18.338J/16.394J: The Mathematics of Infinite Random Matrices
The Stieltjes transform based approach

Raj Rao

Handout #4, Thursday, September 16, 2004.

1 The eigenvalue distribution function

For an $N \times N$ matrix A_N, the eigenvalue distribution function [1] (e.d.f.) $F^{A_N}(x)$ is defined as

$$F^{A_N}(x) = \frac{\text{Number of eigenvalues of } A_N \leq x}{N}. \tag{1}$$

As defined, the e.d.f. is right continuous and possibly atomic i.e. with step discontinuities at discrete points. In practical terms, the derivative of (1), referred to as the (eigenvalue) level density, is simply the appropriately normalized histogram of the eigenvalues of A_N. The `MATLAB` code `histn` we distributed earlier approximates this density.

A surprising result in infinite RMT is that for some matrix ensembles, the expectation $E[F^{A_N}(x)]$ has a well defined i.e. not zero and not infinite limit. We drop the notational dependence on N in (1) by defining the limiting e.d.f. as

$$F^A(x) = \lim_{N \to \infty} E[F^{A_N}(x)]. \tag{2}$$

This limiting e.d.f. [2] is also sometimes referred to in literature as the integrated density of states [2, 3]. Its derivative is referred to as the level density in physics literature [4]. The region of support associated with this limiting density is simply the region where $dF^A(x) \neq 0$. When discussing the limiting e.d.f. we shall often distinguish between, its atomic and non-atomic components.

2 The Stieltjes transform representation

One step removed from the e.d.f. is the Stieltjes transform which has proved to be an efficient tool for determining this limiting density. For all non-real z the Stieltjes (or Cauchy) transform of the probability measure $F^A(x)$ is given by

$$m_A(z) = \int \frac{1}{x - z} dF^A(x) \qquad \text{Im } z \neq 0. \tag{3}$$

The integral above is over the whole [3] or some subset of the real axis since for the matrices of interest, such as the Hermitian or real symmetric matrices, the eigenvalues are real. When we refer to the "Stieltjes transform of A" in this paper, we are referring to $m_A(z)$ defined as in (3) expressed in terms of the limiting density $dF^A(x)$ of the random matrix ensemble A.

---

[1] This is also referred to in literature as the empirical distribution function [1].

[2] Unless we state otherwise any reference to an e.d.f. or the level density. in this paper will refer to the corresponding *limiting* e.d.f. or density respectively.

[3] While the Stieltjes integral is over the positive real axis, the Cauchy integral is more general [5] and can include complex contours as well. This distinction is irrelevant for several practical classes of matrices, such as the sample covariance matrices, where all of the eigenvalues are non-negative. Nonetheless, throughout this paper, (3) will be referred to as the Stieltjes transform with the implicit assumption that the integral is over the *entire* real axis.

The Stieltjes transform in (3) can also be interpreted as an expectation with respect to the measure $F^A(x)$ such that

$$m_A(z) = E_X\left[\frac{1}{x-z}\right]. \tag{4}$$

Since there is a one-to-one correspondence between the probability measure $F^A(x)$ and the Stieltjes transform, convergence of the Stieltjes transform can be used to show the convergence of the probability measure $F^A(x)$. Once this convergence has been established, the Stieltjes transform can be used to yield the density using the so-called Stieltjes-Perron inversion formula [6]

$$\frac{dF^A(x)}{dx} = \frac{1}{\pi}\lim_{\xi\to 0} \operatorname{Im}\, m_A(x+i\xi). \tag{5}$$

When studying the limiting distribution of large random matrices, the Stieltjes transform has proved to be particularly relevant because of its correspondence with the matrix resolvent. The trace of the matrix resolvent, $M_A(z)$, defined as $M_A(z) = (A_N - zI)^{-1}$ can be written as

$$\operatorname{tr}[M_A(z)] = \sum_{i=1}^{N} \frac{1}{\lambda_i - z} \tag{6}$$

where λ_i for $i = 1, 2, \ldots, N$ are the eigenvalues of A_N. For any A_N, $M_A(z)$ is a non-random quantity. However, when A_N is a large random matrix,

$$m_A(z) = \lim_{N\to\infty} \frac{1}{N} tr[M_A(z)]. \tag{7}$$

The Stieltjes transform and its resolvent form in (7) are intimately linked to the classical moment problem [6]. This connection can be observed by noting that the integral in (3) can be expressed as an analytic "multipole" series expansion about $z = \infty$ such that

$$\begin{aligned} m_A(z) &= -\int \sum_{k=0}^{\infty} \frac{x^k}{z^{k+1}} dF^A(x) = -\sum_{k=0}^{\infty} \int \frac{x^k}{z^{k+1}} dF^A(x) \\ &= -\frac{1}{z} - \sum_{k=1}^{\infty} \frac{M_k^A}{z^{k+1}}. \end{aligned} \tag{8}$$

where $M_k^A = \int x^k dF^A(x)$ is the k^{th} moment of x on the probability measure $dF^A(x)$. The analyticity of the Stieltjes transform about $z = \infty$ expressed in (8) is a consequence of our implicit assumption that the region of support for the limiting density $dF^A(x)$ is bounded i.e. $\lim_{x\to\infty} dF^A(x) = 0$.

Incidentally, the η-transform introduced by Tulino and Verdù in [7] can be expressed in terms of $m(z)$ and permits a series expansion about $z = 0$.

Given the relationship in (7), it is worth noting that the matrix moment M_k^A is simply

$$M_k^A = \lim_{N\to\infty} \frac{1}{N} \operatorname{tr}[A_N^k] = \int x^k dF^A(x). \tag{9}$$

Equation (8), written as a multipole series, suggests that a way of computing the density would be to determine the moments of the random matrix as in (9), and then invert the Stieltjes transform using (5). For the famous semi-circular law, Wigner actually used a moment based approach in [8] to determine the density for the standard Wigner matrix though he did not explicitly invert the Stieltjes transform as we suggested above, As the reader may imagine, such a moment based approach is not particularly useful for more general classes of random matrices. We discuss a more relevant and practically useful Stieltjes transform based approach next.

3 Stieltjes transform based approach

Instead of obtaining the Stieltjes transform directly, the so-called Stieltjes transform approach relies instead on finding a canonical equation that the Stieltjes transform satisfies. The Marčenko-Pastur theorem [9] was the first and remains most famous example of such an approach. We include a statement of its theorem in the form found in literature. We encourage you to write this theorem, as an exericse, in a simpler manner.

Theorem 1 (The Marčenko-Pastur Theorem). *Consider an $N \times N$ matrix, B_N. Assume that*

1. *X_n is an $n \times N$ matrix such that the matrix elements X_{ij}^n are independent identically distributed (i.i.d.) complex random variables with mean zero and variance 1 i.e. $X_{ij}^n \in \mathcal{C}$, $E[X_{ij}^n] = 0$ and $E[\|X_{ij}^n\|^2] = 1$.*
2. *$n = n(N)$ with $n/N \to c > 0$ as $N \to \infty$.*
3. *$T_n = diag(\tau_1^n, \tau_2^n, \ldots, \tau_n^n)$ where $\tau_i^n \in \mathbb{R}$, and the e.d.f. of $\{\tau_1^n, \ldots, \tau_n^n\}$ converges almost surely in distribution to a probability distribution function $H(\tau)$ as $N \to \infty$.*
4. *$B_N = A_N + \frac{1}{N} X_n^* T_n X_n$, where A_N is a Hermitian $N \times N$ matrix for which F^{A_N} converges vaguely to $\mathcal{A}$ almost surely, $\mathcal{A}$ being a possibly defective (i.e. with discontinuities) nonrandom distribution function.*
5. *X_n, T_n, and A_N are independent.*

Then, almost surely, F^{B_N} converges vaguely, almost surely, as $N \to \infty$ to a nonrandom d.f. F^B whose Stieltjes transform $m(z)$, $z \in \mathbb{C}$, satisfies the canonical equation

$$m(z) = m_A\left(z - c \int \frac{\tau \, dH(\tau)}{1 + \tau \, m(z)}\right) \tag{10}$$

We now illustrate the use of this theorem with a representative example. This example will help us highlight issues that will be of pedagogical interest throughout this semester.

Suppose $A_N = 0$ i.e. $B_N = \frac{1}{N} X_n^* T_n X_n$. The Stieltjes transform of A_N, by the definition in (3), is then simply

$$m_A(z) = \frac{1}{0 - z} = -\frac{1}{z}. \tag{11}$$

Hence, using the Marčenko-Pastur theorem as expressed in (10), the Stieltjes transform $m(z)$ of B_N is given by

$$m(z) = -\frac{1}{z - c \int \frac{\tau dH(\tau)}{1 + \tau m(z)}}. \tag{12}$$

Rearranging the terms in this equation and using m instead of $m(z)$ for notational convenience, we get

$$z = -\frac{1}{m} + c \int \frac{\tau dH(\tau)}{1 + \tau m}. \tag{13}$$

Equation (13) expresses the dependence between the Stieltjes transform variable m and probability space variable z. Such a dependence, expressed explicitly in terms of $dH(\tau)$, will be referred to throughout this paper as a canonical equation. Equation (13) can also be interpreted as the expression for the functional inverse of $m(z)$.

To determine the density of B_N by using the inversion formula in (5) we need to first solve (13) for $m(z)$. In order to obtain an equation in m and z we need to first know $dH(\tau)$ in (13). In theory, $dH(\tau)$ could be any density that satisfies the conditions of the Marčenko-Pastur theorem. However, as we shall shortly recognize, for an arbitrary distribution, it might not be possible to obtain an analytical or even an easy numerical solution for the density On the other hand, for some specific distributions of $dH(\tau)$, it will indeed be possible to analytically obtain the density We consider one such distribution below.

Suppose $T_n = I$ i.e. the diagonal elements of T_n are non-random with d.f. $dH(\tau) = \delta(\tau - 1)$. Equation (13) then becomes

$$z = -\frac{1}{m} + \frac{c}{1 + m}. \tag{14}$$

Rearranging the terms in the above equation we get

$$zm(1+m) = -(1+m) + cm \tag{15}$$

which, with a bit of algebra, can be written as

$$m^2 z + m(1-c+z) + 1 = 0. \tag{16}$$

Equation (16) is a polynomial equation in m whose coefficients are polynomials in z. We will refer to such polynomials, often derived from canonical equations as Stieltjes (transform) polynomials for the remainder of this paper.

As discussed, to obtain the density using (5) we need to first solve (16) for m in terms of z. Since, from (16), we have a second degree polynomial in m it is indeed possible to analytically solve for its roots and obtain the density.

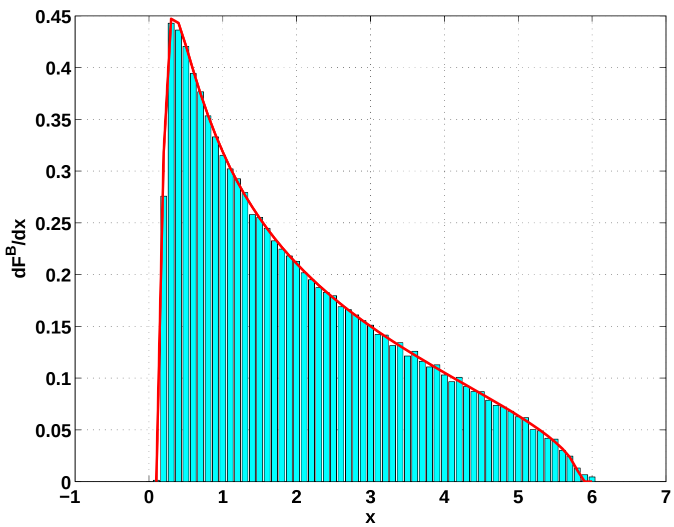

Figure 1: Level density for $B_N = (1/N)X_n^* X_n$ with $c = 2$.

This level density, sometimes referred to in the literature as the Marčenko-Pastur distribution, is given by

$$\frac{dF^B(x)}{dx} = \max\left(0, 1-c\right)\delta(x) + \frac{\sqrt{(x-b_-)(b_+-x)}}{2\pi x} I_{[\ b_-, b_+]} \tag{17}$$

where $b_\pm = (1 \pm \sqrt{c})^2$ and $I_{[b_-,b_+]}$ is the indicator function that is equal to 1 for $b_- < z < b_+$ and 0 elsewhere.

Figure 1 compares the histogram of the eigenvalues of 1000 realizations of the matrix $B_N = \frac{1}{N} X_n^* X_n$ with $N = 100$ and $n = 200$ with the solid line indicating the theoretical density given by (17) for $c = n/N = 2$.

We now consider a modification to the Marčenko-Pastur theorem that is motivated by the sample covariance matrices that appear often in array processing applications.

3.1 The Sample Covariance Matrix

In the previous section we used the Marčenko-Pastur theorem to examine the density of a class of random matrices $B_N = \frac{1}{N} X_n^* T_n X_n$. Suppose we defined the $N \times n$ matrix, $Y_N = X_n^* T_n^{1/2}$, then B_N may be written as

$$B_N = \frac{1}{N} Y_N Y_N^*. \tag{18}$$

Recall that T_n was assumed to be diagonal and non-negative definite so $T_n^{1/2}$ can be constructed uniquely up to the sign. If Y_N were to be interpreted as a matrix of observations, then B_N written as (18) is reminiscent

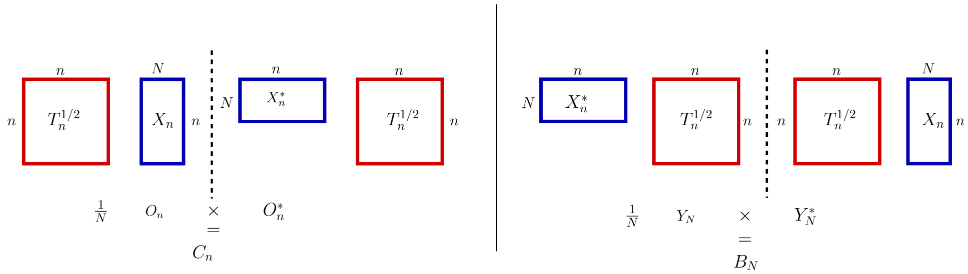

Figure 2: The matrices B_n and C_N when $n > N$.

of sample covariance matrices that appear in many engineering and statistical applications. However, among other things, it is subtly different because of the normalization of the $Y_N Y_N^*$ by the number of rows N of Y_N rather than by the number of columns n. Hence, we need to come up with a definition of a sample covariance matrix that mirrors the manner in which it is used in practical applications.

With engineering, particularly signal processing applications in mind, we introduce the $n \times N$ matrix, $O_n = T_n^{1/2} X_n$ and define

$$C_n = \frac{1}{N} O_n O_n^* \tag{19}$$

to be the *sample covariance matrix* (SCM). By comparing (18) and (19) while recalling the definitions of Y_N and O_n, it is clear that the eigenvalues of B_n are related to the eigenvalues of C_n. For B_N of the form in (18), the Marčenko-Pastur theorem can be used to obtain the canonical equation for $m_B(z)$ given by (13). Recall, that by $m_B(z)$ we mean the Stieltjes transform associated with the limiting e.d.f. $F^B(x)$ of B_N as $N \to \infty$. There is however, an exact relationship between the non-limiting e.d.f.'s $F^{B_N}(x)$ and the $F^{C_n}(x)$ and hence the corresponding Stieltjes transforms $m_{B_N}(z)$ and $m_{C_n}(z)$ respectively. We exploit this relationship below to derive the canonical equation for C_n from the canonical equation for B_N given in (13).

Figure 2 schematically depicts C_n and B_N when $n > N$ i.e. when $c > 1$. In this case, C_n, as denoted in the figure, will have $n - N$ zero eigenvalues. The other N eigenvalues of C_n will, however, be *identically* equal to the N eigenvalues of B_N. Hence, the e.d.f. of C_n can be *exactly* expressed in terms of the e.d.f. of B_N as

$$F^{C_n}(x) = \left(\frac{n-N}{N}\right) I_{(0,\infty]} + \frac{n}{N} F^{B_N}(x) \tag{20}$$

$$= (c-1) I_{(0,\infty]} + c F^{B_N}(x). \tag{21}$$

Recalling the definition of the Stieltjes transform in (3), this implies that the Stieltjes transform $m_{C_n}(z)$ of C_n is related to the Stieltjes transform $m_{B_N}(z)$ of B_N by the expression

$$m_{C_n}(z) = -\frac{c-1}{z} + c\, m_{B_N}(z). \tag{22}$$

Similarly, Figure 3 schematically depicts C_n and B_N when $n < N$ i.e. $c < 1$. In this case, B_N, as denoted in the figure, will have $N - n$ zero eigenvalues. The other n eigenvalues of B_N will, however, be *identically* equal to the n eigenvalues of C_n. Hence, as before, the e.d.f. of B_N can be *exactly* expressed in terms of the e.d.f. of C_n as

$$F^{B_N}(x) = \left(\frac{N-n}{n}\right) I_{(0,\infty]} + \frac{N}{n} F^{C_n}(x) \tag{23}$$

$$= \left(\frac{1}{c} - 1\right) I_{(0,\infty]} + \frac{1}{c} F^{C_n}(x) \tag{24}$$

Once again, recalling the definition of the Stieltjes transform in (3), this implies that the Stieltjes transform $m_{C_n}(z)$ of C_n is related to the Stieltjes transform $m_{B_N}(z)$ of B_N by the expression

$$m_{B_N}(z) = -\left(\frac{1}{c} - 1\right)\frac{1}{z} + \frac{1}{c} m_{C_n}(z). \tag{25}$$

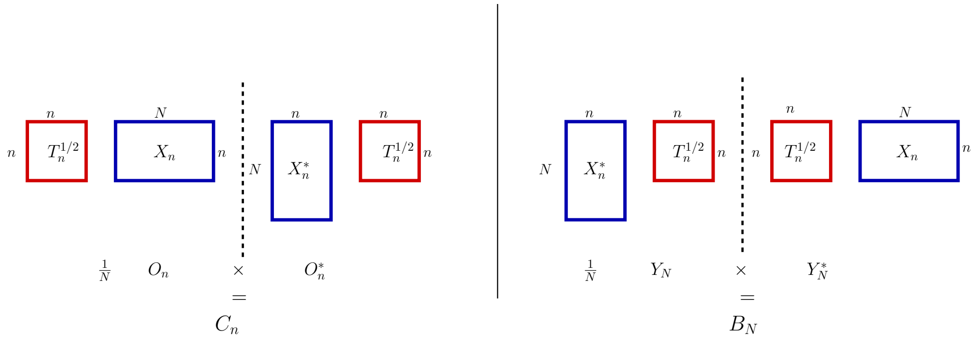

Figure 3: The matrices B_n and C_N when $n < N$.

Multiplying both sides of (25) by c, we get

$$c\, m_{B_N}(z) = -(1-c)\frac{1}{z} + m_{C_n}(z) \tag{26}$$

which upon rearranging the terms is precisely (22). Equation (22) is an *exact* expression relating the Stieltjes transforms of C_n and B_N for *all* n and N. As $N \to \infty$, the Marčenko-Pastur theorem states that $m_{B_N}(z) \to m_B(z)$ which implies that the limiting Stieltjes transform $m_C(z)$, for C_n can be written in terms of $m_B(z)$ using (22) as

$$m_C(z) = -\frac{c-1}{z} + c\, m_B(z). \tag{27}$$

For our purpose of getting the canonical equation for $m_C(z)$ using the canonical equation for $m_B(z)$, it is more useful to rearrange the terms in (27) and express $m_B(z)$ can be written in terms of $m_C(z)$. This relationship is simply

$$m_B(z) = -\left(\frac{1}{c} - 1\right)\frac{1}{z} + \frac{1}{c}\, m_C(z). \tag{28}$$

Hence, to obtain the canonical equation for $m_C(z)$ we simply have to substitute the expression for $m_B(z)$ in (28) into (13). With some fairly straightforward algebra, that we shall omit here, it can be verified that $m_C(z)$ is the solution to the canonical equation

$$m_C(z) = \int \frac{dH(\tau)}{\{(1 - c - c\, z\, m_C(z)\}\tau - z}. \tag{29}$$

Incidentally, (29) was first derived by Silverstein in [10]. He noted that the eigenvalues of $\frac{1}{N}T_n^{1/2}X_nX_n^*T_n^{1/2}$ were the same as those of $\frac{1}{N}X_nX_n^*T_n$ so that (29) was the canonical equation for this class of matrices as well. Additionally, in their proof of the Marčenko-Pastur theorem in [11], Bai and Silverstein dropped the restriction on T_n being diagonal so that T_n could be any matrix whose e.d.f. $F^{T_n} \to H$. As the reader may appreciate, this broadens the class of matrices for which this theorem may be applied.

3.2 The (generalized) Wishart matrix

Revisiting the previous example, when $T_n = I$, $C_n = \frac{1}{N}X_nX_n^*$. This matrix C_n is the generalized version of the famous Wishart matrix ensemble first studied by Wishart in 1928 [12]. In physics literature, the Wishart matrix is also referred to as the Laguerre Ensemble [13]. Strictly speaking, C_n is referred to as a Wishart matrix only when the elements of X_n are i.i.d. Gaussian random variables. The canonical equation for C_n in (29) becomes

$$m = \frac{1}{(1 - c - c\, z\, m) - z} \tag{30}$$

which upon rearranging yields the Stieltjes polynomial

$$c\, z\, m^2 - (1 - c - z)m + 1 = 0. \tag{31}$$

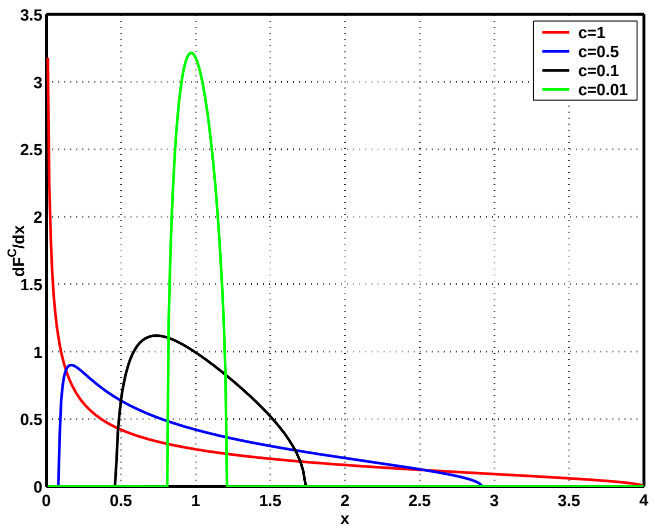

Figure 4: The density of the (generalized) Wishart matrix $\{W_n(c)\}$ for different values of c.

Once again, (31) is a second degree polynomial in m whose coefficients are polynomials in z. Hence, as before, (31) can be solved analytically to yield a solution for m in terms of z. The inversion formula in (5) can then be used to obtain the limiting density for C_n. This is simply

$$\frac{dF^C(x)}{dx} = \max\left(0, 1 - \frac{1}{c}\right)\delta(x) + \frac{\sqrt{(x-b_-)(b_+-x)}}{2\pi x c} I_{[\ b_-, b_+]} \tag{32}$$

where, as before, $b_\pm = (1 \pm \sqrt{c})^2$. As (32) suggests, the limiting density for C_n depends only on the parameter c. Hence, for the remainder of this paper, we will use the notation $\{W(c)\}$ to denote the (generalized) Wishart matrix ensemble [4] defined as $W(c) = \frac{1}{N} X_n X_n^*$ with $c = n/N > 0$.

Figure 4 plots the density in (32) for different values of c. From (32) the reader may notice that as $c \to 0$ i.e. for a fixed n as $N \to \infty$, both the largest and the smallest eigenvalue, b_+ and b_- respectively, approach 1. This validates our intuition about the behavior of $W(c)$ for a fixed n as $N \to \infty$. Additionally, from Figure 4, the reader may notice that the density gets increasingly more symmetrical about $x = 1$ as the value of c decreases. For $c = 0.01$, the density is almost perfectly symmetrical about $x = 1$. The reader may note that with an appropriate scaling and translation, the density of $\{W(c)\}$ as $c \to 0$ could be made to resemble the semi-circular distribution. In fact, in [14] Jonsson used this observation and the correspondence between the distribution in (32) and the generalized β-distribution to infer the moments of $W_n(c)$ from the even moments of the Wigner matrix which are incidentally the Catalan numbers denoted by C_k for an integer k. More recently, Dumitriu recognized [15] that these moments could be written in terms of the (k, r) Narayana numbers [16] defined as

$$N_{k,r} = \frac{1}{r+1}\binom{k}{r}\binom{k-1}{r} \tag{33}$$

so that the individual moments may be obtained from the moment generating function

$$M_k^W(c) = \sum_{r=0}^{k-1} c^r N_{k,r} = \sum_{r=0}^{k-1} c^r \binom{k}{r}\binom{k-1}{r} \tag{34}$$

for which, it may be noted that $M_k^W(1) = C_k = M_{2k}^S$ are also the even moments of the standard Wigner matrix.

[4]For notational convenience, when discussing infinite (generalized) Wishart matrices we will simply refer to $\{W(c)\}$ as the Wishart matrix even when the elements of X_n are i.i.d. but not Gaussian. We will occasionally add a subscript such as $W_1(c)$ to differentiate between different realizations of the ensemble $\{W(c)\}$. When discussing finite Wishart matrices, we will implicitly assume that the elements of X_n are i.i.d. Gaussian. Its use in either manner will be obvious from the context.

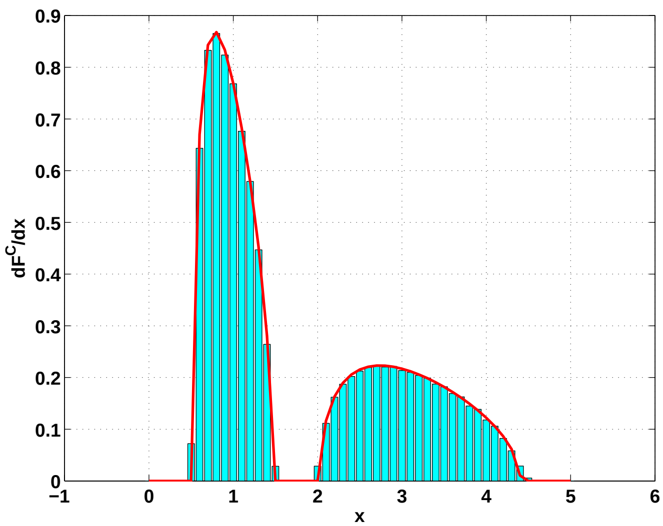

Figure 5: The density of C_n with $c = n/N = 0.1$ and $dH(\tau) = 0.6\,\delta(\tau - 1) + 0.4\,\delta(\tau - 3)$.

The Wishart matrix $\{W(c)\}$ has been studied exhaustively by statisticians and engineers. For infinite Wishart matrices, (32) captures the limiting density and the extreme eigenvalues i.e. the region of support. The limiting moments are given by (34). As it was for the asymptotic Wigner matrix, the limiting density of the asymptotic Wishart matrix did not depend on whether the elements of X_n were real or complex. Similarly, as it was for the Gaussian ensembles, the analytical behavior of the finite Wishart matrices did indeed depend on whether the elements were real or complex. Nonetheless, the limiting moment and eigenvalue behavior could be inferred from the behavior of the finite (real or complex) Wishart matrix counterpart. The reader is directed towards some of the representative literature and the references therein on the distribution of the smallest [17], largest [18, 19, 20, 21], sorted [19, 22, 23], unsorted eigenvalues [19, 22, 23], and condition numbers [22, 23] of the Wishart matrix that invoke this link between the finite and infinite Wishart matrix ensembles. We will now discuss sample covariance matrices for which the behavior of the limiting density can be best, if not solely, analyzed analytically using the Marčenko-Pastur theorem.

3.3 Additional examples of sample covariance matrices

Suppose $dH(\tau) = p\,\delta(\tau - \lambda_1) + (1-p)\,\delta(\tau - \lambda_2)$ i.e. T_n has an atomic mass of weight p at λ_1 and another atomic mass of weight $(1-p)$ at λ_2. The canonical equation in (29) becomes

$$m = \frac{p}{\lambda_1(1-c-cmz) - z} + \frac{1-p}{\lambda_2(1-c-cmz) - z} \tag{35}$$

which upon rearranging yields the Stieltjes polynomial

$$\begin{aligned}\lambda_1\, c^2 m^3 z^2 \lambda_2 &+ \left(-2\lambda_1\lambda_2 cz + \lambda_1 cz^2 + 2\lambda_1 c^2\lambda_2 z + \lambda_2 cz^2\right) m^2 \\ &+ \left(\lambda_1\lambda_2 + \lambda_2 cz + p\lambda_2 cz - \lambda_1 z + \lambda_1 c^2\lambda_2 + z^2 - \lambda_2 z + 2\lambda_1 cz - p\lambda_1 cz - 2\lambda_1\lambda_2\, c\right) m \\ &- \left(p\lambda_2 + z - p\lambda_1 c + \lambda_1 c - \lambda_1 + p\lambda_2 c + p\lambda_1\right) = 0. \end{aligned} \tag{36}$$

It can be readily verified that if $p = 1$ and $\lambda_1 = 1$, then $C_n = \frac{1}{N} T_n^{1/2} X_n X_n^* T_n^{1/2}$ is simply the Wishart matrix we discussed above. Though it might not seem so from a cursory look, for $p = 1$ and $\lambda_1 = 1$, (36) can be shown after some elementary factorization to simplify to (31). Since (36) is a third degree polynomial in m it can conceivably be solved analytically using Cardano's formula.

For general c, p, λ_1 and λ_2 this is cumbersome and cannot be solved analytically as a function of z and c for arbitrary values of p, λ_1, and λ_2. However, for specific values of c, p, λ_1, and λ_2 we can numerically solve the resulting Stieltjes polynomial in (36). For example, when $p = 0.6$, $\lambda_1 = 1$ and $\lambda_2 = 3$, (36) simplifies to

$$3\,c^2 m^3 z^2 + \left(4\,cz^2 - 6\,cz + 6\,c^2 z\right) m^2 + \left(3 - 4\,z - 6\,c + \frac{31}{5}\,cz + 3\,c^2 + z^2\right) m + \frac{11}{5}\,c + z - \frac{11}{5} = 0. \tag{37}$$

For $c = 0.1$, (37) becomes

$$\frac{3}{100}\,m^3 z^2 + \left(-\frac{27}{50}\,z + 2/5\,z^2\right) m^2 + \left(\frac{243}{100} - \frac{169}{50}\,z + z^2\right) m - \frac{99}{50} + z = 0 \tag{38}$$

To determine the density from (38) we need to determine the roots of the polynomial in m and use the inversion formula in (5). Since we do not know the region of support for this density we would conjecture such a region and basically solve the polynomial above for every value of z. Using numerical tools such as the roots command in MATLAB this is not very difficult.

Figure 5 shows the excellent agreement between the theoretical density (solid line) obtained from numerically solving (38) and the histogram of the eigenvalues of 1000 realizations of the matrix C_n with $n = 100$ and $N = n/c = 1000$. Figure 6 shows the behavior of the density for a range of values of c. This figure captures our intuition that as $c \to 0$, the eigenvalues of the sample covariance matrix C_n will be increasingly localized about $\lambda_1 = 1$ and $\lambda_2 = 3$. By contrast, capturing this very same analytic behavior using finite RMT is not as straightforward.

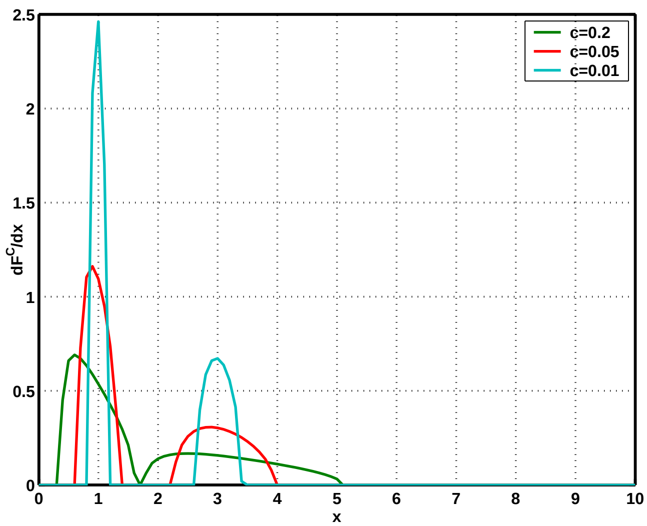

Figure 6: The density of C_n with $dH(\tau) = 0.6\,\delta(\tau - 1) + 0.4\,\delta(\tau - 3)$ for different values of c

Unlike the Wishart matrix, the distribution function (or level density) of finite dimensional covariance matrices, such as the C_n we considered in this example, can only be expressed in terms of zonal or other multivariate orthogonal polynomials that appear frequently in texts such as [19]. Though these polynomials have been studied extensively by multivariate statisticians [24, 25, 26, 27, 28, 29, 30, 31] and more recently by combinatorists [32, 33, 34, 35, 36, 37] the prevailing consensus is that they are unwieldy and not particularly intuitive to work with. This is partly because of their definition as infinite series expansions which makes their numerical evaluation a non-trivial task when dealing with matrices of moderate dimensions. More importantly, from an engineering point of view, the Stieltjes transform based approaches allows us to generate plots of the form in Figure 6 with the implicit assumption that the matrix in question is infinite and yet, predict the behavior of the eigenvalues for the practical finite matrix counterpart with remarkable accuracy, as Figure 5 corroborates. This is the primary motivation for this course's focus on developing *infinite* random matrix theory.

In the lectures that follow we will discuss other techniques that allow us to characterize a very broad class of infinite random matrices that cannot be characterized using finite RMT. We will often be intrigued by and speculate on the link between these infinite matrix ensembles and their finite matrix counterparts. We encourage you to ask us questions on this or to explore them further.

4 Exericses

1. Verified that Silverstein's sample covariance matrix theorem can be inferred directly from the Marčenko-Pastur theorem. Note: You will have to do some substitution tricks to get the parameter "c" to refer to the same quantity.

2. Derive the moments of the Wishart matrix from the observation that as $c \to 0$, the density becomes "approximately" semi-circular. Hint: you will have to make an approximation for the region of support while remembering that for any $a < 1$, $a^2 < a$. There will also be a shifting and rescaling in this problem to get the terms to match up correctly. Recall that the moments of the Wishart matrix are expressed in terms of the Narayana polynomials in (34).

3. Come up with numerical code to compute the theoretical density when $dH(\tau)$ has three atomic masses (e.g. $dH(\tau) = 0.4\,\delta(\tau-1) + 0.4\,\delta(\tau-3) + 0.2\,\delta(\tau-7)$). Plot the limiting density as a function of x for a range of values of c.

4. Do the same when there are four atomic masses in $dH(\tau)$ (e.g. $dH(\tau) = 0.3\,\delta(\tau-1) + 0.25\,\delta(\tau-3) + 0.25\,\delta(\tau-7) + 0.25\,\delta(\tau-10)$) . Verify that the solution obtained matches up with the simulations. Hints: Do *all* the roots match up?

5. What happens if there are atomic masses of *negative* weight in $dH(\tau)$ (e.g. $dH(\tau) = 0.5\,\delta(\tau+1) + 0.5\,\delta(\tau-1)$). Does the limiting theoretical density line up with the experimental results? Check the assumptions of the Marčenko-Pastur theorem! Is this "allowed"?

References

[1] Sang Il Choi and Jack W. Silverstein. Analysis of the limiting spectral distribution of large dimensional random matrices. volume 54(2), pages 295–309. 1995.

[2] Boris A. Khoruzenkho, A.M. Khorunzhy, and L.A. Pastur. Asymptotic properties of large random matrices with independent entries. *Journal of Mathematical Physics*, 37:1–28, 1996.

[3] L. Pastur. A simple approach to global regime of random matrix theory. In *Mathematical results in statistical mechanics*, Marseilles, 1998, pages 429–454. World Sci. Publishing, River Edge, NJ, 1999.

[4] Madan Lal Mehta. *Random Matrices*. Academic Press, Boston, second edition, 1991.

[5] M. Abramowitz and I.A. Stegun, editors. *Handbook of Mathematical Functions*. Dover Publications, New York, 1970.

[6] N. I. Akhiezer. *The classical moment problem and some related questions in analysis*. Hafner Publishing Co., New York, New York, 1965. Translated by N. Kemmer.

[7] A. M. Tulino and S. Verdú. Random matrices and wireless communications. *Foundations and Trends in Communications and Information Theory*, 1(1), June 2004.

[8] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Math.*, 62:548–564, 1955.

[9] V.A. Marčenko and L.A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Math USSR Sbornik*, 1:457–483, 1967.

[10] Jack W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *Journal of Multivariate Analysis*, 55(2):331–339, 1995.

[11] Z. D. Bai and J. W. Silverstein. On the empirical distribution of eigenvalues of a class of large dimensional random matices. *Journal of Multivariate Analysis*, 54(2):175–192, 1995.

[12] J. Wishart. The generalized product moment distribution in samples from a normal multivariate population. *Biometrika*, 20 A:32–52, 1928.

[13] Ioana Dumitriu and Alan Edelman. Matrix Models for Beta Ensembles. *Journal of Mathematical Physics*, (11):5830–5847, November 2002.

[14] Dag Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *J. of Multivar. Anal.*, 12:1–38, 1982.

[15] Ioana Dumitriu and Etienne Rassart. Path counting and random matrix theory. *Electronic Journal of Combinatorics*, 7(7), 2003. R-43.

[16] Robert A. Sulanke. Refinements of the Narayana numbers. *Bull. Inst. Combin. Appl.*, 7:60–66, 1993.

[17] Alan Edelman. The distribution and moments of the smallest eigenvalue of a random matrix of Wishart type. *Lin. Alg. Appl.*, 159:55–80, 1991.

[18] T. Sugiyama. On the distribution of the largest latent root of the covariance matrix. *Annals of Math. Statistics*, 38:1148–1151, 1967.

[19] Robb J. Muirhead. *Aspects of Multivariate Statistical Theory.* John Wiley & Sons, New York, 1982.

[20] Craig A. Tracy and Harold Widom. The Distribution of the Largest Eigenvalue in the Gaussian Ensembles. In J.F. van Diejen and L.Vinet, editors, *Calogero-Moser-Sutherland Models, CRM Series in Mathematical Physics*, volume 4, pages 461–472. Springer-Verlag, Berlin, 2000.

[21] Iain M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, 29(2):295–327, 2001.

[22] Alan Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. on Matrix Analysis and Applications*, 1988:543–560, 1988.

[23] Alan Edelman. *Eigenvalues and Condition Numbers of Random Matrices.* PhD thesis, Massachusetts Institute of Technology, 1989.

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

[25] Zhimin Yan. Generalized hypergeometric functions and Laguerre polynomials in two variables. *Contemp. Math.*, 138:239–259, 1992.

[26] Zhimin Yan. A class of generalized hypergeometric functions in several variables. *Canad. J. Math*, 44:1317–1338, 1992.

[27] P. Bleher and Alexander Its. Semiclassical asymptotics of orthogonal polynomials, Riemann-Hilbert problem, and the universality in the matrix model. *Ann. of Math.*, 150:185–266, 1999.

[28] Percy Deift, T. Kriecherbauer, K. McLaughlin, S. Venakides, and Xin Zhou. Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory. *Comm. Pure Appl. Math.*, 52, 1999.

[29] I.G. Macdonald. *Symmetric Functions and Hall Polynomials.* Oxford Mathematical Monographs. Oxford University Press, 2nd edition, 1998.

[30] Kenneth I. Gross and Donald St. P. Richards. Special functions of matrix argument In: algebraic induction, zonal polynomials, and hypergeometric functions. *Trans. American Math. Society*, 301:781–811, 1987.

[31] Kenneth I. Gross and Donald St. P. Richards. Total positivity, spherical series, and hypergeometric functions of matrix argument. *J. Approx. Theory*, 59, no. 2:224–246, 1989.

[32] Andrei Okounkov and Grigori Olshanski. Asymptotics of Jack polynomials as the number of variables goes to infinity. *IMRN*, 13:641–682, 1998.

[33] Luc Lapointe and Luc Vinet. A Rodrigues formula for the Jack polynomials and the Macdonald-Stanley conjecture.

[34] F. Knop and S. Sahi. A recursion and a combinatorial formula for the Jack polynomials. *Invent. Math.*, 128:9–22, 1997.

[35] Friedrich Knop and Siddhartha Sahi. A recursion and a combinatorial formula for Jack polynomials. *Inventiones Mathematicae*, 128:9–22, 1997.

[36] Richard P. Stanley. Some combinatorial properties of Jack symmetric functions. *Adv. in Math.*, 77:76–115, 1989.

[37] Philip J. Hanlon, Richard P. Stanley, and John R. Stembridge. Some combinatorial aspects of the spectra of normally distributed random matrices. In *Hypergeometric functions on domains of positivity, Jack polynomials, and applications (Tampa, FL, 1991)*, volume 138 of *Contemp. Math.*, pages 151–174. Amer. Math. Soc., Providence, RI, 1992.