Lecture A1

matrix model

matrix integral

Consider an $N \times N$ hermitian matrix M and its potential of the form $\mathrm{tr}W(M)$, where $W(M)$ is a polynomial of M. It is invariant under conjugation, $\mathrm{tr}W(\Omega M \Omega^{-1}) = \mathrm{tr}W(M)$, where $\Omega \in U(N)$. The partition function Z of the matrix model is defined by the integral,

$$Z = \int dM e^{-\mathrm{tr}\ W(M)},$$

where the measure is $dM = 2^{\frac{1}{2}N(N-1)} \prod_i dM_{ii} \prod_{i<j} d\mathrm{Re}M_{ij} d\mathrm{Im}M_{ij}$ and is invariant under the $U(N)$ action. The matrix model has widespread applications in theoretical physics. It is a toy model of the functional integral of the gauge theory (it can be thought of as a zero-dimensional quantum field theory). It was introduced to understand spectra of atomic nuclei, it was used to understand non-Abelian gauge theory in the limit of a large gauge group, it described dynamics of D branes in string theory in certain situations, it has close connections to quantum geometry of Calabi-Yau manifolds, *etc.*

Feynman diagrams

Let us start with the simplest case of $N = 1$ and when the integral is Gaussian, $W(M) = \frac{a}{2}M^2$,

$$Z_0 = \int_{-\infty}^{\infty} dM e^{-\frac{a}{2}M^2} = \sqrt{\frac{2\pi}{a}}.$$

We can also compute *correlation functions*,

$$\langle M^k \rangle_0 = \frac{1}{Z_0} \int_{-\infty}^{\infty} dM\ M^k\ e^{-\frac{a}{2}M^2} = \frac{1}{Z_0}\left(-2\frac{d}{da}\right)^k Z_0 = (2k-1)!!a^{-k},$$

where $(2k-1)!! = 1 \cdot 3 \cdot 5 \cdots (2k-1)$. This can be used to evaluate a matrix integral for a more complicated potential as,

$$\begin{aligned} Z(a,g) &= \int_{-\infty}^{\infty} dM e^{-\frac{a}{2}M^2 - \frac{g}{4!}M^4} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!}\left(-\frac{g}{4!}\right)^n \langle M^{4n} \rangle_0 Z_0(a) \\ &= \sum_{n=0}^{\infty} \frac{(4n-1)!!}{n!(4!)^n}\left(-\frac{g}{a^2}\right)^n Z_0(a). \end{aligned} \quad (1)$$

Note that the sum is an asymptotic expansion in g, and it is not convergent. The combinatorial factor $(4n-1)!!/n!(4!)^n$ is equal to the number of Feynman diagrams with n vertices with 4 legs each. Namely, it is equal to the number of ways to connect the n M^4 vertices using $2n$ lines (propagators). When a diagram has a symmetry, we divide its contribution by the order of the symmetry group. For example, the figure-8 shape diagram for $n = 1$ has $\mathbf{Z_2} \times \mathbf{Z_2} \times \mathbf{Z_2}$ symmetry, and its contribution to the above sum is $1/8$ times $-g/a^2$.

't Hooft counting

Now we move on to $N > 1$. Consider the potential,

$$W(M) = \frac{1}{2}M^2 + \sum_{p=3}^{\infty} \frac{g_p}{p} M^p.$$

Since both the potential and the measure dM is $U(N)$ invariant, it is reasonable to define the partition function by divising the matrix integral by the volume of the $U(N)$ group given by

$$\text{vol } U(N) = \frac{(2\pi)^{\frac{1}{2}N(N+1)}}{G_2(N+1)},$$

where $G_2(z)$ is the Barnes double-Γ function satisfying,

$$G_2(z+1) = \Gamma(z)G_2(z), \quad G_2(1) = 1.$$

The matrix model partition function is then given by,

$$Z = \frac{1}{\text{vol } U(N)} \int dM e^{-\frac{1}{\lambda}\text{tr}W(M)}.$$

Here we introduced the parameter λ to keep track of the perturbative exapansion. Later we will indentify it with the string coupling constant.

Let us evaluate the integral Z in powers of λ. As in the case of $N = 1$, we should first understand the Gaussian integral with $g_p = 0$ ($p = 3, 4, ...$). We can show, for example,

$$\langle M_{ij} M_{kl} \rangle_{Gaussian} = \lambda \delta_{il} \delta_{jk}.$$

Expectation values of higher powers of M are given as sums of products of this two-point function $\langle M_{ij} M_{kl} \rangle_{Gaussian}$. This property of the Gaussian integral is known as the Wick theorem. Since the actio of the $U(N)$ group is given by $M \to \Omega M \Omega^{-1}$, we can think of M_{ij} as in the tensor product of the fundamental representation i and the anti-fundamental representation j of $U(N)$. We can visualize the two-point function $\langle M_{ij} M_{kl} \rangle_{Gaussian}$ by connecting an arrow from i to l and another arrow from k to j, namely from one of the fundamental representation to one of the anti-fundamental representation.

We can compute $\langle \text{tr}M^2 \rangle_{Gaussian}$ by visualizing a pair of loops, one going clockwisely and another going counter-clockwisely. We should also remember that each double-line (a pair of arrows) is weighted by λ. Thus, we see that it should be equal to λN^2.

Question 1: Compute $\langle \text{tr}M^2 \ \text{tr}M^2 \rangle_{Gaussian}$.

To compute $\langle \text{tr}M^3 \ \text{tr}M^3 \rangle_{Gaussian}$, we can visualize 2 cubic-vertices. From each of the cubic vertices, 3 double-ines are emenating, and we need to tie them together, paying attention to the orientations of the arrows. It turns out that there are 2 topologically distinct ways of doing it. One involves 3 loops of arrows, giving $12N^3\lambda^3$ (we have N^3 since there are 3 loops, and we have λ^3 since there are 3 double-lines). Another one involves just 1 loop, giving $3N\lambda^3$ (we have N since there is only 1 loop). These two Feynman diagrams are topologically distinct. For example, the former can be drawn on a plane sheet of paper without any lines crossing each other. It is not possible to do so for the latter. Another way to see the difference is to pay attention to each loop in the diagrams and to identify a disk bounded by the loop so that the loop goes clockwisely as seen from the top of the disk. By attaching such disks, each Feynman diagram can be turned into a closed surface. In our example, this generates two distinct surfaces. For the first type of Feynman diagrams, the resulting surface is the 2-sphere. For the second type, it is a torus.

In general, the perturbative expansion of the partition function Z generates a sum of possibly disjoint diagrams. We can generate a sum of connected diagrams if we take the logarithm $\log Z$. We can then express is as a sum over connected surfaces. It is interesting to note that surfaces generated in this way are always orientable. This is because we started with arrows with definite orientations and is ultimately because we started with the integral over hermitial matrices. If we had started with an integral over anti-symmetric matrices, we would have generated surfaces without orientations (*e.g.* including the Klein bottle).

Each Feynman diagram is weighted as follows. Let us call the number of p-valent vertices as V_p with the total number of vertices $V = \sum_p V_p$, the number of double-lines (propagators) as E, the number of disks (faces) as F. The weight can be counted as,

$$\prod_p \left(-\frac{g_p}{\lambda}\right)^{V_p} \lambda^E N^F = \lambda^{E-V} N^F \prod_p (-g_p)^{V_p} = \lambda^{-V+E-F} (\lambda N)^F \prod_p (-g_p)^{V_p}.$$

It is interesting to note that the combination $(V - E + F)$ is the Euler characteristic of the surface one gets by attaching disks to the Feynman diagram. For a closed surface,

$$V - E + F = 2 - 2g,$$

where g is the number of handles attached to the surface ($g = 0$ for the 2-sphere, $g = 1$ for the torus, *etc*).

The combination $t = \lambda N$ is called the 't Hooft coupling. When we use the perturbative expansion in λ, we assume that $\lambda \ll 1$. The above observation shows that, if we take this limit while keeping the 't Hoof coupling finite, we can express $\log Z$ as a sum of connected surfaces weighted by λ^{2g-2},

$$F = \log Z = \sum_{g=0}^{\infty} F_g(g_p, t) \lambda^{2g-2}.$$

This is called the 't Hooft expansion or the large N expansion (since $\lambda \ll 1$ with $t = \lambda N$ finite means $N \gg 1$). Gerard 't Hooft speculated that the resulting $F_g(g_p, t)$ should have a nice interpretation in terms of a closed string theory. This turned out to be the case for the matrix model.

eigenvalue integral

Since the matrix integral is invariant under $M \to \Omega M \Omega^{-1}$, we can express it as an integral over eigenvalues of M.

$$Z = \frac{1}{\mathrm{vol} U(N)} \int dM \; e^{-\frac{1}{\lambda}\mathrm{tr} W(M)} = \frac{1}{N!} \int \prod_{i=1}^{N} \frac{d\lambda_i}{2\pi} e^{-\frac{1}{\lambda} W(\lambda_i)} \prod_{i<j} (\lambda_i - \lambda_j)^2.$$

This follows from the following identity,

$$\Delta(M)^2 \cdot \int_{U(N)} d\Omega \prod_{i<j} \delta\Big((\Omega M \Omega^{-1})_{ij}\Big) = 1,$$

where $\Delta(M)$ defined so that,

(1) It is an invariant function, $\Delta(\Omega M \Omega^{-1}) = \Delta(M)$.

(2) It is equal to $\prod_{i<j}(\lambda_i - \lambda_j)$ when M is diagonal.

One can think of $\Delta(M)^2$ as the Faddeev-Popov determinant for the gauge condition, $M_{ij} = 0$ $(i < j)$, namely M being diagonal. Using this,

$$\begin{aligned}
\int dM\ e^{-\frac{1}{\lambda}\mathrm{tr}W(M)} &= \int dM\ e^{-\frac{1}{\lambda}\mathrm{tr}W(M)} \Delta(M)^2 \int_{U(N)} d\Omega \prod_{i<j} \delta\Big((\Omega M \Omega^{-1})_{ij}\Big) \\
&= \int dM'\ e^{-\frac{1}{\lambda}\mathrm{tr}W(M')} \Delta(M')^2 \prod_{i<j} \delta(M'_{ij}) \int_{U(N)} d\Omega, \quad (\text{we set } M' = \Omega M \Omega^{-1}) \\
&= \frac{\mathrm{vol}(U(N))}{(2\pi)^N N!} \int \prod_{i=1}^{N} d\lambda_i e^{\frac{1}{\lambda}W(\lambda_i)} \prod_{i<j} (\lambda_i - \lambda_j)^2. \qquad (2)
\end{aligned}$$

This proves the eigenvalue integral expression for Z. One can think of the factor $1/N!$ as taking care of the residual gauge symmetry of exchanging the eigenvalues, which remains after imposing the gauge condition $M_{ij} = 0$ $(i < j)$.

eigenvalue distribution

We can write the eigenvalue integral as,

$$Z = \frac{1}{N!} \int \prod_i \frac{d\lambda_i}{2\pi} e^{-N^2 V(\lambda)},$$

where

$$V(\lambda) = \frac{1}{N} \sum_i t^{-1} W(\lambda_i) - \frac{1}{N^2} \sum_{i<j} \log(\lambda_i - \lambda_j)^2.$$

One can think of this as the potential energy of N particles in one dimensions with coordinate λ in the potential $t^{-1}W(\lambda)$ and with the repulsive potential $-\log(\lambda_i - \lambda_j)^2$.

In the large N limit, the sum over i gives a factor of N, and $V(\lambda)$ is then of the order 1. In fact, by introducing the eigenvalue distribution function,

$$\rho(\lambda) = \frac{1}{N} \sum_i \delta(\lambda - \lambda_i),$$

which is normalized as

$$\int d\lambda \rho(\lambda) = 1,$$

we can write the potential as a functional of ρ,

$$V(\rho) = \frac{1}{t} \int d\lambda \rho(\lambda) W(\lambda) - \int d\lambda d\lambda' \rho(\lambda)\rho(\lambda') log|\lambda - \lambda'|.$$

Since the integrand for the eigenvalue integral is $\exp(-N^2 V)$, the eigenvalues will try to settle in the configuration to minimize the potential V. A variation of V with respect to ρ gives,

$$\frac{1}{2t} W'(\lambda) = \mathrm{P} \int \frac{\rho(\lambda')}{\lambda - \lambda'} d\lambda'.$$

Here P in the right-hand side means that we take the principal value of the integral. To be precise, the above is obtained by taking a derivative of the equation $\delta V/\delta\rho = 0$ with respect to λ. This turns out to be more convenient than the orignal equation.

resolvent

To solve the eigenvalue equation, it is convenient to introduce the resolvent,

$$\omega(p) = \frac{1}{N} \langle \mathrm{tr} \frac{1}{p - M} \rangle.$$

In the large N limit, it becomes

$$\omega_0(p) = \int d\lambda \frac{\rho(\lambda)}{p - \lambda}.$$

Suppose that the eigenvalue distribution $\rho(\lambda)$ has a finite support C on the real line $\mathbf{R}$. The resolvent $\omega_0(p)$ has a branch cut on C. The resolvent has several properties, which can be derived from its definition:

(1) It is analytic, except on C.

(2) $\omega_0(p) \sim 1/p$ for $p \to \infty$.

(3) $\omega_0(\lambda + i\epsilon) - \omega_0(\lambda - i\epsilon) = -2\pi i \rho(\lambda)$ for $\lambda \in C$ and $0 < \epsilon \ll 1$.

The eigenvalue equation then gives,

(4) $\omega_0(\lambda + i\epsilon) + \omega_0(\lambda - i\epsilon) = -\frac{1}{t} W'(\lambda)$.

Assuming the eigenvalue distribution generates a single cut, $C = [b, a]$, there is an explicit solution to these conditions of the form,

$$\omega_0(p) = \frac{1}{2t} \oint \frac{dz}{2\pi i} \frac{W'(z)}{p - z} \sqrt{\frac{(p-a)(p-b)}{(z-a)(z-b)}}.$$

The end points of the cut are determined by the conditions (1) and (2).

As an example, consider the Gaussian model with $W(\lambda) = \frac{1}{2}\lambda^2$. In this case, we expect $a = -b$ by symmetry, and we have

$$\omega_0(p) = \frac{\sqrt{p^2 - a^2}}{2t} \oint \frac{dz}{2\pi i} \frac{z}{(p - z)\sqrt{z^2 - a^2}}.$$

Requiring (2), one finds $a = 2\sqrt{t}$ and

$$\omega_0(p) = \frac{1}{2t} \left(p - \sqrt{p^2 - 4t} \right).$$

By using (3), the eigenvalue distribution is given by

$$\rho(\lambda) = \frac{1}{2\pi t}\sqrt{4t - \lambda^2}.$$

This is the famous semi-circle law of Eugene Wigner.

Generally speaking, if we write

$$\omega_0(p) = \frac{1}{2t}\left(y(p) + W'(p)\right),$$

$y(p)$ obeys

$$y(p)^2 = W'(p)^2 - R(p),$$

where

$$R(p) = 4t \int d\lambda \rho(\lambda) \frac{W'(p) - W'(\lambda)}{p - \lambda}.$$

If $W(p)$ is a polynomial of p of degree k, $R(p)$ is a polynomial of degree $(k-2)$. Thus, (y,p) obeying the above equation defines an algebraic curve (Riemann surface) in $\mathbf{C}^2$. Here we considered a one-cut solution, but in general the curve can have $(k-1)$ branch cuts. This is as expected since it is the same as the number of extrema of $W(p)$, where eigenvalues can congregate.

Given the algebraic curve, we can consider a complex 3-dimensional manifold defined by

$$uv = y^2 - W'(p)^2 + R(p), \qquad (u,v,y,p) \in \mathbf{C}^4.$$

It turns out that this manifold is Calabi-Yau and we can introduce a Ricci-flat Kähler metric on it. We can consider the closed topological string theory of B-type with this Calabi-Yau manifold as its target space, and its partition function is equal to the matrix model partition function Z. In this case, the topological string theory gives the large N dual of the matrix model anticipated by 't Hooft.