 6. Ramírez, Rider, and Virág (2008) extended the Tracy–Widom Laws to all $\beta > 0$.

6.4 Largest Eigenvalue: Two Wishart Matrices

The problem of approximating the distribution of either the largest eigenvalue λ of $(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B}$ or the largest eigenvalue θ of $\mathbf{A}^{-1}\mathbf{B}$ when $\mathbf{A}$ and $\mathbf{B}$ are both $(r \times r)$-matrices and m, n, and r are large has been studied in detail by Johnstone (2008). Johnstone shows that with appropriate centering and scaling of the logit transform of λ, the Tracy-Widom laws F_1 and F_2 continue to hold, F_1 for the real case and F_2 for the complex case, as in the single-Wishart case above.

6.3.1 *Large r, Large m and n: The Tracy–Widom Laws.*

Real case. Suppose $\mathbf{A} \sim \mathcal{W}_r(m, \mathbf{I}_r)$ and $\mathbf{B} \sim \mathcal{W}_r(n, \mathbf{I}_r)$ are independent white Wishart matrices. Assume that $n \geq r$. Then, $\mathbf{A}$ is positive definite. If λ_{1r} is the largest eigenvalue of $(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B}$, then, $0 < \lambda_{1r} < 1$. The following results have been proved for r even only, but empirical results indicate that they may also hold for general r.

Following Johnstone (2008), let

$$\mu_r = 2 \log \tan\left(\frac{\phi+\gamma}{2}\right) \tag{72}$$

and

$$\sigma_r^3 = \frac{16}{(m+n-1)^2} \frac{1}{\sin^2(\phi+\gamma)\sin\phi\sin\gamma} \tag{73}$$

be centering and scaling factors, respectively, for

$$W_r = \text{logit}(\lambda_{1r}) = \log\left(\frac{\lambda_{1r}}{1-\lambda_{1r}}\right), \tag{74}$$

where the angle parameters, γ and ϕ, are defined by

$$\sin^2\left(\frac{\gamma}{2}\right) = \frac{\min(r,m)-1/2}{m+n-1} \tag{75}$$

$$\sin^2\left(\frac{\phi}{2}\right) = \frac{\max(r,m)-1/2}{m+n-1}, \tag{76}$$

respectively. If $m = m_r$, $n = n_r \to \infty$ as $r \to \infty$ in such a way that $\lim_{r\to\infty} \min(r,m)/(m+n) > 0$ and $r/n \to \xi < 1$, then,

$$\frac{W_r - \mu_r}{\sigma_r} \overset{\mathcal{D}}{\Rightarrow} Z_1 \sim F_1, \tag{77}$$

where F_1 is the Tracy-Widom law of order 1 given by (64).

A more precise convergence result can be made: As $m_r, n_r \to \infty$ as $r \to \infty$ through the even integers, then there exists a constant $C > 0$ depending upon (ϕ, γ) such that for large x,

$$\left| P\left\{\frac{W_r - \mu_r}{\sigma_r} \le x\right\} - F_1(x)\right| \le Cr^{-2/3}e^{-x/2}. \tag{78}$$

Complex case. Suppose that we have two independent complex-Wishart random matrices, $\mathbf{A} \sim \mathcal{W}_r^C(m, \mathbf{I}_r)$ and $\mathbf{B} \sim \mathcal{W}_r^C(n, \mathbf{I}_r)$. We are interested in the distribution of the largest eigenvalue of $(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B}$.

Let

$$W^C = \text{logit}(\lambda_r^C) = \log\left(\frac{\lambda_r^C}{1-\lambda_r^C}\right). \tag{79}$$

Assume that $m_r, n_r \to \infty$ as $r \to \infty$ in the same way as for the real case. Defining μ_r^C and σ_r^C as appropriate centering and scaling constants, Johnstone (2008) showed that

$$\frac{W_r^C - \mu_r^C}{\sigma_r^C} \overset{\mathcal{D}}{\Rightarrow} Z_2 \sim F_2, \tag{80}$$

where F_2 is the Tracy-Widom distribution of order 2 given by (55). Let

$$N = \min(m, r), \quad \alpha = n - r, \quad \beta = |m - r|. \tag{81}$$

The centering and scaling constants in (80) are given by

$$\mu_r^C = \frac{\tau_N^{-1}u_N + \tau_{N-1}^{-1}u_{N-1}}{\tau_N^{-1} + \tau_{N-1}^{-1}} \tag{82}$$

$$\frac{1}{\sigma_r^C} = \frac{1}{4}(\tau_N^{-1} + \tau_{N-1}^{-1}), \tag{83}$$

where

$$u_N = 2\log\tan\left(\frac{\phi_N+\gamma_N}{2}\right) \tag{84}$$

$$\tau_N^3 = \frac{16}{(2N+\alpha+\beta+1)^2}\frac{1}{\sin^2(\phi_N+\gamma_N)\sin\phi_N\sin\gamma_N}, \tag{85}$$

$$\sin^2\left(\frac{\gamma_N}{2}\right) = \frac{N+1/2}{2N+\alpha+\beta+1}, \tag{86}$$

$$\sin^2\left(\frac{\phi_N}{2}\right) = \frac{N+\beta+1/2}{2N+\alpha+\beta+1}. \tag{87}$$

Moreover, for large enough x, there exists a constant C depending upon (ϕ,γ) such that

$$\left|P\left\{\frac{W_r^C-\mu_r^C}{\sigma_r^C}\leq x\right\}-F_2(x)\right| \leq Cr^{-2/3}e^{-x/2}. \tag{88}$$

Remark 7. The Tracy-Widom laws have been found to be of such great importance in RMT that they have been said to play a similar role as that of the Gaussian distribution in classical statistical theory (see, e.g., Diaconis, 2003).

6.5 Some Applications of Random Matrix Theory

Each of the following examples in this Section have been modeled by RMT.

Example 1. *The length of the longest increasing subsequence of a random permutation of n objects as $n\to\infty$.* Consider a permutation π of the first n integers $\{1,2,\ldots,n\}$. We can write π as $\{\pi_1,\pi_2,\ldots,\pi_n\}$. Then, π has an increasing subsequence $\ell_n(\pi)$ of length k if there exist indices $1\leq i_i<i_2<\cdots<i_k\leq n$ such that $\pi_{i_1}<\pi_{i_2}<\cdots<\pi_{i_k}$. (By switching the directions of the inequality signs, a similar definition can be given for a decreasing subsequence of length k.)

Assume the $n!$ permutations each of length n are equally likely. For each such permutation, find the length $L_n=\ell_n(\pi)$ of the longest increasing subsequence. For example, let $n=8$ and consider the permutation $\pi=\{2,5,1,3,4,8,6,7\}$; the length of the longest increasing subsequence is $L_8=5$, given by $\{2,3,4,6,7\}$ and $\{1,3,4,6,7\}$. Thus, L_n is a random variable, but the actual subsequence may not be unique. The main questions are: what is the distribution of L_n, and what is its mean, $\mathrm{E}\{L_n\}$, and its variance, $\mathrm{var}\{L_n\}$, for large n?

It has been shown that $\lim_{n\to\infty}\frac{1}{\sqrt{n}}\mathrm{E}\{L_n\}$ exists (Ulam, 1961; Hammersley, 1972) and has the value 2 (Vershik and Kerov, 1977). Thus, for large n, $\mathrm{E}(L_n)\sim 2\sqrt{n}$. Recent work has also showed that

$$\lim_{n\to\infty}\frac{\mathrm{E}\{L_n\}-2\sqrt{n}}{n^{1/6}} = -1.711 \tag{89}$$

and

$$\lim_{N\to\infty}\frac{\mathrm{var}\{L_n\}}{n^{1/3}} = 0.902, \tag{90}$$

where the limiting constants on the rhs of (89) and (90) were obtained by Baik, Deift, and Johansson (1999). Furthermore, the large-n distribution of L_n has been shown (Baik, Deift, and Johansson, 1999) to be

$$P\left\{\frac{L_n-2\sqrt{n}}{n^{1/6}}\leq x\right\} \to F_2(x), \quad \text{as } n\to\infty, \quad x\in\Re, \tag{91}$$

where F_2 is the Tracy–Widom law. Thus, for large n, the distribution of the length of the longest increasing subsequence of a permutation π of $\{1,2,\ldots,n\}$, appropriately centered and scaled, is identical to the distribution of the largest eigenvalue of a random GUE matrix.

The longest increasing subsequence problem is related to the asymptotics of certain non-intersecting paths that describe a random tiling (by closed rectangles called dominoes) of the Aztec diamond of size n. An Aztec diamond is the union of lattice squares $[m, m+1] \times [\ell, \ell+1]$, $m, \ell \in \mathbb{Z}$, that lie inside the region $\{(x, y) : |x|+|y| \leq n+1\}$; It is called an Aztec diamond because its boundary looks like a Mexican pyramid; for details, see, e.g., Johansson (2002).

Example 2. *Patience-Sorting card game. Patience sorting* is an algorithm that derives from a variation of a one-person card game in Britain (called *solitaire* in the United States). It was named by Colin Mallows in 1962 (see Mallows, 1973), who attributed its invention to A.S.C. Ross. The resulting algorithm was developed by John Hammersley (in 1962 but not published until Hammersley, 1972), who recognized that it could be used to compute the length of the longest increasing subsequence (see Example 1 above). It also motivated the development of subadditive ergodic theory. See also Burstein and Lankham (2005).

As explained by Aldous and Diaconis (1999), the game starts out with a "deck" of cards labelled $1, 2, 3, \ldots, n$, and then the deck is shuffled and a card is drawn from the top of the deck and placed into a "pile" according to the following rule. A newly turned-up card can be placed on a card at the top of an existing pile only if its value is lower than that of the top card; otherwise, the new card starts another pile to the right of all existing piles. For example, a 3 can be placed on top of a 5, but a 6 starts a new pile. The game continues until all cards are dealt and placed into piles. The objective is to complete the game with as few piles as possible.

Consider the ordered eight cards from Example 1: 2, 5, 1, 3, 4, 8, 6, 7. Start the first pile with the 2; the 5 is bigger than the 2, so the 5 starts a new pile; the 1 goes on top of the 2, and the 3 goes on top of the 5; the 4 starts a new pile; the 8 starts a new pile; the 6 goes on top of the 8; and the 7 starts a new pile. The piles are as follows:

1	**3**		**6**	
2	5	**4**	8	**7**

This "greedy" strategy, which places each card on top of the most-leftwise pile possible, is the optimal strategy. In this example, we obtained five piles, which is the same result we obtained above for the longest increasing subsequence of the given permutation. In general, the number of piles using the optimal greedy strategy will always equal the length of the longest increasing subsequence.

Note that the top cards in the resulting piles (shown above in boldface) will not necessarily be in permutation order and, hence, will not be an increasing subsequence. For example, consider the sequence 8, 6, 1, 3, 4, 7, 5, 2. The piles are:

1			
6	**2**		**5**
8	3	**4**	7

Clearly, the boldface numbers 1, 2, 4, 5 do not form an increasing subsequence of the original sequence. However, the number of piles (4) is equal to the length (4) of the subsequences 1, 3, 4, 5 or 1, 3, 4, 7.

This correspondence between patience sorting and the length of the longest increasing subsequence yields a very efficient algorithm for computing $L_n = \ell_n(\pi)$.

Example 3. *The spacings of consecutive zeroes of the Riemann zeta function.* Riemann's zeta function is defined by adding up inverse powers of the positive integers:

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \cdots \tag{92}$$

If we set $s = 2$, for example, we have that $\zeta(2) = \pi^2/6$, a result discovered by Leonhard Euler. Euler showed that the zeta function could also be reexpressed as a product over the prime numbers:

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1} = \frac{1}{(1 - \frac{1}{2^s})(1 - \frac{1}{3^s})(1 - \frac{1}{5^s})(1 - \frac{1}{7^s}) \cdots} \tag{93}$$

This is known as the *Euler product formula.* Thus, the zeta function has properties that are intimately connected to the distribution of prime numbers. However, more is true. Riemann showed that the zeta function could be written as a product over its zeroes in the complex plane:

$$\zeta(s) = f(s)\left(1 - \frac{s}{\rho_1}\right)\left(1 - \frac{s}{\rho_2}\right)\left(1 - \frac{s}{\rho_3}\right)\cdots, \tag{94}$$

where $\rho_1, \rho_2, \ldots$ are the complex numbers for which $\zeta(s) = 0$, and $f(s)$ is a simple "fudge factor."

The zeta function has "trivial" zeroes at the negative even integers (i.e., at $s = -2, -4, -6, \ldots$). Riemann showed that all the nontrivial zeroes can be found in an infinite "critical" strip lying above and below the unit interval $0 < Re(s) < 1$ in the complex plane. Moreover, the zeroes are symmetrically located: each zero lying above the unit interval has a mirror-image zero lying below the unit interval. The Riemann Hypothesis says that all the zeroes of the zeta function lie on the vertical line through $Re(s) = \frac{1}{2}$, in which case a zeta zero can be written as $\rho_n = \frac{1}{2} \pm i\gamma_n$, for an ordered sequence of real, positive constants $\{\gamma_n\}$, where $\gamma_n \sim 2\pi n/(\log\ n)$ as $n \to \infty$. George Pólya conjectured around 1913 that the $\{\gamma_n\}$ are the eigenvalues of a certain type of self-adjoint operator and, hence, are real. David Hilbert entered into the discussion and introduced the term "spectrum" for the set of eigenvalues of a Hermitian operator analogous to that of optical spectra found in atoms. In 1972, an important meeting between Hugh Montgomery and Freeman Dyson revealed a surprising connection between spacings of the zeroes of the Riemann zeta function and RMT in the form of GUE eigenvalues (Montgomery, 1973).

Over the next few decades, attempts were made to check the Riemann Hypothesis using numerical methods. Andrew Odlyzko developed extremely efficient algorithms for computing zillions of zeroes in the critical strip, and showed that they all satisfy the Riemann Hypothesis. Odlyzko's data showed that the spacings between consecutive zeroes of the zeta function behave, statistically, like the spacings between consecutive eigenvalues of large, random matrices from the GUE, which added much credibility to Montgomery's proof (Odlyzko, 1987). In other words, the zeroes of the zeta function can be viewed as having a "spectral" interpretation, which agrees with the belief (by physicists) that the zeroes of the zeta function can be interpreted as energy levels in some quantum chaos system. In fact, Hilbert and Pólya had conjectured that the Riemann Hypothesis is true precisely because the zeroes of the zeta function correspond to eigenvalues of a positive linear (Hermitian) operator. For special cases of the zeta function that have been proved, the statistical properties of the eigenvalue spacings and the spacings of the zeroes of the zeta function turn out to be identical.

7. DISCUSSION

This paper reviewed the main ideas and applications of RMT, of which there has been an enormous amount of interest and research in recent years. We now provide some information on software implementations for computing the various types of laws and simulating from the appropriate distributions. We also provide a brief list of recent books and monographs that focus on RMT. We note that there is also a journal *Random Matrices: Theory and Applications* devoted to this topic at the website `worldscientific.com/worldscinet/rmta`.

Several computational packages include routines for computing Marčenko–Pastur "semicircle"-type laws, Wishart-matrix simulations, eigenvalues of a white Wishart matrix, and Tracy–Widom distributions. Iain Johnstone's R package `RMTstat` (Johnstone, Ma, Perry, and Shahram, 2014) is available at `CRAN.R-project.org/package=RMTstat`. There is also N. Raj Rao's `RMTool` (Rao and Edelman, 2007), a publicly available MATLAB symbolic toolbox, which is used to compute the limiting spectral density of a large class of random matrices and can be downloaded from the website `www.mit.edu/~raj/rmtool`. Computation in MATLAB of the Tracy-Widom distributions is slow and cubic spline approximations are preferred (Bejan, 2005).

The classic book in this area is Mehta (2004), now in its third edition. Although it deals at great length with Gaussian ensembles, it does not mention Laguerre ensembles and Wishart matrices and their important roles in mathematical statistics. The book by Porter (1965) is a collection of all the important papers published on RMT prior to 1965. More recent books and monographs include Guionnet (2008), Bai and Silverstein (2010), Anderson, Guionnet, and Zeitouni (2009), Tao (2012), and Erdős and Yau (2017). Edited volumes on RMT include Mezzadri and Snaith (2010) and Akemann, Baik, and Di Francesco (2011).

The entire issue of *The Annals of Statistics* for December 2008 was taken up with the topic of RMT and its use in high-dimensional inference.

An excellent historical account of RMT can be found in Forrester, Snaith, and Verbaarschot (2003), which is acually a Preface to an special issue of the *Journal of Physics* on RMT. For a more technical review of RMT, see Bai (1999). Other excellent reviews of this field include Edelman and Rao (2005) and Paul and Aue (2014).

ACKNOWLEDGEMENTS

The author thanks Cheng Yong Tang for suggesting that a 2008 dormant manuscript on RMT be updated and submitted for publication, Brian Rider for helpful discussions, and Sanat Sarkar for valuable comments and for identifying some inconsistencies in notation. Thanks also go to the two anonymous referees and the Associate Editor whose comments greatly improved this article.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I.A. (1970). *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*, New York: Dover Publications, Inc.

AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle, In: *2nd International Symposium on Information Theory* (B.N. Petrov and F. Csáki, eds.), pp. 267–281, Budapest: Akadémia Kiado.

AKEMANN, G., BAIK, J., and DI FRANCESCO, P. (2011). *The Oxford Handbook of Random Matrix Theory*, Oxford University Press.

ALDOUS, D. and DIACONIS, P. (1999). Longest increasing subsequences from patience sorting to the Baik–Deift–Johansson theorem, *Bulletin of the American Mathematical Society*, New Series, **36**, 413–432.

ANDERSON, G.W., GUIONNET, A., and ZEITOUNI, O. (2009). *An Introduction to Random Matrices*, Cambridge, U.K.: Cambridge University Press.

ANDERSON, T.W. (1963). Asymptotic theory for principal component analysis, *Annals of Mathematical Statistics*, **34**, 122–148.

ANDERSON, T.W. (1984). *An Introduction to Multivariate Statistical Analysis, Second Edition*, New York: Wiley.

ARNOLD, L. (1967). On the asymptotic distribution of the eigenvalues of random matrices, *Journal of Mathematical Analysis and Applications*, **20**, 262–268.

BACHMAT, E., BEREND, D., SAPIR, L., SKIENA, S., and STOLYAROV, N. (2006). Letter to the Editor: Analysis of aeroplane boarding via spacetime geometry and random matrix theory, *Journal of Physics A: General Physics.*

BAI, Z.D. (1999). Methodologies in spectral analysis of large dimensional random matrices, a review, *Statistica Sinica*, **9**, 611–677.

BAI, Z.D. and DING, X. (2012). Estimation of spiked eigenvalues in spiked models, *Random Matrices: Theory and Applications*, **1**, 1150011 (21 pages).

BAI, Z. and SILVERSTEIN, J.W. (2010). *Spectral Analysis of Large Dimensional Random Matrices, Second Edition*, New York: Springer.

BAI, Z.D. and YIN, Y.Q. (1988). Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix, *The Annals of Probability*, **16**, 1729–1741.

BAI, Z.D. and YIN, Y.Q. (1993). Limit of the smallest eigenvalue of large dimensional covariance matrix, *Annals of Probability*, **21**, 1275–1294.

BAI, Z., CHOI, K.P., and FUJIKOSHI, Y. (2018a). Consistency of AIC and BIC in estimating the number of significant components in high-dimensional principal component analysis, *The Annals of Statistics*, **46**, 1050–1076.

BAI, Z., CHOI, K.P., and FUJIKOSHI, Y. (2018b). Limiting behavior of eigenvalues in high-dimensional MANOVA via RMT, *The Annals of Statistics*, **46**, 2985–3013.

BAI, Z., SILVERSTEIN, J., and YIN, Y.Q. (1988). A note on the largest eigenvalue of a large dimensional sample covariance matrix, *Journal of Multivariate Analysis*, **26**, 166–168.

BAIK, J. (2003). Limiting distribution of last passage percolation models, *Proceedings.*

BAIK, J., BEN AROUS, G., and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices, *The Annals of Probability*, **33**, 1643–1697.

BAIK, J., BORODIN, A., DEIFT, P., and SUIDAN, T. (2006). A model for the bus system in Cuernavaca (Mexico), *Journal of Physics A*, **39**, 8965–8975.

BAIK, J., DEIFT, P., and JOHANSSON, K. (1999). On the distribution of the length of the longest increasing subsequence of random permutations, *Journal of the American Mathematical Society*, **12**, 1119–1178.

BAIK, J. and SILVERSTEIN, J.W. (2006). Eigenvalues of large sample covariance matrices of spiked population models, *Journal of Multivariate Analysis*, **97**, 1382–1408.

BAIK, J. and SILVERSTEIN, J.W. (2006). *Spectral Analysis of Large Dimensional Random Matrices*, New York: Springer.

BEENAKKER, C.W.J. (1997). Random-matrix theory of quantum transport, *Reviews of Modern Physics*, **69**, 731–816.

BEJAN, A.I. (2005). Largest eigenvalues and sample covariance matrices. Tracy–Widom and Painleve II: computational aspects and realization in S-Plus with applications, available at `cl.cam.ac.uk/~aib29/Tracy-Widom.pdf`.

BELLMAN, R. (1960). *Introduction to Matrix Analysis*, New York: McGraw-Hill Book Co.

BIRNBAUM, A., JOHNSTONE, I.M., NADLER, B., and PAUL, D. (2018). Minimax bounds for sparse PCA with noisy high-dimensional data, *The Annals of Statistics*, **41**, 1055–1084.

BLOEMENDAL, A., KNOWLES, A., YAU, H.-T., and YIN, J. (2015). On the principal components of sample covariance matrices, unpublished manuscript, available at `arXiv.org/pdf/1404.0788.pdf`.

BOHIGAS, O., GIANNONI, M.J., and SCHMIT, C. (1984). Characterizations of chaotic quantum spectra and universality of level fluctuation laws, *Physics Reviews Letters*, **52**, 1–4.

BRILLINGER, D.R. (1975). *Time Series: Data Analysis and Theory*, New York: Holt, Rinehart & Winston, Inc.

BURSTEIN, A. and LANKHAM, I. (2005). Combinatorics of patience sorting piles. *Proceedings of Formal Power Series and Algebraic Combinatorics*, Taormina, Italy. Available at `arxiv.org/abs/math/0506358v1.pdf`.

CAPITAINE, M., DONATI-MARTIN, C., and FERAL, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices, *The Annals of Probability*, **37**, 1–47.

CONSTANTINE, A.G. (1963). Some noncentral distribution problems in multivariate analysis, *Annals of Mathematical Statistics*, **34**, 1270–1285.

DEIFT, P. (2007). Universality for mathematical and physical systems, In *Proceedings of the International Congress of Mathematicians*, **I**, 125–152.

DELVAUX, S. and KUIJLAARS, A. (2010). A graph-based equilibrium problem for the limiting distribution of nonintersecting Brownian motions at low temperature, *Constructive Approximation*, **32**, 467–512.

DESHPANDE, Y. and MONTANARI, A. (2014). Information-theoretically optimal sparse PCA, *IEEE International Symposium on Information Theory*, 2197–2201.

DIACONIS, P. (2003). Patterns in eigenvalues: The 70th Josiah Willard Gibbs Lecture, *Bulletin of the American Mathematical Society*, **40**, 155–178.

DOBRIBAN, E. (2017). Sharp detection in PCA under correlations: all eigenvalues matter, *The Annals of Statistics*, **45**, 1810–1833.

DONOHO, D., GAVISH, M., and JOHNSTONE, I. (2018). Optimal shrinkage in the spiked covariance model, *The Annals of Statistics*, **46**, 1742–1778.

DYSON, F.J. (1962). The threefold way. Algebraic structure of symmetry groups and ensembles in quantum mechanics, *Journal of Mathematical Physics*, **3**, 1199–1215.

EDELMAN, A. and RAO, N.R. (2005). Random matrix theory, *Acta Numerica*, 1–65.

EL KAROUI, N. (2003). On the largest eigenvalue of Wishart matrices with identity covariance when n, p and p/n tend to infinity, available at
`arXiv:math.ST/0309355`.

EL KAROUI, N. (2007). Tracy–Widom limit for the largest eigenvalue of a large class of complex Wishart matrices, *The Annals of Probability*, **35**, 663–714.

EL KAROUI, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory, *The Annals of Statistics*, **36**, 2757–2790.

ERDŐS, L. (2010). Universality of Wigner random matrices: a survey of recent results, unpublished manuscript, 122 pages, available at
`www.mathphys.org/AZschool/material/AZ10-erdos.pdf`.

ERDŐS, L. and YAU, H-T. (2017). *A Dynamical Approach to Random matrix Theory*, Courant Lecture Notes in Mathematics, **28**, American Mathematical Society.

ERDŐS, L., PÉCHÉ, S., RAMÍREZ, J., SCHLEIN, B., and YAU, H.-T. (2010) Bulk universality for Wigner matrices, *Communications in Pure and Applied Mathematics*, **63**, 895–925.

FAN, Z. and JOHNSTONE, I.M. (2019). Eigenvalue distributions of variance components estimators in high-dimensional random effects models, *The Annals of Statistics*, **47**, 2855–2886.

FORRESTER, P.J., SNAITH, N.C., and VERBAARSCHOT, J.J.M. (2003). Developments in random matrix theory, *Journal of Physics A. Special Edition on Random Matrix Theory*, available at
`arXiv:cond-mat/0303207 v1 11 Mar 2003`.

GEMAN, S. (1980). A limit theorem for the norm of random matrices, *The Annals of Probability*, **8**, 252–261.

GRENANDER, U. (1963). *Probabilities on Algebraic Structures*, New York: John Wiley & Sons, Inc.

GUIONNET, A. (2008). *On Random Matrices.* Lecture Notes in Mathematics, New York: Springer.

HAMMERSLEY, J. (1972). A few seedlings of research, *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, **1** (L.M. LeCam, J. Neyman, and E.L. Scott, eds.), Berkeley, CA: University of California Press, pp. 345–394.

HAN, X., PAN, G., and ZHANG, B. (2016). The Tracy–Widom law for the largest eigenvalue of F type matrices, *The Annals of Statistics*, **44**, 1564–1592.

IZENMAN, A.J. (2013). *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*, New York: Springer.

JAMES, A.T. (1960). The distribution of the latent roots of the covariance matrix, *The Annals of Mathematical Statistics*, **31**, 151–158.

JAMES, A.T. (1961). Zonal polynomials of the real positive definite matrices, *Annals of Mathematics*, **74**, 456–469.

JAMES, A.T. (1964). Distributions of matrix variates and latent roots derived from normal samples, *The Annals of Mathematical Statistics*, **35**, 475–501.

JOHANSSON, K. (2000). Shape fluctuations and random matrices, *Communications in Mathematical Physics*, **209**, 437–476.

JOHANSSON, K. (2002). Non-intersecting paths, random tilings, and random matrices, *Probability Theory and Related Fields*, **223**, 225–280.

JOHNSTONE, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis, *The Annals of Statistics*, **29**, 295–327.

JOHNSTONE, I.M. (2007). High-dimensional statistical inference and random matrices. In *Proceedings of the International Congress of Mathematicians*, **I**, 307–333. Eur. Math. Soc., Zürich.

JOHNSTONE, I.M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits, and rates of convergence, *The Annals of Statistics*, **36**, 2638–2716.

JOHNSTONE, I.M. and LU, A.Y. (2009). On consistency and sparsity for principal components analysis in high dimensions, *Journal of the American Statistical Association*, **104**, 682–693.

JOHNSTONE, I.M. and MA, Z. (2012). Fast approach to the Tracy–Widom law at the edge of GOE and GUE, *Annals of Applied Probability*, **22**, 1962–1988.

JOHNSTONE, I.M. and ONATSKI, A. (2019). Testing in high-dimensional spiked models, to appear in *The Annals of Statistics*.

JOHNSTONE, I.M., MA, Z., PERRY, P.O., and SHAHRAM, M. (2014). RMTstat: Distributions, Statistics and Tests derived from Random Matrix Theory,
`cran.r-project.org/web/packages/RMTstat/index.html`.

KARGIN, V. (2015). On estimation in the reduced-rank regression with a large number of responses and predictors, *Journal of Multivariate Analysis*, **140**, 377–394.

KHATRI, C.G. (1964). Distribution of the largest or the smallest characteristic root under null hypothesis concerning complex multivariate normal populations, *Annals of Mathematical Statistics*, **35**, 1807–1810.

KIECHERBAUER, T., MARKLOF, J., and SOSHNIKOV, A. (2001). Random matrices and quantum chaos, *Proceedings of the National Academy of Sciences*, **98**, 10531-10532.

KRBALEK, M. and SEBA, P. (2000). The statistical properties of the city transport in Cuernavaca (Mexico) and random matrix ensembles, *Journal of Physics A: Mathematical and General*, **33**, L229–L234.

KULESZA, A. and TASKAR, B. (2012). Determinantal point processes for machine learning, *Foundations and Trends in Machine Learning*, **5**, 123–286.

LIVSHYTS, G.V., TIKHOMIROV, K., and VERSHYNIN, R. (2019). The smallest singular value of inhomogeneous square random matrices, unpublished manuscript, available at `arXiv:1909.04219`.

MALLOWS, C.L. (1973). Patience sorting, *Bulletin of the Institute of Mathematics and Its Applications*, **9**, 216–224. Originally introduced by Mallows in 1962 as "Problem 62-2. Patience Sorting," *SIAM Review*, **4**, 148–149.

MARCENKO, V.A, and PASTUR, L.A. (1967). Distributions of eigenvalues of some sets of random matrices, *Math. USSR-Sb*, **1**, 507–536.

MEHTA, M.L. (2004). *Random Matrices, 3rd Edition*, Pure and Applied Mathematics (Amsterdam), 142, Amsterdam, Netherlands: Elsevier/Academic Press.

MEZZADRI, F. and SNAITH, N.C. (eds.) (2010). *Recent Perspectives in Random Matrix Theory and Number Theory*, Cambridge, U.K.: Cambridge University Press.

MINGO, J.A. and SPEICHER, R. (2017). *Free Probability and Random Matrices*, Fields Institute Monograph, **35**, 342 pp.

MIOLANE, L. (2019). Phase transitions in spiked matrix estimation: information-theoretic analysis, unpublished manuscript, available at `arXiv:1806.04343`.

MONTANARI and RICHARD (2016). Non-negative principal component analysis: message passing algorithms and sharp asymptotics, *IEEE Transactions on Information Theory*, **62**. Available at `arxiv.org/abs/1406.4775`.

MONTGOMERY, H.L. (1973). The pair correlation of zeroes of the zeta function, In: *Analytic Number Theory*, Proceedings of the Symposium on Pure Mathematics XXIV, Providence, RI: American Mathematical Society, pp. 181–193.

MUIRHEAD, R.J. (1982). *Aspects of Multivariate Statistical Theory*, New York: Wiley.

NADLER, B. (2008). Finite sample approximation results for principal component analysis: A matrix perturbation approach, *The Annals of Statistics*, **36**, 2791–2817.

ODLYZKO, A.M. (1987). On the distribution of spacings between zeros of the zeta function, *Mathematical Computation*, **48**, 273–308.

ONATSKI, A., MOREIRA, M.J., and HALLIN, M. (2013). Asymptotic power of sphericity tests for high-dimensional data, *The Annals of Statistics*, **41**, 1204–1231.

PAUL, D. (2004). Asymptotics of the leading sample eigenvalues for a spiked covariance model, Technical report, Stanford University. Available at `anson.ucdavis.edu/~debashis/techrep/eigenlimit.pdf`.

PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model, *Statistica Sinica*, **17**, 1617–1642.

PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: A review, *Journal of Statistical Planning and Inference*, **150**, 1–29.

PÉCHÉ, S. (2003). Universality of local eigenvalue statistics for random sample covariance matrices, Thèse No. 2716. École Polytechnique Fédérale de Lausanne.

PÉCHÉ, S. (2008). Universality results for largest eigenvalues of some sample covariance matrix ensembles, *Probability Theory and Related Fields*, **143**, 481–516.

PERRY, A., WEIN, A.S., BANDEIRA, A.S., and MOITRA, A. (2018). Optimality and sub-optimality of PCA I: Spiked random matrix models, *The Annals of Statistics*, **46**, 2416–2451.

PORTER, C.E. (1965). *Statistical Theory of Spectra: Fluctuations*, New York: Academic Press.

PORTER, C.E. and ROSENZWEIG, N. (1960). Statistical properties of atomic and nuclear spectra, *Annals of the Acad. Sci. Fennicae, Serie A VI Physica*, **44**, 1.

RAMÍREZ, J., RIDER, B., and VIRÁG, B. (2008). Beta ensembles, stochastic Airy spectrum, and a diffusion, available at `arXiv:math/0607331v3`.

RAO, J.R. and EDELMAN, A. (2007). The polynomial method for random matrices, *Foundations for Computational Mathematics*,
available at `web.eecs.umich.edu/~rajnrao/rmtool/`.

SCHWARZ, G. (1978). Estimating the dimension of a model, *The Annals of Statistics*, **6**, 461–464.

ŠEBA, P. (2009). Parking and the visual perception of space, *Journal of Statistical Mechanics: Theory and Experiment*, L10002.

SILVERSTEIN, J.W. (1985). The smallest eigenvalue of a large dimensional Wishart matrix, *Journal of Multivariate Analysis*, **12**, 1364–1368.

SILVERSTEIN, J.W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices, *Journal of Multivariate Analysis*, **55**, 331–339.

SINAI, Y. and SOSHNIKOV, A. (1998). A refinement of Wigner's semicircle law in a neighborhood of the spectrum edge, *Functional Analysis and Applications*, **32**, 114–131.

SOSHNIKOV, A. (1999). Universality at the edge of the spectrum in Wigner random matrices, *Communications in Mathematical Physics*, **207**, 697–733.

SOSHNIKOV, A. (2002). A note on universality of the distribution of the largest eigenvalue in certain sample covariance matrices, *Journal of Statistical Physics*, **108**, 1033–1056.

TAO, T. (2012). *Topics in Random Matrix Theory*, Graduate Studies in Mathematics, **132**, Providence, RI: American Mathematical Society.

TAO, T. and VU, V. (2010). Random matrices: universality of local eigenvalue statistics up to the edge, *Communications in Mathematical Physics*, **298**, 549–572.

TAO, T. and VU, V. (2011a). Random matrices: universality of ESDs and the circular law, *The Annals of Probability*, **38**, 2023–2065.

TAO, T. and VU, V. (2011b). Random matrices: universality of local eigenvalue statistics, *Acta Mathematika*, **206**, 127–204.

TAO, T. and VU, V. (2011c). The Wigner–Dyson–Mehta bulk universality conjecture for Wigner matrices, *Electric Journal of Probability*, **16**, 2104–2121.

TAO, T. and VU, V. (2012). Random covariance matrices: universality of local statistics of eigenvalues, *The Annals of Probability*, **40**, 1285–1315.

TIKHOMIROV, K.E. (2015). The limit of the smallest singular value of random matrices with i.i.d. entries, *Advances in Mathematics*, **284**, 1–20.

TRACY, C.A. and WIDOM, H. (1994). Level-spacing distributions and the Airy kernel, *Communications in Mathematical Physics*, **159**, 151–174.

TRACY, C.A. and WIDOM, H. (1994). Level spacing distributions and the Airy kernel, *Communications in Mathematical Physics*, **159**, 33–72.

TRACY, C.A. and WIDOM, H. (1996). On orthogonal and symplectic matrix ensembles, *Communications in Mathematical Physics*, **177**, 727–754.

TRACY, C.A. and WIDOM, H. (2000). Universality of the distribution functions of random matrix theory, In: *Integrable Systems: FRom Classical to Quantum 1999*, **26**, 251–264. American Mathematical Society, Providence, R.I.

TULINO, A.M. and VERDÚ, S. (2004). Random matrix theory and wireless communications, *Foundations and Trends in Communications and Information Theory*, **1**, 1–182.

ULAM, S.M. (1961). Monte Carlo calculations in problems of mathematical physics, In: *Modern Mathematics for the Engineer: Second Series* (E.F. Beckenbach, ed.), New York: McGraw-Hill, pp. 261–281.

VERSHIK, A.M. and KEROV, S.V. (1977). The asymptotic of the Plancherel measure of a symmetric group and the limit form of Young tableaux, *Doklady Akademi Nauk SSSR*, **223**, 1024–1027.

VOICULESCU, D.-V. (1991). Limit laws for random matrices and free products, *Inventions in Mathematics*, **104**, 201–220.

VOICULESCU, D.-V. (1995). Free probability theory: random matrices and von Neumann algebras, *Proceedings of the International Congress of Mathematicians*, 227–242, Basel, Switzerland: Birkhäuser-Verlag.

VON NEUMANN, J. (1937). Some matrix inequalities and metrization of matric space, *Tomsk University Review*, **1**, 286–300.

WACHTER, K.W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements, *The Annals of Probability*, **6**, 1–18.

WANG, W. and FAN, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance, *The Annals of Statistics*, **45**, 1342–1374.

WANG, Q. and YAO, J. (2017). Extreme eigenvalues of large-dimensional spiked Fisher matrices with application, *The Annals of Statistics*, **45**, 415–460.

WIGNER, E.P. (1955). Characteristic vectors of bordered matrices with infinite dimensions, *Annals of Mathematics*, **62**, 548–564.

WIGNER, E.P. (1958). On the distribution of the roots of certain symmetric matrices, *Annals of Mathematics*, **67**, 325–328.

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

YIN, Y.Q., BAI, Z.D., and KRISHNAIAH, P.R. (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix, *Probability Theory and Related Fields*, **78**, 509–521.