4 The resolvent

We saw in the previous lectures how to use the joint PDF of eigenvalues to extract the spectral density for GOE matrices. However, this involved some nontrivial math of solving integral equations. Now we will see a more versatile route to the spectral density that doesn't rely on the JPDF of eigenvalues.

Given a matrix H, consider the complex function G defined by

$$G(z) = \frac{1}{N}\overline{\mathrm{Tr}(zI - H)^{-1}} \tag{1}$$

When averaged over matrices H, this is the *resolvent*, *Green's function*, or *Stieltjes transform*. If $\boldsymbol{x}$ is the vector of eigenvalues of H, then we can equivalently write

$$G(z) = \frac{1}{N}\overline{\sum_{i=1}^{N}\frac{1}{z - x_i}} \tag{2}$$

So, G is a complex function with a simple pole at each of the eigenvalues of H. If H is a symmetric or Hermitian matrix and therefore its eigenvalues are real, then the poles of G will be only on the real line. We can use complex analysis to count the poles in a given region of the complex plane. From the residue theorem, we know that

$$\oint G(z)\,dz = 2\pi i \sum_{z^*\in\circ} \mathrm{Res}(G, z^*) = 2\pi i \sum_{z^*\in\circ}\frac{1}{N} \tag{3}$$

where in the last step we used the fact that the residue of the poles in G are all $\frac{1}{N}$. Consider the rectangular contour in the complex plane defined by

$$x - i\epsilon \to x + dx - i\epsilon \to x + dx + i\epsilon \to x + i\epsilon \to x - i\epsilon \tag{4}$$

On one hand we can use the residue theorem to evaluate the integral of G around this contour, and we find

$$\oint G(z)\,dz = 2\pi i \frac{\text{number of eigenvalues of } H \text{ between } x \text{ and } dx}{N} = 2\pi i \rho(x)\,dx \tag{5}$$

where in the last step we assume to be working at large N so that the empirical density is close to the average density. On the other hand, we can evaluate the integral directly:

$$\begin{aligned}\oint G(z)\,dz = &\frac{1}{2}(G(x - i\epsilon) + G(x + dx - i\epsilon))dx \\ &+\frac{1}{2}(G(x + dx - i\epsilon) + G(x + dx + i\epsilon))2\epsilon \\ &-\frac{1}{2}(G(x + dx + i\epsilon) + G(x + i\epsilon))dx \\ &-\frac{1}{2}(G(x + i\epsilon) + G(x - i\epsilon))2\epsilon \\ &= (G(x - i\epsilon) - G(x + i\epsilon))\,dx\end{aligned} \tag{6}$$

So therefore we find that

$$\rho(x) = \frac{1}{2\pi i} \lim_{\epsilon \to 0} \big(G(x - i\epsilon) - G(x + i\epsilon) \big) \tag{7}$$

In the large-N limit, the accumulation of poles due to the individual eigenvalues becomes a *branch cut*, and the difference in the resolvent G across the cut gives the spectral density. Since we know the spectral density is real-valued, we can conclude from the above formula that the discontinuity is only in the imaginary part of G, or

$$\rho(x) = \frac{1}{2\pi} \lim_{\epsilon \to 0} \operatorname{Im} \big(G(x - i\epsilon) - G(x + i\epsilon) \big) = \frac{1}{\pi} \lim_{\epsilon \to 0} \operatorname{Im} G(x - i\epsilon) \tag{8}$$

where in the final step we have used the fact that the continuous part of the imaginary part of G must be zero for real z.

This is all nice, but it only does us good if we can actually compute G. Luckily, there are a wealth of ways to do this. Now I will show a fairly direct method with deep connections to standard field theory. First, note that G is also the moment generating function for H, since by expanding in z we have

$$\begin{aligned} G(z) &= \sum_{k=0}^{\infty} \frac{\mu_k}{z^{k+1}} = \sum_{k=0}^{\infty} \frac{\frac{1}{N}\overline{\operatorname{Tr} H^k}}{z^{k+1}} = \frac{1}{N}\left(\frac{1}{z} \operatorname{Tr} I + \frac{1}{z^2}\overline{\operatorname{Tr} H} + \frac{1}{z^3}\overline{\operatorname{Tr} H^2} + \cdots \right) \\ &= \frac{1}{N}\left(\frac{N}{z} + \frac{1}{z^2}\sum_{i=1}^{N} \overline{H_{ii}} + \frac{1}{z^3}\sum_{i,j=1}^{N} \overline{H_{ij}H_{ji}} + \frac{1}{z^4}\sum_{i,j,k=1}^{N} \overline{H_{ij}H_{jk}H_{ki}} + \cdots \right) \end{aligned} \tag{9}$$

Because for the GOE ensemble the distribution of the Hs are Gaussian with variance $\frac{1}{N}$, we know that

$$\overline{H_{ij}H_{kl}} = \frac{1}{2}\frac{1}{N}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) \tag{10}$$

Further, because the Hs are Gaussian, we can use Wick's theorem to expand each of the correlations between multipoint correlators. All odd correlations are zero since the distribution of elements is centered, and for even ones we have, e.g.,

$$\overline{H_{ij}H_{jk}H_{kl}H_{li}} = \overline{H_{ij}H_{jk}}\;\overline{H_{kl}H_{li}} + \overline{H_{ij}H_{kl}}\;\overline{H_{jk}H_{li}} + \overline{H_{ij}H_{li}}\;\overline{H_{jk}H_{kl}} \tag{11}$$

We can then represent this series expansion using diagrams. We have two kinds of lines: straight horizontal lines that carry $\frac{1}{z}\delta_{ij}$, where the number of such lines gives the order of the diagram in z. Curved double lines carry $\overline{H_{ij}H_{kl}} = \frac{1}{2}\frac{1}{N}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})$. We have no vertices: the curved double lines can only intersect the straight horizontal lines. We have no vertices: the curved double lines can only intersect the straight horizontal lines, and where they meet their indices coincide.[1] The value of a diagram is then given by summing over all indices and dividing by N.

[1] If we had a non-Gaussian theory, we would have a construction identical to this, but with vertices present. The figures here I stole from PRE 84, 061125 (2011), where they do consider non-Gaussian theories and therefore the figures have vertices.

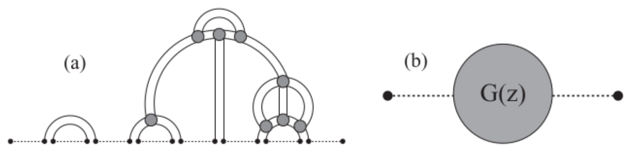

FIG. 1. (a) An example of a diagram contributing to the generating function $G_{ij}(z)$. Two end points should be labeled by indices ij. Each horizontal dashed line corresponds to $\frac{1}{z}\delta_{ab}$ while a double line represents the expectation value (the propagator) $\langle H_{ab}H_{cd}\rangle_0 = \frac{1}{Ng_2}\delta_{ad}\delta_{bc}$. Since all lines in the diagram are proportional to the δ function this equation reduces to a scalar equation for $G(z)$. Each horizontal dashed line corresponds after this reduction to $\frac{1}{z}$, each double line to $\frac{1}{g_2}$, each vertex to g_n. The diagram has eight horizontal lines so it contributes to the term $\frac{\mu_7}{z^8}$ in the series expansion (9). The diagram contains seven cubic vertices g_3^7 and one quartic vertex g_4 that are generated by the perturbative expansion of the residual part of (15). Each pair of dots on the horizontal line corresponds to a factor H_{ab} inside the average $\langle \mathrm{Tr} H^7\rangle_0$. (b) The graphical notation for the generating function $G(z)$. It generates diagrams having two end points which include, for example, the one shown on the left.

First, we want to do the standard field theory simplification. Consider defining $\Sigma(z)$ as the sum over contributions from all diagrams that are one-line-irreducible, meaning that they cannot be split into two disconnected diagrams by cutting one line. It follows that G, which sums over all diagrams, is given by the sum over copies of $\Sigma(z)$ joined together by horizontal lines, implying

$$G(z) = \frac{1}{z} + \frac{1}{z}\Sigma(z)\frac{1}{z} + \frac{1}{z}\Sigma(z)\frac{1}{z}\Sigma(z)\frac{1}{z} + \cdots = \frac{1}{z-\Sigma(z)} \tag{12}$$

In field theory Σ is known as the self-energy.

FIG. 2. Diagrams $G(z)$ can be obtained from one-line-irreducible diagrams $\Sigma(z)$ (see Fig. 3) by joining them one after another.

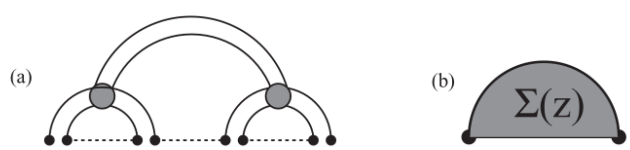

FIG. 3. (a) An example of a one-line-irreducible diagram. (b) The graphical notation for the generating function $\Sigma(z)$ of one-line-irreducible diagrams.

Second, we recognize a special simplification of this field theory. The value of every diagram is given by the sum over many indices of δ-functions, but depending on the way the δ-functions are strung together they can result in different powers of N. In the above, for instance, consider

$$\sum_{ijkl} \overline{H_{ij}H_{jk}}\,\overline{H_{kl}H_{li}} = \frac{1}{(2N)^2}\sum_{ijkl}(\delta_{ij}\delta_{jk} + \delta_{ik}\delta_{jj})(\delta_{kl}\delta_{li} + \delta_{ki}\delta_{ll}) \tag{13}$$

$$= \frac{1}{(2N)^2}\sum_{ijkl}(\delta_{ij}\delta_{jk}\delta_{kl}\delta_{li} + \delta_{ik}\delta_{jj}\delta_{kl}\delta_{li} + \delta_{ij}\delta_{jk}\delta_{ki}\delta_{ll} + \delta_{ik}\delta_{jj}\delta_{ki}\delta_{ll})$$

$$= \frac{1}{(2N)^2}\sum_{ijk}(\delta_{ij}\delta_{jk}\delta_{ki} + \delta_{ik}\delta_{jj}\delta_{ki} + N\delta_{ij}\delta_{jk}\delta_{ki} + N\delta_{ik}\delta_{jj}\delta_{ki})$$

$$= \frac{1}{(2N)^2}\sum_{ij}(\delta_{ij}\delta_{ji} + \delta_{ii}\delta_{jj} + N\delta_{ij}\delta_{ji} + N\delta_{ii}\delta_{jj})$$

$$= \frac{1}{(2N)^2}\sum_{i}(\delta_{ii} + N\delta_{ii} + N\delta_{ii} + N^2\delta_{ii})$$

$$= \frac{1}{(2N)^2}(N + 2N^2 + N^3) = \frac{N}{2^2} + o(N)$$

versus

$$\sum_{ijkl} \overline{H_{ij}H_{kl}}\,\overline{H_{jk}H_{li}} = \frac{1}{(2N)^2}\sum_{ijkl}(\delta_{ik}\delta_{jl} + \delta_{jk}\delta_{il})(\delta_{jl}\delta_{ki} + \delta_{ji}\delta_{kl}) \tag{14}$$

$$= \frac{1}{(2N)^2}\sum_{ijkl}(\delta_{ik}\delta_{jl}\delta_{jl}\delta_{ki} + \delta_{jk}\delta_{il}\delta_{jl}\delta_{ki} + \delta_{ik}\delta_{jl}\delta_{ji}\delta_{kl} + \delta_{jk}\delta_{il}\delta_{ji}\delta_{kl})$$

$$= \frac{1}{(2N)^2}\sum_{ijk}(\delta_{ik}\delta_{jj}\delta_{ki} + \delta_{jk}\delta_{ij}\delta_{ki} + \delta_{ik}\delta_{jk}\delta_{ji} + \delta_{jk}\delta_{ik}\delta_{ji})$$

$$= \frac{1}{(2N)^2}\sum_{ij}(\delta_{ii}\delta_{jj} + \delta_{ji}\delta_{ij} + \delta_{ij}\delta_{ji} + \delta_{ji}\delta_{ji})$$

$$= \frac{1}{(2N)^2}\sum_{i}(N\delta_{ii} + \delta_{ii} + \delta_{ii} + \delta_{ii}) = \frac{1}{(2N)^2}(N^2 + 3N) = o(N)$$

Some diagrams are factors of N smaller than others. What differentiates those that contribute at largest order in N? The answer is given by their topology: only *planar* diagrams contribute to largest order in N. Why is this true? Each chain of δ-functions ending in a loop, i.e., $\delta_{i_1 i_2}\delta_{i_2 i_3}\cdots\delta_{i_n i_1}$, contributes a factor of N when summed. Therefore, the diagrams that contribute most are those that have the largest number of loops, which turn out to be the planar ones.

Why is the planar property interesting? If we wanted to analyze this field theory in the normal way, by summing over all diagrams order by order, it would dramatically reduce the number of diagrams to consider, so this is nice. However, it also gives us the ability to make an independent relationship between G and Σ at large N. Consider the following argument: G is the sum over all diagrams, irreducible or not. However, we can take any non-irreducible diagram and produce an irreducible one at two higher orders by adding a curved double line connecting one end to the other. Every such diagram must be included in Σ, by definition, and every *planar* diagram must be of this form, since any other possibility would result in crossing. Therefore, considering only planar diagrams, it must be true that

$$\Sigma(z) = \frac{1}{2}G(z) \tag{15}$$

This allows us to directly find G, since

$$G(z) = \frac{1}{z - \frac{1}{2}G(z)} \mapsto 0 = -\frac{1}{2}G(z)^2 + zG(z) - 1 \tag{16}$$

which gives from the quadratic equation

$$G(z) = z \mp \sqrt{z^2 - 2} \tag{17}$$

The branch of G must be chosen so that $G(z)$ vanishes as $\|z\|$ gets large, since we know the resolvent decays for z larger than any of the eigenvalues. Therefore, we have

$$\rho(x) = \frac{1}{\pi} \lim_{\epsilon \to 0} \operatorname{Im} G(x - i\epsilon) = \frac{1}{\pi}\sqrt{2 - z^2} \tag{18}$$

the Wigner semicircle.

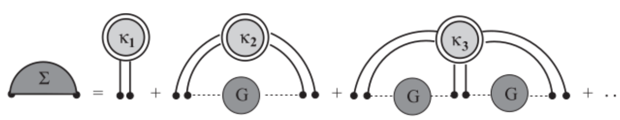

FIG. 4. A one-line-irreducible diagram can be obtained from a one-line-reducible diagram by adding to it a minimal diagrammatic structure complying with the measure (15), which makes it one-line-irreducible. Such a minimal structure is provided by diagrams corresponding to planar connected moments κ_k (free cumulants) (see Fig. 5), which we indicated by bubbles surrounded by double circles in the figure. This double ring around the bubble is chosen to make it similar to double brackets used in our notation for connected averages. Diagrams in such a bubble are connected. The difference between diagrams corresponding to planar connected moments (cumulants) and spectral moments is explained in Fig. 5.