15

Addition and Multiplication: Recipes and Examples

In the second part of this book, we have built the necessary tools to compute the spectrum of sums and products of free random matrices. In this chapter we will review the results previously obtained and show how they work on concrete, simple examples. More sophisticated examples, and some applications to real world data, will be developed in subsequent chapters.

15.1 Summary

We introduced the concept of freeness which can be summarized by the following intuitive statement: two large matrices are free if their eigenbases are related by a random rotation. In particular a large matrix drawn from a rotationally invariant ensemble is free with respect to any matrix independent of it, for example a deterministic matrix.[1] For example $\mathbf{A}$ and $\mathbf{OBO}^T$ are free when $\mathbf{O}$ is a random rotation matrix (in the large dimension limit). When $\mathbf{A}$ and $\mathbf{B}$ are free, their R- and S-transforms are, respectively, additive and multiplicative:

$$R_{\mathbf{A}+\mathbf{B}}(x) = R_{\mathbf{A}}(x) + R_{\mathbf{B}}(x), \qquad S_{\mathbf{AB}}(t) = S_{\mathbf{A}}(t) S_{\mathbf{B}}(t). \tag{15.1}$$

The free product needs some clarification as $\mathbf{AB}$ is in general not a symmetric matrix, the S-transform $S_{\mathbf{AB}}(t)$ in fact relates to the eigenvalues of the matrix $\sqrt{\mathbf{A}}\mathbf{B}\sqrt{\mathbf{A}}$, which are the same as those of $\sqrt{\mathbf{B}}\mathbf{A}\sqrt{\mathbf{B}}$ when both $\mathbf{A}$ and $\mathbf{B}$ are positive semi-definite (otherwise the square-root is ill defined).

15.1.1 R- and S-Transforms

The R- and S-transforms are defined by the following relations:

$$\mathfrak{g}_{\mathbf{A}}(z) = \tau\left[(z - \mathbf{A})^{-1}\right], \tag{15.2}$$

$$\mathfrak{t}_{\mathbf{A}}(\zeta) = \tau\left[(1 - \zeta^{-1}\mathbf{A})^{-1}\right] - 1 = \zeta\,\mathfrak{g}_{\mathbf{A}}(\zeta) - 1; \tag{15.3}$$

$$R_{\mathbf{A}}(x) = \mathfrak{z}_{\mathbf{A}}(x) - \frac{1}{x}, \quad S_{\mathbf{A}}(t) = \frac{t+1}{t\,\zeta_{\mathbf{A}}(t)} \text{ if } \tau(\mathbf{A}) \neq 0, \tag{15.4}$$

[1] By *large*, we mean that all normalized moments computed using freeness are correct up to corrections that are $O(1/N)$.

https://doi.org/10.1017/9781108768900.019 Published online by Cambridge University Press

where $\mathfrak{z}_{\mathbf{A}}(x)$ and $\zeta_{\mathbf{A}}(t)$ are the inverse functions of $\mathfrak{g}_{\mathbf{A}}(z)$ and $\mathfrak{t}_{\mathbf{A}}(\zeta)$, respectively.

Under multiplication by a scalar they behave as

$$R_{\alpha\mathbf{A}}(x) = \alpha R_{\mathbf{A}}(\alpha x), \qquad S_{\alpha\mathbf{A}}(t) = \alpha^{-1} S_{\mathbf{A}}(t). \tag{15.5}$$

While the R-transform behaves simply under a shift by a scalar,

$$R_{\mathbf{A}+\alpha\mathbf{1}}(x) = \alpha + R_{\mathbf{A}}(x), \tag{15.6}$$

there is no simple formula for computing the S-transform of a shifted matrix. On the other hand, the S-transform is simple under matrix inversion:

$$S_{\mathbf{A}^{-1}}(x) = \frac{1}{S_{\mathbf{A}}(-x-1)}. \tag{15.7}$$

The two transforms are related by the following equivalent identities:

$$S_{\mathbf{A}}(t) = \frac{1}{R_{\mathbf{A}}(t S_{\mathbf{A}}(t))}, \quad R_{\mathbf{A}}(x) = \frac{1}{S_{\mathbf{A}}(x R_{\mathbf{A}}(x))}. \tag{15.8}$$

The identity matrix has particularly simple transforms:

$$\mathfrak{g}_{\mathbf{1}}(z) = \frac{1}{z-1} \qquad \mathfrak{t}_{\mathbf{1}}(\zeta) = \frac{1}{\zeta - 1}; \tag{15.9}$$

$$R_{\mathbf{1}}(x) = 1, \qquad S_{\mathbf{1}}(t) = 1. \tag{15.10}$$

The R- and S-transforms have the following Taylor expansion for small arguments:

$$R_{\mathbf{A}}(x) = \kappa_1 + \kappa_2 x + \kappa_3 x^2 + \cdots, \qquad S_{\mathbf{A}}(x) = \frac{1}{\kappa_1} - \frac{\kappa_2}{\kappa_1^3} x + \frac{2\kappa_2^2 - \kappa_1\kappa_3}{\kappa_1^5} x^2 + \cdots, \tag{15.11}$$

where κ_n are the free cumulants of $\mathbf{A}$:

$$\kappa_1 = \tau(\mathbf{A}), \qquad \kappa_2 = \tau(\mathbf{A}^2) - \tau^2(\mathbf{A}), \qquad \kappa_3 = \tau(\mathbf{A}^3) - 3\tau(\mathbf{A})\tau(\mathbf{A}^2) + 2\tau^3(\mathbf{A}). \tag{15.12}$$

Combining Eqs. (15.7) and (15.11), we can obtain the inverse moments of $\mathbf{A}$ from its S-transform. In particular,

$$\tau(\mathbf{A}^{-1}) = S_{\mathbf{A}}(-1), \quad \tau(\mathbf{A}^{-2}) = S_{\mathbf{A}}(-1)\left(S_{\mathbf{A}}(-1) - S'_{\mathbf{A}}(-1)\right). \tag{15.13}$$

15.1.2 Computing the Eigenvalue Density

The R-transform provides a systematic way to obtain the spectrum of the sum $\mathbf{C}$ of two independent matrices $\mathbf{A}$ and $\mathbf{B}$, where at least one of them is rotationally invariant. Here is a simple recipe to compute the eigenvalue density of a free sum of matrices:

1 Find $\mathfrak{g}_{\mathbf{B}}(z)$ and $\mathfrak{g}_{\mathbf{A}}(z)$.
2 Invert $\mathfrak{g}_{\mathbf{B}}(z)$ and $\mathfrak{g}_{\mathbf{A}}(z)$ to get $\mathfrak{z}_{\mathbf{B}}(g)$ and $\mathfrak{z}_{\mathbf{A}}(g)$, and hence $R_{\mathbf{B}}(x)$ and $R_{\mathbf{A}}(x)$.

https://doi.org/10.1017/9781108768900.019 Published online by Cambridge University Press

3 $R_{\mathbf{C}}(x) = R_{\mathbf{B}}(x) + R_{\mathbf{A}}(x)$, which gives $\mathfrak{z}_{\mathbf{C}}(g) = R_{\mathbf{C}}(g) - g^{-1}$.
4 Solve $\mathfrak{z}_{\mathbf{C}}(g) = z$ for $\mathfrak{g}_{\mathbf{C}}(z)$.
5 Use Eq. (2.47) to find the density:

$$\rho_{\mathbf{C}}(\lambda) = \frac{\lim_{\eta\to 0^+} \operatorname{Im} \mathfrak{g}_{\mathbf{C}}(\lambda - \mathrm{i}\eta)}{\pi}. \tag{15.14}$$

In the multiplicative case ($\mathbf{C} = \mathbf{A}^{\frac{1}{2}}\mathbf{B}\mathbf{A}^{\frac{1}{2}}$), the recipe is similar:

1 Find $\mathfrak{t}_{\mathbf{B}}(\zeta)$ and $\mathfrak{t}_{\mathbf{A}}(\zeta)$.
2 Invert $\mathfrak{t}_{\mathbf{B}}(\zeta)$ and $\mathfrak{t}_{\mathbf{A}}(\zeta)$ to get $\zeta_{\mathbf{B}}(t)$ and $\zeta_{\mathbf{A}}(t)$, and hence $S_{\mathbf{B}}(t)$ and $S_{\mathbf{A}}(t)$.
3 $S_{\mathbf{C}}(t) = S_{\mathbf{B}}(t)S_{\mathbf{A}}(t)$, which gives $\zeta_{\mathbf{C}}(t)S_{\mathbf{C}}(t)t = t + 1$.
4 Solve $\zeta_{\mathbf{C}}(t) = \zeta$ for $\mathfrak{t}_{\mathbf{C}}(\zeta)$.
5 Equation (2.47) for $\mathfrak{g}_{\mathbf{C}}(z) = (\mathfrak{t}_{\mathbf{C}}(z) + 1)/z$ is equivalent to

$$\rho_{\mathbf{C}}(\lambda) = \frac{\lim_{\eta\to 0^+} \operatorname{Im} \mathfrak{t}_{\mathbf{C}}(\lambda - \mathrm{i}\eta)}{\pi\lambda}. \tag{15.15}$$

In some cases, the equation in step 4 is exactly solvable. But it is usually a high order polynomial equation, or worse, a transcendental equation. In these cases numerical solution is still possible. There always exists at least one solution that satisfies

$$\mathfrak{g}(z) = z^{-1} + O(z^{-2}) \tag{15.16}$$

for $z \to \infty$. Since the eigenvalues of $\mathbf{B}$ and $\mathbf{A}$ are real, their R- and S-transforms are real for real arguments. Hence the equation in step 4 is an equation with real coefficients. In order to find a non-zero eigenvalue density we need to find solutions with a strictly positive imaginary part when the parameter η goes to zero. When the equation is quadratic or cubic, complex solutions come in complex conjugated pairs: therefore, at most one solution will have a strictly positive imaginary part. As a numerical trick, $\pi\rho(\lambda)$ can be equated with the maximum of the imaginary part of all two or three solutions (the density will be zero when all solutions are real). For higher order polynomial and transcendental equations, we have to be more careful as there can be spurious complex solutions with positive imaginary part.

Exercise 15.2.1 shows how to do these computations in concrete cases.

15.2 R- and S-Transforms and Moments of Useful Ensembles

15.2.1 Wigner Ensemble

The Wigner ensemble is rotationally invariant, therefore a Wigner matrix is free from any matrix from which it is independent. For a Wigner matrix $\mathbf{X}$ of variance σ^2, the R-transform reads

$$R_{\mathbf{X}}(x) = \sigma^2 x. \tag{15.17}$$

The Wigner matrix is stable under free addition, i.e. the free sum of two Wigner matrices of variance σ_1^2 and σ_2^2 is a Wigner with variance $\sigma_1^2 + \sigma_2^2$.

https://doi.org/10.1017/9781108768900.019 Published online by Cambridge University Press

The Wigner matrix is traceless ($\tau(\mathbf{X}) = 0$), so its S-transform is ill-defined. However, we can shift the mean of the entries of $\mathbf{X}$ by a certain parameter m. We then have $R_{\mathbf{X}+m}(x) = m + \sigma^2 x$. We can use Eq. (11.97) and compute the S-transform:

$$S_{\mathbf{X}+m}(t) = \frac{\sqrt{m^2 + 4\sigma^2 t} - m}{2\sigma^2 t} = \frac{m}{2\sigma^2 t}\left(\sqrt{1 + \frac{4\sigma^2 t}{m^2}} - 1\right). \tag{15.18}$$

It is regular at $t = 0$ whenever $m > 0$ and tends to $(\sigma\sqrt{t})^{-1}$ when $m \to 0$.

Finally, let us recall the formula for the positive moments of Wigner matrices:

$$\tau(\mathbf{X}^{2k}) = \frac{(2k)!}{(k+1)k!^2}\sigma^{2k}; \qquad \tau(\mathbf{X}^{2k+1}) = 0. \tag{15.19}$$

The negative moments of $\mathbf{X}$ are all infinite, because the density of zero eigenvalues is positive.

15.2.2 Wishart Ensemble

For a white Wishart matrix $\mathbf{W}_q$ with parameter $q = N/T$, one has (see Section 10.1)

$$R_{\mathbf{W}_q}(x) = \frac{1}{1 - qx}. \tag{15.20}$$

To compute its S-transform we first remember that its Stieltjes transform $\mathfrak{g}(z)$ satisfies Eq. (4.37), which can be written as an equation for $\mathfrak{t}(\zeta)$ or its inverse $\zeta(t)$:

$$\zeta t - (1 + qt)(t + 1) = 0 \quad \Rightarrow \quad S_{\mathbf{W}_q}(t) = \frac{1}{1 + qt}. \tag{15.21}$$

The first few moments of a white Wishart matrix are given by

$$\tau(\mathbf{W}_q) = 1, \quad \tau(\mathbf{W}_q^2) = 1 + q; \qquad \tau(\mathbf{W}_q^{-1}) = \frac{1}{1-q}, \quad \tau(\mathbf{W}_q^{-2}) = \frac{1}{(1-q)^3}. \tag{15.22}$$

15.2.3 Inverse-Wishart Ensemble

We take the opportunity of this summary of R- and S-transforms to introduce a very useful ensemble of matrices, namely the inverse-Wishart ensemble. We will call an inverse-Wishart matrix[2] the inverse of a white Wishart matrix, which, we recall, has unit normalized trace.

For a Wishart matrix to be invertible we need to have $q < 1$. Let $\mathbf{W}_q$ be such a matrix. Using Eq. (11.116) we can show that

$$S_{\mathbf{W}_q^{-1}}(t) = 1 - q - qt. \tag{15.23}$$

[2] More generally the inverse of a Wishart matrix with any covariance $\mathbf{C}$ can be called an inverse-Wishart but we will only consider *white* inverse-Wishart matrices.

https://doi.org/10.1017/9781108768900.019 Published online by Cambridge University Press

Since $\tau(\mathbf{W}_q^{-1}) = 1/(1-q)$, we define the (normalized) inverse-Wishart as $\mathbf{M}_p = (1-q)\mathbf{W}_q^{-1}$ and call $p := q/(1-q)$. Rescaling and changing variable we obtain

$$S_{\mathbf{M}_p}(t) = 1 - pt. \tag{15.24}$$

By construction the inverse-Wishart has mean 1 and variance p. Using Eq. (15.11), we find that it has $\kappa_3(\mathbf{M}_p) = 2p^2$, which is higher than the skewness of a white Wishart matrix with the same variance ($\kappa_3(\mathbf{W}_q) = q^2$).

From the S-transform we can find the R-transform using Eq. (15.8):

$$R_{\mathbf{M}_p}(x) = \frac{1 - \sqrt{1-4px}}{2px}. \tag{15.25}$$

To find the Stieltjes transform and the density, it is easier to compute the T-transform from Eq. (11.63) and convert the result into a Stieltjes transform:

$$\mathfrak{g}_{\mathbf{M}_p}(z) = \frac{(1+2p)z - 1 - \sqrt[\oplus]{(z-1)^2 - 4pz}}{2pz^2}. \tag{15.26}$$

We can use Eq. (2.47) to find the density of eigenvalues or do the following change of variable in the white Wishart density (Eq. (4.43)):

$$x = \frac{1-q}{\lambda} \quad \text{and} \quad p = \frac{q}{1-q}. \tag{15.27}$$

Both methods give (see Fig. 15.1)

$$\rho_{\mathbf{M}_p}(x) = \frac{\sqrt{(x_+ - x)(x - x_-)}}{2\pi p x^2}, \quad x_- < x < x_+, \tag{15.28}$$

with the edges of the spectrum given by

$$x_\pm = 2p + 1 \pm 2\sqrt{2(p+1)}. \tag{15.29}$$

From the Stieltjes transform we can obtain

$$\tau(\mathbf{M}_p^{-1}) = -\lim_{z\to 0} \mathfrak{g}_{\mathbf{M}_p}(z) = 1 + p. \tag{15.30}$$

Other low moments of the inverse-Wishart matrix read

$$\tau(\mathbf{M}_p) = 1, \quad \tau(\mathbf{M}_p^2) = 1 + p; \qquad \tau(\mathbf{M}_p^{-1}) = 1 + p, \quad \tau(\mathbf{M}_p^{-2}) = (1+p)(1+2p). \tag{15.31}$$

Finally, the large N inverse-Wishart matrix potential can be obtained from the real part of the Stieltjes transform using (5.38)

$$V_{\mathbf{M}_p}(x) = \frac{1}{px} + \frac{1+2p}{p}\log x. \tag{15.32}$$

https://doi.org/10.1017/9781108768900.019 Published online by Cambridge University Press

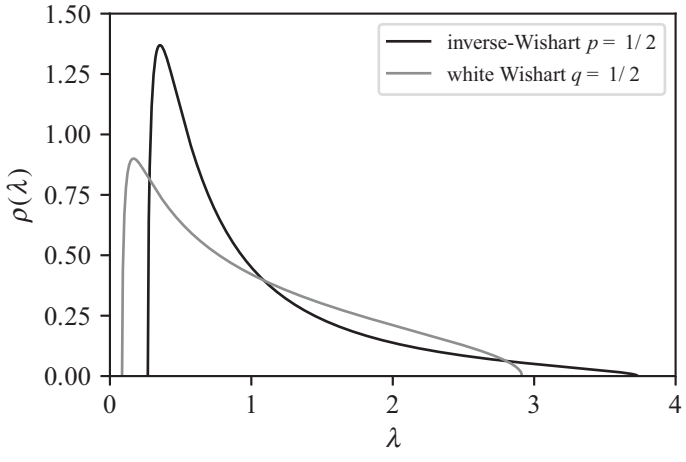

Figure 15.1 Density of eigenvalues for an inverse-Wishart distribution with $p = 1/2$. The white Wishart distribution ($q = 1/2$) is shown for comparison. Both laws are normalized, have unit mean and variance $1/2$.

For completeness we give here the law of the elements of a general (not necessarily white) inverse-Wishart matrix at finite N. We recall the law of a general Wishart matrix Eq. (4.16):

$$P(\mathbf{E}) = \frac{(T/2)^{NT/2}}{\Gamma_N(T/2)} \frac{(\det \mathbf{E})^{(T-N-1)/2}}{(\det \mathbf{C})^{T/2}} \exp\left[-\frac{T}{2} \operatorname{Tr}\left(\mathbf{E}\mathbf{C}^{-1}\right)\right], \tag{15.33}$$

where $\mathbf{E}$ is an $N \times N$ Wishart matrix measured over T time steps with true correlations $\mathbf{C}$ and normalized such that $\mathbb{E}[\mathbf{E}] = \mathbf{C}$. We define the inverse-Wishart as $\mathbf{M} = \mathbf{E}^{-1}$. Note that a finite N Wishart matrix has

$$\mathbb{E}\left[\mathbf{E}^{-1}\right] = \frac{T}{T-N-1}\mathbf{C}^{-1} =: \Sigma, \tag{15.34}$$

where we have defined the matrix Σ such that $\mathbb{E}[\mathbf{M}] = \Sigma$. To do the change of variable $\mathbf{E} \to \mathbf{M}$ in the joint probability density, we need to multiply by the Jacobian $(\det \mathbf{M})^{-N-1}$ (see Eq. (1.41)). Putting everything together we obtain

$$P(\mathbf{M}) = \frac{(T-N-1)^{NT/2}}{2^{NT/2}\Gamma_N(T/2)} \frac{(\det \Sigma)^{T/2}}{(\det \mathbf{M})^{(T+N+1)/2}} \exp\left[-\frac{T-N-1}{2} \operatorname{Tr}\left(\mathbf{M}^{-1}\Sigma\right)\right]. \tag{15.35}$$

In the scalar case $N = 1$, the inverse-Wishart distribution reduces to an inverse-gamma distribution:

$$P(m) = \frac{b^a}{\Gamma(a)} m^{-a-1} \mathrm{e}^{-b/m}, \tag{15.36}$$

with $b = (T-2)\Sigma/2$ and $a = T/2$.

https://doi.org/10.1017/9781108768900.019 Published online by Cambridge University Press

Exercise 15.2.1 Free product of two Wishart matrices

In this exercise, we will compute the eigenvalue distribution of a matrix $\mathbf{E} = (\mathbf{W}_{q_0})^{\frac{1}{2}}\mathbf{W}_q(\mathbf{W}_{q_0})^{\frac{1}{2}}$; as we will see in Section 17.1, this matrix would be the sample covariance matrix of data with true covariance given by a Wishart with q_0 observed over T samples such that $q = N/T$.

(a) Using Eq. (15.21) and the multiplicativity of the S-transform, write the S-transform of $\mathbf{E}$.

(b) Using the definition of the S-transform write an equation for $\mathfrak{t}_{\mathbf{E}}(z)$. It is a cubic equation in t. If either q_0 or q goes to zero, it reduces to the standard Marčenko–Pastur quadratic equation.

(c) Use Eq. (15.15) and a numerical root finder to plot the eigenvalue density of $\mathbf{E}$ for $q_0 = 1/2$ and $q \in \{1/4, 1/2, 3/4\}$. In practice you can work with $\eta = 0$; of the three roots of your cubic equation, at most one will have a positive imaginary part. When all three solutions are real $\rho_{\mathbf{E}}(\lambda) = 0$.

(d) Generate numerically two independent Wishart matrices with $q = 1/2$ ($N = 1000$ and $T = 2000$) and compute $\mathbf{E} = (\mathbf{W}_{q_0})^{\frac{1}{2}}\mathbf{W}_q(\mathbf{W}_{q_0})^{\frac{1}{2}}$. Note that the square-root of a matrix is obtained by applying the square-roots to its eigenvalues. Diagonalize your $\mathbf{E}$ and compare its density with your result in (c).

15.3 Worked-Out Examples: Addition

15.3.1 The Arcsine Law

Consider the free sum of two symmetric orthogonal matrices, i.e. matrices with eigenvalues ± 1 with equal weights. Let $\mathbf{M}_1$ and $\mathbf{M}_2$ be two such matrices, their Stieltjes and R-transforms are given by

$$\mathfrak{g}(z) = \frac{z}{z^2 - 1} \qquad \text{and} \qquad R(g) = \frac{\sqrt{1 + 4g^2} - 1}{2g}, \tag{15.37}$$

from which we can deduce that $\mathbf{M} = \frac{1}{2}(\mathbf{M}_1 + \mathbf{M}_2)$ has an R-transform given by

$$R_{\mathbf{M}}(g) = \frac{\sqrt{1 + g^2} - 1}{g}, \tag{15.38}$$

where we have used the scaling $R_{\alpha\mathbf{A}}(x) = \alpha R_{\mathbf{A}}(\alpha x)$ with $\alpha = 1/2$.

The corresponding Stieltjes transform reads

$$\mathfrak{g}_{\mathbf{M}}(z) = \frac{1}{z\sqrt{1 - 1/z^2}}. \tag{15.39}$$

https://doi.org/10.1017/9781108768900.019 Published online by Cambridge University Press

From this expression we deduce that the density of eigenvalues is given by the centered arcsine law:

$$\rho_{\mathbf{M}}(\lambda) = \frac{1}{\pi}\frac{1}{\sqrt{1-\lambda^2}}, \qquad \lambda \in (-1,1), \tag{15.40}$$

and zero elsewhere. This corresponds to a special case of the Jacobi ensemble that we have encountered in Section 7.1.3.

15.3.2 Sum of Uniform Densities

Suppose now we want to compute the eigenvalue distribution of a matrix $\mathbf{M} = \mathbf{U} + \mathbf{O}\mathbf{U}\mathbf{O}^T$, where $\mathbf{U}$ is a diagonal matrix with entries uniformly distributed between -1 and 1 (e.g. $[\mathbf{U}]_{kk} = 1 + (1-2k)/N$) and $\mathbf{O}$ a random orthogonal matrix. This is the free sum of two matrices with uniform eigenvalue density.

First we need to compute the Stieltjes transform of $\mathbf{U}$. We have

$$\rho_{\mathbf{U}}(\lambda) = \frac{1}{2}, \qquad \lambda \in (-1,1). \tag{15.41}$$

The corresponding Stieltjes transform is[3]

$$\mathfrak{g}_{\mathbf{U}}(z) = \frac{1}{2}\int_{-1}^{1}\frac{\mathrm{d}\lambda}{z-\lambda} = \frac{1}{2}\log\left(\frac{z+1}{z-1}\right). \tag{15.42}$$

Note that when $-1 < \lambda < 1$ the argument of the log in $\mathfrak{g}_{\mathbf{U}}(z)$ is negative so $\operatorname{Im}\mathfrak{g}(\lambda - \mathrm{i}\eta) = \mathrm{i}\pi/2$, consistent with a uniform distribution of eigenvalues. We then compute the R-transform by finding the inverse of $\mathfrak{g}_{\mathbf{U}}(z)$:

$$\mathfrak{z}(g) = \frac{\mathrm{e}^{2g}+1}{\mathrm{e}^{2g}-1} = \coth(g). \tag{15.43}$$

And so the R-transform of $\mathbf{U}$ is given by

$$R_{\mathbf{U}}(g) = \coth(g) - \frac{1}{g}. \tag{15.44}$$

The R-transform of $\mathbf{M}$ is twice that of $\mathbf{U}$. To find the Stieltjes transform of $\mathbf{U}$ we thus need to solve

$$z = R_{\mathbf{M}}(g) + \frac{1}{g} = 2\coth(g) - \frac{1}{g}, \tag{15.45}$$

for $\mathfrak{g}(z)$. This is a transcendental equation and we need to solve it for complex z near the real axis. Before attempting to do this, it is useful to plot $\mathfrak{z}(g)$ (Fig. 15.2). The region where $z = \mathfrak{z}(g)$ does not have real solutions is where the eigenvalues are. This region is between a local maximum and a local minimum of $\mathfrak{z}(g)$. We should look for complex solutions of Eq. (15.45) near the real axis for $\operatorname{Re}(z)$ between -1.54 and 1.54. We can then put this equation into a complex non-linear solver. The density will be given by $\operatorname{Im}\mathfrak{g}(z)/\pi$ for $\operatorname{Im}(z)$

[3] A more general uniform density between $[m-a, m+a]$ has mean m, variance $a^2/3$ and $\mathfrak{g}_{\mathbf{U}}(z) = \log((z-m+a)/(z-m-a))/(2a)$.

https://doi.org/10.1017/9781108768900.019 Published online by Cambridge University Press

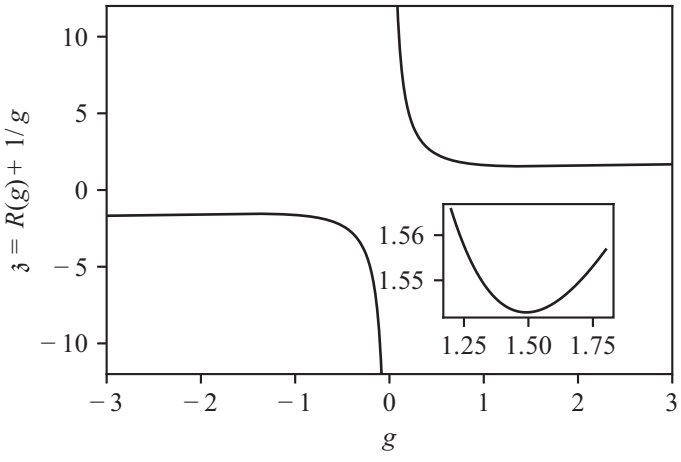

Figure 15.2 The function $\mathfrak{z}(g) = R_{\mathbf{M}}(g) + 1/g$ for the free sum of two flat distributions. Note that there is a region of z near $[-1.5, 1.5]$ when $z = \mathfrak{z}(g)$ does not have real solutions. This is where the eigenvalues lie. The inset shows a zoom of the region near $z = 1.5$, indicating more clearly that $\mathfrak{z}(g)$ has a minimum at $g_+ \approx 1.49$, so $\lambda_+ = \mathfrak{z}(g_+)$. The exact edges of the spectrum are $\lambda_\pm \approx \pm 1.5429$.

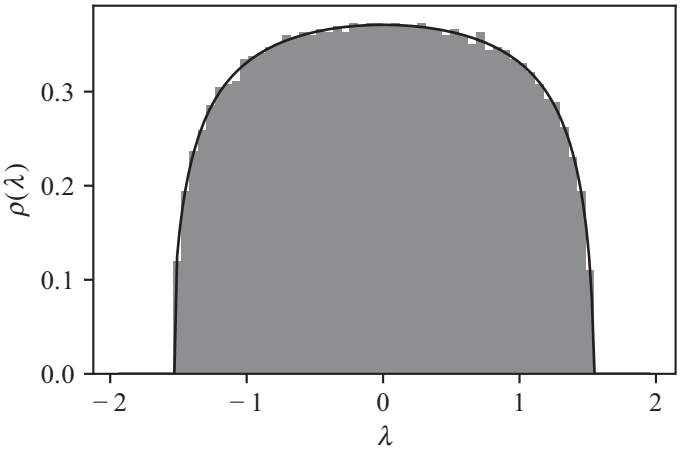

Figure 15.3 Density of eigenvalues for the free sum of two uniform distributions. Continuous curve was computed using a numerical solution of Eq. (15.45). The histogram is a numerical simulation with $N = 5000$.

very small and $\mathrm{Re}(z)$ in the desired range. Note that complex solutions come in conjugated pairs, and it is hard to force the solver to find the correct one. This is not a problem; since their imaginary parts have the same absolute value, we can just use

$$\rho(\lambda) = \frac{|\operatorname{Im} \mathfrak{g}(\lambda - i\eta)|}{\pi} \text{ for some small } \eta. \tag{15.46}$$

We have plotted the resulting density in Figure 15.3.

https://doi.org/10.1017/9781108768900.019 Published online by Cambridge University Press

15.4 Worked-Out Examples: Multiplication

15.4.1 Free Product of a Wishart and an Inverse-Wishart

Consider the free product of a Wishart $\mathbf{W}_q$ of parameter q and an independent inverse-Wishart $\mathbf{M}_p$ of parameter p, i.e. $\mathbf{E} = \sqrt{\mathbf{M}_p}\mathbf{W}_q\sqrt{\mathbf{M}_p}$. We already have the building blocks:

$$S_{\mathbf{M}_p}(t) = 1 - pt; \qquad S_{\mathbf{W}_q}(t) = \frac{1}{1+qt} \quad \Rightarrow \quad S_{\mathbf{E}}(t) = \frac{1-pt}{1+qt}, \tag{15.47}$$

leading to

$$\zeta_{\mathbf{E}}(t) = \frac{(t+1)(1+qt)}{t(1-pt)}. \tag{15.48}$$

Inverting this relation to obtain $\mathrm{t}_{\mathbf{E}}(\zeta)$ leads to a quadratic equation for t:

$$(t+1)(1+qt) = \zeta t(1-pt), \tag{15.49}$$

which can be explicitly solved as

$$\mathrm{t}_{\mathbf{E}}(z) = \frac{z-q-1-\overset{\oplus}{\sqrt{(q+1-z)^2-4(q+zp)}}}{2(q+zp)}. \tag{15.50}$$

Using Eq. (15.15) finally yields

$$\rho_{\mathbf{E}}(\lambda) = \frac{\sqrt{4(p\lambda+q)-(1-q-\lambda)^2}}{2\pi\lambda(p\lambda+q)}. \tag{15.51}$$

The edges of the support are given by

$$\lambda_{\pm} = \left[1+q+2p \pm 2\sqrt{(1+p)(q+p)}\right]. \tag{15.52}$$

One can check that the limit $p \to 0$ recovers the trivial case $\mathbf{M}_0 = \mathbf{1}$ for which the Marčenko–Pastur edges indeed read

$$\lambda_{\pm} = (1+q) \pm 2\sqrt{q}. \tag{15.53}$$

Exercise 15.4.1 Free product of Wishart and inverse-Wishart

(a) Generate numerically a normalized inverse-Wishart $\mathbf{M}_p$ for $p = 1/4$ and $N = 1000$. Check that $\tau(\mathbf{M}_p) = 1$ and $\tau(\mathbf{M}_p^2) = 1.25$. Plot a normalized histogram of the eigenvalues of $\mathbf{M}_p$ and compare with Eq. (15.28).

(b) Generate an independent Wishart $\mathbf{W}_q$ with $q = 1/4$ and compute $\mathbf{E} = \sqrt{\mathbf{M}_p}\mathbf{W}_q\sqrt{\mathbf{M}_p}$. To compute $\sqrt{\mathbf{M}_p}$, diagonalize $\mathbf{M}_p$, take the square-root of its eigenvalues and reconstruct $\sqrt{\mathbf{M}_q}$. Check that $\tau(\mathbf{E}) = 1$ and $\tau(\mathbf{E}^2) = 1.5$. Plot a normalized histogram of the eigenvalues of $\mathbf{E}$ and compare with Eq. (15.51).

https://doi.org/10.1017/9781108768900.019 Published online by Cambridge University Press

(c) For every eigenvector $\mathbf{v}_k$ of $\mathbf{E}$ compute $\xi_k := \mathbf{v}_k^T \mathbf{M}_p \mathbf{v}_k$, and make a scatter plot of ξ_k vs λ_k, the eigenvalue of $\mathbf{v}_k$. Your scatter plot should show a noisy straight line. We will see in Chapter 19, Eq. (19.49), that this is related to the fact that linear shrinkage is the optimal estimator of the true covariance from the sample covariance when the true covariance is an inverse-Wishart.

15.4.2 Free Product of Projectors

As our last simple example, consider a space of large dimension N, and a projector $\mathbf{P}_1$ on a subspace of dimension $N_1 \leq N$, i.e. a diagonal matrix with N_1 diagonal elements equal to unity, and $N - N_1$ elements equal to zero. We now introduce a second projector $\mathbf{P}_2$ on a subspace of dimension $N_2 \leq N$, and would like to study the eigenvalues of the free product of these two projectors, $\mathbf{P} = \mathbf{P}_1 \mathbf{P}_2$. Clearly, all eigenvalues of $\mathbf{P}$ must lie in the interval $(0, 1)$.

As now usual, we first need to compute the T-transform of $\mathbf{P}_1$ and $\mathbf{P}_2$. We define the ratios $q_a = N_a/N$, $a = 1, 2$. Since $\mathbf{P}_a$ has N_a eigenvalues equal to unity and $N - N_a$ eigenvalues equal to zero, one finds

$$g_{\mathbf{P}_a}(z) = \frac{1}{N}\left[\frac{N_a}{z-1} + \frac{N-N_a}{z}\right] = \frac{q_a - 1 + z}{z(z-1)} \quad \Rightarrow \quad \mathrm{t}_{\mathbf{P}_a}(\zeta) = \frac{q_a}{\zeta - 1}. \tag{15.54}$$

Therefore, the inverse of the T-transforms just read $\zeta_{\mathbf{P}_a}(t) = 1 + q_a/t$, and

$$S_{\mathbf{P}_a}(t) = \frac{t+1}{t+q_a} \quad \Rightarrow \quad S_{\mathbf{P}} = \frac{(t+1)^2}{(t+q_1)(t+q_2)}. \tag{15.55}$$

Now, going backwards,

$$\zeta_{\mathbf{P}}(t) = \frac{(t+q_1)(t+q_2)}{t(t+1)}, \tag{15.56}$$

again leading to a quadratic equation for $\mathrm{t}_{\mathbf{P}}(\zeta)$:

$$(\zeta - 1)t^2 + (\zeta - q_1 - q_2)t - q_1 q_2 = 0, \tag{15.57}$$

whose solution is

$$\mathrm{t}(\zeta) = \frac{(q_1 + q_2 - \zeta) + \overset{\oplus}{\sqrt{}}\zeta^2 - 2\zeta(q_1 + q_2 - 2q_1 q_2) + (q_1 - q_2)^2}{2(\zeta - 1)}, \tag{15.58}$$

where the notation $\overset{\oplus}{\sqrt{}}\cdot$ defined by Eq. (4.56) ensures that we pick the correct root. Note that the argument under the square-root has zeros for

$$\lambda_\pm = q_1 + q_2 - 2q_1 q_2 \pm 2\sqrt{q_1 q_2 (1-q_1)(1-q_2)}. \tag{15.59}$$

One can check that $\lambda_- \geq 0$, the zero bound being reached for $q_1 = q_2$. Note also that $\lambda_+ \lambda_- = (q_1 - q_2)^2$.

https://doi.org/10.1017/9781108768900.019 Published online by Cambridge University Press

The Stieltjes transform of $\mathbf{P}$ can thus be written as

$$\mathfrak{g}_{\mathbf{P}}(z) = \frac{1}{z} + \frac{(q_1+q_2-z) + \sqrt[\oplus]{z^2 - 2z(q_1+q_2-2q_1q_2) + (q_1-q_2)^2}}{2z(z-1)}. \tag{15.60}$$

This quantity has poles at $z = 0$ and $z = 1$, and an imaginary part when $z \in (\lambda_+, \lambda_-)$. The spectrum of $\mathbf{P}$ therefore has a continuous part, given by

$$\rho_{\mathbf{P}}(\lambda) = \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi\lambda(1-\lambda)}, \tag{15.61}$$

and two delta peaks, $A_0\delta(\lambda)$ and $A_1\delta(\lambda - 1)$. To find the amplitude of the potential poles, we need to compute the numerator of (15.60) at the values $z = 0$ and $z = 1$. Remember that $\sqrt[\oplus]{\cdot}$ equals $-\sqrt{\cdot}$ on the real axis left of the left edge and $\sqrt{\cdot}$ right of the right edge. The amplitude of the $z = 0$ pole of $\mathfrak{g}_{\mathbf{P}}(z)$ is

$$A_0 = 1 - \frac{(q_1+q_2) - \sqrt{(q_1-q_2)^2}}{2} = 1 - \min(q_1, q_2), \tag{15.62}$$

while the amplitude of the $z = 1$ pole is

$$\begin{aligned} A_1 &= \frac{(q_1+q_2-1) + \sqrt{1 - 2(q_1+q_2-2q_1q_2) + (q_1-q_2)^2}}{2} \\ &= \frac{(q_1+q_2-1) + \sqrt{(q_1+q_2-1)^2}}{2} = \max(q_1+q_2-1, 0). \end{aligned} \tag{15.63}$$

This makes a lot of sense geometrically: our product of two projectors can only have a unit eigenvalue if the sum of the dimensions of space spanned by these two projectors exceeds the total dimension N, i.e. when $N_1 + N_2 > N$. Otherwise, there cannot be (generically) any eigenvalue beyond λ_+.

When $q_1 + q_2 < 1$, the density of non-zero eigenvalues (15.61) is the same (up to a normalization) as the density of eigenvalues of a Jacobi matrix (7.20). If we match the edges of the spectrum we find the identification $c_1 = q_{\max}/q_{\min}$ and $c_+ = 1/q_{\min}$. The ratio of normalization $1/c_+ = q_{\min}$ implies that the product of projectors density has a missing mass of $1 - q_{\min}$, which is precisely the Dirac mass at zero. The special case $q_1 = q_2 = 1/2$ was discussed in the context of 2×2 matrices in Exercise 12.5.2. In that case, half of the eigenvalues are zero and the other half are distributed according to the arcsine law: the arcsine law is the limit of a Jacobi matrix with $c_1 \to 1$ and $c_+ \to 2$.

There is an alternative, geometric interpretation of the above calculation that turns out to be useful in many different contexts, see Section 17.4 for an extended discussion. The eigenvectors of projector $\mathbf{P}_1$ form a set of N_1 orthonormal vectors $\mathbf{x}_\alpha$, $\alpha = 1, \ldots, N_1$, from which an $N_1 \times N$ matrix $\mathbf{X}$ of components $(x_\alpha)_i$ can be formed. Similarly, we define an $N_2 \times N$ matrix $\mathbf{Y}$ of components $(y_\beta)_i$, $\beta = 1, \ldots, N_2$. Now, one can write $\mathbf{P}$ as

$$\mathbf{P} = \mathbf{X}^T\mathbf{X}\mathbf{Y}^T\mathbf{Y}. \tag{15.64}$$

https://doi.org/10.1017/9781108768900.019 Published online by Cambridge University Press

The non-zero eigenvalues of $\mathbf{P}$ are the same as the non-zero eigenvalues of $\mathbf{M}^T\mathbf{M}$ (or those of $\mathbf{M}\mathbf{M}^T$), where $\mathbf{M}$ is the $N_1 \times N_2$ matrix of overlaps:

$$\mathbf{M}_{\alpha,\beta} := \sum_{i=1}^{N} (x_\alpha)_i (y_\beta)_i. \tag{15.65}$$

The eigenvalues of $\mathbf{P}$ correspond to the square of the singular values s of $\mathbf{M}$. The geometrical interpretation of these singular values is as follows: the largest singular value corresponds to the maximum overlap between any normalized linear combination of the $\mathbf{x}_\alpha$ on the one hand and of the $\mathbf{y}_\beta$ on the other hand. These two linear combinations define two one-dimensional subspaces of the spaces spanned by $\mathbf{x}_\alpha$ and $\mathbf{y}_\beta$. Once these optimal directions are removed, one can again ask the same question for the remaining $N_1 - 1$ and $N_2 - 1$ dimensional subspaces, defining the second largest singular value of $\mathbf{M}$, and so on.

15.4.3 The Jacobi Ensemble Revisited

We saw in the previous section that the free product of two random projectors has a very simple S-transform and that its non-zero eigenvalues are given by those of a Jacobi matrix. We suspect that the Jacobi ensemble has itself a simple S-transform. Rather than computing its S-transform from its Stieltjes transform (7.18), let us just use the properties of the S-transform to compute it directly from the definition of a Jacobi matrix.

Recall from Chapter 7, an $N \times N$ Jacobi matrix $\mathbf{J}$ is defined as $(\mathbf{1} + \mathbf{E})^{-1}$ where the matrix $\mathbf{E}$ is the free product of the inverse of an unnormalized Wishart matrix $\mathbf{W}_1$ with $T_1 = c_1 N$ and another unnormalized Wishart $\mathbf{W}_2$ with $T_2 = c_2 N$.

The two Wishart matrices have S-transforms given by

$$S_{\mathbf{W}_{1,2}}(t) = T_{1,2} \frac{1}{1 + c_{1,2}^{-1} t} = N \frac{1}{c_{1,2} + t}. \tag{15.66}$$

Using the relation for inverse matrices (15.7), we find

$$S_{\mathbf{W}_1^{-1}}(t) = N^{-1}(c_1 - t - 1). \tag{15.67}$$

The S-transform of $\mathbf{E}$ is just the product

$$S_{\mathbf{E}}(t) = S_{\mathbf{W}_1^{-1}}(t) S_{\mathbf{W}_2}(t) = \frac{c_1 - t - 1}{c_2 + t}. \tag{15.68}$$

The next step is to shift the matrix $\mathbf{E}$ by $\mathbf{1}$. As mentioned earlier, there is no easy rule to compute the S-transform of a shifted matrix. So this will be the hardest part of the computation.

The R-transform behaves simply under shift. The trick is to use one of Eqs. (15.8) to write an equation for $R_{\mathbf{E}}(x)$, shift $\mathbf{E}$ by $\mathbf{1}$ and use the other R-S relation to find back an equation for $S_{\mathbf{E}+\mathbf{1}}(t)$. First we write

$$(c_2 + t) S_{\mathbf{E}} = c_1 - t - 1. \tag{15.69}$$

The second of Eqs. (15.8) can be interpreted as the replacements $S \to 1/R$ and $t \to xR$ and gives

$$(1 + x - c_1 + R_{\mathbf{E}}) R_{\mathbf{E}} + c_2 = 0. \tag{15.70}$$

Now $R_{\mathbf{E}}(x) = R_{\mathbf{E}+\mathbf{1}}(x) - 1$, so

$$(x - c_1 + R_{\mathbf{E}+\mathbf{1}})(R_{\mathbf{E}+\mathbf{1}} - 1) + c_2 = 0. \tag{15.71}$$

https://doi.org/10.1017/9781108768900.019 Published online by Cambridge University Press

Following the first of Eqs. (15.8), we make the replacements $R \to 1/S$ and $x \to tS$ and find

$$1 - c_1 + t + (c_2 + c_1 - t - 1)S_{\mathbf{E+1}} = 0 \quad \Rightarrow \quad S_{\mathbf{E+1}}(t) = \frac{t + 1 - c_1}{t + 1 - c_2 - c_1}. \tag{15.72}$$

Finally, using the relation for inverse matrices, Eq. (15.7) gives

$$S_{\mathbf{J}}(t) = \frac{t + c_2 + c_1}{t + c_1}. \tag{15.73}$$

We can verify that the T-transform of the Jacobi ensemble

$$\mathrm{t}_{\mathbf{J}}(\zeta) = \frac{c_1 + 1 - c_+\zeta + \overset{\oplus}{\sqrt{}}c_+^2\zeta^2 - 2(c_1c_+ + c_+ - 2c_1)\zeta + (c_1 - 1)^2}{2(\zeta - 1)} \tag{15.74}$$

is compatible with our previous result on the Stieltjes transform, Eq. (7.18). We can use the Taylor series of the S-transform (15.11) to find the first few cumulants:

$$\kappa_1 = \frac{c_1}{c_1 + c_2}, \quad \kappa_2 = \frac{c_1c_2}{(c_1 + c_2)^3}, \quad \kappa_3 = \frac{(c_1 - c_2)c_1c_2}{(c_1 + c_2)^5}. \tag{15.75}$$

From the S-transform we can compute the R-transform using Eq. (15.8):

$$R_{\mathbf{J}}(x) = \frac{x - c_1 - c_2 - \sqrt{x^2 + 2(c_1 - c_2)x + (c_1 + c_2)^2}}{2x}. \tag{15.76}$$

Finally, we note that the arcsine law is a Jacobi matrix with $c_1 = c_2 = 1$ and has the following transform:

$$S(t) = \frac{t + 2}{t + 1}, \quad R(x) = \frac{x - 2 - \sqrt{x^2 + 4}}{2x}. \tag{15.77}$$

For the centered arcsine law we have $R_{\mathrm{s}}(t) = 2R(2x) + 1$ and we recover Eq. (15.38).

https://doi.org/10.1017/9781108768900.019 Published online by Cambridge University Press