Assignment 3

Random Matrix Theory in Data Science and Statistics

(EN.553.796, Fall 2025)

Assigned: October 21, 2025   Due: 11:59pm EST, November 5, 2025

Solve all problems. Each problem is worth an equal amount towards your grade.

Submit solutions in LaTeX. Write in complete sentences. Include and justify all steps of your arguments, but avoid writing excessive discussion that is not contributing to your solution. You are welcome to include images if you think that will help explain your solutions.

Problem 1 (Free probability)**.** Define the 2×2 matrix

$$\boldsymbol{A} := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Let $t \sim \mathsf{Unif}([0, \pi])$ and define the random rotation matrix

$$\boldsymbol{U} := \begin{bmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{bmatrix}.$$

Finally, define $\boldsymbol{X}^{(2d)} := \boldsymbol{I}_d \otimes \boldsymbol{A} \in \mathbb{R}^{2d \times 2d}$ and $\boldsymbol{Y}^{(2d)} := \boldsymbol{I}_d \otimes (\boldsymbol{U}\boldsymbol{A}\boldsymbol{U}^\top) \in \mathbb{R}^{2d \times 2d}$ random matrices.

1. Show that the sequences $\boldsymbol{X}^{(2d)}$ and $\boldsymbol{Y}^{(2d)}$ converge in expected moments (i.e., that $\lim_{d\to\infty} \frac{1}{2d} \mathbb{E}\,\mathsf{Tr}\, \boldsymbol{X}^{(2d)^k}$ exists for all k and likewise for $\boldsymbol{Y}^{(2d)}$) and that the pair of sequences is asymptotically free. (View the definition of asymptotic freeness as restricted to a sequence of matrices in only even dimensions.)

   (Hint: Boil this down to a statement about the 2×2 matrices $\boldsymbol{A}$ and $\boldsymbol{U}$.)

2. To what measure must the empirical spectral distribution of $\boldsymbol{X}^{(2d)} + \boldsymbol{Y}^{(2d)}$ then converge in expected moments? Why?

   (Hint: You do not need to calculate an additive free convolution by hand if you paid attention to lecture or the lecture notes.)

3. Show that the empirical spectral distribution of $\boldsymbol{X}^{(2d)} + \boldsymbol{Y}^{(2d)}$ almost surely consists of at most two atoms. Therefore, qualitatively, it will never resemble the measure you described in Part 2. For example, in a histogram of the eigenvalues, only at most two bins will ever be non-empty. Explain formally and precisely why this is not a contradiction to Part 2.

Problem 2 (Eigenvector perturbation bound)**.** Write $\boldsymbol{v}_1(\boldsymbol{X})$ for the unit-norm eigenvector of $\lambda_1(\boldsymbol{X})$ for $\boldsymbol{X} \in \mathbb{R}^{d\times d}_{\mathsf{sym}}$. Whenever this notation is used below, you may assume that $\lambda_1(\boldsymbol{X})$ occurs with multiplicity 1 as an eigenvalue of $\boldsymbol{X}$.

Suppose $\boldsymbol{M} \in \mathbb{R}^{d\times d}_{\mathsf{sym}}$, and that $\boldsymbol{\Delta}$ has the same dimensions as $\boldsymbol{M}$ and satisfies the operator norm bound $\|\boldsymbol{\Delta}\| < \lambda_1(\boldsymbol{M}) - \lambda_2(\boldsymbol{M})$. You will show the perturbation inequality

$$\langle \boldsymbol{v}_1(\boldsymbol{M}), \boldsymbol{v}_1(\boldsymbol{M}+\boldsymbol{\Delta})\rangle^2 \geq 1 - \left(\frac{\|\boldsymbol{\Delta}\|}{\lambda_1(\boldsymbol{M}) - \lambda_2(\boldsymbol{M}) - \|\boldsymbol{\Delta}\|}\right)^2.$$

Follow these steps, where we abbreviate $\boldsymbol{v} := \boldsymbol{v}_1(\boldsymbol{M})$ and $\widetilde{\boldsymbol{v}} := \boldsymbol{v}_1(\boldsymbol{M}+\boldsymbol{\Delta})$.

1. Show that $\lambda_1(\boldsymbol{M}) - \lambda_i(\boldsymbol{M}+\boldsymbol{\Delta}) \geq \lambda_1(\boldsymbol{M}) - \lambda_2(\boldsymbol{M}) - \|\boldsymbol{\Delta}\|$ for all $i \geq 2$.

   (Hint: You may use the Courant-Fischer min-max theorem. Look it up and take a minute to internalize it if you are not familiar with this.)

2. Using Part 1, show that $\|\boldsymbol{\Delta v}\| \geq (\lambda_1(\boldsymbol{M}) - \lambda_2(\boldsymbol{M}) - \|\boldsymbol{\Delta}\|) \cdot \|(\boldsymbol{I} - \widetilde{\boldsymbol{v}}\widetilde{\boldsymbol{v}}^\top)\boldsymbol{v}\|$.

   (Hint: Expand $\boldsymbol{v}$ in the orthonormal basis of eigenvectors of $\boldsymbol{M}+\boldsymbol{\Delta}$.)

3. Complete the proof.

Also show the following application:

4. Suppose that ν is a probability measure on $\mathbb{R}$ with mean 0, variance 1, and that is σ^2-subgaussian for some $\sigma^2 > 0$. Let $\boldsymbol{W} = \boldsymbol{W}^{(d)} \sim \mathsf{Wig}(d, \nu)$, and let $\boldsymbol{x} = \boldsymbol{x}^{(d)} \in \mathbb{R}^d$ with $\|\boldsymbol{x}\| = 1$. Let $\lambda \geq 0$ be a constant not depending on d and consider the matrix $\boldsymbol{Y} = \lambda\sqrt{d}\,\boldsymbol{x}\boldsymbol{x}^\top + \boldsymbol{W}$. Give a function $f : \mathbb{R}_{\geq 0} \to [0,1]$ such that f is non-decreasing, $f(\lambda) \to 1$ as $\lambda \to \infty$, and such that, for any fixed $\lambda > 0$, we have that

   $$\lim_{d\to\infty} \mathbb{P}[\langle \boldsymbol{v}_1(\boldsymbol{Y}), \boldsymbol{x}\rangle^2 \geq f(\lambda)] = 1.$$

   You may use the limit theorems for $\|\boldsymbol{W}^{(d)}\|$ from class and the lecture notes. Informally, this says that the top eigenvector of $\boldsymbol{Y}$ can achieve an arbitrarily good estimate of a rank-one perturbation of $\boldsymbol{W}$ in the $\boldsymbol{x}$ direction, provided the magnitude λ of the perturbation is large enough. We will see more precise versions in class soon.

Problem 3 (More on Gaussian random vectors)**.** This problem is a continuation of Problem 2 from Homework 1. In the next homework, a final problem in the sequence will have you derive powerful consequences of these ideas for random matrices. For now, you will derive some more general tools and a first application.

1. Suppose $F : \mathbb{R}^d \to \mathbb{R}$ is a smooth function with $\max\{|F(\boldsymbol{x})|, \|\nabla F(\boldsymbol{x})\|_2^2, \|\nabla^2 F(\boldsymbol{x})\|_F^2\} \leq C(1+\|\boldsymbol{x}\|)^K$ for some $C, K > 0$ and all $\boldsymbol{x} \in \mathbb{R}^d$, where $\nabla^2 F$ is the $d \times d$ Hessian matrix of second derivatives. Let $\boldsymbol{\Sigma}, \boldsymbol{\Lambda} \in \mathbb{R}^{d\times d}_{\mathsf{sym}}$ be positive semidefinite. Define $\boldsymbol{\Sigma}(t) := (1-t)\boldsymbol{\Sigma} + t\boldsymbol{\Lambda}$ for $t \in [0,1]$, and write

   $$f(t) := \mathop{\mathbb{E}}_{\boldsymbol{g}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{\Sigma}(t))} F(\boldsymbol{g}).$$

That is, we are considering the value of an expectation of a general function of a Gaussian vector as the covariance matrix moves along a line in matrix space. Show that the derivative of this value is

$$f'(t) = \frac{1}{2}\left\langle \mathbf{\Lambda} - \mathbf{\Sigma}, \mathop{\mathbb{E}}_{\boldsymbol{g}\sim\mathcal{N}(\mathbf{0},\mathbf{\Sigma}(t))} \nabla^2 F(\boldsymbol{g}) \right\rangle.$$

Here, $\langle \boldsymbol{A}, \boldsymbol{B}\rangle = \mathsf{Tr}(\boldsymbol{A}\boldsymbol{B}) = \sum_{i,j} A_{ij}B_{ij}$ is the Frobenius inner product.

You may differentiate under the expectation (i.e., bring a derivative inside an expectation) without justification, but you should consider on your own time what the justification would be.

(**Hint:** If $\boldsymbol{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ and $\boldsymbol{h} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Lambda})$ independently, construct a Gaussian vector with covariance $\mathbf{\Sigma}(t)$ to make differentiating under the expectation easier. Then, use Problem 2 of Homework 1.)

2. Show that, if F as above is also convex, and $\boldsymbol{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ and $\boldsymbol{h} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma})$ are independent Gaussian vectors (that is, the entries of $\boldsymbol{g}$ may be correlated with one another, and likewise for $\boldsymbol{h}$, but entries of $\boldsymbol{g}$ are independent of entries of $\boldsymbol{h}$) for any $\mathbf{\Sigma}, \mathbf{\Gamma} \in \mathbb{R}^{d\times d}_{\mathsf{sym}}$ positive semidefinite, then
$$\mathbb{E}F(\boldsymbol{g}) \le \mathbb{E}F(\boldsymbol{g} + \boldsymbol{h}).$$
Informally, expectations of convex functions of Gaussians are only increased by adding independent noise. Show that the same also holds for $F(\boldsymbol{x}) = \mathsf{max}_{i\in[d]} x_i$, though it is not smooth.

(**Hint:** $\mathsf{Law}(\boldsymbol{g} + \boldsymbol{h}) = \mathcal{N}(\mathbf{0}, \mathbf{\Lambda})$ for some $\mathbf{\Lambda}$. Write this out and use Part 1. For the last part, consider the "soft-max" function $F(\boldsymbol{x}) = \beta^{-1}\log(\sum_{i=1}^d \exp(\beta x_i))$ and take $\beta \to \infty$.)

3. Suppose that $\boldsymbol{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ and $\boldsymbol{h} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Lambda})$ are arbitrary centered Gaussian vectors as in Part 1. Show that $d_{\boldsymbol{g}}(i,j) := (\mathbb{E}(g_i - g_j)^2)^{1/2}$ defines a metric on $[d]$ provided it is not the case that $g_i = g_j$ almost surely for any distinct $i, j \in [d]$, and similarly for $d_{\boldsymbol{h}}(i,j) := (\mathbb{E}(h_i - h_j)^2)^{1/2}$. Suppose that, for all $i, j \in [d]$, we have $d_{\boldsymbol{g}}(i,j) \le d_{\boldsymbol{h}}(i,j)$. Show that
$$\mathbb{E}\max_{i\in[d]} g_i \le \mathbb{E}\max_{i\in[d]} h_i.$$

(**Hint:** For the second part, expand the condition on $\boldsymbol{g}$ and $\boldsymbol{h}$ into a condition on $\mathbf{\Sigma}$ and $\mathbf{\Lambda}$. Again consider the soft-max function and use Part 1, but now explicitly calculate the Hessian.)