



Review

Random matrix theory in statistics: A review

Debashis Paul*, Alexander Aue

Department of Statistics, University of California, Davis, United States



ARTICLE INFO

Article history:

Received 1 March 2013

Received in revised form

6 September 2013

Accepted 10 September 2013

Available online 27 September 2013

Keywords:

MANOVA

Marčenko–Pastur law

Principal component analysis

Stieltjes transform

Wishart matrix

Tracy–Widom law

ABSTRACT

We give an overview of random matrix theory (RMT) with the objective of highlighting the results and concepts that have a growing impact in the formulation and inference of statistical models and methodologies. This paper focuses on a number of application areas especially within the field of high-dimensional statistics and describes how the development of the theory and practice in high-dimensional statistical inference has been influenced by the corresponding developments in the field of RMT.

© 2014 Elsevier B.V. All rights reserved.

Contents

| | |
|---|----|
| 1. Introduction | 2 |
| 2. Background and motivation | 3 |
| 3. Large random matrices | 5 |
| 3.1. Behavior of the bulk spectrum | 5 |
| 3.1.1. Method of moments | 7 |
| 3.1.2. Stieltjes transform | 7 |
| 3.1.3. Convergence rate of the ESD | 9 |
| 3.1.4. Extension to non-i.i.d. settings | 9 |
| 3.1.5. Linear spectral statistics | 11 |
| 3.1.6. Spectrum separation | 13 |
| 3.2. Behavior at the edge of the spectrum | 13 |
| 3.2.1. Connection to orthogonal polynomials | 13 |
| 3.2.2. Tracy–Widom laws for the extreme eigenvalues | 14 |
| 3.3. Universality | 16 |
| 3.4. Behavior of the eigenvectors | 17 |
| 3.5. Matrix ensembles with heavy-tails | 17 |
| 4. Applications | 18 |
| 4.1. Applications to statistics | 18 |
| 4.1.1. Signal detection using the Tracy–Widom law | 18 |
| 4.1.2. PCA under the spiked covariance model | 18 |

* Corresponding author. Tel.: +1 530 564 0513.

E-mail addresses: debashis@wald.ucdavis.edu, debpaual@ucdavis.edu (D. Paul), alexaue@wald.ucdavis.edu (A. Aue).

| | | |
|--------|---|----|
| 4.1.3. | Application to MANOVA | 20 |
| 4.1.4. | Tests of hypothesis involving the bulk spectrum | 20 |
| 4.1.5. | Spectrum estimation | 20 |
| 4.1.6. | Eigen-subspace estimation | 21 |
| 4.2. | Applications to wireless communication | 21 |
| 4.3. | Applications to finance | 21 |
| 4.4. | Other applications | 22 |
| 4.5. | Computational tools | 22 |
| 5. | Sparse PCA, CCA, LDA and covariance estimation | 22 |
| 5.1. | Regularization of covariance and concentration matrices | 22 |
| 5.2. | Covariance estimation using sparse factor models | 23 |
| 5.3. | Sparse canonical correlation analysis | 23 |
| 5.4. | Sparse discriminant analysis | 24 |
| 6. | Future directions | 24 |
| | Acknowledgements | 24 |
| | References | 24 |

1. Introduction

Statistics has entered into an age where an increasingly larger volume of more complex data is being generated, often through automated measuring devices, in a wide array of disciplines such as genomics, atmospheric sciences, communications, biomedical imaging, economics and many others. The representation of such data in any nominal coordinate system often leads to so-called high-dimensional data that are frequently associated with phenomena transcending the boundary of classical multivariate statistical analysis. The continued growth of these new data sources has given rise to the incorporation of different mathematical tools into the realms of statistical analysis, which include convex analysis, Riemannian geometry and combinatorics. Random matrix theory has emerged as a particularly useful framework for posing many theoretical questions associated with the analysis of high-dimensional multivariate data.

In this paper, we mainly focus on several application areas of random matrix theory (RMT) in statistics. These include problems in dimension reduction, hypothesis testing, clustering, regression analysis and covariance estimation. We also briefly describe the important role played by RMT in enabling certain theoretical analyses in wireless communications and econometrics. Different themes emerging from these problems have in turn led to further investigation of some classical RMT phenomena. Among these, the notion of *universality* has profound implications in the context of high-dimensional data analysis in terms of the applicability of many statistical techniques beyond the classical framework built upon the multivariate Gaussian distribution. With these perspectives, the treatment of the topics will focus on those aspects of RMT relevant to the statistical questions. Thus, the topics in RMT that receive most attention in this paper are those related to the behavior of the bulk spectrum, i.e., the empirical spectral distribution, and the behavior of the edge of the spectrum, i.e., the extreme eigenvalues, of random matrices. Also, the sample covariance matrix being a dominant object of study in most of the multivariate analyses, much of the paper is devoted to the study of its spectral behavior. For more detailed accounts of these topics, the reader may refer to [Anderson et al. \(2009\)](#), [Bai and Silverstein \(2009\)](#) and [Pastur and Shcherbina \(2011\)](#). More complete and self-containing treatments of a number of “core” topics of RMT not covered in this paper, including the Riemann–Hilbert approach to the asymptotics of orthogonal polynomials and random matrices, the distribution of spacings and correlation functions of eigenvalues and their connections with determinantal point processes, and the role of the eigenvalue statistics in physics, can be found in the monographs ([Akemann et al., 2011](#); [Anderson et al., 2009](#); [Deift, 2000](#); [Deift and Gioev, 2009](#); [Forrester, 2010](#); [Guionnet, 2009](#); [Mehta, 2004](#); [Tao, 2012](#)), and the survey articles ([Diaconis, 2003](#); [Soshnikov, 2000](#)). The connection between RMT and free probability theory, another significant topic not discussed here, is explored in detail in [Anderson et al. \(2009\)](#), [Hiai and Petz \(2000\)](#), [Mingo and Speicher \(2006\)](#), [Nica and Speicher \(2006\)](#) and [Edelman and Rao \(2005\)](#). Finally, wireless communications and finance are two areas beyond physics and statistics where tools and concepts from RMT have been successfully applied and thus we give a brief overview of these topics in [Section 4](#). Applications of RMT in wireless communications are the focus of [Coulliet and Debbah \(2011\)](#) and [Tulino and Verdú \(2004\)](#), while for a detailed look at applications in finance one may refer to [Bouchaud et al. \(2003\)](#) and [Bouchaud and Potters \(2009\)](#). A survey of some of the statistical topics covered in this review can be found in [Johnstone \(2007\)](#).

We now give a brief outline of this paper. There are two different ways in which RMT has impacted modern statistical procedures. On one hand, most of the mathematical treatment of RMT have focused on matrices with high degree of independence in the entries, which one may refer to as “unstructured” random matrices. The results from the corresponding theory have been used primarily in the context of hypothesis testing where the null hypothesis corresponds to the absence of any directionality, or signal component, in the data. On the other hand, in high-dimensional statistics, we are primarily interested in problems where there are lower dimensional structures buried under random noise. An effective treatment of the latter problem often requires going beyond the realms of the classical RMT framework and into the domain of statistical regularization schemes. Keeping these perspectives in mind, we devote [Sections 2](#) and [3](#) to the motivations and theoretical developments in RMT, while focusing on the statistical applications of RMT in [Section 4](#). Finally, in [Section 5](#), we focus on modern statistical regularization schemes based on various forms of sparse structures for dealing with high-dimensional statistical problems.

RMT does not play any direct role in this context, except possibly in the theoretical analysis of some estimation schemes, but provides guidance for potential implications of violating the structural assumptions underlying the inferential procedures.

2. Background and motivation

Random matrices play a central role in statistics in the context of analysis of multivariate data. There are numerous books on classical multivariate analysis, most notably [Anderson \(1984\)](#), [Mardia et al. \(1980\)](#), and [Muirhead \(1982\)](#), that describe the major problems addressed through the use of analysis of random matrices. Most of these problems are naturally formulated in terms of the eigen-decomposition of certain Hermitian or symmetric matrices. These problems can be broadly categorized into two groups – the first group involves the eigen-analysis of a single Hermitian matrix, often referred to as a *single Wishart problem*; and the second group involves a generalized eigenvalue problem involving two independent Hermitian matrices of the same dimension, often referred to as a *double Wishart problem*. The first group includes principal component analysis (PCA), factor analysis and tests for population covariance matrices in one-sample problems. The second group includes multivariate analysis of variance (MANOVA), canonical correlation analysis (CCA), tests for equality of covariance matrices and tests for linear hypotheses in multivariate linear regression problems. In addition, random matrices play a natural role in defining and characterizing estimates in multivariate linear regression problems and in classification (within sample covariance matrices) and clustering problems (pairwise distance or similarity matrices). Thus, analyzing the behavior of eigenvalues and eigenvectors of random symmetric or Hermitian matrices has a precedence in statistics that goes back to the work of [Pearson \(1901\)](#) who introduced the notion of dimension reduction of multivariate data through PCA. In this section, we briefly discuss each of the classical problems mentioned in the previous paragraph.

Principal component analysis (PCA) is a highly versatile nonparametric tool for data reduction and model building. For an extensive discussion of the various applications and variants of PCA, one may refer to [Jolliffe \(2002\)](#). The formulation of PCA in classical multivariate analysis at the population level is as follows. Suppose that we measure p variables (assume real-valued, for simplicity), expressed as a random vector $X = (X^{(1)}, \dots, X^{(p)})^T$. Suppose also that the random vector X has finite variance $\Sigma := \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T]$. The primary goal of PCA is to obtain a lower dimensional representation of the data in the form of linear transformations of the original variable, subject to the condition that the residual variance is as small as possible. This can be achieved by considering a sequence of linear transformations given by $v_k^T X$, $k = 1, 2, \dots, p$, satisfying the requirement that $\text{Var}(v_k^T X)$ is maximized subject to the conditions that v_k are unit norm vectors in \mathbb{R}^p (or \mathbb{C}^p if the data is complex-valued), and v_k is orthonormal to $\{v_j : j = 1, \dots, k-1\}$, i.e., $v_k^T v_j = 0$ for $j = 1, \dots, k-1$. This optimization problem can be solved in terms of the spectral decomposition of the nonnegative definite Hermitian matrix Σ :

$$\Sigma v_k = \ell_k v_k, \quad k = 1, \dots, p, \quad (1)$$

where v_k are the orthonormal vectors. Here ℓ_k (always real-valued) is an eigenvalue associated with v_k . Note that in this formulation the eigenvalues ℓ_k are ordered, i.e., $\ell_1 \geq \dots \geq \ell_p \geq 0$. If ℓ_k is of multiplicity one, then v_k is unique up to a sign change. In practice, we do not know Σ and we typically observe a sample X_1, \dots, X_n for the variable X . In that case, the empirical version of PCA involves replacing Σ by its natural estimate $S_n = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$, and performing the spectral decomposition for S_n analogous to (1). The corresponding eigenvectors \hat{v}_k are often referred to as the sample principal components. The corresponding ordered eigenvalues $\hat{\ell}_k$ are typically used to detect the dimension of the reduction subspace. One of the commonly used techniques is to plot the eigenvalues against their indices (so-called “scree plot”) and then look for an “elbow” in the plot. There are formal tests based on likelihood ratios (see e.g., [Anderson, 1984](#); [Muirhead, 1982](#)) that assume that, after a certain index, the eigenvalues are all equal and that the observations are Gaussian. Notice that the name *single Wishart* arises from the fact that if X_1, \dots, X_n are i.i.d. $N_p(\mu, \Sigma)$, then $(n-1)S_n$ has Wishart $(p, n-1, \Sigma)$ distribution. Under Gaussianity, one of the commonly used tests for *sphericity*, i.e., the hypothesis $H_0 : \Sigma = \mathbf{I}_p$, is Roy's largest root test ([Roy, 1953](#)) which rejects H_0 if $\hat{\ell}_1$, the largest eigenvalue of S exceeds a threshold determined by the level of significance. If $H_0 : \Sigma = \sigma^2 \mathbf{I}_p$ for some unknown σ^2 , the corresponding generalized likelihood ratio test, under an alternative that assumes Σ to be a rank-one perturbation of $\sigma^2 \mathbf{I}_p$, rejects for large values of $\hat{\ell}_1 / (\sum_{j=2}^p \hat{\ell}_j)$ ([Graybill and Johnson, 1972](#); [Nadler et al., 2008](#)). We return to these problems in [Section 4](#).

A factor analysis problem can be seen as a generalization of PCA in that it assumes a certain signal-plus-noise decomposition of the observation vector X :

$$X - \mu = \mathbf{L}f + \varepsilon, \quad (2)$$

where f is an $m \times 1$ dimensional random vector, \mathbf{L} is a $p \times m$ dimensional nonrandom matrix, f and ε are uncorrelated, and ε has mean 0 and variance Ψ , a $p \times p$ diagonal matrix. For identifiability, it is typically assumed that $\mathbb{E}[f] = 0$ and $\mathbb{E}[ff^T] = \mathbf{I}_m$. Under this setting, the covariance matrix of X is of the form $\Sigma = \mathbf{L}\mathbf{L}^T + \Psi$. Thus, if Ψ is a multiple of the identity, the problem of estimating \mathbf{L} from data can be formulated in terms of a PCA of the sample covariance matrix. One important distinction between PCA and factor analysis is that, in the latter case, the practitioner implicitly assumes a causal model for the data. In general, factor analysis problems are often solved through a maximum likelihood approach (see [Tipping and Bishop, 1999](#)). A more enhanced version of the factor analysis model, the so-called dynamic factor model, is used extensively in econometrics, where the factors f are taken to be time-dependent ([Forni et al., 2000, 2004, 2005](#)).

A detailed discussion of various versions of the double Wishart eigen-problem, including a summary of the associated distribution theory when the observations are Gaussian, can be found in [Johnstone \(2008, 2009\)](#). Within this framework, we first consider the canonical correlation analysis (CCA) problem. Again, first we deal with the formulation at the population level. Suppose that real-valued random vectors X and Y are jointly observed, where X is of dimension p and Y is of dimension q . Then a generalization of the notion of correlation between X and Y is expressed in terms of the sequence of canonical correlation coefficients defined as

$$\rho_k = \max_{(u,v) \in S_k} |\text{Cor}(u^T X, v^T Y)|, \quad k = 1, 2, \dots, \min\{p, q\}, \quad (3)$$

where

$$S_k := \{(u, v) \in \mathbb{R}^{p+q} : u^T \Sigma_{XX} u = v^T \Sigma_{YY} v = 1; u^T \Sigma_{XX} u_j = v^T \Sigma_{YY} v_j = 0, j = 1, \dots, k-1\},$$

with $\Sigma_{XX} = \text{Var}(X)$, $\Sigma_{YY} = \text{Var}(Y)$, and (u_k, v_k) denoting the pair of vectors for which the maximum in (3) is attained. If $\Sigma_{XY} = \text{Cov}(X, Y)$, then the optimization problem (3) can be formulated as the following generalized eigenvalue problem: the successive canonical correlations $\rho_1 \geq \dots \geq \rho_{\min\{p,q\}} \geq 0$ satisfy the generalized eigen-equations

$$\det(\Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX} - \rho^2 \Sigma_{XX}) = 0. \quad (4)$$

When we have n samples $\{(X_i, Y_i) : i = 1, \dots, n\}$ we can replace Σ_{XX} , Σ_{XY} and Σ_{YY} by their sample counterparts and, assuming $n > \max\{p, q\}$, the corresponding sample canonical correlations $r_1 \geq \dots \geq r_{\min\{p,q\}} \geq 0$ satisfy the sample version of (4). It is shown in [Mardia et al. \(1980\)](#) that in the latter case, we can reformulate the corresponding generalized eigen-analysis problem as solving

$$\det(\mathbf{U} - r^2(\mathbf{U} + \mathbf{V})) = 0, \quad (5)$$

where \mathbf{U} and \mathbf{V} are independent Wishart matrices if (X_i, Y_i) are i.i.d. Gaussian and $\Sigma_{XY} = \mathbf{0}$, i.e., X and Y are independently distributed.

Next, we consider the problem of testing hypotheses in a multivariate linear regression model given by

$$\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{E}, \quad (6)$$

where \mathbf{Y} is a $p \times n$ matrix consisting of n measurements on p response variables, \mathbf{X} is the $q \times n$ nonrandom (known) matrix of measurements on the q predictor variables, and \mathbf{B} is the $p \times q$ matrix of regression coefficients, while \mathbf{E} is the matrix of the residuals which is typically assumed to have i.i.d. columns with mean 0 and variance Σ , say. Then, as described in [Mardia et al. \(1980\)](#), the union-intersection test for the linear hypothesis of the form $H_0 : \mathbf{C}\mathbf{B} = \mathbf{0}$ where \mathbf{C} is $c \times p$ and \mathbf{D} is $q \times d$, and both are specified, can be expressed in terms of the largest eigenvalue of $\mathbf{U}(\mathbf{U} + \mathbf{V})^{-1}$ where \mathbf{U} and \mathbf{V} are appropriately specified independent Wishart matrices (under Gaussianity of the entries of \mathbf{E}).

The two sample test for equality of variances assumes that we have i.i.d. samples from two normal populations $N_p(\mu_1, \Sigma_1)$ and $N_p(\mu_2, \Sigma_2)$ of sizes n_1 and n_2 , say. Then several tests for the hypothesis $H_0 : \Sigma_1 = \Sigma_2$ can be formulated in terms of functionals of the eigenvalues of $\mathbf{U}(\mathbf{U} + \mathbf{V})^{-1}$ where $\mathbf{U} = (n_1 - 1)\mathbf{S}_1$ and $\mathbf{V} = (n_2 - 1)\mathbf{S}_2$ are the sample covariances for the two samples, which would follow independent Wishart distributions in p dimension with d.f. $n_1 - 1$ and $n_2 - 1$ and dispersion matrix $\Sigma_1 = \Sigma_2$ under H_0 .

The MANOVA (multivariate analysis of variance) problem supposes that there are g normal populations for p -dimensional variables with means μ_1, \dots, μ_g and common covariance Σ , and we have independent samples of sizes n_i , $i = 1, \dots, g$. The problem is to test $H_0 : \mu_1 = \dots = \mu_g$. A natural test is to reject H_0 for large values of the largest eigenvalue of $\mathbf{V}^{-1}\mathbf{U}$ where \mathbf{U} is the between-sample covariance matrix and \mathbf{V} is the within-sample covariance matrix, which are independent Wishart (up to normalization) under H_0 .

It should be noted here that until the analysis of large random matrices came into play, the theoretical analysis of each of the methods mentioned above was restricted to one of the following settings: (i) the observations are independent and have multivariate normal distributions, in which case closed-form expressions of the distribution of certain statistics could be derived; or, (ii) the observations are independent, though not necessarily Gaussian, and asymptotic analyses are carried out in the framework where the dimensionality of the observations remains fixed while the amount of data, or sample size, increases to infinity. Both these frameworks have their limitations, even though they have played very significant roles in formulating statistical techniques over a long period of time. The limitations of these procedures become more tangible when one notices that many of the distributional results, either exact or approximate, do not really apply in the context of data that are of only moderately large dimension, in comparison with the sample size. And so, in order to accommodate the analysis of a large variety of data being collected nowadays, arising in fields such as genomics, economics, atmospheric science, chemometrics and astronomy, to name a few, it is imperative to either modify or reformulate some of the statistical techniques. This is where RMT has been playing a significant role, especially over the last decade. The relevant asymptotics take both data dimensionality and sample size to infinity. A third and less explored alternative, particularly relevant from the point of view of signal processing, keeps both dimensionality and sample size fixed, but takes the noise level to zero. The next two sections deal with the development of RMT that provide at least partial answers to some of the issues. Then, in [Section 4](#), we return to the application of those results in the context of high-dimensional statistical inference.

3. Large random matrices

In this section, we deal with two kinds of random matrices that have been central to most of the developments in RMT – (i) the sample covariance matrix, often referred to as the Wishart matrix, and (ii) the Wigner matrix. Both being symmetric or Hermitian matrices, depending on the entries of the matrix being real- or complex-valued, there are similarities in the type of results derived about their spectra in the RMT literature, although there are interesting differences in their asymptotic behavior. We first give a basic introduction to these matrix models and then proceed towards dealing with some of the key questions associated with the eigenvalues and eigenvectors of such matrices.

The classical RMT model for Wishart matrices requires specifying two sequences of integers n , the sample size, and $p = p(n)$, the dimension, so that $p \rightarrow \infty$ as $n \rightarrow \infty$, and

$$\lim_{n \rightarrow \infty} \frac{p}{n} \rightarrow \gamma \in (0, \infty). \quad (7)$$

Suppose that $\mathbf{X} = (X_{ij} : 1 \leq i \leq p, 1 \leq j \leq n)$ is a $p \times n$ matrix with real- or complex-valued entries, such that the columns $X_j = (X_{ij})_{i=1}^p$ of \mathbf{X} are independent. Then the $p \times p$ matrix $\mathbf{S} = n^{-1} \mathbf{X} \mathbf{X}^T$ is defined as the (uncentered) sample covariance matrix. The unnormalized version $n\mathbf{S} = \mathbf{X} \mathbf{X}^T$ is often referred to as the Wishart matrix. The name derives from the fact that if the columns of \mathbf{X} are independently distributed as $N_p(0, \Sigma)$ for some positive semidefinite matrix Σ , then the distribution of $n\mathbf{S}$ is the *Wishart distribution* $W_p(n, \Sigma)$ where n stands for the degrees of freedom, p denotes the dimension, and Σ is the scale parameter. The distribution was first studied by [Wishart \(1928\)](#) and has remained a central object of study in multivariate statistical analysis. Detailed study on the properties of this distribution can be found in [Anderson \(1984\)](#) and [Muirhead \(1982\)](#). In statistics, the sample covariance matrix is more commonly defined as $\hat{\mathbf{S}} = (n-1)^{-1}(\mathbf{X} \mathbf{X}^T - n \bar{\mathbf{X}} \bar{\mathbf{X}}^T)$ where $\bar{\mathbf{X}} = (1/n) \sum_{j=1}^n X_j$ is the sample mean. However, for mostly all of the interesting results in RMT under the framework (7), the distinction between \mathbf{S} and $\hat{\mathbf{S}}$ is not material. Hence, except when the latter is explicitly needed, we concern ourselves with the behavior of the spectrum and eigenvectors of \mathbf{S} .

The Wigner matrix has also been an object of elaborate study in the RMT literature starting with its introduction by [Wigner \(1955, 1958\)](#) to model the spectra of heavy atoms. In quantum mechanics, the energy states of a system is described by the eigenvalues of a Hamiltonian H (a Hermitian operator). Wigner's hypothesis was that the spacings between the lines in the spectrum of a heavy atom should resemble the spacings between the eigenvalues of a symmetric or Hermitian random matrix. In effect his proposal was to replace the Hamiltonian H with a large $n \times n$ Hermitian matrix with random entries. This led to the investigation of the spectrum of large dimensional random Hermitian matrices by physicists such as [Dyson \(1962a,b,c\)](#). Subsequently random matrices have found applications in different branches of physics including nuclear physics, solid state physics ([Bahcall, 1996](#)) and quantum chaos ([Bohigas et al., 1984](#)). An overview of some of the physical applications of matrix models can be found in [Forrester \(2010\)](#) and [Mehta \(2004\)](#). A Wigner matrix $\mathbf{X} = (X_{ij} : 1 \leq i, j \leq n)$ is an $n \times n$ matrix with real or complex entries such that (i) $X_{ij} = \bar{X}_{ji}$ (where the bar denotes the complex conjugate) for the complex case and $X_{ij} = X_{ji}$ for the real case, for $1 \leq i < j \leq n$; and (ii) the entries $\{X_{ij} : 1 \leq i < j \leq n\}$ are independent random variables with mean 0 and variance 1. Since \mathbf{X} is symmetric (in the real case) or Hermitian (in the complex case), the diagonal entries $\{X_{ii} : 1 \leq i \leq n\}$ are necessarily real-valued random variables. It may be noted that, [Wigner \(1955\)](#) considered a random sign matrix with zero diagonal entries. Quite often the diagonal entries of a Wigner random matrix have variances different from the variance of the off-diagonal entries.

3.1. Behavior of the bulk spectrum

Suppose that \mathbf{X} is an $N \times N$ matrix with eigenvalues $\lambda_1, \dots, \lambda_N \in \mathbb{C}$. The empirical distribution of the eigenvalues of \mathbf{X} , usually referred to as the *empirical spectral distribution* (ESD) of \mathbf{X} , is the function $N^{-1} \sum_{i=1}^N \delta_{\lambda_i}$ where δ_y denotes the Dirac mass at y . If \mathbf{X} is Hermitian, so that the eigenvalues of \mathbf{X} are real, we can define the empirical distribution function of \mathbf{X} as $F^{\mathbf{X}}(x) = N^{-1} \sum_{j=1}^N \mathbf{1}_{\{\lambda_j \leq x\}}$ for $x \in \mathbb{R}$. In RMT, the ESD of a random matrix plays a central role in studying the properties of the spectrum. One motivation is that many statistics associated with a random matrix \mathbf{X} can be expressed as a linear functional of its ESD, or a *linear spectral statistic*, i.e., a function of the form $\int g(x) dF^{\mathbf{X}}(x)$ for some suitably regular function g . For example, the logarithm of the determinant of a sample covariance matrix \mathbf{S} (of dimension $p \times p$, say) is an important object of study in wireless communication (cf. [Tulino and Verdú, 2004](#)), and it can be expressed as $\sum_{j=1}^p \log \lambda_j = p \int \log x dF^{\mathbf{S}}(x)$, where the λ_j 's are the eigenvalues of \mathbf{S} . Many classical test procedures in multivariate statistics use traces of polynomials of the sample covariance matrix, which also fall in this category. Thus, detailed knowledge of the asymptotic behavior of the ESD can help in understanding the behavior of linear spectral statistics.

One of the first questions one can ask about the ESD is whether, after appropriate normalization of the matrix, this random distribution converges to a probability distribution in an appropriate sense as the dimension of the matrix grows. The celebrated *semicircle law* provides a first answer to this in the context of Wigner matrices, i.e., square Hermitian matrices with independent entries on and above the diagonal with zero mean and unit variance. [Wigner \(1958\)](#) showed that the expected ESD of an $n \times n$ Wigner matrix with Gaussian entries, multiplied by $1/\sqrt{n}$, converges in distribution to the semicircle law that has p.d.f.

$$f(x) = \frac{1}{2\pi} \sqrt{4 - x^2} \mathbf{1}_{[-2,2]}(x). \quad (8)$$

Arnold (1967, 1971) showed that the ESD of the normalized Wigner matrix itself almost surely converges in distribution to the semicircle law. There have been numerous further developments that determined in particular the necessary and sufficient conditions for the convergence of the ESD. We have the following version that states the result under the weakest moment conditions.

Theorem 3.1. Suppose that \mathbf{X} is an $n \times n$ Wigner matrix whose diagonal entries are i.i.d. real random variables with mean 0 and variance 1, and those above the diagonal are i.i.d. complex random variables with mean 0 and variance 1. Then, as $n \rightarrow \infty$, the ESD of \mathbf{X}/\sqrt{n} almost surely converges in distribution to the semicircle law with p.d.f. given by (8).

An analogous result was derived for the sample covariance matrix by Marčenko and Pastur (1967) assuming that the fourth moments of the entries of the data matrix are finite. The important facet of this result is the dependence of the limiting distribution on the limiting ratio $\gamma = \lim_{n \rightarrow \infty} p/n$. Since then, numerous researchers, notably Wachter (1978), Yin (1986), and Yin et al. (1983), have contributed to weakening the conditions on the matrix entries as well as extending the class of matrices for which the ESD has a nonrandom limit. The following version is under the minimal moment conditions.

Theorem 3.2. Suppose that \mathbf{X} is a $p \times n$ matrix with i.i.d. real- or complex-valued entries with mean 0 and variance 1. Suppose also that (7) holds. Then, as $n \rightarrow \infty$, the ESD of $\mathbf{S} = n^{-1} \mathbf{X} \mathbf{X}^*$ converges almost surely in distribution to a nonrandom distribution, known as the Marčenko–Pastur law and denoted by F_γ . If $\gamma \in (0, 1]$, then F_γ has the p.d.f.

$$f_\gamma(x) = \frac{\sqrt{(b_+(\gamma) - x)(x - b_-(\gamma))}}{2\pi\gamma x} \mathbf{1}_{[b_-(\gamma), b_+(\gamma)]}(x), \quad (9)$$

where $b_\pm(\gamma) = (1 \pm \sqrt{\gamma})^2$. If $\gamma \in (1, \infty)$, then F_γ is a mixture of a point mass 0 and the p.d.f. $f_{1/\gamma}$ with weights $1 - 1/\gamma$ and $1/\gamma$, respectively.

An obvious implication of Theorem 3.2 is the spreading of the eigenvalues of the sample covariance matrix around their population counterpart, and the increase in this spread as the limiting dimension-to-sample-size ratio γ increases from 0 to 1. When $p/n \rightarrow 0$, both the largest and the smallest eigenvalues of \mathbf{S} converge to 1 and thus the Marčenko–Pastur law does not hold for the ESD of \mathbf{S} (Fig. 1). In Bai and Yin (1988), it was proved that after suitable normalization, the ESD of \mathbf{S} tends to the semicircular law in this setting. Bao (2012) and Pan and Gao (2009) extended this result when dealing with sample covariance matrices that can be expressed as $n^{-1} \mathbf{A}^{1/2} \mathbf{X} \mathbf{X}^* \mathbf{A}^{1/2}$, where \mathbf{A} denotes a sequence of $p \times p$ nonnegative definite matrices whose eigenvalues have certain regularity as p increases, and the $n \times p$ data matrix \mathbf{X} has i.i.d. standardized entries, while $p, n \rightarrow \infty$ such that $p/n \rightarrow 0$. Incidentally, these results also provide a connection between the Wishart and Wigner matrices from the point of view of statistical analysis, and underline the importance of studying Wigner matrices in statistics when dealing with moderately large dimensional data.

Theorems 3.1 and 3.2 have been cornerstones of an increasingly growing body of literature in RMT. One aspect of these investigations has been to relax the assumptions on the entries of the matrices, for example removing the condition of independence of the entries in a Wigner matrix, while allowing for more complex structures in the matrices, for example dealing with the ESD of the sample covariance matrix corresponding to the data matrix $\mathbf{Y} = \mathbf{A}^{1/2} \mathbf{X} \mathbf{B}^{1/2}$ where \mathbf{A} and \mathbf{B} are nonnegative definite matrices and \mathbf{X} has i.i.d. entries. The other aspect of these studies has been the exploration of finer issues associated with the ESD, including rates of convergence, fluctuations of linear spectral statistics, spectrum separation, etc. Both these aspects have had significant impacts on the recent developments of inferential techniques for high-dimensional statistics. Before turning to these topics, we briefly summarize the two main techniques that have been widely

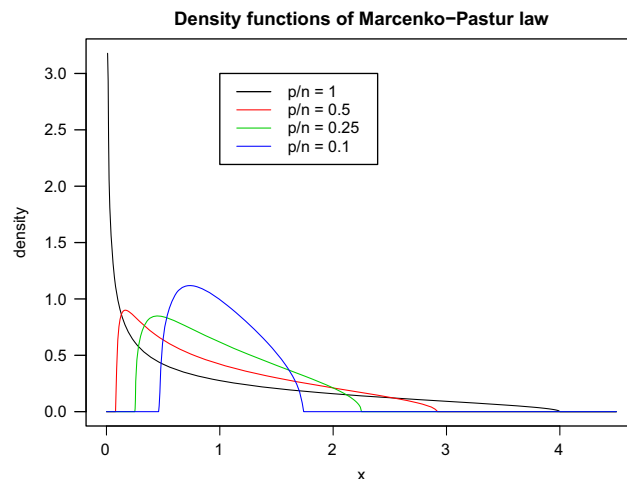


Fig. 1. Marčenko–Pastur density functions for $\gamma = p/n = 0.1, 0.25, 0.5, 1$.

used in RMT for proving limit theorems about the ESD of a random matrix – the method of moments and the method based on Stieltjes transforms.

3.1.1. Method of moments

The method of moments has been used extensively for determining the behavior of the ESD as well as extreme eigenvalues of a random matrix. For detailed accounts on the application of this approach to RMT, which depends heavily on the combinatorics of graph partitioning, one may refer to [Anderson et al. \(2009\)](#), [Bai \(1999\)](#), [Bai and Silverstein \(2009\)](#), [Guionnet \(2009\)](#) and [Sodin \(2007\)](#). The following result (quoted from [Bai and Silverstein, 2009](#)), known as *Carleman condition*, is a key ingredient in proving the limiting behavior of the ESD of a Wigner matrix or a sample covariance matrix ([Theorems 3.1 and 3.2](#)).

Lemma 3.1. Let $\beta_k = \beta_k(F) := \int x^k dF(x)$ for $k = 1, 2, \dots$, where F is a probability distribution on \mathbb{R} . If the Carleman condition

$$\sum_{k=1}^{\infty} \beta_{2k}^{-1/2k} = \infty \quad (10)$$

is satisfied, then F is uniquely determined by the moment sequence $\{\beta_k : k = 1, 2, \dots\}$.

This result is used in conjunction with the following lemma.

Lemma 3.2. A sequence $\{F_n\}$ of probability distributions on \mathbb{R} converges in distribution to a probability distribution F if the following conditions are satisfied:

1. Each F_n has moments of all order.
2. For each fixed integer k , $\beta_{n,k} = \int x^k dF_n(x)$ converges to $\beta_k = \int x^k dF(x)$ as $n \rightarrow \infty$.
3. The coefficients $\{\beta_k\}$ of the limiting distribution F satisfies (10).

For a symmetric or Hermitian random matrix \mathbf{X} , the k -th moment of the ESD of \mathbf{X} equals $\text{tr}(\mathbf{X}^k)$. Thus, in view of [Lemma 3.2](#), in order to prove that the ESD of \mathbf{X} almost surely converges in distribution to some nonrandom distribution F , one needs to check that its moment sequence $\{\beta_k = \beta_k(F)\}$ satisfies (10), and that for every $k \geq 1$, $\text{tr}(\mathbf{X}^k) \rightarrow \beta_k$ almost surely. In both [Theorems 3.1 and 3.2](#), the entries of the relevant matrices are only assumed to have finite second moments. Thus, the results are first proved for settings where the entries of the random matrix have all the moments. Then the entries of the original data matrix are appropriately truncated and centered and it is shown that the difference between the ESD of the original matrix and that of its truncated version vanishes as the dimensionality becomes large.

3.1.2. Stieltjes transform

The Stieltjes transform plays nearly as useful a role in RMT as the Fourier transform in classical probability theory. The Stieltjes transform of a measure μ on the real line is defined as the function

$$S_\mu(z) = \int \frac{1}{x-z} \mu(dx), \quad z \in \mathbb{C}^+, \quad (11)$$

where $\mathbb{C}^+ := \{x + iy : x \in \mathbb{R}, y > 0\}$. Note that S_μ is analytic on \mathbb{C}^+ and maps into \mathbb{C}^+ . The following inversion formula allows one to reconstruct the distribution function from its Stieltjes transform.

Lemma 3.3. Let P be a probability measure on the real line. If $a < b$ are points of continuity of the associated distribution function, then

$$P((a, b)) = \frac{1}{\pi} \lim_{v \rightarrow 0^+} \int_a^b \Im(S_P(u + iv)) du. \quad (12)$$

The following lemma (cf. [Geronimo and Hill, 2003](#)) gives a necessary and sufficient condition for the limit of Stieltjes transforms of a sequence of probability measures to be the Stieltjes transform of a probability measure.

Lemma 3.4. Suppose that $\{P_n\}$ is a sequence of Borel probability measures on the real line with Stieltjes transforms $\{s_n\}$. If $\lim_{n \rightarrow \infty} s_n(z) = s(z)$ for all $z \in \mathbb{C}^+$, then there exists a Borel probability measure P with Stieltjes transform $S_P = s$ if and only if

$$\lim_{v \rightarrow \infty} i v s(iv) = -1, \quad (13)$$

in which case P_n converges to P in distribution.

To see why the Stieltjes transform plays such an important role in the study of the asymptotic behavior of ESDs, suppose that for each $N \geq 1$, \mathbf{W}_N is an $N \times N$ Hermitian random matrix, so that its eigenvalues are all real, with ESD $F^{\mathbf{W}_N}$. Then, the Stieltjes transform of $F^{\mathbf{W}_N}$, say s_N , is given by $s_N(z) = N^{-1} \text{tr}((\mathbf{W}_N - z\mathbf{I}_N)^{-1})$. Notice that $(\mathbf{W}_N - z\mathbf{I}_N)^{-1}$ is the resolvent of the matrix \mathbf{W}_N and its points of singularity are at the eigenvalues of \mathbf{W}_N . In view of [Lemma 3.4](#), in order to prove that the

sequence of ESDs $F^{\mathbf{W}_n}$ converges to a distribution F , say (in probability or almost surely), one needs to check that $\{s_N\}$ satisfies the conditions of the lemma (in probability or almost surely).

In many classical problems in RMT, especially those where the random matrix of interest is a Wigner or Wishart-type matrix, it is relatively easy to derive an approximate iterative equation for the Stieltjes transform of its ESD by using appropriate inversion formulas for block matrices. As an illustration, suppose that $\mathbf{W}_n = \mathbf{X}_n/\sqrt{n}$ where \mathbf{X}_n is the $n \times n$ Wigner matrix described in Theorem 3.1. Also, for simplicity, assume that the entries X_{ij} are uniformly bounded. For each $k=1, \dots, n$, let $\alpha_{n,k}$ be the $(n-1) \times 1$ vector that is the k -th column of \mathbf{W}_n with the k -th element removed, and let $\mathbf{W}_{n,k}$ be the $(n-1) \times (n-1)$ matrix derived by removing the k -th row and k -th column from \mathbf{W}_n . If s_n denotes the Stieltjes transform of \mathbf{W}_n , then, by using a standard matrix inversion formula:

$$s_n(z) = \frac{1}{n} \operatorname{tr}((\mathbf{W}_n - z\mathbf{I}_n)^{-1}) = \frac{1}{n} \sum_{k=1}^n \left(\frac{X_{kk}}{\sqrt{n}} - z - \alpha_{n,k}^* (\mathbf{W}_{n,k} - z\mathbf{I}_{n-1})^{-1} \alpha_{n,k} \right)^{-1}.$$

By the description of \mathbf{W}_n , for each k , $\alpha_{n,k}$ and $\mathbf{W}_{n,k}$ are independent and $\alpha_{n,k}$ has i.i.d. entries with zero mean and variance $1/n$. Hence, the quadratic form $\alpha_{n,k}^* (\mathbf{W}_{n,k} - z\mathbf{I}_{n-1})^{-1} \alpha_{n,k}$ concentrates around $n^{-1} \operatorname{tr}((\mathbf{W}_{n,k} - z\mathbf{I}_{n-1})^{-1})$. Notice also that the $\mathbf{W}_{n,k}$'s are identically distributed as $1/\sqrt{n}$ times an $(n-1) \times (n-1)$ Wigner matrix so that for each k , $n^{-1} \operatorname{tr}((\mathbf{W}_{n,k} - z\mathbf{I}_{n-1})^{-1})$ can be approximated by $s_{n-1}(z)$ and subsequently by $s_n(z)$. Finally, noting that the contribution from the terms X_{kk}/\sqrt{n} can be neglected, we have the following approximate identity:

$$s_n(z) \approx \frac{1}{-z - s_n(z)}$$

for large enough n . From these heuristics, it is expected that for each $z \in \mathbb{C}^+$, $s_n(z)$ converges almost surely to $s(z)$ which satisfies the identity $s(z)(z + s(z)) = -1$, the latter being satisfied by the Stieltjes transform of the semicircle law with p.d.f. (8).

A different decomposition is used to deal with the ESD of a Wishart matrix $\mathbf{S}_n = n^{-1} \mathbf{X}_n \mathbf{X}_n^*$, where \mathbf{X}_n is $p \times n$ and has i.i.d. standardized entries, when $p/n \rightarrow \gamma \in (0, \infty)$. An important ingredient here is the following representation of the resolvent $\mathbf{R}_n(z) := (\mathbf{S}_n - z\mathbf{I}_p)^{-1}$:

$$z\mathbf{R}_n(z) + \mathbf{I}_p = \mathbf{R}_n(z)\mathbf{S}_n = \frac{1}{n} \sum_{j=1}^n \mathbf{R}_n(z)X_jX_j^*. \quad (14)$$

Define $\mathbf{R}_{(-j)}(z) = (\mathbf{S}_n - (1/n)X_jX_j^* - z\mathbf{I}_p)^{-1}$, and use the rank one perturbation formula for inverses to write

$$\mathbf{R}_n(z) = \mathbf{R}_{(-j)}(z) - \frac{\frac{1}{n}\mathbf{R}_{(-j)}(z)X_jX_j^*\mathbf{R}_{(-j)}(z)}{1 + \frac{1}{n}X_j^*\mathbf{R}_{(-j)}(z)X_j}, \quad j = 1, \dots, n.$$

Substituting this in (14), after straightforward algebra, one has

$$\frac{z}{p} \operatorname{tr}(\mathbf{R}_n(z)) + 1 = \frac{n}{p} - \frac{1}{p} \sum_{j=1}^n \frac{1}{1 + \frac{1}{n}X_j^*\mathbf{R}_{(-j)}(z)X_j}.$$

Now, $s_n(z) = p^{-1} \operatorname{tr}(\mathbf{R}_n(z))$ is the Stieltjes transform of the ESD of \mathbf{S}_n . In addition, for each j , by the structure of X_j and the fact that it is independent of $\mathbf{S}_n - n^{-1}X_jX_j^*$, we have the approximation $X_j^*\mathbf{R}_{(-j)}(z)X_j \approx \operatorname{tr}(\mathbf{R}_{(-j)}(z))$ which holds in a probabilistic sense. Indeed, the concentration of random quadratic forms of this kind appears repeatedly in the analysis of random Hermitian matrices. Further, by approximating $\operatorname{tr}(\mathbf{R}_{(-j)}(z))$ by $\operatorname{tr}(\mathbf{R}_n(z))$, which can be justified whenever $z \in \mathbb{C}^+$, and replacing p/n by its limiting value γ , we have the approximate equation: $\gamma z s_n(z) + \gamma \approx 1 - (1 + \gamma s_n(z))^{-1}$, which, after a simplification, can be expressed as

$$s_n(z) \approx \frac{1}{1 - \gamma - \gamma z s_n(z) - z}, \quad z \in \mathbb{C}^+.$$

A limiting version of this equation is satisfied by the Stieltjes transform of the Marčenko–Pastur law (see Eq. (20) and Theorem 3.4).

For a rigorous proof of these results, using a formal application of the principles indicated here, one may refer to Chapters 2 and 3 of Bai and Silverstein (2009). The key steps are the following:

- (i) A truncation and centralization of the entries of the random matrix so that the resulting entries have zero mean and are uniformly bounded while the ESDs of the original and truncated matrices are asymptotically equivalent.
- (ii) Establishing that $s_n(z) - \mathbb{E}[s_n(z)] \rightarrow 0$ a.s. for each $z \in \mathbb{C}^+$ by using martingale decomposition techniques.
- (iii) Establishing that $\mathbb{E}[s_n(z)] \rightarrow s(z)$ for each $z \in \mathbb{C}^+$ for some limiting function $s(z)$.
- (iv) Proving that the limiting function $s(z)$ is the unique solution of a functional equation satisfied by the Stieltjes transform of the limiting ESD.

3.1.3. Convergence rate of the ESD

Since the derivation of limiting ESDs in the context of Wigner and Wishart matrices, there have been numerous investigations to obtain the rates of convergence of the ESDs to their limits. Pioneering works by Bai (1993a,b) established Berry–Esseen type inequalities for the difference of the two distributions in terms of their Stieltjes transforms for Wigner and Wishart matrices when the data matrices have independent entries. Applying this inequality, a convergence rate for the expected ESD of a large Wigner matrix was proved to be $O(n^{-1/4})$ and that for the sample covariance matrix was shown to be $O(n^{-1/4})$ if the ratio of the dimension to the degrees of freedom is far from 1, and $O(n^{-5/48})$ if the ratio is close to 1. Some further developments can be found in Bai et al. (1997, 1999, 2002). In Bai et al. (2003), the rate of convergence for the ESD of a sample covariance matrix when $\gamma = 1$ was improved to $O(n^{-1/8})$. In Götze and Tikhomirov (2003, 2004), for both Wigner and Wishart matrices, assuming sufficient number of moments for the entries, it was shown that as long as γ is away from 1, the rate of convergence of expected ESD and the ESD to the limiting ESD is $O(n^{-1/2})$. The works of Tao and Vu (2011, 2012a) show that the rate of convergence is at most $O(n^{-1} \log n)$. In Tao and Vu (2012b), they derived a sharp concentration inequality for the number of sample eigenvalues of a normalized Wigner matrix in any interval, which shows that the eigenvalues in the bulk spectrum are localized to an interval of width $O(n^{-1}(\log n)^{O(1)})$. These results are then used to provide sharp concentration bounds for individual sample eigenvalues around the corresponding quantiles of the semicircle law. These also extend certain results in Erdős et al. (2009, 2012) on the local behavior of the sample eigenvalues of a Wigner matrix. Among related works, Gustavsson (2005) established asymptotic normality of the k -th eigenvalue of a Gaussian complex Wigner (GUE) matrix as well as the joint distribution of several such eigenvalues, when both k and $n-k$ tend to infinity as $n \rightarrow \infty$. O’Rourke (2010) extended this result to the setting of Gaussian real Wigner matrices (GOE). Tao and Vu (2010b, 2011) extended the result of Gustavsson (2005) for the GUE to a class of Hermitian Wigner matrices with non-Gaussian entries whose first four moments match the Gaussian moments, which is an instance of the universality phenomena discussed in detail in Section 3.3. In addition, Pillai and Jin (2011) showed that, if the entries of the data matrix have sub-exponential decay, then the Stieltjes transform of the normalized standard Wishart matrix converges to the Marčenko–Pastur law at rate $O(n^{-1})$, with rate constant depending on the imaginary part of the complex number at the Stieltjes transform being evaluated.

3.1.4. Extension to non-i.i.d. settings

There have been several extensions of Theorems 3.1 and 3.2 that established asymptotic limits of ESDs corresponding to data matrices that do not necessarily have i.i.d. entries. If in Theorems 3.1 and 3.2, the assumptions are weakened to require that X_{jk} ’s are only independent, then one assumes the additional Lindeberg-type condition

$$\frac{1}{\eta^2 n^2} \sum_{j,k} \mathbb{E} \left[|X_{jk}|^2 \mathbf{1}_{\{|X_{jk}| > \eta \sqrt{n}\}} \right] \rightarrow 0 \quad \text{as } n \rightarrow \infty \text{ for any } \eta > 0. \quad (15)$$

Since in statistics, random vectors commonly have arbitrary positive definite matrices as their population covariances, a lot of effort has gone into deriving asymptotic results for ESDs of matrices of the form $\mathbf{W} = n^{-1} \mathbf{A}^{1/2} \mathbf{X} \mathbf{X}^* \mathbf{A}^{1/2}$ where \mathbf{X} has independent entries with zero mean and unit variance, and \mathbf{A} is a random matrix which is positive definite and is independent of \mathbf{X} , and $\mathbf{A}^{1/2}$ is the Hermitian square-root of \mathbf{A} . Clearly, if $\mathbf{Y} = \mathbf{A}^{1/2} \mathbf{X}$, then conditionally on \mathbf{A} , the columns of \mathbf{Y} are independent with zero mean and common covariance \mathbf{A} , a setting commonly studied in multivariate analysis. As discussed in Section 2, a context in which the study of eigenvalues of such matrices becomes important is in MANOVA where we encounter the generalized eigenvalue problem

$$\det(\mathbf{X}_1 \mathbf{X}_1^* - \lambda \mathbf{X}_2 \mathbf{X}_2^*) = 0, \quad (16)$$

where \mathbf{X}_j are $p \times n_j$ random matrices with entries having zero mean and finite variance, and \mathbf{X}_1 and \mathbf{X}_2 are independent. In this case, the roots λ of (16) are the eigenvalues of the matrix $n_1^{-1} \mathbf{X}_1 \mathbf{X}_1^* \mathbf{A}$, or equivalently of $n_1^{-1} \mathbf{A}^{1/2} \mathbf{X}_1 \mathbf{X}_1^* \mathbf{A}^{1/2}$ where $\mathbf{A} = (n_1/n_2)(n_2^{-1} \mathbf{X}_2 \mathbf{X}_2^*)^{-1}$. A matrix of the latter type is referred to as a multivariate F -matrix since it is a generalization of the F -statistic used in univariate hypothesis tests. We remark that the eigenvalues of $\mathbf{X}_1 \mathbf{X}_1^* (\mathbf{X}_2 \mathbf{X}_2^*)^{-1}$ are in one to one correspondence with the eigenvalues of $\mathbf{X}_1 \mathbf{X}_1^* (\mathbf{X}_1 \mathbf{X}_1^* + \mathbf{X}_2 \mathbf{X}_2^*)^{-1}$, the so-called *double Wishart matrix*. Since the pioneering work of Wachter (1980), who considered the limiting behavior of the solution of (16) when the entries of \mathbf{X}_1 and \mathbf{X}_2 are i.i.d. $N(0, 1)$, many researchers, including Silverstein (1985), Yin (1986), Yin et al. (1983, 1988), and Yin and Krishnaiah (1983), have investigated the limiting spectral distribution of the multivariate F -matrices, or more generally of products of two independent random matrices. Yin and Krishnaiah (1983) established the existence of the limiting ESD of the matrix sequence $\mathbf{S}_n \mathbf{A}_p$, where \mathbf{S}_n is a standard Wishart matrix of dimension p and degrees of freedom n with $p/n \rightarrow \gamma \in (0, \infty)$, \mathbf{A}_p is a sequence of $p \times p$ positive definite matrix satisfying $p^{-1} \text{tr}(\mathbf{A}_p^k) \rightarrow h_k$ as $p \rightarrow \infty$, and the sequence $\{h_k\}$ satisfies the Carleman condition. Yin (1986) extended the result to the setting where \mathbf{S}_n is the sample covariance matrix corresponding to a $p \times n$ data matrix with i.i.d. real random variables with zero mean and unit variance. Yin et al. (1983) proved the existence of the limiting ESD for a sequence of multivariate F -matrices and Silverstein (1985) derived the functional form of the limiting ESD. Note, however, that the correspondence between the sample canonical correlation coefficients based on two independent sets of variables and the generalized eigenvalue problem (16) is exactly valid only if the variables are jointly Gaussian. Yang and Pan (2012) proved the existence of the limiting ESD of canonical correlation coefficients under much weaker

distributional assumptions. Recently, [Bai et al. \(2012\)](#) proved the existence of the limiting ESD for the eigenvalues of *beta matrices*, i.e., matrices of the form $\mathbf{U}(\mathbf{U} + \alpha\mathbf{V})^{-1}$ where \mathbf{U} and \mathbf{V} are independent Wishart-type matrices and α is a positive constant.

A result on the existence of the limiting ESD for products of random matrices is given below.

Theorem 3.3. Suppose that the entries of the $p \times n$ matrix \mathbf{X}_n are independent complex random variables satisfying (15), that \mathbf{A}_p is a sequence of Hermitian matrices independent of \mathbf{X}_n , and that the ESD of \mathbf{A}_p tends to a nonrandom limit $F^{\mathbf{A}}$ in probability (or almost surely). If $p/n \rightarrow \gamma \in (0, \infty)$, then the ESD of the product $(n^{-1}\mathbf{X}_n\mathbf{X}_n^*)\mathbf{A}_p$ tends to a nonrandom limit in probability (or almost surely, accordingly).

A generalization of this result, which is particularly useful in statistics, was proved by [Silverstein and Bai \(1995\)](#).

Theorem 3.4. Suppose that the entries of the $p \times n$ matrix \mathbf{X}_n are complex random variables that are independent for each n and identically distributed for all n and satisfy $\mathbb{E}[|X_{11} - E(X_{11})|^2] = 1$. Also, assume that $\mathbf{A}_p = \text{diag}(\tau_1, \dots, \tau_p)$, where $\tau_j \in \mathbb{R}$ and the empirical distribution function of $\{\tau_1, \dots, \tau_p\}$ converges almost surely to a probability distribution function H as $n \rightarrow \infty$. Let $\tilde{\mathbf{W}}_n = \mathbf{B}_n + n^{-1}\mathbf{X}_n^*\mathbf{A}_p\mathbf{X}_n$ where \mathbf{B}_n is an $n \times n$ Hermitian matrix satisfying that $F^{\mathbf{B}_n}$ converges to $F^{\mathbf{B}}$ almost surely, where $F^{\mathbf{B}}$ is a distribution function on \mathbb{R} . Assume further that \mathbf{X}_n , \mathbf{A}_p and \mathbf{B}_n are independent. Then as $n, p \rightarrow \infty$ such that $p/n \rightarrow \gamma \in (0, \infty)$, the ESD $F^{\tilde{\mathbf{W}}_n}$ of $\tilde{\mathbf{W}}_n$ converges to a nonrandom distribution \tilde{F} , where, for any $z \in \mathbb{C}^+$, its Stieltjes transform $\underline{s} = \underline{s}(z)$ is the unique solution in \mathbb{C}^+ of the equation

$$\underline{s} = s_B \left(z - \gamma \int \frac{\tau dH(\tau)}{1 + \underline{s}\tau} \right), \quad (17)$$

where $s_B(z)$ is the Stieltjes transform of $F^{\mathbf{B}}$.

When \mathbf{B}_n is the zero matrix, (17) reduces to

$$z = -\frac{1}{\underline{s}} + \gamma \int \frac{\tau dH(\tau)}{1 + \underline{s}\tau} \quad (18)$$

which gives an explicit inverse function for $\underline{s}(z)$. Assume further that \mathbf{A}_n is positive definite. Defining $s(z) = (1/\gamma)(\underline{s}(z) + (1-\gamma)/z)$ and noticing that the nonzero eigenvalues of

$$\mathbf{W}_n := \frac{1}{n} \mathbf{A}_p^{1/2} \mathbf{X}_n \mathbf{X}_n^* \mathbf{A}_p^{1/2} \quad (19)$$

coincide with those of $n^{-1}\mathbf{X}_n^*\mathbf{A}_p\mathbf{X}_n$, it can be easily deduced that, under the assumptions of [Theorem 3.4](#), the ESD $F^{\mathbf{W}_n}$ of \mathbf{W}_n converges to a nonrandom distribution F almost surely, where the Stieltjes transform $s = s(z)$ of F is the unique solution in the set $\{s \in \mathbb{C} : -(1-\gamma)/z + \gamma s \in \mathbb{C}^+\}$ of

$$s = \int \frac{dH(\tau)}{\tau(1 - \gamma - \gamma z s) - z}. \quad (20)$$

Eqs. (18) and (20) are sometimes referred to as the *Marčenko–Pastur equations*. These are the building blocks for downstream analyses about the behavior of the limiting ESDs of covariance matrices considered here, a topic first studied by [Silverstein and Combettes \(1992\)](#). They characterized the analytic properties of the limiting distribution F , in particular, proving the existence of a continuous density on \mathbb{R}^+ for F when $\gamma \in (0, 1)$, and determining the support of F in terms of the zeros of the derivative, with respect to s (restricted to \mathbb{R}), of z satisfying (18). These equations have also been very useful in estimating the spectrum of \mathbf{A}_p when one observes data of the form $\mathbf{Y}_n = \mathbf{A}_p^{1/2}\mathbf{X}_n$, which is a topic studied in [Section 4](#).

Significant relaxations on the conditions of [Theorem 3.4](#) have been made by [Bai and Zhou \(2008\)](#) who dealt with matrices of the form $n^{-1}\mathbf{Y}_n\mathbf{Y}_n^*$ where the columns of the $p \times n$ matrix \mathbf{Y}_n are independent but there may be arbitrary dependence within each column. They proved the existence of the limiting ESD requiring only that $\mathbb{E}[\bar{Y}_{jk}Y_{lk}] = a_{lj}$ for $1 \leq l, j \leq p$ and the quadratic forms $Y_k^*\mathbf{C}Y_k$ have suitable concentration around $\text{tr}(\mathbf{C}\mathbf{A}_p)$ for every $p \times p$ matrix \mathbf{C} with bounded norm, where Y_k denotes the k -th column of \mathbf{Y}_n and $\mathbf{A}_p = ((a_{jl}))$. They used this result to prove that the ESD of a sample correlation matrix converges almost surely to the Marčenko–Pastur law, thus extending the scope of a result by [Jiang \(2004\)](#). Further applications include proving the existence of the limiting ESD of a sample covariance matrix when the columns of \mathbf{Y}_n are i.i.d. stationary time series, e.g., a causal VARMA process ([Jin et al., 2009](#)), and more generally, a linear process ([Pfaffel and Schlemm, 2012](#); [Yao, 2012](#)).

A powerful method for extending the class of random matrices for which nonrandom limiting ESD exists has been developed by [Chatterjee \(2006\)](#) using the so-called “Lindeberg principle”, which allows one to compare the expectations $\mathbb{E}[f(X_1, \dots, X_n)]$ and $\mathbb{E}[f(Y_1, \dots, Y_n)]$, for a smooth function f and sequences of random variables $\{X_j\}$ and $\{Y_j\}$, by replacing an X variable by a Y variable in a telescoping sum. In applications, it is assumed that one of the sequences, say $\{Y_j\}$, has independent entries, and the behavior of $\mathbb{E}[f(Y_1, \dots, Y_n)]$ is well-understood, while the first and second moments of X_j , conditional on X_1, \dots, X_{j-1} , are close to the first and second moments of Y_j , for all j . [Chatterjee \(2006\)](#) used this approach to generalize the Wigner’s semicircle law by showing that, after appropriate normalization, the ESD of the $n \times n$ matrix \mathbf{X}_n/\sqrt{n} converges to the semicircle law, where \mathbf{X}_n is symmetric and the entries on and above the diagonal are exchangeable random variables with finite fourth moment.

Apart from relaxing the assumption on the entries of the matrices, there have been several attempts to extend the structure of the matrices for which limiting ESDs exist. [Bryc et al. \(2006\)](#) used combinatorial methods to prove the existence of limiting ESDs for random matrices of Hankel, Toeplitz and Markov types. Further theoretical developments for such matrices were made by [Bose and Sen \(2008\)](#). Recently, [El Karoui \(2009a\)](#), [Paul and Silverstein \(2009\)](#) and [Zhang \(2006\)](#), under slightly different assumptions, have proved the existence of the limiting ESD for matrices of the form

$$\mathbf{W}_n = \frac{1}{n} \mathbf{A}_p^{1/2} \mathbf{X}_n \mathbf{B}_n \mathbf{X}_n^* \mathbf{A}_p^{1/2}, \quad (21)$$

where \mathbf{X}_n is $p \times n$ with complex entries with zero mean and unit variance, \mathbf{A}_p and \mathbf{B}_n are positive definite matrices of dimension $p \times p$ and $n \times n$, respectively, and the ESDs $F^{\mathbf{A}_p}$ and $F^{\mathbf{B}_n}$ converge to distributions $F^{\mathbf{A}}$ and $F^{\mathbf{B}}$, respectively, while (7) holds. \mathbf{W}_n given by (21) may be seen as the sample covariance matrix corresponding to a data matrix having a separable covariance structure commonly assumed in spatio-temporal statistics. [El Karoui \(2009a\)](#), [Paul and Silverstein \(2009\)](#) and [Zhang \(2006\)](#) showed that the Stieltjes transform of the limiting ESD is obtained as the unique solution of a coupled functional equation. In related developments, [Hachem et al. \(2006\)](#) studied the limiting ESD of the sample covariance matrix corresponding to a $p \times n$ data matrix \mathbf{Y}_n whose entries are given by $Y_{jk} = \sigma(j/p, k/n)X_{jk}$ where X_{jk} are i.i.d. with mean zero and unit variance and $\sigma: [0, 1] \times [0, 1] \rightarrow (0, \infty)$ is a continuous function. Similar ideas have been used by [Hachem et al. \(2005\)](#) to derive the limiting ESD of $n^{-1} \mathbf{Y}_n \mathbf{Y}_n^*$ when the entries of \mathbf{Y}_n have the moving average representation $Y_{jk} = \sum_{(l,l') \in \mathbb{Z}^2} h_{l,l'} Z_{l-l', l'-k}$, where $Z_{l,l'}$ are i.i.d. complex Gaussian with zero mean and unit variance, and the deterministic complex sequence $\{h_{l,l'}\}$ satisfies $\sum_{(l,l') \in \mathbb{Z}^2} |h_{l,l'}| < \infty$.

Another class of matrices that has become popular in RMT, partly due to its applicability in wireless communications, is the sample covariance matrix corresponding to an “information-plus-noise” matrix, i.e.,

$$\mathbf{W}_n = \frac{1}{n} (\mathbf{R}_n + \sigma \mathbf{X}_n) (\mathbf{R}_n + \sigma \mathbf{X}_n)^*, \quad (22)$$

where the “information matrix” \mathbf{R}_n is a $p \times n$ matrix which is independent of the “noise matrix” \mathbf{X}_n which has independent real or complex entries with zero mean and unit variance, and $\sigma > 0$ is a scale parameter. [Dozier and Silverstein \(2007b,a\)](#) proved the existence of the limiting ESD of \mathbf{W}_n given by (22), and analyzed the behavior of the limit if (7) holds and the ESD of $n^{-1} \mathbf{R}_n \mathbf{R}_n^*$ converges to a nonrandom limit distribution on \mathbb{R}^+ . [Hachem et al. \(2007\)](#) considered a further generalization and derived deterministic equivalents of the Stieltjes transform of the ESD of \mathbf{W}_n given by $\mathbf{W}_n = (\mathbf{R}_n + \mathbf{Y}_n) (\mathbf{R}_n + \mathbf{Y}_n)^*$ where \mathbf{Y}_n is a $p \times n$ matrix with $Y_{jk} = \sigma_{jk}(n) X_{jk} / \sqrt{n}$, $1 \leq j \leq p$, $1 \leq k \leq n$, where X_{jk} are i.i.d. with zero mean and unit variance, and $\{\sigma_{jk}(n)\}$ is an array of uniformly bounded nonnegative numbers satisfying certain regularity conditions, and \mathbf{R}_n is a $p \times n$ deterministic matrix.

3.1.5. Linear spectral statistics

Let F_n be the ESD of an $N_n \times N_n$ Hermitian random matrix with eigenvalues $\lambda_1, \dots, \lambda_{N_n}$. Specifically, for Wishart-type matrices $N_n = p$ and for Wigner-type matrices $N_n = n$. We define a linear spectral statistics (LSS) corresponding to the function g defined on the real line to be the quantity

$$\int g(x) dF_n(x) = \frac{1}{N_n} \sum_{j=1}^{N_n} g(\lambda_j). \quad (23)$$

If, as $n \rightarrow \infty$, $N_n \rightarrow \infty$ and F_n converges to a limiting distribution F (in probability or almost surely) and g is a continuous function, then $\int g(x) dF_n(x) \rightarrow \int g(x) dF(x)$ (in probability or almost surely, respectively). Under such a scenario, it is of interest to study the fluctuations of $\int g(x) dF_n(x)$ around $\int g(x) dF(x)$. One may expect that the process $G_n(x) = \alpha_n(F_n(x) - F(x))$, when viewed as a random element in $\mathbb{D}[0, \infty)$ (the metric space of functions with discontinuities of the first kind along with the Skorohod metric) to converge to some limiting process for an appropriate normalizing sequence $\alpha_n \rightarrow \infty$. Unfortunately, that this is not possible in general follows from the results by [Bai and Silverstein \(2004\)](#) for Wishart matrices and those by [Diaconis and Evans \(2001\)](#) for Haar matrices. However, it may still be possible to find an appropriate sequence α_n such that the random variables

$$G_n(g) := \alpha_n \int g(x) (dF_n(x) - dF(x)) \quad (24)$$

may converge to some limit law for a suitably regular class of functions g . One of the first works in this direction was done by [Jonsson \(1982\)](#), who took $g(x) = x^r$, $r \geq 1$ an integer, and F_n as the ESD of a normalized standard Wishart matrix with dimension p and degrees of freedom n , assuming that (7) holds. Similar results for the Wigner matrix were obtained by [Sinai and Soshnikov \(1998\)](#) who also allowed the power of the monomial to grow slowly with n . Further research proving similar results under a variety of settings has been carried out by [Anderson and Zeitouni \(2006\)](#), [Bai and Silverstein \(2004\)](#), [Bai and Yao \(2005\)](#), [Chatterjee \(2006\)](#), [Johansson \(1998\)](#), [Lytova and Pastur \(2009\)](#) and [Shcherbina \(2011\)](#), among others. Among results involving non-smooth functionals, for Gaussian complex Wishart matrices, [Costin and Lebowitz \(1995\)](#) and [Soshnikov \(2002b\)](#) proved a CLT for the number of eigenvalues falling within an interval (cf. Theorem 4 in [O'Rourke, 2010](#)).

A central limit theorem (CLT) for the Stieltjes transforms for the classical Wigner and sample covariance matrices was proved by [Girko](#) as early as 1975 (see [Girko, 2001](#) for references) and further refinements on these results were made by

Khorunzhy et al. (1996). Bai and Silverstein (2004) proved a CLT for the random vector $(G_n(g_1), \dots, G_n(g_K))$ with $\alpha_n = n$ under (7), when the following assumptions hold:

- (i) $F_n = F^{\mathbf{W}_n}$ is the ESD of $\mathbf{W}_n = n^{-1} \mathbf{A}_p^{1/2} \mathbf{X}_n \mathbf{X}_n^* \mathbf{A}_p^{1/2}$, where \mathbf{A}_p is a $p \times p$ random positive definite matrix whose ESD converges to a nonrandom distribution $F^{\mathbf{A}}$, and is independent of the $p \times n$ matrix \mathbf{X}_n with i.i.d. real- or complex-valued entries X_{jk} satisfying $\mathbb{E}[X_{11}] = 0$, $\mathbb{E}[|X_{11}|^2] = 1$, $\mathbb{E}[|X_{11}|^4] < \infty$; and $\mathbb{E}[X_{11}^4] = 3$ if X_{11} and \mathbf{A}_p are real, while $\mathbb{E}[X_{11}^2] = 0$ and $\mathbb{E}[|X_{11}|^4] = 2$ if X_{11} is complex;
- (ii) g_1, \dots, g_K are analytic functions on an open interval containing $[\liminf \lambda_{\min}^{\mathbf{A}} \mathbf{1}_{(0,1)}(\gamma)(1 - \sqrt{\gamma})^2, \limsup \lambda_{\max}^{\mathbf{A}}(1 + \sqrt{\gamma})^2]$, where $\lambda_{\min}^{\mathbf{A}}$ and $\lambda_{\max}^{\mathbf{A}}$ denote the smallest and the largest eigenvalues of \mathbf{A}_p , respectively.

Certain restrictions are also imposed on the rate of convergence of the ESD of \mathbf{A}_p . A central idea in the proof is the following representation which uses the *Cauchy integral formula*:

$$G_n(g) = -\frac{1}{2\pi i} \oint_C g(z)(n(s_n(z) - s(z))) dz,$$

where C is a positively oriented contour enclosing the support of F and F_n , and s and s_n are the Stieltjes transforms of F and F_n , respectively. So, the problem of finding the limit distribution of $G_n(g)$ reduces to finding the limiting process of $M_n(z) = n(s_n(z) - s(z))$. Bai and Silverstein (2004) constructed a truncated version of $M_n(z)$ on a suitably chosen contour and applied martingale decomposition techniques on the latter process to derive the final result.

For Gaussian Wigner matrices, Johansson (1998) characterized a large class of functions g for which the CLT holds by making use of the explicit form of the joint density of the eigenvalues. Bai and Yao (2005) proved an analogous CLT for linear spectral statistics when F_n is the ESD of the matrix \mathbf{X}_n/\sqrt{n} where \mathbf{X}_n is a Wigner matrix whose entries on and above the diagonal are independent with zero mean and finite fourth moments, the entries above the diagonal have unit variance while the entries on the diagonal have equal variance. Bai et al. (2009) and Bai et al. (2010) used Bernstein polynomial approximation for g , which is now required to be only four times continuously differentiable, to extend the results in Bai and Silverstein (2004) on the CLT of (24) for Wigner and Wishart matrices, respectively. Lytova and Pastur (2009) and Shcherbina (2011) extended the scope of the results of Bai and Silverstein (2004) and Bai and Yao (2005) by relaxing the conditions on the moments of the entries of the data matrices, while requiring that the tails of the Fourier transform of g decay at certain rates, which translate into the requirement that g has bounded derivatives of a certain finite order. In contrast to the Stieltjes transform-based approaches, Lytova and Pastur (2009) used the Fourier transform of the LSS as the basic building block. Their method depends on first proving the results for matrices with Gaussian entries and then utilizing an interpolation between the random matrix ensemble of interest and a conveniently chosen Gaussian matrix, and then making use of an extension of the well-known *Stein's Lemma* (see Proposition 3.1 of Lytova and Pastur, 2009). Chatterjee (2006) used an approach based on *Stein's method for normal approximation* (see Diaconis and Holmes, 2004 for an exposition) for proving CLTs of LSS that additionally gives a bound on the total variation distance between the distribution of the normalized LSS and the limiting Gaussian distribution. Further developments in terms of deriving limit laws for normalized individual entries of the matrix $g(\mathbf{W}_n)$,¹ where \mathbf{W}_n is a normalized Wigner or Wishart matrix, have been made in O'Rourke et al. (2011a,b), Pastur and Lytova (2011) and Pizzo et al. (2012). A CLT for the log-determinant of a Wigner matrix with first four moments matching with either the GUE or GOE ensemble and with explicit mean and variance terms has been established by Tao and Vu (2012c). A rate of convergence for the CLT of the log-determinant of random matrices with independent random variables has been established by Nguyen and Vu (2012). CLTs for certain special classes of linear spectral statistics, for example the log-determinant of the sample covariance matrix, have been proved in many different contexts, including for information-plus-noise type data matrices (Hachem et al., 2012), and for data matrices having independent entries with a given variance profile (Hachem et al., 2008). The latter results are applicable to problems involving models arising in wireless communication. Zheng (2012) proved a CLT for the LSS of F -matrices (or double Wishart matrices) and Bai et al. (2012) dealt with related "Beta" matrices, which are applicable to problems involving testing equality of the covariance matrices of two populations.

In addition to limit theory, there have been several recent works yielding finite sample probability inequalities on the fluctuations of LSS around their means under minimal smoothness assumptions on g , and under appropriate tail behavior on the entries of the data matrix. One of the simplest versions is the following result (Theorem 6.1 of Guionnet, 2009) about LSS for Wigner matrices.

Theorem 3.5. Suppose that $\mathbf{W}_n = \mathbf{X}_n/\sqrt{n}$ where \mathbf{X}_n is a Wigner matrix and the entries $\{X_{jk} : 1 \leq j \leq k \leq n\}$ are independent and their probability distributions satisfy a logarithmic Sobolev inequality with constant $c < \infty$.² Then, for any Lipschitz function g on

¹ If \mathbf{X} an $N \times N$ Hermitian matrix with spectral decomposition $\mathbf{X} = \mathbf{E} \text{diag}(\lambda_1, \dots, \lambda_N) \mathbf{E}^*$ where \mathbf{E} is unitary, then $g(\mathbf{X}) = \mathbf{E} \text{diag}(g(\lambda_1), \dots, g(\lambda_N)) \mathbf{E}^*$.

² A probability measure P on \mathbb{R}^N is said to satisfy the logarithmic Sobolev inequality with constant c if, for any differentiable function $f : \mathbb{R}^N \rightarrow \mathbb{R}$

$$\int f^2 \log \frac{f^2}{\int f^2 dP} \leq 2c \int \|\nabla f\|_2^2 dP$$

where $\|\nabla f\|_2^2 = \sum_{j=1}^N (\partial_{x_j} f)^2$. A distribution satisfying a logarithmic Sobolev inequality has sub-Gaussian tails (see Ledoux, 2001 for further characterizations).

\mathbb{R} with Lipschitz norm $|g|_{\mathcal{L}}$, and for any $\delta > 0$

$$\mathbb{P}\left(\left|\int g(x) dF^{\mathbf{W}_n}(x) - \mathbb{E}\left(\int g(x) dF^{\mathbf{W}_n}(x)\right)\right| > \delta\right) \leq 2 \exp\left(-\frac{n^2 \delta^2}{4c|g|_{\mathcal{L}}^2}\right). \quad (25)$$

A survey of such concentration inequalities for a wide class of random matrices can be found in [Anderson et al. \(2009\)](#), [Guionnet \(2009\)](#) and [Guionnet and Zeitouni \(2000\)](#).

3.1.6. Spectrum separation

A natural question is *how does the finite sample behavior of the ESD compare with the limiting distributions?* This question is relevant, since the limiting ESD results do not clarify whether a vanishingly small fraction of the extreme eigenvalues behaves somewhat differently from the bulk spectrum. A significant step towards answering this was taken by [Bai and Silverstein \(1998\)](#) who worked in the setting of [Theorem 3.4](#) with \mathbf{B}_n being the zero matrix, and showed that if $\|\mathbf{A}_n\|$ is bounded, then almost surely, for a large enough n , none of the eigenvalues of the sample covariance matrix \mathbf{W}_n given by (19) lies in a closed interval outside the support of the limiting ESD. A further refinement of this result was made by [Bai and Silverstein \(1999\)](#). To describe their main result, first note that if the limiting spectrum of the population covariance (i.e., \mathbf{A}_n) consists of a finite number of disjoint components, then as is to be expected, the support of the limiting ESD of \mathbf{W}_n also splits into the same number of disjoint components provided $\gamma = \lim_{n,p \rightarrow \infty} p/n$ is sufficiently small. [Bai and Silverstein \(1999\)](#) showed that, under this setting, for large n , almost surely exactly the same fractions of eigenvalues lie in disjoint intervals containing the different components of the support of the limiting ESD. These results were proved by showing the convergence of Stieltjes transforms at an appropriate rate, uniform with respect to the real part of z over certain intervals, while the imaginary part of z converges to zero. They have important implications in array signal processing, where a simple model is that an unknown number q of sources emit signals onto an array of p sensors in a noise-filled environment ($q < p$) and one goal is to estimate q from independent samples. The result by [Bai and Silverstein \(1999\)](#) allows one to determine the fraction of sources to sensors, i.e., q/p , provided the sample size is large enough so that the support of the limiting ESD of the sample covariance matrix of the data splits into two components corresponding to the signal and noise. Such results have also been used to estimate the spectrum of the underlying population covariance matrix from the spectrum of the sample covariance matrix (see, e.g. [Mestre, 2008](#)). These results have been extended to other classes of matrices. For example, [Bai and Silverstein \(2012\)](#) and [Paul and Silverstein \(2009\)](#) generalized the result of [Bai and Silverstein \(1998\)](#) to sample covariance matrices of the kind (21) and (22), respectively.

3.2. Behavior at the edge of the spectrum

Extreme eigenvalues of a sample covariance matrix play an important role in statistical analysis. Several questions arise naturally from the description given in [Section 3.1](#) about the limiting ESD of a random matrix. *What happens to the extreme eigenvalues of a random matrix? Do they converge to the extreme points of the support of the limiting ESD?* This question is non-trivial as one may see from [Theorem 3.2](#), the convergence of the ESD of \mathbf{S}_n to the Marčenko–Pastur law F_γ does not rule out the possibility that $o(n)$ number of sample eigenvalues remain outside the support of F_γ even when $n \rightarrow \infty$. That this does not happen was first shown by [Geman \(1980\)](#) who proved that the largest eigenvalue of \mathbf{S}_n converges almost surely to $(1 + \sqrt{\gamma})^2$ under a growth condition on all the moments of the underlying distribution. Later, [Yin et al. \(1988\)](#) proved the same result only assuming finiteness of the fourth moment, a condition that was shown to be necessary by [Bai et al. \(1988\)](#). Under the same assumption, [Bai and Yin \(1993\)](#) proved that the smallest eigenvalue of the sample covariance matrix converges to $(1 - \sqrt{\gamma})^2$ when $p < n$. [Bai and Yin \(1988\)](#) found the necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix to 2, which, by symmetry of Wigner matrices, is also valid for the almost sure convergence of the smallest eigenvalue to -2 . A generalization of these results was obtained by [Bai and Silverstein \(1999\)](#) for matrices of the form $n^{-1}\mathbf{A}_p^{1/2}\mathbf{X}_n\mathbf{X}_n^*\mathbf{A}_p^{1/2}$ where \mathbf{A}_p is a $p \times p$ positive definite matrix with ESD converging to a limiting distribution on \mathbb{R}^+ as (7) holds, and the $p \times n$ matrix \mathbf{X}_n has i.i.d. entries with zero mean and unit variance.

These results brought to focus the question of fluctuations of the extreme eigenvalues of random matrices around their limits. Since classical test procedures in multivariate analysis like Roy's largest root test are defined in terms of the extreme eigenvalues of single or double Wishart matrices, obtaining precise quantitative answers to such questions, for example by obtaining limiting distributions for appropriately normalized extreme eigenvalues for certain classes of random matrices, is of immense importance from the point of view of obtaining precise cut-off thresholds or p -values for these tests. The last decade has seen many exciting new developments towards deriving such results, starting from the somewhat idealized scenario of independent Gaussian data and followed by subsequent study of universality of such phenomena. We present a summary of these developments in the rest of this subsection.

3.2.1. Connection to orthogonal polynomials

Joint distributions of eigenvalues of the three classical random matrix models – Wigner, Wishart and double Wishart, where the entries of the data matrix are real or complex Gaussian, are available in close form. The densities (for the

Table 1

Weight functions for random matrix ensembles.

| $w(x)$ | Domain of $w(x)$ | Orthogonal polynomial | Matrix ensemble |
|-----------------------------------|---------------------|-----------------------|-----------------|
| $e^{-x^2/2}$ | $(-\infty, \infty)$ | Hermite | Wigner |
| $x^a e^{-x}$, ($a > 0$) | $(0, \infty)$ | Laguerre | Wishart |
| $(1-x)^a(1+x)^b$, ($a, b > 0$) | $(0, 1)$ | Jacobi | Double Wishart |

unordered eigenvalues of unnormalized matrices) are given in the following general form:

$$f_{w,\beta}(x_1, \dots, x_N) = c_{w,\beta} \prod_{j=1}^N w(x_j)^{\beta/2} \prod_{1 \leq j < k \leq N} |x_j - x_k|^\beta, \quad (26)$$

where w is a nonnegative weight function on \mathbb{R} , $\beta = 1$ or 2 corresponds to the entries of the data matrix being real- or complex-valued, and $c_{w,\beta}$ is a normalizing constant. Note that the weight functions w determine some classical families of orthogonal polynomials as described in Table 1. Since these matrices are orthogonally or unitarily invariant according to whether the entries of the data matrix are real (corresponding to $\beta = 1$) or complex (corresponding to $\beta = 2$) Gaussian, the standard terminology for referring to such matrix ensembles is as follows. Wigner matrices with i.i.d. real entries are referred to as the GOE (Gaussian Orthogonal Ensemble) and those with complex Gaussian entries are referred to as the GUE (Gaussian Unitary Ensemble). In each case the entries of the data matrix are i.i.d. with mean zero and variance one. Similarly, the Wishart matrices, for which the entries of the data matrix are i.i.d. real or complex Gaussian (with zero mean and unit variance), are referred to as LOE and LUE, respectively, where 'L' stands for Laguerre. Correspondingly, the double Wishart matrices, for which the entries of the data matrix are i.i.d. real or complex Gaussian, are referred to as JOE and JUE, respectively, where 'J' stands for Jacobi.

3.2.2. Tracy–Widom laws for the extreme eigenvalues

In a collection of celebrated papers Tracy and Widom (1994a,b, 1996), and Widom (1999) analyzed the limiting distributions of the extreme eigenvalues of a Wigner matrix and derived formulae for describing the distribution of the extreme eigenvalues of a Wishart matrix by utilizing the joint density of the eigenvalues (26). They derived that for the GOE and GUE, as the dimension increases to infinity, the largest eigenvalue converges in distribution to the respective Tracy–Widom laws defined below. Johnstone (2001) showed that the largest eigenvalues for the LUE and LOE, after appropriate normalization, also converge to the corresponding Tracy–Widom laws, as $p, n \rightarrow \infty$ such that $p/n \rightarrow \gamma \in (0, 1]$ (and so for $\gamma \in [1, \infty)$, by interchanging the role of n and p). Slightly earlier, Johansson (2000) proved that the scaling limit for the largest eigenvalue for the LUE is the Tracy–Widom law F_2 given by (27) below.

The c.d.f. of the Tracy–Widom distribution corresponding to GUE and LUE, denoted by F_2 , is given by

$$F_2(s) = \exp\left(-\int_s^\infty (x-s)q^2(x) dx\right), \quad s \in \mathbb{R}, \quad (27)$$

while the c.d.f. of the Tracy–Widom distribution corresponding to GOE and LOE, denoted by F_1 , is given by

$$F_1(s) = \exp\left(-\frac{1}{2} \int_s^\infty (q(x) + (x-s)q^2(x)) dx\right), \quad s \in \mathbb{R}; \quad (28)$$

where $q(x)$ satisfies the Painlevé II differential equation $q''(x) = xq(x) + 2q^3(x)$ with the feature that $q(x) - A(x) \rightarrow 0$ as $x \rightarrow \infty$, where $A(x)$ denotes the Airy function (for properties of the Airy function, one may refer to Olver, 1974). In terms of these distributions, the main results of Johnstone (2001) can be stated as follows.

Theorem 3.6. Suppose that the entries of the $p \times n$ matrix \mathbf{X} are i.i.d. complex Gaussian with mean zero and variance one. Let $l_{1,p}$ denote the largest eigenvalue of $\mathbf{X}\mathbf{X}^*$. If $p/n \rightarrow \gamma \in (0, 1]$, as $n \rightarrow \infty$, then

$$\frac{l_{1,p} - \mu_{n,p}}{\sigma_{n,p}} \Rightarrow W_2, \quad (29)$$

where

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

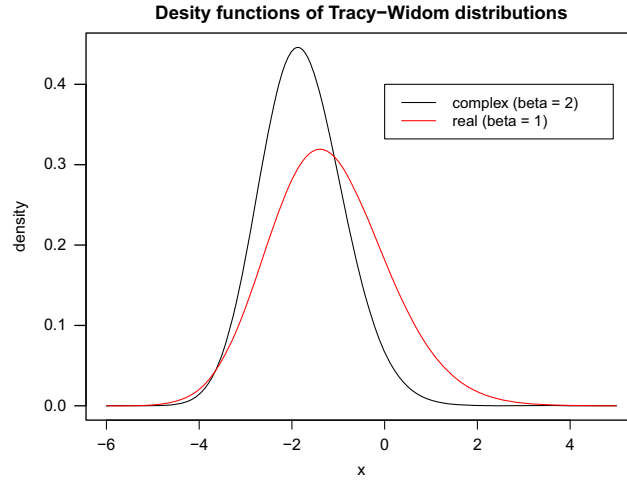


Fig. 2. Tracy–Widom density functions for $\beta = 1, 2$.

If the entries of \mathbf{X} are i.i.d. real Gaussian with mean zero and variance one, and $l_{1,p}$ denotes the largest eigenvalue of $\mathbf{X}\mathbf{X}^T$, then as $n \rightarrow \infty$, so that $p/n \rightarrow (0, 1]$,

$$\frac{l_{1,p} - \mu'_{n,p}}{\sigma'_{n,p}} \Rightarrow W_1, \quad (30)$$

where

$$\mu'_{n,p} = (\sqrt{n-1} + \sqrt{p})^2, \quad \sigma'_{n,p} = (\sqrt{n-1} + \sqrt{p}) \left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}} \right)^{1/3},$$

here the random variables W_1 and W_2 have distributions with c.d.f. F_1 and F_2 , respectively.

An explanation of the appearance of the Tracy–Widom law in terms of the limit of the Fredholm determinant of certain integral operator is given at the end of this subsection. Johnstone (2008) proved similar scaling limits for the largest eigenvalues of JOE and JUE, i.e., the double Wishart matrix ensembles: $\mathbf{U}(\mathbf{U} + \mathbf{V})^{-1}$ where $\mathbf{U} = \mathbf{X}\mathbf{X}^*$ (respectively, $\mathbf{X}\mathbf{X}^T$) and $\mathbf{V} = \mathbf{Y}\mathbf{Y}^*$ (respectively, $\mathbf{Y}\mathbf{Y}^T$), where \mathbf{X} and \mathbf{Y} are independent matrices of dimension $p \times m$ and $p \times n$, respectively, with i.i.d. complex or real standard Gaussian entries. He showed that under the assumption that

$$m = m(p) \rightarrow \infty, \quad n = n(p) \rightarrow \infty \text{ as } p \rightarrow \infty \quad \text{s.t.} \quad \lim_{p \rightarrow \infty} \frac{\min\{p, n\}}{m+n} > 0 \text{ and } \lim_{p \rightarrow \infty} \frac{p}{m} < 1, \quad (31)$$

the normalized quantity $[\log(\theta_{1,p}/(1-\theta_{1,p})) - \mu_p]/\sigma_p$, where $\theta_{1,p}$ denotes the largest eigenvalue of $\mathbf{U}(\mathbf{U} + \mathbf{V})^{-1}$, and μ_p and σ_p are appropriate centering and scaling sequences, converges in distribution to the Tracy–Widom laws F_2 and F_1 , in the complex and real settings, respectively. Jiang (2009) proved Tracy–Widom limits for the largest eigenvalue of the Jacobi ensemble under a different asymptotic regime where $m^2/n \rightarrow \infty$ while $p/n \rightarrow \gamma \in (0, \infty)$, which essentially corresponds to having $m^{-1}\mathbf{U} \approx \mathbf{I}_p$ (Fig. 2).

A striking feature of all these results is the $O(n^{2/3})$ scaling for the fluctuations of the normalized largest eigenvalue (analogous scaling holds also for GOE and GUE), instead of the $O(n)$ scaling for the classical extreme value theory for i.i.d. random variables with appropriate tail behavior, which reflects the fact that eigenvalues of a random matrix repel each other. The latter can be seen from the presence of the Vandermonde determinant in the expression (26) for the joint density of the eigenvalues, which ensures that the eigenvalues of these ensembles are more regularly spaced than what would be the case for i.i.d. random variables.

Many generalizations of these results have been achieved. One class of models for which explicit Tracy–Widom-type limit laws for the extreme eigenvalues have been found is the class of β -ensembles (Dumitriu, 2003; Dumitriu and Edelman, 2002) which correspond to random matrices that can be expressed as $\mathbf{W}\mathbf{W}^*$, where \mathbf{W} is a bi-diagonal matrix with independent χ -distributed entries on the diagonal and the main subdiagonal, with degrees of freedom controlled by the parameter β ($\beta = 1$ and 2 correspond to the real and complex Wishart ensembles). Ramírez et al. (2011) described such limit laws in terms of the eigenstates of a stochastic Airy operator introduced by Edelman and Sutton (2007) and Sutton (2005). Sodin (2010) studied the asymptotic distribution of the eigenvalues of random Hermitian periodic band matrices. Sodin (2009) proved a Tracy–Widom limit law for the largest eigenvalue of a class of sparse random matrices obtained from the adjacency matrix of a random graph by multiplying every entry by a random sign.

We briefly describe the key concepts leading to the derivation of Tracy–Widom laws and their subsequent refinements for the Wishart ensemble. The works of Tracy and Widom showed that for the Wishart ensemble, the c.d.f. of the extreme

eigenvalues can be expressed in terms of Fredholm determinants of integral operators (for relevant operator theory one may refer to Gohberg and Krein, 1969 or Reed and Simon, 1972) whose kernels have integral representations in terms of weighted Laguerre polynomials. In addition, the limiting c.d.f. can be represented in terms of Fredholm determinants of integral operators with kernels represented in terms of the Airy function. The representations are simpler in the complex case than in the real case and hence the former is easier to deal with. Johnstone (2001) proved Theorem 3.6 by utilizing these representations and applying Liouville–Green transformation theory for the convergence of weighted Laguerre polynomials to Airy functions to show that the corresponding integral operators converge in trace class norm. His derivation made use of the important fact that the convergence of an operator in trace class norm implies convergence of its Fredholm determinant. A modified form of the asymptotics of the weighted Laguerre polynomials was used by El Karoui (2003) to extend the domain of Theorem 3.6 to the setting when $p/n \rightarrow 0$ as $p, n \rightarrow \infty$. Paul (2011) used the technique of El Karoui (2003) to prove that under the latter setting, the scaling limits of the normalized smallest eigenvalues are reflected Tracy–Widom laws, which paralleled a result of Baker et al. (1998) that showed that the reflection of F_2 about the origin is the scaling limit for the largest eigenvalue of the LUE. El Karoui (2007) made further refinements of the asymptotic argument of Johnstone (2001) to prove a Berry–Esseen type rate of convergence result for the largest eigenvalue of the LUE and Choup (2006) derived Edgeworth expansions for the same. Ma (2012) derived rates of convergence result, with the appropriate $O(n^{-2/3})$ bound, for both the largest and the smallest eigenvalues of LOE and LUE. Johnstone (2008) derived that under (31), the rate of convergence of the (appropriately normalized) largest eigenvalue of the JOE and JUE to the corresponding Tracy–Widom limits is $O(n^{-2/3})$. Johnstone and Ma (2012) used similar techniques to derive the $O(n^{-2/3})$ bound on the rate of convergence for the largest eigenvalue of GOE and GUE.

3.3. Universality

In the last decade, a lot of effort has been devoted to derive universality of results on the behavior of the eigenvalues of random matrices. Universality essentially means that the limiting behavior of the eigenvalue statistics does not depend on the distribution of the entries. While the statement does not hold in complete generality, in many settings, the behavior of both bulk and edge eigenvalues depend essentially on the first four moments of the distribution of the entries. The investigation on the existence of limiting ESDs as carried out by Bai, Silverstein and contemporary researchers already showed that at the level of the first order convergence (convergence of the Stieltjes transform of the ESD), the behavior is universal, as long as the entries of the matrices are standardized independent random variables satisfying a Lindeberg-type condition. The finer characteristics, such as the limiting distribution of normalized extreme (or edge) eigenvalues, started receiving increased attention with the works of Soshnikov who proved the Tracy–Widom limit of the normalized largest eigenvalues on Wigner (Soshnikov, 1999) and Wishart (Soshnikov, 2002a) matrices. However, these results still required the existence of all moments (in particular, sub-Gaussian tails), symmetry of the distribution of the entries, and for the Wishart case, assumed that the dimension to sample size ratio approaches one. Pécché (2009) extended the results of Soshnikov (2002a) by allowing the dimension to sample size ratio to approach any nonnegative value. Significant progress was made on the relaxation of the symmetry requirement in Pécché and Soshnikov (2008). Proofs of these results used sophisticated combinatorial techniques. Bulk universality, expressed for example in terms of the limiting behavior of the correlation functions of the eigenvalues, has been achieved through a different set of approaches. First results of this kind were proved by Johansson (2001) and were subsequently improved by Ben Arous and Pécché (2005) and Johansson (2009) in the setting of Gauss divisible Hermitian ensembles for which the matrices are in the form of a Wigner matrix perturbed by a Gaussian Wigner matrix.

Significant new developments on the universality phenomena have been made by Erdős et al. (2010, 2009, 2012), Erdős (2010), Erdős and Yau (2012), and Tao and Vu (2010a,b, 2011, 2012a,b) who used analytical techniques to study the question of both bulk and edge universality and managed to remove much of the restrictions on the distribution of the entries. Indeed, the “four moments theorems” of Tao and Vu assert effectively that the limiting behavior of the local statistics of Wigner and Wishart matrices are the same as when the entries are i.i.d. standard Gaussian, provided the first four moments of the entries match with those of the standard Gaussian. The technical arguments differ somewhat between the bulk spectrum and the edge of the spectrum. Tao and Vu (2010b) and Wang (2012) proved the universality of local eigenvalue statistics at the edge of the spectrum for the Wigner and Wishart cases, respectively. An instance of such results is the following “Four Moments Theorem” (partly restating Theorem 5 in Tao and Vu, 2012a).

Theorem 3.7. Let $\mathbf{X} = ((X_{ij}))$ and $\mathbf{X}' = ((X'_{ij}))$ be $p \times n$ matrices with $p, n \rightarrow \infty$ such that $p/n \rightarrow y \in (0, 1]$. The entries X_{ij} (resp. X'_{ij}) are jointly independent, have mean zero and variance 1, and obey the moment condition $\sup_{i,j} \mathbb{E}[|X_{ij}|^{C_0}] < C$ for a sufficiently large constant $C_0 \geq 2$ and some C independent of p, n . Moreover, all the moments of order up to 4 are identical for X_{ij} and X'_{ij} . Let \mathbf{S} and \mathbf{S}' denote the associated covariance matrices. Then the following holds for sufficiently small c_0 and for every $\varepsilon \in (0, 1)$ and for every $k \geq 1$.

Let $G : \mathbb{R}^k \rightarrow \mathbb{R}$ be a smooth function obeying the derivative bound

$$|\nabla^j G(x)| \leq n^{c_0}$$

for all $0 \leq j \leq 5$ and $x \in \mathbb{R}^k$. Then for any $\varepsilon p \leq i_1 < i_2 < \dots < i_k \leq (1 - \varepsilon)p$, and for sufficiently large n depending on ε, k, c_0 , we have

$$\mathbb{E}[|G(n\lambda_{i_1}(\mathbf{S}), \dots, n\lambda_{i_k}(\mathbf{S})) - G(n\lambda_{i_1}(\mathbf{S}'), \dots, n\lambda_{i_k}(\mathbf{S}'))|] \leq n^{-c_0}, \quad (32)$$

where λ_j denotes the j -th largest eigenvalue.

The key concepts involved in deriving these results are (i) quantification of the gap between successive ordered eigenvalues at the edge of the spectrum; (ii) concentration of the fraction of eigenvalues over any interval around the integral of the p.d.f. of the limiting ESD over the same interval; (iii) delocalization of the singular vectors of the data matrix; and (iv) asymptotic negligibility of the difference between the expected value of suitably smooth functions of the normalized eigenvalues at the edge of the spectrum. Steps (i)–(iii) are derived through a very careful analysis of the Stieltjes transform of the ESD of the random matrix, while step (iv) is derived by an application of the Lindeberg principle mentioned in Section 3.1.4. Erdős and coauthors' approach relies on deriving a local semicircle law (for Wigner matrices) and then proving universality for Gauss divisible matrices through utilizing the notion of *Dyson Brownian motion*. The final step involves approximating all Wigner matrix ensembles with Gauss divisible ensembles. This step and the establishment of local semicircle laws require establishing tight bounds for the individual entries of the resolvent matrices. Erdős et al. (2012) gave a detailed overview of these techniques and also extended the universality results to so-called generalized Wigner matrices, where the entries are independent but have different variances, and to certain classes of banded Hermitian random matrices. Universality at the edge of the spectrum of sample covariance and correlation matrices has been studied by Feldheim and Sodin (2010), Bao et al. (2012) and Pillai and Jin (2012). Lee and Yin (2012) established necessary and sufficient conditions for edge universality of a Wigner matrix. Large deviations of the extreme eigenvalues have been studied by Benaych-Georges et al. (2012).

A different kind of universality that generalizes the weight function w in the joint density of the eigenvalues (26), by considering fairly arbitrary smooth functions with certain rates of decay, has been studied extensively in the mathematical physics community. However, the intersection of the class of models commonly encountered in statistical problems with this universality class is rather limited, and so we do not delve into these interesting mathematical developments. A detailed account of the results and techniques related to this class of problems can be found in the monograph (Deift and Gioev, 2009).

3.4. Behavior of the eigenvectors

The study of the behavior of the eigenvectors of a sample covariance matrix arises in the context of PCA. In the *null setting* (i.e., LOE or LUE) the data matrices have independent standard Gaussian entries, and the joint distributions of the entries are invariant under rotation by orthogonal or unitary matrices. This implies that the matrix of eigenvectors is Haar distributed, i.e., the distribution is uniform on the space of orthogonal (in the real case) or unitary (in the complex case) matrices. Several attempts have been made to derive analogous results even when the entries are not Gaussian. A first result of this kind was proved by Silverstein (1984) who showed that if the first four moments of the entries of the data matrix match those of the standard Gaussian, then the matrix of eigenvectors is *asymptotically Haar distributed* as the dimension and sample size increase while the dimension-to-sample size ratio approaches a positive, finite constant. This was proved by using the fact that if an $N \times N$ orthogonal matrix \mathbf{U} is Haar distributed then for any unit vectors $x \in \mathbb{R}^N$, $y = \mathbf{U}x$ is uniformly distributed over the unit sphere in \mathbb{R}^N , and consequently, for any sequence of unit vectors x , the process $Y_N(t) = \sqrt{N/2} \sum_{j=1}^{\lfloor Nt \rfloor} (|y_j|^2 - 1/N)$, $t \in [0, 1]$, converges in distribution (in the space $\mathbb{D}[0, 1]$) to a Brownian bridge process. This result was further strengthened in Silverstein (1989, 1990). Bai et al. (2011) provided an additional analytical tool for proving that the matrix of eigenvectors is asymptotically Haar distributed. Bai et al. (2007) extended the results of Silverstein (1989, 1990) for matrices of the form $n^{-1} \mathbf{A}_p^{1/2} \mathbf{X}_n \mathbf{X}_n^* \mathbf{A}_p^{1/2}$ where \mathbf{A}_p is a $p \times p$ nonnegative definite matrix and the $p \times n$ data matrix \mathbf{X}_n has i.i.d. standardized entries. A different approach that deals with projections of sample eigen-subspaces onto population eigen-subspaces has been adopted by Ledoit and Pêché (2011) (see Section 4.1).

3.5. Matrix ensembles with heavy-tails

While most of the attention in RMT has been devoted to studying the behavior of eigenvalues of random matrices whose entries have finite fourth moments, random matrices with heavy-tailed i.i.d. entries (so-called *Lévy matrices*) arise naturally in the context of financial data (Bouchaud et al., 2003). Soshnikov (2004, 2006) showed that under the standard asymptotic framework on the dimension to sample size ratio, the behavior of the largest eigenvalues of appropriately normalized Wigner-type matrices with heavy-tailed entries is surprisingly simple. Indeed, in the absence of a finite fourth moment, the asymptotic behavior of the top eigenvalues is determined by the behavior of the largest entries of the matrix. In particular, the largest normalized sample eigenvalue has a Fréchet limit distribution. Soshnikov analyzed the setting when the tail index α (the rate at which tails of the probability distribution decay) of the entries is between 0 and 2. Auffinger et al. (2009) extended the results by Soshnikov (2004), to prove that for $\alpha \in (0, 4)$, and for both Wigner and Wishart-type matrices, the point process of the largest eigenvalues (properly normalized) converges to an inhomogeneous Poisson point process whose intensity function is determined by the tail index of the (common) distribution of the entries of the matrix. A version of

these results was also proved by [Biroli et al. \(2007\)](#) at a physical level of rigor. More recently, [Davis et al. \(2011\)](#) considered sample covariance matrices when the data matrix has i.i.d. rows of linear time series $\sum_{j=1}^{\infty} c_j Z_{i,t-j}$ where $Z_{i,t}$'s are i.i.d. random variables with tail index $\alpha \in (0, 2)$. They showed that the point process of eigenvalues of the sample covariance matrix converges in distribution to a Poisson point process with intensity measure depending on α and $\sum_{j=1}^{\infty} c_j^2$.

4. Applications

In this section we discuss applications of RMT to statistics and allied fields. The focus is more on the practical implications than on further theoretical insight. The latter may be obtained from the multitude of references cited.

4.1. Applications to statistics

As mentioned in [Section 2](#), one of the prime motivations for the study of random matrices in statistics is to find appropriate modifications of the classical multivariate analysis techniques to find approximations and develop new methodologies that are applicable for samples of moderate to large dimension. In this section, we summarize some of these applications in the context of PCA, MANOVA, multivariate tests and classification and discriminant analysis problems. In [Section 5](#), we briefly discuss some statistical methodologies that do not directly use RMT results, but whose developments are guided by issues associated with large random matrix phenomena.

4.1.1. Signal detection using the Tracy–Widom law

One of the earliest uses of the distribution of the largest eigenvalue of the sample covariance matrix is in testing the hypothesis $H_0 : \Sigma = \mathbf{I}_p$ when i.i.d. samples are drawn from a $N(\mu, \Sigma)$ distribution. The Tracy–Widom law for the largest sample eigenvalue under the null Wishart case, i.e., when the population covariance matrix $\Sigma = \mathbf{I}_p$ allows a precise determination of the cut-off value for this test, which, with a careful calibration of the centering and normalizing sequences, is very accurate even for relatively small p and n ([Johnstone, 2001, 2009](#)). The behavior of the power of the test requires formulating suitable alternative models. A formulation of such a model with useful practical implications is discussed in the following subsections. In addition, Tracy–Widom law for the largest eigenvalue has been extensively used for signal detection ([Bianchi et al., 2011](#); [Kritchman and Nadler, 2008](#); [Nadler et al., 2008](#); [Onatski, 2009](#)). Many of these approaches use a sequential hypothesis testing framework whereby the Tracy–Widom law is used to determine the null distribution for testing the presence of an additional signal direction.

4.1.2. PCA under the spiked covariance model

In the previous sections, we discussed the behavior of the eigenvalues and eigenvectors of the sample covariance matrix when the eigenvalues of the population covariance matrix are either identical or are evenly spread out so that none of them “sticks out” from the bulk. However, in high-dimensional statistics, often the variation in the data is modeled as the combined effect of a low-dimensional “signal” buried in a “high-dimensional” noise. If one further assumes an additive structure, then we obtain a convenient description of the data in terms of a factor model. Such models are quite useful from the point of view of signal detection and estimation when the signal is low-dimensional and embedded in isotropic or near isotropic noise. Natural statistical questions arising in problems such as dimension reduction can then be easily translated in terms of the behavior of the eigenvalues and the eigenvectors of the sample covariance matrix. A particularly useful idealized model of this kind, named *spiked covariance model* by [Johnstone \(2001\)](#), has been in use for quite some time in statistics. Under this model, the population covariance matrix Σ is expressed as

$$\Sigma = \sum_{j=1}^M \lambda_j \theta_j \theta_j^T + \sigma^2 \mathbf{I}_p, \quad \text{where } \theta_1, \dots, \theta_M \text{ are orthonormal; } \lambda_1 \geq \dots \geq \lambda_M > 0 \text{ and } \sigma^2 > 0. \quad (33)$$

This model implies that, expect for M leading eigenvalues $\ell_j = \lambda_j + \sigma^2$ for $j=1, \dots, M$, the rest of the eigenvalues are all equal. This model has been studied extensively in the context of high-dimensional PCA since it brings out a number of key issues associated with dimension reduction in the high-dimensional context. [Lu \(2002\)](#) first demonstrated the inconsistency of the sample PCA under (33). In the statistical physics literature, the phase transition behavior of the leading sample eigenvalues has been established by [Hoyle and Rattray \(2004\)](#), [Reimann et al. \(1996\)](#) and [Watkin and Nadal \(1994\)](#). This phase transition phenomenon is described in its simplest form in the following theorem, where, for convenience, we assume $\sigma^2 = 1$.

Theorem 4.1. Suppose that Σ is a $p \times p$ positive definite matrix with eigenvalues $\ell_1 \geq \dots \geq \ell_M > 1 = \dots = 1$, and let $\hat{\ell}_1 \geq \dots \geq \hat{\ell}_p$ be the eigenvalues of the sample covariance matrix $\mathbf{S} = n^{-1} \mathbf{\Sigma}^{1/2} \mathbf{Z} \mathbf{Z}^* \mathbf{\Sigma}^{1/2}$ where the $p \times n$ data matrix \mathbf{Z} has i.i.d. real or complex entries with zero mean, unit variance and finite fourth moment. Suppose that $p, n \rightarrow \infty$ such that $p/n \rightarrow \gamma \in (0, \infty)$. Then, for each

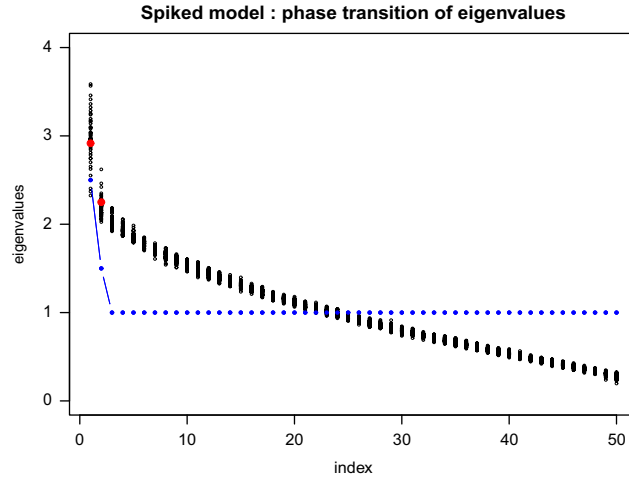


Fig. 3. An illustration of the phase transition of eigenvalues in a spiked covariance model: here, $p=50$, $n=200$ and eigenvalues of the covariance matrix are $\ell_1=2.5$, $\ell_2=1.5$, $\ell_j=1$ for $j=3, \dots, p$. So, $\ell_1 > 1 + \sqrt{p/n}$ and $\ell_2 = 1 + \sqrt{p/n}$. Blue dots correspond to the population eigenvalues. Black circles correspond to the sample eigenvalues (based on i.i.d. Gaussian samples) for 50 replicates. Solid red circles indicate the theoretical limits of the first two eigenvalues for $\gamma = p/n = 0.25$. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

fixed $j=1, \dots, M$

$$\hat{\ell}_j \xrightarrow{\text{a.s.}} \begin{cases} (1 + \sqrt{\gamma})^2 & \text{if } \ell_j \leq 1 + \sqrt{\gamma}, \\ \ell_j \left(1 + \frac{\gamma}{\ell_j - 1}\right) & \text{if } \ell_j > 1 + \sqrt{\gamma}. \end{cases} \quad (34)$$

This result shows that the phase transition of the sample eigenvalues depends only on the magnitude of the corresponding population eigenvalue. This result was first proved rigorously by Baik and Silverstein (2006) by utilizing a finite-sample version of the Marčenko–Pastur equation (18).

These phase transition results have been used to explain certain phenomena in various disciplines. For example, Harding (2008) used this to explain the single factor bias of arbitrage pricing models in finite samples. Patterson et al. (2006) used the phase transition behavior to explain the nature of difficulties of inferring the population structure from genetic data.

Assuming Gaussianity of the observations, Paul (2007) proved further that a phase transition of the corresponding eigenvectors also occurs. Specifically, if $\ell_j \leq 1 + \sqrt{\gamma}$ and the j -th largest eigenvalue is of multiplicity one, then the angle between the j -th sample and population eigenvectors converges to $\pi/2$ a.s., while the a.s. limit is less than $\pi/2$ if $\ell_j > 1 + \sqrt{\gamma}$. He also proved Gaussian fluctuations of the sample eigenvalues and projections of the sample eigenvectors onto the M -dimensional signal subspace in the supercritical regime (population eigenvalue above the phase transition limit). Further refinements of these results have been carried out by various authors. Bai and Zhou (2008) extended the results of Paul (2007) by dropping the Gaussianity assumption. Nadler (2008) used a finite sample asymptotic framework by allowing $\sigma^2 \rightarrow 0$ in (33). Onatski (2012) gave a more comprehensive description of phase transition phenomenon under a factor model. Benaych-Georges and Nadakuditi (2011) gave a more intuitive explanation of the phase transition phenomenon that brings out the central role played by quadratic forms involving the resolvent of a null Wishart matrix. Bai and Yao (2011) extended the scope of the results further by considering a “generalized spiked model”, where the “noise” eigenvalues of the population covariance matrix decay slowly rather than remaining constant (Fig. 3).

The behavior of the sample eigenvalues when the corresponding population eigenvalues is at (critical) or below (subcritical) the phase transition limit has also been a topic of intense study. Baik et al. (2005) applied stationary phase methods to the joint distribution of the sample eigenvalues when the data were complex Gaussian, and showed that in the subcritical regime, the limiting distribution of the leading sample eigenvalues is an interpolated Tracy–Widom law. El Karoui (2007) extended this result to a class of generalized spiked covariance model for complex Gaussian data. Analysis under critical or sub-critical regime for real-valued observations is technically more challenging. Féral and Pécché (2009) used combinatorial methods to prove such results. Benaych-Georges et al. (2011) gave an analytical proof that depends on the decomposition by Benaych-Georges and Nadakuditi (2011) and by controlling the gaps between successive eigenvalues of a Wishart ensemble near the edge, similar to what is used in proving universality results in Tao and Vu (2010b).

There have been several publications on the closely related problem of characterizing the behavior of eigenvalues for deformed Wigner ensembles with the associated phase transition behavior exhibited by the extreme eigenvalues (see Benaych-Georges et al., 2011; Capitaine et al., 2009; Féral and Pécché, 2007; Pizzo et al., 2013; Renfrew and Soshnikov, 2013; Knowles and Yin, 2011, 2012 and the references therein). Chapon et al. (2012) dealt with a low rank deformed signal-plus-noise covariance matrix. Bloemendal and Virág (2010, 2011) studied the limiting distribution of the largest sample

eigenvalues for certain finite rank deformed Wigner and Wishart ensembles, and characterized the limit laws in terms of stochastic Airy operators and Dyson Brownian motion.

4.1.3. Application to MANOVA

Johnstone (2008, 2009) gave an extensive account of the use of Roy's largest root test with Tracy–Widom limit distribution under the various classical double Wishart problems, including CCA, MANOVA and testing of linear hypothesis under multivariate linear models. The striking feature of this approximation is that even though the result is asymptotic in nature, it is quite accurate for sample sizes as low as ten with comparable dimensions. However, it is worthwhile mentioning that for other test statistics such as $\hat{\ell}_1 / \sum_{j=2}^p \hat{\ell}_j$ the convergence to the Tracy–Widom law for finite dimension and sample size may not be as good. Nadler (2011) discussed possible corrections for this. The detection limit of the largest root test in the double Wishart setting under a spiked alternative was studied by Nadakuditi and Silverstein (2010), who established a phase transition behavior qualitatively similar to what is observed in PCA under the spiked covariance model. Behavior under spiked alternatives but with finite dimensions and sample size, and a noise scale parameter converging to zero has been investigated by Nadler and Johnstone (2011).

4.1.4. Tests of hypothesis involving the bulk spectrum

Results on the behavior of the limiting ESD for Wishart and double Wishart type matrices have also been applied to various high-dimensional inferences. We briefly describe a few examples.

Tests for linear hypotheses: Wilk's likelihood ratio test (LRT) is among the most widely used classical test procedure for testing linear hypothesis about the regression parameter \mathbf{B} in a multivariate linear model (6), where ϵ has i.i.d. columns distributed as $N_p(0, \Sigma)$ and \mathbf{B} is a $p \times q$ matrix of unknown regression coefficients. However, the computation of the quantiles of the Wilk's statistic under the null or the alternative requires complex analytic approximations and these distributional approximations are meaningful only for moderate dimension of the dependent variable. Bai et al. (2012) proposed a modification to Wilk's test in a high-dimensional context based on a CLT for linear statistic of F -matrices recently developed by Zheng (2012).

Test for equality of covariance matrices: Bai et al. (2009) considered the problem of testing $H_0 : \Sigma_1 = \Sigma_2$ where Σ_i 's are the population covariance matrices of two independent samples of p -dimensional i.i.d. Gaussian observations. They proposed corrections to the classical likelihood ratio test statistic when p is proportional to the sample size. Under this setting, the traditional χ^2 approximation of the LR test statistic does not work. Their test is based on the CLT for linear spectral statistics of sample covariance matrices and applies even to non-Gaussian populations under the framework that the samples are of the form $\Sigma_i^{1/2} Z_{ij}$ where the vectors Z_{ij} 's have i.i.d. standardized entries. They showed that corrected LR tests have realized size close to the nominal level for both moderate and high dimensions.

Hotelling's T^2 and generalizations: Suppose that X_1, \dots, X_n are i.i.d. $N_p(\mu, \Sigma_p)$, where $\mu \in \mathbb{R}^p$ and Σ_p is positive definite, and we want to test the null hypothesis $H_0 : \mu = 0$ against $H_A : \mu \neq 0$. The classical test is Hotelling's T^2 test that rejects H_0 for large values of the statistic $T^2 := n\bar{X}^T \mathbf{S}^{-1} \bar{X}$, where \bar{X} is the sample mean and \mathbf{S} is the sample covariance matrix. For $p \geq n$, one cannot define T^2 in this way since \mathbf{S}_n is not invertible. Even when $p < n$ but p/n is not small, the dimensionality strongly influences the sampling distribution of the test statistic under both null and alternative hypotheses. Bai and Saranadasa (1996) addressed this problem, assuming that $p/n \rightarrow \gamma \in (0, 1)$ as $n \rightarrow \infty$, and proved a CLT for the T^2 statistic, in addition to showing that Hotelling's T^2 test is inconsistent in this setting. To fix this problem, tests based on certain quadratic functionals of the sample covariance and correlation matrices have been proposed by Chen and Qin (2010) and Srivastava and Du (2008). In contrast, Chen et al. (2011) proposed a regularized Hotelling's T^2 statistic: $RHT(\lambda) := n\bar{X}_n(\mathbf{S}_n + \lambda \mathbf{I}_p)^{-1} \bar{X}_n$, which is well-defined for all $\lambda > 0$. Assuming Gaussianity, and working under the setting considered by Bai and Saranadasa (1996), they constructed an asymptotic test by proving a CLT for $RHT(\lambda)$. Under the same framework, Lopes et al. (2012) proposed a test constructed from averaged Hotelling's T^2 statistics based on random projections of the data into lower-dimensional subspaces. Limit distributions of a generalized version of the Hotelling's T^2 statistic have been derived by Pan and Zhou (2011) under less restrictive distributional assumptions.

4.1.5. Spectrum estimation

The problem of estimating the spectrum of the population covariance matrix Σ based on i.i.d. observations from a population with mean zero and covariance Σ has received a lot of attention due to their use in many high-dimensional inference procedures. Many different approaches to this problem in the existing literature start by modeling the ESD of the population covariance matrix as a mixture of point masses. El Karoui (2008b) used the Marčenko–Pastur equation (18) and employed a linear programming method to minimize the sup-norm difference between z and the expression on the right-hand side with the Stieltjes transform of the dual sample covariance matrix as input. Rao et al. (2008) used the asymptotic distribution of the sample spectral moments to form a pseudo-likelihood and then maximized this with respect to the parameter. Bai et al. (2010) exploited the relationship between the population spectral moments and the sample spectral moments to formulate estimating equations. Chen et al. (2011) proposed an improved version of the latter procedure. Mestre (2008) proposed an estimator based on contour integrals of functions defined in terms of the Stieltjes transform of the ESD of \mathbf{S}_n under the assumption that the sample spectrum separates in clusters corresponding to the mixture proportions in the population spectrum.

One component of estimating the covariance matrix of high-dimensional data is the estimation of the distribution of its eigenvalues, or at least the leading eigenvalues (under a “spiked covariance” setting). The leading eigenvectors can be estimated with a certain precision under a “spiked covariance” model if the leading population eigenvalues are significantly larger than the background noise. However, if any simplifying structural assumption on the eigenvectors cannot be reliably made and the distribution of the eigenvalues of the underlying population covariance matrix may be approximated well by a finite discrete distribution, efforts have been made to estimate the corresponding eigen-subspaces. The approach of [Mestre \(2008\)](#) for eigen-subspace estimation falls in this category. Significant progress in formulating a general framework for inference on eigen-subspaces has been made by [Ledoit and P  ch   \(2011\)](#) in the case where $\mathbf{X} = \mathbf{B}^{1/2}\mathbf{Z}$ with \mathbf{Z} having i.i.d. standardized entries. They proved results about the projections of a group of eigenvectors of \mathbf{S}_n , indexed by the sample eigenvalues, onto a group of population eigenvectors (i.e., eigenvectors of \mathbf{B}) indexed by the population eigenvalues. A key step in their analysis is establishing the limiting behavior of the functional $p^{-1}\text{tr}(\mathbf{g}(\mathbf{B})(\mathbf{S}_n - \mathbf{z}\mathbf{I}_p)^{-1})$ where $\mathbf{g} : \mathbb{R} \rightarrow \mathbb{R}$ is a bounded function with finitely many discontinuities. These results have been used by [Ledoit and Wolf \(2012\)](#) to develop shrinkage estimators of the population covariance matrix that are rotationally invariant and whose construction only requires nonlinear shrinkage of the sample eigenvalues. [Wang et al. \(2012\)](#) utilized similar results to develop a shrinkage estimator of the precision matrix, i.e., the inverse of the population covariance matrix. [Coulliet et al. \(2012\)](#) formulate a different shrinkage strategy motivated by Huber-type M -estimators ([Huber, 1964](#); [Maronna, 1976](#)) to produce a robust estimate of the population covariance matrix.

One encounters large dimensional random matrices in various problems of signal processing, particularly in wireless communication. [Bai and Silverstein \(2009\)](#), [Coulliet and Debbah \(2011\)](#) and [Tulino and Verdú \(2004\)](#) list several such problems, which include (i) determination of the channel capacity of a MIMO (multiple-input-multiple-output) channel, which can be expressed in terms of the logarithm of the determinant of the matrix $\mathbf{I} + \mathbf{S}$ where \mathbf{S} is a random Wishart-type matrix related to the signal-to-noise ratio in the transmission channels; (ii) determination of the limiting SINR (signal-to-interference noise ratio) in random channels and in linearly precoded systems, such as CDMA (code-division-multiple-access) systems ([Bai and Silverstein, 2007](#)); (iii) asymptotic performance analysis of receivers; (iv) energy estimation from multiple sources ([Coulliet et al., 2011](#)). Random matrices also arise more generally in various signal processing problems, e.g., in the detection of input signals ([Nadakuditi and Silverstein, 2010](#); [Silverstein and Combettes, 1992](#)), subspace estimation in a sensory network ([Hachem et al., 2012, 2013](#)).

RMT has also become popular in dealing with financial data, typically those involving a large number of variables, say stock prices, and a not so large number of measurements, which may represent the number of trading days. [Bouchaud et al. \(2003\)](#); [Bouchaud and Potters \(2009\)](#) give many instances of such problems. [Frahm and Jaekel \(2005\)](#) considered a model for financial data (typically log-returns) that accounts for over-dispersion and elliptically symmetric distribution for the variables. The observation vectors under this model can be expressed i.i.d. samples Y of the form $Y = \mu + \Lambda \Gamma U$, where $\mu \in \mathbb{R}^p$, Γ is a $p \times q$ matrix, U is uniformly distributed on the unit sphere \mathbb{S}^{q-1} and Λ is nonnegative random variable independent of U . The behavior of the corresponding sample covariance matrix has been analyzed in detail by [El Karoui \(2009a\)](#) who derived Marčenko–Pastur type limiting equations for the ESD under the asymptotic framework where $n, p, q \rightarrow \infty$ such that n/q and p/q have finite limits. This model is also closely related to the separable covariance model [\(21\)](#). A related model has been used by [Zheng and Li \(2011\)](#) in the context of estimation of integrated covolatility matrix based on high-dimensional, high-frequency financial data.

One of the classical problems in finance is the *Markowitz portfolio optimization* problem (Markowitz, 1952, 1956). This problem is about the optimal strategy for investing in p assets subject to certain linear constraints. If the mean and covariance of the returns from the assets are given by μ and Σ , assuming that both are known, the optimization problem can be formulated as a quadratic programming problem: $\min_w w^T \Sigma w$ subject to $w^T \mu = m_p$ and $w^T \mathbf{1} = 1$, where $\mathbf{1}$ denotes the vector of all 1's and m_p denotes a guaranteed mean return. The solution, denoted w_{opt} , is called the optimal allocation, and the curve $w_{opt}^T \Sigma w_{opt}$, quantifies the risk associated with the optimal strategy, as a function of m_p . Bai et al. (2009) and El Karoui (2009b, 2010c) analyzed this statistical problem from the point of view of RMT when p/n is approximately constant and the observations are i.i.d. A key finding from these is the underestimation of the true risk if one uses the empirical versions of the mean and the covariance matrix. El Karoui (2010c) also suggested ways of correcting this bias.

4.4. Other applications

Many statistical problems involve ridge-type shrinkage and/or quadratic forms involving a Wishart-type matrix. A detailed study of such shrinkage schemes has been carried out by [ElKaroui and Koesters \(2011\)](#). Random matrices have also been used to describe properties of large random graphs, especially the behavior of eigenvalues of graph Laplacian and the adjacency matrix ([Ding and Jiang, 2010](#); [Jiang, 2012](#)). Moreover, RMT has been utilized to understand the analytic properties of compressed sensing ([Cai and Jiang, 2011](#); [Vershynin, 2012](#)). Effective uses of concepts and tools from RMT have also been made in the analysis of spectral clustering methods with a growing number of clusters ([Rohe et al., 2011](#)), in community detection problems in large random networks ([Nadakuditi and Newman, 2012](#)), and in kernel PCA methods for dimension reduction when the observations are high-dimensional and the argument of the kernel function depends on the inner products of pairs of observations ([El Karoui, 2010b,a](#)).

4.5. Computational tools

One of the recent developments that will go a long way in making the results of RMT accessible to professional statisticians and students is the development of computational tools. The densities of Tracy–Widom laws satisfy complicated nonlinear differential equations and hence are not easily accessible. [Johnstone \(2009\)](#) gave a nice survey of the applications of Tracy–Widom laws in the context of different hypothesis testing problems arising from classical statistics. Johnstone and some of his students have developed an R package `RMTstat` (available from the CRAN website at <http://cran.r-project.org/web/packages/RMTstat/>; see [Johnstone et al., 2009](#) for documentation), which can be used to find p -values of these tests. The same package can also be used to compute the c.d.f., p.d.f. and quantiles of the Marčenko–Pastur and Tracy–Widom distributions, and draw random samples from the same. It has functions for generating the limit distributions under a spiked covariance model, i.e., when the population covariance is a finite rank perturbation of a multiple of the identity matrix. Other significant contributors include Raj Rao Nadakuditi (`MATLAB` toolbox `RMTool`), Momar Dieng (`MATLAB` toolbox `RMLab`) and Plamen Koev (`mhg` package for computing hypergeometric functions of a matrix argument).

5. Sparse PCA, CCA, LDA and covariance estimation

The lack of consistency of classical inferential procedures for dealing with problems such as PCA, CCA, LDA and covariance estimation, as outlined by results from RMT, induced a flurry of activity in the statistical community to design regularized estimation schemes that can be effectively utilized in high-dimensional settings where additional structural information on the parameters describing the statistical models is available. This approach benefited from increasingly sophisticated uses of various convex optimization procedures for penalization/regularization and fast algorithms for implementing them, most notably the l_1 -norm penalization schemes. These developments were also motivated partly by the desire to have interpretable predictive models that use a relatively small number of features. This is particularly important in dealing with high-dimensional data arising in genomic or proteomic studies where the number of variables (genes, proteins) can run into thousands whereas the sample sizes are often moderate. The popularity of l_1 -penalized regression procedures and their variants, such as SCAD, grouped lasso, etc., has underlined the success of sparse parametrization in statistical models. During the last decade, the notion of sparsity has also been utilized in dealing with regularization of classical multivariate statistical methodologies which, at a base level, are dependent on some form of the covariance matrix. An important distinction from regression problems is that, covariance matrices and their inverses, being intrinsically higher-dimensional objects, allow multiple ways in which structural information may be incorporated. Therefore, several different notions of sparsity have been utilized in formulating the regularization schemes. In the rest of this section, we briefly describe some of these approaches and the corresponding theoretical developments.

Since [Stein's \(1956\)](#) influential work, which showed the sample covariance matrix based on multivariate normal observations to be a poor estimate of the population covariance matrix, several regularization procedures for estimating the population covariance matrix and their inverses have been proposed. The inverse covariance matrix, or precision matrix, or concentration matrix is very important in Gaussian graphical models since the location of zeros in the precision matrix correspond to the absence of edges in the conditional dependency graph among the variables. Earlier approaches, which were mostly based on some form of shrinkage of the eigenvalues of the sample covariance matrices, adopted the classical fixed dimensional multivariate Gaussian framework, and evaluated the performance of the estimator in terms of the minimax risk under appropriate loss functions. Prominent works of this kind include ([Dey and Srinivasan, 1985, 1986](#); [Haff, 1977, 1980](#); [Loh, 1991](#); [Stein, 1976](#)). With the advent of high-dimensional data and the phenomena associated with the eigenvalues of high-dimensional Wishart-type matrices, new considerations were coming into play while adopting regularization schemes.

5.1. Regularization of covariance and concentration matrices

[Bickel and Levina \(2008a\)](#) proposed two approaches based on banding and tapering applied to the sample covariance matrix and its Cholesky decomposition to estimate the covariance matrix and its inverse. They also derived rates of

convergence of the proposed estimator for certain classes of covariance and precision matrices whose entries essentially decay away from the diagonal at certain algebraic rates. The remarkable aspect of this approach is that the asymptotic analysis suggested that the method is applicable as long as $\log p = o(n)$, provided the structural assumptions on the parameters hold. Related regularization methods have been proposed by Wu and Pourahmadi (2003, 2009). Cai et al. (2010) proved the minimax rates for tapering estimators of covariance and precision matrices in terms of the spectral norm loss. Bhattacharjee and Bose (2013) established convergence rates for banded and tapered estimates of large dimensional covariance matrices under weak dependence among the entries. Cai and Yuan (2012) suggested an adaptive block thresholding approach and showed that it attains the minimax rate.

In contrast with the banding/tapering/block thresholding approaches that depend on assuming a decay of covariances away from the diagonal, and thereby require an implicit ordering of the coordinates (variables), a different class of estimators of the covariance and precision matrices has been formulated based on the notion of sparsity. This notion assumes that a large fraction of the entries of the covariance (correspondingly, precision) matrix are zero, or of small magnitude. Thresholding individual entries of the sample covariance matrix is an effective approach when the covariance matrix is sparse, and with proper choice of thresholds it has been shown to be consistent in the operator norm by Bickel and Levina (2008b) and El Karoui (2008a). Rothman et al. (2009) considered more general thresholding strategies. Cai and Liu (2011a) proposed an adaptive thresholding rule and showed that it attains the minimax rate under the spectral norm over a class of covariance matrices with uniform l_q constraint on the rows. Cai and Zhou (2012a) proved the minimax rates for the thresholding estimator in terms of the matrix l_1 norm loss over the same class. A different form of sparse regularization that depends on imposing an l_1 constraint on the off-diagonal entries of the precision matrix through a penalized likelihood framework has also been investigated by various authors, including Friedman et al. (2008), Rothman et al. (2008), and Yuan and Lin (2007). Lam and Fan (2009) compared different penalization and thresholding methods for sparse precision matrix estimation and derived rates of convergence under the Frobenius norm, while Cai et al. (2011) obtained rates of convergence of the l_1 -constrained estimator under the spectral norm. Indirect l_1 -type penalization schemes for determining the zero entries in the precision matrix have been proposed by Meinshausen and Bühlmann (2006) and Peng et al. (2009).

5.2. Covariance estimation using sparse factor models

A structural assumption on the covariance matrix that is popular in dealing with high-dimensional problems is a factor model or spiked covariance model structure. Under a spiked covariance model, Σ is modeled by (33). Under this framework, estimation of the leading, or “signal” eigenvectors (corresponding eigenvalues greater than σ^2) and estimation of the covariance or the precision matrix are closely related problems. A key conclusion from RMT is that when $p/n \rightarrow 0$, the eigenvectors of standard PCA are inconsistent estimators of the corresponding population eigenvectors. Therefore, various regularization schemes, implicitly or explicitly assuming some form of sparsity of the leading eigenvectors, have been proposed in the literature. For example, Witten et al. (2009) and Zou et al. (2006) imposed l_1 -type sparsity constraints directly on the eigenvector estimates and proposed optimization procedures for obtaining them. d’Aspremont et al. (2007) suggested a semi-definite programming (SDP) problem as a relaxation to the l_0 -penalty for sparse population eigenvectors. Assuming a single spike (i.e., $M=1$ in (33)) and l_0 -sparsity for the first eigenvector, Amini and Wainwright (2008) studied the asymptotic properties of the resulting leading eigenvector of the covariance estimator and obtained the optimal rates of convergence. Karuthgamer et al. (2013) discussed limitations of the semidefinite programming approach and related statistical procedures in extracting the signal that is not very sparse.

Combinations of thresholding and eigen-analysis have proved to be effective tools in sparse PCA. Birnbaum et al. (2012) studied estimation of the leading eigenvectors under (33), assuming l_q -sparsity of the leading eigenvectors, and established lower bounds on the minimax risk for any eigenvector estimator under l_2 loss. Under the same model, Johnstone and Lu (2009) developed an estimation scheme that pre-selects coordinates by thresholding the diagonal of the sample covariance matrix, followed by the spectral decomposition of the submatrix corresponding to the selected coordinates, and proved consistency of this estimator assuming p grows at most polynomially with n . Ma (2013) developed iterative thresholding sparse PCA (ITSPCA), which is based on repeated filtering, thresholding and orthogonalization steps that result in sparse estimators of the subspaces spanned by the leading eigenvectors. Birnbaum et al. (2012) also proposed an estimation procedure named ASPCA which is based on a two-stage coordinate selection scheme. Both these estimators are rate-optimal, so long as the diagonal thresholding scheme is consistent (see Ma, 2013; Paul and Johnstone, 2012). A different two-stage estimation scheme, based on a regression framework, has been proposed by Cai et al. (2012). Incidentally, in order to establish upper bounds on rates of convergence of the estimators, Amini and Wainwright (2008), Ma (2013) and Paul and Johnstone (2012) all utilize results on the concentration of the extreme singular values of a rectangular matrix with i.i.d. standard Gaussian entries, which translates into finite sample probabilistic bounds on the extreme eigenvalues of a Wishart matrix. Similar results have been utilized extensively in the literature on compressed sensing and sparse signal recovery. One may refer to Vershynin (2011) or Vershynin (2012) for detailed discussions on such applications.

5.3. Sparse canonical correlation analysis

Sparse versions of CCA, where the canonical weight vectors (weight vectors for the X and Y variables in the canonical correlation computation), are constrained to be sparse through imposition of sparse penalty, have also been proposed in the

literature, notably by [Hardoon and Shawe-Taylor \(2009\)](#), [Lee et al. \(2009\)](#), [Parkhomenko et al. \(2009\)](#) and [Witten et al. \(2009\)](#). These methods typically formulate the sparse estimation of the canonical vectors in terms of a sequence of iterative convex or non-convex optimization problems that yield sparse solution depending upon the degree of the penalty. However, there does not appear to be any rigorous mathematical treatment of these procedures at this point.

5.4. Sparse discriminant analysis

One of the important applications of estimated covariance or precision matrices is in discriminant analysis, where the interest is in classifying new observations to one of the two classes which are assumed to have possibly different means μ_1, μ_2 , and either the same or different covariances. The result of [Bickel and Levina \(2004\)](#) proved inconsistency of Fisher's classical linear discriminant analysis (LDA), which is applicable when the two populations have the same covariance matrix, under the setting when $p/n \rightarrow 0$. This can again be explained in terms of the behavior of the eigenvalues of the pooled sample covariance matrix. Motivated by this, regularized versions of LDA in the high-dimensional context, for which the regularization is imposed either through sparse penalization or thresholding of the discriminant function, often in combination with sparse estimation of the precision matrix, have been proposed and analyzed by [Cai and Liu \(2011b\)](#), [Clemmensen et al. \(2011\)](#), [Shao et al. \(2011\)](#), [Witten and Tibshirani \(2009, 2011\)](#). A key requirement for good performance of these procedures is the sparsity of the vector $\Sigma^{-1}(\mu_1 - \mu_2)$, where Σ denotes the common population covariance matrix. A different approach that ignores the covariance structure, but assumes sparsity of $\mu_1 - \mu_2$, has been proposed by [Fan and Fan \(2008\)](#).

6. Future directions

There are plenty of multivariate statistical techniques which require certain modifications to be effective in dealing with moderate to high-dimensional data. Here, we briefly discuss some areas where the enhancement of RMT may be beneficial:

- One striking aspect of typical economic/financial problems is that the data are dependent on time, while much of the theory in this field is under the setting of i.i.d. observations. Thus, a thorough investigation of the potential for extending the current theory on the eigenvalues of Wishart-type matrices, when the columns of the data matrix can be viewed as a realization of a high-dimensional multivariate time series, can have a significant impact on econometrics and finance.
- One important open problem is the ability of RMT to handle various forms of missing data in the high-dimensional context. Although there are some works on Hadamard products of random matrices ([Bai and Zhang, 2006](#); [Hachem et al., 2007, 2008](#)) which are applicable to *missing at random* scenarios, a more comprehensive approach using techniques from RMT is likely to be fruitful.
- The ubiquity of massive streaming data, especially in problems involving array signal processing, trade of stocks and various online trading schemes, etc., poses an enticing prospect of using tools and concepts from RMT. Processing of such massive streaming data may be facilitated immensely by reducing computation times through judicious coupling of random projection techniques with tools from RMT.
- A potentially useful avenue for the application of RMT is in numerical optimization algorithms that use gradient based methods for large dimensional data. While there has been an explosive growth in mathematical descriptions in the RMT literature, computational tools have not kept pace with the theoretical developments. Integration of computational tools with tools for analysis of large dimensional data using RMT principles has the potential of creating a new paradigm for statistical practices.

Acknowledgements

We thank Boaz Nadler and two anonymous referees for their helpful suggestions. The research is supported by the National Science Foundation grants DMR-1035468, DMS-1106690, DMS-1209226 and DMS-1305858.

References

- Akemann, G., Baik, J., Di Francesco, P., 2011. *The Oxford Handbook of Random Matrix Theory*. Oxford University Press, USA.
- Amini, A., Wainwright, M., 2008. High-dimensional analysis of semidefinite relaxations for sparse principal components. *Annals of Statistics* 37, 2877–2921.
- Anderson, G., Guionnet, A., Zeitouni, O., 2009. *An Introduction to Random Matrices*. Cambridge University Press.
- Anderson, G., Zeitouni, O., 2006. A CLT for a band matrix model. *Probability Theory and Related Fields* 134, 283–338.
- Anderson, T.W., 1984. *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York.
- Arnold, L., 1967. On the asymptotic distribution of the eigenvalues of random matrices. *Journal of Mathematical Analysis and Applications* 20, 262–268.
- Arnold, L., 1971. On Wigner's semicircle law for the eigenvalues of random matrices. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 19, 191–198.
- Auffinger, A., Ben Arous, G., Pécché, S., 2009. Poisson convergence for the largest eigenvalues of heavy tailed random matrices. *Annales de l'Institut Henri Poincaré—Probabilités et Statistiques* 45, 589–610.

- Bahcall, S.R., 1996. Random matrix model for superconductors in a magnetic field. *Physical Review Letters* 77, 5276–5279.
- Bai, J., 2003. Inferential theory for factor models for large dimensions. *Econometrica* 71, 135–171.
- Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.
- Bai, J., Ng, S., 2007. Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics* 25, 52–60.
- Bai, Z.D., 1993a. Convergence rate of expected spectral distributions of large random matrices. Part I. Wigner matrices. *Annals of Probability* 21, 625–648.
- Bai, Z.D., 1993b. Convergence rate of expected spectral distributions of large random matrices. Part II. Sample covariance matrices. *Annals of Probability* 21, 649–672.
- Bai, Z.D., 1999. Methodologies in spectral analysis of large dimensional random matrices, a review. *Statistica Sinica* 9, 611–677.
- Bai, Z.D., Chen, J.Q., Yao, J.-F., 2010. On estimation of the population spectral distribution from a high-dimensional sample covariance matrix. *Australian and New Zealand Journal of Statistics* 52, 423–437.
- Bai, Z.D., Hu, J., Pan, G.M., Zhou, W., 2012. Convergence of the Empirical Spectral Distribution Function of Beta Matrices. Technical Report. arXiv:1208.5953.
- Bai, Z.D., Jiang, D., Yao, J.-F., Zheng, S., 2009. Corrections to LRT on large-dimensional covariance matrix by RMT. *Annals of Statistics* 37, 3822–3840.
- Bai, Z.D., Jiang, D., Yao, J.-F., Zheng, S., 2012. Testing Linear Hypothesis in High-Dimensional Regression. Technical Report. arXiv:1206.0867.
- Bai, Z.D., Liu, H.X., Wong, W.K., 2009. On the Markowitz mean-variance analysis of self-financing portfolios. *Risk and Decision Analysis* 1, 35–42.
- Bai, Z.D., Liu, H.X., Wong, W.K., 2011. Asymptotic properties of eigenmatrices of a large sample covariance matrix. *Annals of Applied Probability* 21, 1994–2014.
- Bai, Z.D., Miao, B.Q., Pan, G.M., 2007. On asymptotics of eigenvectors of large sample covariance matrix. *Annals of Statistics* 35, 1532–1572.
- Bai, Z.D., Miao, B.Q., Tsay, J., 1997. A note on the convergence rate of the spectral distributions of large dimensional random matrices. *Statistics and Probability Letters* 34, 95–102.
- Bai, Z.D., Miao, B.Q., Tsay, J., 1999. Remarks on the convergence rate of the spectral distributions of Wigner matrices. *Journal of Theoretical Probability* 12, 301–311.
- Bai, Z.D., Miao, B.Q., Tsay, J., 2002. Convergence rates of the spectral distributions of large Wigner matrices. *International Journal of Mathematics* 1, 65–90.
- Bai, Z.D., Miao, B.Q., Yao, J.F., 2003. Convergence rates of spectral distributions of large sample covariance matrices. *SIAM Journal of Matrix Analysis and Applications* 25, 105–127.
- Bai, Z.D., Saranadasa, H., 1996. Effect of high dimension, by an example of a two sample problem. *Statistica Sinica* 6, 311–329.
- Bai, Z.D., Silverstein, J.W., 1998. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Annals of Probability* 26, 316–345.
- Bai, Z.D., Silverstein, J.W., 1999. Exact separation of eigenvalues of large dimensional sample covariance matrices. *Annals of Probability* 27, 1536–1555.
- Bai, Z.D., Silverstein, J.W., 2004. CLT for linear spectral statistics of large dimensional sample covariance matrix. *Annals of Probability* 32, 553–605.
- Bai, Z.D., Silverstein, J.W., 2007. On the signal-to-interference ratio of CDMA systems in wireless communications. *Annals of Applied Probability* 17, 81–101.
- Bai, Z.D., Silverstein, J.W., 2009. *Spectral Analysis of Large Dimensional Random Matrices*. Springer.
- Bai, Z.D., Silverstein, J.W., 2012. No eigenvalues outside the support of the limiting spectral distribution of information-plus-noise type matrices. *Random Matrices: Theory and Applications* 1, 1150004.
- Bai, Z.D., Silverstein, J.W., Yin, Y.Q., 1988. A note on the largest eigenvalue of a large dimensional sample covariance matrix. *Journal of Multivariate Analysis* 26, 166–168.
- Bai, Z.D., Wang, X., Zhou, W., 2009. CLT for linear spectral statistics of Wigner matrices. *Electronic Journal of Probability* 14, 2391–2417.
- Bai, Z.D., Wang, X., Zhou, W., 2010. Functional CLT for sample covariance matrices. *Bernoulli* 16, 1086–1113.
- Bai, Z.D., Yao, J.-F., 2005. On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli* 11, 1059–1092.
- Bai, Z.D., Yao, J.-F., 2011. On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis* 106, 167–177.
- Bai, Z.D., Yin, Y.Q., 1988. Convergence to the semicircle law. *Annals of Probability* 16, 863–875.
- Bai, Z.D., Yin, Y.Q., 1988. Necessary and sufficient conditions for the almost sure convergence of the largest eigenvalue of Wigner matrices. *Annals of Probability* 16, 1729–1741.
- Bai, Z.D., Yin, Y.Q., 1993. Limit of the smallest eigenvalue of large dimensional covariance matrix. *Annals of Probability* 21, 1275–1294.
- Bai, Z.D., Zhang, L.X., 2006. Semicircle law for Hadamard products. *SIAM Journal of Matrix Analysis and Applications* 29, 473–495.
- Bai, Z.D., Zhou, W., 2008. Large sample covariance matrices without independence structures in columns. *Statistica Sinica* 18, 425–442.
- Baik, J., Ben Arous, G., Pécché, S., 2005. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability* 33, 1643–1697.
- Baik, J., Silverstein, J.W., 2006. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis* 97, 1382–1408.
- Baker, T.H., Forrester, P.J., Pearce, P.A., 1998. Random matrix ensembles with an effective extensive external charge. *Journal of Physics A: Mathematical and General* 31, 6087–6101.
- Bao, Z., 2012. Strong convergence of ESD for the generalized sample covariance matrices when $p/n \rightarrow 0$. *Statistics and Probability Letters* 82, 894–901.
- Bao, Z., Pan, G.M., Zhou, W., 2012. Tracy–Widom law for the extreme eigenvalues of sample correlation matrices. *Electronic Journal of Probability* 17 (88), 1–32.
- Ben Arous, G., Pécché, S., 2005. Universality of local eigenvalue statistics for some sample covariance matrices. *Communications in Pure and Applied Mathematics* 53, 1316–1357.
- Benaych-Georges, F., Guionnet, A., Maida, M., 2011. Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electronic Journal of Probability* 16, 1621–1662.
- Benaych-Georges, F., Guionnet, A., Maida, M., 2012. Large deviations of the extreme eigenvalues of random deformations of matrices. *Probability Theory and Related Fields* 154, 703–751.
- Benaych-Georges, F., Nadakuditi, R.R., 2011. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics* 227, 494–521.
- Bhattacharjee, M., Bose, A., 2013. Consistency of large dimensional sample covariance matrix under weak dependence. *Statistical Methodology*.
- Bianchi, P., Debbah, M., Maida, M., Najim, J., 2011. Performance of statistical tests for single source detection using random matrix theory. *IEEE Transactions on Information Theory* 57, 2400–2419.
- Bickel, P.J., Levina, E., 2004. Some theory for Fisher's linear discriminant function, “naive Bayes”, and some alternatives when there are many more variables than observations. *Bernoulli* 10, 989–1010.
- Bickel, P.J., Levina, E., 2008b. Covariance regularization by thresholding. *Annals of Statistics* 36, 2577–2604.
- Bickel, P.J., Levina, E., 2008a. Regularized estimation of large covariance matrices. *Annals of Statistics* 36, 199–227.
- Birnbaum, A., Johnstone, I.M., Nadler, B., Paul, D., 2013. Minimax bounds for sparse PCA with noisy high-dimensional data. *Annals of Statistics* 41, 1055–1084.
- Biroli, G., Bouchaud, J.P., Potters, M., 2007. On the top eigenvalue of heavy-tailed random matrices. *Europhysics Letters* 78, 10001.
- Bloemendal, A., Virág, B., 2010. Limits of Spiked Random Matrices I. Technical Report. arXiv:1011.1877v2.
- Bloemendal, A., Virág, B., 2011. Limits of Spiked Random Matrices II. Technical Report. arXiv:1109.1877v1.
- Bohigas, O., Giannoni, M.J., Schmit, C., 1984. Characterization of chaotic quantum spectra and universality of level fluctuation laws. *Physical Review Letters* 52, 1–4.
- Bose, A., Sen, A., 2008. Another look at the moment method for large dimensional random matrices. *Electronic Journal of Probability* 13, 588–628.
- Bouchaud, J.P., Mézard, M., Potters, M., 2003. *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*, 3rd ed. Cambridge University Press.
- Bouchaud, J.P., Potters, M., 2009. Financial Applications of Random Matrix Theory: A Short Review. Technical Report. arXiv:0910.1205v1.

- Bryc, W., Dembo, A., Jiang, T., 2006. Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Annals of Probability* 34, 1–38.
- Cai, T., Jiang, T., 2011. Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Annals of Statistics* 39, 1496–1525.
- Cai, T.T., Liu, W., 2011a. Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association* 106, 672–684.
- Cai, T.T., Liu, W., 2011b. A direct estimation approach to sparse linear discriminant analysis. *Journal of the American Statistical Association* 106, 1566–1577.
- Cai, T.T., Liu, W., Luo, X., 2011. A constrained l_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association* 106, 594–607.
- Cai, T.T., Ma, Z., Wu, Y., 2012. Sparse PCA: Optimal Rates and Adaptive Estimation. Technical Report. arXiv:1211.1309v2.
- Cai, T.T., Yuan, M., 2012. Adaptive covariance matrix estimation through block thresholding. *Annals of Statistics* 40, 2012–2014.
- Cai, T.T., Zhang, C.-H., Zhou, H.H., 2010. Optimal rates of convergence for covariance matrix estimation. *Annals of Statistics* 38, 2118–2144.
- Cai, T.T., Zhou, H.H., 2012a. Minimax estimation of large covariance matrices under l_1 norm. *Statistica Sinica* 22, 1319–1378.
- Capitaine, M., Donati-Martin, C., Féral, D., 2009. The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations. *Annals of Probability* 37, 1–47.
- Chapon, F., Couillet, R., Hachem, W., Mestre, X., 2012. On the Isolated Eigenvalues of Large Gram Random Matrices with a Fixed Rank Deformation. Technical Report. arXiv:1207.0471v1.
- Chatterjee, S., 2006. A generalization of the Lindeberg principle. *Annals of Probability* 34, 2061–2076.
- Chatterjee, S., 2009. Fluctuations of eigenvalues and second order Poincaré inequalities. *Probability Theory and Related Fields* 143, 1–40.
- Chen, J.Q., Delyon, B., Yao, J.-F., 2011. On a model selection problem from high-dimensional sample covariance matrices. *Journal of Multivariate Analysis* 102, 1388–1398.
- Chen, L., Paul, D., Prentice, R.L., Wang, P., 2011. A regularized Hotelling's T^2 test for pathway analysis in proteomic studies. *Journal of the American Statistical Association* 106, 1345–1360.
- Chen, S.X., Qin, Y.-L., 2010. A two-sample test for high-dimensional data with applications to gene-set testing. *Annals of Statistics* 38, 808–835.
- Choup, L.N., 2006. Edgeworth expansion of the largest eigenvalue distribution function of GUE and LUE. *International Mathematics Research Notices*, 61049.
- Clemmensen, L., Hastie, T., Witten, D., Ersboll, B., 2011. Sparse discriminant analysis. *Technometrics* 53, 406–413.
- Costin, O., Lebowitz, J., 1995. Gaussian fluctuations in random matrices. *Physical Review Letters* 75, 69–72.
- Couillet, R., Debbah, M., 2011. *Random Matrix Methods for Wireless Communications*. Cambridge University Press.
- Couillet, R., Pascal, F., Silverstein, J.W., 2012. Robust m-Estimation for Array Processing: A Random Matrix Approach. Technical Report. arXiv:1204.5320v1.
- Couillet, R., Silverstein, J.W., Bai, Z.D., Debbah, M., 2011. Eigen-inference for energy estimation of multiple sources. *IEEE Transactions on Information Theory* 57, 2420–2439.
- d'Aspremont, A., ElGhaoui, L., Jordan, M.I., Lanckriet, G.R.G., 2007. A direct formulation of sparse PCA using semidefinite programming. *SIAM Review* 49 (3), 434–448.
- Davis, R.A., Pfaffel, O., Stelzer, R., 2011. Limit Theory for the Largest Eigenvalues of Sample Covariance Matrices with Heavy-Tails. Technical Report. arXiv:1108.5464v1.
- Deift, P., 2000. Orthogonal Polynomials and Random Matrices: A Riemann–Hilbert Approach. American Mathematical Society.
- Deift, P., Gioev, D., 2009. *Random Matrix Theory: Invariant Ensembles and Universality*. American Mathematical Society.
- Dey, D.K., Srinivasan, C., 1985. Estimation of a covariance matrix under Stein's loss. *Annals of Statistics* 13, 1581–1591.
- Dey, D.K., Srinivasan, C., 1986. Trimmed minimax estimator of a covariance matrix. *Annals of the Institute of Statistical Mathematics* 38, 101–108.
- Diaconis, P., 2003. Patterns in eigenvalues: the 70th Josiah Willard Gibbs lecture. *Bulletin of the American Mathematical Society* 40, 155–178.
- Diaconis, P., Evans, S.N., 2001. Linear functionals of eigenvalues of random matrices. *Transactions of the American Mathematical Society* 353, 2615–2633.
- Diaconis, P., Holmes, S. (Eds.), 2004. *Stein's Method: Expository Lectures and Applications*. Institute of Mathematical Statistics.
- Ding, X., Jiang, T., 2010. Spectral distribution of adjacency and Laplacian matrices of random graphs. *Annals of Applied Probability* 20, 2086–2117.
- Dozier, R.B., Silverstein, J.W., 2007b. Analysis of the limiting spectral distribution of large dimensional information-plus-noise type matrices. *Journal of Multivariate Analysis* 98, 1099–1122.
- Dozier, R.B., Silverstein, J.W., 2007a. On the empirical distribution of eigenvalues of large dimensional information-plus-noise type matrices. *Journal of Multivariate Analysis* 98, 678–694.
- Dumitriu, I., 2003. *Eigenvalue Statistics for the Beta-Ensembles*. Ph.D. Thesis. Massachusetts Institute of Technology.
- Dumitriu, I., Edelman, A., 2002. Matrix models for beta ensembles. *Journal of Mathematical Physics* 43, 5830–5847.
- Dyson, F.J., 1962a. Statistical theory of energy levels of complex systems I. *Journal of Mathematical Physics* 3, 140–156.
- Dyson, F.J., 1962b. Statistical theory of energy levels of complex systems II. *Journal of Mathematical Physics* 3, 157–165.
- Dyson, F.J., 1962c. Statistical theory of energy levels of complex systems III. *Journal of Mathematical Physics* 3, 166–175.
- Edelman, A., Rao, N.R., 2005. Random matrix theory. *Acta Numerica* 14, 233–297.
- Edelman, A., Sutton, B.D., 2007. From random matrices to stochastic operators. *Journal of Statistical Physics* 127, 1121–1165.
- El Karoui, N., 2003. On the Largest Eigenvalue of Wishart Matrices with Identity Covariance When n , p and $p/n \rightarrow \infty$. Technical Report. arXiv:math/0309335v1.
- El Karoui, N., 2007. Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Annals of Probability* 35, 663–714.
- El Karoui, N., 2008a. Operator norm consistent estimation of large dimensional sparse covariance matrices. *Annals of Statistics* 36, 2717–2756.
- El Karoui, N., 2008b. Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Annals of Statistics* 36, 2757–2790.
- El Karoui, N., 2009a. Concentration of measure and spectra of random matrices: applications to correlation matrices, elliptical distributions and beyond. *Annals of Applied Probability* 19, 2362–2405.
- El Karoui, N., 2009. On the Realized Risk of High-Dimensional Markowitz Portfolios. Technical Report. University of California, Berkeley.
- El Karoui, N., 2010c. High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: risk underestimation. *Annals of Statistics* 38, 3487–3566.
- El Karoui, N., 2010b. On information plus noise kernel random matrices. *Annals of Statistics* 38, 3191–3216.
- El Karoui, N., 2010a. The spectrum of kernel random matrices. *Annals of Statistics* 38, 1–50.
- ElKaroui, N., Koesters, H., 2011. Geometric Sensitivity of Random Matrix Results: Consequences for Shrinkage Estimators of Covariance and Related Statistical Methods. Technical Report. arXiv:1105.1404v1.
- Erdős, L., Péché, S., Ramírez, J.A., Schlein, B., Yau, H.-T., 2010. Bulk universality for Wigner matrices. *Communications on Pure and Applied Mathematics* 63, 895–925.
- Erdős, L., Ramírez, J.A., Schlein, B., Tao, T., Vu, V., Yau, H.-T., 2010. Bulk universality for Wigner Hermitian matrices with subexponential decay. *Mathematics Research Letter* 17, 667–674.
- Erdős, L., Schlein, B., Yau, H.-T., 2009. Local semicircle law and complete delocalization for Wigner random matrices. *Communications in Mathematical Physics* 287, 641–655.
- Erdős, L., Yau, H.-T., 2012. Universality of local spectral statistics of random matrices. *Bulletin of the American Mathematical Society* 49, 377–414.
- Erdős, L., Yau, H.-T., Yin, J., 2012. Rigidity of eigenvalues of generalized Wigner matrices. *Advances in Mathematics* 229, 1435–1515.
- Fan, J., Fan, Y., 2008. High-dimensional classification using features annealed independence rules. *Annals of Statistics* 36, 2605–2637.
- Feldheim, O.N., Sodin, S., 2010. A universality result for the smallest eigenvalues of certain sample covariance matrices. *Geometric and Functional Analysis* 20, 88–123.
- Féral, D., Péché, S., 2007. The largest eigenvalue of rank one deformation of large Wigner matrices. *Communications in Mathematical Physics* 272, 185–228.

- Féral, D., Péch , S., 2009. The largest eigenvalues of sample covariance matrices for a spiked population: diagonal case. *Journal of Mathematical Physics* 50, 073302.
- Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic factor model: identification and estimation. *Review of Economics and Statistics* 111, 62–85.
- Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2004. The generalized dynamic factor model: consistency and rates. *Journal of Econometrics* 119, 231–255.
- Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2005. The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association* 100, 830–840.
- Forrester, P.J., 2010. *Log-Gases and Random Matrices*. Princeton University Press.
- Frahm, G., Jaekel, U., 2005. Random Matrix Theory and Robust Covariance Estimation for Financial Data. Technical Report. arXiv:physics/0503007.
- Friedman, J., Hastie, T., Tibshirani, R., 2008. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9, 432–441.
- Geman, S., 1980. A limit theorem for the norm of random matrices. *Annals of Probability* 8, 252–261.
- Geronimo, J.S., Hill, T.P., 2003. Necessary and sufficient condition that the limit of Stieltjes transforms is a Stieltjes transform. *Annals of Probability* 31, 54–60.
- Girko, V.L., 2001. *Theory of Stochastic Canonical Volumes I and II*. Kluwer Academic Publishers Group, Dordrecht.
- Gohberg, I., Krein, M., 1969. *Introduction to the Theory of Linear Nonselfadjoint Operators*. American Mathematical Society.
- G tze, F., Tikhomirov, A., 2003. Rate of convergence to the semi-circular law. *Probability Theory and Related Fields* 127, 228–276.
- G tze, F., Tikhomirov, A., 2004. Rate of convergence in probability to the Marchenko–Pastur law. *Bernoulli* 10, 503–548.
- Graybill, F.A., Johnson, D.E., 1972. An analysis of a two-way model with interaction and no replication. *Journal of the American Statistical Association* 67, 862–868.
- Guionnet, A., 2009. *Large Random Matrices: Lectures on Macroscopic Asymptotics*. Lecture Notes in Mathematics / cole d’ t  de Probabilit s de Saint-Flour. Cambridge University Press.
- Guionnet, A., Zeitouni, O., 2000. Concentration of the spectral measure for large matrices. *Electronic Communications in Probability* 5, 119–136.
- Gustavsson, J., 2005. Gaussian fluctuations of eigenvalues in the GUE. *Annales de l’Institut Henri Poincar  – Probabilit s et Statistiques* 41, 151–178.
- Hachem, W., Kharouf, M., Najim, J., Silverstein, J.W., 2012. A CLT for information-theoretic statistics of non-centered Gram random matrices. *Random Matrices, Theory and Applications* 1, 1150010.
- Hachem, W., Loubaton, P., Mestre, X., Najim, J., Vallet, P., 2012. Large information plus noise random matrix models and consistent subspace estimation in large sensor networks. *Random Matrices, Theory and Applications* 1, 1150006.
- Hachem, W., Loubaton, P., Mestre, X., Najim, J., Vallet, P., 2013. A subspace estimator for fixed rank perturbations of large random matrices. *Journal of Multivariate Analysis* 114, 427–447.
- Hachem, W., Loubaton, P., Najim, J., 2005. The empirical eigenvalue distribution of a Gram matrix: from independence to stationarity. *Markov Processes and Related Fields* 11, 629–648.
- Hachem, W., Loubaton, P., Najim, J., 2006. The empirical distribution of the eigenvalues of a Gram matrix with a given variance profile. *Annales de l’Institut Henri Poincar  – Probabilit s et Statistiques* 42, 649–670.
- Hachem, W., Loubaton, P., Najim, J., 2007. Deterministic equivalents for certain functionals of large random matrices. *Annals of Applied Probability* 17, 875–930.
- Hachem, W., Loubaton, P., Najim, J., 2008. A CLT for information-theoretic statistics of Gram random matrices with a given variance profile. *Annals of Applied Probability* 18, 2071–2130.
- Haff, L.R., 1977. Minimax estimation for a multinormal precision matrix. *Journal of Multivariate Analysis* 7, 374–385.
- Haff, L.R., 1980. Empirical Bayes estimation of the multivariate normal covariance matrix. *Annals of Statistics* 8, 586–597.
- Harding, M.C., 2008. Explaining the single factor bias of arbitrage pricing models in finite samples. *Economics Letters* 99, 85–88.
- Hardoon, D.R., Shawe-Taylor, J., 2009. Sparse Canonical Correlation Analysis. Technical Report. arXiv:0908.2724v1.
- Hiai, M., Petz, D., 2000. *The Semicircle Law, Free Random Variables, and Entropy*. American Mathematical Society.
- Hoyle, D.C., Rattray, M., 2004. Principal-component-analysis eigenvalue spectra from data with symmetry breaking structure. *Physical Review E* 69, 026124.
- Huber, P.J., 1964. Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35, 73–101.
- Jiang, T., 2004. The limiting distributions of eigenvalues of sample correlation matrices. *Sankhy * 66, 35–48.
- Jiang, T., 2009. Approximation of Haar distributed matrices and limiting distributions of eigenvalues of Jacobi ensembles. *Probability Theory and Related Fields* 144, 221–246.
- Jiang, T., 2012. Low eigenvalues of large random graphs. *Probability Theory and Related Fields* 153, 671–690.
- Jin, B., Wang, C., Miao, B., LoHuang, M.-N., 2009. Limiting spectral distribution of large-dimensional sample covariance matrices generated by VARMA. *Journal of Multivariate Analysis* 100, 2112–2125.
- Johansson, K., 1998. On fluctuations of eigenvalues of random Hermitian matrices. *Duke Mathematical Journal* 91, 151–204.
- Johansson, K., 2000. Shape fluctuations and random matrices. *Communications in Mathematical Physics* 209, 437–476.
- Johansson, K., 2001. Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices. *Communications in Mathematical Physics* 215, 683–705.
- Johansson, K., 2009. Universality of Certain Hermitian Wigner Matrices under Weak Moment Conditions. Technical Report. arXiv:0910.4467.
- Johnstone, I.M., 2007. High Dimensional Statistical Inference and Random Matrices.
- Johnstone, I.M., 2001. On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics* 29, 295–327.
- Johnstone, I.M., 2008. Multivariate analysis and Jacobi ensembles: largest eigenvalue, Tracy–Widom limits and rates of convergence. *Annals of Statistics* 36, 2638–2716.
- Johnstone, I.M., 2009. Approximate null distribution of the largest root in multivariate analysis. *Annals of Applied Statistics* 3, 1616–1633.
- Johnstone, I.M., 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., 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., Ma, Z., Perry, P.O., Shahram, M., 2009. RMTstat: distributions, statistics and tests derived from random matrix theory. R Package Version 0.2.
- Jolliffe, I., 2002. *Principal Component Analysis*. Springer, Berlin.
- Jonsson, D., 1982. Some limit theorems for the eigenvalues of a sample covariance matrix. *Journal of Multivariate Analysis* 12, 1–38.
- Karuthgamer, R., Nadler, B., Vilenchik, D., 2013. Do Semidefinite Relaxations Really Solve Sparse PCA. Technical Report. arXiv:1306.3690.
- Khorunzhy, A.M., Khoruzhenko, B.A., Pastur, L.A., 1996. Asymptotic properties of large random matrices with independent entries. *Journal of Mathematical Physics* 37, 5033–5060.
- Knowles, A., Yin, J., 2011. The Isotropic Semicircle Law and Deformation of Wigner Matrices. Technical Report. arXiv:1110.6449.
- Knowles, A., Yin, J., 2012. The Outliers of a Deformed Wigner Matrix. Technical Report. arXiv:1207.5619.
- Kritchman, S., Nadler, B., 2008. Determining the number of components in a factor model from limited noisy data. *Chemometrics and Intelligent Laboratory Systems* 94, 19–32.
- Lam, C., Fan, J., 2009. Sparsistency and rates of convergence in large covariance matrix estimation. *Annals of Statistics* 37, 4254–4278.
- Ledoit, O., P   , S., 2011. Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields* 151, 233–264.
- Ledoit, O., Wolf, M., 2012. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics* 40, 1024–1060.
- Ledoux, M., 2001. *The Concentration of Measure Phenomenon*. American Mathematical Society.
- Lee, J.O., Yin, J., 2012. A Necessary and Sufficient Condition for Edge Universality of Wigner Matrices. Technical Report. arXiv:1206.2251.

- Lee, W., Lee, D., Lee, Y., Pawitan, Y., 2009. Sparse canonical correlation analysis for high-throughput data. *Statistical Applications to Genetics and Molecular Biology* 10.
- Loh, W.-L., 1991. Estimating covariance matrices. *Annals of Statistics* 19, 283–296.
- Lopes, M.E., Jacob, L., Wainwright, M.J., 2012. A More Powerful Two-Sample Test in High Dimensions using Random Projection. Technical Report. arXiv:1108.2401v2.
- Lu, A.Y., 2002. Sparse Principal Components Analysis for Functional Data. Ph.D. Thesis. Stanford University.
- Lytova, A., Pastur, L., 2009. Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Annals of Probability* 37, 1778–1840.
- Ma, Z., 2012. Accuracy of the Tracy–Widom limits for the extreme eigenvalues in white Wishart matrices. *Bernoulli* 18, 322–359.
- Ma, Z., 2013. Sparse principal component analysis and iterative thresholding. *Annals of Statistics* 41, 772–801.
- Mardia, K.V., Kent, J.T., Bibby, J.M., 1980. *Multivariate Analysis*. Academic Press.
- Markowitz, H., 1952. Portfolio selection. *Journal of Finance* 7, 77–91.
- Markowitz, H., 1956. The optimization of a quadratic function subject to linear constraints. *Naval Research Logistics Quarterly* 3, 111–133.
- Maronna, R.A., 1976. Robust m -estimators of multivariate location and scatter. *Annals of Statistics* 4, 51–76.
- Marčenko, V., Pastur, L., 1967. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik* 1, 457–483.
- Mehta, M.L., 2004. *Random Matrices*, 3rd ed. Elsevier.
- Meinshausen, N., Bühlmann, P., 2006. High dimensional graphs and variable selection with the lasso. *Annals of Statistics* 34, 1436–1462.
- Mestre, X., 2008. Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Transactions on Information Theory* 54, 5113–5129.
- Mingo, J.A., Speicher, R., 2006. Second order freeness and fluctuations of random matrices. I. Gaussian and Wishart matrices and cyclic Fock spaces. *Journal of Functional Analysis* 235, 226–270.
- Muirhead, R.J., 1982. *Aspects of Multivariate Statistical Theory*. Wiley, New York.
- Nadakuditi, R.R., Newman, M.E.J., 2012. Graph spectra and detectability of community structure in networks. *Physical Review Letters* 108, 188701.
- Nadakuditi, R.R., Silverstein, J.W., 2010. Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples. *IEEE Journal of Selected Topics in Signal Processing* 4, 468–480.
- Nadler, B., 2008. Finite sample approximation results for principal component analysis: a matrix perturbation approach. *Annals of Statistics* 36, 2791–2817.
- Nadler, B., 2011. On the distribution of the ratio of the largest eigenvalue to the trace of a Wishart matrix. *Journal of Multivariate Analysis* 102, 363–371.
- Nadler, B., Johnstone, I.M., 2011. On the Distribution of Roy's Largest Root Test in MANOVA and in Signal Detection in Noise. Technical Report. Stanford University.
- Nadler, B., Penna, B., Garello, R., 2008. Performance of eigenvalue-based signal detectors with known and unknown noise level. In: *IEEE International Conference on Communications ICC2011*.
- Nguyen, H.H., Vu, V., 2012. Random Matrices: Law of the Determinant. Technical Report. arXiv:1112.0752v2.
- Nica, A., Speicher, R., 2006. *Lectures on the Combinatorics of Free Probability*. London Mathematical Society Lecture Note Series. Cambridge University Press.
- Olver, F.W.J., 1974. *Asymptotics and Special Functions*. Academic Press.
- Onatski, A., 2009. Testing hypotheses about the number of factors in large factor models. *Econometrica* 77, 1447–1479.
- Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics* 92, 1004–1016.
- Onatski, A., 2012. Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics* 168, 244–258.
- O'Rourke, S., 2010. Gaussian fluctuations of eigenvalues in Wigner random matrices. *Journal of Statistical Physics* 138, 1045–1066.
- O'Rourke, S., Renfrew, D., Soshnikov, A., 2011. Fluctuations of Matrix Entries of Regular Functions of Sample Covariance Random Matrices. Technical Report. arXiv:1106.0320v1.
- O'Rourke, S., Renfrew, D., Soshnikov, A., 2011. On Fluctuations of Matrix Entries of Regular Functions of Wigner Matrices with Non-identically Distributed Entries. Technical Report. arXiv:1104.1663v2.
- Pan, G.M., Gao, J., 2009. Asymptotic Theory for Sample Covariance Matrices under Cross-Sectional Dependence. Technical Report. Nanyang Technological University, Singapore.
- Pan, G.M., Zhou, W., 2011. Central limit theorem for Hotelling's T^2 statistic under large dimension. *Annals of Applied Probability* 21, 1860–1910.
- Parkhomenko, E., Tritchler, D., Beyene, J., 2009. Sparse canonical correlation analysis with application to genomic data integration. *Statistical Applications to Genetics and Molecular Biology* 8.
- Pastur, L., Lytova, A., 2011. Non-Gaussian Limiting Laws for the Entries of Regular Functions of the Wigner Matrices. Technical Report. arXiv:1103.2345.
- Pastur, L., Shcherbina, M., 2011. Eigenvalue Distribution of Large Random Matrices. American Mathematical Society.
- Patterson, N., Price, A.L., Reich, D., 2006. Population structure and eigenanalysis. *PLoS Genetics* 2, 2074–2093.
- Paul, D., 2007. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica* 17, 1617–1642.
- Paul, D., 2011. Asymptotic distribution of the smallest eigenvalue of $\text{Wishart}(N, n)$ when $N, n \rightarrow \infty$ such that $N/n \rightarrow 0$. In: Jiang, J., Roussas, G.G., Samaniego, F.J. (Eds.), *Nonparametric Statistical Methods and Related Topics: A Festschrift in Honor of Professor P K Bhattacharya on the Occasion of His 80th Birthday*. World Scientific, pp. 423–458.
- Paul, D., Johnstone, I.M., 2012. Augmented Sparse Principal Component Analysis for High Dimensional Data. Technical Report.
- Paul, D., Silverstein, J.W., 2009. No eigenvalues outside the limiting empirical spectral distribution of a separable covariance matrix. *Journal of Multivariate Analysis* 100, 37–57.
- Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* 2 (1), 559–572.
- Péché, S., 2009. Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probability Theory and Related Fields* 143, 481–516.
- Péché, S., Soshnikov, A., 2008. On the lower bound of the spectral norm of symmetric random matrices with independent entries. *Electronic Communications in Probability* 13, 280–290.
- Peng, P., Wang, P., Zhou, N., Zhu, J., 2009. Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association* 104, 735–746.
- Pfaffel, O., Schlemm, E., 2012. Eigenvalue Distribution of Large Sample Covariance Matrices of Linear Processes. Technical Report. arXiv:1201.3828.
- Pillai, N., Jin, J., 2011. Universality of Covariance Matrices. Technical Report. arXiv:1110.2501.
- Pillai, N., Jin, J., 2012. Edge universality of correlation matrices. *Annals of Statistics* 40, 1737–1763.
- Pizzo, A., Renfrew, D., Soshnikov, A., 2012. Fluctuations of Matrix Entries of Regular Functions of Wigner Matrices.
- Pizzo, A., Renfrew, D., Soshnikov, A., 2013. On finite rank deformations of Wigner matrices. *Annales de l'Institut Henri Poincaré—Probabilités et Statistiques* 49, 64–94.
- Ramírez, J.A., Rider, B., Virág, B., 2011. Beta ensembles, stochastic Airy spectrum, and a diffusion. *Journal of the American Mathematical Society* 24, 919–944.
- Rao, N.R., Mingo, J.A., Speicher, R., Edelman, A., 2008. Statistical eigen-inference from large Wishart matrices. *Annals of Statistics* 36, 2850–2885.
- Reed, M., Simon, B., 1972. *Methods of Modern Mathematical Physics*. Academic Press.
- Reimann, R., Van den Broeck, C., Bex, G.J., 1996. A gaussian scenario for unsupervised learning. *Journal of Physics A* 29, 3521–3535.
- Renfrew, D., Soshnikov, A., 2013. On finite rank deformations of Wigner matrices II: delocalized perturbations. *Random Matrices: Theory and Applications* 2, 1250015.
- Rohe, K., Chatterjee, S., Yu, B., 2011. Spectral clustering and high-dimensional stochastic blockmodel. *Annals of Statistics* 39, 1878–1915.

- Rothman, A.J., Bickel, P.J., Levina, E., Zhu, J., 2008. Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics* 2, 494–515.
- Rothman, A.J., Levina, E., Zhu, J., 2009. Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association* 104, 177–186.
- Roy, S.N., 1953. On a heuristic method of test construction and its use in multivariate analysis. *Annals of Mathematical Statistics* 24, 220–238.
- Shao, J., Wang, Y., Deng, X., Wang, S., 2011. Sparse linear discriminant analysis by thresholding for high-dimensional data. *Annals of Statistics* 39, 1241–1265.
- Shcherbina, M., 2011. Central Limit Theorem for Linear Eigenvalue Statistics of Wigner and Sample Covariance Random Matrices.
- Silverstein, J.W., 1984. Some limit theorems on the eigenvectors of large dimensional sample covariance matrices. *Journal of Multivariate Analysis* 15, 295–324.
- Silverstein, J.W., 1985. The limiting eigenvalue distribution of a multivariate F matrix. *SIAM Journal of Mathematical Analysis* 16, 641–646.
- Silverstein, J.W., 1989. On the eigenvectors of large-dimensional sample covariance matrices. *Journal of Multivariate Analysis* 15, 295–324.
- Silverstein, J.W., 1990. Weak convergence of random functions defined by the eigenvectors of sample covariance matrices. *Annals of Probability* 18, 1174–1193.
- Silverstein, J.W., Bai, Z.D., 1995. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate Analysis* 54, 175–192.
- Silverstein, J.W., Combettes, P.L., 1992. Signal detection via spectral theory of large dimensional random matrices. *IEEE Transactions on Signal Processing* 40, 2100–2104.
- Sinai, Y., Soshnikov, A., 1998. Central limit theorem for traces of large random symmetric matrices with independent matrix elements. *Boletim da Sociedade Brasileira de Matemática (N.S.)* 29, 1–24.
- Sodin, S., 2007. Random matrices, nonbacktracking walks, and orthogonal polynomials. *Journal of Mathematical Physics* 48, 123503.
- Sodin, S., 2009. Tracy–Widom law for some sparse random matrices. *Journal of Statistical Physics* 136, 834–841.
- Sodin, S., 2010. The spectral edge of some random band matrices. *Annals of Mathematics* 172, 2223–2251.
- Soshnikov, A., 1999. Universality at the edge of the spectrum in Wigner random matrices. *Communications in Mathematical Physics* 207, 697–733.
- Soshnikov, A., 2000. Determinantal random point fields. *Russian Mathematical Surveys* 55, 923–975.
- Soshnikov, A., 2002b. Gaussian limit for determinantal random point fields. *Annals of Probability* 30, 171–187.
- Soshnikov, A., 2002a. A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *Journal of Statistical Physics* 108, 1033–1056.
- Soshnikov, A., 2004. Poisson statistics for the largest eigenvalue of Wigner random matrices with heavy tails. *Electronic Communications in Probability* 9, 82–91.
- Soshnikov, A., 2006. Poisson statistics for the largest eigenvalues in random matrix ensembles. In: *Mathematical Physics of Quantum Mechanics, Lecture Notes in Physics*, vol. 690. Springer, Berlin, pp. 351–364.
- Srivastava, M.S., Du, M., 2008. A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis* 99, 386–402.
- Stein, C., 1956. Some Problems in Multivariate Analysis I. Technical Report. University of California, Berkeley.
- Stein, C., 1976. Estimation of a Covariance Matrix. Technical Report, Rietz Lecture, 39th Annual IMS Meeting. Atlanta, Georgia.
- Sutton, B.D., 2005. The Stochastic Operator Approach to Random Matrix Theory. Ph.D. Thesis. Massachusetts Institute of Technology.
- Tao, T., 2012. Topics in Random Matrix Theory. American Mathematical Society.
- Tao, T., Vu, V., 2010a. Random matrices: the distribution of the smallest singular values. *Geometric Functional Analysis* 20, 260–297.
- Tao, T., Vu, V., 2010b. Random matrices: universality of local eigenvalue statistics up to the edge. *Communications in Mathematical Physics* 298, 549–572.
- Tao, T., Vu, V., 2011. Random matrices: universality of local eigenvalue statistics. *Acta Mathematica* 206, 127–204.
- Tao, T., Vu, V., 2012c. A central limit theorem for the determinant of a Wigner matrix. *Advances in Mathematics* 231, 74–101.
- Tao, T., Vu, V., 2012a. Random covariance matrices: universality of local statistics of eigenvalues. *Annals of Probability* 40, 1285–1315.
- Tao, T., Vu, V., 2012. Random Matrices: Sharp Concentration of Eigenvalues. Technical Report. arXiv:1201.4789v3.
- Tipping, M.E., Bishop, C.M., 1999. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B* 61.
- Tracy, C., Widom, H., 1994a. Fredholm determinants, differential equations and matrix models. *Communications in Mathematical Physics* 163, 33–72.
- Tracy, C., Widom, H., 1994b. Level spacing distributions and the airy kernel. *Communications in Mathematical Physics* 159, 151–174.
- Tracy, C., Widom, H., 1996. On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics* 177, 727–754.
- Tulino, A.M., Verdú, S., 2011. *Random Matrix Methods and Wireless Communications*. Now Publishers Inc.
- Vershynin, R., 2011. Spectral norm of products of random and deterministic matrices. *Probability Theory and Related Fields* 150, 471–509.
- Vershynin, R., 2012. Introduction to the non-asymptotic analysis of random matrices. In: Eldar, Y., Kutyniok, G. (Eds.), *Compressed Sensing, Theory and Applications*. Cambridge University Press, pp. 210–268.
- Wachter, K.W., 1978. The strong limits of random matrix spectra for sample matrices of independent elements. *Annals of Probability* 6, 1–18.
- Wachter, K.W., 1980. The limiting empirical measure of multiple discriminant ratios. *Annals of Statistics* 8, 937–957.
- Wang, C., Pan, G.M., Cao, L., 2012. A Shrinkage Estimation of Large Dimensional Precision Matrices using Random Matrix Theory. Technical Report. arXiv:1211.2400v1.
- Wang, K., 2012. Random covariance matrices: universality of local statistics of eigenvalues up to the edge. *Random Matrices: Theory and Applications* 1, 1150005.
- Watkin, T.H., Nadal, J.-P., 1994. Optimal unsupervised learning. *Journal of Physics A* 27, 1899–1915.
- Widom, H., 1999. On the relation between orthogonal, symplectic and unitary ensembles. *Journal of Statistical Physics* 94, 347–363.
- 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 generalised product moment distribution in samples from a normal multivariate population. *Biometrika* 20A, 32–52.
- Witten, D.M., Tibshirani, R., 2009. Covariance-regularized regression and classification for high-dimensional problems. *Journal of the Royal Statistical Society, Series B* 71, 615–636.
- Witten, D.M., Tibshirani, R., 2011. Penalized classification using Fisher's linear discriminant. *Journal of the Royal Statistical Society, Series B* 73, 753–772.
- Witten, D.M., Tibshirani, R., Hastie, T., 2009. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* 10, 515–534.
- Wu, W.B., Pourahmadi, M., 2003. Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* 90, 831–844.
- Wu, W.B., Pourahmadi, M., 2009. Banding sample covariance matrices of stationary processes. *Statistica Sinica* 19, 1755–1768.
- Yang, Y., Pan, G.M., 2012. The convergence of the empirical distribution of canonical correlation coefficients. *Electronic Journal of Probability* 17, 1–13.
- Yao, J.-F., 2012. A note on a Marcenko–Pastur type theorem for time series. *Statistics & Probability Letters* 82, 22–28.
- Yin, Y.Q., 1986. Limiting spectral distribution for a class of random matrices. *Journal of Multivariate Analysis* 20, 50–68.
- Yin, Y.Q., Bai, Z.D., Krishnaiah, P.R., 1983. Limiting behavior of the eigenvalues of a multivariate F matrix. *Journal of Multivariate Analysis* 13, 508–516.
- Yin, Y.Q., Bai, Z.D., 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.
- Yin, Y.Q., Krishnaiah, P.R., 1983. A limit theorem for the eigenvalues of product of two random matrices. *Journal of Multivariate Analysis* 13, 489–507.
- Yuan, M., Lin, Y., 2007. Model selection and estimation in the Gaussian graphical model. *Biometrika* 94, 19–35.
- Zhang, L., 2006. Spectral Analysis of Large Dimensional Random Matrices. Ph.D. Thesis. National University of Singapore.
- Zheng, S., 2012. Central limit theorems for linear spectral statistics of large dimensional F -matrices. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques* 48, 444–476.
- Zheng, X., Li, Y., 2011. On the estimation of integrated covariance matrices of high dimensional diffusion processes. *Annals of Statistics* 39, 3121–3151.
- Zou, H., Hastie, T., Tibshirani, R., 2006. Sparse principal component analysis. *Journal of Computational and Graphical Statistics* 15, 265–286.