These notes have not received the scrutiny of publication. They could be missing important references, etc.

Lecture 9: Random Matrices III

9 Rigorous Upper Bound

Recap: A Gaussian orthogonal ensemble (GOE) matrix $J \in \mathbb{R}^{n\times n}$ is a random symmetric matrix with entries distributed as

$$J_{ij} = J_{ji}, \quad J_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \frac{1}{n}\right) \text{ for } i < j, \quad J_{ii} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \frac{2}{n}\right) \text{ for } i = 1, \dots, n.$$

In the previous two lectures, we analyzed the largest eigenvalue $\lambda_{\max}(J)$ and derived

$$\lim_{n\to\infty} \mathbb{E}\left[\lambda_{\max}(J)\right] = 2$$

using the replica method.
In this lecture, we provide a rigorous proof of the corresponding upper bound:

$$\mathbb{E}\left[\lambda_{\max}(J)\right] \le 2.$$

9.1 Motivation

For a GOE matrix J, consider the collection[1]

$$g = (\langle x, Jx\rangle)_{x\in S^{n-1}}, \quad S^{n-1} = \left\{x \in \mathbb{R}^n : \|x\|_2 = 1\right\},$$

which defines a mean-zero Gaussian process. The largest eigenvalue of J can then be written as

$$\lambda_{\max}(J) = \max_{x\in S^{n-1}} \langle x, Jx\rangle =: \max_{x\in\mathcal{X}} g_x,$$

where $\mathcal{X} = S^{n-1}$. Our goal is to derive a uniform upper bound for this Gaussian process:

$$\mathbb{E}\left[\max_{x\in\mathcal{X}} g_x\right] \le ?.$$

Since the exact value of $\mathbb{E}\left[\max_{x\in\mathcal{X}} g_x\right]$ in the GOE case is difficult to compute, our strategy is to construct a stochastically dominating Gaussian process h on the same index set $\mathcal{X}$ such that

$$\mathbb{E}\left[\max_{x\in\mathcal{X}} g_x\right] \le \mathbb{E}\left[\max_{x\in\mathcal{X}} h_x\right],$$

and $\mathbb{E}\left[\max_{x\in\mathcal{X}} h_x\right]$ is tractable to compute.
We begin with a general comparison inequality for Gaussian processes, which intuitively states that a process exhibiting greater variability has a larger expected maximum. We will then apply this result to identify a suitable dominating process for the GOE case and complete the rigorous upper bound.

[1] In class, we used an assymetric version of the process which slightly breaks some calculations. These notes have the fixed and simplified version.

Theorem 1 (Sudakov-Fernique Inequality). *Let $(g_x)_{x\in\mathcal{X}}$ and $(h_x)_{x\in\mathcal{X}}$ be two mean-zero Gaussian processes on the same index set $\mathcal{X}$. Suppose that for all $x, y \in \mathcal{X}$,*

$$\mathrm{Var}\left(g_x - g_y\right) \leq \mathrm{Var}\left(h_x - h_y\right).$$

Then

$$\mathbb{E}\left[\log \sum_{x\in\mathcal{X}} \exp(g_x)\right] \leq \mathbb{E}\left[\log \sum_{x\in\mathcal{X}} \exp(h_x)\right], \tag{1}$$

$$\mathbb{E}\left[\max_{x\in\mathcal{X}} g_x\right] \leq \mathbb{E}\left[\max_{x\in\mathcal{X}} h_x\right]. \tag{2}$$

Example 1. Consider two Gaussian random vectors

$$g \sim \mathcal{N}\left(0, \mathbb{1}\mathbb{1}^\mathsf{T}\right), \quad h \sim \mathcal{N}\left(0, I_n\right),$$

which can also be viewed as two Gaussian processes indexed by the finite set $\mathcal{X} = [n]$.
For any $i \neq j \in [n]$,

$$\mathrm{Var}(g_i - g_j) = 0, \quad \mathrm{Var}(h_i - h_j) = 2.$$

By the Sudakov–Fernique inequality,

$$\mathbb{E}\left[\max_{i\in[n]} g_i\right] \leq \mathbb{E}\left[\max_{i\in[n]} h_i\right].$$

In fact, for the random vector g, all coordinates are identical:

$$g_1 = g_2 = \cdots = g_n, \quad g_1 \sim \mathcal{N}\left(0, 1\right), \quad \mathbb{E}\left[\max_{i\in[n]} g_i\right] = \mathbb{E}[g_1] = 0.$$

In contrast, for h, the components $\{h_i\}_{i=1}^n$ are i.i.d $\mathcal{N}\left(0, 1\right)$ random variables, and there exsits a small absolute constant $c > 0$ such that

$$\mathbb{E} \max_{i\in[n]} h_i \geq c\sqrt{\log n}.$$

Intuitively, the randomness in g arises from a single standard Gaussian variable shared across all coordinates, whereas the randomness in h comes from n independent standard Gaussian variables, which aligns with the intuition behind the Sudakov–Fernique inequality.

9.2 Gaussian interpolation

The proof of the Sudakov-Fernique inequality relies on the Gaussian interpolation trick. Let $(g_x)_{x\in\mathcal{X}}$ and $(h_x)_{x\in\mathcal{X}}$ be independent mean-zero Gaussian processes on the same index set $\mathcal{X}$. For any $x \in \mathcal{X}$, define the interpolating process as

$$G_x(t) = \sqrt{t} g_x + \sqrt{1-t} h_x, \quad t \in [0, 1].$$

Then $G_x(0) = h_x, G_x(1) = g_x$, and

$$\mathrm{Var}\left(G_x(t)\right) = t\, \mathrm{Var}\left(g_x\right) + (1-t)\, \mathrm{Var}\left(h_x\right).$$

The motivation for the Gaussian interpolation trick is that, instead of proving the conclusions of Theorem 1 directly, one can show that the corresponding quantities of interest (such as the expected maximum) associated with the interpolated process $G_x(t)$ are decreasing in t. This monotonicity immediately implies the desired inequalities. The elegance of this method is that this can be investigated locally by considering the derivative with respect to t.
Before proceeding to the proof of Theorem 1 via the Gaussian interpolation trick, we introduce a useful lemma called Gaussian integration by parts.

Lemma 1 (**Gaussian Integration By Parts**). *Let $X \sim \mathcal{N}(0, \Sigma)$, where Σ is an $n \times n$ covariance matrix. Then for any differentiable function $f : \mathbb{R}^n \to \mathbb{R}$,*

$$\mathbb{E}\left[X_i f(X)\right] = \sum_{j=1}^{n} \Sigma_{ij} \mathbb{E}\left[\frac{\partial f}{\partial x_j}(X)\right],$$

assuming the expectations above exist and are finite.

Proof. We first establish the result in one dimension. Let $\xi \sim \mathcal{N}(0, 1)$, for any differentiable function $\varphi : \mathbb{R} \to \mathbb{R}$ with compact support,

$$\begin{aligned}
\mathbb{E}\left[\varphi'(\xi)\right] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \varphi'(x) e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \varphi(x) e^{-\frac{x^2}{2}} \Big|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x\varphi(x) e^{-\frac{x^2}{2}} dx \\
&= \mathbb{E}[\xi\varphi(\xi)]
\end{aligned}$$

by integration by parts. By an approximation argument, the result extends to all φ such that $\xi\varphi(\xi)$ and $\varphi'(\xi)$ are integrable.
Write $X = \Sigma^{1/2} Z$, where $Z \sim \mathcal{N}(0, I_n)$. Then $X_i = \sum_{k=1}^{n} \Sigma_{ik}^{1/2} Z_k$, and hence

$$\mathbb{E}\left[X_i f(X)\right] = \sum_{k=1}^{n} \Sigma_{ik}^{1/2} \mathbb{E}\left[Z_k f\left(\Sigma^{1/2} Z\right)\right] = \sum_{k=1}^{n} \Sigma_{ik}^{1/2} \mathbb{E}\left[\mathbb{E}\left[Z_k f\left(\Sigma^{1/2} Z\right) \mid Z_{-k}\right]\right],$$

where $Z_{-k} = \{Z_j\}_{j \neq k}$.
Since $\{Z_k\}_{k=1}^{n}$ are independent, we have $Z_k \mid Z_{-k} \sim \mathcal{N}(0, 1)$. Applying the one-dimensional result together with the chain rule gives

$$\mathbb{E}\left[Z_k f\left(\Sigma^{1/2} Z\right) \mid Z_{-k}\right] = \sum_{j=1}^{n} \Sigma_{jk}^{1/2} \mathbb{E}\left[\frac{\partial f}{\partial x_j}\left(\Sigma^{1/2} Z\right) \mid Z_{-k}\right].$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[X_i f(X)\right] &= \sum_{k=1}^{n} \Sigma_{ik}^{1/2} \sum_{j=1}^{n} \Sigma_{jk}^{1/2} \mathbb{E}\left[\frac{\partial f}{\partial x_j}\left(\Sigma^{1/2} Z\right)\right] \\
&= \sum_{j=1}^{n} \left(\sum_{k=1}^{n} \Sigma_{ik}^{1/2} \Sigma_{kj}^{1/2}\right) \mathbb{E}\left[\frac{\partial f}{\partial x_j}\left(\Sigma^{1/2} Z\right)\right] \\
&= \sum_{j=1}^{n} \Sigma_{ij} \mathbb{E}\left[\frac{\partial f}{\partial x_j}(X)\right],
\end{aligned}$$

which completes the proof. □

9.3 Proof of the Sudakov-Fernique inequality

Proof of Theorem 1. If for any $t \in (0, 1)$ and $\beta > 0$,

$$\frac{d}{dt} \mathbb{E}\left[\log \sum_{x \in \mathcal{X}} \exp(\beta G_x(t))\right] \leq 0, \tag{3}$$

then integrating over t yields

$$\begin{aligned}\mathbb{E}\left[\log\sum_{x\in\mathcal{X}}\exp(\beta G_x(1))\right] &= \mathbb{E}\left[\log\sum_{x\in\mathcal{X}}\exp(\beta G_x(0))\right] + \int_0^1 \frac{d}{dt}\mathbb{E}\left[\log\sum_{x\in\mathcal{X}}\exp(\beta G_x(t))\right]dt \\ &\le \mathbb{E}\left[\log\sum_{x\in\mathcal{X}}\exp(\beta G_x(0))\right].\end{aligned}$$

Taking $\beta = 1$ gives the first inequality in 1. Moreover, since

$$\lim_{\beta\to\infty}\frac{1}{\beta}\log\sum_{x\in\mathcal{X}}\exp(\beta G_x(t)) = \max_{x\in\mathcal{X}} G_x(t),$$

letting $\beta\to\infty$ yields the second inequality in 2.
Therefore, it suffices to prove that for any $t\in(0,1)$ and $\beta>0$, inequality 3 holds. We have

$$\frac{d}{dt}G_x(t) = \frac{1}{2}\left(\frac{g_x}{\sqrt{t}} - \frac{h_x}{\sqrt{1-t}}\right),$$

and hence

$$\begin{aligned}\frac{d}{dt}\log\sum_{x\in\mathcal{X}}\exp(\beta G_x(t)) &= \frac{\sum_{x\in\mathcal{X}}\frac{d}{dt}\exp(\beta G_x(t))}{\sum_{x\in\mathcal{X}}\exp(\beta G_x(t))} \\ &= \frac{\beta}{2}\cdot\frac{\sum_{x\in\mathcal{X}}\exp(\beta G_x(t))\cdot\left(g_x/\sqrt{t} - h_x/\sqrt{1-t}\right)}{\sum_{x\in\mathcal{X}}\exp(\beta G_x(t))}\end{aligned}$$

Since g and h are independent, conditioning on h leaves g as the same Gaussian process. Conditioned on any realization of h, define the softmax weights

$$\psi_{h,x}(g) = \frac{\exp(\beta G_x(t))}{\sum_{x\in\mathcal{X}}\exp(\beta G_x(t))}.$$

By Lemma 1, for each fixed x

$$\begin{aligned}\mathbb{E}\left[\psi_{h,x}(g)\cdot g_x\middle| h\right] &= \sum_{y\in\mathcal{X}}\mathrm{Cov}(g_x,g_y)\mathbb{E}\left[\frac{\partial\psi_{h,x}}{\partial g_y}(g)\middle| h\right] \\ &= \sum_{y\in\mathcal{X}}\mathrm{Cov}(g_x,g_y)\mathbb{E}\left[\frac{\mathbb{1}\left\{y=x\right\}\cdot\beta\sqrt{t}\exp(\beta G_y(t))\sum_{x'\in\mathcal{X}}\exp(\beta G_{x'}(t))}{\left(\sum_{x'\in\mathcal{X}}\exp(\beta G_{x'}(t))\right)^2}\right. \\ &\qquad\left. - \frac{\beta\sqrt{t}\exp(\beta G_x(t))\exp(\beta G_y(t))}{\left(\sum_{x'\in\mathcal{X}}\exp(\beta G_{x'}(t))\right)^2}\middle| h\right] \\ &= \beta\sqrt{t}\left(\mathrm{Cov}(g_x,g_x)\sum_{y\in\mathcal{X}}\mathbb{E}\left[\frac{\exp(\beta G_x(t)+\beta G_y(t))}{\left(\sum_{x'\in\mathcal{X}}\exp(\beta G_{x'}(t))\right)^2}\middle| h\right]\right. \\ &\qquad\left. - \sum_{y\in\mathcal{X}}\mathrm{Cov}(g_x,g_y)\mathbb{E}\left[\frac{\exp(\beta G_x(t)+\beta G_y(t))}{\left(\sum_{x'\in\mathcal{X}}\exp(\beta G_{x'}(t))\right)^2}\middle| h\right]\right)\end{aligned}$$

Define

$$H(x,y) = \mathbb{E}\left[\frac{\exp(\beta G_x(t)+\beta G_y(t))}{\left(\sum_{x'\in\mathcal{X}}\exp(\beta G_{x'}(t))\right)^2}\right] > 0.$$

Then by the law of total expectation,

$$\begin{aligned}\mathbb{E}\left[\frac{\beta}{2}\sum_{x\in\mathcal{X}}\psi_{h,x}(g)\cdot\frac{g_x}{\sqrt{t}}\right] &= \frac{\beta^2}{2}\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{X}}\left(\operatorname{Cov}(g_x,g_x)-\operatorname{Cov}(g_x,g_y)\right)\cdot H(x,y)\\ &= \frac{\beta^2}{2}\sum_{x\neq y}\left(\operatorname{Cov}(g_x,g_x)-\operatorname{Cov}(g_x,g_y)\right)\cdot H(x,y)\\ &= \frac{\beta^2}{4}\sum_{x\neq y}\left(\operatorname{Cov}(g_x,g_x)-2\operatorname{Cov}(g_x,g_y)+\operatorname{Cov}(g_y,g_y)\right)\cdot H(x,y)\\ &= \frac{\beta^2}{4}\sum_{x\neq y}\operatorname{Var}(g_x-g_y)\cdot H(x,y).\end{aligned}$$

By a symmetric argument for the h-term, we obtain

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\left[\log\sum_{x\in\mathcal{X}}\exp(\beta G_x(t))\right] &= \frac{\beta^2}{4}\sum_{x\neq y}\left(\operatorname{Var}(g_x-g_y)-\operatorname{Var}(h_x-h_y)\right)\cdot H(x,y)\\ &\le 0,\end{aligned}$$

where the last inequality follows from the assumption. This verifies inequality (3) and thus completes the proof of Theorem 1. □

9.4 Application: sharp upper bound for GOE

In our GOE case, we compute the variance of the Gaussian process g. For any $x\in\mathcal{X}$ and $y\in\mathcal{X}$,

$$\begin{aligned}\operatorname{Var}(g_x-g_y) &= \operatorname{Var}(\langle x,Jx\rangle-\langle y,Jy\rangle)\\ &= \operatorname{Var}(\langle J,xx^\top-yy^\top\rangle)\\ &= \frac{\|xx^\top-yy^\top\|_F^2+\operatorname{tr}\left(\left(xx^\top-yy^\top\right)^2\right)}{n}\\ &= \frac{4\left(1-\langle x,y\rangle^2\right)}{n}.\end{aligned}$$

To apply the Sudakov–Fernique inequality, we construct another mean-zero Gaussian process with larger pairwise variance. For any $x=(u,v)\in\mathcal{X}$, define

$$h_x=\frac{2\langle Z,x\rangle}{\sqrt{n}},$$

where $Z\sim\mathcal{N}(0,I_n)$. Then for $x\in\mathcal{X}$ and $y\in\mathcal{X}$,

$$\begin{aligned}\operatorname{Var}(h_x-h_y) &= \operatorname{Var}\left(\frac{2\langle Z,x-y\rangle}{\sqrt{n}}\right)\\ &= \frac{4\|x-y\|_2^2}{n}\\ &= \frac{8\left(1-\langle x,y\rangle\right)}{n}.\end{aligned}$$

Since $x,y\in S^{n-1}$, we have $t:=\langle x,y\rangle\le 1$. Hence

$$\begin{aligned}\operatorname{Var}(h_x-h_y)-\operatorname{Var}(g_x-g_y) &= \frac{4(2-2t-1+t^2)}{n}\\ &= \frac{4(1-t)^2}{n}\\ &\ge 0.\end{aligned}$$

By the Sudakov-Fernique inequality,

$$\mathbb{E}\left[\max_{x\in\mathcal{X}} g_x\right] \le \mathbb{E}\left[\max_{x\in\mathcal{X}} h_x\right] = \mathbb{E}\left[\frac{2\|Z\|_2}{\sqrt{n}}\right] \le 2,$$

where the last inequality follows from

$$\mathbb{E}\left[\|Z\|_2\right] = \mathbb{E}\left[\sqrt{\|Z\|_2^2}\right] \le \sqrt{\mathbb{E}\left[\|Z\|_2^2\right]} = \sqrt{n},$$

by Jensen's inequality. Consequently,

$$\mathbb{E}\left[\lambda_{\max}(J)\right] = \mathbb{E}\left[\max_{x\in\mathcal{X}} g_x\right] \le 2,$$

which is the desired upper bound.