

Mathematics of Data Science, Fall 2025 - Homework 3

Due at 11:49PM on Sunday Oct/19 (Gradescope)

Your submitted solutions to assignments should be your own work. You are allowed to discuss homework problems with other students, but should carry out the execution of any thoughts/directions discussed independently, on your own. Acknowledge any source you consult. **Do not use any type of Large Language Model, e.g., ChatGPT, to blindly answer this assignment. If you do, your submission will be voided and you will get zero as a grade.**

Problem 1 - Fun with eigenvalues

Let $A \in \mathbf{R}^{n \times m}$ be a matrix with $n \leq m$. Prove the following.

- (a) The eigenvalues of AA^\top correspond to $\sigma_1(A)^2, \dots, \sigma_n(A)^2$ and similarly, $A^\top A$ has eigenvalues $\sigma_1(A)^2, \dots, \sigma_n(A)^2$ together with $m - n$ copies of the eigenvalue zero. Based on this observation give an alternate proof of the singular value decomposition theorem from Lecture 9 using the spectral decomposition theorem.
- (b) Just for this question suppose further that $A \in \mathcal{S}^n$ (i.e., the matrix is symmetric $n \times n$). Somebody lands you the eigenvalue decomposition $A = U\Lambda U^\top$ ($U \in O(n)$ and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$). Provide an algorithm, and a justification of its correctness, to compute the singular value decomposition of A from its eigenvalue decomposition.
- (c) Establish the Courant–Fischer min–max formula

$$\sigma_i(A) = \sup_{\substack{V \subset \mathbb{R}^m \\ \dim V = i}} \inf_{\substack{v \in V \\ \|v\|_2 = 1}} \|Av\|_2, \quad (1)$$

for all $1 \leq i \leq n$, where the supremum ranges over all i -dimensional subspaces V of \mathbb{R}^m .

- (d) Use (c) to show the next two inequalities.

1. For all $1 \leq i, j$ with $i + j - 1 \leq n$,

$$\sigma_{i+j-1}(A + B) \leq \sigma_i(A) + \sigma_j(B).$$

2. For all $1 \leq i \leq n$,

$$|\sigma_i(A + B) - \sigma_i(A)| \leq \|B\|_{\text{op}}.$$

Problem 2 - Operating at large scale

Let $A \in \mathbf{R}^{n \times m}$ be a matrix with $n \leq m$. Suppose we wanted to compute its operator norm. One natural strategy would be to first find the singular value decomposition of A and then extract the top singular value. However, this strategy is often too expensive for large matrices. Instead, here are two computationally friendly ways to approximate the operator norm of a matrix. Consider a random vector $x \sim \mathcal{N}(0, I_m)$.

(a) Show that, with probability one,

$$\lim_{k \rightarrow \infty} \|(A^\top A)^k x\|_2^{1/k} = \|A\|_{\text{op}}^2.$$

(b) Consider an iterative method that starts at $x_0 = x$, and at each iteration updates

$$x_{k+1} \leftarrow \frac{A^\top A x_k}{\|A^\top A x_k\|_2}.$$

Show that with probability one, we have $\lim_{k \rightarrow \infty} \|A^\top A x_k\|_2 = \|A\|_{\text{op}}^2$.

(c) Generate $A \in \mathbf{R}^{n \times m}$ a random matrix with i.i.d. entries where $A_{11} \sim N(0, 1)$ with $n = 100$ and $m = 200$. Compute using $\sigma_1(A)$ an SVD decomposition and plot the errors of both methods as a function of k . That is, plot $\left| \|(A^\top A)^k x\|_2^{1/k} - \sigma_1(A)^2 \right|$ and $\left| \|A^\top A x_k\|_2 - \sigma_1(A)^2 \right|$ versus the number of iterations k .

Problem 3 - Yet another variant of Davis-Kahan $\sin \theta$

Here is yet another variant that is often useful. Let $\|\cdot\|_F$ be the Frobenius norm.

Theorem 1. Let $M, M^* \in \mathbf{R}^{n \times n}$ be symmetric with eigenvalues $\lambda_1 \geq \dots \geq \lambda_n$ and $\lambda_1^* \geq \dots \geq \lambda_n^*$, respectively, and let $E = M - M^*$. Fix integers $1 \leq r \leq s \leq n$ and assume

$$\Delta := \min\{\lambda_{r-1}^* - \lambda_r^*, \lambda_s^* - \lambda_{s+1}^*\} > 0, \quad \text{where by convention } \lambda_0^* := +\infty, \text{ and } \lambda_{n+1}^* := -\infty.$$

Let $d := s - r + 1$, and let

$$U := [u_r, \dots, u_s] \in \mathbf{R}^{n \times d}, \quad U^* := [u_r^*, \dots, u_s^*] \in \mathbf{R}^{n \times d}, \quad \Lambda = (\lambda_r, \dots, \lambda_s), \quad \Lambda^* = (\lambda_r^*, \dots, \lambda_s^*),$$

where $M u_j = \lambda_j u_j$ and $M^* u_j^* = \lambda_j^* u_j^*$ for $j = r, \dots, s$. Further, let $U_\perp \in \mathbf{R}^{n \times (n-d)}$ and $U_\perp^* \in \mathbf{R}^{n \times (n-d)}$ be such that $[U, U_\perp] \in O(n)$ and $[U^*, U_\perp^*] \in O(n)$. Then,

$$\|U_\perp^\top U^*\|_F \leq \frac{2\|E\|_F}{\Delta}.$$

The proof is somewhat similar to the one in class, and you will develop it in this exercise.

(a) Show that $\|U \Lambda^* - M^* U\|_F \leq \|E U\|_F + \|U(\Lambda - \Lambda^*)\|_F$ and use this inequality to conclude $\|U \Lambda^* - M^* U\|_F \leq 2\|E\|_F$.

(b) Prove that $\|U \Lambda^* - M^* U\|_F \geq \|U_\perp^{*\top} U \Lambda^* - \Lambda_\perp^* U_\perp^{*\top} U\|_F$. **Hint:** Pythagoras always comes to the rescue.

(c) We will use a simple linear algebra fact that requires some additional notation. Let $A \in \mathbf{R}^{n \times m}$ and $B \in \mathbf{R}^{p \times q}$, then their Kronecker product is a matrix $(A \otimes B) \in \mathbf{R}^{pn \times qm}$ defined by blocks via

$$A \otimes B = \begin{bmatrix} A_{11}B & \cdots & A_{1m}B \\ \vdots & \ddots & \vdots \\ A_{n1}B & \cdots & A_{nm}B \end{bmatrix}.$$

Fact 1. *The map that sends a matrix W to the matrix product $W \mapsto BWA^\top$ can be written as $BWA^\top = (A \otimes B) \text{vec}(W)$ where vec stacks the columns of W into a vector.*

Use this fact to show that $\left\| U_\perp^{\star\top} U \Lambda^\star - \Lambda_\perp^\star U_\perp^{\star\top} U \right\|_F \geq \Delta \|U_\perp^\top U^\star\|_F$.

(d) Leverage these inequalities to prove this version of the Davis-Kahan $\sin \theta$ theorem.

Problem 4 - Fixing the spectrum

In Lecture 13 we proved a guarantee for the recovery of communities in the stochastic block model. This guarantee requires the probability of a link across communities q to be sufficiently large. In this exercise, we will fix this issue. Let $G \sim G(n, p, q)$ be a graph drawn using this distribution and A be its random adjacency matrix.

(a) Consider the following quantity

$$\hat{\lambda}_1 = \frac{2}{n(n-1)} \sum_{i < j} A_{ij}.$$

Show that with high probability (i.e., the probability goes to zero as $n \rightarrow \infty$)

$$\left| \hat{\lambda}_1 - \frac{p+q}{2} \right| \leq \frac{1}{n}.$$

(b) Design a modified spectral algorithm such that misclassifies at most $C/(p-q)^2$ misclassified vertices with high probability. **Hint:** Consider $\hat{A} = A - \hat{\lambda}_1 \mathbf{1}\mathbf{1}^\top$.