High Dimensional Statistics

Spring 2025

Sharif University of Technology

Dr. Amir Najafi

Homework 4 | Random matrices and covariance estimation | Due: 1404/3/14

Problem 1: Bounds on Eigenvalues *[10 points]*

Given two symmetric matrices $\mathbf{A}$ and $\mathbf{B}$, show directly, without citing any other theorems, that

$$|\gamma_{\max}(\mathbf{A}) - \gamma_{\max}(\mathbf{B})| \leq \|\mathbf{A} - \mathbf{B}\|_2 \quad \text{and} \quad |\gamma_{\min}(\mathbf{A}) - \gamma_{\min}(\mathbf{B})| \leq \|\mathbf{A} - \mathbf{B}\|_2.$$

Problem 2: Matrix Monotone Functions *[10 points]*

A function $f : \mathbb{S}_+^{d\times d} \to \mathbb{S}_+^{d\times d}$ on the space of symmetric positive semidefinite matrices is said to be *matrix monotone* if

$$f(\mathbf{A}) \preceq f(\mathbf{B}) \quad \text{whenever } \mathbf{A} \preceq \mathbf{B}.$$

Here $\preceq$ denotes the positive semidefinite ordering on $\mathbb{S}_+^{d\times d}$.

(a) Show by counterexample that the function $f(\mathbf{A}) = \mathbf{A}^2$ is *not* matrix monotone. (Hint: Note that $(\mathbf{A} + t\mathbf{C})^2 = \mathbf{A}^2 + t^2\mathbf{C}^2 + t(\mathbf{AC} + \mathbf{CA})$, and search for a pair of positive semidefinite matrices such that $\mathbf{AC} + \mathbf{CA}$ has a negative eigenvalue.)

(b) Show by counterexample that the matrix exponential function $f(\mathbf{A}) = e^{\mathbf{A}}$ is *not* matrix monotone. (Hint: Part (a) could be useful.)

(c) Show that the matrix logarithm function $f(\mathbf{A}) = \log \mathbf{A}$ is matrix monotone on the cone of strictly positive definite matrices. (Hint: You may use the fact that $g(\mathbf{A}) = \mathbf{A}^p$ is matrix monotone for all $p \in [0, 1]$.)

Problem 3: Sub-Gaussian Random Matrices *[10 points]*

Consider the random matrix $\mathbf{Q} = g\mathbf{B}$, where $g \in \mathbb{R}$ is a zero-mean σ-sub-Gaussian variable.

(a) Assume that g has a distribution symmetric around zero, and $\mathbf{B} \in \mathbb{S}^{d\times d}$ is a deterministic matrix. Show that $\mathbf{Q}$ is sub-Gaussian with matrix parameter $\mathbf{V} = c^2\sigma^2\mathbf{B}^2$, for some universal constant c.

(b) Now assume that $\mathbf{B} \in \mathbb{S}^{d\times d}$ is random and independent of g, with $\|\mathbf{B}\|_2 \leq b$ almost surely. Prove that $\mathbf{Q}$ is sub-Gaussian with matrix parameter given by $\mathbf{V} = c^2\sigma^2 b^2 \mathbf{I}_d$.

Problem 4: Random Packings *[10 points]*

The goal of this exercise is to prove that there exists a collection of vectors $\mathcal{P} = \{\theta^1, \ldots, \theta^M\}$ belonging to the sphere $\mathbb{S}^{d-1}$ such that:

(a) the set $\mathcal{P}$ forms a 1/2-packing in the Euclidean norm;

(b) the set $\mathcal{P}$ has cardinality $M \geq e^{c_0 d}$ for some universal constant c_0;

(c) the inequality $\left\| \frac{1}{M} \sum_{j=1}^{M} (\theta^j \otimes \theta^j) \right\|_2 \leq \frac{2}{d}$ holds.

(Note: You may assume that d is larger than some universal constant so as to avoid annoying subcases.)

Problem 5: Semicircle Law with Correlated Entries *[15 points]*

In this problem, we explore whether the *Wigner semicircle law* continues to hold for ensembles of random matrices with *weakly dependent entries*, rather than independent ones. Our focus is on analyzing a concrete example of such a matrix ensemble to determine how this dependence influences the limiting spectral distribution.

Let $A_n \in \mathbb{R}^{n \times n}$ be a symmetric random matrix, and define the normalized matrix

$$W_n = \frac{1}{\sqrt{n}} A_n, \qquad n \geq 1.$$

Assume that the upper–triangular entries $\{A_{ij}\}_{1 \leq i \leq j \leq n}$ are jointly Gaussian with $\mathbb{E}[A_{ij}] = 0$ and $\mathrm{Var}(A_{ij}) = 1$. Let

$$\rho_{ij,kl} := \mathbb{E}[A_{ij} A_{kl}]$$

denote the covariance between matrix entries. We consider two different models for the dependence structure:

(a) Entries are more strongly correlated when their indices are close. Specifically,

$$\rho_{ij,kl} = r_{\min\{|i-k|, |j-l|\}},$$

where the sequence $(r_\ell)_{\ell \geq 0}$ is non–increasing. A typical example is exponential decay, $r_\ell = c\, e^{-\beta \ell}$ for constants $c > 0$, $\beta > 0$.

(b) Entries in different columns are uncorrelated:

$$\rho_{ij,kl} = r_{|i-k|} \cdot \delta_{jl},$$

where δ_{jl} is the Kronecker delta, and $(r_\ell)_{\ell \geq 0}$ is again non–increasing. This means columns of A_n are pairwise uncorrelated.

As $n \to \infty$, how does the spectrum of W_n behave in each case? In particular, does the empirical spectral distribution converge to the Wigner semicircle law despite the presence of dependence?

Problem 6: Heteroscedastic Wigner Matrix *[15 points]*

In this problem we want to investigate the same Wigner–type random–matrix ensemble, but now the variance of each entry is allowed to depend on its position in the matrix rather than being identically equal to 1.

Let $A_n \in \mathbb{R}^{n\times n}$ be a random symmetric matrix whose upper–triangular entries $\{A_{ij}\}_{1\le i\le j\le n}$ are independent and satisfy

$$\mathbb{E}[A_{ij}] = 0, \qquad \mathbb{E}[A_{ij}^2] = \sigma_{ij}^2 \le \sigma_0^2,$$

for some fixed constant $\sigma_0 > 0$. Set

$$W_n = \frac{1}{\sqrt{n}}\,A_n, \qquad \lambda_{\max}(W_n) = \text{largest eigenvalue of } W_n.$$

(a) Let $\widetilde{A}_n$ be a symmetric matrix with independent centred entries of variance σ_0^2, and define $\widetilde{W}_n = \frac{1}{\sqrt{n}}\widetilde{A}_n$. Investigate the correctness of the following inequality:

$$\mathbb{P}\big(\lambda_{\max}(W_n) \ge x\big) \;\le\; \mathbb{P}\big(\lambda_{\max}(\widetilde{W}_n) \ge x\big), \qquad \forall\, x \ge 0$$

(b) Now suppose $\sigma_{ij}^2 = 1$ for all i, j. Derive an explicit tail bound for $\lambda_{\max}(W_n)$ or show that it is unbounded.

Problem 7: Spectral Detection of Mixture Complexity *[15 points]*

In this problem, the goal is to determine when the number of mixture components can be consistently recovered from spectral information in high dimensions. The analysis centers on how assumptions about variance, separation of means, etc. influence the spectral behavior of the data and enable reliable detection.

Let

$$\mathcal{F} = \Big\{\, x \mapsto \prod_{j=1}^{d} p_{\theta_j}(x_j) \;\Big|\; \theta_1, \dots, \theta_d \in \Theta \Big\}$$

denote a class of product measures on $\mathbb{R}^d$. For each $\theta \in \Theta$ set

$$m(\theta) = \mathbb{E}_{X\sim p_\theta}[X], \qquad V(\theta) = \mathbb{E}_{X\sim p_\theta}\big[(X - m(\theta))^2\big].$$

Assume we observe i.i.d. vectors $X_1, \dots, X_n \in \mathbb{R}^d$ distributed according to

$$X_i \sim \sum_{k=1}^{K} \pi_k \Big(\prod_{j=1}^{d} p_{\theta_j^{(k)}}\Big), \qquad (\pi_1, \dots, \pi_K) \in \Delta^{K-1},\ \theta_j^{(k)} \in \Theta,$$

where both the parameters $\{\theta_j^{(k)}\}$ and the number of components K are unknown. Throughout, suppose $K \ll d$ and regard the regime $d \to \infty$.

(a) Assume $V(\theta) = 1$ for all $\theta \in \Theta$. Under what additional constraints (e.g. on the means $m(\theta_j^{(k)})$ or mixing weights π_k) can the true K be consistently inferred from the spectrum of the (centred or uncentred) sample covariance matrix? Specify which random-matrix theorems you would invoke and explain why the condition $K \ll d$ is essential.

(b) Now suppose merely $V(\theta) \leq \sigma^2$ for some known constant $\sigma^2 < \infty$. How does this relaxation affect your answer to (a)? Can K still be detected spectrally, and if so under which assumptions?

(c) Pinpoint the main technical difficulties encountered in (a) or (b) and try to outline at least one approach that might bypass them.

Reading Material

Read the paper *Learning Rates as a Function of Batch Size: A Random Matrix Theory Approach to Neural Network Training* and answer the following problems.

Problem 8: Paper Summarization *[15 points]*

In this problem, you should write a clear and concise overview of the paper in **no more than two pages**. Please ensure that your summary:

- **Highlights the paper's main goals and contributions.** What are the authors aiming to achieve, and why is it important?
- **Describes the key technical ideas and methods.** How do the authors approach the problem? What techniques or frameworks do they employ?
- **Discusses the significance and implications of the results.** Why are these findings impactful, and how might they influence future research or applications?

Above all, your summary should be well-structured, straightforward, and demonstrate a thorough understanding of the paper's content.