Statistics 291: Lecture 6 (February 8, 2024)

Kac-Rice II: spherical spin glasses without external field

Instructor: Mark Sellke

Scribe: Rushil Mallarapu

1 Spherical Derivatives of the Hamiltonian and Kac-Rice

Let $x \in S_N$. Then $P_x^\perp = I_n - \frac{xx^\top}{N}$ is the rank $N-1$ projection matrix from $\mathbf{R}^N$ onto the tangent plane of S_N at x. Formally, this should be a map from $T_x\mathbf{R}^N \cong \mathbf{R}^N$ to T_xS_N, but we'll be somewhat loose with the identification between the latter and $\mathbf{R}^{N-1}$. Using this, we can define the spherical gradient, (naive) tangential Hessian, radial derivative, and (correct) spherical Hessian.

$$\begin{aligned}
\nabla_{\text{sph}} H_{N,p}(x) &= P_x^\perp \nabla H_{N,p}(x) \\
\nabla^2_{\text{tan}} H_{N,p}(x) &= P_x^\perp \nabla^2 H_{N,p}(x) P_x^\perp \\
\nabla_{\text{rad}} H_{N,p}(x) &= \frac{1}{N}\langle \nabla H_{N,p}(x), x\rangle = \frac{p}{N} H_{N,p}(x) \\
\nabla^2_{\text{sph}} H_{N,p}(x) &= \nabla^2_{\text{tan}} H_{N,p}(x) - \nabla_{\text{rad}} H_{N,p}(x) P_x^\perp.
\end{aligned}$$

Note that the radial derivative being a rescaling of the Hamiltonian is special to working with a p-spin model, by homogeneity.

Recall that last class we derived the *Kac-Rice formula* for the expected number of critical points of the p-spin Hamiltonian $H_{N,p}(x)$:

$$\mathbf{E}\left[\text{Crt}_{S_N}(H_{N,p})\right] = \int_{S_N} \varphi_{\nabla_{\text{sph}} H_{N,p}(x)}(\vec{0})\, \mathbf{E}\left[\left|\det \nabla^2_{\text{sph}} H_{N,p}(x)\right| \,\middle|\, \nabla_{\text{sph}} H_{N,p}(x) = \vec{0}\right] d\mu^{\text{vol}}_{N-1}(S_N) \tag{1}$$

Because we're looking for critical points on the sphere, we needed to define the gradient and Hessian from the perspective of the sphere to get the right formula. Also note that we're taking this integral with respect to the unnormalized $N-1$-dimensional volume measure on S_N. Luckily, all the terms here are relatively easy to compute, as the spherical derivatives are all Gaussian-flavored:

Proposition 1.1. *For all $x \in S_N$, $(H_{N,p}(x), \nabla_{\text{sph}} H_{N,p}(x), \nabla^2_{\text{tan}} H_{N,p}(x))$ are independent, with*

(a) $H_{N,p}(x) \sim \mathcal{N}(0, N)$,

(b) $\nabla_{\text{sph}} H_{N,p}(x) \sim \sqrt{p}\,\mathcal{N}(0, P_x^\perp \cong I_{N-1})$,

(c) $\nabla^2_{\text{tan}} H_{N,p}(x) \sim \sqrt{p(p-1)\left(\frac{N-1}{N}\right)}\,\text{GOE}(N-1)$.

The last distribution is the *Gaussian orthogonal ensemble*, where $A \sim \text{GOE}(M)$ if A is a symmetric random matrix with $A_{ii} \sim \mathcal{N}(0, 2/M)$, and $A_{ij} = A_{ji} \sim \mathcal{N}(0, 1/M)$, with entries otherwise independent.

Proof. Recall from lecture 2 that the Hamiltonian is rotationally invariant in distribution. Thus, we can fix WLOG $x = (\sqrt{N}, 0, \dots, 0)$ to be the "north pole." Now we can compute everything above directly:

(a) We have $H_{N,p}(x) = N^{-(p-1)/2} \cdot g_{1\dots1} N^{p/2} = \sqrt{N} g_{1\dots1} \sim \mathcal{N}(0, N)$.

(b) Fix any $j \neq 1$, and let's look at $[\nabla_{\text{sph}} H_{N,p}(x)]_j$. Because we're projecting onto the tangent space at the north pole, this is basically computing $\langle \nabla H_{N,p}(x), e_j \rangle$, for e_j the jth standard basis vector. By a similar computation, this is

$$\begin{aligned}\langle \nabla H_{N,p}(x), e_j \rangle &= N^{-(p-1)/2} \langle G_N^{(p)}, x^{\otimes p-1} \otimes y + x^{\otimes p-2} \otimes y \otimes x + \cdots + y \otimes x^{\otimes p-1} \rangle \\ &= g_{1\dots j} + g_{1\dots j1} + \cdots + g_{j1\dots1} \sim \sqrt{p}\mathcal{N}(0,1),\end{aligned}$$

as we're adding p i.i.d. standard Gaussians. Thus, we get $\sqrt{p}$ times a standard Gaussian $N-1$-dimensional vector. Let's write that last sum as $\sum_{\text{sym}} g_{1\dots1j}$, where this indicates summing over permutations without eating multiplicity factors.

(c) Fix $j, k \neq 1$. Then repeating the same process as above gives

$$\begin{aligned}[\nabla^2_{\text{tan}} H_{N,p}(x)]_{jk} &= N^{-1/2} \sum_{\text{sym}} g_{1\dots1jk} \\ &\sim \begin{cases} \mathcal{N}(0, p(p-1)/N) & j \neq k \\ \mathcal{N}(0, 2p(p-1)/N) & j = k \end{cases}\end{aligned}$$

When $j \neq k$, there are $p(p-1)$ distinct i.i.d. factors; when $j = k$, there are $p(p-1)/2$ distinct factors, but each is summed twice, so overall we double the variance. Everything here is independent and symmetric in j, k, and the $N-1$ above then comes from the GOE scaling; its not that important.

□

Let's plug this into (1): the integrand is symmetric across x, so fixing x to be the north pole, we get

$$\mathbf{E}\left[\text{Crt}_{S_N}(H_{N,p})\right] = \text{Vol}_{N-1}(S_N)\, \varphi_{\nabla_{\text{sph}} H_{N,p}(x)}(\vec{0})\, \mathbf{E}\left[\left|\det \sqrt{p(p-1)\frac{N-1}{N}}\,\text{GOE}(N-1) - pH_{N,p}(x)/N I_{N-1}\right|\right].$$

Note that we don't need to condition in the expectation because of independence! To get at the inner expectation, let's integrate over $E = H_{N,p}(x)/N$; this is the (dimensionless) energy, i.e. its $O(1)$ as $N \to \infty$. This gives, by fixing the radial derivative,

$$\begin{aligned}&\mathbf{E}\left[\text{Crt}_{S_N}(H_{N,p})\right] \\ &= \text{Vol}_{N-1}(S_N) \varphi_{\nabla_{\text{sph}} H_{N,p}(x)}(\vec{0}) \int_{\mathbf{R}} \varphi_{H_{N,p}(x)/N}(E)\, \mathbf{E}\left[\left|\det \sqrt{p(p-1)\frac{N-1}{N}}\,\text{GOE}(N-1) - pEI_{N-1}\right|\right] dE.\end{aligned}$$

2 Determinants of GOE Matrices

Using the Kac-Rice formula (1) now reduces to understanding (up to a simple rescaling), expressions of the form $\mathbf{E}[|\det(\text{GOE}(N) - tI_N)|]$. If we're experienced with random matrix theory, we can guess that there's some function $\psi(t)$ for which

$$\mathbf{E}[|\det(\text{GOE}(N) - tI_N)|] \stackrel{?}{=} \exp\left(N\psi(t) + o(N)\right).$$

The two ingredients that go into this guess are, first, taking logs, so that

$$\log|\det(\text{GOE}(N) - tI_N)| = \sum_{i=1}^{N} \log|\lambda_i - t|,$$

with $\lambda_1 \geq \cdots \geq \lambda_N$ being the (random) eigenvalues of GOE(N) matrices. Next, we might know that the empirical eigenvalue distribution should be a semicircle distribution. Specifically, the empirical eigenvalue distribution is

$$\nu_N \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i},$$

and the semicircle distribution ν_{SC}, with $d\nu_{SC}(x) = \frac{1}{2\pi}\sqrt{4-x^2}\,dx$, for $x \in [-2,2]$. We can guess that with high probability, ν_N is close to ν_{SC}. This means that, for any fixed constant $\epsilon > 0$, then as $N \to \infty$, the following two statements hold w.h.p.:

(a) $d_{BL}(\nu_N, \nu_{SC}) \leq \epsilon$, where d_{BL} is the Bounded Lipschitz metric:

$$d_{BL}(\nu_N, \nu_{SC}) = \sup_{1\text{-Lip } f:\, \mathbf{R}\to\mathbf{R}} \left| \int f\, d\nu_N - \int f\, d\nu_{SC} \right| \leq \epsilon.$$

This metric is "equivalent to Wasserstein distance" (i.e. it induces the same topology on probability measures supported in any fixed compact set) and captures cases where maybe one distribution has lots of point densities and the other is continuous and concentrated around those points.

(b) $\lambda_1, \ldots, \lambda_N \in [-2-\epsilon, 2+\epsilon]$.

Remark. In general, we can use random matrix theory to evaluate these Kac-Rice integrals; luckily our situation is particularly nice.

Thus, we should guess that $\psi(t) = \int \log|u-t|\, d\nu_{SC}$. If log was 1-Lipschitz, then we would have

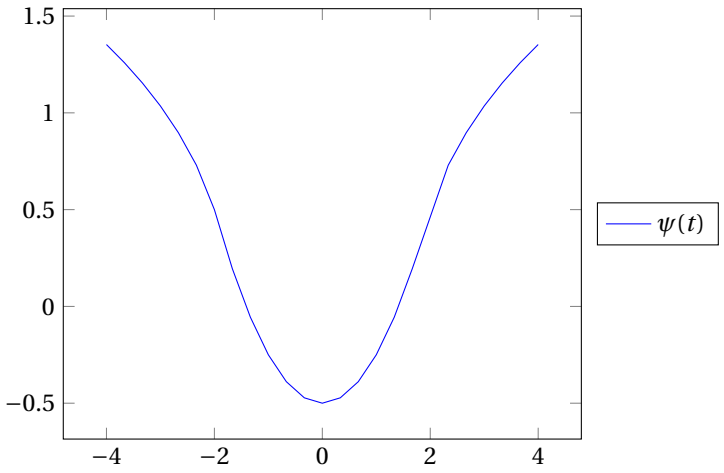
$$\frac{1}{N} \sum_{i=1}^{N} \log|\lambda_i - t| = \int \log|u-t|\, d\nu_N$$

is within ϵ of $\psi(t)$ w.h.p. As it turns out, this guess is correct! We'll start proving the upper bound later; the lower bound requires more heavy random matrix theory and we won't cover it now.

Also, ψ has an explicit formula:

$$\psi(t) = \frac{t^2}{4} - \frac{1}{2} + 1_{|t|\geq 2}\left(-\frac{|t|\sqrt{t^2-4}}{4} + \log\left(\frac{\sqrt{t^2-4}+|t|}{2} \right) \right).$$

The correction term in parentheses, which we'll denote $\theta(t)$, only kicks in for $t \notin (-2,2)$. Graphing $\psi(t)$



we see that outside from $[-2,2]$, it diverges logarithmically. As an aside, if you differentiate $\psi'(t)$ under the integral, you get the "Stieltjes transform" of ν_{SC}.

Combining everything, we get the following exponential growth rate of (1), counting the critical points over any interval of energies $a \le H_{N,p}/N \le b$.

Theorem 2.1 (Annealed exponential growth rate for the number of critical points)**.** *For $-\infty \le a < b \le \infty$,*

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbf{E}\left[\mathrm{Crt}_{S_N}(H_{N,p}) \text{ with } H_{N,p}(x)/N \in [a,b]\right] = \max_{E\in(a,b)} \left\{ \frac{\log(p-1)}{2} - E^2\left(\frac{p-2}{4p-4}\right) + 1_{|E|\ge 2\sqrt{\frac{p-1}{p}}}\, \theta\left(E\sqrt{\frac{p}{p-1}}\right)\right\}.$$

*(**In class there was a small mistake, there should not have been a factor of** $1/2$ **on the last term.**)*

Proof. Recall that

$$\begin{aligned}&\mathbf{E}\left[\mathrm{Crt}_{S_N}(H_{N,p}) \text{ with } H_{N,p}(x)/N \in [a,b]\right] \\ &= \mathrm{Vol}_{N-1}(S_N)\varphi_{\nabla_{\mathrm{sph}} H_{N,p}(x)}(\vec{0}) \int_a^b \varphi_{H_{N,p}(x)/N}(E)\, \mathbf{E}\left[\left|\det\sqrt{p(p-1)\frac{N-1}{N}}\,\mathrm{GOE}(N-1) - pEI_{N-1}\right|\right] dE.\end{aligned}$$

For a fixed $E \in [a,b]$, there are 4 contributions to the Kac-Rice formula, up to $e^{o(N)}$ factors:

(a) $\mathrm{Vol}_{N-1}(S_N) \approx (2\pi e)^{N/2}$.

(b) $\phi_{\nabla_{\mathrm{sph}} H_{N,p}(x)}(\vec{0}) = (2\pi p)^{-N/2}$.

(c) $\phi_{H_{N,p}(x)/N}(E) \approx e^{-NE^2/2}$.

(d) the determinant factor: $(p(p-1))^{N/2} \exp\left(N\psi\left(E\sqrt{\frac{p-1}{p}}\right)\right)$.

After a lot of miraculous cancellations and applying the Laplace approximation, the result follows. □

Remark. This gives an upper bound for the maximum value of $\max_{x\in S_N} H_{N,p}(x)$, which is asymptotically sharp for pure p-spin models. Let $\Phi(E)$ be the function of E in the curly braces in the statement of Theorem 2.1. Say we know that $\Phi(E_0) = 0$. Then, if $a = E_0 + \epsilon$ for small $\epsilon > 0$ and $b = \infty$, we know $\max_{E\ge a} \Phi(E) < 0$. Then, by Markov's inequality, with high probability there are no critical points x with $H_{N,p}(x) \ge N(E_0+\epsilon)$. Thus, $\max_{x\in S_N} H_{N,p}(x)/N \le E_0 + \epsilon$. We'll return to this later.

Remark. It turns out we don't know much about the concentration of the number of critical points!

3 The Hoffman-Wielandt Inequality

It remains to prove the upper bound

$$\mathbf{E}\left[|\det(\mathrm{GOE}(N) - tI_N)|\right] \le \exp\left(N\psi(t) + o(N)\right).$$

We'll use concentration of measure of the eigenvalues. Today we'll prove the Hoffman-Wielandt inequality, which roughly says that the eigenvalues of a matrix depend in a Lipschitz way on the matrix entries. Next time we'll see how this implies the upper bound we want!

Lemma 3.1 (Hoffman-Wielandt Inequality)**.** *Let $A_N, \tilde{A}_N$ be $N \times N$ symmetric matrices with eigenvalues $\lambda_1 \ge \cdots \ge \lambda_N$, $\tilde{\lambda}_1 \ge \ldots \tilde{\lambda}_N$ respectively. Then*

$$\sum_{i=1}^N |\lambda_i - \tilde{\lambda}_i|^2 \le \left\|A_N - \tilde{A}_N\right\|_F^2 = \sum_{i,j=1}^N (A_N - \tilde{A}_N)_{ij}^2.$$

This is the L^2 analog of the "Weyl inequality," which is the easier fact that

$$\max_{i=1,\dots,n} |\lambda_i - \tilde{\lambda}_i| \le \|A_N - \tilde{A}_N\|_{\text{op}}.$$

This can be proven using the Courant-Fischer min-max characterization of eigenvalues.

Proof. The idea is to fix λ_i, $\tilde{\lambda}_i$, and minimize the RHS $\|A_N - \tilde{A}_N\|_F^2$. The lemma is effectively the claim that this is minimized when these matrices have the same eigenvectors.

Without loss of generality, let $A_N = \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{bmatrix}$ and let $\tilde{A}_N = O\tilde{\Lambda}O^\top$, for O an orthogonal $N \times N$ matrix. Then, we can compute

$$\begin{aligned} \|A_N - \tilde{A}_N\|_F^2 &= \operatorname{Tr}\left((A_N - \tilde{A}_N)(A_N - \tilde{A}_N)^\top\right) \\ &= \operatorname{Tr}(\Lambda^2 + \tilde{\Lambda}^2) - 2\operatorname{Tr}(\Lambda O \tilde{\Lambda} O^\top). \end{aligned}$$

Minimizing this is equivalent to maximizing the objective

$$\operatorname{Tr}(\Lambda O \tilde{\Lambda} O^\top) = \sum_{i,j=1}^N \lambda_i \tilde{\lambda}_j O_{ij}^2,$$

which is a linear function of $O^{(2)}$, the entrywise square of O; note that for $O = I$, this expression becomes exactly the LHS above in the original inequality. Now, note that $O^{(2)}$ is a doubly stochastic matrix, with all entries between 0 and 1 and $\sum_i O_{ij}^2 = \sum_j O_{ij}^2 = 1$. Thus, we can maximize our objective over the space of all doubly stochastic matrices.

By the Birkhoff–von Neumann Theorem, this space is a convex polytope, with vertices given by permutation matrices, i.e. matrices of the form P^π for $\pi \in \Sigma_N$, the symmetric group on $1, \dots, N$, where

$$P_{ij}^\pi = \begin{cases} 1 & \pi(i) = j \\ 0 & \text{else} \end{cases}.$$

Thus, this maximum is achieved for a permutation matrix:

$$\max_{O \in O_N} \sum_{i,j=1}^N \lambda_i \tilde{\lambda}_j O_{ij}^2 = \max_{\pi \in \Sigma_N} \sum_{i,j=1}^N \lambda_i \tilde{\lambda}_j P_{ij}^\pi = \max_{\pi \in \Sigma_N} \sum_{i=1}^N \lambda_i \tilde{\lambda}_{\pi(i)}$$

Finally, by the rearrangement inequality, we conclude that the latter maximum is achieved for $\pi = \text{id}$. (For any other π, there exists i such that $\pi(i+1) < \pi(i)$, and then switching π on those values would increase the sum.) Thus, the minimum value of the Frobenius norm above is

$$\sum_{i=1}^N \lambda_i^2 + \tilde{\lambda}_i^2 - 2\lambda_i \tilde{\lambda}_i = \sum_{i=1}^N |\lambda_i - \tilde{\lambda}_i|^2,$$

thus completing the proof. □