14

Edge Eigenvalues and Outliers

In many instances, the eigenvalue spectrum of large random matrices is confined to a single interval of finite size. This is of course the case for Wigner matrices, where the correctly normalized eigenvalues fall between $\lambda_- = -2$ and $\lambda_+ = +2$, with a semi-circular distribution between the two edges, and, correspondingly, a square-root singularity of the density of eigenvalues close to the edges. This is also the case for Wishart matrices, for which again the density of eigenvalues has square-root singularities close to both edges. As discussed in Section 5.3.2, this is a generic property, with a few notable exceptions. One example is provided by Wishart matrices with parameter $q = 1$, for which the eigenvalue spectrum extends down to $\lambda = 0$ with an inverse square-root singularity there. Another case is that of Wigner matrices constrained to have all eigenvalues positive: the spectrum also has an inverse square-root singularity – see Eq. (5.94). One speaks of a "hard edge" in that case, because the minimum eigenvalue is imposed by a strict constraint. The Wigner semi-circle edge at $\lambda_+ = 2$, on the other hand, is "soft" and appears naturally as a result of the minimization of the energy of a collection of interacting Coulomb charges in an external potential.[1]

Consider for example Wigner matrices of size N. The existence of sharp edges delimiting a region where one expects to see a non-zero density of eigenvalues from a region where there should be none is only true in the asymptotically large size limit $N \to \infty$. For large but finite N, on the other hand, one expects that the probability to find an eigenvalue beyond the Wigner sea is very small but non-zero. The width of the transition region, and the tail of the density of states was investigated a while ago, culminating in the beautiful results by Tracy and Widom on the distribution of the *largest* eigenvalue of a random matrix, which we will describe in the next section. The most important result is that the width of the region around λ_+ within which one expects to observe the largest eigevalue of a Wigner matrix goes down as $N^{-2/3}$.

Hence the largest eigenvalue $\lambda_{\max}$ does not fluctuate very far away from the classical edge λ_+. Take for example $N = 1000$; $\lambda_{\max}$ is within $1000^{-2/3} = 0.01$ away from $\lambda_+ = 2$. In real applications the largest eigenvalue can deviate quite substantially from the classical

[1] Note that there are also cases where the soft edge has a different singularity, see Section 5.3.3, and cases where the eigenvalue spectrum extends up to infinity, for example "Lévy matrices" with IID elements of infinite variance.

https://doi.org/10.1017/9781108768900.017 Published online by Cambridge University Press

edge. The origin of such a large eigenvalue is usually not an improbably large Tracy–Widom fluctuation but rather a true outlier that should be modeled separately. This is the goal of the present chapter. We will see in particular that perturbing a Wigner (or Wishart) matrix with a deterministic, low-rank matrix of sufficient amplitude $a > a_c$ generates "true" outliers, which remain at a distance $O(a)$ from the upper edge. For $a < a_c$ on the other hand, the largest eigenvalue remains at distance $N^{-2/3}$ from λ_+.

14.1 The Tracy–Widom Regime

The Tracy–Widom result characterizes precisely the distance between the largest eigenvalue $\lambda_{\max}$ of Gaussian Wigner or Wishart matrices and the upper edge of the spectrum which we denoted by λ_+. This result can be (formally) stated as follows: the rescaled distribution of $\lambda_{\max} - \lambda_+$ converges, for $N \to \infty$, towards the Tracy–Widom distribution, usually noted F_1:

$$\mathbb{P}\left(\lambda_{\max} \leq \lambda_+ + \gamma N^{-2/3} u\right) = F_1(u), \tag{14.1}$$

where γ is a constant that depends on the problem and $F_1(u)$ is the $\beta = 1$ Tracy–Widom distribution. For the Wigner problem, $\lambda_+ = 2$ and $\gamma = 1$, whereas for Wishart matrices, $\lambda_+ = (1 + \sqrt{q})^2$ and $\gamma = \sqrt{q}\lambda_+^{2/3}$. In fact, Eq. (14.1) holds for a much wider large class of $N \times N$ random matrices, for example symmetric random matrices with arbitrary IID elements with a finite fourth moment. The Tracy–Widom distribution for all three values of β is plotted in Figure 14.1. (The case where the fourth moment is infinite is discussed in Section 14.3 below.)

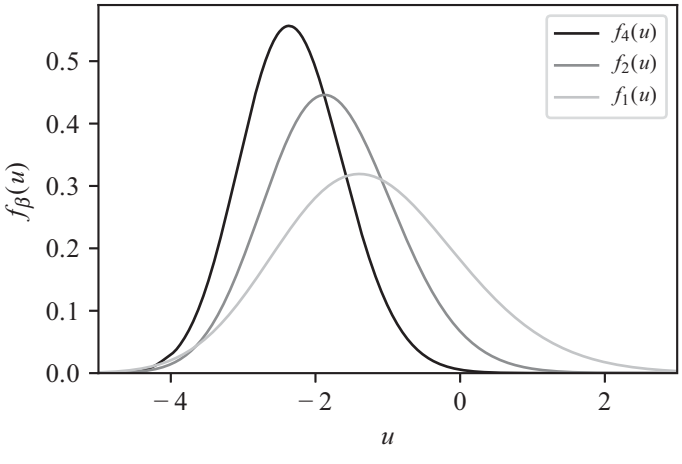

Figure 14.1 Rescaled and shifted probability density of the largest eigenvalue for a large class of random matrices such as Wigner and Wishart: the Tracy–Widom distribution. The distribution depends on the Dyson index (β) and is shown here for $\beta = 1, 2$ and 4.

https://doi.org/10.1017/9781108768900.017 Published online by Cambridge University Press

Everything is known about the Tracy–Widom density $f_1(u) = F_1'(u)$, in particular its left and right far tails:

$$\ln f_1(u) \propto -u^{3/2}, \quad (u \to +\infty); \qquad \ln f_1(u) \propto -|u|^3, \quad (u \to -\infty). \tag{14.2}$$

One notices that the left tail is much thinner than the right tail: pushing the largest eigenvalue inside the Wigner sea implies compressing the whole Coulomb gas of repulsive charges, which is more difficult than pulling one eigenvalue away from λ_+. Using this analogy and the formalism of Section 5.4.2, the large deviation regime of the Tracy–Widom problem (i.e. for $\lambda_{\max} - \lambda_+ = O(1)$) can be obtained. Note that the result is exponentially small in N as the $u^{3/2}$ behavior for $u \to \infty$ combines with $N^{2/3}$ to give a linear in N dependence.

The distribution of the smallest eigenvalue $\lambda_{\min}$ around the lower edge λ_- is also Tracy–Widom, except in the particular case of Wishart matrices with $q = 1$. In this case $\lambda_- = 0$, which is a "hard" edge since all eigenvalues of the empirical matrix must be non-negative.[2]

The behavior of the width of the transition region can be understood using a simple heuristic argument. Suppose that the $N = \infty$ density goes to zero near the upper edge λ_+ as $(\lambda_+ - \lambda)^\theta$ (generically, $\theta = 1/2$ as is the case for the Wigner and the Marčenko–Pastur distributions). For finite N, one expects not to be able to say whether the density is zero or non-zero when the probability to observe an eigenvalue is of order $1/N$, i.e. when the $O(1)$ eigenvalue is within the "blurred" region. This leads to a blurred region of width

$$|\lambda^* - \lambda_+|^{\theta+1} \propto \frac{1}{N} \to \Delta\lambda^* \sim N^{-\frac{1}{1+\theta}}, \tag{14.3}$$

which goes to zero as $N^{-2/3}$ in the generic square-root case $\theta = 1/2$. More precisely, for Gaussian ensembles, the average density of states at a distance $\sim N^{-2/3}$ from the edge behaves as

$$\rho_N(\lambda \approx \lambda_+) = N^{-1/3}\Phi_1\left[N^{2/3}(\lambda - \lambda_+)\right], \tag{14.4}$$

with $\Phi_1(u \to -\infty) \approx \sqrt{-u}/\pi$ so as to recover the asymptotic square-root singularity, since the N dependence disappears in that limit. Far from the edge, $\ln \Phi_1(u \to +\infty) \propto -u^{3/2}$, showing that the probability to find an eigenvalue outside of the allowed band decays exponentially with N and super-exponentially with the distance to the edge. The function $\Phi_1(u)$ is not known analytically for real Wigner matrices ($\beta = 1$) but an explicit expression is available for complex Hermitian Wigner matrices, and reads (Fig. 14.2)

$$\Phi_2(u) = \mathrm{Ai}'^2(u) - u\,\mathrm{Ai}^2(u), \tag{14.5}$$

with the same asymptotic behaviors as $\Phi_1(u)$. ($\mathrm{Ai}(u)$ is the standard Airy function.)

[2] This special case is treated in Péché [2003].

https://doi.org/10.1017/9781108768900.017 Published online by Cambridge University Press

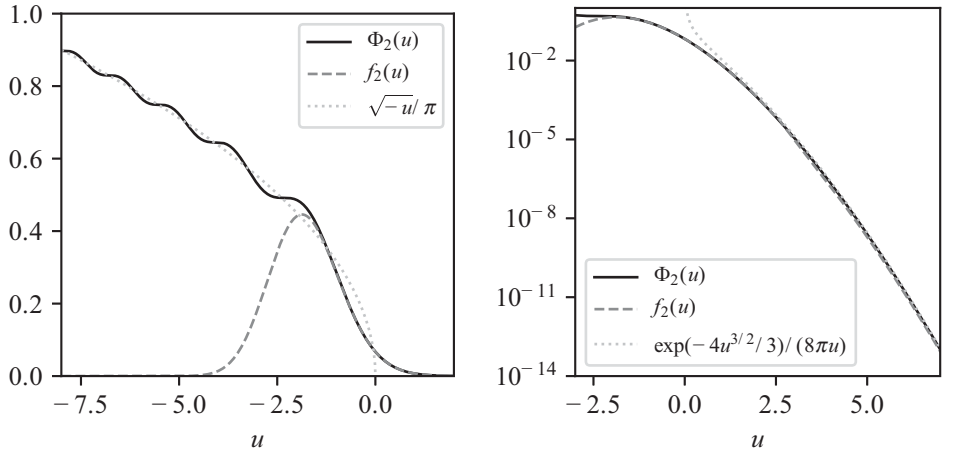

Figure 14.2 Behavior of the density near the edge λ_+ at the scale $N^{-2/3}$ for complex Hermitian Wigner matrices given by Eq. (14.5). For comparison the probability of the largest eigenvalue $f_2(u)$ is also shown. For positive values, the two functions are almost identical and behave as $\exp(-4u^{3/2}/3)/(8\pi u)$ for large u (right). For negative arguments the functions are completely different: $\Phi_2(u)$ behaves as $\sqrt{-u}/\pi$ for large negative u while $f_2(u) \to 0$ as the largest eigenvalue cannot be in the bulk (left).

14.2 Additive Low-Rank Perturbations

14.2.1 Eigenvalues

We will now study the outliers for an additive perturbation to a large random matrix. Take a large symmetric random matrix $\mathbf{M}$ (e.g. Wigner or Wishart) with a well-behaved asymptotic spectrum that has a deterministic right edge λ_+. We would like to know what happens when one adds to $\mathbf{M}$ a low-rank (deterministic) perturbation. For simplicity, we only consider the rank-1 perturbation $a\mathbf{u}\mathbf{u}^T$ with $\|\mathbf{u}\| = 1$ and a of order 1, but the results below easily generalize to the case of a rank-n perturbation with $n \ll N$.

We want to know whether there will be an isolated eigenvalue of $\mathbf{M} + a\mathbf{u}\mathbf{u}^T$ outside the spectrum of $\mathbf{M}$ (i.e. an "outlier") or not. To answer this question, we calculate the matrix resolvent

$$\mathbf{G}_a(z) = \left(z - \mathbf{M} - a\mathbf{u}\mathbf{u}^T\right)^{-1}. \tag{14.6}$$

The matrix $\mathbf{G}_a(z)$ has a pole at every eigenvalue of $\mathbf{M} + a\mathbf{u}\mathbf{u}^T$. An alternative approach would have been to study the zeros of the function $\det\left(z - \mathbf{M} - a\mathbf{u}\mathbf{u}^T\right)$, but the full resolvent $\mathbf{G}_a(z)$ also gives us information about the eigenvectors.

Now we apply the Sherman–Morrison formula (1.28); taking $\mathbf{A} = z - \mathbf{M}$, we get

$$\mathbf{G}_a(z) = \mathbf{G}(z) + a\frac{\mathbf{G}(z)\mathbf{u}\mathbf{u}^T\mathbf{G}(z)}{1 - a\mathbf{u}^T\mathbf{G}(z)\mathbf{u}}, \tag{14.7}$$

https://doi.org/10.1017/9781108768900.017 Published online by Cambridge University Press

where $\mathbf{G}(z)$ is the resolvent of the original matrix $\mathbf{M}$. We are looking for a real eigenvalue such that $\lambda_1 > \lambda_+$. Let us take $z = \lambda_1 \in \mathbb{R}$ outside the spectrum of $\mathbf{M}$, so $\mathbf{G}(\lambda)$ is real and regular. To have an outlier at λ_1, we need a pole of $\mathbf{G}_a$ at λ_1, i.e. the following equation needs to be satisfied:

$$1 - a\mathbf{u}^T\mathbf{G}(\lambda_1)\mathbf{u} = 0. \tag{14.8}$$

Assume now that $\mathbf{M}$ is drawn from a rotationally invariant ensemble or, equivalently, that the vector $\mathbf{u}$ is an independent random vector uniformly distributed on the unit sphere. In the language of Chapter 11, we say that the perturbation $a\mathbf{u}\mathbf{u}^T$ is free from the matrix $\mathbf{M}$. We then have, in the eigenbasis of $\mathbf{G}$,

$$\mathbf{u}^T\mathbf{G}(z)\mathbf{u} = \sum_i \mathbf{u}_i^2\mathbf{G}_{ii}(z) \approx \frac{1}{N}\operatorname{Tr}\mathbf{G}(z) = g_N(z) \overset{N\to\infty}{\to} \mathfrak{g}(z). \tag{14.9}$$

Thus we have a pole when

$$a\mathfrak{g}(\lambda_1) = 1 \Rightarrow \mathfrak{g}(\lambda_1) = 1/a. \tag{14.10}$$

If $\mathfrak{z}(g)$, the inverse function of $\mathfrak{g}(z)$, exists, we arrive at

$$\lambda_1 = \mathfrak{z}\left(\frac{1}{a}\right). \tag{14.11}$$

The condition for the invertibility of $\mathfrak{g}(z)$ happens to be precisely the same as the condition to have an outlier, i.e. $\lambda_1 > \lambda_+$ – see Section 10.4. We have established there that $\lambda_1 = \mathfrak{z}(1/a)$ is monotonically increasing in a, and $\lambda_1 = \lambda_+$ when $a = a^* = 1/\mathfrak{g}(\lambda_+)$, which is the critical value of a for which an outlier first appears. Generically, $g_+ = \mathfrak{g}(\lambda_+)$ is a minimum of $\mathfrak{z}(g)$:

$$\left.\frac{\mathrm{d}\mathfrak{z}(g)}{\mathrm{d}g}\right|_{g_+} = 0 \quad \text{when} \quad \mathfrak{z}(g_+) = \lambda_+. \tag{14.12}$$

For instance, for Wigner matrices, we have $\mathfrak{z}(g) = \sigma^2 g + g^{-1}$, for which

$$\sigma^2 - g_+^{-2} = 0 \Rightarrow g_+ = \sigma^{-1}, \tag{14.13}$$

and $\lambda_+ = \mathfrak{z}(\sigma^{-1}) = 2\sigma$, which is indeed the right edge of the semi-circle law.

In sum, for $a > a^* = 1/g_+$, there exists a unique outlier eigenvalue that is increasing with a. The smallest value for which we can have an outlier is $a^* = 1/g_+$, corresponding to $\lambda_1 = \lambda_+$. For $a < a^*$ there is no outlier to the right of λ_+.[3]

Using the relation between the inverse function $\mathfrak{z}(g)$ and the R-transform (10.10), we can express the position of the outlier as

$$\lambda_1 = R\left(\frac{1}{a}\right) + a \qquad \text{for} \qquad a > a^* = \frac{1}{g_+}. \tag{14.14}$$

[3] Outliers such that $\lambda < \lambda_-$ behave similarly, we just need to consider the matrix $-\mathbf{M} - a\mathbf{u}\mathbf{u}^T$ and follow the same logic.

https://doi.org/10.1017/9781108768900.017 Published online by Cambridge University Press

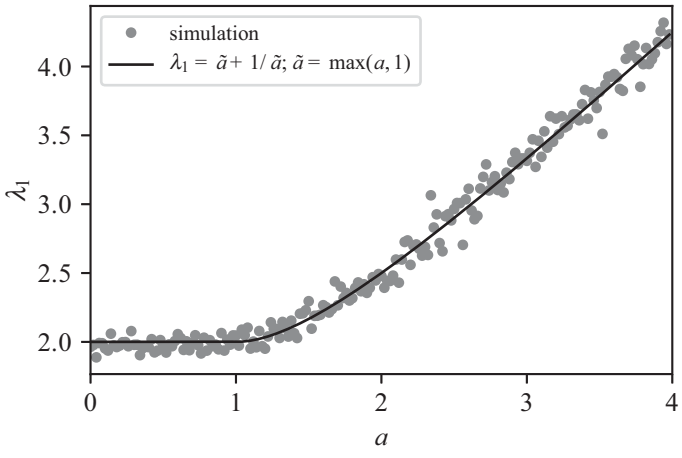

Figure 14.3 Largest eigenvalue of a Gaussian Wigner matrix with $\sigma^2 = 1$ with a rank-1 perturbation of magnitude a. Each dot is the largest eigenvalue of a single random matrix with $N = 200$. Equation (14.16) is plotted as the solid curve. For $a < 1$, the fluctuations follow a Tracy–Widom law with $N^{-2/3}$ scaling, while for $a > 1$ the fluctuations are Gaussian with $N^{-1/2}$ scaling. From the graph, we see fluctuations that are indeed smaller when $a < 1$. They also have a negative mean and positive skewness, in agreement with the Tracy–Widom distribution.

Using the cumulant expansion of the R-transform (11.63), we then get a general expression for large a:

$$\lambda_1 = a + \tau(\mathbf{M}) + \frac{\kappa_2(\mathbf{M})}{a} + O(a^{-2}). \tag{14.15}$$

For Wigner matrices, we actually have for all a (see Fig. 14.3)

$$\lambda_1 = a + \frac{\sigma^2}{a} \qquad \text{for} \qquad a > a^* = \sigma. \tag{14.16}$$

When $a \to a^*$, on the other hand, one has

$$\left.\frac{\mathrm{d}\lambda_1(a)}{\mathrm{d}a}\right|_{a^*} = \left.\frac{\mathrm{d}\mathfrak{z}(g)}{\mathrm{d}g}\right|_{g_+ = 1/a^*} = 0. \tag{14.17}$$

Hence, one has, for $a \to a^*$ and for generic square-root singularities,

$$\lambda_1 = \lambda_+ + C(a - a^*)^2 + O\left((a - a^*)^3\right), \tag{14.18}$$

where C is some problem dependent coefficient.

By studying the fluctuations of $\mathbf{u}\mathbf{G}(\lambda)\mathbf{u}^T$ around $\mathfrak{g}(\lambda)$, one can show that the fluctuations of the outlier around $\lambda_1 = R(a^{-1}) + a$ are Gaussian and of order $N^{-1/2}$. This is to be contrasted with the fluctuations of the largest eigenvalue when there are no outliers ($a < g_+$), which are Tracy–Widom and of order $N^{-2/3}$. The transition between the two regimes is called the Baik–Ben Arous–Péché (BBP) transition.

https://doi.org/10.1017/9781108768900.017 Published online by Cambridge University Press

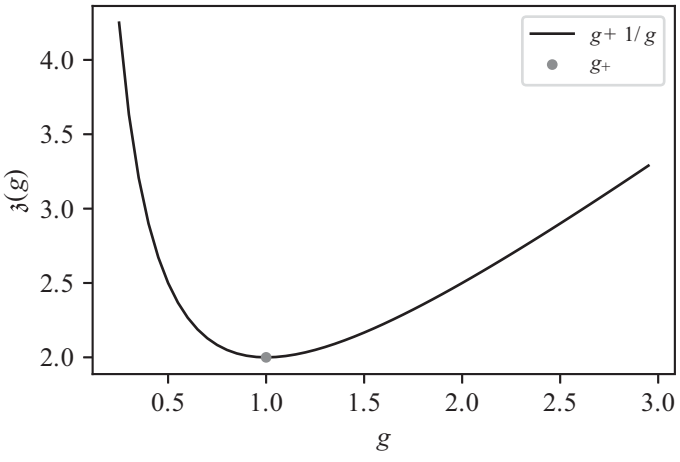

Figure 14.4 Plot of the inverse function $\mathfrak{z}(g) = g + 1/g$ for the unit Wigner function $\mathfrak{g}(z)$. The gray dot indicates the point (g_+, λ_+). The line to the left of this point is the true inverse of $\mathfrak{g}(z)$: $\mathfrak{z}(g)$ is defined on $[0, g_+)$ and is monotonously decreasing in g. The line to the right is a spurious solution introduced by the R-transform. Note that the point $g = g_+$ is a minimum of $\mathfrak{z}(g) = g + 1/g$.

We finish this section with two remarks.

- One concerns the solutions to Eq. (14.14), and the way to find the value a^* beyond which an outlier appears. The point is that, while the function $R(g = 1/a)$ is well defined for $g \in [0, g_+)$, it also often makes sense even beyond g_+ (see the discussion in Section 10.4). In that case, one will find spurious solutions: Figure 14.4 shows a plot of $\mathfrak{z}(g) = R(g) + 1/g$ in the unit Wigner case, which is still well defined for $g > g_+ = 1$ even if this function is no longer the inverse of $\mathfrak{g}(z)$ (Section 10.4). There are two solutions to $\mathfrak{z}(g) = \lambda_1$, one such that $g < g_+$ and the other such that $g > g_+$. As noted above, the point g_+ is a minimum of $\mathfrak{z}(g)$, beyond which the relation between λ_1 and a is monotonically increasing because $\mathfrak{g}(z)$ is monotonically decreasing for $z > \lambda_+$.
- The second concerns the case of a free rank-n perturbation, when $n \ll N$. In this case one cannot use the Sherman–Morrison formula but one can compute the R-transform of the perturbed matrix, and infer the $1/N$ correction to the Stieltjes transform. The poles of this correction term give the possible outliers. To each eigenvalue a_k ($k = 1, \ldots, n$) of the perturbation, one can associate a candidate outlier λ_k given by

$$\lambda_k = R\left(\frac{1}{a_k}\right) + a_k \quad \text{when} \quad a_k > \frac{1}{g_+}. \tag{14.19}$$

14.2.2 Outlier Eigenvectors

The matrix resolvent in Eq. (14.7) can also tell us about the eigenvectors of the perturbed matrix. We expect that, for a very strong rank-1 perturbation $a\mathbf{u}\mathbf{u}^T$, the eigenvector $\mathbf{v}_1$

https://doi.org/10.1017/9781108768900.017 Published online by Cambridge University Press

associated with the outlier λ_1 will be very close to the perturbation vector $\mathbf{u}$. On the other hand, for $\lambda_1 \approx \lambda_+$, the vector $\mathbf{u}$ will strongly mix with bulk eigenvectors of $\mathbf{M}$ so the eigenvector $\mathbf{v}_1$ will not contain much information about $\mathbf{u}$.

To understand this phenomenon quantitatively, we will study the squared overlap $|\mathbf{v}_1^T\mathbf{u}|^2$. With the spectral decomposition of $\mathbf{M} + a\mathbf{u}\mathbf{u}^T$, we can write

$$\mathbf{G}_a(z) = \sum_{i=1}^{N} \frac{\mathbf{v}_i\mathbf{v}_i^T}{z - \lambda_i}, \tag{14.20}$$

where λ_1 denotes the outlier and $\mathbf{v}_1$ its eigenvector, and $\lambda_i, \mathbf{v}_i$, $i > 1$ all other eigenvalues/eigenvectors. Thus we have

$$\lim_{z\to\lambda_1} \mathbf{u}^T\mathbf{G}_a(z)\mathbf{u} \cdot (z - \lambda_1) = |\mathbf{v}_1^T\mathbf{u}|^2. \tag{14.21}$$

Hence, by (14.7) and (14.9), we get

$$\begin{aligned} |\mathbf{v}_1^T\mathbf{u}|^2 &= \lim_{z\to\lambda_1} \left(\mathfrak{g}(z) + a\frac{\mathfrak{g}(z)^2}{1 - a\mathfrak{g}(z)}\right)(z - \lambda_1) \\ &= \lim_{z\to\lambda_1} \mathfrak{g}(z)\frac{z - \lambda_1}{1 - a\mathfrak{g}(z)}. \end{aligned} \tag{14.22}$$

We cannot simply evaluate the fraction above at $z = \lambda_1$, for at that point $\mathfrak{g}(\lambda_1) = a^{-1}$ and we would get $0/0$. We can however use l'Hospital's rule[4] and find

$$|\mathbf{v}_1^T\mathbf{u}|^2 = -\frac{\mathfrak{g}(\lambda_1)^2}{\mathfrak{g}'(\lambda_1)}, \tag{14.24}$$

where we have used $a^{-1} = \mathfrak{g}(\lambda_1)$. The right hand side is always positive since $\mathfrak{g}$ is a decreasing function for $\lambda > \lambda_+$.

We can rewrite Eq. (14.24) in terms of the R-transform and get a more useful formula. To compute $g'(z)$, we take the derivative with respect to z of the implicit equation $z = R(\mathfrak{g}(z)) + \mathfrak{g}^{-1}(z)$ and get

$$1 = R'(\mathfrak{g}(z))\mathfrak{g}'(z) - \frac{\mathfrak{g}'(z)}{\mathfrak{g}^2} \Rightarrow \mathfrak{g}'(z) = \frac{1}{R'(\mathfrak{g}(z)) - \mathfrak{g}^{-2}(z)}. \tag{14.25}$$

Hence we have

$$\begin{aligned} |\mathbf{v}_1^T\mathbf{u}|^2 &= 1 - \mathfrak{g}(\lambda_1)^2 R'(\mathfrak{g}(\lambda_1)) \\ &= 1 - a^{-2}R'(a^{-1}). \end{aligned} \tag{14.26}$$

We can now check our intuition about the overlap for large and small perturbations. For a large perturbation $a \to \infty$, Eq. (14.26) gives

[4] L'Hospital's rule states that

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} = \frac{f'(x_0)}{g'(x_0)} \text{ when } f(x_0) = g(x_0) = 0. \tag{14.23}$$

https://doi.org/10.1017/9781108768900.017 Published online by Cambridge University Press

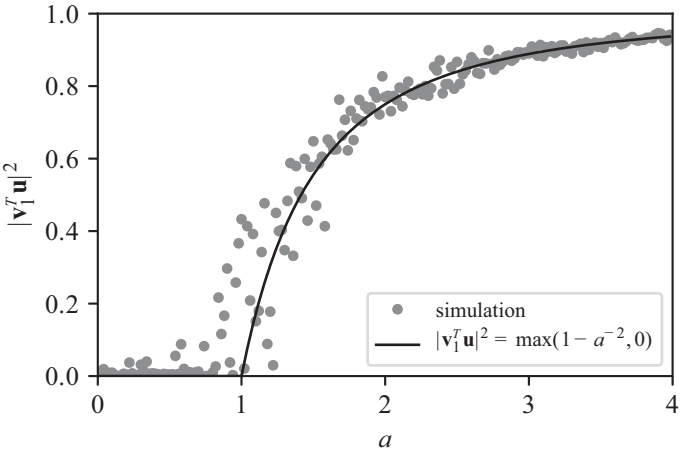

Figure 14.5 Overlap between the largest eigenvector and the perturbation vector of a Gaussian Wigner matrix with $\sigma^2 = 1$ with a rank-1 perturbation of magnitude a. Each dot is the overlap for a single random matrix with $N = 200$. Equation (14.31) is plotted as the solid curve.

$$|\mathbf{v}_1^T\mathbf{u}|^2 = 1 - \frac{\kappa_2(\mathbf{M})}{a^2} + O(a^{-3}) \quad \text{when} \quad a \to \infty. \tag{14.27}$$

As expected, $\mathbf{v}_1 \to \mathbf{u}$ when $a \to \infty$: the angle between the two vectors decreases as $1/a$.

The overlap near the transition $\lambda_1 \to \lambda_+$ can be analyzed as follows. The derivative of $\mathfrak{g}(z)$ can be written as

$$\mathfrak{g}'(z) = -\int_{\lambda_-}^{\lambda_+} \frac{\rho(x)}{(z-x)^2} \mathrm{d}x. \tag{14.28}$$

For a density that vanishes at the edge as $\rho(\lambda) \sim (\lambda_+ - \lambda)^\theta$ with exponent θ between 0 and 1, we have that $\mathfrak{g}(z)$ is finite at $z = \lambda_+$ but $\mathfrak{g}'(z)$ diverges at the same point, as $|z - \lambda_+|^{\theta-1}$. From Eq. (14.24), we have in that case[5]

$$|\mathbf{v}_1^T\mathbf{u}|^2 \propto (\lambda_1 - \lambda_+)^{1-\theta} \quad \text{when} \quad \lambda_1 \to \lambda_+. \tag{14.29}$$

In the generic case, one has $\theta = 1/2$ and, from Eq. (14.18), $\lambda_1 - \lambda_+ \propto (a - a^*)^2$, leading to

$$|\mathbf{v}_1^T\mathbf{u}|^2 \propto a - a^* \quad \text{when} \quad a \to a^*. \tag{14.30}$$

These general results are nicely illustrated by Wigner matrices, for which $R(x) = \sigma^2 x$. The overlap is explicitly given by (see Fig. 14.5)

$$|\mathbf{v}_1^T\mathbf{u}|^2 = 1 - \left(\frac{a^*}{a}\right)^2 \quad \text{when} \quad a > a^* = \sigma. \tag{14.31}$$

[5] In Chapter 5, we encountered a *critical density* where $\rho(\lambda)$ behaves as $(\lambda_+ - \lambda)^\theta$ with an exponent $\theta = \frac{3}{2} > 1$. In this case $\mathfrak{g}'(z)$ does not diverge as $z \to \lambda_+$ and the squared overlap at the edge of the BBP transition does not go to zero (first order transition). For example for the density given by Eq. (5.59) we find $|\mathbf{v}_1^T\mathbf{u}|^2 = \frac{4}{9}$ at the edge $\lambda_1 = 2\sqrt{2}$.

https://doi.org/10.1017/9781108768900.017 Published online by Cambridge University Press

As $a \to a^*$, $\lambda_1 \to 2\sigma$ and $|\mathbf{v}_1^T \mathbf{u}|^2 = 2(a - a^*)/a^* \to 0$: the eigenvector becomes delocalized as the eigenvalue merges with the bulk. For $a < a^*$, one can rigorously show that there is no information left in the eigenvalues of the perturbed matrix that would allow us to reconstruct $\mathbf{u}$.

Note that for $\lambda_1 > \lambda_+$, $|\mathbf{v}_1^T \mathbf{u}|^2$ is of order unity. In Chapter 19 we will see that this is not the case for the overlaps between perturbed and unperturbed eigenvectors in the bulk, which have typical sizes of order N^{-1}.

Exercise 14.2.1 Additive perturbation of a Wishart matrix

Define a modified Wishart matrix $\mathbf{W}_1$ such that every element $(\mathbf{W}_1)_{ij} = \mathbf{W}_{ij} + a/N$, where $\mathbf{W}$ is a standard Wishart matrix and a is a constant of order 1. $\mathbf{W}_1$ is a standard Wishart matrix plus a rank-1 perturbation $\mathbf{W}_1 = \mathbf{W} + a\mathbf{u}\mathbf{u}^T$.

(a) What is the normalized vector $\mathbf{u}$ in this case?

(b) Using Eqs. (14.14) and (10.15) find the value of the outlier and the minimal a in the Wishart case.

(c) The square-overlap between the vector $\mathbf{u}$ and the new eigenvector $\mathbf{v}_1$ is given by Eq. (14.26). Give an explicit expression in the Wishart case.

(d) Generate a large modified Wishart ($q = 1/4$, $N = 1000$) for a few a in the range $[1, 5]$. Compute the largest eigenvalue λ_1 and associated eigenvector $\mathbf{v}_1$. Plot λ_1 and $|\mathbf{v}_1^T \mathbf{u}|^2$ as a function of a and compare with the predictions of (b) and (c).

14.3 Fat Tails

The previous section allows us to discuss the very interesting situation of real symmetric random matrices $\mathbf{X}$ with IID elements X_{ij} that have a fat-tailed distribution, but with a finite variance (the case of infinite variance will be alluded to at the end of the section). In order to have eigenvalues of order unity, the random elements must be of typical size $N^{-1/2}$, so we write

$$X_{ij} = \frac{x_{ij}}{\sqrt{N}}, \tag{14.32}$$

with x_{ij} distributed according to some density $P(x)$ of mean zero and variance unity, but that decays as $\mu|x|^{-1-\mu}$ for large x. This means that most elements X_{ij} are small, of order $N^{-1/2}$, with some exceptional elements that are of order unity. The probability that $|X_{ij}| > 1$ is actually given by

$$\mathbb{P}(|X_{ij}| > 1) \approx 2 \int_{\sqrt{N}}^{\infty} \mathrm{d}x \frac{\mu}{x^{1+\mu}} = \frac{2}{N^{\mu/2}}. \tag{14.33}$$

Since there are in total $N(N-1)/2 \approx N^2/2$ such random variables, the total number of such variables that exceed unity is given by $N^{2-\mu/2}$. Hence, for $\mu > 4$, this number tends

https://doi.org/10.1017/9781108768900.017 Published online by Cambridge University Press

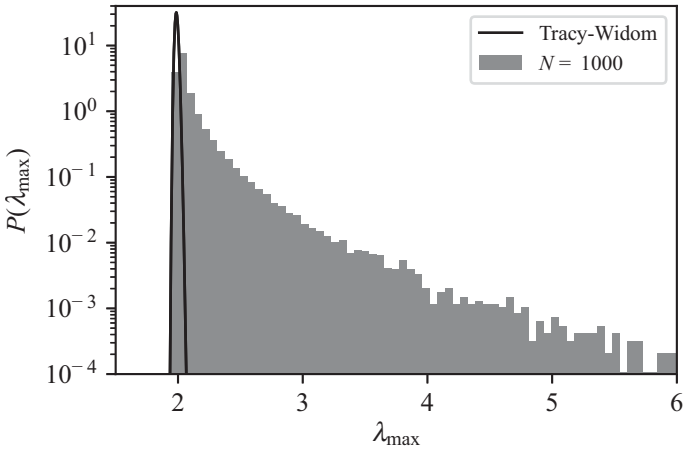

Figure 14.6 Distribution of the largest eigenvalue of $N = 1000$ Wigner matrices with elements drawn from a distribution with $\mu = 5$ compared with the prediction from the Tracy–Widom distribution. Even though as $N \to \infty$ this distribution converges to Tracy–Widom, at $N = 1000$ there is no agreement between the laws as the power law tail $\mathbb{P}(\lambda_{\max} > x) \sim x^{-5}/\sqrt{N}$ still dominates.

to zero with N: there are typically *no* such large elements in the considered random matrix. Since each pair of large entries $X_{ij} = X_{ji}$ can be considered as a rank-2 perturbation of a matrix with all elements of order $N^{-1/2}$, one concludes that for $\mu > 4$ there are no outliers, and the statistics of the largest eigenvalue is given by the Tracy–Widom distribution around $\lambda_+ = 2$. This hand-waving argument can actually be made rigorous: in the large N limit, the finiteness of the fourth moment of the distribution of matrix elements is sufficient to ensure that the largest eigenvalue is given by the Tracy–Widom distribution. However, one should be careful in interpreting this result, because although very large elements appear with vanishing probability, they still dominate the tail of the Tracy–Widom distribution for finite N. The reason is that, whereas the former decreases as $N^{2-\mu/2}$, the latter decreases much faster, as $\exp(-N(\lambda_{\max} - \lambda_+)^{3/2})$ (see Fig. 14.6).

Now consider the case $2 < \mu < 4$. Since $\mu > 2$, the variance of X_{ij} is finite and one knows that the asymptotic distribution of eigenvalues of $\mathbf{X}$ is given by the Wigner semi-circle, with $\lambda_+ = 2$. But now the number of large entries in the matrix $\mathbf{X}$ grows with N as $N^{2-\mu/2}$, which is nevertheless still much smaller than N. Each large pair of entries $X_{ij} = X_{ji} = a$ larger than unity (in absolute value) contributes to two outliers, given by $\lambda = \pm(a + 1/a)$. So there are in total $O(N^{2-\mu/2})$ outliers, the density of which is given by

$$\rho_{\text{out}}(\lambda > 2) = N^{1-\mu/2} \int_1^\infty \mathrm{d}x \frac{\mu}{x^{1+\mu}} \delta\left(\lambda - x - \frac{1}{x}\right). \tag{14.34}$$

https://doi.org/10.1017/9781108768900.017 Published online by Cambridge University Press

This is a rather strange situation where the density of outliers goes to zero as $N \to \infty$ as soon as $\mu > 2$, but at the same time the largest eigenvalue in this problem goes to infinity as

$$\lambda_{\max} \sim N^{\frac{2}{\mu}-\frac{1}{2}}, \qquad 2 < \mu < 4. \tag{14.35}$$

Finally, let us briefly comment on the case $\mu < 2$, for which the variance of the entries of $\mathbf{X}$ diverges (a case called "Lévy matrices" in the literature). For eigenvalues to remain of order unity, one needs to scale the matrix elements differently, as

$$X_{ij} = \frac{x_{ij}}{N^{\frac{1}{\mu}}}, \tag{14.36}$$

with

$$P(x) \underset{|x|\to\infty}{\sim} \frac{\Gamma(1+\mu)\sin(\frac{\pi\mu}{2})}{|x|^{1+\mu}}, \tag{14.37}$$

where the funny factor involving the gamma function is introduced for convenience only. The eigenvalue distribution is no longer given by a semi-circle. In fact, the support of the distribution is in this case unbounded. For completeness, we give here the exact expression of the distribution in terms of Lévy stable laws $L_{\mu}^{C,\beta}(u)$, where β is called the asymmetry parameter and C the scale parameter.[6] For a given value of λ, one should first solve the following two self-consistent equations for C and β:

$$\begin{aligned} C &= \int_{-\infty}^{+\infty} \mathrm{d}g |g|^{\mu/2-2} L_{\mu/2}^{C,\beta}(\lambda - 1/g), \\ C\beta &= \int_{-\infty}^{+\infty} \mathrm{d}g \,\mathrm{sign}(g) |g|^{\mu/2-2} L_{\mu/2}^{C,\beta}(\lambda - 1/g). \end{aligned} \tag{14.38}$$

Finally, the distribution of eigenvalues $\rho_L(\lambda)$ of Lévy matrices is obtained as

$$\rho_L(\lambda) = L_{\mu/2}^{C(\lambda),\beta(\lambda)}(\lambda). \tag{14.39}$$

One can check in particular that this distribution decays for large λ exactly as $P(x)$ itself. In other words, the tail of the eigenvalue distribution is the same as the tail of the independent entries of the Lévy matrix.

14.4 Multiplicative Perturbation

In data-analysis applications, we often need to understand the largest eigenvalue of a sample covariance matrix. A true covariance with a few isolated eigenvalues can be treated as a matrix $\mathbf{C}_0$ with no isolated eigenvalue plus a low-rank perturbation. The passage from the true covariance to the sample covariance is equivalent to the free product of the true covariance with a white Wishart matrix with appropriate aspect ratio $q = N/T$. To understand such matrices, we will now study outliers for a multiplicative process.

Consider the free product of a certain covariance matrix $\mathbf{C}_0$ with a rank-1 perturbation and another matrix $\mathbf{B}$:

$$\mathbf{E} = \mathbf{B}^{\frac{1}{2}} \mathbf{C}_0^{\frac{1}{2}} \left(\mathbf{1} + a\mathbf{u}\mathbf{u}^T\right) \mathbf{C}_0^{\frac{1}{2}} \mathbf{B}^{\frac{1}{2}}, \tag{14.40}$$

[6] More precisely, $L_{\mu}^{C,\beta}(x)$ is the Fourier transform of $\exp(-C|k|^{\mu}(1 + \mathrm{i}\beta\,\mathrm{sign}(k)\tan(\pi\mu/2)))$ for $\mu \neq 1$, and of $\exp(-C|k|(1 + \mathrm{i}(2\beta/\pi)\mathrm{sign}(k)\log|k|))$ for $\mu = 1$.

https://doi.org/10.1017/9781108768900.017 Published online by Cambridge University Press

where $\mathbf{u}$ is a normalized eigenvector of $\mathbf{C}_0$ with eigenvalue λ_0, and $\mathbf{B}$ is positive semi-definite, free from $\mathbf{C}_0$, and with $\tau(\mathbf{B}) = 1$. In the special case where $\mathbf{B}$ is a white Wishart, our problem corresponds to a noisy observation of a perturbed covariance matrix, where one of the modes (the one corresponding to $\mathbf{u}$) has a variance boosted by a factor $(1 + a)$.

The matrix $\mathbf{E}_0 := \mathbf{B}^{\frac{1}{2}}\mathbf{C}_0\mathbf{B}^{\frac{1}{2}}$ has an unperturbed spectrum with a lower edge λ_- and an upper edge λ_+. We want to establish, as in the additive case, the condition for the existence of an outlier $\lambda_1 > \lambda_+$ or $\lambda_1 < \lambda_-$, and the exact position of the outlier when it exists.

The eigenvalues of $\mathbf{E}$ are the zeros of its characteristic polynomial, in particular for the largest eigenvalue λ_1 we have

$$\det(\lambda_1\mathbf{1} - \mathbf{E}_0 - a\mathbf{B}^{\frac{1}{2}}\mathbf{C}_0\mathbf{u}\mathbf{u}^T\mathbf{B}^{\frac{1}{2}}) = 0. \tag{14.41}$$

We are looking for an eigenvalue outside the spectrum of $\mathbf{E}_0$, i.e. $\lambda_1 > \lambda_+$ or $\lambda_1 < \lambda_-$. For such a λ_1, the matrix $\lambda_1\mathbf{1} - \mathbf{E}_0$ is invertible and we can use the matrix determinant lemma Eq. (1.30):

$$\det(\mathbf{A} + \mathbf{u}\mathbf{v}^T) = \det\mathbf{A} \cdot \left(1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}\right), \tag{14.42}$$

with $\mathbf{A} = \lambda_1\mathbf{1} - \mathbf{E}_0$ and $\mathbf{u} = -\mathbf{v} = \sqrt{a}\mathbf{B}^{\frac{1}{2}}\mathbf{C}_0^{\frac{1}{2}}\mathbf{u}$. Equation (14.41) becomes

$$\det(\lambda_1\mathbf{1} - \mathbf{E}_0) \cdot \left(1 - a\mathbf{u}^T\mathbf{C}_0^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}}\mathbf{G}_0(\lambda_1)\mathbf{B}^{\frac{1}{2}}\mathbf{C}_0^{\frac{1}{2}}\mathbf{u}\right) = 0, \tag{14.43}$$

where we have introduced the matrix resolvent $\mathbf{G}_0(\lambda_1) := (\lambda_1\mathbf{1} - \mathbf{E}_0)^{-1}$. As we said, the matrix $\lambda_1\mathbf{1} - \mathbf{E}_0$ is invertible so its determinant is non-zero. Thus any outlier needs to solve

$$a\lambda_0\mathbf{u}^T\mathbf{B}^{\frac{1}{2}}\mathbf{G}_0(\lambda_1)\mathbf{B}^{\frac{1}{2}}\mathbf{u} = 1. \tag{14.44}$$

Again we assume that $\mathbf{B}$ is a rotationally invariant matrix with respect to $\mathbf{C}_0$. Then we know that in the large N limit $\mathbf{G}_0(z)$ is diagonal in the basis of $\mathbf{B}$ and reads (see, *mutatis mutandis*, Eq. (13.47))

$$\mathbf{G}_0(z) \approx S^*(z)\mathbf{G}_\mathbf{B}(zS^*(z)), \qquad S^*(z) := S_{\mathbf{C}_0}(z\mathrm{g}_0(z) - 1). \tag{14.45}$$

Furthermore, since $\mathbf{u}$ and $\mathbf{B}$ are also free,

$$\mathbf{u}^T\mathbf{B}^{\frac{1}{2}}\mathbf{G}_0(\lambda_1)\mathbf{B}^{\frac{1}{2}}\mathbf{u} \approx N^{-1}\,\mathrm{Tr}\left(\mathbf{B}^{\frac{1}{2}}G_0(\lambda_1)\mathbf{B}^{\frac{1}{2}}\right) \equiv S^*(\lambda_1)\mathrm{t}_\mathbf{B}(\lambda_1 S^*(\lambda_1)), \tag{14.46}$$

where we have recognized the T-transform of the matrix $\mathbf{B}$:

$$\mathrm{t}_\mathbf{B}(z) := \tau\left(\mathbf{B}(z - \mathbf{B})^{-1}\right). \tag{14.47}$$

Thus, the position of the outlier λ_1 is given by the solution of

$$a\lambda_0 S^*(\lambda_1)\mathrm{t}_\mathbf{B}(\lambda_1 S^*(\lambda_1)) = 1. \tag{14.48}$$

In order to keep the calculation simple, we now assume that $\mathbf{C}_0 = \mathbf{1}$. In this case, $S^* = 1$ and $\lambda_0 = 1$, so the equation simplifies to

https://doi.org/10.1017/9781108768900.017 Published online by Cambridge University Press

$$a\mathrm{t}_{\mathbf{B}}(\lambda_1) = 1. \tag{14.49}$$

To know whether this equation has a solution we need to know if $\mathrm{t}_{\mathbf{B}}(\zeta)$ is invertible. The argument is very similar to the one for $\mathrm{g}(z)$ in the additive case. In the large N limit, $\mathrm{t}_{\mathbf{B}}(\zeta)$ converges to

$$\mathrm{t}_{\mathbf{B}}(\zeta) = \int_{\lambda_-}^{\lambda_+} \frac{\rho_{\mathbf{B}}(x)x}{\zeta - x}\mathrm{d}x. \tag{14.50}$$

So $\mathrm{t}_{\mathbf{B}}(\zeta)$ is monotonically decreasing for $\zeta > \lambda_+$ and is therefore invertible. We then have

$$\lambda_1 = \zeta(a^{-1}) \qquad \text{when} \qquad \lambda_1 > \lambda_+, \tag{14.51}$$

where we use the notation $\zeta(t)$ for the inverse of the T-transform of $\mathbf{B}$, in the region where it is invertible.

The inverse function $\zeta(t)$ can be expressed in terms of the S-transform via Eq. (11.92). We get

$$\lambda_1 = \zeta(a^{-1}) = \frac{a+1}{S_{\mathbf{B}}(a^{-1})} \qquad \text{when} \qquad a > \frac{1}{\mathrm{t}_{\mathbf{B}}(\lambda_+)}. \tag{14.52}$$

Applying the theory to a Wishart matrix $\mathbf{B} = \mathbf{W}$ with

$$S_{\mathbf{W}}(x) = \frac{1}{1+qx}, \quad \lambda_\pm = (1 \pm \sqrt{q})^2, \tag{14.53}$$

one finds that an outlier appears to the right of λ_+ for $a > \sqrt{q}$, with

$$\lambda_1 = (a+1)\left(1 + \frac{q}{a}\right). \tag{14.54}$$

For large a, we have $\lambda_1 \approx a + 1 + q$, i.e. a large eigenvalue $a+1$ in the covariance matrix $\mathbf{C}$ will appear shifted by q in the eigenvalues of the sample covariance matrix.

Nothing prevents us from considering negative values of a, such that $a > -1$ to preserve the positive definite nature of $\mathbf{C}$. In this case, an outlier appears to the left of λ_- when $a < -\sqrt{q}$. Its position is given by the very same equation (14.54) as above.

Exercise 14.4.1 Transpose version of multiplicative perturbation

Consider a positive definite rotationally invariant random matrix $\mathbf{B}$ and a normalized vector $\mathbf{u}$. In this exercise, we will show that the matrix $\mathbf{F}$ defined by

$$\mathbf{F} = (\mathbf{1} + c\mathbf{u}\mathbf{u}^T)\mathbf{B}(\mathbf{1} + c\mathbf{u}\mathbf{u}^T), \tag{14.55}$$

with $c > 0$ sufficiently large, has an outlier λ_1 given by Eq. (14.52) with $b+1 = (c+1)^2$.

(a) Show that for two positive definite matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{B}^{\frac{1}{2}}\mathbf{A}\mathbf{B}^{\frac{1}{2}}$ has the same eigenvalues as $\mathbf{A}^{\frac{1}{2}}\mathbf{B}\mathbf{A}^{\frac{1}{2}}$.

https://doi.org/10.1017/9781108768900.017 Published online by Cambridge University Press

(b) Show that for a normalized vector $\mathbf{u}$

$$\left(\mathbf{1} + (a-1)\mathbf{u}\mathbf{u}^T\right)^{\frac{1}{2}} = \mathbf{1} + (\sqrt{a}-1)\mathbf{u}\mathbf{u}^T. \tag{14.56}$$

(c) Finish the proof of the above statement.

Exercise 14.4.2 Multiplicative perturbation of an inverse-Wishart matrix

We will see in Section 15.2.3 that the inverse-Wishart matrix $\mathbf{M}_p$ is defined as

$$\mathbf{M}_p = (1-q)\mathbf{W}_q^{-1}, \tag{14.57}$$

where $\mathbf{W}_q$ is a Wishart matrix with parameter q and p the variance of the inverse-Wishart is given by $p = \frac{q}{1-q}$. The S-transform of $\mathbf{M}_p$ is given by

$$S_{\mathbf{M}_p}(t) = 1 - pt. \tag{14.58}$$

Consider the diagonal matrix $\mathbf{D}$ with $\mathbf{D}_{11} = d$ and all other diagonal entries equal to 1.

(a) $\mathbf{D}$ can be written as $\mathbf{1} + c\mathbf{u}\mathbf{u}^T$. What is the normalized vector $\mathbf{u}$ and the constant c?

(b) Using the result from Exercise 14.4.1, find the value of the largest eigenvalue of the matrix $\mathbf{D}\mathbf{M}_p\mathbf{D}$ as a function of d. Note that your expression will only be valid for sufficiently large d.

(c) Numerically generate matrices $\mathbf{M}_p$ with $N = 1000$ and $p = 1/2$ ($q = 1/3$). Find the largest eigenvalue of $\mathbf{D}\mathbf{M}_p\mathbf{D}$ for various values of d and make a plot of λ_1 vs d. Superimpose your analytical result.

(d) (Harder) Find analytically the minimum value of d to have an outlier λ_1.

14.5 Phase Retrieval and Outliers

Optical detection devices like CCD cameras or photosensitive films measure the photon flux but are blind to the phase of the incoming light. More generally, it is often the case that one can only measure the power spectral density of a signal, which is the magnitude of its Fourier transform. Can one recover the full signal based on this partial information? This problem is called phase retrieval and can be framed mathematically as follows. Let an unknown vector $\mathbf{x} \in \mathbb{R}^N$ be "probed" with T vectors $\mathbf{a}_k$, in the sense that the measurement apparatus gives us $y_k = |\mathbf{a}_k^T\mathbf{x}|^2$ with $k = 1, \ldots, T$.[7] Vectors $\mathbf{x}$ and $\mathbf{a}_k$ are taken to be real but they can easily be made complex.

The phase retrieval problem is

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left(\sum_k \left| |\mathbf{a}_k^T\mathbf{x}|^2 - y_k \right|^2 \right). \tag{14.59}$$

[7] We could consider that there is some additional noise in the measurement of y_k but for simplicity we keep here with the noiseless version.

https://doi.org/10.1017/9781108768900.017 Published online by Cambridge University Press

It is a difficult non-convex optimization problem with many local minima. To efficiently find an acceptable solution, we need a starting point $\mathbf{x}_0$ that somehow points in the direction of the true solution $\mathbf{x}$. The problem is that in large dimensions the probability that a random vector $\mathbf{x}_0$ has an overlap $|\mathbf{x}^T\mathbf{x}_0| > \varepsilon$ is exponentially small in N as soon as $\varepsilon > 0$. We will explore here a technique that allows one to find a vector $\mathbf{x}_0$ with non-vanishing overlap with the true $\mathbf{x}$.

The idea is to build some sort of weighted Wishart matrix such that this matrix will have an outlier with non-zero overlap with the unknown true vector $\mathbf{x}$. Consider the following matrix:

$$\mathbf{M} = \frac{1}{T}\sum_{k=1}^{T} f(y_k)\,\mathbf{a}_k\mathbf{a}_k^T, \tag{14.60}$$

where the T vectors $\mathbf{a}_k$ are of size N and $f(y)$ is a function that we will choose later. The function $f(y)$ should be bounded above, otherwise we might have outliers dominated by a few large values of $f(y_k)$. One such function that we will study is the simple threshold $f(y) := \Theta(y-1)$. In large dimensions the results should not depend on the precise statistics of the vectors $\mathbf{a}_k$ provided they are sufficiently random. Here we will assume that all their components are standard IID Gaussian. This assumption makes the problem invariant by rotation. Without loss of generality, we can assume the true vector $\mathbf{x}$ is in the canonical direction $\mathbf{e}_1$. The weights $f(y_k)$ are therefore assumed to be correlated to $|[\mathbf{a}_k]_1|^2 = |\mathbf{a}_k^T\mathbf{e}_1|^2$ and independent of all other components of the vectors $\mathbf{a}_k$.

Given that the first row and column of $\mathbf{M}$ contain the element $[\mathbf{a}_k]_1$, we write the matrix in block form as in Section 1.2.5:

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{pmatrix}, \tag{14.61}$$

with the (11) block of size 1×1 and the (22) block of size $(N-1)\times(N-1)$. To find a potential outlier, we look for the zeros of the Stieltjes transform $g_{\mathbf{M}}(z) = \tau((z\mathbf{1}-\mathbf{M})^{-1})$. Combining Eqs. (1.32) and (1.33),

$$Ng_{\mathbf{M}}(z) = \operatorname{Tr}\mathbf{G}_{22}(z) + \frac{1+\operatorname{Tr}\left[\mathbf{G}_{22}(z)\mathbf{M}_{21}\mathbf{M}_{12}\mathbf{G}_{22}(z)\right]}{z-\mathbf{M}_{11}-\mathbf{M}_{12}\mathbf{G}_{22}(z)\mathbf{M}_{21}}, \tag{14.62}$$

where $\mathbf{G}_{22}(z)$ is the matrix resolvent of the rotationally invariant matrix $\mathbf{M}_{22}$ (i.e. the matrix $\mathbf{M}$ without its first row and column). In the large N limit we expect M_{22} to have a continuous spectrum with an edge λ_+. When the condition for the denominator to vanish, i.e.

$$\lambda_1 - \mathbf{M}_{11} - \mathbf{M}_{12}\mathbf{G}_{22}(\lambda_1)\mathbf{M}_{21} = 0, \tag{14.63}$$

has a solution for $\lambda_1 > \lambda_+$ we can say that the matrix $\mathbf{M}$ has an outlier. The overlap between the corresponding eigenvector $\mathbf{v}_1$ and $\mathbf{x}$ is given by the residue

$$\varrho := \frac{|\mathbf{v}_1^T\mathbf{x}|^2}{|\mathbf{x}|^2} = |\mathbf{v}_1^T\mathbf{e}_1|^2 = \lim_{z\to\lambda_1}\frac{z-\lambda_1}{z-\mathbf{M}_{11}-\mathbf{M}_{12}\mathbf{G}_{22}(z)\mathbf{M}_{21}}. \tag{14.64}$$

In the large N limit the scalar equation (14.63) becomes self-averaging. We have

$$\mathbf{M}_{11} = \frac{1}{T}\sum_{k=1}^{T} f(y_k)([\mathbf{a}_k]_1)^2 \overset{T\to\infty}{\to} \mathbb{E}\left[f(y)([\mathbf{a}]_1)^2\right]. \tag{14.65}$$

https://doi.org/10.1017/9781108768900.017 Published online by Cambridge University Press

For the second term we have

$$\mathbf{M}_{12}\mathbf{G}_{22}(z)\mathbf{M}_{21} = \sum_{k,\ell=1}^{T} \frac{1}{T^2} f(y_k) f(y_\ell) [\mathbf{a}_k]_1 [\mathbf{a}_\ell]_1 \sum_{i,j>1}^{N} [\mathbf{a}_k]_i [\mathbf{G}_{22}(z)]_{ij} [\mathbf{a}_\ell]_j$$
$$\overset{T\to\infty}{\to} q\mathbb{E}\left[f^2(y)([\mathbf{a}]_1)^2\right] h(z), \tag{14.66}$$

where $q = N/T$ and

$$h(z) = \tau\left(\frac{\mathbf{H}\mathbf{H}^T}{T}\mathbf{G}_{22}(z)\right), \quad [\mathbf{H}]_{ik} = [\mathbf{a}_k]_i \quad i > 1. \tag{14.67}$$

We can now put everything together and use l'Hospital's rule to compute the residue. For convenience we define the constants $c_n := \mathbb{E}\left[f^n(y)([\mathbf{a}]_1)^2\right]$. There will be an outlier with overlap

$$\varrho = \frac{1}{1 - qc_2 h'(\lambda_1)} \tag{14.68}$$

when there is a solution $\lambda_1 > \lambda_+$ of

$$\lambda_1 = c_1 + qc_2 h(\lambda_1). \tag{14.69}$$

We will come back later to the computation of $h(z)$. In the $q \to 0$ limit the matrix $\mathbf{M}$ becomes proportional to the identity $\mathbf{M} = \mathbb{E}[f(y)]\mathbf{1} := m_1\mathbf{1}$, so $\mathfrak{g}_{\mathbf{M}}(z) = 1/(z - m_1)$ and $h(z) = 1/(z - m_1)$. For $q = 0$ we have a solution $\lambda_1 = c_1$ which satisfies $c_1 \geq m_1$. In this limit the overlap tends to one. The linear correction in q is easily obtained as we only need $h(z)$ to order zero. We obtain

$$\lambda_1 = c_1 + q\frac{c_2}{c_1 - m_1} + O(q^2), \quad \varrho = 1 - q\frac{c_2}{(c_1 - m_1)^2} + O(q^2). \tag{14.70}$$

Note that $c_2/(c_1 - m_1)^2$ is always positive so the overlap decreases with q starting from $\varrho = 1$ at $q = 0$. For the unit thresholding function $f(y) = \Theta(y - 1)$ we have $m_1 = \operatorname{erfc}(1/\sqrt{2}) \approx 0.317$ and $c_1 = c_2 = m_1 + \sqrt{2/(\mathrm{e}\pi)} \approx 0.801$ (see Fig. 14.7).

Since we have the freedom to choose any bounded function $f(y)$ we should choose the one that gives the largest overlap for the value of q given by our dataset. We will do an easier computation, namely minimize the slope of the linear approximation in q. We want

$$f_{\text{opt}}(y) = \operatorname*{argmin}_{f(y)} \frac{c_2}{(c_1 - m_1)^2} = \operatorname*{argmin}_{f(y)} \frac{\mathbb{E}_a\left[f^2(a^2)a^2\right]}{\mathbb{E}_a\left[f(a^2)(a^2 - 1)\right]^2}, \tag{14.71}$$

where the law of a is $\mathcal{N}(0, 1)$. A variational minimization gives

$$f_{\text{opt}}(y) = 1 - \frac{1}{y}. \tag{14.72}$$

The optimal function is not bounded below and therefore the distribution of eigenvalues is singular with $c_2 \to \infty$ and $m_1 \to -\infty$. One should think of this function as the limit of a series of functions such that $c_2/(c_1 - m_1)^2 \to 0$. In Figure 14.7 we see that numerically this function has indeed an overlap as a function of q with zero slope at the origin. As a consequence it has non-zero overlap for much greater values of q (fewer data T) than the simple thresholding function. It turns out that our small q optimum $f(y) = 1 - 1/y$ is actually the optimal function for all values of q.

https://doi.org/10.1017/9781108768900.017 Published online by Cambridge University Press

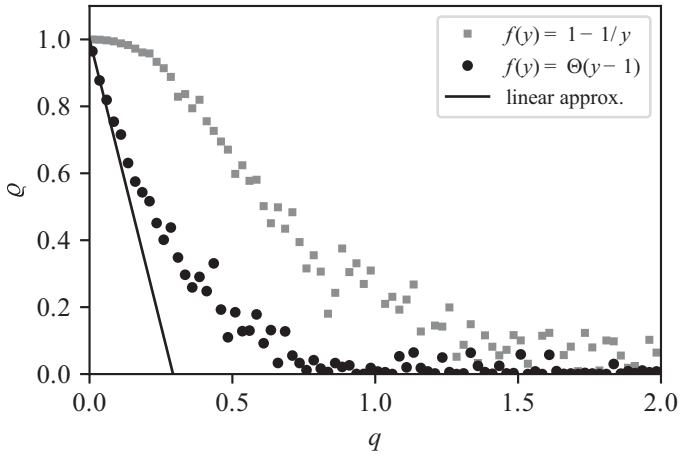

Figure 14.7 Overlap $\varrho := \left|\mathbf{v}_1^T \mathbf{x}\right|^2 / |\mathbf{x}|^2$ between the largest eigenvector and the true signal as a function of $q := N/T$ for two functions: the simple $f(y) = \Theta(y-1)$ and the optimal $f(y) = 1 - 1/y$. Each dot corresponds to a single matrix of aspect ratio q and $NT = 10^6$. The solid line corresponds to the linear q approximation Eq. (14.70) in the thresholding case. For the optimal case, the slope at the origin is zero.

For completeness we show here how to compute the function $h(z)$. We have the following subordination relation (for the (22) block of the relevant matrices):

$$\mathrm{t}_{\mathbf{M}}(\zeta) = \mathrm{t}_{\mathbf{W}_q}\left(S_f(q\mathrm{t}_{\mathbf{M}}(\zeta))\zeta\right), \tag{14.73}$$

where $S_f(t)$ is the S-transform of a diagonal matrix with entries $f(y_k)$. We then have

$$\begin{aligned} h(z) = \tau(\mathbf{W}_q \mathbf{G}_{22}(z)) &= S\tau\left[(Sz\mathbf{1} - \mathbf{M})^{-1}(\mathbf{M} - Sz + Sz)\right] \\ &= S(z\mathfrak{g}_{\mathbf{M}}(z) - 1), \end{aligned} \tag{14.74}$$

with $S := S_f(q(z\mathfrak{g}_{\mathbf{M}}(z) - 1))$. Since

$$S_{\mathbf{M}}(t) = \frac{S_f(qt)}{1+qt} = \frac{t+1}{t\zeta}, \tag{14.75}$$

we have

$$h(z) = \mathfrak{g}_{\mathbf{M}}(z)[1 + q(z\mathfrak{g}_{\mathbf{M}}(z) - 1)], \tag{14.76}$$

where the function $\mathfrak{g}_{\mathbf{M}}(z)$ can be obtained by inverting the relation

$$\mathfrak{z}(g) = \int_{-\infty}^{\infty} \frac{\mathrm{d}x}{\sqrt{2\pi}} \frac{f(x^2)\mathrm{e}^{-x^2/2}}{1 - qgf(x^2)} + \frac{1}{g}. \tag{14.77}$$

Bibliographical Notes

- For a recent review on the subject of free low-rank additive or multiplicative perturbations, the interested reader should consult

https://doi.org/10.1017/9781108768900.017 Published online by Cambridge University Press

  - M. Capitaine and C. Donati-Martin. Spectrum of deformed random matrices and free probability. *preprint arXiv:1607.05560*, 2016.
- Equation (14.16) was first published by both physicists and mathematicians, respectively, in
  - S. F. Edwards and R. C. Jones. The eigenvalue spectrum of a large symmetric random matrix. *Journal of Physics A: Mathematical and General*, 9(10):1595, 1976,
  - Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.
- In 2005, Baik, Ben Arous and Péché studied the transition in the statistics of the largest eigenvector in the presence of a low-rank perturbation. The transition from no outlier to one outlier is now called the BBP (Baik–Ben Arous–Péché) transition; see
  - J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, 33(5):1643–1697, 2005,
  - D. Féral and S. Péché. The largest eigenvalue of rank one deformation of large Wigner matrices. *Communications in Mathematical Physics*, 272(1):185–228, 2007.
- The case of outliers with a hard edge is treated in
  - S. Péché. *Universality of local eigenvalue statistics for random sample covariance matrices*. PhD thesis, EPFL, 2003.
- The generalization to spiked tensors has recently been studied in
  - G. Ben Arous, S. Mei, A. Montanari, and M. Nica. The landscape of the spiked tensor model. *Communications on Pure and Applied Mathematics*, 72(11):2282–2330, 2019.
- The case of additive or multiplicative perturbation to a general matrix (beyond Wigner and Wishart) was worked out by
  - F. Benaych-Georges and R. R. Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011.
- For studies about the edge properties of Wigner matrices, see e.g.
  - M. J. Bowick and E. Brzin. Universal scaling of the tail of the density of eigenvalues in random matrix models. *Physics Letters B*, 268(1):21–28, 1991,

  and for the specific problem of the largest eigenvalue, the seminal paper of
  - C. A. Tracy and H. Widom. Level-spacing distributions and the Airy kernel. *Communications in Mathematical Physics*, 159(1):151–174, 1994,

  which has led to a flurry of activity, see e.g.
  - S. Majumdar. Random matrices, the Ulam problem, directed polymers and growth models, and sequence matching. In *Les Houches Lecture Notes "Complex Systems"*, volume 85. Elsevier Science, 2007.
- When the elements of the random matrix are fat tailed, the Tracy–Widom law is modified, see
  - G. Biroli, J.-P. Bouchaud, and M. Potters. On the top eigenvalue of heavy-tailed random matrices. *Europhysics Letters (EPL)*, 78(1):10001, 2007,

- A. Auffinger, G. Ben Arous, and S. Péché. Poisson convergence for the largest eigenvalues of heavy tailed random matrices. *Annales de l'I.H.P. Probabilités et statistiques*, 45(3):589–610, 2009.

- For studies of infinite variance Lévy matrices, see
  - P. Cizeau and J. P. Bouchaud. Theory of Lévy matrices. *Physical Review E*, 50:1810–1822, 1994,
  - Z. Burda, J. Jurkiewicz, M. A. Nowak, G. Papp, and I. Zahed. Free random Lévy and Wigner-Lévy matrices. *Physical Review E*, 75:051126, 2007,
  - G. Ben Arous and A. Guionnet. The spectrum of heavy tailed random matrices. *Communications in Mathematical Physics*, 278(3):715–751, 2008,
  - S. Belinschi, A. Dembo, and A. Guionnet. Spectral measure of heavy tailed band and covariance random matrices. *Communications in Mathematical Physics*, 289(3):1023–1055, 2009.
- On the outlier method for the phase recovery problem, the reader is referred to
  - Y. M. Lu and G. Li. Phase transitions of spectral initialization for high-dimensional nonconvex estimation. *preprint arXiv:1702.06435*, 2017,
  - W. Luo, W. Alghamdi, and Y. M. Lu. Optimal spectral initialization for signal recovery with applications to phase retrieval. *IEEE Transactions on Signal Processing*, 67(9):2347–2356, 2019,

  and for a review, see
  - A. Stern, editor. *Optical Compressive Imaging*. CRC Press, Boca Raton, Fla., 2017.