$\boxed{\text{Section 6}}$ Dyson Brownian Motion and Bulk Universality

I. Various notions of Bulk universality

The local eigenvalue statistics can either be expressed in terms of "local correlation functions" rescaled around some energy E or the "gap statistics" for a gap $\lambda_{j+1}-\lambda_j$ with a given label j. They are called "fixed energy" and "fixed gap" universalites, and they do not coincide. In fact, eigenvalues fluctuate on a scale much larger than the typical eigenvalue spacing, the label j of the eigenvalue λ_j closest to a fixed energy E is not a deterministic function of E. Moreover, the two concepts both have natural averaged versions, which are generall easier to establish.

Rmk: Recall that $\int_{\gamma_j}^{\gamma_{j+1}} \rho_{sc}(x)\,dx = \frac{1}{N} \Rightarrow \gamma_{j+1}-\gamma_j \sim \frac{1}{N\rho_{sc}(\gamma_j)}$. Hence, the ~~fluctuation~~ gaps and correlation functions need to be rescaled by the local density ρ_{sc} to get an universal limit. This holds in more general setups, such as sample covariance matrices.

① Fixed energy universality: $\forall n \in \mathbb{N}$, $F: \mathbb{R}^n \to \mathbb{R}$ is C_c^∞. For any const $\kappa>0$, we have that uniformly in $E \in [-2+\kappa, 2-\kappa]$,

$$\lim_{N\to\infty} \frac{1}{\rho_{sc}(E)^n} \int_{\mathbb{R}^n} d\vec{\alpha}\, F(\vec{\alpha})\, p_N^{(n)}\Big(E + \frac{\vec{\alpha}}{N\rho_{sc}(E)}\Big) = \int_{\mathbb{R}^n} d\vec{\alpha}\, F(\vec{\alpha})\, q^{(n)}_{GOE/GUE}(\vec{\alpha}),$$

where $\vec{\alpha} = (\alpha_1, \cdots, \alpha_n)$, $p_N^{(n)}$ is the n-point correlation function, and $q^{(n)}_{GOE/GUE}(\vec{\alpha}) = \det\big(S(\alpha_i - \alpha_j)\big)_{i,j=1}^n$ is the determinant of sine-kernel we derived before.

② Averaged bulk universality (on scale $N^{-1+\varepsilon}$):

$$\lim_{N\to\infty} \frac{1}{\rho_{sc}(E)^n} \int_{E-b}^{E+b} \frac{dx}{2b} \int_{\mathbb{R}^n} d\vec{\alpha}\, F(\vec{\alpha})\, p_N^{(n)}\Big(x + \frac{\vec{\alpha}}{N\rho_{sc}(E)}\Big) = \int_{\mathbb{R}^n} d\vec{\alpha}\, F(\vec{\alpha})\, q^{(n)}_{GOE/GUE}(\vec{\alpha})$$

where $b = N^{-1+\varepsilon}$ $\forall$ const $\varepsilon>0$.

③ Fixed gap universality: Fixed any small constant $\delta>0$ and $n \in \mathbb{N}$. For any $F: \mathbb{R}^n \to \mathbb{R}$, $F \in C_c^\infty$ and any $k, m \in [\delta N, (1-\delta)N]$, we have that

$$\lim_{N\to\infty} \Big| \mathbb{E}_{H_N} F\big(N\rho_{sc}(\gamma_k)(\lambda_{k+1}-\lambda_k), \cdots, N\rho_{sc}(\gamma_k)(\lambda_{k+n}-\lambda_k)\big) - \mathbb{E}_{GOE/GUE} F\big(N\rho_{sc}(\gamma_m)(\lambda_{m+1}-\lambda_m), \cdots, N\rho_{sc}(\gamma_m)(\lambda_{m+n}-\lambda_m)\big)\Big| = 0.$$

④ Averaged gap universality: For $l = N^\varepsilon$, $\forall$ const $\varepsilon>0$,

$$\lim_{N\to+\infty} \Big| \frac{1}{2l+1} \sum_{j=k-l}^{k+l} \mathbb{E}_{H_N} F\big(N\rho_{sc}(\gamma_k)(\lambda_{j+1}-\lambda_j), \cdots, N\rho_{sc}(\gamma_k)(\lambda_{j+n}-\lambda_j)\big) - \mathbb{E}_{GOE/GUE} F\big(N\rho_{sc}(\gamma_m)(\lambda_{m+1}-\lambda_m), \cdots, N\rho_{sc}(\gamma_m)(\lambda_{m+n}-\lambda_m)\big)\Big| = 0.$$

Rmk: Fixed energy $\underset{\not\Leftarrow}{\Rightarrow}$ Averaged energy, Fixed gap $\underset{\not\Leftarrow}{\Rightarrow}$ Averaged gap,

Fixed energy $\underset{\not\Leftarrow}{\not\Rightarrow}$ Fixed gap, Averaged energy $\Leftrightarrow$ Averaged gap.

We will focus on proving the averaged energy universality.

Theorem 6.1 The averaged energy universality holds on scales $N^{-1+\varepsilon}$ $\forall$ $0<\varepsilon<1$.

This is a version of the famous "Wigner-Dyson-Mehta conjecture".

II. The three-step strategy

Step 1: Local semicircle law. At this step, we get precise estimates on the matrix elements of the resolvent, the rigidity of eigenvalues, and delocalization of eigenvectors.

Step 2: Universality for Gaussian divisible ~~ensembles~~ ensembles.

Gaussian divisible ensembles are random matrices that can be written as $H_0 + \sqrt{t}\, H^G$, where H_0 is Wigner, H^G is GOE / GUE and independent of H_0, and $t>0$ is a parameter.

A convenient way to generate H_t is the "matrix Ornstein-Uhlenbeck (OU) Process":

$$dH_t = \frac{1}{\sqrt{N}} dB_t - \frac{1}{2} H_t\, dt, \quad H_t = H_0,$$

Where B_t is a matrix Brownian motion whose entries are independent BMs up to symmetry $B_t^* = B_t$, and $\frac{1}{\sqrt{N}} B_t \overset{\text{law}}{=} \sqrt{t}$ GOE / GUE. For each entry,

$$dh_{ij}(t) = \frac{1}{\sqrt{N}} db_{ij}(t) - \frac{1}{2} h_{ij}(t)\, dt.$$

It has a unique strong solution:

$$h_{ij}(t) = h_{ij}(0)\, e^{-t/2} + \int_0^t \frac{e^{-(t-t')/2}}{\sqrt{N}}\, db_{ij}(t').$$

Note that $\int_0^t \frac{e^{-(t-t')/2}}{\sqrt{N}}\, db_{ij}(t')$ is centered Gaussian of variance $\frac{1}{N}\int_0^t e^{-(t-t')}\, dt' = \frac{1}{N}(1-e^{-t})$.

Hence, with a slight abuse of notation, we write it as $h^G_{ij} \cdot \sqrt{1-e^{-t}}$.

This gives a solution

$$H_t \overset{\text{law}}{=} e^{-t/2} H_0 + \sqrt{1-e^{-t}}\, H^G.$$

A big advantage of this form is that variances are preserved throughout the process:

$$\mathbb{E}|h_{ij}(t)|^2 = e^{-t}\, \mathbb{E}|h_{ij}(0)|^2 + (1-e^{-t})\, \mathbb{E}|h^G_{ij}|^2 = \mathbb{E}|h_{ij}(0)|^2,$$

for $\mathbb{E}|h_{ij}(0)|^2 = 1+\delta_{ij}$ in the real case, and $\mathbb{E}|h_{ij}(0)|^2 = 1$ in the complex case.

The purpose of Step 2 is to show ~~that~~ the bulk universality of H_t for $t = N^{-1+\varepsilon}$ for any $0<\varepsilon<1$.

Approximation by a Gaussian ~~div~~ divisible ensemble

Step 3: Given a Wigner matrix, H, there exists a Wigner H_0 such that H_t has asymptotically identical local eigenvalue statistics as H. This is usually done through a Green's function comparison argument by using certain moment matching conditions. Alternatively, one can also use a continuity estimate of the matrix OU process.

The "three-step strategy" is now (one of) the most standard in proving the bulk universality of random matrices (for edge universality, the Step 2 sometimes is not necessary).
Here, Step 1 is model-specific and generally is the "hardest" step. The Steps 2 and 3 are more standard, where general methods / proofs / arguments are known and work for "most" models.
In particular, the strongest result for Step 2 has been established for very general initial conditions H_0 (not necessarily a random matrix).

III. Dyson Brownian Motion

The matrix Brownian motion introduces SPDE tools to study the evolution of the eigenvalues of H_t: $\lambda_1(t) \le \lambda_2(t) \le \cdots \le \lambda_N(t)$. A classical theorem below will guarantee that the eigenvalues are simple and continuous functions of t. So the labelling is preserved along the evolution.

In principle, the eigenvalues $\{\lambda_i(t)\}$ and eigenvectors $\{\vec{u}_i(t)\}$ of H_t are correlated strongly, and we expect a couple system of stochastic differential equations for them (which is indeed the case if B_t is not chosen to have the law of GOE/GUE).
But, Dyson observe that the eigenvalues themselves satisfy an autonomous system of SDEs that does not involve eigenvectors, which is called the Dyson Brownian Motion (DBM).

Theorem 6.2 The eigenvalues $\{\lambda_i(t)\}$ of H_t satisfy the following system of SDEs:

$$d\lambda_i = \frac{\sqrt{2}}{\sqrt{\beta N}} dB_i(t) + \Big(-\frac{\lambda_i}{2} + \frac{1}{N}\sum_{j\neq i}\frac{1}{\lambda_i-\lambda_j}\Big)dt, \quad 1\le i\le N, \quad \begin{cases}\beta=1 & \text{for real,}\\ \beta=2 & \text{for complex.}\end{cases}$$

$\{B_i(t): 1\le i\le N\}$ is a collection of independent BMs. The solution to the above equation is called a DBM (where there is not necessarily an underlying matrix model).

Proof: Let $\lambda_\alpha^{(t)}$ be an eigenvalue of $H(t) = (h_{ij}(t))$ with eigenvector $\vec{u}_\alpha(t)$. Almost surely, all eigenvalues are simple. We apply Ito's formula to $\lambda_\alpha(t)$ to derive the DBM.
We only consider the real case with $\beta=1$.
Differentiating: $H\vec{u}_\alpha = \lambda_\alpha \vec{u}_\alpha$, $\vec{u}_\alpha^* \vec{u}_\beta = \delta_{\alpha\beta}$, we obtain that

$$(1)\quad \frac{\partial H}{\partial h_{ij}}\vec{u}_\alpha + H\frac{\partial \vec{u}_\alpha}{\partial h_{ij}} = \frac{\partial \lambda_\alpha}{\partial h_{ij}}\vec{u}_\alpha + \lambda_\alpha\frac{\partial \vec{u}_\alpha}{\partial h_{ij}},$$

$$(2)\quad \frac{\partial \vec{u}_\alpha^*}{\partial h_{ij}}\vec{u}_\beta + \vec{u}_\alpha^* \frac{\partial \vec{u}_\beta}{\partial h_{ij}} = 0, \qquad \vec{u}_\alpha^* \frac{\partial \vec{u}_\alpha}{\partial h_{ij}} = 0 .$$

Taking inner product $\vec{u}_\alpha$ with (1) and using (2), we get

$$\vec{u}_\alpha^* \frac{\partial H}{\partial h_{ij}} \vec{u}_\alpha + \lambda_\alpha \vec{u}_\alpha^* \frac{\partial \vec{u}_\alpha}{\partial h_{ij}} = \frac{\partial \lambda_\alpha}{\partial h_{ij}} + \lambda_\alpha \vec{u}_\alpha^* \frac{\partial \vec{u}_\alpha}{\partial h_{ij}} .$$

$$\Rightarrow \frac{\partial \lambda_\alpha}{\partial h_{ij}} = \vec{u}_\alpha^* \frac{\partial H}{\partial h_{ij}} \vec{u}_\alpha . \qquad (*)$$

Taking inner product $\vec{u}_\beta$ with (1) and using (2), we get (for $\beta \neq \alpha$)

$$\vec{u}_\beta^* \frac{\partial H}{\partial h_{ij}} \vec{u}_\alpha + \lambda_\beta \vec{u}_\beta^* \frac{\partial \vec{u}_\alpha}{\partial h_{ij}} = \lambda_\alpha \vec{u}_\beta^* \frac{\partial \vec{u}_\alpha}{\partial h_{ij}} .$$

This implies that $\displaystyle \frac{\partial \vec{u}_\alpha}{\partial h_{ij}} = \sum_{\beta \neq \alpha} \vec{u}_\beta \left(\vec{u}_\beta^* \frac{\partial \vec{u}_\alpha}{\partial h_{ij}} \right) = \sum_{\beta \neq \alpha} \vec{u}_\beta \frac{1}{\lambda_\alpha - \lambda_\beta} \left(\vec{u}_\beta^* \frac{\partial H}{\partial h_{ij}} \vec{u}_\alpha \right) \qquad (\dagger)$

For $(*)$, it writes:
$$\begin{cases} \dfrac{\partial \lambda_\alpha}{\partial h_{ij}} = (2-\delta_{ij})\, u_\alpha(i)\, u_\alpha(j) . \\[2ex] \dfrac{\partial u_\alpha(k)}{\partial h_{ij}} = \displaystyle\sum_{\beta \neq \alpha} \frac{u_\beta(i) u_\alpha(j) + u_\beta(j) u_\alpha(i)(1-\delta_{ij})}{\lambda_\alpha - \lambda_\beta} u_\beta(k) . \end{cases}$$
For $(\dagger)$, it writes

With these two formulas, we can also compute the second order partial derivatives:

$$\cancel{\frac{\partial^2 \lambda_\alpha}{\partial h_{ij} \partial h_{kl}}} \quad \frac{\partial^2 \lambda_\alpha}{\partial h_{ik} \partial h_{lj}} = (2-\delta_{ik}) \left[\frac{\partial u_\alpha(i)}{\partial h_{lj}} u_\alpha(k) + u_\alpha(i) \frac{\partial u_\alpha(k)}{\partial h_{lj}} \right]$$

$$= (2-\delta_{ik}) \sum_{\beta \neq \alpha} \frac{1}{\lambda_\alpha - \lambda_\beta} \Big[\big(u_\beta(j) u_\alpha(l) + u_\beta(l) u_\alpha(j)(1-\delta_{lj})\big) u_\beta(i) u_\alpha(k)$$
$$+ \big(u_\beta(j) u_\alpha(l) + u_\beta(l) u_\alpha(j)(1-\delta_{lj})\big) u_\beta(k) u_\alpha(i) \Big]$$

$$= (2-\delta_{ik}) \sum_{\beta \neq \alpha} \frac{1}{\lambda_\alpha - \lambda_\beta} \big(u_\beta(j) u_\alpha(l) + u_\beta(l) u_\alpha(j)(1-\delta_{lj})\big)\big(u_\beta(i) u_\alpha(k) + u_\beta(k) u_\alpha(i)\big) .$$

Now, using Ito's formula, we obtain that

$$d\lambda_\alpha = \sum_{i \le k} \frac{\partial \lambda_\alpha}{\partial h_{ik}} dh_{ik} + \frac{1}{2} \sum_{i \le k} \sum_{j \le l} \frac{\partial^2 \lambda_\alpha}{\partial h_{ik} \partial h_{lj}} [dh_{ik}, dh_{lj}]$$

$$= \sum_{i,k} u_\alpha(i) u_\alpha(k) \left[\frac{db_{ik}}{\sqrt{N}} - \frac{1}{2} h_{ik}\, dt \right] + \sum_{i,k} \sum_{\alpha \neq \beta} \frac{1}{2N} \frac{1}{\lambda_\alpha - \lambda_\beta} \left[u_\beta(i) u_\alpha(k) + u_\beta(k) u_\alpha(i) \right]^2 dt$$

$$= \frac{1}{\sqrt{N}} \sum_{i,k} u_\alpha(i) u_\alpha(k)\, db_{ik} - \frac{1}{2} \lambda_\alpha\, dt + \frac{1}{N} \sum_{\alpha \neq \beta} \frac{1}{\lambda_\alpha - \lambda_\beta} dt ,$$

where we used $[dh_{ik}, dh_{lj}] = \frac{1}{N} \delta_{il} \delta_{kj} (1+\delta_{ik})\, dt$, and $\sum_k h_{ik} u_\alpha(k) = \lambda_\alpha u_\alpha(i)$.

Now, we define a new Gaussian process: $\tilde{B}_\alpha := \sum_{i,k} u_\alpha(i) u_\alpha(k)\, db_{ik}$. Clearly, $\mathbb{E} \tilde{B}_\alpha = 0$.
We now calculate its covariance:

$$\mathbb{E}[(d\tilde{B}_\alpha),(d\tilde{B}_\beta)] = \sum_{i,k}\sum_{\ell,j} u_\alpha(i)\, u_\alpha(k)\, u_\beta(\ell)\, u_\beta(j)\,[db_{ik}, db_{\ell j}]$$

$$= \sum_{i,k}\sum_{\ell,j} u_\alpha(i)\, u_\alpha(k)\, u_\beta(\ell)\, u_\beta(j)\, (\delta_{i\ell}\delta_{kj} + \delta_{ij}\delta_{k\ell})\, dt$$

$$= 2 \sum_{i,k} u_\alpha(i)\, u_\beta(i)\, u_\alpha(k)\, u_\beta(k)\, dt \;=\; 2\,\delta_{\alpha\beta}\, dt .$$

Thus, $\tilde{B}_\alpha = \sqrt{2}\, B_\alpha$, where $B_\alpha(t)$ is a standard real BM and B_α's are independent of each other. This gives the DBM with $\beta = 1$. $\square$

A standard SPDE argument shows that there is a strong solution to the DBM:

$$(\ddagger) \quad d\lambda_i = \frac{\sqrt{2}}{\sqrt{\beta N}}\, dB_i(t) + \Big(-\frac{\lambda_i}{2} + \frac{1}{N}\sum_{j\neq i} \frac{1}{\lambda_i - \lambda_j}\Big)\, dt, \quad 1 \le i \le N, \quad \beta \ge 1.$$

Note: if $\lambda_j < \lambda_i$, then the "interaction force" $\frac{1}{\lambda_i - \lambda_j}$ is >0, while if $\lambda_j > \lambda_i$, $\frac{1}{\lambda_i - \lambda_j} < 0$. This gives a repulsion between particles $\{\lambda_i\}$.

Theorem 6.3. Let $\bar{\Delta}_N := \{\vec{\lambda} : \lambda_1 \le \lambda_2 \le \cdots \le \lambda_N\}$. Let $\beta \ge 1$ and suppose that the initial cond. $\vec{\lambda}(0) \in \bar{\Delta}_N$. Then, there exists a unique strong solution to $(\ddagger)$ in the space of continuous functions $(\vec{\lambda}(t))_{t\ge 0} \in C(\mathbb{R}_+, \bar{\Delta}_N)$. Moreover, $\forall\, t > 0$, we have $\vec{\lambda}(t) \in \Delta_N$ and $\vec{\lambda}(t)$ depends continuously on $\vec{\lambda}(0)$. In particular, if $\vec{\lambda}(0) \in \Delta_N$, then $(\vec{\lambda}(t))_{t\ge 0} \in C(\mathbb{R}_+, \Delta_N)$, i.e., the particles are separated ~~for all~~ for all times along the evolution.

Rmk: The DBM can be regarded as a Itô drift-diffusion process. Hence, we can mimic the proof of the existence and uniqueness of the strong solution there. But, one needs to deal with the sigularities $(\lambda_i - \lambda_j)^{-1}$. The "level repulsion mechanism" will play a significant role in the proof.

IV. Strong local ergodicity of DBM

~~DBM~~ The Gaussian measure (GOE / GUE) is the only stationary measure of DBM and the DBM dynamics converges to this equilibrium from any initial condition.

Recall the invariant β-ensemble: $\vec{\lambda} = (\lambda_1, \cdots, \lambda_N)$,

$$\mu_N(d\lambda) = \frac{1}{Z_N} \exp(-\beta N \mathcal{H}_N(\vec{\lambda}))\, d\vec{\lambda},$$

$$\mathcal{H}_N(\vec{\lambda}) = \frac{1}{2}\sum_{i=1}^N V(\lambda_i) - \frac{1}{N}\sum_{i<j} \log|\lambda_i - \lambda_j| .$$

For us, we are interested in the GOE / GUE case with $V(\lambda) = \frac{1}{2}\lambda^2$.

Then, we define the Dirichlet form associated with μ_N:

$$D_\mu(f) := \frac{1}{\beta N}\sum_{i=1}^{N}\int (\partial_i f)^2 d\mu = \frac{1}{\beta N}\|\nabla f\|^2_{L^2(\mu_N)}, \qquad \partial_i := \partial_{\lambda_i}.$$

The symmetric operator associated with the Dirichlet form is called generator and denoted by $\mathcal{L}_\mu \equiv \mathcal{L}$. It is defined through ($\langle , \rangle$: inner product)

$$D_\mu(f) = \langle f, (-\mathcal{L})f\rangle_{L^2(\mu)} = -\int f\mathcal{L}f\, d\mu_N. \quad (-\mathcal{L} \text{ is a positive operator})$$

Note that $\mathcal{L}$ can be chosen as $\mathcal{L} = \frac{1}{\beta N}\Delta - (\nabla \mathcal{H})\cdot\nabla$:

$$-\int f\mathcal{L}f\ \frac{1}{Z_N}\exp(-\beta N \mathcal{H}_N(\vec\lambda))\,d\vec\lambda = -\int f\frac{1}{\beta N}\Delta f\ \frac{1}{Z_N}\exp(-\beta N\mathcal{H}_N)\,d\vec\lambda + \int f(\nabla\mathcal{H})\cdot\nabla f\ \frac{1}{Z_N}\exp(-\beta N\mathcal{H}_N)\,d\vec\lambda$$

$$= D_\mu(f) + \frac{1}{\beta N}\int f(\nabla f)\cdot\nabla\left(\frac{1}{Z_N}\exp(-\beta N\mathcal{H}_N)\right)d\vec\lambda + \int f(\nabla\mathcal{H})\cdot\nabla f\, d\mu_N = D_\mu(f).$$

In components: $\mathcal{L} = \sum_{i=1}^{N}\frac{1}{\beta N}\partial_i^2 + \sum_{i=1}^{N}\left(-\frac{1}{2}V'(\lambda_i) + \frac{1}{N}\sum_{j\neq i}\frac{1}{\lambda_i-\lambda_j}\right)\partial_i$.

For $V(\lambda) = \frac{1}{2}\lambda^2$, we have $\mathcal{L}_G = \frac{1}{\beta N}\sum_{i=1}^{N}\partial_i^2 + \sum_{i=1}^{N}\left(-\frac{1}{2}\lambda_i + \frac{1}{N}\sum_{j\neq i}\frac{1}{\lambda_i-\lambda_j}\right)\partial_i$.

~~Con~~ With the DBM and Ito's formula, we notice that for any twice differentiable g,

$$\partial_t \mathbb{E}\, g(\vec\lambda(t)) = \mathbb{E}\,\mathcal{L}_G\, g(\vec\lambda(t)).$$

Write the distribution of $\vec\lambda(t)$ at time t as $f_t(\vec\lambda)\mu_N(d\vec\lambda)$. We have

$$\partial_t\int g(\vec\lambda)f_t(\vec\lambda)\mu_N(d\vec\lambda) = \int (\mathcal{L}_G g(\vec\lambda))f_t(\vec\lambda)\mu_N(d\vec\lambda) = \int g(\vec\lambda)[\mathcal{L}_G f_t(\vec\lambda)]\mu_N(d\vec\lambda).$$

In other words, the density $f_t(\vec\lambda)$ satisfies $\partial_t f_t(\vec\lambda) = \mathcal{L}_G f_t(\vec\lambda)$. (☆)
Note that $f(\vec\lambda)\equiv 1$ is a solution to this equation, i.e., $\mu_N(d\vec\lambda)$ is a stationary measure of the DBM. Our goal is to show that for any initial condition f_0, $f_t \to f_\infty \equiv 1$. A much harder and more important question is: how fast the dynamics reach equilibrium?

$\boxed{\text{Dyson's conjecture}}$ The global equilibrium of DBM is reached in time of order 1 and the local equilibrium (in the bulk) is reached in time of order $\frac{1}{N}$.

From $H_t = e^{-t/2}H_0 + \sqrt{1-e^{t}}\,H_G$, we see that the global equilibrium is indeed reached within a time of order 1. The key is that the local equilibrium is achieved much faster if an a priori estimate on the initial locations of the eigenvalues holds, which verifies Dyson's conjecture.

This a priori estimate is the "rigidity of eigenvalues".

Theorem 6.4 (Relaxation of DBM) Suppose for some exponent $\xi\in(0,\frac{1}{2})$, the rigidity of the eigenvalues ~~in the~~ of H_t holds on scale $N^{-1+\xi}$, i.e.,

$$\max_j |\lambda_j(t)-\gamma_j(t)| \prec N^{-1+\xi} \quad \cancel{\max_j |\lambda_j^{(t)}-\gamma_j^{(t)}| \prec N^{-1+\xi}} \quad \text{uniformly in } t\in[N^{-1+2\xi},N].$$

Let $E\in[-2+\kappa,2-\kappa]$ and $b\equiv b_N>0$ such that $[E-b,E+b]\in(-2,2)$. Then, $\forall\, n\geq 1$ and $F\in C_c^\infty:\mathbb{R}^n\to\mathbb{R}$, we have

$$\left|\int_{E-b}^{E+b}\frac{dx}{2b}\int_{\mathbb{R}^n} d\vec{\alpha}\; F(\vec{\alpha})\,(p_t^{(n)}-p_G^{(n)})(x+\frac{\vec{\alpha}}{N})\right|$$

$$\leq N^{\varepsilon}\left[\frac{N^{-1+\xi}}{b}+\sqrt{\frac{1}{bNt}}\right]\|F\|_{C^1}, \quad \forall \text{ const } \varepsilon>0 \text{ and } t\in[N^{-1+2\xi},N].$$

Here, $p_t^{(n)}$ is the n-point correlation function of H_t, $p_G^{(n)}$ —— " —— of H_G. | $\|F\|_{C^1}=\|F\|_\infty+\sup_{x\in\mathbb{R}^n}\|\nabla F(x)\|_\infty$

Rmk: The upper bound N on t is not essential. It can be replaced by N^{ε}, $\forall\varepsilon>0$, where $H_t=e^{-t/2}H_0+\sqrt{1-e^{-t}}H_G$ is super-close to H_G with an exponentially error $e^{-O(N^{\varepsilon})}$. For us, the most interesting case will be $t\ll 1$.

The above thm says that if we have rigidity on scale $N^{-1+\xi}$, then the DBM has averaged bulk universality for any $t\gg N^{-1+2\xi}$ on scales $b\gg\max\{N^{-1+\xi},(Nt)^{-1}\}$. If H_0 is a Wigner matrix, we can choose ξ as small as possible. ~~This gives Theorem 6.1.~~

~~Proof of Theorem 6.4: For any constant $\varepsilon>0$, we choose $\xi=\varepsilon$ and $t=N^{-1+2\varepsilon}$, $b=N^{-1+3\varepsilon}$.~~

~~As long as ε is sufficiently small (e.g., $\varepsilon<$~~

In the sense of large energy windows of size $b=N^{-\varepsilon}$, the above theorem essentially establishes the Dyson's conjecture the time to local equilibrium is $N^{-1+\varepsilon}$ $\forall\varepsilon>0$.

V: Entropy

To analyze the convergence of f_t to f_∞, a classical tool is entropy and the log-Sobolev inequality (LSI).

Def: Given two probability measures μ and ν, we define the relative entropy of ν w.r.t. μ as

$$S(\nu|\mu)=\int\log\frac{d\nu}{d\mu}\,d\nu=\int\frac{d\nu}{d\mu}\log\frac{d\nu}{d\mu}\,d\mu$$

$\longrightarrow$ (Radon-Nikodym derivative)

if ν is absolutely continuous w.r.t. μ. (Otherwise, we set $S(\nu|\mu)=\infty$.)

If $\nu = f\mu$, then $\frac{d\nu}{d\mu} = f$. Then we write $S_\mu(f) := S(\nu|\mu) = \int f \log f \, d\mu$, the entropy of f.

Since the function $x \mapsto x\log x$ is convex on $\mathbb{R}_+$, by Jensen's inequality,

$$\int f \log f \, d\mu \geq \left(\int f \, d\mu\right) \log\left(\int f \, d\mu\right) = 0,$$ i.e., the relative entropy is always positive.

We now present some important inequalities related to the entropy.

Prop (Gibbs inequality) Let X be a random variable defined on the probability space of μ and ν. For any $\alpha > 0$, we have:

$$\mathbb{E}^\nu[X] \leq \alpha^{-1} S(\nu|\mu) + \alpha^{-1} \log \mathbb{E}^\mu e^{\alpha X}.$$

Proof: Without loss of generality, we can take $\alpha = 1$ by setting $X \to \alpha X$. By Jensen's ineq,

$$\mathbb{E}^\nu X - S(\nu|\mu) = \int X \, d\nu - \int \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu} \, d\mu$$

$$= \int \log\left[e^X \frac{d\mu}{d\nu}\right] d\nu \leq \log\left[\int e^X \frac{d\mu}{d\nu} \, d\nu\right] = \log\left[\mathbb{E}^\mu e^X\right]. \quad \square$$

Rmk: In fact, we have $S(\nu|\mu) = \sup_X \left[\mathbb{E}^\nu X - \log \mathbb{E}^\mu e^X\right]$.

Recall that the L^p distance between $f\mu$ and μ is defined by: $\left[\int |f-1|^p d\mu\right]^{\frac{1}{p}}$. When $p = 1$, it is also called total variation (TV) norm. Entropy is a weaker measure of distance between two probability measures than the L^p distance $\forall p > 1$:

$$|f \log f| = |[1+(f-1)] \log[1+(f-1)]| \leq C_p(|f-1| + |f-1|^p)$$

$$\Rightarrow \int f \log f \, d\mu \leq C_p \left(\int |f-1|^p d\mu\right)^{1/p} + C_p \int |f-1|^p d\mu.$$

(This can be chosen as 2) (This can be chosen as $\frac{2}{p-1}$)

But, it is stronger than the TV norm. Also, notice the following simple relation:

$$\frac{d}{dp}\Big|_{p=1} \left[\int f^p d\mu\right]^{\frac{1}{p}} = \int f \log f \, d\mu.$$

Prop (Pinsker inequality) Suppose $\int f \, d\mu = 1$ and $f \geq 0$. Then, we have that

$$\left[\int |f-1| d\mu\right]^2 \leq 2\int f \log f \, d\mu.$$

Proof: Recall the equivalent form of TV norm: $\int |f-1| d\mu = \sup_{|g| \leq 1} \left[\int fg \, d\mu - \int g \, d\mu\right]$

Using the Gibbs inequality (with $X = g$ and $\nu = f\mu$), we get $\forall t > 0$,

$$(*) \quad \int fg \, d\mu - \int g \, d\mu \leq t^{-1} \log \int e^{tg} d\mu + t^{-1} \int f \log f \, d\mu - \int g \, d\mu.$$

Denote $h(t) := \log \int e^{tg} d\mu$, $t \geq 0$. A direct calculation gives that

$$h'(t) = \frac{\int g e^{tg}\, d\mu}{\int e^{tg}\, d\mu} = \int g\, dw_t\,, \quad \text{where} \quad dw_t := \frac{e^{tg}}{\int e^{tg} d\mu}\, d\mu\,,$$

$$h''(t) = \int g^2\, dw_t - \left(\int g\, dw_t\right)^2 = \mathrm{Var}_{w_t}(g) \le 1 \quad \text{since } |g| \le 1.$$

~~By Cauchy-Schwarz inequality, we get $\mathrm{Var}_{w_t}(g)$~~ By Taylor expansion,

$$h(t) \le h(0) + t h'(0) + \tfrac{1}{2}t^2 \Rightarrow \quad \tfrac{1}{t}\log \int e^{tg} d\mu \le \int g\, d\mu + \tfrac{1}{2}t.$$

Then, using (*), we get

$$\int f g\, d\mu - \int g\, d\mu \le \tfrac{1}{2}t + \tfrac{1}{t}\int f \log f\, d\mu \quad \forall\, t>0.$$

Optimize over t, we get: $\longrightarrow \le \sqrt{2\int f \log f\, d\mu}$.

Finally, taking sup over g, we get: $\int |f-1|\, d\mu \le \sqrt{2\int f\log f\, d\mu}$. $\square$

Entropy is particularly useful in studying product measures on a "high-dimensional" space. Then, the Pinsker inequality gives a good (and actually sharp in many cases) bound on the TV distance between two measures on high-d spaces.

Consider product probability measures $\mu = \mu_1 \otimes \mu_2$, $\nu = \nu_1 \otimes \nu_2$, $\nu_j = f_j \mu_j$, $j=1,2$,

$$\begin{aligned} S(\nu_1\otimes\nu_2 \mid \mu_1\otimes\mu_2) &= \int\int f_1(x) f_2(y) \log[f_1(x) f_2(y)]\, \mu_1(dx)\, \mu_2(dy) \\ &= \int\int \left[f_1(x) f_2(y) \log f_1(x)\, \mu_1(dx)\, \mu_2(dy) + f_1(x) f_2(dy) \log f_2(y)\, \mu_1(dx)\, \mu_2(dy)\right] \\ &= \int f_1(x) \log f_1(x)\, \mu_1(dx) + \int f_2(y) \log f_2(y)\, \mu_2(dy) \\ &= S(\nu_1|\mu_1) + S(\nu_2|\mu_2). \end{aligned}$$

This makes entropy a good tool to measure distances between measures in high dimensions, since the relative entropy grows linearly in N:

$$S_\mu(\mu_1\otimes\cdots\otimes\mu_N \mid \nu_1\otimes\cdots\otimes\nu_N) = \sum_{i=1}^N S(\mu_i|\nu_i).$$

On the other hand, $\forall\, p>1$, the L^p distance grows exponentially in N, which is often useless. The entropy is ~~already~~ stronger than L^1 distance, while its growth in N is much more ~~manageable~~ manageable. In addition, entropy is easier to calculate than L^p-norm for $p\neq 2$.

VI: LSI (Logarithmic Sobolev inequality)

Def: A probability measure μ on $\mathbb{R}^N$ satisfies the <u>LSI</u> if there exists a constant γ such that:

$$S(f) = \int f\log f\, d\mu \le \gamma \int \underbrace{|\nabla\sqrt{f}|^2}_{\frac{1}{4}\frac{|\nabla f|^2}{f}}\, d\mu = \gamma D(\sqrt{f})$$

for any smooth density $f\ge 0$ with $\int f\, d\mu = 1$.

The smallest such γ is called the LSI constant of μ.

For our purpose, we focus on Gibbs measure defined by a Hamiltonian $\mathcal{H}$:
$$d\mu(x) = \frac{e^{-\mathcal{H}(x)}}{Z} dx . \qquad (*)$$

The Dirichlet norm is defined by $D_\mu(f) = \int |\nabla f|^2 d\mu$. The generator associated with μ is $\mathcal{L} = \Delta - (\nabla \mathcal{H}) \cdot \nabla$. Recall that $\mathcal{L}$ satisfies
$$\int f (\mathcal{L} g) = \int (\mathcal{L} f) g = - \int \nabla f \cdot \nabla g \, d\mu .$$

☆ <u>Theorem 6.5</u> (Bakry - Emery) Suppose $\mathcal{H}$ in $(*)$ satisfies a uniform convexity condition:
$$\nabla^2 \mathcal{H}(x) \geq K$$
for some constant $K > 0$ and any x. (∇^2 denotes the Hessian, and "$\geq$" is used in the following sense: ~~for two Hermitian matrices~~ the smallest eigenvalue of $\nabla^2 \mathcal{H}$ is at least K.)
Then the LSI holds for μ with an LSI const $\gamma \leq 2/K$, i.e.,
$$S(f) \leq \frac{2}{K} D(\sqrt{f}) \quad \text{for any density } f \text{ with } \int f d\mu = 1 .$$

Furthermore, the dynamics $\partial_t f_t = \mathcal{L} f_t$, $t > 0$, relaxes to equilibrium on the time scale $\frac{1}{K}$ in the following senses:
$$S(f_t) \leq e^{-2tK} S(f_0), \qquad D(\sqrt{f_t}) \leq \frac{1}{2t} e^{-tK} S(f_0) .$$

Proof: Let f_t be a solution to $\partial_t f_t = \mathcal{L} f_t$ with a given smooth initial condition f_0.
We can check that
$$\partial_t S(f_t) = \partial_t \int f_t \log f_t \, d\mu = \int (\mathcal{L} f_t) \log f_t \, d\mu + \overbrace{\int f_t \frac{\mathcal{L} f_t}{f_t} d\mu}^{=0}$$
$$= - \int (\nabla f_t) \cdot \nabla (\log f_t) d\mu = - \int \frac{|\nabla f_t|^2}{f_t} d\mu = -4 D(\sqrt{f_t}) . \quad (+)$$

Let $h_t := \sqrt{f_t}$, then $\partial_t h_t = \frac{1}{2h_t} \partial_t (h_t^2) = \frac{1}{2h_t} \mathcal{L}(h_t^2) = \mathcal{L} h_t + \frac{1}{h_t} |\nabla h_t|^2$. Then, we can compute the evolution of the Dirichlet form:
$$\partial_t D(\sqrt{f_t}) = \partial_t \int |\nabla h_t|^2 d\mu = 2 \int (\nabla h_t) \cdot (\nabla \partial_t h_t) d\mu$$
$$= 2 \int (\nabla h_t) \cdot (\nabla \mathcal{L} h_t) d\mu + 2 \int (\nabla h_t) \cdot \nabla \frac{|\nabla h_t|^2}{h_t} d\mu$$
$$= 2 \int (\nabla h_t) \cdot (\nabla \mathcal{L} - \mathcal{L} \nabla) h_t \, d\mu + 2 \int (\nabla h_t) \cdot \mathcal{L}(\nabla h_t) d\mu + 2 \int \sum_{i,j} (\partial_i h) \left[\frac{2(\partial_j h)(\partial_i \partial_j h)}{h} - \frac{(\partial_j h)^2 \partial_i h}{h^2} \right]$$
$$= -2 \int (\nabla h_t) \cdot (\nabla^2 \mathcal{H}) \cdot \nabla h_t \, d\mu - 2 \int \sum_{i,j} (\partial_i \partial_j h)^2 d\mu + 2 \int \sum_{i,j} \left[\frac{2(\partial_i h)(\partial_j h)(\partial_{ij} h)}{h} - \frac{(\partial_i h)^2 (\partial_j h)^2}{h^2} \right] d\mu$$
$$= -2 \int (\nabla h_t) \cdot (\nabla^2 \mathcal{H}) \nabla h_t \, d\mu - 2 \int \sum_{i,j} \left(\partial_{ij} h - \frac{(\partial_i h)(\partial_j h)}{h} \right)^2 d\mu$$

$$\leq -2\int (\nabla h_t)\cdot(\nabla^2 \Omega)\cdot(\nabla h_t)\,d\mu \leq -2K\int |\nabla h_t|^2 d\mu = -2K\,D(\sqrt{f_t}).$$

~~Integrating With Gronw~~ In sum, $\partial_t D(\sqrt{f_t}) \leq -2K\,D(\sqrt{f_t}) \Rightarrow D(\sqrt{f_t}) \leq e^{-2tK} D(\sqrt{f_0})$.
This shows that the equilibrium is achieved at $t=\infty$ with $f_\infty = 1$, where both the entropy and Dirichlet form are zero. Integrating the equality $(+)$ from $t=0$ to $t=\infty$, we get

$$-S(f_0) = -4\int_0^\infty D(\sqrt{f_t})\,dt \geq -4D(\sqrt{f_0})\int_0^\infty e^{-2tK}\,dt = -\frac{2}{K}D(\sqrt{f_0}).$$

This proves the LSI for any $f=f_0$. In particular, it also holds for $f=f_t$. Then,

$$\partial_t S(f_t) = -4D(\sqrt{f_t}) \leq -2KS(f_t) \Rightarrow S(f_t) \leq e^{-2Kt} S(f_0).$$

Finally, $S(f_t) - S(f_{t/2}) = -4\int_{t/2}^{t} D(\sqrt{f_{t'}})\,dt' \overset{\leq -2t}{} D(\sqrt{f_t})$

$$\Rightarrow 2t\,D(\sqrt{f_t}) \leq S(f_{t/2}) \Rightarrow D(\sqrt{f_t}) \leq \frac{1}{2t}S(f_{t/2}) \leq \frac{1}{2t}e^{-tK}S(f_0). \qquad \square$$

<u>Example</u> (LSI for Gaussian measure) Consider a Gaussian measure on $\mathbb{R}^N$,

$$d\mu(x) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx.$$

By Bakry-Emery, LSI holds for μ with $\gamma \leq 2\sigma^2$. This is actually sharp: $\gamma = 2\sigma^2$.

<u>Prop</u> (LSI implies spectral gap / Poincaré inequality) Let μ satisfy LSI with LSI const γ. Then, $\forall f \in L^2(\mu)$ with $\int f\,d\mu = 0$, we have

$$\int f^2 d\mu \leq \frac{\gamma}{2}\int |\nabla f|^2 d\mu = \frac{\gamma}{2}D(f),$$

i.e., μ has a spectral gap of size at least $\frac{\gamma}{2}$.

<u>Pf</u>: By def, $\forall$ density u, $\int u\log u\,d\mu \leq \gamma D(\sqrt{u})$. Define $u = 1+\varepsilon f$ for small $\varepsilon>0$. Then,

$$\int (1+\varepsilon f)\log(1+\varepsilon f)\,d\mu \leq \frac{\gamma}{4}\int \frac{\varepsilon^2|\nabla f|^2}{1+\varepsilon f}\,d\mu$$

$$\Rightarrow \int \frac{1}{\varepsilon^2}(1+\varepsilon f)\log(1+\varepsilon f)\,d\mu \leq \frac{\gamma}{4}\int \frac{|\nabla f|^2}{1+\varepsilon f}\,d\mu.$$ Taking $\varepsilon \to 0$, we get

$$\frac{1}{2}\int f^2 d\mu \leq \frac{\gamma}{4}\int |\nabla f|^2 d\mu. \qquad \square$$

<u>Prop</u> (LSI implies large deviation : Herbst bound) Suppose μ satisfies LSI with const γ. Let F be a function with $\mathbb{E}_\mu F = 0$. Then,

$$\mathbb{E}_\mu e^F \leq \exp\left(\frac{\gamma}{4}\|\nabla F\|_\infty^2\right), \quad \|\nabla F\|_\infty := \sup_x |\nabla F(x)|.$$

In particular, we have $\mathbb{P}_\mu(|F| \geq \alpha) \leq \exp\left(-\frac{\alpha^2}{\gamma\|\nabla F\|_\infty^2}\right)$.