Exercises in Mathematics of Data Science

Afonso S. Bandeira, Pedro Abdalla Teixeira, Kevin Lucca, Anastasia Kireeva, Petar Nizić-Nikolac

October 11, 2025

This is a list of exercises given as homework exercises or exam questions in classes taught by the first author over the last decade (following, e.g., [Ban16, BSS25, Ban25]), Most exercises are meant for graduate level courses in Mathematics of Data Science, while some have have also appeared in Bachelor level courses in related topics (see [Ban25]). This list was prepared together with the other authors who have been TAs for these courses in ETH Zürich at various times and have came up with many of these exercises. Since the first author is often asked for exercises around these topics, we thought it would be most useful to others to make a list available online. We hope these are useful to educators, and for students to use as an additional practice ground. It is also a good resource for readers of the text by the first author, Singer, and Strohmer [BSS25], designed for a graduate level course. There are several other excellent resources for exercises, you can find some within the references of [BSS25, Ban25]. For the readers looking for more challenging problems, we recommend [BKMR25].[1]

A couple of disclaimers: We found it essentially impossible to trace back the origin of each exercise. For many of these exercises we were inspired by several other sources, including homework sets and exams that we solved ourselves when we were students. Also, as you see below, we tried to mark the expected difficulty of problems. We advise the reader to keep in mind that rating difficulty of problems tends to be unreliable.

Contents

References

[Ban16] Afonso S. Bandeira. Ten lectures and forty-two open problems in the mathematics of data science. *Available online at: https://people.math.ethz.ch/~abandeira//TenLecturesFortyTwoProblems.pdf*, 2016.

[Ban25] Afonso S. Bandeira. A tour through the "mathematics of signals, networks, and learning". *Lecture Notes. Available online: https://people.math.ethz.ch/~abandeira//MathofSNLnotes2025.pdf*, 2025.

[BKMR25] Afonso S. Bandeira, Anastasia Kireeva, Antoine Maillard, and Almut Rödder. Randomstrasse101: Open problems of 2024. *arXiv preprint arXiv:2504.20539*, 2025.

[BSS25] Afonso S. Bandeira, Amit Singer, and Thomas Strohmer. Topics in mathematics of data science. *Preprint available online: https://people.math.ethz.ch/~abandeira//BandeiraSingerStrohmer-MDS-draft.pdf*, 2025.

[1] The blog https://randomstrasse101.math.ethz.ch/ has several open problems in this area.

2 Curses, Blessings, and Surprises in High Dimensions

Problem 2.1 (•• | Drawing points in the unit ball)**.** Let $d \geq 2$ and $x_1, \ldots, x_n \in \mathbb{R}^d$ are sampled uniformly at random from the unit ball B^d. The goal of this exercise is to show that with high probability all points will be contained in an annulus of width $2 \ln n / d$ (A events), and every pair of points will be nearly orthogonal (O events). Namely, for any distinct numbers $i, j \in [n]$ define the events

$$A_i = \left\{ \|x_i\| \geq 1 - \frac{2 \ln n}{d} \right\}, \qquad O_{i,j} = \left\{ |\langle x_i, x_j \rangle| \leq \frac{\sqrt{6 \ln n}}{\sqrt{d-1}} \right\}.$$

Prove that there exists a constant $C > 0$ (independent of both n and d) such that

$$\mathbb{P}\left(A_i \text{ for all } i, O_{i,j} \text{ for all } i \neq j\right) \geq 1 - \frac{C}{n}.$$

Problem 2.2 (• | Integral identity and Chebyshev's inequality)**.** Let X be a non-negative integrable random variable.

(a) Prove the integral identity:

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > t)\, dt.$$

(b) Generalize it for an arbitrary integrable random variable Y (not necessarily non-negative).

(c) Assume that Y has a finite absolute moment $\mathbb{E}|Y|^p$ for some $p > 0$. Find an integral expression for $\mathbb{E}|Y|^p$.

(d) Let X be any random variable with finite expected value and finite p-th central moment $\mathbb{E}\,|X - \mathbb{E}X|^p$, for some $p \geq 1$. Prove that for any $t > 0$:

$$\mathbb{P}\left(|X - \mathbb{E}X| \geq t\right) \leq \frac{\mathbb{E}\,|X - \mathbb{E}X|^p}{t^p}.$$

Problem 2.3 (• | MGF)**.** Let X be a real random variables. Recall the definition of the MGF:

$$M_X(\lambda) = \mathbb{E}\left[e^{\lambda X}\right].$$

(a) Show that $M_{X+X'}(\lambda) = M_X(\lambda) M_{X'}(\lambda)$, where X' is another random variable independent of X.

(b) Show that $M_X(\lambda) \geq \exp\left(\lambda\, \mathbb{E}[X]\right)$.

(c) Let Z be a standard Gaussian random variable, i.e. the one having the probability density function

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

Show that $M_Z(\lambda) = e^{\lambda^2/2}$.

Problem 2.4 (•• | Cantelli's inequality)**.** Let X be a real random variable with finite mean and finite variance $\mathrm{Var}(X)$. Then for any $t > 0$,

$$\mathbb{P}(X - \mathbb{E}X \geq t) \leq \frac{\mathrm{Var}(X)}{\mathrm{Var}(X) + t^2}.$$

Hint. *Work with the shifted random variable $Y = X - \mathbb{E}X + u$, for arbitrary shift u.*

Problem 2.5 (•• | Paley–Zygmund inequality)**.** Let X be a non-negative random variable with finite variance. Prove that for any $0 \leq \theta \leq 1$:

$$\mathbb{P}\left(X > \theta\, \mathbb{E}[X]\right) \geq (1-\theta)^2 \frac{(\mathbb{E}[X])^2}{\mathbb{E}[X^2]}.$$

Hint. *Note $\mathbb{E}[X] = \mathbb{E}\left[X \mathbb{1}_{\{X \leq \theta \mathbb{E}[X]\}}\right] + \mathbb{E}\left[X \mathbb{1}_{\{X > \theta \mathbb{E}[X]\}}\right]$. How could one obtain the term $\mathbb{E}\left[X^2\right]$?*

Problem 2.6 (• | log-MGF and Cramér transform)**.** Let X be a centered (i.e. $\mathbb{E}[X] = 0$) random variable. We define the logarithm of the moment generating function (log-MGF) for $\lambda \in \mathbb{R}$ as

$$\psi_X(\lambda) = \log \mathbb{E}\left[e^{\lambda X}\right].$$

Suppose that it exists in an open neighbourhood around zero. The Cramér transform of X is defined for $t \in \mathbb{R}$:

$$\psi_X^*(t) = \sup_{\lambda \in \mathbb{R}} \left(\lambda t - \psi_X(\lambda)\right).$$

(a) Prove that for any $t \geq 0$, one has

$$\mathbb{P}(X \geq t) \leq \exp\left(-\psi_X^*(t)\right).$$

(b) Suppose that $X_1, \ldots, X_n$ are n i.i.d. copies of X, and let $S_n = X_1 + \ldots + X_n$. Prove that for any $t \in \mathbb{R}$:

$$\psi_{S_n}^*(t) = n\psi_X^*(t/n).$$

Problem 2.7 (•• | Chernoff bound for polynomial vs. exponential moments)**.** Let X be a non-negative real random variable whose MGF is finite over $\mathbb{R}$. Fix $t > 0$. Show that

$$\inf_{p \in \mathbb{N} \cup \{0\}} \frac{\mathbb{E}\left[X^p\right]}{t^p} \leq \inf_{\lambda > 0} \frac{\mathbb{E}\left[e^{\lambda X}\right]}{e^{\lambda t}}.$$

Problem 2.8 (•• | Poisson tail bound)**.** Let X be a Poisson random variable with parameter $\mu \in (0, \infty)$. Prove that for any $t > 0$:

$$\mathbb{P}(X > \mu + t) \leq \exp\left(-\mu\, h(t/\mu)\right),$$

where $h(x) = (1 + x)\log(1 + x) - x$.

Problem 2.9 (•• | Weak bound for Komlós Conjecture)**.** Let $A \in \mathbb{R}^{n \times n}$ matrix whose columns $a_1, \ldots, a_n$ satisfy $\|a_i\|_2 = 1$ for all $i \in \{1, \ldots, n\}$. Prove that there exists an absolute constant $C > 0$ such that

$$\min_{\varepsilon \in \{-1,+1\}^n} \|A\varepsilon\|_\infty \leq C\sqrt{n}.$$

Hint. *Introduce randomness on ε and take expectation.*

Problem 2.10 (• | Properties of subgaussian random variables)**.** Recall [BSS25, Definition 2.9].

(a) Show that if X and X' are independent mean-zero subgaussian random variables with variance parameters σ and σ' respectively, then $X + X'$ is a subgaussian random variable with variance parameter $\sqrt{\sigma^2 + \sigma'^2}$.

(b) Is this still true if we drop the independence assumption?

(c) Show that for any $\lambda \in \mathbb{R}$ it holds

$$\mathrm{Var}(X) \leq \frac{2}{\lambda^2}\left(\exp\left(\frac{\lambda^2\sigma^2}{2}\right) - 1\right).$$

(d) Using L'Hôpital's rule, or otherwise, taking $\lambda \to 0$ deduce that

$$\mathrm{Var}(X) \leq \sigma^2.$$

Hint. *This inequality might be useful: $2 + t^2 \leq e^t + e^{-t}$.*

Problem 2.11 (•• | Hoeffding's lemma)**.** Let X be a random variables such that $X \in [a, b]$ a.s., for some real numbers $a \leq b$. In the proof of Hoeffding's theorem [BSS25, Theorem 2.15] it was shown that X is $(b-a)/2$-subgaussian. We will show here that this is optimal, and prove a slighly weaker version using symmetrization.

(a) Let X' be an independent copy of X, and set $Y = X - X'$. This is usually called the *symmetrization* of X. Show that for any $\lambda \in \mathbb{R}$, the following inequality between MGFs holds:

$$\mathbb{E}\left[e^{\lambda(X - \mathbb{E}X)}\right] \leq \mathbb{E}\left[e^{\lambda Y}\right].$$

(b) Show that

$$\mathbb{E}\left[e^{\lambda Y}\right] = \mathbb{E}\left[\cosh(\lambda Y)\right].$$

(c) Using approximation $\cosh(x) \leq e^{x^2/2}$ (no proof needed) conclude that X is $(b-a)$-subgaussian.

(d) Using Problem 2.10 (no proof needed) show that for any real numbers $a \leq b$, there is a random variable X such that $X \in [a, b]$ a.s., and for any $\sigma < (b-a)/2$, X is not subgaussian with parameter σ.

Remark. *Note that in [BSS25, Definition 2.9] the subgaussianity condition is given in terms of $X - \mu$, where μ is the mean of X.*

Problem 2.12 (• | Hoeffding's inequality for subgaussians)**.** Let $X_1, \ldots, X_n$ be independent random variables such that X_i is subgaussian with parameter σ_i, and let $S_n = X_1 + \ldots + X_n$ be the sum. Fix $t > 0$.

(a) Show that the Cramér transform of S_n is lower bounded by

$$\psi^*_{S_n}(t) \geq \frac{t^2}{2\sum \sigma_i^2}.$$

(b) Using Problem 2.6 deduce that

$$\mathbb{P}\left(|S_n - \mathbb{E}S_n| > t\right) \leq 2\exp\left(-\frac{t^2}{2\sum_{i=1}^n \sigma_i^2}\right).$$

Problem 2.13 (●●● | Bernstein's inequality - bounded moments)**.** Let $X_1, \ldots, X_n$ be independent centered random variables such that for all $i \in [n]$ and integers $m \geq 2$, one has

$$\mathbb{E}\,|X_i|^m \leq \frac{\sigma_i^2 R^{m-2}}{2} m!,$$

where $R > 0$ and $\sigma_i > 0$ are constants that may depend only on distribution of X_i.

(a) Prove that, for all $t > 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| > t\right) \leq 2\exp\left(-\frac{t^2}{2(\nu^2 + Rt)}\right),$$

where $\nu^2 = \sum_{i=1}^n \sigma_i^2$.

(b) Deduce Bernstein's inequality for bounded variables (Theorem 2.17 in lecture notes). Namely, show that for independent centered random variables $X_1, \ldots, X_n$ satisfying $|X_i| \leq a$ a.s. and $\mathbb{E}X_i^2 \leq \sigma^2$, it holds

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| > t\right) \leq 2\exp\left(-\frac{t^2}{2n\sigma^2 + \frac{2}{3}at}\right)$$

for any $t > 0$.

Hint. *After applying the Chernoff bound, choose* $\lambda = \frac{t}{\nu^2 + Rt}$.

3 Singular Value Decomposition and Principal Component Analysis

Problem 3.1 (• | Equivalent definitions of spectral/operator norms). Given a matrix $M \in \mathbb{R}^{m\times n}$, prove that all of the following quantities are equal:

(a) $\sup_{\|v\|_2=1} \|Mv\|_2$, the *operator norm* of M, which is commonly denoted by $\|M\|$;

(b) $\sup_{v\neq 0} \frac{\|Mv\|_2}{\|v\|_2}$;

(c) $\sup_{\|u\|_2=\|v\|_2=1} u^\top Mv$;

(d) $\sigma_1(M)$, the largest singular value of M;

(e) $\sqrt{\lambda_1(MM^\top)}$, the square root of the largest eigenvalue of $MM^\top$;

(f) $\sqrt{\lambda_1(M^\top M)}$, the square root of the largest eigenvalue of $M^\top M$.

Problem 3.2 (• | Maximal entry bound). Given a matrix $X \in \mathbb{R}^{n\times m}$, show that for any $i \in [n]$ and $j \in [m]$ we have

$$-\|X\| \leq |X_{ij}| \leq \|X\|\,.$$

Problem 3.3 (• | Symmetrization of matrices). We are going to explore three ways in which an $m \times n$ (with $m \leq n$) real-valued matrix M can be symmetrized.

(a) Let A be an $m \times m$ matrix defined by

$$A \coloneqq MM^\top.$$

Check that A is symmetric, and show that its m eigenvalues are given by: $\sigma_1(M)^2, \sigma_2(M)^2, \ldots, \sigma_m(M)^2$.

(b) Show that A and $B \coloneqq M^\top M$ have the same non-zero eigenvalues, up to multiplicities.

(c) Let C be an $(m+n)\times(m+n)$ matrix defined by

$$C \coloneqq \begin{pmatrix} 0_{m\times m} & M \\ M^\top & 0_{n\times n} \end{pmatrix},$$

where $0_{r\times r}$ is an $r \times r$ all-zeros matrix. Check that C is symmetric, and show that its $m+n$ eigenvalues are given by:

(1) $\sigma_1(M), \ldots, \sigma_m(M)$; (2) $-\sigma_1(M), \ldots, -\sigma_m(M)$; (3) $n - m$ of them are zeros.

Problem 3.4 (• • | Gershgorin circle theorem). Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix with entries $(a_{ij})_{i,j\in[n]}$. For $i \in [n]$ let R_i be the sum of the absolute values of the non-diagonal entries in the i-th row:

$$R_i = \sum_{j\neq i} |a_{ij}|\,.$$

Prove that every eigenvalue of A lies within at least one of the Gershgorin discs $D(a_{ii}, R_i)$, i.e. for any eigenvalue λ of A we can find $i \in [n]$ such that $|\lambda - a_{ii}| \leq R_i$.

Problem 3.5 (• | Low rank approximation). Let $A \in \mathbb{R}^{m\times n}$ and k be an integer such that $1 \leq k \leq \operatorname{rank}(A)$.

(a) Prove that there exists a matrix $B \in \mathbb{R}^{m\times n}$ of rank k such that

$$\|A - B\| \leq \frac{\|A\|_F}{\sqrt{k}}.$$

(b) Does the statement (a) hold if the operator norm on the left hand side is replaced with the Frobenius norm $\|A - B\|_F$?

Problem 3.6 (• • | Quadratic form optimization). Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n$. Given $r \in \{1, 2, \ldots, n\}$, consider the following optimization problem:

$$\max_{v_1,\ldots,v_r\in\mathbb{R}^n} \sum_{i=1}^{r} v_i^\top A v_i \quad \text{s.t.} \quad v_i^\top v_j = \delta_{ij} \text{ for } 1 \leq i, j \leq r,$$

where δ_{ij} is the Kronecker delta defined as

$$\delta_{ij} = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

(a) Show that $\mathrm{Tr}(A)$ is the solution of the problem when $r = n$.

(b) Determine the solution of the problem in terms of the eigenvalues of A when $r < n$.

Problem 3.7 (• | Inner product between matrices)**.** For any two matrices $A, B \in \mathbb{R}^{n\times n}$, consider the map

$$\langle A, B\rangle := \mathrm{Tr}\left(AB^\top\right).$$

(a) Prove that $\langle \cdot, \cdot \rangle$ is an inner product on the space of $n \times n$ matrices.

(b) Show that $\|A\|_F^2 = \langle A, A\rangle$.

(c) Deduce the *matrix Cauchy-Schwarz inequality*: $\langle A, B\rangle \leq \|A\|_F \|B\|_F$.

Problem 3.8 (•• | Rotation minimisation)**.** Let $A, B \in \mathbb{R}^{m\times n}$ be two arbitrary matrices. Find the solution, in terms of A and B, or their SVD decompositions, of the following optimization problem:

$$\arg\min_{\Omega\in\mathcal{O}(m)} \|\Omega A - B\|_F .$$

Here $\mathcal{O}(m)$ denotes the set of all $m \times m$ orthogonal matrices.

Problem 3.9 (• | Polar decomposition)**.** Let $A \in \mathbb{R}^{n\times n}$. Prove that there exists a positive semi-definite matrix P and an orthogonal matrix Q such that $A = PQ$.

Problem 3.10 (•• | Power method)**.** Let $A \in \mathbb{R}^{n\times n}$ be a symmetric positive semi-definite matrix with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n \geq 0$ and associated eigenvectors $v_1, \ldots, v_n \in \mathbb{R}^n$ (that form an orthonormal basis). In this exercise, the goal is to show that the power method converges exponentially fast.

(a) Let $y_0 \in \mathbb{R}^n$ be an initial vector that satisfies $Ay_0 \neq 0$. Define the power method iteration for $k \geq 0$:

$$y_{k+1} = \frac{Ay_k}{\|Ay_k\|}.$$

Prove that these iterations are well-definied, i.e. that $\|Ay_k\| \neq 0$ for any $k \geq 1$.

(b) Define the Rayleigh quotient as $\xi_k = y_k^\top A y_k$ and its relative error by

$$\mathrm{err}(\xi_k) = \frac{\lambda_1 - \xi_k}{\lambda_1}.$$

If A is diagonal, i.e., $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and $\lambda_1 = 1$, show that we can represent the error as

$$\mathrm{err}(\xi_k) = \frac{\sum_{i=2}^n w_i^2 \lambda_i^{2k}(1-\lambda_i)}{w_1^2 + \sum_{i=2}^n w_i^2\lambda_i^{2k}},$$

where $w_i = \langle y_0, v_i\rangle$ for all $i \in [n]$.

(c) If A is diagonal with $1 = \lambda_1 > \lambda_2 > \lambda_3$, and we start from y_0 such that $w_1 \neq 0$ and $w_2 \neq 0$, show that

$$\frac{\mathrm{err}(\xi_{k+1})}{\mathrm{err}(\xi_k)} \to \left(\frac{\lambda_2}{\lambda_1}\right)^2 \quad \text{as } k \to \infty.$$

(d) Generalize the result in (c) for an arbitrary matrix (not necessarily diagonal) having $\lambda_1 > \lambda_2 > \lambda_3$.

Problem 3.11 (•• | Moore-Penrose Pseudoinverse)**.** Let A be an $n \times m$ real matrix. A *pseudoinverse* of A is an $m \times n$ matrix A^+ such that the following three conditions are simultaneously met:

- $AA^+A = A$.
- $A^+AA^+ = A^+$.
- Both AA^+ and A^+A are symmetric.

(a) Let Σ be an $n \times m$ rectangular diagonal matrix ($\Sigma_{ij} = 0$ for $i \neq j$) with non-negative entries. Find an $m \times n$ rectangular diagonal matrix Σ^+ that is a pseudoinverse of Σ.

(b) Given a general $n \times m$ matrix A, consider the singular value decomposition of $A = U\Sigma V^T$ with U and V being orthogonal matrices, and Σ being (rectangular) diagonal. Prove that the matrix A^+, given by $A^+ = V\Sigma^+U^T$, is a pseudoinverse of A.

(c) Prove that if A is an invertible $n \times n$ matrix, then A^{-1} is a pseudoinverse of A.

(d) Prove that if A has full column rank (its columns are linearly independent) then its pseudoinverse is given by

$$A^{+} = \left(A^{\top} A\right)^{-1} A^{\top}.$$

(e) Prove that the pseudoinverse is unique.

Definition. *Given an integer n, a* standard Gaussian Wigner matrix $W \in \mathbb{R}^{n \times n}$ *is a symmetric random matrix whose diagonal and upper-diagonal entries are jointly independent Gaussian variables, such that $W_{ii} \sim \mathcal{N}(0,2)$ and, for $i < j, W_{ij} = W_{ji} \sim \mathcal{N}(0,1)$.*

Problem 3.12 (**•••** | BBP for spiked Wigner model)**.** In the lectures you learned about BBP transition for the Wishart model, i.e., when we observe $Y = \frac{1}{n} X X^{\top}$, where X is an $p \times n$ matrix with columns drawn independently from $\mathcal{N}(0, I_p + \beta u u^{\top})$. We will explore the similar type of phase transition for another model.

Let W be an $n \times n$ Wigner matrix, v be a unit-norm vector in $\mathbb{R}^n$ and $\xi \geq 0$. We define the *spiked Wigner model* as observing $Y = \frac{1}{\sqrt{n}} W + \xi v v^{\top}$, with the aim of recovering the *signal* v. This model exhibits the following phase transition (as $n \to \infty$) for

1. the largest eigenvalue $\lambda_{\max}$ of Y:

$$\lambda_{\max} \to \begin{cases} 2 & \text{if } \xi \leq 1, \\ \xi + \frac{1}{\xi} & \text{if } \xi > 1; \end{cases}$$

2. the leading eigenvector $v_{\max}$ of Y:

$$\left|\langle v_{\max}, v \rangle\right|^2 \to \begin{cases} 0 & \text{if } \xi \leq 1, \\ 1 - \frac{1}{\xi^2} & \text{if } \xi > 1. \end{cases}$$

To measure the quality of the recovery procedure we define the *mean squared error* of an estimate $w \in \mathbb{R}^n$ as

$$\operatorname{mse}(w) = \mathbb{E}\left[\left\| w w^{\top} - v v^{\top} \right\|^2\right].$$

Find the asymptotic behaviour of the mean squared error for the PCA estimator, i.e., the value of

$$\lim_{n \to \infty} \operatorname{mse}(v_{\max})$$

as a function of ξ.

Remark. *The mean square error defined as above might appear unnatural for this problem since one can measure the difference between two vectors using ℓ_2 norm (up to a sign). However, this type of MSE definition can be more useful when the perturbation is not rank-one and non-symmetric (e.g., in low-rank matrix estimation problems). Additionally, it addresses the issue of sign invariance in the model, where the observation remains the same whether the signal is v or $-v$, and therefore, we can recover the vector only up to a sign.*

4 Graphs, Networks, and Clustering

Definition (Irreducible matrix)**.** *A matrix $A \in \mathbb{R}^{n\times n}$ is called irreducible if there is no permutation matrix P such that*

$$P^\top AP = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix},$$

where A_{11} and A_{22} are square matrices (not necessarily of same dimensions). In other words, an irreducible matrix cannot be transformed into block upper-triangular matrix by simultaneous row/column permutations.

Problem 4.1 (•• | Irreducibility and graphs)**.** Let $A \in \mathbb{R}^{n\times n}$ be a matrix with non-negative entries. We define $G(A)$, a directed graph associated to A, in the following way: there is a link from i to j if and only if $A_{ij} > 0$.

(a) Prove that if A is irreducible, and x is its eigenvector with non-negative entries, then x has only positive entries.

(b) Show that the statement above is false if we drop the assumption that A is irreducible.

(c) Show that A is irreducible if and only if the associated graph $G(A)$ is strongly connected, which means that for every ordered pair of nodes (i, j) there is a path from i to j (of any length).

Problem 4.2 (••• | PageRank and Random Teleports)**.** In the lectures, we considered the PageRank algorithm designed for ranking pages based on their importance by analysing their ingoing and outgoing links. However, there exist graphs such that PageRank fails to predict meaningful scores. We will consider a simple fix for this problem which is often referred to as Random Teleports.

Let $n > k > 1$. Consider a directed graph on $n + 1$ vertices, labelled $0, 1, \ldots, n$, with the following links. Vertex 0 links only to itself, and any other vertex $j \in [n]$ has k outgoing edges to its next k vertices: $j+1, \ldots, j+k \mod n$, and an edge to vertex 0. See Figure 4.1 for an example of such a graph.

(a) Compute the rank of vertices according to the PageRank scheme described in Section 4.1 of Lecture Notes.

(b) We define PageRank with Random Teleports as follows: with probability β a random walker follows a link at random, and with probability $1 - \beta$, jumps to a random vertex (link or vertex is chosen uniformly at random). We form a new random walk matrix $M \in \mathbb{R}^{(n+1)\times(n+1)}$ whose entries m_{ij} equal to the probability of going from vertex j to vertex i. The ranking is then defined as the leading eigenvector of the constructed matrix M. For $k = 1$ and fixed $0 < \beta < 1$ compute the PageRank scores for nodes with the teleport probability $1 - \beta$.

Remark. *Have you recognized the connection with the notion of irreducibility in this exercise?*

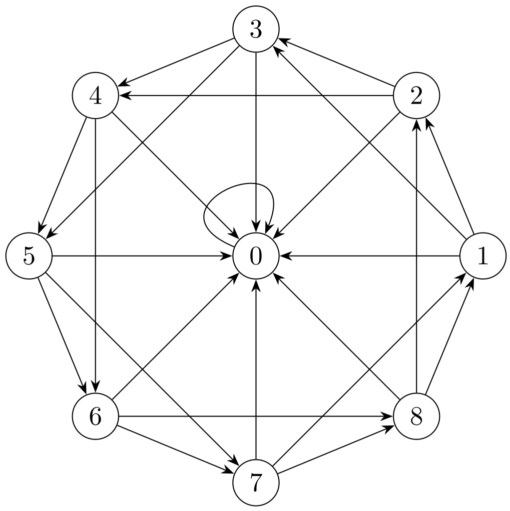

Figure 4.1: Example of the directed graph from Problem 4.2 with $n = 8$ and $k = 2$.

Problem 4.3 (• | Lloyd's algorithm - monotonicity)**.** Recall the problem of k-means clustering: given $x_1, \ldots, x_n \in \mathbb{R}^p$ (with $n \geq k$), we want to minimize the following objective function

$$\mathrm{cost}_2(S_1, \ldots, S_k; \mu_1, \ldots, \mu_k) := \sum_{l=1}^{k} \sum_{i\in S_l} \|x_i - \mu_l\|_2^2 \tag{4.1}$$

that depends on the clusters $S_1, \ldots, S_k$ with centers $\mu_1, \ldots, \mu_k \in \mathbb{R}^p$. Denote the minimum value by

$$\mathrm{opt}_2 \coloneqq \min_{\substack{\text{partition } S_1,\ldots,S_k \\ \text{centers } \mu_1,\ldots,\mu_k}} \mathrm{cost}_2(S_1, \ldots, S_k; \mu_1, \ldots, \mu_k)$$

Prove the following two properties.

(a) Given a choice for the partition $S_1, \ldots, S_k$ (of non-empty sets), the centers that minimize (4.1) are given by

$$\mu_l = \frac{1}{|S_l|} \sum_{i \in S_l} x_i.$$

(b) Given the centers $\mu_1, \ldots, \mu_k \in \mathbb{R}^p$ the partition that minimizes (4.1) assigns each point x_i to the cluster

$$l = \arg\min_{l=1,\ldots,k} \|x_i - \mu_l\|_2 .$$

Problem 4.4 (• | k-means objective - equivalent problem)**.** Consider the same setting of Problem 4.3. Show that

$$\mathrm{opt}_2 = \min_{S_1,\ldots,S_k} \sum_{l=1}^{k} \frac{1}{|S_l|} \sum_{i,j \in S_l} \|x_i - x_j\|_2^2. \tag{4.2}$$

Hint. *Expand the square in (4.2) at the optimal choice of centers* $\mu_l = \frac{1}{|S_l|} \sum_{i \in S_l} x_i$.

Problem 4.5 (• | Lloyd's algorithm - convergence)**.** Given any set of n points in $\mathbb{R}^p$, prove that Lloyd's algorithm stops after a finite number of iterations, in other words, that the objective eventually stops decreasing.

Hint. *There are finitely many partitions of* n *points.*

Problem 4.6 (•• | Lloyd's algorithm - different objective)**.** Let $n \in \mathbb{N}$ be odd, $k \leq n$, and $x_1, \ldots, x_n \in \mathbb{R}^p$. Instead of minimizing sum-of-squares of ℓ_2 norms (4.1), suppose we want to minimize an objective function with ℓ_1 norms:

$$\mathrm{cost}_1(S_1, \ldots, S_k; \mu_1, \ldots, \mu_k) \coloneqq \sum_{l=1}^{k} \sum_{i \in S_l} \|x_i - \mu_l\|_1 . \tag{4.3}$$

Denote the minimum value by

$$\mathrm{opt}_1 \coloneqq \min_{\substack{\text{partition } S_1,\ldots,S_k \\ \text{centers } \mu_1,\ldots,\mu_k}} \mathrm{cost}_1(S_1, \ldots, S_k; \mu_1, \ldots, \mu_k).$$

(a) Given a choice for the partition $S_1, \ldots, S_k$ (of non-empty sets), which centers do minimize the alternative objective function (4.3)? A proof of minimality needs to be provided.

(b) Develop an algorithm analogous to Lloyd's algorithm using the alternative objective function (4.3).

(c) Prove that it is always the case that $\mathrm{opt}_2 \leq \mathrm{opt}_1^2$.

Problem 4.7 (• | Laplacian and connectivity)**.** Given an undirected graph G, with the associated Laplacian L_G, show that $\lambda_2(L_G) > 0$ if and only if G is connected.

Problem 4.8 (• | Normalized Laplacian)**.** Given an undirected weighted graph $G = (V, E, W)$, we define the normalized Laplacian matrix $\mathcal{L}_G = D^{-1/2} L_G D^{-1/2}$, where D is the degree matrix and L_G is the graph Laplacian.

(a) Show that $\mathcal{L}_G$ is symmetric and PSD (positive semi-definite).

(b) Show that all the eigenvalues of $\mathcal{L}_G$ are real numbers, between 0 and 2.

Problem 4.9 (•• | Tightness of the upper bound in Cheeger's inequality)**.** Let n be an even number greater than 2. Let C be a cycle graph on n vertices, labelled 1 to n. As usual, we set $w_{ij} = \mathbf{1}[\{i,j\} \in E]$, so that $W = A$.

(a) Prove that for every cut S, with $\emptyset \subsetneq S \subsetneq [n]$, its Cheeger's cut is lower bounded as

$$h(S) \geq \frac{2}{n}.$$

(b) Denote by $\lambda_2(C)$ the second smallest eigenvalue of the Laplacian of the graph C. Prove that

$$\lambda_2(C) \leq \frac{c}{n^2},$$

where c is an absolute constant.

(c) Conclude that the upper bound in Cheeger's inequality is tight up to an absolute constant.

Hint. *Consider the quadratic form $x^\top L_C x$ for the vector $x \in \mathbb{R}^n$ given by $x_i = \left|i - \frac{n}{2}\right| - \frac{n}{4}$.*

Problem 4.10 (•• | Tightness of the lower bound in Cheeger's inequality)**.** Let $d \geq 2$ be an integer, $G = (V, E)$ be the d-dimensional hypercube, and $\mathcal{L}_G$ its normalized Laplacian. We index the $n = 2^d$ vertices of the hypercube by d-dimensional $\{0,1\}$-vectors, i.e. $V = \{0,1\}^d$, and given any $x, y \in V$, we have $\{x, y\} \in E$ if and only if x and y differ in exactly one coordinate. The example for $d = 3$ is given in Figure 4.2.

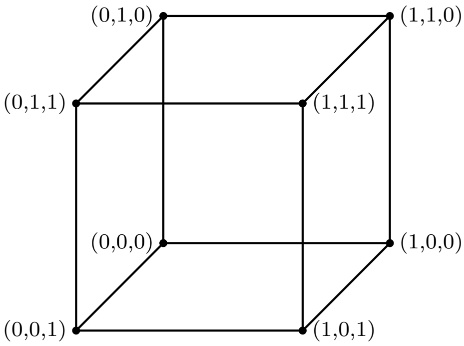

Figure 4.2: 3-dimensional hypercube.

Given a (possibly empty) subset $T \subseteq [d]$ let $v_T \in \mathbb{R}^n$ be a vector, whose coordinates are indexed by $n = 2^d$ vertices of the hypercube and defined by

$$v_T(x) = (-1)^{\sum_{i \in T} x_i},$$

where x_i is the i-th coordinate of the vertex $x \in \{0,1\}^d$. Also, let $S_T \subseteq V$ be the subset of vertices given by

$$S_T = \{x \in V \colon v_T(x) = 1\}.$$

[When $T = \emptyset$, we interpret the empty sum as zero, i.e. $\sum_{i \in \emptyset} x_i = 0$.]

(a) Compute $h(S_{\{1\}})$, the Cheeger's cut of the subset $S_{\{1\}}$.

(b) Show that for any $T \subseteq [d]$, v_T is an eigenvector of $\mathcal{L}_G$ with eigenvalue $\frac{2|T|}{d}$.

(c) Show that if $T' \subseteq [d]$ is distinct from T, then v_T and $v_{T'}$ are orthogonal.

(d) Conclude that for any $0 \leq k \leq d$, the eigenspace corresponding to the eigenvalue $\frac{2k}{d}$ has dimension $\binom{d}{k}$.

(e) Compute h_G, the Cheeger's constant of G.

5 Nonlinear Dimension Reduction and Diffusion Maps

Problem 5.1 (• | Random walks: eigenvalues)**.** Let $G = (V, E, W)$ be an undirected weighted graph.

(a) Suppose that $A, B \in \mathbb{R}^{n\times n}$ are similar matrices, which means that there exists an invertible matrix $P \in \mathbb{R}^{n\times n}$ such that $B = P^{-1}AP$. Prove that they have the same eigenvalues, with equal geometric multiplicities.

(b) Show that the transition probability matrix $M \coloneqq D^{-1}W$ is similar to the matrix $S \coloneqq D^{-1/2}WD^{-1/2}$.

(c) Deduce that all eigenvalues of M are real.

(d) Prove that every eigenvalue of M belongs to the interval $[-1, 1]$, and show that 1 is an eigenvalue.

Problem 5.2 (•• | Random walks: connectivity)**.** Let $G = (V, E, W)$ be an undirected weighted graph. Show that the largest eigenvalue of $M = D^{-1}W$ has multiplicity one if and only if the graph is connected. Here, two vertices are connected if and only if there exists a path from one to another along which all edges have positive weights.

Problem 5.3 (••• | Lazy walk)**.** Let $G = (V, E, W)$ be an undirected weighted graph.

(a) Let $M' = \frac{1}{2}(M + I)$ be the transition probability matrix of the associated lazy walk. Show that M' is not necessarily symmetric, but it is always positive semi-definite metrix.

(b) Prove that the lazy random walk is aperiodic, which means that for any vertex $i \in V$ there is no (period) integer $k > 1$ such that for any (time) $t \geq 1$:

$$\left((M')^t\right)_{ii} > 0 \implies k \mid t.$$

(c) Suppose that W is an irreducible matrix. Prove that there exists $T \in \mathbb{N}$ such that for every $t \geq T$, all entries of $(M')^t$ are positive. (This means that the associated Markov Chain is regular.)

Problem 5.4 (• | Equilibrium distribution)**.** Let $G = (V, E, W)$ be an undirected weighted graph. Suppose that there is an equilibrium distribution π on V, which means that starting from any $i \in V$:

$$\mathbb{P}\left(X(t) = j \mid X(0) = i\right) \to \pi_j \text{ as } t \to \infty,$$

holds for any $j \in V$. Prove that $\lambda_2(M) < 1$.

Hint. *This means that for any $i_1, i_2, j \in V$, we have* $\mathbb{P}(X(t) = j \mid X(0) = i_1) - \mathbb{P}(X(t) = j \mid X(0) = i_2) \to 0$.

Problem 5.5 (•• | Truncated diffusion map of a cycle graph)**.** Let C be a cycle graph of length n, where $n \geq 3$. Find the diffusion map truncated to 2 dimensions.

Problem 5.6 (•• | Diffusion map of a complete graph)**.** Let K_n be complete graph on n nodes, where $n \geq 3$. Find the diffusion map. (Since there are more bases of real-valued eigenvectors, it is sufficient to pick one.)

Problem 5.7 (•• | Diffusion map of a lazy walk)**.** Let $G = (V, E, W)$ be an undirected weighted graph with the associated transition probability matrix M. Let $\varphi_t \colon V \to \mathbb{R}^{n-1}$ be the diffusion map made from M. Now let $M' = \frac{1}{2}(M + I)$ be the transition probability matrix of the associated lazy walk, and $\varphi'_t \colon V \to \mathbb{R}^{n-1}$ be the diffusion map made from M'. Prove that for any $t \geq 1$:

$$\varphi'_t = 2^{-t}\sum_{u=0}^{t}\binom{t}{u}\varphi_u.$$

Problem 5.8 (•• | Hitting times and semi-supervised learning)**.** Let $G = (V, E, W)$ be an undirected, connected graph with non-negative weights w_{ij}. The vertex set is partitiond as $V = V_+ \cup V_- \cup V^\star$, where V_+ are labeled as 1, V_- are labeled as 0 and $V^\star$ are unlabeled. Suppose that every unlabelled vertex ($V_\star$) is connected to at least one labelled vertex ($V_+ \cup V_-$) by an edge.

To predict the label of the unlabeled vertices, you wish to find a function $f^\star : V \to \mathbb{R}$ which agrees on the labeled vertices and predicts the values of the unlabeled vertices as values in $\mathbb{R}$ as smoothly as possible:

$$f^\star \coloneqq \arg \min_{\substack{f:V\to\mathbb{R}:\\ f(i)=1, i\in V^+\\ f(i)=0, i\in V^-}} \sum_{i<j} w_{ij}(f(i) - f(j))^2. \tag{5.1}$$

Now, consider a random walk X on V given by the following transition probabilities:

$$\mathbb{P}\left(X(t+1)=j \mid X(t)=i\right)=\frac{w_{ij}}{\deg(i)}.$$

Given a node $i \in V$, let $g(i)$ be the probability that a random walker starting at i reaches a node in V_+ before reaching one in V_-. I.e. if $T_+ = \inf\{t \geq 0 \colon X(t) \in V_+\}$ and $T_- = \inf\{t \geq 0 \colon X(t) \in V_-\}$, then

$$g(i) \coloneqq \mathbb{P}\left(T_+ < T_- \mid X(0)=i\right).$$

(a) Show that for any $i \in V$ and $t \geq 0$, $\sum_{j\in V} \mathbb{P}\left(X(t+1)=j \mid X(t)=i\right)=1$, so that X is a random walk.

(b) Show that g satisfies constraints of the optimization problem: $g(i)=1$ for $i \in V_+$, and $g(i)=0$ for $i \in V_-$.

(c) Prove that g satisfies the following equality for any unlabelled $i \in V^\star$:

$$g(i)=\frac{1}{\deg(i)}\sum_{j\in V_+} w_{ij}+\frac{1}{\deg(i)}\sum_{j\in V^\star} w_{ij}g(j). \tag{5.2}$$

(d) By analyzing first-order optimality conditions (it is enough to state the formula from the notes, without proving) of the optimization problem (5.1), show that $f^\star$ also satisfies (5.2), and conclude that $f^\star = g$.

6 Linear Dimension Reduction via Random Projections

Problem 6.1 (•• | Johnson-Lindenstrauss Lemma: Alternative Version)**.** The goal of this exercise is to prove the random projection lemma and then to use this result to prove another version of the Johnson-Lindenstrauss lemma.

(a) Let P be the coordinate projection, which maps a vector in $\mathbb{R}^n$ onto its first m coordinates in $\mathbb{R}^m$. Let $z \in S^{n-1}$ be a random vector sampled uniformly on the sphere S^{n-1}. Show that

$$\mathbb{E}\|Pz\|_2^2 = \frac{m}{n}\|z\|_2^2.$$

(b) Prove the following statement using the result in Problem 8.5: There exists an absolute constant $c > 0$, such that for any $\varepsilon > 0$ with probability at least $1 - 2\exp(-c\varepsilon^2 m)$ it holds

$$(1-\varepsilon)\sqrt{\frac{m}{n}}\|z\|_2 \leq \|Pz\|_2 \leq (1+\varepsilon)\sqrt{\frac{m}{n}}\|z\|_2. \tag{6.1}$$

(c) Note that the result in (b) is stated for a random vector, while in dimension reduction we wish to find a randomized projection such that it preserves geometry for a fixed set of points. However, it can be shown that the same result (6.1) holds when $z \in \mathbb{R}^n$ is a fixed vector, and P is a orthogonal projection onto an m-dimensional subspace chosen uniformly at random from all m-dimensional subspaces in $\mathbb{R}^n$. In fact, these two models are equivalent.

You can use the mentioned fact **without proof**. Using (b), show the following result: Let $\mathcal{X}$ be a set of n points in $\mathbb{R}^n$ and let $\varepsilon > 0$. Suppose that

$$m \geq \frac{C}{\varepsilon^2}\log n.$$

Consider a random subspace E of dimension m chosen uniformly from all m-dimensional subspaces in $\mathbb{R}^n$, and let P be an orthogonal projection on this set. Then with probability at least $1 - 2\exp(-c\varepsilon^2 m)$, the scaled projection $Q = \sqrt{\frac{n}{m}}P$ is an ε-approximate isometry for $\mathcal{X}$, i.e., for all $x, y \in \mathcal{X}$,

$$(1-\varepsilon)\|x-y\| \leq \|Qx - Qy\| \leq (1+\varepsilon)\|x-y\|.$$

Here $C, c > 0$ are universal constants.

Problem 6.2 (••• | Optimality of the Johnson-Lindenstrauss Lemma)**.** Recall that Johnson-Lindenstrauss lemma states that the geometry of the data is well preserved when we choose a random subspace of dimension $m \asymp \log n$. In this problem, we will show that this dependency is optimal.

Find an example of a set of n points in $\mathbb{R}^n$, for which it is not possible construct an ε-isometry for $\varepsilon = 1/2025$ onto a subspace of dimension m such that $m/\log n \to 0$ as n goes to infinity. You are expected to use results from other Problem sections (for instance section 8).

7 Community Detection and the Power of Convex Relaxation

Problem 7.1 (•• | Random MaxCut and Boosting)**.** We consider the following naive (but surprisingly effective) procedure to find a large cut in a graph G with an even number of vertices: choose a set S of $n/2$ vertices uniformly at random in G. We want to show that the partition (S, S^c) cuts a large number of edges with some (small but positive) probability. Then we boost the procedure to increase the probability of finding a large cut.

(a) Show that for any fixed $\varepsilon \in (0, \frac{1}{2})$

$$\mathbb{P}\left(\text{cut}(S) > \left(\frac{1}{2} - \varepsilon\right)|E|\right) \geq \varepsilon,$$

where $|E|$ is the number of edges in graph G.

Hint. *For (a) it may be easier to study the set of edges that are NOT being cut.*

(b) The result of the previous subproblem is rather unsatisfying, since if we want to find a cut with $0.49|E|$ edges, the probability of success may be as low as 0.01. For this reason we will modify our procedure, namely to improve our accuracy we sample S several times.

Suppose we run the procedure k times and get sets $S_1, \dots, S_k$. We want to construct a set S^* from these outputs such that for any $\delta \in (0, 1)$,

$$\mathbb{P}\left(\text{cut}(S^*) > \left(\frac{1}{2} - \varepsilon\right)|E|\right) \geq 1 - \delta$$

Find such cut and give an estimate on the required number of trials $k(\delta, \epsilon)$ depending on probability parameter δ and approximation parameter ϵ.

Remark: observe that the same technique can be applied to other randomized algorithms as well.

Problem 7.2 (• | Dual SDP)**.** To find the solution to the community detection problem in SBM, in the course we introduce convex relaxation of the problem and subsequently use convex duality to certify the optimality. In this problem, we will find the dual problem using the Lagrangian function.

Recall the definition of a semidefinite program (SDP).

Definition. *A semidefinite program (SDP) is an optimization problem of the following type:*

$$\max_{X \in \mathbb{R}^{n\times n}} \langle A, X\rangle \quad \text{subject to } X \succeq 0, \langle B_i, X\rangle = b_i, \quad i = 1, \dots, m, \tag{7.1}$$

where $A, B_1, \dots, B_m \in \mathbb{R}^{n\times n}$ and $b_i \in \mathbb{R}$ are given.

In SDPs, one of the constraints is positive semidefiniteness of a matrix. This constraint can be incorporated in the Lagrangian function as follows:

$$\mathcal{L}(X, \nu, Y) = \langle A, X\rangle + \sum_{i=1}^{m} \nu_i (b_i - \langle B_i, X\rangle) + \langle Y, X\rangle,$$

where $Y \in \mathbb{R}^{n\times n}$ is positive semidefinite matrix, and $\nu \in \mathbb{R}^m$. Using this Lagrangian, we can easily check that

$$p^* = \max_{X} \min_{\substack{\nu, Y\\ Y \succeq 0}} \mathcal{L}(X, \nu, Y)$$

coincides with the optimal value of the original SDP (7.1).

(a) Using the expression for the Lagrangian function, find the dual function

$$g(Y, \nu) = \max_{X \in \mathbb{R}^{n\times n}} \mathcal{L}(X, \nu, Y)$$

defined for PSD matrices $Y \in \mathbb{R}^{n\times n}$ and $\nu \in R^m$ (note that the dual function may be infinite for certain values of Y). Then write the dual program of the SDP (7.1) (the dual program just minimizes the dual function and contains the constraints that prevent the dual function from being infinite).

(b) Using (a), find the dual of the following semidefinite program:

$$\begin{aligned} \max \quad & \text{tr}(BX) \\ \text{s.t.} \quad & X_{ii} = 1 \quad \text{for each } i \\ & X \succeq 0. \end{aligned}$$

Problem 7.3 (●●● | Connectedness of the Erdős-Rényi graph)**.** We define the Erdős-Rényi graph as a random graph $G \sim \mathcal{G}(n,p)$ with n vertices generated by placing each possible edge independently at random with probability p. The Erdős-Rényi graph is a popular model to study the performance of several optimization algorithms on graphs. Many of these algorithms rely on the graphs being connected and in this problem we study when this is the case for the Erdős-Rényi model.

We define $p := \frac{\lambda \log n}{n}$ for some constant $\lambda > 0$.

(a) Prove that if $\lambda \leq 1 - c$, where $c > 0$ is an absolute constant, then the graph G has an isolated vertex with probability $1 - o(1)$. (We use the standard asymptotic notation, $f(n) = o(1)$ if $\lim_{n\to\infty} f(n) = 0$.)

Hint. *For (a) consider the random variable that counts the number of isolated vertices.*

Hint. *Use Problem 2.5.*

(b) Now observe the following: A graph is disconnected if and only if there exits a set of k nodes such that $k \leq \left\lfloor \frac{n}{2} \right\rfloor$ and there is no edge connecting the set of k nodes with the complement set of $n-k$ nodes. Use this fact to prove that if $\lambda \geq 1 + c$ for an absolute constant $c > 0$, then the graph is connected with probability $1 - o(1)$.

Problem 7.4 (● | Sum of Squares Proof)**.** Let x, y be real numbers, prove

$$x^4 + y^4 + 4xy + 2 \geq 0.$$

Hint. *Add and subtract an appropriate monomial.*

Problem 7.5 (● | Smallest Eigenvalue Program)**.** Let $A \in \mathbb{R}^{d\times d}$ be a symmetric matrix. Prove that the following optimization problem has the smallest eigenvalue of A as optimal value:

$$\begin{aligned} \min \quad & \mathrm{Tr}(AX) \\ \text{s.t.} \quad & \mathrm{Tr}(X) = 1 \\ & X \succeq 0. \end{aligned}$$

(Recall that $X \succeq 0$ means that X is PSD.)

Problem 7.6 (●● | Discrepancy Relaxation)**.** Let $A \in \mathbb{R}^{d\times m}$ be a matrix. We define its discrepancy as the optimal value of the following minimization problem:

$$\mathrm{disc}(A) = \min_{\varepsilon \in \{-1,1\}^m} \|A\varepsilon\|_\infty .$$

The vector discrepancy of A is the minimal value of the problem

$$\mathrm{vecdisc}(A) = \min_{u_1,\ldots,u_m \in \mathbb{S}^{m-1}} \max_{1\leq i\leq d} \left\| \sum_{j=1}^m A_{i,j} u_j \right\|_2 ,$$

where $\mathbb{S}^{m-1} \subset \mathbb{R}^m$ denotes the euclidean unit sphere, so $\|u_i\|_2 = 1$ for all $1 \leq i \leq m$. The goal of this exercise is to show that vector discrepancy is a convex relaxation of discrepancy, which can be solved using a semidefinite program.

(a) Prove the inequality

$$\mathrm{vecdisc}(A)^2 \leq \mathrm{disc}(A)^2.$$

(b) Prove that the quantity $\mathrm{vecdisc}(A)^2$ is the optimal value of the following semidefinite program:

$$\begin{aligned} \min \quad & D \in \mathbb{R} \\ \text{s.t.} \quad & (AXA^\top)_{i,i} \leq D \quad \forall 1 \leq i \leq d \\ \text{and} \quad & X_{i,i} = 1 \quad \forall 1 \leq i \leq m \\ & X \succeq 0 \in \mathbb{R}^{m\times m}. \end{aligned}$$

Hint. *Use the square-root of the matrix X to construct your unit vectors.*

Problem 7.7 (•••|Minimum Bisection and Community Detection)**.** The goal of this exercise is to relate the minimum bisection problem with exact recovery in the community detection problem. Assume that n is even and consider the graph with n vertices drawn from the stochastic block model with two balanced communities, i.e, each community has size $n/2$, moreover, the two communities are chosen uniformly at random. Let p be the probability that an edge is placed inside the communities and q across the communities with $p > q$.

Our goal is to estimate the partition Ω induced by the communities with an estimator $\hat{\Omega}(G)$ that depends only on one sample of the random graph G. Prove that the estimator that minimizes the probability of error is equivalent to solve the minimum bisection of the observed graph G (the minimum bisection is a partition into two equally-sized subsets, such that the number of edges being cut by such a partition is minimal). The probability of error P_e is given by

$$P_e := \mathbb{P}(\hat{\Omega} \neq \Omega) = \sum_g \mathbb{P}(\hat{\Omega}(G) \neq \Omega | G = g)\mathbb{P}(G = g).$$

Here the sum is taken over all possible realizations of the random graph G.

Hint. *Use Bayes' rule to simplify and note that you can ignore terms that do not depend on $\hat{\Omega}$.*

Problem 7.8 (•|PSD Set Convexity)**.** Show that the set $S_n^+ = \{A \in \mathbb{R}^{n\times n} : A \succeq 0\}$ is convex and that it is invariant under multiplication with a positive scalar.

Problem 7.9 (••|Spectral Algorithm for Planted Clique)**.** We want to analyze parts of a spectral algorithm, which is used to find the largest clique in a graph G on n vertices. This algorithm is often analyzed for the planted clique model, where $G \sim \mathcal{G}(n, 1/2)$ is random Erdős-Rényi graph and then k vertices of G are randomly uniformly selected and then edges will be added to G until these k vertices become a clique (fully connected amongst each other). We call the graph we get after this procedure $\tilde{G}$. The goal is to find this planted clique with the so called "AKS spectral algorithm", which relies on computing the top eigenvector of the matrix $M := A - \frac{1}{2}\mathbf{1}_n\mathbf{1}_n^\top$, where A is the adjacency matrix of $\tilde{G}$ and $\mathbf{1}_n \in \mathbb{R}^n$ is the all-ones vector. The idea behind this algorithm is that the matrix M is typically close to the matrix $\frac{1}{2}\mathbf{1}_S\mathbf{1}_S^T$, where $\mathbf{1}_S$ is the indicator vector of the planted clique S, and if the matrices are close, then their top eigenvectors should in some sense also be close. This is the part that you will prove in this exercise:

(a) Let $0 < \varepsilon < 1$, and suppose there exists a symmetric matrix $M \in \mathbb{R}^{n\times n}$ and a subset $S \subset [n]$ with the property $|S| > 2(1 + \varepsilon^{-1}) \left\| M - \frac{1}{2}\mathbf{1}_S\mathbf{1}_S^\top \right\|$, where $\mathbf{1}_S \in \mathbb{R}^n$ is the indicator vector of S (so 1 whenever the coordinate is in S and 0 otherwise). If v is an eigenvector corresponding to the largest eigenvalue of M with norm $\|v\|_2^2 = |S|$, prove that the following inequality holds:

$$\min\{\|v - \mathbf{1}_S\|_2^2, \|-v - \mathbf{1}_S\|_2^2\} \leq 2|S|\varepsilon^2.$$

You may use the following Theorem without proof:

Theorem. *Let $M \in \mathbb{R}^{n\times n}$ be a symmetric matrix and let v be an eigenvector corresponding to the largest eigenvalue of M. Let $y \in \mathbb{R}^n$ be any vector and let θ be the angle between y and v, then*

$$|\sin(\theta)| \leq \frac{\left\| M - yy^\top \right\|}{\|yy^\top\| - \|M - yy^\top\|}$$

Hint. *Use the fact that the dot product of $\mathbf{1}_S$ and v depends on the angle between these vectors.*

Problem 7.10 (••|Little Grothendieck Problem)**.** Let $C \succeq 0$ ($C \in \mathbb{R}^{n\times n}$ is positive semidefinite). In this problem you will show an approximation ratio of $\frac{2}{\pi}$ to the problem

$$\max_{x_i = \pm 1} \sum_{i,j=1}^n C_{ij}x_ix_j.$$

Similarly to `Max-Cut`, we consider

$$\max_{\substack{v_i \in \mathbb{R}^n \\ \|v_i\|^2 = 1}} \sum_{i,j=1}^n C_{ij}v_i^Tv_j.$$

The goal is to show that, for $g \sim \mathcal{N}(0, I_{n\times n})$, taking $x_i = \operatorname{sign}(v_i^Tg)$ a randomized rounding,

$$\mathbb{E}\left[\sum_{i,j=1}^n C_{ij}x_ix_j\right] \geq \frac{2}{\pi}\sum_{i,j=1}^n C_{ij}v_i^Tv_j \tag{7.2}$$

The difficulty lies in the fact that $\mathbb{E}[x_ix_j]$ is not easy to compute, which is why we divide this exercise into two parts.

(a) Compute the quantity $\mathbb{E}[\text{sign}(v_i^T g)\langle v_j, g\rangle]$.

Hint. *This number should only depend on the inner product between v_j and v_i.*

(b) Define the matrix $S \in \mathbb{R}^{n \times n}$ with entries $S_{i,j} = (\langle v_i, g\rangle - \sqrt{\pi/2}\,\text{sign}(v_i^T g))(\langle v_j, g\rangle - \sqrt{\pi/2}\,\text{sign}(v_j^T g))$. Show that

$$\text{Tr}(CS) \geq 0$$

holds, and use this fact to prove the inequality (7.2).

8 Concentration of Measure and Gaussian Analysis

Problem 8.1 (• | Moments of Gaussians)**.** Let Z be a standard gaussian random variable. Recall that Gaussian integration by parts states the following: given any differentiable function $f\colon \mathbb{R} \to \mathbb{R}$ whose derivative is absolutely integrable with respect to the standard normal measure, we have

$$\mathbb{E}\left[Zf(Z)\right] = \mathbb{E}\left[f'(Z)\right].$$

Let $p \geq 1$ be an integer and Z be a standard gaussian random variable. Show that

$$\mathbb{E}\left[Z^p\right] = \begin{cases} (p-1)!! & \text{if } p \text{ is even;} \\ 0 & \text{if } p \text{ is odd.} \end{cases}$$

Here, !! denotes the double factorial, defined as $n!! = n \cdot (n-2) \cdots 3 \cdot 1$ for an odd natural number n.

Note. *There are $(p-1)!!$ possible pairings of p elements (when p even), and this is not a coincidence!*

Problem 8.2 (•• | Maximum of Gaussians)**.** Let $g_1, \dots, g_d$ be a collection of (not necessarily independent) Gaussian random variables with zero mean and variance σ^2.

(a) Prove that the following bound holds

$$\mathbb{E} \max_{i=1,\dots,d} g_i \leq \sigma\sqrt{2\log d}.$$

(If you do not manage to prove the inequality with sharp constant 2 in the square root, you can replace it by an absolute constant $C > 0$.)

(b) Prove that the bound in the previous item is sharp up to an absolute constant if we assume that all the Gaussian random variables are independent, i.e. there exists a universal constant $c > 0$ such that

$$\mathbb{E} \max_{i=1,\dots,d} g_i \geq c\,\sigma\sqrt{\log d}.$$

(c) Show that the conclusion of (b) is false if we drop the assumption that $g_1, \dots, g_d$ are independent.

Problem 8.3 (••• | Application of Slepian's lemma)**.** Let W be a $d \times d$ Gaussian Wigner matrix [BSS25, §3.3.2].

(a) Prove that

$$\mathbb{E} \sup_{v \in S^{d-1}} \langle g, v\rangle = \mathbb{E}\|g\|_2 \leq \sqrt{d},$$

where $g \in \mathbb{R}^d$ is a standard Gaussian random vector.

(b) Apply Slepian's lemma to prove that

$$\mathbb{E}\lambda_{\max}(W) \leq 2\sqrt{d}.$$

(c) Show that the upper bound above is tight up to an absolute constant.

Hint. *For (b), consider the stochastic process $Y_v := 2\langle g, v\rangle$. For (c), first show $\mathbb{E}\|g\|_2 \leq c\sqrt{d}$ by integration.*

Lemma (Gamma Function Bound)**.** *Let $x \geq \frac{1}{2}$, we have*

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt \leq 3x^x.$$

Problem 8.4 (• | Moments of Subgaussian Variables)**.** Let Y be a σ^2-subgaussian random variable, so for all $t \geq 0$ it holds

$$\mathbb{P}(|Y| \geq \sigma t) \leq 2e^{-\frac{t^2}{2}}.$$

Prove that for all $p \geq 1$ we have

$$\mathbb{E}[|Y|^p]^{\frac{1}{p}} \leq C\sigma\sqrt{p},$$

where $C > 0$ is a universal constant.

Hint. *Use Problem 2.2(a) and bounds on the gamma function.*

Theorem (Gaussian Lipschitz concentration)**.** *Let $g_1, \ldots, g_n$ be i.i.d standard Gaussian random variables. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a L-Lipschitz function with respect to the Euclidean norm. Then, for all $t > 0$,*

$$\mathbb{P}(|f(g_1, \ldots, g_n) - \mathbb{E}f(g_1, \ldots, g_n)| \geq Lt) \leq 2e^{-t^2/2}. \tag{8.1}$$

Theorem (Lipschitz Concentration on the Sphere)**.** *Let $f : \sqrt{n}S^{n-1} \to \mathbb{R}$ be a L-Lipschitz function and let X be a random vector uniformly distributed on the sphere $\sqrt{n}S^{n-1}$. Then, for all $t > 0$,*

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| \geq Lt) \leq 2e^{-ct^2}. \tag{8.2}$$

Here $c > 0$ is an absolute constant.

Problem 8.5 ($\bullet\bullet$ | Lipschitz Concentration around L_p norms)**.** In this problem we will extend the Lipschitz concentration on the sphere to concentration around L_p norms for $p \geq 1$, i.e., we will prove that under the same assumptions as in (8.2) and additionally assuming that f is non-negative,

$$\mathbb{P}(|f(X) - \|f(X)\|_{L_p}| \geq Lt) \leq 2e^{-c_p t^2}, \tag{8.3}$$

where $c_p > 0$ is a constant only depending on p and $\|Z\|_{L_p} := (\mathbb{E}|Z|^p)^{1/p}$ for a random variable Z such that its p-th absolute moment is well-defined.

We will split the proof into several steps.

(a) Let Y be an L^2-subgaussian random variable in the sense of Problem 8.4. Prove that for any $A \geq 0$ there exists a constant c_A only depending on A, such that

$$\mathbb{P}(|Y - LA| \geq Lt) \leq 2e^{-c_A t^2}$$

holds for some constant $c_A > 0$ only depending on A.

(b) Show that for any non-negative random variable Z,

$$|\mathbb{E}Z - (\mathbb{E}Z^p)^{1/p}| \leq (\mathbb{E}\,|Z - \mathbb{E}Z|^p)^{1/p}.$$

You can assume that all the moments are well-defined. Use Problem 8.4 and (8.2) to conclude

$$|\mathbb{E}[f(X)] - \mathbb{E}[f(X)^p]^{1/p}| \leq CL\sqrt{p}$$

for some universal constant $C > 0$.

(c) Now we have all the ingredients to prove the theorem. Using (a), (b), and (8.2), complete the proof of inequality (8.3).

Problem 8.6 ($\bullet\bullet\bullet$ | Duality and Covering Numbers)**.** We will present the principle of duality in a completely different context. Suppose you have a set $T \subseteq \mathbb{R}^n$ and some number $\varepsilon > 0$. We call a subset $S \subseteq T$ an ε-covering of T, if for every $t \in T$ there exists an $s \in S$, such that $\|s - t\|_2 \leq \varepsilon$. We call a subset $S \subseteq T$ an ε-separated set, if for every $s \neq s' \in S$ we have $\|s - t\|_2 > \varepsilon$. We define the following optimal values:

$$\mathcal{N}(T, \varepsilon) := \min_{\substack{S \subseteq T \\ S\,\varepsilon-\text{covering of } T}} |S| \qquad \mathcal{D}(T, \varepsilon) := \max_{\substack{S \subseteq T \\ S\,\varepsilon-\text{separated}}} |S|$$

(a) Show that these two optimization problems are duals in the following sense:

$$\mathcal{N}(T, \varepsilon) \leq \mathcal{D}(T, \varepsilon) \leq \mathcal{N}(T, \varepsilon/2)$$

(b) Use part (a) to show that for the euclidean ball in d-dimensions one has the following covering number estimates for every $0 < \varepsilon < 1$:

$$\left(\frac{1}{\varepsilon}\right)^d \leq \mathcal{N}(\mathbb{B}_2^d, \varepsilon) \leq \left(\frac{3}{\varepsilon}\right)^d$$

Problem 8.7 ($\bullet$ | Norm of a Gaussian Vector)**.** Given a standard Gaussian vector $g \in \mathbb{R}^d$, we saw in Problem 8.3 that $\mathbb{E}\|g\|_2 \leq \sqrt{d}$. The goal of this exercise is to give a simple proof that this is sharp using Gaussian Lipschitz concentration inequality (8.1).

(a) Prove that the variance of $\|g\|_2$ is at most an absolute constant.

(b) Show that $\mathbb{E}\|g\|_2/\sqrt{d}$ converges to one as d goes to infinity.

Problem 8.8 (● | Gaussian Width of the Simplex)**.** The gaussian width of a set $S \subseteq \mathbb{R}^d$ is defined as

$$\omega(S) := \mathbb{E}\left[\sup_{s\in S} \langle s, g\rangle\right],$$

where $g \in \mathbb{R}^d$ is a standard gaussian vector, so all coordinates of g are independent and $\mathcal{N}(0,1)$-distributed. Consider the $d-1$-dimensional simplex

$$S_d := \left\{ x \in \mathbb{R}^d \,|\, 0 \leq x_i \leq 1,\, \sum_{i=1}^{d} x_i = 1 \right\}$$

Our goal is to show that there exist universal constants $c, C > 0$, such that

$$c\sqrt{\log(d)} \leq \omega(S_d) \leq C\sqrt{\log(d)}.$$

(a) For any subset $T \subseteq \mathbb{R}^d$ we define its convex hull $\operatorname{conv}(T)$ as follows:

$$\operatorname{conv}(T) = \left\{ \sum_{i=1}^{k} x_i t_i \,|\, k \in \mathbb{N}_{>0},\, x_i \in S_k,\, t_i \in T \right\}$$

Prove the equality

$$\omega(\operatorname{conv}(T)) = \omega(T).$$

(b) Find a finite set $T \subseteq \mathbb{R}^d$, such that $S_d = \operatorname{conv}(T)$. Use the result of another problem in this section to deduce the desired inequalities.

9 Matrix Concentration Inequalities

Problem 9.1 (• | Constructing Wigner)**.** Let Z be a $d \times d$ random matrix whose entries are all independent standard gaussians (in total d^2 of them). Show that $\frac{1}{\sqrt{2}}\left(Z + Z^\top\right)$ is a Gaussian Wigner matrix [BSS25, §3.3.2].

Problem 9.2 (• | Computing the σ parameter)**.** Let W be a $d \times d$ Gaussian Wigner matrix and let D be a $d \times d$ diagonal matrix with independent standard gaussians on the diagonal. Show that $\left\|\mathbb{E}W^2\right\|^{\frac{1}{2}} = \sigma(W) = \sqrt{d+1}$ and $\left\|\mathbb{E}D^2\right\|^{\frac{1}{2}} = \sigma(D) = 1$, and upper bound $\mathbb{E}\|W\|$ and $\mathbb{E}\|D\|$ using the Non-commutative Khintchine inequality.

Problem 9.3 (• • | Intrinsically free Non-commutative Khintchine Inequality)**.** In fact, a stronger version of the Non-commutative Khintchine inequality is known. Let $A_1, \dots, A_n \in \mathbb{R}^{d\times d}$ by symmetric matrices and $g_1, \dots, g_n \in \mathcal{N}(0,1)$ i.i.d. The gaussian series $X = \sum_{t=1}^n g_t A_t$ satisfies

$$\mathbb{E}\|X\| \le 2\sigma + C\, v\,(\log d)^{\frac{3}{2}},$$

where $C > 0$ is an absolute constant, $\sigma^2 = \left\|\sum_{i=1}^n A_i^2\right\|$, and v is given by

$$v^2 = \|\mathrm{Cov}(X)\|.$$

Here, matrix covariance $\mathrm{Cov}(X)$ is a $d^2 \times d^2$ matrix, whose row and column coordinates are indexed by pairs of indices, and entries are given by

$$\mathrm{Cov}(X)_{ij,kl} = \mathbb{E}\left[X_{ij}X_{kl}\right] \quad \text{for } i,j,k,l \in [d].$$

Compute $\mathrm{Cov}(W)$ and $\mathrm{Cov}(D)$, where W and D are as in Problem 9.2, and deduce using the intrinsically free Non-commutative Khintchine Inequality that there is an absolute constant $C' > 0$, such that

$$\mathbb{E}\|W\| \le C'\sqrt{d}.$$

Hint. *As a start, show that* $\mathrm{Cov}(X)_{ij,kl} = \sum_{t=1}^n [A_t]_{ij}[A_t]_{kl}$.

Problem 9.4 (• | Hermitian dilation)**.** In order to extend the matrix Bernstein inequality in the book [BSS25, Theorem 9.13] from symmetric to general rectangular matrices, we will use Hermitian dilation. For a matrix $S \in \mathbb{R}^{d_1\times d_2}$, the Hermitian dilation $\mathcal{H}(S)$ is defined as

$$\mathcal{H}(S) := \begin{pmatrix} 0_{d_1\times d_1} & S \\ S^\top & 0_{d_2\times d_2}\end{pmatrix} \in \mathbb{R}^{(d_1+d_2)\times(d_1+d_2)}.$$

In Problem 3.3(c) we showed that $\mathcal{H}(S)$ is symmetric and that $\|\mathcal{H}(S)\| = \|S\|$.

Let $S_1, \dots, S_n \in \mathbb{R}^{d_1\times d_2}$ be random rectangular matrices satisfying $\mathbb{E}[S_i] = 0$ for every $i \in [n]$. Show that

$$\mathbb{E}\left\|\sum_{i=1}^n S_i\right\| \le \sqrt{C(d)}\,\sigma + C(d)\,L,$$

where $d = d_1 + d_2$, $C(d) = 4 + 8\lceil \log d\rceil$ and

$$\sigma^2 = \max\left\{\left\|\sum_{i=1}^n \mathbb{E}\left[S_i S_i^\top\right]\right\|, \left\|\sum_{i=1}^n \mathbb{E}\left[S_i^\top S_i\right]\right\|\right\}, \qquad L^2 = \mathbb{E}\max_i \|S_i\|^2.$$

Problem 9.5 (• • • | Bernstein's inequality - expectation bound)**.** Let $\{X_k\}_{k=1}^n$ be a sequence of independent random symmetric $d \times d$ matrices. Assume that each X_k satisfies:

$$\mathbb{E}X_k = 0 \text{ and } \|X_k\| \le R \text{ almost surely.}$$

In this exercise, the goal is to show that

$$\mathbb{E}\left\|\sum_{k=1}^n X_k\right\| \le C\left(\sigma\sqrt{\log(d+1)} + R\log(d+1)\right). \tag{9.1}$$

To get a bound on expectation from the tail bound in the book [BSS25, Theorem 7.9], we will use an integral identity for the expectation (see Problem 2.2(a)).

(a) Show that for any $a, b > 0$

$$\exp\left(-\frac{2}{a+b}\right) \le \max\left\{\exp\left(-\frac{1}{a}\right), \exp\left(-\frac{1}{b}\right)\right\} \le \exp\left(-\frac{1}{a}\right) + \exp\left(-\frac{1}{b}\right).$$

Apply it to the exponent in the right-hand side of matrix Bernstein's inequality so that you get integrals that are easier to compute.

(b) For very small t, the tail bound is loose since the exponent is close to 1. Think how we can isolate the case of t close to 0.

(c) Once you split the integral, it remains only to compute the individual parts and choose the right constants in your argument. To find it, you can closely examine the final bound (9.1).

Problem 9.6 ($\bullet\bullet$ | Randomized matrix multiplication)**.** Let $A \in \mathbb{R}^{n\times m}$ be a real-valued matrix with unit spectral norm $\|A\| = 1$. The cost of computing $AA^\top$, using the standard matrix multiplication method, is of the order of $n^2 m$, which can be prohibitive when n and m are very large. In some cases, it is sufficient to obtain only an approximate solution which allows us to reduce the costs significantly. In this problem you will show that by using randomness we can get an approximation of the product more efficiently.

Denote $a_1, \dots, a_m \in \mathbb{R}^n$ columns of matrix A. Define a random matrix X such that $\mathbb{P}\left(X = m \cdot a_k a_k^\top\right) = \frac{1}{m}$.

(a) Suppose we draw s independent copies of X, denoted by $X_1, \dots, X_s$, and then average them $\hat{X}_s = \frac{1}{s}\sum_{k=1}^s X_k$. Prove that $\hat{X}_s$ in an unbiased estimator of $AA^\top$, meaning that $\mathbb{E}\hat{X}_s = AA^\top$.

(b) Define the coherence statistic as $\mu(A) = m \cdot \max\limits_{k=1,\dots,m} \|a_k\|^2$. Show that:

$$\max_{k=1,\dots,s} \left\|\mathbb{E}\left[(X_k - \mathbb{E}X_k)^2\right]\right\| \le 2\mu(A)$$

and

$$\max_{k=1,\dots,s} \|X_k - \mathbb{E}X_k\| \le 2\mu(A), \text{ almost surely.}$$

(c) Use Problem 9.5 (proof not needed) to show that, for an absolute constant C, if the number of samples s satisfies

$$s \ge C \max\left\{\frac{1}{\varepsilon}, \frac{1}{\varepsilon^2}\right\} \mu \log n$$

then the procedure achieves ε-accuracy, i.e., $\mathbb{E}\left\|\hat{X}_s - AA^\top\right\| \le \varepsilon$.

Problem 9.7 ($\bullet$ | Commuting vs. simultaneously diagonalizable)**.** Let $A, B \in \mathbb{R}^{d\times d}$ be two symmetric matrices.

(a) If A and B are simultaneously diagonalizable, show that they commute.

(b) If A and B commute, and B has all eigenvalues distinct, show that they are simultaneously diagonalizable.

(c) Generalize (a) and (b) for n symmetric matrices $A_1, \dots, A_n \in \mathbb{R}^{d\times d}$.

Hint. *For (b) show that if $Bv = \lambda v$ for some $v \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ then $B(Av) = \lambda(Av)$.*

Problem 9.8 ($\bullet\bullet$ | NCK for commuting matrices)**.** In the book we discussed the role of commutativity of the matrices for the upper bound of the expected value of the spectral norm of a random matrix. Recall that for $X := \sum_{i=1}^n g_i A_i$, where $A_1, \dots, A_n \in \mathbb{R}^{d\times d}$ are symmetric matrices and $g_1, \dots, g_n \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$, it holds

$$\sigma \lesssim \mathbb{E}\|X\| \lesssim \sigma\sqrt{\log d},$$

where $\sigma^2 := \left\|\sum_{i=1}^n A_i^2\right\|$.

(a) Suppose that $A_1, \dots, A_n \in \mathbb{R}^{d\times d}$ are symmetric commuting matrices. This means that they are simultaneously diagonalizable (well known fact, no proof needed), so there is an orthogonal matrix $Q \in \mathbb{R}^{d\times d}$ so that $D_i := QA_iQ^{-1}$ is a diagonal matrix for any $i \in [n]$. Let $\lambda_1^{(i)}, \dots, \lambda_d^{(i)}$ be the entries that appear, in that order, on the diagonal of D_i. Show that

$$\mathbb{E}\|X\| = \mathbb{E}\max_{k=1,\dots,d}\left|\sum_{i=1}^n g_i \lambda_k^{(i)}\right|.$$

(b) Deduce using Problem 8.2(a) that

$$\mathbb{E}\,\|X\| \lesssim \sigma\sqrt{\log d}.$$

(c) Find an example of commuting matrices $A_1, \dots, A_n$ such that Problem 8.2(b) implies

$$\mathbb{E}\,\|X\| \gtrsim \sigma\sqrt{\log d}.$$

Problem 9.9 ($\bullet\bullet$ | Trace Commutativity Inequality)**.** The goal of this exercise is to show the following key inequality that proves why commuting matrices perform worse than non-commuting matrices in trace moment estimations. One can actually follow a slightly different approach than the one shown in the book if one follows the hint in (b).
Let $X, A \in \mathbb{R}^{d\times d}$ be symmetric matrices and let k, l be nonnegative integers with $k + l$ being even, then

$$\mathrm{Tr}(AX^kAX^l) \leq \mathrm{Tr}(A^2X^{k+l}).$$

(a) Let $X = \sum_{i=1}^d \lambda_i u_i u_i^\top$ be the the eigenvalue decomposition of X, prove

$$\mathrm{Tr}(AX^kAX^l) \leq \sum_{i,j=1}^d |\lambda_i|^k\, |\lambda_j|^l\, (u_i^\top A u_j)^2.$$

(b) Finish the proof by showing

$$\sum_{i,j=1}^d |\lambda_i|^k\, |\lambda_j|^l\, (u_i^\top A u_j)^2 \leq \mathrm{Tr}(A^2X^{k+l}).$$

Hint. *Use Young's Inequality on the product of eigenvalues.*

10 Compressive Sensing and Sparsity

Problem 10.1 (•• | Sparse vector approximation in ℓ^2)**.** Let $N \geq s \geq 1$ be integers and $x \in \mathbb{C}^N$ be a vector. Show that there exists an s-sparse vector $y \in \mathbb{C}^N$ such that

$$\|x - y\|_2 \leq \frac{1}{2\sqrt{s}} \|x\|_1 .$$

Problem 10.2 (• | ℓ_0-Minimization Recovery)**.** Let $A \in \mathbb{C}^{m\times p}$ be a matrix. Suppose that every s-sparse vector x can be uniquely recovered by A via $\|.\|_0$ minimization, i.e, we choose x^* that minimizes $\|z\|_0$ subject to the constraint $Ax = Az$ and x is the unique minimum of this problem if x has at most s nonzero entries. Here $\|.\|_0$ is the ℓ_0 "norm", it counts the number of nonzero entries of the input vector.

(a) Prove that every $2s$ columns of A are linearly independent.

(b) Prove that $m \geq 2s$.

(c) Prove that if a matrix $B \in \mathbb{C}^{m\times p}$ satisfies the condition that every $2s$ columns are linearly independent, then every s-sparse vector x can be uniquely recovered by B via $\|.\|_0$ minimization.

Problem 10.3 (•• | Stable Nullspace Property)**.** A fundamental fact in compressed sensing is that in order to recover an s-sparse vector $x \in \mathbb{R}^N$ by minimizing the ℓ_1 norm, the measurement matrix $\Phi \in \mathbb{R}^{d\times N}$ needs to satisfy the null space property: For every non-trivial vector v in the kernel of Φ and all sets S such that $|S| \leq s$, it holds that $\|v_S\|_1 < \|v_{S^c}\|_1$ [BSS25, Definition 10.3]. Here v_S denotes the vector in $\mathbb{R}^{|S|}$ corresponding to the restriction of v to the index set S. The goal of this exercise is to study the compressed sensing problem when x is approximately sparse.

We say that a matrix $\Phi \in \mathbb{R}^{d\times N}$ satisfies *the* (s, ρ)*-stable null space property* if for every non-zero vector $v \in \ker(\Phi)$ and all sets S such that $|S| \leq s$, the following holds

$$\|v_S\|_1 \leq \rho \|v_{S^c}\|_1.$$

Prove the following facts

(a) Given a set $S \subset \{1, \dots, N\}$ and vectors $x, z \in \mathbb{R}^N$,

$$\|(x - z)_{S^c}\|_1 \leq \|z\|_1 - \|x\|_1 + 2\|x_{S^c}\|_1 + \|(x - z)_S\|_1.$$

(b) Prove that if $\Phi \in \mathbb{R}^{d\times N}$ satisfies the (s, ρ)-stable nullspace property with $\rho \in (0, 1)$, then the solution of the optimization program

$$\hat{x} := \operatorname{argmin} \|z\|_1 \quad \text{subject to} \quad \Phi z = \Phi x$$

satisfies

$$\|\hat{x} - x\|_1 \leq 2\sigma_s(x)\frac{1 + \rho}{1 - \rho},$$

where $\sigma_s(x)$ is the s-best term approximation error of x, given by $\sigma_s(x) := \inf_{z:\|z\|_0\leq s} \|x - z\|_1$.

(c) Show that the stable nullspace property with $\rho < 1$ is sufficient for exact recovery when the vector x is s-sparse.