# BULK SPECTRUM: WIGNER AND MARCHENKO-PASTUR THEOREMS

## 1. Gaussian Orthogonal and Unitary Ensembles

**Definition 1.** A random symmetric, $N \times N$ matrix $X$ chosen from the *Gaussian orthogonal ensemble* (abbreviated *GOE*) is a symmetric matrix whose entries $X_{i,i}$ and $X_{i,j} = X_{j,i}$ (for $j > i$) are independent real random variables, with distributions

(1) $$X_{i,i} \sim N(0,1) \quad \text{and} \quad X_{i,j} \sim N(0,1/2)\ \forall\ j > i.$$

The term *orthogonal* derives from the fact that the distribution of the random matrix $X$ is invariant under orthogonal transformations of $\mathbb{R}^N$, that is, if $U$ is an orthogonal matrix then $U^T X U$ has the same distribution as $X$. This is a consequence of the trace rule $\text{Tr}\,(AB) = \text{Tr}\,(BA)$ and the definition $U^T U = U U^T$ of an orthogonal matrix, together with the following expression for the density of the GOE:

**Proposition 1.** *If $X$ is chosen from the Gaussian orthogonal ensemble, then the joint probability density of the entries $X_{i,i}$ and $X_{i,j}$ relative to the Lebesgue measure $\prod_i dx_{i,i} \prod_{i<j} dx_{i,j}$ is*

$$e^{-\text{Tr}\,(X^2)/2}/C_N$$

*where* $C_N = (2\pi)^{N/2}\pi^{\binom{N}{2}/2}$.

**Exercise 1.** $^{**}$ Show that if $Y$ is a random symmetric matrix with real entries $Y_{i,j} = Y_{j,i}$ whose distribution is invariant under orthogonal transformations and whose entries $Y_{i,i}$ and $Y_{i,j}$ are independent with marginal distributions $F_{\text{diag}}$ and $F_{\text{upper}}$, respectively, then the joint density of the entries is

$$\exp\{-a\text{Tr}\,(Y^2) - b\text{Tr}\,(Y)\}/C_{a,b,N}.$$

The study of the Gaussian orthogonal ensemble was begun by E. Wigner, a physicist who used the random matrix $X$ as a model for the Hamiltonian of a large atomic nucleus. Physical considerations dictated that the distribution should be invariant under orthogonal transformations. Random matrices with complex entries are also of physical interest; these should have distributions that are invariant by *unitary* transformations.

**Definition 2.** A complex random variable $Z = X + iY$ is said to have the complex normal distribution with mean $\mu$ and variance $\sigma^2 > 0$, denoted by $N_{\mathbb{C}}(\mu, \sigma^2)$, if its real and imaginary parts $X, Y$ are independent real normal random variables, and its first two moments are

(2) $$EZ := EX + iEY = \mu \quad \text{and} \quad \text{var}(Z) := E(Z-\mu)(Z-\mu)^* = \sigma^2.$$

**Notation:** The complex conjugate of a complex number $z$ will be denoted by either $\bar{z}$ or $z^*$. The adjoint (conjugate transpose) of a square matrix $X$ will be denoted by $X^*$. (Recall that a matrix $X$ is *Hermitian* if $X^* = X$.)

**Exercise 2.** If $Z \sim N_{\mathbb{C}}(0,1)$ then

(A) $e^{i\theta} Z \sim N_{\mathbb{C}}(0,1)$ for every $\theta \in \mathbb{R}$;
(B) $\mathscr{D}(\arccos(X/|Z|\,|\,|Z|)) \sim \text{Uniform-}[-\pi,\pi]$; and

(C) $EZ^m \bar{Z}^n = 0$ if $m \neq n$.

**Definition 3.** A random $N \times N$ matrix $Z$ chosen from the *Gaussian unitary ensemble* is a Hermitian matrix whose entries $Z_{i,i}$ and $Z_{i,j}$ (for $j > i$) are independent, with distributions

$$Z_{i,i} \sim N(0,1) \quad \text{and} \quad Z_{i,j} \sim N_{\mathbb{C}}(0,1). \tag{3}$$

**Proposition 2.** *If $Z$ is chosen from the Gaussian unitary ensemble, then the joint probability density of the entries $Z_{i,i}$, $X_{i,j} := Re(Z_{i,j})$, and $Y_{i,j} := Re(Z_{i,j})$ relative to the Lebesgue measure $\prod_i dz_{i,i} \prod_{i<j} dx_{i,j} dy_{i,j}$ is*

$$e^{-\text{Tr}\,(X^2)/2} / D_N$$

*where $D_N = (2\pi)^{N/2}(2\pi)^{\binom{N}{2}}$. Consequently, the distribution of $Z$ is invariant under unitary transformations, that is, if $U$ is unitary, then $U^* ZU \sim Z$.*

Since a random matrix $X$ chosen from either the Gaussian Orthogonal Ensemble or the Gaussian Unitary Ensemble is necessarily Hermitian, it is diagonalizable, with *real* eigenvalues $\lambda_i$. (This is the content of the *Spectral Theorem*.) Hence, the trace of $X$ is the sum $\sum_i \lambda_i$ of the eigenvalues, and so Propositions 1–2 imply that the joint densities of the entries are symmetric functions of the eigenvalues.

Wigner's main result concerned the asymptotic behavior of the *empirical spectral distribution* of a random matrix chosen from the Gaussian Orthogonal Ensemble or the Gaussian Unitary Ensemble. (The *empirical spectral distribution* $F^M$ of a diagonizable matrix $M$ is defined to be the uniform distribution on the set $\{\lambda_i\}_{1 \le i \le N}$ of eigenvalues [listed according to multiplicity].) The result holds more generally – and is no harder to prove – for random Hermitian matrices with independent entries:

**Theorem 3.** *(Wigner) Let $G, H$ be probability distributions on $\mathbb{R}$ and $\mathbb{C}$, respectively, such that $H$ has second moment $1$, that is,*

$$\int_{z \in \mathbb{C}} |z|^2 \, dH(z) = 1. \tag{4}$$

*Let $X$ be a random $N \times N$ Hermitian matrix whose entries $X_{i,i}$ and $X_{i,j} = X^*_{j,i}$, for $j > i$, are independent, with diagonal entries $X_{i,i}$ distributed according to $G$ and off-diagonal entries $X_{i,j}$ distributed according to $H$. Define*

$$M = X / \sqrt{N}.$$

*Then as $N \to \infty$ the empirical spectral distribution $F^M$ of $M$ converges weakly to the* semi-circle law *with density*

$$p(t) = \frac{\sqrt{4-t^2}}{2\pi} \mathbf{1}_{[-2,2]}(t). \tag{5}$$

*More precisely, if $\{\lambda_i\}_{1 \le i \le N}$ are the eigenvalues of $M$, then for any bounded, continuous function $f : \mathbb{R} \to \mathbb{R}$ and any $\varepsilon > 0$,*

$$\lim_{N \to \infty} P\{|N^{-1} \sum_{i=1}^{N} f(\lambda_i) - \int f(t) p(t) \, d\tau| \ge \varepsilon\} = 0 \tag{6}$$

The proof will be broken into two distinct parts: first, in sections 4–5 we will show that the theorem is true under the additional hypotheses that the random variables $X_{i,i}$ and $X_{i,j}$ have expectation zero and finite moments of all orders; second, in section 7 we will show that these hypotheses are extraneous. For ease of exposition we will consider only the case where the random variables $X_{i,j}$ are real-valued; the complex-valued case is really no different, but some additional accounting for imaginary parts is necessary. It might at first seem strange that the theorem holds without the hypothesis that the random variables $X_{i,i}$ and $X_{i,j}$ have mean zero. The explanation, as we will show in section 7, is that adding a constant to every entry of the matrix $X$ doesn't change the empirical spectral distribution very much, even though it might change one eigenvalue by a large amount.

## 2. Sample Covariance Matrices

**Definition 4.** Let $\Sigma$ be a positive definite, symmetric $p \times p$ matrix, and let $X$ be a random $p \times n$ matrix whose columns are independent, identically distributed random vectors each with the multivariate Gaussian distribution $N(0, \Sigma)$. Assume that $n \geq p$. Then the random $p \times p$ matrix

$$S = XX^T \tag{7}$$

has the *Wishart distribution* $W(\Sigma, n)$.

**Exercise 3.** The *sample covariance matrix* of a sample $Y_1, Y_2, \dots, Y_N$ of independent, identically distributed random (column) vectors is usually defined by

$$S = (N-1)^{-1} \sum_{j=1}^{N} (Y_j - \bar{Y})(Y_j - \bar{Y})^T.$$

Show that if the random vectors $Y_j$ have the multivariate normal distribution $N(\mu, \Sigma)$ then $S$ has the Wishart distribution $W((N-1)^{-1}\Sigma, N-1)$. Note: Here $\bar{Y}$ denotes the sample mean $\bar{Y} = N^{-1}\sum_j Y_j$. (This notation will not be used elsewhere in the notes.)

**Proposition 4.** *If $S \sim W(\Sigma, n)$ then the joint density of the entries $S_{i,i}, S_{i,j}$ with $j > i$ relative to Lebesgue measure is*

$$f_{\Sigma,n}(S) = |\det(S)|^{(n-p-1)/2} \exp\{-\mathrm{Tr}\,(\Sigma^{-1}S)/2\}/C_{\Sigma,n} \tag{8}$$

*for an appropriate normalizing constant $C_{\Sigma,n}$:*

$$C_{\Sigma,n} = \sqrt{2}^{np} \sqrt{\pi}^{\binom{p}{2}} \det(\Sigma)^{n/2} \prod_{i=1}^{p} \Gamma((n+1-i)/2) \tag{9}$$

The proof will be given later in the course. Note the formal similarity with the $\chi^2$ density. If $\Sigma = I$, the exponential factor reduces to $e^{-\sum \lambda_i^2/2}$, where $\lambda_i$ are the eigenvalues of $S$, and the determinant is $\det(S) = \prod \lambda_i$, so the density is a symmetric function of the eigenvalues.

**Theorem 5.** *(Marchenko & Pastur) Assume that $S \sim W(I/n, n)$ where $I$ is the $p \times p$ identity matrix. If $p, n \to \infty$ in such a way that $p/n \to y \in (0,1)$ then the empirical spectral distribution $F^S$ converges weakly to the* Marchenko-Pastur distribution *with density*

$$f_y(x) = \frac{\sqrt{(b-x)(x-a)}}{2\pi xy} \mathbf{1}_{[a,b]}(x) \tag{10}$$

*where*

$$a = a_y = (1-\sqrt{y})^2 \quad \textit{and}$$
$$b = b_y = (1+\sqrt{y})^2.$$

*More generally, let $X$ be a random $p \times n$ matrix whose entries $X_{i,j}$ are independent, identically distributed random variables with distribution $G$. Assume that $G$ has mean zero and variance* 1*. Define*

$$S = XX^T/n. \tag{11}$$

*Then as $p, n \to \infty$ in such a way that $p/n \to y \in (0,1)$, the empirical spectral distribution $F^S$ converges weakly to the Marchenko-Pastur distribution with density $f_y$ defined by* (10)*.*

## 3. Method of Moments; Szego's Theorem

3.1. **Method of Moments.** The proofs of Theorems 3 and 5 will be based on the *method of moments* (see Background Notes). The strategy will be to show that for each integer $k = 1, 2, \dots$, the $k$th moment of the empirical spectral distribution converges (in probability) to the $k$th moment of the corresponding limit law (the semi-circle law in Theorem 3, the Marchenko-Pastur law in Theorem 5). The key to this strategy is that moments of the empirical spectral distribution can be represented as matrix traces: in particular, if $M$ is an $N \times N$ diagonizable matrix with empirical spectral distribution $F^M$, then

$$\boxed{\int \lambda^k F^M(d\lambda) = N^{-1} \mathrm{Tr}\, M^k}. \tag{12}$$

The trace $\mathrm{Tr}\, M^k$ has, in turn, a combinatorial interpretation as a sum over *closed paths* of length $k$ in the *augmented* complete graph $K_N^+$. (A *closed path* is a path in a graph that begins and ends at the same vertex. The *complete graph* $K_N$ on $N$ vertices is the graph with vertex set $[N] = \{1, 2, 3, \dots, N\}$ and an edge connecting every pair of distinct vertices. The *augmented complete graph* $K_N^+$ has the same vertices and edges as the complete graph $K_N$, and in addition a loop connectng each vertex to itself. Thus, a path in $K_N^+$ is an arbitrary sequence of vertices $i_1 i_2 \cdots i_m$, whereas a path in $K_N$ is a sequence of vertices in which no vertex appears twice consecutively.) Denote by $\mathscr{P}_i^k$ the set of all length$-k$ closed paths

$$\gamma = i = i_0, i_1, i_2, \dots, i_{k-1}, i_k = i$$

in $K_N^+$ that begin and end at vertex $i$, and for each such path $\gamma$ define the *weight*

$$w(\gamma) = w_M(\gamma) = \prod_{j=0}^{k-1} M_{i_j, i_{j+1}}. \tag{13}$$

Then

$$\boxed{\mathrm{Tr}\, M^k = \sum_{i=1}^{N} \sum_{\gamma \in \mathscr{P}_i^k} w(\gamma)}. \tag{14}$$

3.2. **Szego's Theorem.** To illustrate how the method of moments works in a simple case, we prove a classical theorem of Szego regarding the empirical spectral distribution of a *Toeplitz matrix*. A Toeplitz matrix is a square matrix whose entries are constant down diagonals:

$$T = T_N = \begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_{N-1} & a_N \\ a_{-1} & a_0 & a_1 & \cdots & a_{N-2} & a_{N-1} \\ a_{-2} & a_{-1} & a_0 & \cdots & a_{N-3} & a_{N-2} \\ & & & \cdots & & \\ a_{-N} & a_{-N+1} & a_{-N+2} & \cdots & a_{-1} & a_0 \end{pmatrix}. \tag{15}$$

The Fourier series $A(\theta) := \sum_{n=-\infty}^{\infty} a_n e^{in\theta}$ whose coefficients are the entries of $T$ is sometimes called the *symbol* of the infinite Toeplitz matrix $T_\infty$. The *spectral distribution* is the distribution of $A(\Theta)$, where $\Theta$ is a random variable uniformly distributed on the interval $[-\pi,\pi]$.

**Theorem 6.** *(Szego) Assume that only finitely many of the coefficients $a_n$ are nonzero, that is, the symbol $A(\theta)$ is a trig polynomial. Then the empirical spectral distribution $F_N$ of the Toeplitz matrix $T_N$ converges as $N\to\infty$ to the spectral distribution, that is, for every bounded continuous function $f:\mathbb{C}\to\mathbb{R}$,*

$$\lim_{N\to\infty}\int f(t)\,F_N(dt) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(A(\theta))\,d\theta. \tag{16}$$

*Proof.* The $k$th moment of the empirical spectral distribution $F_N$ is $N^{-1}\times$the trace of $T_N^k$, by (12), and by (14), the trace $\mathrm{Tr}\ T_N^k$ is the sum of the $T-$weights of all length-$k$ closed paths. By hypothesis, there exists $M<\infty$ such that $a_n = 0$ if $|n| > M$. Thus, only closed paths in which the steps are of size $\le M$ in magnitude will have nonzero weights. Moreover, because $T_N$ is Toeplitz, for each $M < i < N-M$,

$$\sum_{\gamma\in\mathscr{P}_i^k} w(\gamma) = (a * a * \cdots * a)_0 = (2\pi)^{-1}\int_{-\pi}^{\pi} A(\theta)^k\,d\theta$$

where $*$ denotes sequence convolution. Thus, as $N\to\infty$,

$$N^{-1}\mathrm{Tr}\ T_N^k \longrightarrow (2\pi)^{-1}\int_{-\pi}^{\pi} A(\theta)^k\,d\theta.$$

This proves that the moments of the empirical spectral distribution converge to the corresponding moments of the spectral distribution. Since the spectral distribution is uniquely determined by its moments (this follows because it has bounded support), the result follows. □

**Exercise 4.** Show that Szego's Theorem is true under the weaker hypothesis that $\sum_{-\infty}^{\infty}|a_n| < \infty$.

**Exercise 5.** Let $M_N$ be the random symmetric tridiagonal matrix

$$M_N = \begin{pmatrix} X_{1,1} & X_{1,2} & 0 & 0 & \cdots & 0 & 0 \\ X_{1,2} & X_{2,2} & X_{2,3} & 0 & \cdots & 0 & 0 \\ 0 & X_{2,3} & X_{3,3} & X_{3,4} & \cdots & 0 & 0 \\ & & & & \cdots & & \\ 0 & 0 & 0 & 0 & \cdots & X_{N-1,N-1} & X_{N-1,N} \\ 0 & 0 & 0 & 0 & \cdots & X_{N-1,N} & X_{N,N} \end{pmatrix}$$

whose entries $X_{i,j}$, where $j=i$ or $j=i+1$, are independent, identically distributed $N(0,1)$. (Note: Since $M_N$ is symmetric, the eigenvalues are real.)

(A) * Prove that as $N\to\infty$ the empirical spectral distribution of $M_N$ converges to a nontrivial, nonrandom limit. HINT: WLLN for $m-$dependent random variables.
(B) ** Identify the limit distribution.

## 4. Semi-Circle Law and the Catalan Numbers

**Definition 5.** The $n$th *Catalan number* is defined by

$$\kappa_n = \binom{2n}{n}\Big/(n+1). \tag{17}$$

**Exercise 6.** (A) Show that $\kappa_n$ is the number of *Dyck words* of length $2n$. A Dyck word is a sequence $x_1 x_2 \cdots x_{2n}$ consisting of n +1's and n −1's such that no initial segment of the string has more −1's than +1's, that is, such that for every $1 \le m \le 2n$,

$$\sum_{j=1}^{m} x_j \ge 0$$

HINT: Reflection Principle. (B) Show that $\kappa_n$ is the number of expressions containing $n$ pairs of parentheses which are correctly matched: For instance, if $n = 3$,

((())) ()(()) ()()() (())() (()())

(C) Show that $\kappa_n$ is the number of rooted ordered binary trees with $n+1$ leaves. (See your local combinatoricist for the definition.) HINT: Show that there is a bijection with the set of Dyck words, and use the result of part (A). (D) Do problem 6.19 in R. Stanley, *Enumerative Combinatorics*, vol. 2.

**Proposition 7.** *Let $p(t)$ be the semi-circle density* (5). *Then all odd moments of p are zero, and the even moments are the Catalan numbers:*

$$\kappa_n = \int_{-2}^{2} t^{2n} p(t)\, dt \quad \forall\ n = 0, 1, 2, \ldots \tag{18}$$

*Proof.* Calculus exercise. Hint: Use the change of variable $t = 2\cos\theta$ to rewrite the integral as an integral involving powers of $\sin\theta$ and $\cos\theta$. Then try either integration by parts, or rewrite the sines and cosines in terms of $e^{\pm i\theta}$ and use the orthogonality of the complex exponentials. □

*Remark* 1. If the semi-circle density (5) seems vaguely familiar, it is because it is closely related to the probability generating function of the first return time to the origin by a simple random walk on the integers. Let $S_n$ be a simple random walk started at 0; thus, the increments $X_n = S_n - S_{n-1}$ are independent, identically distributed Rademacher-$\frac{1}{2}$ (that is, $\pm 1$ with probabilities $\frac{1}{2}, \frac{1}{2}$). Define

$$T = \min\{n \ge 1 : S_n = 0\} \tag{19}$$

Then the probability generating function of $T$ is (see 312 notes)

$$Ez^T = 2 - \sqrt{4 - z^2} = \sum_{n=1}^{\infty} z^{2n} \binom{2n}{n} \Big/ \{(n+1)2^{2n}\}. \tag{20}$$

The last power series expansion follows either from Newton's binomial formula or a direct combinatorial argument (exercise). Note the appearance of the Catalan numbers.

## 5. PROOF OF WIGNER'S THEOREM

In this section we will prove Theorem 3 under the following additional assumption on the distributions of the entries $X_{i,j}$. Later we will show how truncation methods can be used to remove this assumption.

**Assumption 1.** The random variables $(X_{i,j})_{j \ge i}$ are independent. The (possibly complex) random variables $(X_{i,j})_{j > i}$ have distribution $F$ with mean zero, variance 1, and all moments finite, and the (real) andom variables $X_{i,i}$ have distribution $G$ with mean zero and all moments finite.

5.1. **Combinatorial Preliminaries.** The proof will be based on the method of moments. We will show that for each $k = 1,2,\dots,$ the random variable $N^{-1}\mathrm{Tr}\, M^k$ converges in probability to the $k$th moment of the semi-circle law. To accomplish this, it suffices, by Chebyshev's inequality, to show that the mean $N^{-1}E\mathrm{Tr}\, M^k$ converges to the $k$th moment of the semi-circle law, and that the variance $N^{-2}\mathrm{var}(\mathrm{Tr}\, M^k)$ converges to 0. By equation (14), the expected trace is a sum over closed paths of length $k$ in the augmented complete graph $K_N^+$. We will begin by showing that the contribution to the expectation of an closed path $\gamma$ depends only on its *type*.

**Definition 6.** The *type* of a path $\gamma = i_0 i_1 i_2 \cdots i_k$ is the sequence $\tau(\gamma) = m_0 m_1 m_2 \cdots m_k$, where for each $j$ the vertex $i_j$ is the $m_j$th *distinct* vertex encountered in the sequence $i_0, i_1, \dots, i_k$.

**Example 1.** The type of the path $uvquuquv$ is $1,2,3,1,1,3,1,2$.

**Lemma 8.** *If $\gamma$ and $\gamma'$ are two closed paths in the augmented complete graph $K_N^+$ with the same type, then there is a permutation $\sigma$ of the the vertex set $[N]$ that maps $\gamma$ to $\gamma'$.*

*Proof.* Exercise. □

**Corollary 9.** *If $\gamma$ and $\gamma'$ are two closed paths in $K_N^+$ with the same type, then*

$$Ew_M(\gamma) = Ew_M(\gamma') \tag{21}$$

*Proof.* This follows directly from the preceding lemma, because for any permutation $\sigma$ of the index set $[N]$, the distribution of the random matrix $(M_{\sigma(i),\sigma(j)})$ is the same as that of $(M_{i,j})$. This is obvious, because the off-diagonal entries of $M$ are i.i.d., as are the diagonal entries. Note: In fact, this holds more generally for any random matrix ensemble that is invariant under either unitary or orthogonal transformations, because a permutation matrix is both orthogonal and unitary. □

**Corollary 10.** *Let $\mathscr{T}_k$ be the set of all closed path types of length $k+1$. For each type $\tau$, let $H(\tau,N)$ be the number of closed paths in $K_N^+$ with type $\tau$, and let $Ew_M(\tau)$ be the common value of the expectations* (21)*. Then*

$$E\mathrm{Tr}\, M^k = \sum_{\tau\in\mathscr{T}_k} H(\tau,N)Ew_M(\tau). \tag{22}$$

*Moreover, if $G(\tau)$ is the number of distinct integers in the sequence $\tau$, then*

$$H(\tau,N) = N(N-1)(N-2)\cdots(N-G(\tau)+1) := (N)_{G(\tau)}. \tag{23}$$

*Proof.* Exercise. □

**Lemma 11.** *Let $\gamma$ be a closed path in $K_N^+$. If $Ew_M(\gamma) \neq 0$ then every edge or loop e crossed by $\gamma$ at least once must be crossed at least twice.*

*Proof.* If the edge $ij$ is crossed by $\gamma$ just once (say in the forward direction), then the factor $X_{i,j}$ occurs to the first power in the weight $w_M(\gamma)$ (see (13)), and the factor $X_{j,i} = X_{i,j}^*$ does not ocur at all. Since the random variables $X_{i,j}$ above or on the main diagonal are independent, all with mean zero, it follows that the expectation $Ew_M(\gamma) = 0$. □

For each closed path $\gamma$ in the *augmented* complete graph $K_N^+$, define $\Gamma(\gamma)$ to be the subgraph of the complete graph $K_N$ consisting of the vertices visited and edges crossed (in either direction) by $\gamma$. (Note: Loops are not included.) Clearly $\Gamma(\gamma)$ is itself a connected graph, because the path $\gamma$ visits every vertex and edge. By Lemma 8, if two closed paths $\gamma,\gamma'$ are of the same type, then $\Gamma(\gamma)$

and $\Gamma(\gamma')$ are isomorphic; moreover, if $\gamma$ crosses every edge in $\Gamma(\gamma)$ at least twice, then $\gamma'$ crosses every edge in $\Gamma(\gamma')$ at least twice.

**Lemma 12.** *If $\Gamma$ is a connected graph , then*

$$\#(\textit{vertices}) \leq \#(\textit{edges}) + 1, \tag{24}$$

*and equality holds if and only if $\Gamma$ is a* tree *(that is, a graph with no nontrivial cycles).*

*Remark* 2. Together with the equation (23), this will be used to show that the dominant contribution to the sum (22) comes from path-types for which representative paths $\gamma$ are such that $\Gamma(\gamma)$ is a tree and $\gamma$ contains no loops. The key observation is that if $\gamma$ is a closed path of length $k+1$ (that is, it makes $k$ steps) that crosses every edge in $\Gamma(\gamma)$ at least twice, then $\Gamma(\gamma)$ has no more than $[k/2]$ edges.

*Proof of Lemma 12.* Consider first the case where $\Gamma$ is a tree. I will show, by induction on the number of vertices in the tree $\Gamma$, that

$$\#(\text{vertices}) = \#(\text{edges}) + 1.$$

First note that, since $\Gamma$ has no nontrivial cycles, the removal of any edge must disconnect $\Gamma$ (why?). Also, since $\Gamma$ contains only finitely many edges, it must have an *end*, that is, a vertex $v$ that is incident to only one edge $e$ (why?) Now remove both $v$ and $e$ from the graph; the resulting graph $\Gamma'$ is still a tree (why?) but has one fewer vertex than $\Gamma$. Hence, the induction hypothesis implies that the desired formula holds for $\Gamma'$, and it therefore follows that it must also hold for $\Gamma$.

Now consider the case where the graph $\Gamma$ has at least one cycle. Then removal of one edge on this cycle will leave the graph *connected*, with the same number of vertices but one fewer edge. Continue to remove edges from cycles, one at a time, until there are no more cycles. Then the resulting graph will be a tree, with the same vertex set as the original graph $\Gamma$ but with strictly fewer edges. □

*Remark* 3. Here is another argument for inequality (24). This argument has the virtue that it extends easily to *pairs* of paths $\gamma, \gamma'$ – see the proof of Proposition 15 below. Let $\gamma = i_0 i_1 \dots i_{2k}$ be a path of length $2k$. Define a corresponding 0-1 sequence $\nu(\gamma) = n_0 n_1 \cdots n_{2k}$ as follows: If $\gamma$ visits vertex $i_j$ for the first time at step $j$, set $n_j = 1$, otherwise set $n_j = 0$. Now consider the $j$th step: If it is across an edge that was previously crossed (in either direction), then vertex $i_j$ must have been visited earlier, and so $n_j = 0$. Hence, if the path $\gamma$ crosses every edge at least twice, then

$$\sum_{j=1}^{2k} n_j \leq k,$$

because there are $2k$ edge crossings. Thus, the total number $\sum_{j=0}^{2k} n_j$ of vertices visited by $\gamma$ cannot be larger than $k+1$.

**Lemma 13.** *Let $\mathscr{G}_k$ be the set of closed path types $\tau \in \mathscr{T}_{2k}$ with the following properties:*

(i) *For at least one (and hence every) closed path $\gamma$ of type $\tau(\gamma) = \tau$, the graph $\Gamma(\gamma)$ is a tree.*
(ii) *The path $\gamma$ contains no loops, and crosses every edge in $\Gamma(\gamma)$* exactly *twice, once in each direction.*

*Then $\mathscr{G}_k$ has cardinality*

$$\#\mathscr{G}_k = \kappa_k, \tag{25}$$

*where $\kappa_k$ is the $k$th Catalan number.*

**Note:** If $\Gamma(\gamma)$ is a tree and if $\gamma$ crosses every edge in $\Gamma(\gamma)$ exactly twice, then it must necessarily cross every edge once in each direction, because otherwise the graph $\Gamma(\gamma)$ would contain a nontrivial cycle.

*Proof.* Let $\gamma = i_0 i_1 \cdots i_{2k}$ be an element of $\mathscr{G}_k$. By Lemma 12, the path $\gamma$ visits $k+1$ distinct vertices, including the initial vertex $i_0$. Define a $\pm 1$ sequence $s(\gamma) = s_1 s_2 \cdots s_{2k}$ as follows:

(+) $s_j = +1$ if $\gamma$ visits $i_{j-1}$ for the first time on the $(j-1)$th step;
(-) $s_j = -1$ otherwise.

This is a Dyck word, because for every $m \le 2k$ the sum $\sum_{j=1}^{m} s_j$ counts the number of vertices visited exactly once by $\gamma$ in the first $m$ steps (why?). Conversely, for every Dyck word $s$ of length $2k$, there is a closed path $\gamma$ in the graph $K_N$ such that $s(\gamma) = s$ *provided* $N \ge k+1$ (why?). Hence, the result follows from Exercise 6. □

5.2. **Convergence of Means.**

**Proposition 14.** *If Assumption 1 holds, then for every integer $k \ge 0$,*

$$\lim_{N\to\infty} N^{-1} E\mathrm{Tr}\, M^{2k} = \kappa_k \quad \textit{and} \tag{26}$$

$$\lim_{N\to\infty} N^{-1} E\mathrm{Tr}\, M^{2k+1} = 0 \tag{27}$$

*where $\kappa_k$ is the $k$th Catalan number.*

*Proof.* The starting point is formula (22). According to this formula, the expectation is a sum over closed path types of the appropriate length. By Lemma 11, the only path types $\tau$ that contribute to this sum are those for which representative closed paths $\gamma$ cross every edge twice. I will use Lemma 12 to determine the types that make the *dominant* contribution to the sum (22).

**Case 1: Odd Moments.** Since only paths $\gamma$ that make $2k+1$ steps and cross every edge at least twice contribute to the expectation $E\mathrm{Tr}\, M^{2k+1}$, such paths have graphs $\Gamma(\gamma)$ with no more than $k$ edges. Consequently, by Lemma 12, $\Gamma(\gamma)$ has no more than $k+1$ vertices, and so equation (23) implies that

$$H(\tau(\gamma), N) \le N^{k+1}.$$

Now consider the expectation $Ew_M(\gamma)$. The factors in the product

$$w_M(\gamma) = \prod_{j=0}^{2k} M_{i_j,i_{j+1}} = N^{-(2k+1)/2} \prod_{j=0}^{2k} X_{i_j,i_{j+1}}$$

can be grouped by edge (or loop, if there are factors $X_{i,i}$ from the diagonal). By Assumption 1, the distributions of the diagonal and off-diagonal entries have finite moments of all orders, so Hölder's inequality implies that

$$|Ew_M(\tau(\gamma))| = |Ew_M(\gamma)| \le N^{-(2k+1)/2} \max(E|X_{1,2}|^{2k+1}, E|X_{1,1}|^{2k+1}).$$

Therefore,

$$\begin{aligned} N^{-1}|E\mathrm{Tr}\, M^{2k+1}| &\le N^{-1} \sum_{\tau} H(\tau, N)|Ew_M(\tau)| \\ &\le N^{-1} N^{k+1} N^{-(2k+1)/2} \max(E|X_{1,2}|^{2k}, E|X_{1,1}|^{2k}) \\ &= O(N^{-1/2}) \end{aligned}$$

**Note:** The upper bound can be sharpened by a factor of $N^{-1}$, because with a bit more work (Exercise!) it can be shown that if $\gamma$ is a closed path with an odd number $2k+1$ of steps that crosses every edge at least twice, then $\Gamma(\gamma)$ must contain a cycle, and so cannot have more than $k$ vertices.

**Case 2: Even Moments.** In this case, I will use use Lemma 12 to show that for large $N$, the *dominant* contribution comes from types $\tau \in \mathcal{G}_k$ First, observe that if $\tau \in \mathcal{G}_k$, then for any closed path $\gamma$ of type $\tau(\gamma) = \tau$ the product contains $k$ *distinct* factors $|X_{l,m}|^2/N$, all with $m > l$. This is because the path $\gamma$ crosses every edge in $\Gamma(\gamma)$ exactly twice, once in each direction, and contains no loops. (Recall that the random matrix $X$ is Hermitian, so $X_{l,m} = X^*_{m,l}$.) Since the $k$ factors are distinct, they are independent, and since they are also identically distributed with variance 1 it follows that

$$Ew_M(\tau) = N^{-k}(E|X_{1,2}|^2)^k = N^{-k}.$$

Now consider an arbitrary path type $\tau \in \mathcal{T}_{2k}$, and let $\gamma$ be a closed path of type $\tau$. As in Case 1, Assumption 1 and Hölder's inequality imply that

$$|Ew_M(\gamma)| \le N^{-k} \max(E|X_{1,2}|^{2k}, E|X_{1,1}|^{2k}).$$

Observe that this is of the same order of magnitude as the expectation $N^{-1}$ for types $\tau \in \mathcal{G}_k$.

Finally, consider the factors $H(\tau, N) = (N)_{G(\tau)}$ in the formula (22). By Lemma 12, $G(\tau) \le k+1$ for all types, and the inequality is strict *except* for types $\tau \in \mathcal{G}_k$. Hence, the dominant contribution to the sum (22) comes from types $\tau \in \mathcal{G}_k$. Therefore, by Lemma 13,

$$E\text{Tr } M^{2k} = \sum_{\tau \in \mathcal{T}_{2k}} H(\tau, N) E w_M(\tau) \sim \sum_{\tau \in \mathcal{G}_k} H(\tau, N) N^{-k} \sim N\kappa_k.$$

□

5.3. **Bound on the Variance.** By Lemma 11 and Proposition 14, the *expected* moments of the empirical spectral distribution $F^M$ converge to the corresponding moments of the semi-circle law. Thus, to complete the proof of Wigner's theorem, it suffices to show that the *variances* of these moments converge to zero as $N \to \infty$.

**Proposition 15.** *If Assumption 1 holds, then there exist constants $C_k < \infty$ such that*

$$\text{var}(N^{-1}\text{Tr } M^k) \le C_k/N. \tag{28}$$

*Proof.* Use the basic formula (12) to compute the second moment and the square of the first moment:

$$E(\text{Tr } M^k)^2 = \sum_\gamma \sum_{\gamma'} E w_M(\gamma) w_M(\gamma') \quad \text{and}$$
$$(E\text{Tr } M^k)^2 = \sum_\gamma \sum_{\gamma'} E w_M(\gamma) E w_M(\gamma'),$$

where the sums are over the set of all pairs $\gamma, \gamma'$ of closed paths of length $k$ in the complete graph $K_N$. Now for any pair $\gamma, \gamma'$ with no vertices (and hence no edges) in common,

$$E w_M(\gamma) w_M(\gamma') = E w_M(\gamma) E w_M(\gamma')$$

because the products $w_M(\gamma)$ and $w_M(\gamma')$ have no factors in common. Consequently, the difference $E(\text{Tr } M^k)^2 - (E\text{Tr } M^k)^2$ is entirely due to pairs $\gamma, \gamma'$ that have vertices in common. Furthermore, for the same reason as in Lemma 11, $E w_M(\gamma) w_M(\gamma') = 0$ unless every edge in $\Gamma(\gamma) \cup \Gamma(\gamma')$

is crossed at least twice (either twice by $\gamma$, twice by $\gamma$, or once by each). Consider a pair $\gamma, \gamma'$ with both these properties: by the argument in Remark 3 above, the total number of vertices in $\Gamma(\gamma) \cup \Gamma(\gamma')$ cannot exceed $k+1$. Consequently, the total number of such pairs is no larger than $N^{k+1}$, and so their aggregate contribution to the expectation $E(\text{Tr } M^k)^2$ is bounded in magnitude by

$$N^{2k+1} \max_{\gamma,\gamma'} |E w_M(\gamma) w_M(\gamma')| \leq NE|X_{1,2}|^{2k}.$$

□

**Exercise 7.** * Show that the bound (28) can be improved to

$$\text{var}(N^{-1}\text{Tr } M^k) \leq C_k / N^2.$$

## 6. Proof of the Marchenko-Pastur Theorem

6.1. **Trace Formula for Covariance Matrices.** The basic strategy will be the same as for Wigner's theorem, but the combinatorics is somewhat different. The starting point is once again a combinatorial formula for Tr $S^k$, where $S$ is the sample covariance matrix (11). This may be rewritten as

$$S = YY^T \quad \text{where} \quad Y = X/\sqrt{n}. \tag{29}$$

Recall that the matrix $X$ (and therefore also the matrix $Y$) is $p \times n$. Thus, the entries $Y_{i,j}$ of $Y$ have row indices $i \in [p]$ and column indices $j \in [n]$. Define $B = B(p,n)$ to be the *complete bipartite graph* with partitioned vertex set $[p] \cup [n]$: that is, the graph whose edges join arbitrary pairs $i \in [p]$ and $j \in [n]$. It is helpful to think of the vertices as being arranged on two levels, an upper level $[p]$ and a lower level $[n]$; with this convention, paths in the graph make alternating up and down steps. Denote by $\mathscr{Q}$ the set of all closed paths

$$\gamma = i_0, j_1, i_1, j_2, \cdots i_{k-1}, j_k, i_k = i_0 \tag{30}$$

in $B(p,n)$ of length $2k$ beginning and ending at a vertex $i \in [p]$, and for each such path define the *weight* (or $Y-$*weight*) of $\gamma$ by

$$w(\gamma) = w_Y(\gamma) = \prod_{m=1}^{k} Y_{i_{m-1},j_m} Y_{j_m,i_m}. \tag{31}$$

Then

$$\boxed{\text{Tr } S^k = \sum_{\gamma \in \mathscr{Q}} w_Y(\gamma)}. \tag{32}$$

6.2. **Enumeration of Paths.** The proof of the Marchenko-Pastur Theorem will be accomplished by showing that for each $k \geq 0$ (a) the expectation $E\text{Tr } S^k$ converges to the $k$th moment of the Marchenko-Pastur distribution (10), and (b) the variance varTr $S^k$ converges to 0. Given (a), the proof of (b) is nearly the same as in the proof of Wigner's theorem, so I will omit it. The main item, then, is the analysis of the expectation $E\text{Tr } S^k$ as $n, p \to \infty$ with $p/n \to y$.

**Lemma 16.** *Let $\gamma \in \mathscr{Q}$ be a closed path in $B(p,n)$ that begins and ends at a vertex $i \in [p]$. If $Ew_Y(\gamma) \neq 0$ then every edge crossed by $\gamma$ must be crossed at least twice.*

*Proof.* See Lemma 11. □

For any path $\gamma$ in $B(n,p)$ with initial vertex in $[p]$, the type $\tau(\gamma)$ and the associated subgraph $\Gamma(\gamma)$ of $B(n,p)$ are defined as earlier. Since paths $\gamma$ jump back and forth between the disjoint vertex sets $[p]$ and $[n]$, even entries of $\tau(\gamma)$ count vertices in $[p]$ and odd entries count vertices in $[n]$.( Note: Indices start at 0.) Thus, it will be convenient to extract the even- and odd- entry subsequences $\tau_E$ and $\tau_O$ of $\tau$, and to write $\tau = (\tau_E, \tau_O)$ (this is an abuse of notation – actually $\tau$ is the sequence obtained by *shuffling* $\tau_E$ and $\tau_O$). Lemma 8 translates to the bipartite setting as follows:

**Lemma 17.** *If $\gamma$ and $\gamma'$ are two closed paths in the complete bipartite graph $B(n,p)$ with the same type $\tau(\gamma)=\tau(\gamma')$, then there are permutations $\sigma,\sigma'$ of the vertex sets $[n]$ and $[p]$ that jointly map $\gamma$ to $\gamma'$. Consequently, if $\gamma,\gamma'$ have the same type then*

$$Ew_M(\gamma) = Ew_M(\gamma'). \tag{33}$$

**Corollary 18.** *Let $\mathscr{T}_k^B$ be the set of all closed path types of length $2k+1$. For each $\tau \in \mathscr{T}_K^B$ let $H^B(\tau,n,p)$ be the number of closed paths in $B(n,p)$ of type $\tau$, and let $Ew_S(\tau)$ be the common value of the expectations* (33)*. Then*

$$E\mathrm{Tr}\, S^k = \sum_{\tau\in\mathscr{T}_k^B} H^B(\tau,n,p) Ew_S(\tau). \tag{34}$$

*Moreover, if $G_O(\tau)$ and $G_E(\tau)$ are the number of distinct entries in the even/odd sequences $\tau_O$ and $\tau_E$, respectively, then*

$$H(\tau,n,p) = (p)_{G_E(\tau)}(n)_{G_O(\tau)}. \tag{35}$$

Recall that $(n)_m = n(n-1)\cdots(n-m+1)$. The proof of Corollary 18 is trivial, but the difference in formulas (23) and (35) is what ultimately accounts for the difference between the limit distributions in the Wigner and Marchenko-Pastur theorems. Note that $G_O(\tau)$ and $G_E(\tau)$ are the numbers of distinct vertices in the lower and upper levels $[n]$ and $[p]$, respectively, visited by a path $\gamma$ of type $\tau$.

**Lemma 19.** *Let $\gamma$ be a closed path in the bipartite graph $B(n,p)$ of length $2k+1$ (thus, $\gamma$ makes $2k$ steps, each across an edge). If $\gamma$ crosses each edge of $\Gamma(\gamma)$ at least twice, then the number of distinct vertices in $\Gamma(\gamma)$ is no larger than $k+1$:*

$$G_O(\tau(\gamma)) + G_E(\tau(\gamma)) \le k+1, \tag{36}$$

*and equality holds if and only if $G(\gamma)$ is a tree and $\gamma$ crosses every edge of $\Gamma(\gamma)$ exactly twice, once in each direction.*

This can be proved by the same argument as in Lemma 12, and the result serves the same purpose in the proof of the theorem – it implies that the main contribution to the sum (34) comes from path types $\tau$ for which the corresponding graph $\Gamma$ is a tree. To see this, keep in mind that $p$ and $n$ are of the same order of magnitude, so (35) will be of maximum order of magnitude when $G_O(\tau(\gamma)) + G_E(\tau(\gamma))$ is maximal. Enumeration of these types is the main combinatorial task:

**Lemma 20.** *Let $\mathscr{G}_k^B(r)$ be the set of closed path types $\tau \in \mathscr{T}_{2k}^B$ with the following properties:*

(i) *For at least one (and hence every) closed path $\gamma$ of type $\tau(\gamma)=\tau$, the graph $\Gamma(\gamma)$ is a tree.*
(ii) *The path $\gamma$ crosses every edge in $\Gamma(\gamma)$* exactly *twice, once in each direction.*
(iii) $G_E(\tau) = r+1$ *and* $G_O(\tau) = k-r$.

*(Note that $G_E(\tau) \geq 1$ for any allowable type, because all paths start and end on the upper level $[p]$.) Then $\mathscr{G}_k^B(r)$ has cardinality*

$$\#\mathscr{G}_k^B(r) = \binom{k-1}{r}\binom{k}{r}\Big/(r+1). \tag{37}$$

*Proof.* First, I will show that $\mathscr{G}_k^B(r)$ can be put into one-to-one correspondence with the set $\mathscr{A}_k^+(r,r)$ of sequences

$$s = d_1 u_1 d_2 u_2 \cdots d_k u_k$$

satisfying the following properties:

- (A) Each $d_i \in \{0,-1\}$, and each $u_i \in \{0,1\}$.
- (r,r) $\sum_{i=1}^k u_1 = r$ and $\sum_{i=1}^k d_i = -r$.
- (+) $\sum_{i=1}^m (u_i + d_i) + d_{m+1} \geq 0$ for all $m < k$.

Then I will show that the set $\mathscr{A}_k^+(r,r)$ has cardinality (37).

**Step 1: Bijection $\mathscr{G}_k^B(r) \leftrightarrow \mathscr{A}_k^+(r,r)$.** Let $\gamma$ be a path satisfying hypotheses (i)–(iii) of the lemma. This path crossses $k$ edges, once in each direction, and alternates between the vertex sets $[p]$ and $[n]$. After an even number $2i$ of edge crossings, $\gamma$ is at a vertex in $[p]$; if this vertex is new – that is, if it is being visited for the first time – then set $u_i = +1$, and otherwise set $u_i = 0$. After an odd number $2i-1$ of edge crossings, $\gamma$ is at a vertex in $[n]$, having just exited a vertex $v \in [p]$. If this edge crossing is the *last* exit from vertex $v$, then set $d_i = -1$; otherwise, set $d_i = 0$. It is clear that this sequence $s$ has properties (1), (2), (3) above, and it is also clear that if $\gamma$ and $\gamma'$ have the same type, then the corresponding sequences $s$ and $s'$ are the same. Thus, we have constructed a mapping $\varphi : \mathscr{G}_k^B(r) \to \mathscr{A}_k^+(r,r)$.

It remains to prove that $\varphi$ is a bijection. This we will accomplish by exhibiting an inverse mapping $\psi : \mathscr{A}_k^+(r,r) \to \mathscr{G}_k^B(r)$. Several facts about the algorithm used to produce $\varphi$ are relevant: If $\gamma$ is a path through vertices $i_0, j_1, i_1, j_2, \dots$ as in (30), with $i_m \in [p]$ amd $j_m \in [n]$, then

- (a) Vertex $j_m$ is new *unless* $d_m = -1$.
- (b) Vertex $j_m$ is exited for the last time if and only if $u_m = 0$.
- (c) If vertex $i_m$ is *not* new, that is, if $u_m = 0$, then $i_m$ is the *last* vertex in $[p]$ visited by $\gamma$ that has not been exited for the last time, that is, by an edge marked $d = -1$.
- (d) If vertex $j_m$ is *not* new, then $j_m$ is the *last* vertex in $[n]$ visited by $\gamma$ that has not been exited for the last time.

These all follow from the hypotheses that $\Gamma(\gamma)$ is a tree, and that $\gamma$ crosses every edge of $\Gamma(\gamma)$ exactly twice, once in each direction. Consider, for instance, assertion (a): If $d_m = -1$, then vertex $i_{m-1}$ is being exited for the last time, across the edge $i_{m-1}j_m$; this edge must have been crossed once earlier, so $j_m$ is not new. Similarly, if $d_m = 0$, then vertex $i_{m-1}$ will be revisited at some later time; if $j_m$ were *not* new, then either $\gamma$ would have a nontrivial cycle or edge $i_{m-1}j_m$ would be crossed $\geq 3$ times.

**Exercise 8.** Give detailed proofs of (a)–(d).

Assertions (a)–(d) provide an algorithm for constructing the type sequences $\tau_E, \tau_O$ from the sequence $s$. By (a) and the definition of the sequence $s$, the markers $u_m = 1$ and $d_l = 0$ indicate which entries of $\tau_E$ and $\tau_O$ correspond to new vertices; each such entry must be the max+1 of the previous entries. By (b) and the definition of $s$, the markers $u_m = 0$ and $d_l = -1$ indicate which entries of $\tau_E$ and $\tau_O$ correspond to vertices being exited for the last time. Consequently, (c) and (d) allow one to deduce which previously visited vertex is being visited when the current vertex is not new.

**Step 2: Enumeration of $\mathscr{A}_k^+(r,r)$.** This will make use of the *Reflection Principle*, in much the same way as in the classical ballot problem. Note first that sequences $s \in \mathscr{A}_k^+(r,r)$ must have $d_1 = 0$ and $u_k = 0$, because of the constraints (+) and (r,r), so we are really enumerating the set of sequences

$$s = 0(u_1 d_2)(u_2 d_3)\cdots(u_{k-1} d_k)0 \tag{38}$$

that satisfy the properties (A), (r,r), and (+).

The first step is to enumerate the sequences of the form (38) that satisfy property (A) and count constraints $(r,s)$ such as the constraint $(r,r)$ above. Thus, for each integer $k \geq 1$ and integers $r,s \geq 0$, let $\mathscr{A}_{k-1}(r,s)$ be the set of sequences $s$ of the form (38) such that:

(A) Each $d_i \in \{0,-1\}$, and each $u_i \in \{0,1\}$.
(r,s) $\sum_{i=1}^k u_1 = r$ and $\sum_{i=1}^k d_i = -s$.

This set is easily enumerated: There are $k-1$ "down" slots, which must be filled with $k-s-1$ 0's and $s$ (-1)'s, and there are $k-1$ "up" slots, which must be filled with $k-r-1$ 0's and $r$ (+1)'s, so

$$\#(\mathscr{A}_{k-1}(r,s)) = \binom{k-1}{r}\binom{k-1}{s}. \tag{39}$$

The proof of (37) will be completed by showing that

$$\#\mathscr{A}_k^+(r,r) = \#\mathscr{A}_{k-1}(r,r) - \#\mathscr{A}_{k-1}(r-1,r+1). \tag{40}$$

To see this, observe that every sequence in $\mathscr{A}_k^+(r,r)$ is an element of $\mathscr{A}_{k-1}(r,r)$, by (38). Hence, it is enough to show that the set of sequences in $\mathscr{A}_{k-1}(r,r)$ that do *not* satisfy the nonnegativity constraint (+) is in one-to-one correspondence with $\mathscr{A}_{k-1}(r-1,r+1)$. If a sequence $s \in \mathscr{A}_k(r,r)$ does not satisfy (+), then there must be a *first* time $\tau < k$ such that

$$\sum_{i=1}^{\tau-1}(u_i + d_{i+1}) = -1; \tag{41}$$

moreover, the remaining terms must sum to $+1$, because the overall sum is 0:

$$\sum_{i=\tau}^{k-1}(u_i + d_{i+1}) = +1. \tag{42}$$

Reflect this part of the path by reversing all pairs $u_j d_{j+1}$ and then negating: thus, for each $u_j d_{j+1}$ with $j \geq \tau$,

$$\begin{aligned}(+,0) &\mapsto (0,-)\\ (+,-) &\mapsto (+,-)\\ (0,-) &\mapsto (+,0)\\ (0,0) &\mapsto (0,0).\end{aligned}$$

This mapping sends sequences $s$ in $\mathscr{A}_{k-1}(r,r)$ that do not satisfy (+) to sequences $s'$ in $\mathscr{A}_{k-1}(r-1),r+1)$; it is invertible because for any sequence $s' \in \mathscr{A}_{k-1}(r-1,r+1)$, the reflection rule can be applied in reverse – after the first time $\tau$ such that the sum (41) reaches the level $-1$, flip the remaining pairs according to the same rules as above to recover a sequence $s \in \mathscr{A}_{k-1}(r-1, r-1)$. □

### 6.3. **Moments of the Marchenko-Pastur Law.**

**Proposition 21.** *The $k$th moment of the Marchenko-Pastur density $f_y(x)$ defined by* (10) *is*

$$\int x^k f_y(x)\,dx = \sum_{r=0}^{k} y^r \binom{k-1}{r}\binom{k}{r}\Big/(r+1). \tag{43}$$

*Proof.* Another calculus exercise, this one somewhat harder than the last. □

### 6.4. **Convergence of Means.**

**Proposition 22.** *Let $X$ be a random $p \times n$ matrix whose entries $X_{i,j}$ are independent, identically distributed random variables with distribution $G$ with mean zero, variance* 1*, and all moments finite. Define $S = XX^T/n$. As $n, p \to \infty$ with $p/n \to y \in (0,\infty)$,*

$$Ep^{-1}\mathrm{Tr}\, S^k \longrightarrow \sum_{r=0}^{k} y^r \binom{k-1}{r}\binom{k}{r}\Big/(r+1).$$

*Proof.* This is very similar to the proof of the corresponding result in the Wigner matrix setting. Lemma 19, together with a simple argument based on the Hölder inequality, implies that the dominant contribution to the sum (34) comes from types $\tau$ such that paths $\gamma$ of type $\tau(\gamma) = \tau$ have graphs $\Gamma(\gamma)$ that are trees, and cross every edge exactly twice, once in each direction. For these types $\tau$, the products $w_S(\tau)$ contain $k$ factors $|X_{i,j}|^2/n^k$; these are independent, with mean 1, so the expectation $Ew_S(\gamma) = 1/n^k$ for each type $\tau \in \mathscr{G}_k^B(r)$. Consequently,

$$E\mathrm{Tr}\, S^k \sim \sum_{r=0}^{k-1} \sum_{\tau \in \mathscr{G}_k^B(r)} H^B(\tau, n, p)/n^k = \sum_{r=0}^{k-1} \sum_{\tau \in \mathscr{G}_k^B(r)} (p)_{r+1}(n)_{k-r}/n^k$$

The result now follows directly from Lemma 20. □

## 7. TRUNCATION TECHNIQUES

7.1. **Perturbation Inequalities.** In section 5 we proved Wigner's theorem under the assumption that the entries $X_{i,j}$ have expectation zero and all moments finite. In this section, we show how these hypotheses can be removed. The main tools are the following inequalities on the difference between the empirical spectral distributions of two Hermitian matrices $A, B$. In the first, $L(F, G)$ denotes the Lévy distance between the probability distributions $F, G$ (see Background notes).

**Proposition 23.** *(Perturbation Inequality) Let $A$ and $B$ be Hermitian operators on $V = \mathbb{C}^n$ relative to the standard inner product, and let $F^A$ and $F^B$ be their empirical spectral distributions. Then*

$$L(F^A, F^B) \le n^{-1}\mathrm{rank}(A-B). \tag{44}$$

**Proposition 24.** *Let $A$ and $B$ be $n \times n$ Hermitian matrices with eigenvalues $\lambda_i$ and $\mu_i$, respectively, listed in decreasing order. Then*

$$\sum_{i=1}^{n} (\lambda_i - \mu_i)^2 \le \mathrm{Tr}\,(A-B)^2. \tag{45}$$

See the Background Notes for proofs.

7.2. **Wigner's Theorem for Real Symmetric Matrices.** Assume now that the entries $X_{i,j}$ are *real*, so that $X_{i,j} = X_{j,i}$,and that the hypotheses of Theorem 3 are satisfied. Thus, the common distribution $G$ of the off-diagonal entries $X_{i,j}$ has finite second moment, but the distribution $H$ of the diagonal entries is arbitrary (it need not even have a finite first moment). The first step will be to show that *truncating* either the diagonal or off-diagonal entries has only a small effect on the empirical spectral distribution .

**Lemma 25.** *For a fixed constant* $0 < C < \infty$, *set*

$$\tilde{X}_{i,i} = X_{i,i}\mathbf{1}_{[-C,C]}(X_{i,i})$$

*and define* $\tilde{X}$ *and* $\tilde{M} = \tilde{X}/\sqrt{N}$ *to be the matrices obtained from* $X$ *by changing the diagonal entries* $X_{i,i}$ *of* $X$ *to* $\tilde{X}_{i,i}$. *Then for any* $\varepsilon > 0$ *there exists* $C = C_{\varepsilon,F} < \infty$ *such that with probability approaching* 1 *as* $N \to \infty$,

$$\|F^M - F^{\tilde{M}}\|_\infty < \varepsilon. \tag{46}$$

*Proof.* This follows from the perturbation inequality (44). For any $\varepsilon > 0$, if $C$ is sufficiently large then with probability approaching one as $N \to \infty$, fewer than $\varepsilon N$ of the diagonal entries will be changed. Consequently, the difference $M - \tilde{M}$ will have fewer than $\varepsilon N$ nonzero entries, and hence will be of rank $< \varepsilon N$. Therefore, the perturbation inequality (44) implies (46). □

**Lemma 26.** *Assume that the distribution* $G$ *of the off-diagonal entries* $X_{i,j}$ *has finite second moment. For any constant* $0 < C < \infty$, *let*

$$\tilde{X}_{i,j} = X_{i,j}\mathbf{1}_{[-C,C]}(X_{i,j}) \quad \text{for } j \neq i, \tag{47}$$

*and define* $\tilde{X}$ *and* $\tilde{M} = \tilde{X}/\sqrt{N}$ *to be the matrix obtained from* $X$ *by changing the off-diagonal entries* $X_{i,j}$ *of* $X$ *to* $\tilde{X}_{i,j}$. *Then for any* $\varepsilon > 0$ *there exists* $C = C_{\varepsilon,F} < \infty$ *such that with probability approaching* 1 *as* $N \to \infty$,

$$\sup_{x\in\mathbb{R}} |F^M(x+\varepsilon) - F^{\tilde{M}}(x)| < \varepsilon \quad \text{and} \tag{48}$$

$$\sup_{x\in\mathbb{R}} |F^M(x-\varepsilon) - F^{\tilde{M}}(x)| < \varepsilon. \tag{49}$$

*Remark* 4. The slightly odd-looking conclusion is a way of asserting that the empirical spectral distributions of $M$ and $\tilde{M}$ are close in the *weak topology* (the topology of convergence in distribution).

*Proof.* This follows from the trace inequality (45). Denote by $\lambda_i$ and $\mu_i$ the eigenvalues of $M$ and $\tilde{M}$, respectively (in decreasing order). The difference $M - \tilde{M}$ has nonzero entries only in those $i, j$ for which $i \neq j$ and $|X_{i,j}| > C$, and so

$$\text{Tr }(M - \tilde{M})^2 = N^{-1} \sum_{i\neq j} X_{i,j}^2 \mathbf{1}\{|X_{i,j}| > C\}.$$

Taking expectation gives

$$\begin{aligned} E\text{Tr }(M - \tilde{M})^2 &= N^{-1} \sum_{i\neq j} EX_{i,j}^2 \mathbf{1}\{|X_{i,j}| > C\} \\ &\leq N^{-1}N(N-1)EX_{1,2}^2\mathbf{1}\{|X_{1,2}| > C\} \\ &\leq \delta^2 N \end{aligned}$$

provided $C$ is sufficiently large. (This follows because we have assumed that the distribution of the random variables $X_{i,j}$ has finite second moment). The Markov inequality now implies that

$$P\{\mathrm{Tr}\,(M-\tilde{M})^2 \geq \delta N\} \leq \delta.$$

Hence, by the trace inequality (45),

$$P\{\sum_{i=1}^{N}(\lambda_i-\mu_i)^2 \geq \delta N\} \leq \delta.$$

This (with $\delta=\varepsilon^2$) implies the inequalities (48)-(**??**). (Exercise: Explain why.) □

For any constant $C<\infty$, the truncated random variables $\tilde{X}_{i,i}$ and $\tilde{X}_{i,j}$ defined in Lemmas 25 and 26 have finite moments of all order. However, they need not have expectation zero, so the version of Wigner's theorem proved in section 5 doesn't apply directly.

**Lemma 27.** *Suppose that Wigner's theorem holds for Wigner matrices $X_{i,i}\sim G$ and $X_{i,j}\sim H$. Then for any constants $a,b\in\mathbb{R}$, the theorem also holds for Wigner matrices with diagonal entries $X_{i,i}+a$ and off-diagonal entries $X_{i,j}+b$.*

*Proof.* First, consider the effect of adding a constant $b$ to *all* of the entries of $X$: this changes the matrix to $\tilde{X}=X+b\mathbf{1}$, where $\mathbf{1}$ is the matrix with all entries 1. The matrix $\mathbf{1}$ has rank 1, so the perturbation inequality (44) implies that the Lévy distance between the empirical spectral distributions $F^M, F^{\tilde{M}}$ of $M=X/\sqrt{N}$ and $\tilde{M}=\tilde{X}/\sqrt{N}$ is bounded by $1/N$. Hence, if $F^M$ is close to the semi-circle law in Lévy distance, then so is $F^{\tilde{M}}$.

Now consider the effect of adding a constant $a$ to every diagonal entry. This changes $X$ to $\tilde{X}=X+aI$, and thus $M$ to $\tilde{M}=M+aI/N$. Clearly,

$$\mathrm{Tr}\,(M-\tilde{M})^2=a^2/N;$$

consequently, by Proposition 24, the eigenvalues $\lambda_i$ and $\mu_i$ (listed in order) satisfy

$$\sum_i |\lambda_i-\mu_i|^2 \leq a^2/N.$$

This implies that the Lévy distance between the empirical spectral distributions of $M$ and $\tilde{M}$ converges to 0 as $N\to\infty$. □

*Proof of Wigner's Theorem.* This is a direct consequence of Lemmas 25, 26, and 27. □