Random Matrix Theory, homework 1, due October 8.

Exercise 1: Klein's lemma. Let $f:\mathbb{R}\to\mathbb{R}$ be convex, and $\mathcal{H}_N$ the set of $N\times N$ Hermitian matrices. Prove that

$$\begin{cases} \mathcal{H}_N & \to \mathbb{R} \\ H & \mapsto \operatorname{Tr} f(H) \end{cases}$$

is convex.

Exercise 2: concentration of the Stieltjes transform. Let H be a random matrix from the $N\times N$ Gaussian orthogonal ensemble with the usual normalization, i.e. it has density $e^{-\frac{N}{4}\operatorname{Tr}(H^2)}$ with respect to the Lebesgue measure on the symmetric matrices. Prove that there exists $c>0$ such that for any $t>0$, $N\geq 1$ and $z=E+\mathrm{i}\eta\in\mathbb{H}$ we have

$$\mathbb{P}\left(\left|\frac{1}{N}\operatorname{Tr}\frac{1}{H-z}-\mathbb{E}\frac{1}{N}\operatorname{Tr}\frac{1}{H-z}\right|\geq t\right)\leq c^{-1}e^{-cN^2t^2\eta^4}.$$

Same question when H is symmetric with independent entries (in the upper triangle) bounded with $N^{-1/2}$.

Exercise 3: concentration of individual eigenvalues. Let H be a $N\times N$ symmetric random matrix such that $|H_{ij}|\leq 1$ for any i,j. The goal of this exercise is to prove that for any $1\leq k\leq N$ and $t>0$ we have

$$\mathbb{P}(|\lambda_k-\mathrm{M}(\lambda_k)|\geq t)\leq 4e^{-\frac{t^2}{32k^2}},$$

where $\lambda_N\leq\dots\leq\lambda_1$, the λ_k's are the eigenvalues of H and $\mathrm{M}(\lambda_k)$ is the median of λ_k.

(i) Let $\mathcal{S}$ be the set of $N\times N$ symmetric random matrices, $t>0$, $\mathcal{A}=\{A\in\mathcal{S}:\lambda_k(A)\leq M(\lambda_k))\}$ and $\mathcal{B}=\{B\in\mathcal{S}:\lambda_k(B)\geq M(\lambda_k)+t)\}$. Let $(v_\ell)_{1\leq\ell\leq N}$ be the normalized eigenvectors of some $B\in\mathcal{B}$ ($Bv_\ell=\lambda_k(B)v_\ell$) and

$$\alpha_{ii}=\sum_{\ell=1}^k|v_\ell(i)|^2,\quad \alpha_{ij}=2\Big(\sum_{\ell=1}^k|v_\ell(i)|^2\Big)^{1/2}\cdot\Big(\sum_{\ell=1}^k|v_\ell(j)|^2\Big)^{1/2}.$$

Prove that $\sum_{1\leq i\leq j\leq k}\alpha_{ij}^2\leq 2k^2$.

(ii) Prove that for any $A\in\mathcal{A}$ we have $\sum_{1\leq i\leq j\leq N:A_{ij}\neq B_{ij}}\alpha_{ij}\geq\frac{t}{2}$.

(iii) Conclude.

(iv) Open problem: is there an absolute c such that $\mathbb{P}(|\lambda_k-\mathrm{M}(\lambda_k)|\geq t)\leq c^{-1}e^{-ct^2}$ for any $k,t>0$?

Exercise 4: eigenvalues distribution for the Ginibre ensemble. Read and reproduce the proof of Theorem 15.1.1 in *Log gases and random matrices*, to obtain the following fact. If G has density $e^{-\operatorname{Tr}(GG*)}$ with respect to the Lebesgue density on $N\times N$ matrices (complex entires, no symmetry assumption), then writing $X=UTU^*$ for its Schur form, the matrix T has density ($\lambda_i=T_{ii}$ and $T_{ij},i<j$ are the only non-zero entries) w.r.t the Lebesgue measure proportional to

$$\prod_{i<j}|\lambda_i-\lambda_j|^2e^{-\sum_{i=1}^N|\lambda_i|^2}e^{-\sum_{i<j}|T_{ij}|^2}.$$

In particular the spectrum has density $\prod_{i<j}|\lambda_i-\lambda_j|^2e^{-\sum_{i=1}^N|\lambda_i|^2}$.

Problem 1: the Selberg integral. The purpose of this problem is to calculate the partition function of Gaussian β-ensembles ($\beta \geq 0$), i.e. proving that

$$Z_N^{(\beta)} := \int_{\mathbb{R}^N} |\Delta(\lambda_1, \dots, \lambda_N)|^\beta e^{-N\frac{\beta}{4}\sum_{i=1}^N \lambda_i^2} \mathrm{d}\lambda_1 \dots \mathrm{d}\lambda_N = (2\pi)^{N/2} \left(\frac{\beta N}{2}\right)^{-\frac{N(N-1)\beta}{4} - \frac{N}{2}} \prod_{j=1}^N \frac{\Gamma(1 + j\frac{\beta}{2})}{\Gamma\left(1 + \frac{\beta}{2}\right)}, \tag{1}$$

where $\Delta(\lambda_1, \dots, \lambda_N) = \prod_{1 \leq i < j \leq N} (\lambda_i - \lambda_j)$ is the Vandermonde determinant. Here, $\Gamma(z) = \int e^{-t} t^{z-1}$ for $\mathfrak{Re}(z) > 0$. First, we will prove the Selberg integral formula: for any $\gamma_1, \gamma_2 > -1$ and $\gamma \geq 0$,

$$\begin{aligned} S_N(\gamma_1, \gamma_2, \gamma) &:= \int_{[0,1]^N} \left(\prod_{i=1}^N t_i\right)^{\gamma_1} \left(\prod_{i=1}^N (1 - t_i)\right)^{\gamma_2} |\Delta(t_1, \dots, t_N)|^{2\gamma} \mathrm{d}t_1 \dots \mathrm{d}t_N \\ &= \prod_{j=0}^{N-1} \frac{\Gamma(1 + \gamma_1 + j\gamma)\Gamma(1 + \gamma_2 + j\gamma)\Gamma(1 + (j+1)\gamma)}{\Gamma(2 + \gamma_1 + \gamma_2 + (N + j - 1)\gamma)\Gamma(1 + \gamma)}. \end{aligned} \tag{2}$$

(i) Prove the Euler integral formula:

$$\int_{[0,1]} t^{\gamma_1} (1 - t)^{\gamma_2} \mathrm{d}t = \frac{\Gamma(1 + \gamma_1)\Gamma(1 + \gamma_2)}{\Gamma(2 + \gamma_1 + \gamma_2)},$$

by writing $\Gamma(1+\gamma_1)\Gamma(1+\gamma_2)$ as a double integral and making an appropriate change of variables.

(ii) In question (ii) to (vii), assume $\gamma \in \mathbb{N}$. Prove that

$$S_N(\gamma_1, \gamma_2, \gamma) = \sum_{0 \leq n_1, \dots, n_N \leq 2\gamma(N-1)} c_{n_1, \dots, n_N} \prod_{j=1}^N \frac{\Gamma(1 + \gamma_1 + n_j)\Gamma(1 + \gamma_2)}{\Gamma(2 + \gamma_1 + \gamma_2 + n_j)}$$

for some coefficients $c_{n_1, \dots, n_N}$ independent of γ_1 and γ_2.

(iii) Prove that if $c_{n_1, \dots, n_N} \neq 0$ then $\sum_{i=1}^N n_i = N(N-1)\gamma$. Assuming additionally that $n_1 \leq \dots \leq n_N$, prove that for any $j \in [\![1, N]\!]$ we have

$$(j - 1)\gamma \leq n_j \leq (N + j - 2)\gamma.$$

For the first inequality, you can first consider $j = N$ and then observe that $\Delta(t_1, \dots, t_j)$ divides $\Delta(t_1, \dots, t_N)$. For the second inequality, you can write $\Delta(t_1, \dots, t_j)$ in terms of $\Delta(t_1^{-1}, \dots, t_j^{-1})$.

(iv) Prove that

$$S_N(\gamma_1, \gamma_2, \gamma) = \frac{P(\gamma_1, \gamma_2)}{Q(\gamma_2)} \prod_{j=0}^{N-1} \frac{\Gamma(1 + \gamma_1 + j\gamma)\Gamma(1 + \gamma_2 + j\gamma)\Gamma(1 + (j+1)\gamma)}{\Gamma(2 + \gamma_1 + \gamma_2 + (N + j - 1)\gamma)\Gamma(1 + \gamma)}.$$

where P and Q are polynomials with the same degree in γ_2.

(v) By symmetry in γ_1 and γ_2, prove that P/Q is actually a constant $c(\gamma, N)$.

(vi) By ordering $t_1 \leq \dots \leq t_N$ and conditioning on t_N, prove that

$$S_N(0, 0, \gamma) = \frac{1}{\gamma(N-1) + 1} S_{N-1}(0, 2\gamma, \gamma)$$

(vii) Conclude that (2) holds for any $\gamma \in \mathbb{N}$.

(viii) Prove that (2) holds for any $\gamma > 0$. You can assume the following theorem by Carlson.

If f is analytic on $\mathfrak{Re}(z) \geq 0$, vanishes on $\mathbb{N}$ and $f(z) = \mathrm{O}(e^{\mu z})$ with $\mu < \pi$, then $f = 0$ on $\mathfrak{Re}(z) \geq 0$.

(ix) Prove (1). Hint: $e^{-c\lambda^2} = \lim_{L \to \infty} (1 - \lambda/L)^{cL^2} (1 + \lambda/L)^{cL^2}$.

Problem 2. Loop equations and linear statistics for the Gaussian Unitary Ensemble. Consider the probability distribution of eigenvalues from the Gaussian Unitary Ensemble:

$$\mu(\mathrm{d}\boldsymbol{\lambda}) = \frac{1}{Z_N} \prod_{1\leq k<\ell\leq N} |\lambda_k - \lambda_\ell|^2 e^{-\frac{N}{2}\sum_{k=1}^N \lambda_k^2} \mathrm{d}\lambda_1 \dots \mathrm{d}\lambda_N$$

on the simplex $\lambda_1 < \cdots < \lambda_N$. For a smooth $f : \mathbb{R} \to \mathbb{R}$ supported on $(-2+\kappa, 2-\kappa)$ $(\kappa > 0)$ we consider the general linear statistics $S_N(f) = \sum_{k=1}^N f(\lambda_k) - N \int f(s)\varrho(s)\mathrm{d}s$, where $\varrho(s) = (2\pi)^{-1}\sqrt{(4-s^2)_+}$. We want to prove the weak convergence of $S_N(f)$ to a Gaussian random variable for large N, with no need of any normalization.

We are interested in the Fourier transform $Z(u) = \mathbb{E}_\mu(e^{\mathrm{i}uS_N(f)})$. We will need a complex modification of the GUE, namely $\mathrm{d}\mu^u(\boldsymbol{\lambda}) = \frac{e^{\mathrm{i}uS_N(f)}}{Z(u)}\mathrm{d}\mu(\boldsymbol{\lambda})$, assuming that $Z(u) \neq 0$. Let $s_N(z) = \frac{1}{N}\sum_k \frac{1}{z-\lambda_k}$ and $m_{N,u}(z) = \mathbb{E}^{\mu^u}(s_N(z))$. The Stieltjes transform of the semicircle distribution is $m(z) = \int \frac{\varrho(s)}{z-s}\mathrm{d}s = \frac{z-\sqrt{z^2-4}}{2}$, where the square root is chosen so that m is holomorphic on $[-2,2]^{\mathrm{c}}$ and $m(z) \to 0$ as $|z| \to \infty$.

(i) We denote the first correlation function $\varrho_1^{(N,u)}$, i.e. this is the unique continuous function such that $\mathbb{E}^{\mu^u}[\sum_i f(\lambda_i)] = N \int f\varrho_1^{(N,u)}$ for any continuous, bounded f. Prove that

$$(m_{N,u}(z) - m(z))^2 - \sqrt{z^2-4}\,(m_{N,u}(z) - m(z)) + \frac{\mathrm{i}u}{N}\int_{\mathbb{R}} \frac{f'(s)}{z-s}\varrho_1^{(N,u)}(s)\mathrm{d}s = -\mathrm{var}_{\mu^u}\left(s_N(z)\right).$$

This is called the (first) loop equation. To derive it, you may first prove that

$$m_{N,u}(z)^2 + \int_{\mathbb{R}} \frac{-s + \mathrm{i}uN^{-1}f'(s)}{z-s}\varrho_1^{(N,u)}(s)\mathrm{d}s = -\mathrm{var}_{\mu^u}\left(s_N(z)\right).$$

Hint: integrate by parts or change variables $\lambda_k = y_k + \varepsilon(\mathfrak{Re}/\mathfrak{Im})\frac{1}{z-y_k}$ and note $\partial_{\varepsilon=0}\log Z(u) = 0$.

(ii) Remember the rigidity for Wigner matrices, in particular for GUE: for any $\xi, D > 0$ there exists $C > 0$ such that uniformly in $N \geq 1$ and $k \in [\![1,N]\!]$ we have $\mu\left(|\lambda_k - \gamma_k| > N^{-\frac{2}{3}+\xi}(\hat{k})^{-\frac{1}{3}}\right) \leq CN^{-D}$, where $\int_{-\infty}^{\gamma_k}\varrho(s\mathrm{d}s) = \frac{k}{N}$ and $\hat{k} = \min(k, N+1-k)$. Assume $Z(u) \neq 0$. Prove that

$$|\mu^u|\left(|\lambda_k - \gamma_k| > N^{-\frac{2}{3}+\xi}(\hat{k})^{-\frac{1}{3}}\right) \leq C\frac{N^{-D}}{|Z(u)|},$$

where $|\mu^u|$ is the total variation of the complex measure μ^u. Conclude that uniformly in $z = E+\mathrm{i}\eta$, $-2+\kappa < E < 2-\kappa$, $0 < |\eta| < 1$, we have

$$|\mathrm{var}_{\mu^u}\left(s_N(z)\right)| = \mathrm{O}\left(\frac{N^{-2+2\xi}}{\eta^2|Z(u)|^2}\right).$$

(iii) Prove that uniformly in $-2+\kappa < E < 2-\kappa$, $N^{-1+\xi} \leq \eta \leq 1$, we have

$$m_{N,u}(z) - m(z) = \frac{1}{\sqrt{z^2-4}}\frac{\mathrm{i}u}{N}\int_{\mathbb{R}}\frac{f'(s)}{z-s}\varrho(s)\mathrm{d}s + \mathrm{O}\left(\frac{N^{-2+3\xi}}{\eta^2|Z(u)|^2}\right).$$

(iv) Let $\chi : \mathbb{R} \to \mathbb{R}^+$ be a smooth function such that $\chi(y) = 1$ for $|y| < 1/2$ and $\chi(y) = 0$ for $|y| > 1$. Prove that for any $\lambda \in \mathbb{R}$, we have

$$f(\lambda) = -\frac{1}{2\pi}\iint_{\mathbb{R}^2}\frac{\mathrm{i}yf''(x)\chi(y) + \mathrm{i}(f(x)+\mathrm{i}yf'(x))\chi'(y)}{x+\mathrm{i}y-\lambda}\mathrm{d}x\mathrm{d}y,$$

where the right hand side converges absolutely. For this, you can reproduce the proof of Cauchy's integral formula based on Green's theorem, considering the quasi-analytic extension $(f(x) + \mathrm{i}yf'(x))\chi(y)$.

(v) Note that $\partial_u \log Z(u) = \mathbb{E}_{\mu^u}(\mathrm{i}S_N(f))$. Conclude that bulk linear statistics converge to a Gaussian random variable.