CHAPTER 2

Wigner's semicircle law

1. Wigner matrices

Definition 12. A *Wigner matrix* is a random matrix $X = (X_{i,j})_{i,j \le n}$ where

(1) $X_{i,j}$, $i < j$ are i.i.d (real or complex valued).
(2) $X_{i,i}$, $i \le n$ are i.i.d real random variables (possibly a different distribution)
(3) $X_{i,j} = \overline{X}_{j,i}$ for all i,j.
(4) $\mathbf{E}[X_{1,2}] = 0$, $\mathbf{E}[|X_{1,2}|^2] = 1$. $\mathbf{E}[X_{1,1}] = 0$, $\mathbf{E}[X_{1,1}^2] < \infty$.

Definition 13. Let A have i.i.d $CN(0,1)$ entries and let H have i.i.d $N(0,1)$ entries. Set $X = \frac{A+A^*}{\sqrt{2}}$ and $Y = \frac{H+H^*}{\sqrt{2}}$. X is called the *GUE matrix* and Y is called the *GOE matrix*. Equivalently, we could have defined X (or Y) as a Wigner matrix with $X_{1,2} \sim CN(0,1)$ (resp. $Y_{1,2} \sim N(0,1)$) and $X_{1,1} \sim N(0,2)$ (resp. $Y_{1,1} \sim N(0,2)$). GUE and GOE stand for Gaussian unitary ensemble and Gaussian orthogonal ensemble, respectively.

The significance of GUE and GOE matrices is that their eigenvalue distributions can be computed *exactly*! We shall see that later in the course. However, for the current purpose of getting limits of ESDs, they offer dispensable, but helpful, simplifications in calculations. The following exercise explains the reson for the choice of names.

Exercise 14. Let X be a GOE (or GUE) matrix. Let P be a non-random orthognal (respectively, unitary) $n \times n$ matrix. Then $P^*XP \stackrel{d}{=} P$.

Let X be a Wigner matrix and let $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$ denote the eigenvalues of X (real numbers, since X is Hermitian). Observe that $\sum \tilde{\lambda}_k^2 = \text{tr}(X^2) = \sum_{i,j} |X_{i,j}|^2$. By the law of large numbers, the latter converges in probability if we divide by n^2 and let $n \to \infty$. Hence, if we let $\lambda_k = \tilde{\lambda}_k / \sqrt{n}$ be the eigenvalues of $X/\sqrt{n}$, then $\mathbf{n}^{-1} \sum_{k=1}^n \lambda_k^2$ converges in probability to a constant. This indicates that we should scale X down by $\sqrt{n}$[1]. Let L_n and $\overline{L}_n$ denote the ESD and the expected ESD of $X/\sqrt{n}$ respectively. Note that we used the finiteness of variance of entries of X in arguing for the $1/\sqrt{n}$ scaling. For heavy tailed entries, the scaling will be different.

Theorem 15. *Let X_n be an $n \times n$ Wigner random matrix. Then $\overline{L}_n \to \mu_{s.c}$ and $L_n \stackrel{P}{\to} \mu_{s.c}$.*

In this chapter we shall see three approaches to proving this theorem.

(a) The method of moments.
(b) The method of Stieltjes' transforms
(c) The method of invariance principle.

---

[1]Recall that for a sequence of probability measures to converge, it must be *tight*. Often the simplest way to check tightness is to check that the variances or second moments are bounded. This is what we did here.

Roughly, these methods can be classified as combinatorial, analytic and probabilistic, respectively. The first two methods are capable of proving Theorem 15 fully. The last method is a general probabilistic technique which does not directly prove the theorem, but easily shows that the limit must be the same for all Wigner matrices.

Since part of the goal is to introduce these techniques themselves, we shall not carry out each proof to the end, particularly as the finer details get more technical than illuminating. For example, with the method of moments we show that expected ESD of GOE matrices converges to semi-circular law and only make broad remarks about general Wigner matrices. Similarly, in the Stieltjes transform proof, we shall assume the existence of fourth moments of $X_{i,j}$. However, putting everthing together, we shall havea complete proof of Theorem 15. These techniques can be applied with minimal modifications to several other models of random matrices, but these will be mostly left as exercises.

2. The method of moments for expected ESD of GOE and GUE matrix

The idea behind the method of moments is to show that $\mu_n \to \mu$, whene $\mu_n, \mu \in \mathcal{P}(\mathbb{R})$ by showing that the moments $\int x^p \mu_n(dx) \to \int x^p \mu(dx)$ for all non-negative integer p. Of course this does not always work. In fact one can find two probability measures μ and ν with the same moments of all orders. Taking $\mu_n = \nu$ gives a counterexample.

Result 16. Let $\mu_n, \mu \in \mathcal{P}(\mathbb{R})$ and ssume that $\int x^p \mu_n(dx) \to \int x^p \mu(dx)$ for all $p \geq 1$. If μ is determined by its moments, then $\mu_n \to \mu$.

Checking if a probability measure is determined by its moments is not easy. An often used sufficient condition is summability of $(\int x^{2p} \mu(dx))^{-1/2p}$, called *Carlemann's condition*. An even easier version which suffices for our purposes (for example when the limit is the semicircle distribution) is in the following exercise.

Exercise 17. Let $\mu_n, \mu \in \mathcal{P}(\mathbb{R})$. Suppose μ is compactly supported. If $\int x^p \mu_n(dx) \to \int x^p \mu(dx)$ for all $p \geq 1$, then $\mu_n \to \mu$.

The first technique we shall use to show Wigner's semicircle law is the method of moments as applied to L_n. Since $\mu_{s.c}$ is compactly supported, exercise 17 shows that it is sufficient to prove that $\int x^p L_n(dx) \to \int x^p \mu_{s.c}(dx)$ for all p. The key observation is the formula

$$\int x^p L_n(dx) = \frac{1}{n} \sum_{k=1}^{n} \lambda_k^p = \frac{1}{n} \text{tr}(X/\sqrt{n})^p = \frac{1}{n^{1+\frac{p}{2}}} \sum_{i_1,\ldots,i_p=1}^{n} X_{i_1,i_2} \ldots X_{i_p,i_1} \tag{4}$$

which links spectral quantities to sums over entries of the matrix X. By taking expectations, we also get

$$\int x^p \overline{L}_n(dx) = \frac{1}{n} \sum_{k=1}^{n} \lambda_k^p = \frac{1}{n} \text{tr}(X/\sqrt{n})^p = \frac{1}{n^{1+\frac{p}{2}}} \sum_{i_1,\ldots,i_p=1}^{n} \mathbf{E}[X_{i_1,i_2} \ldots X_{i_p,i_1}] \tag{5}$$

which will help in showing that $\overline{L}_n \to \mu_{s.c.}$. We first carry out the method of moments for the expected ESD of a GOE matrix, and later go on to the more involved statement about the ESD of a general Wigner matrix. The first goal is to see how the semicircle distribution arises.

The idea is to use the formula (5) and evaluate the expectation on the right hand side with the help of the Wick formula of exercise 2. The rest of the work is in keeping track of the combinatorics to see how the semicircle moments emerge. To get the idea, we first do it by hand for a few small values of q in (5). We work with the GOE matrix X. Remember that $X_{i,i} \sim N(0,2)$ and $X_{i,j} \sim N(0,1)$ for $i < j$.

(i) Case, q=1. $\mathbf{E}[X_{i,j}X_{j,i}] = 1$ for $j \neq i$ and 2 for $j = i$. Hence $\mathbf{E}[\mathrm{tr}(X^2)] = 2n + 2\binom{n}{2} = n^2 + n$ and

$$\int x^2\, \overline{L}_n(dx) = \frac{1}{n^2}\mathbf{E}[\mathrm{tr}X^2] = 1.$$

(ii) Case $q = 2$. From the Wick formula for real Gaussians, $\mathbf{E}[X_{i,j}X_{j,k}X_{k,\ell}X_{\ell,i}]$ becomes

$$\begin{aligned} &= \mathbf{E}[X_{i,j}X_{j,k}]\mathbf{E}[X_{k,\ell}X_{\ell,i}] + \mathbf{E}[X_{i,j}X_{k,\ell}]\mathbf{E}[X_{j,k}X_{\ell,i}] + \mathbf{E}[X_{i,j}X_{\ell,i}]\mathbf{E}[X_{j,k}X_{k,\ell}] \\ &= (\delta_{i,k} + \delta_{i,j,k}) + (\delta_{i,k}\delta_{j,\ell} + \delta_{i,\ell}\delta_{j,k})(\delta_{i,k}\delta_{j,\ell} + \delta_{i,j}\delta_{k,\ell}) + (\delta_{j,\ell} + \delta_{i,j,\ell})(\delta_{j,\ell} + \delta_{j,k,\ell}) \end{aligned}$$

corresponding to the three matchings $\{\{1,2\},\{3,4\}\}$, $\{\{1,3\},\{2,4\}\}$, $\{\{1,4\},\{2,3\}\}$ respectively. Observe that the diagonal entries are also taken care of, since their variance is 2. This looks messy, but look at the first few terms. When we sum over all i, j, k, ℓ, we get

$$\sum_{i,j,k,\ell} \delta_{i,k} = n^3, \qquad \sum_{i,j,k,\ell} \delta_{i,j,k} = n^2, \qquad \sum_{i,j,k,\ell} (\delta_{i,k}\delta_{j,\ell})^2 = n^2.$$

It is clear that what matters is how many of the indices i, j, k, ℓ are forced to be equal by the delta functions. The more the constraints, the smaller the contribution upon summing. Going back, we can see that only two terms ($\delta_{i,k}$ in the first summand and $\delta_{j,\ell}$ term in the third summand) contribute n^3, while the other give n^2 or n only.

$$\int x^4\, \overline{L}_n(dx) = \frac{1}{n^3}\mathbf{E}[\mathrm{tr}X^4] = \frac{1}{n^3}\sum_{i,j,k,\ell}(\delta_{i,k} + \delta_{j,\ell}) \;+\; \frac{1}{n^3}O(n^2) = 2 + O(n^{-1}).$$

Observe that the two non-crossing matchings $\{\{1,2\},\{3,4\}\}$ and $\{\{1,4\},\{2,3\}\}$ contributed 1 each, while the crossing-matching $\{\{1,3\},\{2,4\}\}$ contributed zero in the limit. Thus, recalling exercise 2, we find that $\int x^4\, \overline{L}_n(dx) \to \int x^4\, \mu_{s.c.}(dx)$

(iii) Case $q = 3$. We need to evaluate $\mathbf{E}[X_{i_1,i_2}X_{i_2,i_3}\ldots X_{i_6,i_1}]$. By the wick formula, we get a sum over matching of $[6]$. Consider two of these matchings.

(a) $\{1,4\},\{2,3\},\{5,6\}$: This is a non-crossing matching. We get

$$\begin{aligned} &\mathbf{E}[X_{i_1,i_2}X_{i_4,i_5}]\mathbf{E}[X_{i_2,i_3}X_{i_3,i_4}]\mathbf{E}[X_{i_5,i_6}X_{i_6,i_1}] \\ =\; &(\delta_{i_1,i_4}\delta_{i_2,i_5} + \delta_{i_1,i_5}\delta_{i_2,i_4})(\delta_{i_2,i_4} + \delta_{i_2,i_3,i_4})(\delta_{i_5,i_1} + \delta_{i_5,i_1,i_6}) \\ =\; &\delta_{i_1,i_5}\delta_{i_2,i_4} + [\ldots]. \end{aligned}$$

When we sum over $i_1, \ldots, i_6$, the first summand gives n^4 while all the other terms (pushed under $[\ldots]$) give $O(n^3)$. Thus the contribution from this matching is $n^4 + O(n^3)$.

(b) $\{1,5\},\{2,6\},\{3,4\}$: A crossing matching. We get which is equal to

$$\begin{aligned} &\mathbf{E}[X_{i_1,i_2}X_{i_5,i_6}]\mathbf{E}[X_{i_2,i_3}X_{i_6,i_1}]\mathbf{E}[X_{i_3,i_4}X_{i_4,i_5}] \\ =\; &(\delta_{i_1,i_5}\delta_{i_2,i_6} + \delta_{i_1,i_6}\delta_{i_2,i_5})(\delta_{i_2,i_6}\delta_{i_3,i_1} + \delta_{i_2,i_1}\delta_{i_3,i_6})(\delta_{i_3,i_5} + \delta_{i_3,i_4,i_5}) \end{aligned}$$

It is easy to see that all terms are $O(n^3)$. Thus the total contribution from this matching is $O(n^3)$.

We leave it as an exercise to check that all crossing matchings of $[6]$ give $O(n^3)$ contribution while the non-crossing ones give $n^4 + O(n^3)$. Thus,

$$\int x^6\, \overline{L}_n(dx) = \frac{1}{n^4}\mathbf{E}[\mathrm{tr}X^6] = \frac{1}{n^4}(C_6 n^4 + O(n^3)) \to C_6 = \int x^6 \mu_{s.c}(dx).$$

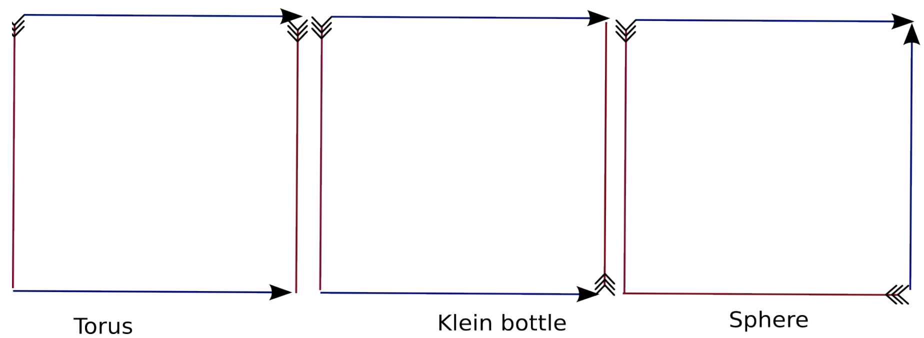

FIGURE 1. Three of the four surfaces that can be got by gluing a quadrilateral.

3. Expected ESD of GOE or GUE matrix goes to semicircle

Proposition 18. *Let $X = (X_{i,j})_{i,j\le n}$ be the GOE matrix and let L_n be the ESD of $X_n/\sqrt{n}$. Then $\overline{L}_n \to \mu_{s.c}$.*

To carry out the convergence of moments $\int x^{2q}\overline{L}_n(dx) \to \int x^{2q}\mu(dx)$ for general q, we need some preparation in combinatorics.

Definition 19. Let P be a polygon with $2q$ vertices labeled $1,2,\dots,2q$. A *gluing of P* is a matching of the edges into pairs along with an assignment of sign $\{+,-\}$ to each matched pair of edges. Let $\mathcal{M}^{\dagger}_{2q}$ denote the set of all gluings of P. Thus, there are $2^q(2q-1)!!$ gluings of a polygon with $2q$ sides.

Further, let us call a gluing $M \in \mathcal{M}^{\dagger}_{2q}$ to be *good* if the underlying matching of edges is non-crossing and the orientations are such that matched edges are oriented in opposite directions. That is, $[r,r+1]$ can be matched by $[s+1,s]$ but not with $[s,s+1]$. The number of good matchings is C_q, by part (3) of exercise 9.

Example 20. Let P be a quadrilateral with vertices $1,2,3,4$. Consider the gluing $M = \{\{[1,2],[4,3]\},\{[2,3],[1,4]\}\}$. It means that the edge $[1,2]$ is identified with $[4,3]$ and the edge $[2,3]$ is identified with $[1,4]$. If we actually glue the edges of the polygon according to these rules, we get a torus[2]. The gluing $M' = \{\{[1,2],[3,4]\},\{[2,3],[1,4]\}\}$ is different from M. What does the gluing give us? We identify the edges $[2,3]$ and $[1,4]$ as before, getting a cylinder. Then we glue the two circular ends in *reverse* orientation. Hence the resulting surface is Klein's bottle. See Figure 1.

For a polygon P and a gluing M, let V_M denote the number of distinct vertices in P after gluing by M. In other words, the gluing M gives an equivalence relationship on the vertices of P, and V_M is the number of equivalence classes.

Lemma 21. *Let P be a polygon with $2q$ edges and let $M \in \mathcal{M}^{\dagger}_{2q}$. Then $V_M \le q+1$ with equality if and only if M is good.*

Assuming the lemma we prove the convergence of $\overline{L}_n$ to semicircle.

[2]Informally, gluing means just that. Formally, gluing means that we fix homeomorphism $f:[1,2]\to[3,4]$ such that $f(1)=3$ and $f(2)=4$ and a homeomorphism $g:[2,3]\to[1,4]$ such that $g(2)=1$ and $g(3)=4$. Then define the equivalences $x\sim f(x)$, $y\sim g(y)$. The resulting quotient space is what we refer to as the glued surface. It is locally homeomorphic to $\mathbb{R}^2$ which justifies the word "surface". The quotient space does not depend on the choice of homeomorphisms f and g. In particular, if we reverse the orientations of all the edges, we get the same quotient space.

PROOF OF PROPOSITION 18.

$$\begin{aligned}\mathbf{E}[X_{i_1,i_2}\dots X_{i_{2q},i_1}] &= \sum_{M\in\mathcal{M}_{2q}}\prod_{\{r,s\}\in M}\mathbf{E}[X_{i_r,i_{r+1}}X_{i_s,i_{s+1}}]\\ &= \sum_{M\in\mathcal{M}_{2q}}\prod_{\{r,s\}\in M}(\delta_{i_r,i_s}\delta_{i_{r+1},i_{s+1}}+\delta_{r,s+1}\delta_{r+1,s})\\ &= \sum_{M\in\mathcal{M}^{\dagger}_{2q}}\prod_{\{e,f\}\in M}\delta_{i_e,i_f}. \qquad (6)\end{aligned}$$

Here for two edges e,f, if $e=[r,r+1]$ and $s=[s,s+1]$ (or $f=[s+1,s]$), then δ_{i_e,i_f} is just $\delta_{i_r,i_s}\delta_{i_{r+1},i_{s+1}}$ (respectively $\delta_{i_r,i_{s+1}}\delta_{i_{r+1},i_s}$). Also observe that diagonal entries are automatically taken care of since they have have variance 2 (as opposed to variance 1 for off-diagonal entries).

Sum (6) over $i_1,\dots,i_{2q}$ and compare with Recall (5) to get

$$\int x^{2q}\overline{L}_n(dx)=\frac{1}{n^{1+q}}\sum_{M\in\mathcal{M}^{\dagger}_{2q}}\sum_{i_1,\dots,i_{2q}}\prod_{\{e,f\}\in M}\delta_{i_e,i_f}=\frac{1}{n^{1+q}}\sum_{M\in\mathcal{M}^{\dagger}_{2q}}n^{V_M}. \qquad (7)$$

We explain the last equality. Fix M, and suppose some two vertices r,s are identified by M. If we choose indices $i_1,\dots,i_{2q}$ so that some $i_r\neq i_s$, then the δ-functions force the term to vanish. Thus, we can only choose one index for each equivalence class of vertices. This can be done in n^{V_M} ways.

Invoke Lemma 21, and let $n\to\infty$ in (7). Good matchings contribute 1 and others contribute zero in the limit. Hence, $\lim_{n\to\infty}\int x^{2q}\overline{L}_n(dx)=C_q$. The odd moments of $\overline{L}_n$ as well as $\mu_{s.c}$ are obviously zero. By exercise 5, and employing exercise 17 we conclude that $\overline{L}_n\to\mu_{s.c}$. ■

It remains to prove Lemma 21. If one knows a little algebraic topology, this is clear. First we describe this "high level picture". For the benefit of those not unfamiliar with Euler characteristic and genus of a surface, we give a self-contained proof later[3].

A detour into algebraic topology: Recall that a *surface* is a topological space in which each point has a neighbourhood that is homeomorphic to the open disk in the plane. For example, a polygon (where we mean the interior of the polygon as well as its boundary) is not a surface, since points on the boundary do not have disk-like neighbourhoods. A sphere, torus, Klein bottle, projective plane are all surfaces. In fact, these can be obtained from the square P_4 by the gluing edges appropriately.

(1) Let $P=P_{2q}$ and $M\in\mathcal{M}^{\dagger}_{2q}$. After gluing P according to M, we get a surface (means a topological space that is locally homeomorphic to an open disk in the plane) which we denote P/M. See examples 20.
(2) If we project the edges of P via the quotient map to P/M, we get a graph G_M drawn (or "embedded") on the surface P/M. A graph is a combinatorial object, defined by a set of vertices V and a set of edges E. An *embedding of a graph* on

[3]However, the connection given here is at the edge of something deep. Note the exact formula for GOE $\int t^{2q}d\overline{L}_n(t)=\sum_{g=0}^{q}n^{-g}A_{q,g}$, where $A_{q,g}$ is the number of gluings of P_{2q} that lead to a surface with Euler characteristic $2-2g$. The number g is called the genus. The right hand side can be thought of as a generating function for the number $A_{q,g}$ in the variable n^{-1}. This, and other related formulas express generating functions for maps drawn on surfaces of varying genus in terms of Gaussian integrals over hermitian matrices, which is what the left side is. In particular, such formulas have been used to study "random quadrangulations of the sphere", and other similar objects, using random matrix theory. Random planar maps are a fascinating and active research are in probability, motivated by the notion of "quantum gravity" in physics.

a surface is a collection of function $f : V \to S$ and $f_e : [0,1] \to S$ for each $e \in E$ such that f is one-one, for $e = (u,v)$ the function f_e is a homeomorphism such that $f_e(0) = f(u)$ and $f_e(1) = f(v)$, and such that $f_e((0,1))$ are pairwise disjoint. For an embedding, each connected component of $S \setminus \cup_{e \in E} f_e[0,1]$ is called a *face*. A *map* is an embedding of the graph such hat each face is homeomorphic to a disk.

(3) For any surface, there is a number χ called the *Euler characteristic* of the surface, such that for any map drawn on the surface, $V - E + F = \chi$, where V is the number of vertices, E is the number of edges and F is the number of faces of the graph. For example, the sphere has $\chi = 2$ and the torus has $\chi = 0$. The Klein bottle also has $\chi = 0$. The *genus* of the surface is related to the Euler characteristic by $\chi = 2 - 2g$.

(4) A general fact is that $\chi \leq 2$ for any surface, with equality if and only if the surface is simply connected (in which case it is homeomorphic to the sphere).

(5) The graph G_M has $F = 1$ face (the interior of the polygon is the one face, as it is homeomorphically mapped under the quotient map), $E = q$ edges (since we have merged $2q$ edges in pairs) and $V = V_M$ vertices. Thus, $V_M = \chi(G_M) - 1 + q$. By the previous remark, we get $V_M \leq q + 1$ with equality if and only if P/M is simply connected.

(6) Only good gluings lead to simply connected P/M.

From these statements, it is clear that Lemma 21 follows. However, for someeone unfamiliar with algebraic topology, it may seem that we have restated the problem without solving it. Therefore we give a self-contained proof of the lemma now.

PROOF OF LEMMA 21. After gluing by M, certain vertices of P are identified. If $V_M > q$, there must be at least one vertex, say r, of P that was not identified with any other vertex. Clearly, then M must glue $[r-1, r]$ with $[r, r+1]$. Glue these two edges, and we are left with a polygon Q with $2q - 2$ sides with an edge sticking out. For r to remain isolated, it must not enter the gluing at any future stage. This means, the gluing will continue within the polygon Q. Inductively, we conclude that Q must be glued by a good gluing. Retracing this to P, we see that M must be a good gluing of P. Conversely, if M is a good gluing, it is easy to see that $V_M = q + 1$[4]. ■

Exercise 22. Show that the expected ESD of the GUE matrix also converges to $\mu_{s.c.}$.

4. Wishart matrices

The methods that we are going to present, including the moment method, are applicable beyond the simplest model of Wigner matrices. Here we remark on what we get for Wishart matrices. Most of the steps are left as exercises.

Definition 23. Let $m < n$ and let $X_{m \times n}$ be a random matrix whose entries are i.i.d. If $\mathbf{E}[X_{i,j}] = 0$ and $\mathbf{E}[|X_{i,j}|^2] = 1$, we say that the $m \times m$ matrix $A = XX^*$ is a *Wishart matrix*. If in addition, $X_{i,j}$ are i.i.d $N(0,1)$ (or $CN(0,1)$), then A is called a real (or complex, respectively) Gaussian Wishart matrix.

[4]Thanks to R. Deepak for this neat proof. Another way to state it is as follows. Consider the polygon P (now a topological space homeomorphic to the closed disk). Glue it by M to get a quotient space P/M. Consider the graph G formed by the edges of P (so G is a cycle). Project to G to P/M. The resulting graph G_M is connected (since G was), and has q edges. Hence it can have at most $q+1$ vertices, and it has $q+1$ vertices if and only if the G_M is a tree. Work backwards to see that M must be good. The induction step is implicit in proving that a graph has $V \leq E + 1$ with equality for and only for trees.

Note that X is not hermitian, but A is. The positive square roots of the eigenvalues of A are called the *singular values* of X. Then the following is true.

Theorem 24. *Let $X_{m,n}$ be a real or complex Gaussian Wishart matrix. Suppose m and n go to infinity in such a way that $m/n \to c$ for a finite positive constant c. Let L_n be the ESD of A_n/n. Then, the expected ESD $\overline{L}_n \to \mu^c_{m.p}$ which is the* Marcenko-Pastur distribution, *defined as the probability measure with density*

$$\frac{d\mu^c_{m.p}(t)}{dt} = \frac{1}{2\pi c}\frac{\sqrt{(b-t)(t-a)}}{t}, \qquad b = 1+\sqrt{c},\ a = 1-\sqrt{c},\ \textit{for } t \in [a,b].$$

Exercise 25. Prove Theorem 24.
Hint: The following trick is not necessary, but often convenient. Given an $m \times n$ matrix X, define the $(m+n)\times(m+n)$ matrix

$$B = \begin{bmatrix} 0_{m\times m} & X_{m\times n} \\ X^t_{n\times m} & 0_{n\times n} \end{bmatrix}.$$

Assume $m \leq n$. By exercise 26 below, to study the ESD of $A = XX^*$, one might as well study the ESD of B.

Exercise 26. For A and B as in the hint for the previous exercise, suppose $m < n$. If s_k^2, $k \leq m$ are the eigenvalues of A, then the eigenvalues of B are $\pm s_k$, $k \leq m$ together with $n-m$ zero eigenvalues.