EECS 127/227AT **Fall 2020**
Homework 2
Homework 2 is due on Gradescope by Friday 9/18 at 11.59 p.m.

1 The Laplacian Matrix

We are given an undirected simple graph $G = (V, E)$ where $V = \{1, \ldots, n\}$ is the set of vertices and $E \subseteq V \times V$ is the set of oriented edges (recall that an undirected graph is called simple if it has no self-loops). Since G is undirected, for $i, j \in V$, we have $(i, j) \in E \Leftrightarrow (j, i) \in E$. Note that in this problem each unordered pair of vertices $\{i, j\}$ with $j \neq i$ for which i is connected to j (such an unordered pair is called an edge in graph theory) is being counted twice in E, once as (i, j) and once as (j, i), which is why we call E the set of oriented edges rather than just edges. Let d_i be the degree of vertex $i \in V$, i.e. the number of edges (i.e. half the number of oriented edges) incident at the vertex. We define the Laplacian matrix $L \in \mathbb{R}^{n \times n}$ as

$$L_{ij} = \begin{cases} d_i & i = j \\ -1 & (i, j) \in E \\ 0 & \text{otherwise} \end{cases}.$$

Equivalently, $L = D - A$, where D is the diagonal matrix defined by $D_{ii} = d_i$, and A is the adjacency matrix of the graph.

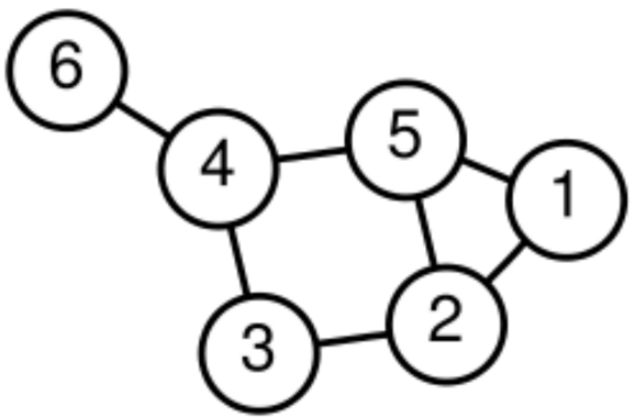

Figure 1: An undirected simple graph.

(a) Form the Laplacian matrix for the graph shown in Figure 1.

(b) Show that in general the Laplacian matrix L for any simple undirected graph $G = (V, E)$ is a symmetric matrix.

(c) Show that for every $x \in \mathbb{R}^n$ we have

$$x^\top L x = \frac{1}{2} \sum_{(i,j) \in E} (x_i - x_j)^2.$$

Deduce that $L \succeq 0$ (i.e. $L \in \mathbb{S}^n_+$).

(d) Show that zero is always an eigenvalue of L. Exhibit a corresponding eigenvector.

(e) Suppose that G is connected. That is, for any two vertices $u, v \in V$, they are connected by a path

$$(u, w_1), (w_1, w_2), \ldots, (w_{k-1}, w_k), (w_k, v) \in E,$$

for some $k \in \mathbb{N}$ and $w_1, \ldots, w_k \in V$. In this case, show that the rank of L is exactly $n - 1$.

2 Symmetric and Positive Semidefinite (PSD) Matrices

(a) A square matrix $B \in \mathbb{R}^{n \times n}$ is said to be *skew-symmetric* if $B^T = -B$. In particular, all the diagonal entries of B must be 0.

Show that any square matrix P can be written as a sum of a symmetric and a skew-symmetric matrix in a unique way. Explicitly write the expression for each of the matrices in this sum in terms of P.

(b) Suppose P and Q are two positive semidefinite matrices of the same size. Show that $Trace(PQ) = 0$ if and only if $PQ = 0$.

(c) For two matrices P and Q of the same size, their *Hadamard product* $P \circ Q$ is the matrix of the same size resulting from the elementwise product of the entries of P and Q. Namely, the (ij) entry of $P \circ Q$ is $p_{ij}q_{ij}$.

Show that if P and Q are positive semidefinite matrices of the same size, then $P \circ Q$ is positive semidefinite.

3 Ellipsoids Associated with Positive Semidefinite Matrices

(a) Show that if $A \in \mathbb{S}^n_{++}$ then $A^{-1} \in \mathbb{S}^n_{++}$.

(b) Recall that for $A, B \in \mathbb{S}^n$ we write $A \succeq B$ if $A - B \in \mathbb{S}^n_+$ so, for instance, $A \in \mathbb{S}^n_+$ can be written as $A \succeq 0$.

For the following parts of this problems assume that $A, B \in \mathbb{S}^n_{++}$.

Show that $B \succeq A$ if and only if $A^{-1} \succeq B^{-1}$. To make sense of the statement of the claim, recall that from part (a) of the problem we know that A^{-1} and B^{-1} exist and both are in $\mathbb{S}^n_{++}$.

(c) Let $\mathcal{E}_A := \{x \in \mathbb{R}^n : x^T A^{-1} x \leq 1\}$ and $\mathcal{E}_B := \{x \in \mathbb{R}^n : x^T B^{-1} x \leq 1\}$.
Show that $\mathcal{E}_A \subseteq \mathcal{E}_B$ iff $B \succeq A$.

4 Subspace Condition Number

Let $A \in \mathbb{R}^{n \times n}$ be a full-rank square matrix with singular values $\sigma_1 \geq \ldots \geq \sigma_n$. Recall that

$$\sigma_1 = \sup_{x \in \mathbb{R}^n : x \neq 0} \frac{\|Ax\|_2}{\|x\|_2},$$

and

$$\sigma_n = \inf_{x \in \mathbb{R}^n : x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

Further, since A has full rank, we have $\sigma_n > 0$.

We define the *condition number* $\kappa(A)$ as

$$\kappa(A) := \frac{\sigma_1}{\sigma_n} = \frac{\sup_{x \in \mathbb{R}^n : x \neq 0} \|Ax\|_2 / \|x\|_2}{\inf_{x \in \mathbb{R}^n : x \neq 0} \|Ax\|_2 / \|x\|_2}.$$

The condition number is a measure of "how singular" A is (if $\kappa(A)$ is large, then A is close to singular).

Given a subspace $\mathcal{V} \subseteq \mathbb{R}^n$, we can analogously define the *subspace condition number*

$$\kappa_{\mathcal{V}}(A) := \frac{\sup_{x \in \mathcal{V} : x \neq 0} \|Ax\|_2 / \|x\|_2}{\inf_{x \in \mathcal{V} : x \neq 0} \|Ax\|_2 / \|x\|_2},$$

which is just the condition number of A restricted to $\mathcal{V}$.

(a) Let $A = U \Sigma V^\top$ be the SVD of A. Show that

$$\kappa_{\mathcal{V}}(A) = \kappa_{\mathcal{W}}(\Sigma),$$

for some other subspace $\mathcal{W} \subseteq \mathbb{R}^n$ such that $\dim(\mathcal{W}) = \dim(\mathcal{V})$.

(b) The standard basis vector e_i is defined as the vector that is one at element i and zero elsewhere. Given a diagonal matrix Σ and distinct standard basis vectors e_i, e_j, suppose that $\|\Sigma e_i\|_2 = \sigma_i$ and $\|\Sigma e_j\|_2 = \sigma_j$ with $\sigma_i \leq \sigma_j$. For any b such that $\sigma_i \leq b \leq \sigma_j$, exhibit a linear combination u of e_i and e_j such that $\|\Sigma u\|_2 = b$ and $\|u\|_2 = 1$.

(c) Show that

$$\inf_{\mathcal{V} : \dim(\mathcal{V}) \geq \lceil n/2 \rceil} \kappa_{\mathcal{V}}(A) = 1.$$

Remark: Consider the special case when $A \succ 0$, i.e. $A \in \mathbb{S}^n_{++}$. Then we can write $A = U \Lambda U^T$ with $U \in \mathbb{R}^{n \times n}$ orthogonal and $\Lambda \in \mathbb{R}^{n \times n}$ diagonal with strictly positive entries. So, if the diagonal entries of Λ are in decreasing order, we can treat the representation $U \Lambda U^T$ of A itself as being its singular value decomposition and we see that the singular values of A are just its eigenvalues (which are all strictly positive) in decreasing order. The image of the unit ball under A is

$$\{Ax \;:\; x^T x \leq 1\} = \{y \in \mathbb{R}^n \;:\; y^T A^{-T} A^{-1} y \leq 1\}.$$

This is an ellipsoid with major (longest) and minor (shortest) axes σ_1 and σ_n respectively, where $\sigma_1 \geq \ldots \geq \sigma_n > 0$ are the singular values or, equivalently in this case, the eigenvalues of A. Any ellipsoid in $\mathbb{R}^n$ can be associated to a matrix in $\mathbb{S}^n_{++}$ in this way, so a corollary of this result is that any ellipsoid in $\mathbb{R}^n$ must have a $\lceil n/2 \rceil$-dimensional cross-section that is a sphere. For example, any ellipsoid in $\mathbb{R}^3$ has a 2-dimensional circular cross-section.

5 PCA and senate voting data

In this problem, we consider a matrix of senate voting data, which we manipulate in Python. The data is contained in a $n \times m$ data matrix X, where each row corresponds to a senator and each column to a bill. Each entry of X is either $1, -1$ or 0, depending on whether the senator voted for the bill, against the bill, or abstained, respectively. Please compute your answers using the attached Jupyter Notebook.

1. Suppose we want to assign a *score* to each senator based on their voting pattern, and then observe the empirical variance of these scores. To describe this, let us choose a $a \in \mathbb{R}^m$ and a scalar $b \in \mathbb{R}$. We define the score for senator i as:

$$f(x_i, a, b) = x_i^\top a + b, \quad i = 1, 2, \ldots, n.$$

Note that $x_i^\top$ denotes the i^{th} row of X and is a row vector of length m, as in the problem above.

Let us denote by $z = f(X, a, b)$ the column vector of length n obtained by stacking the scores for each senator. Then

$$z = f(X, a, b) = Xa + b\mathbb{1} \in \mathbb{R}^n$$

where $\mathbb{1}$ is a vector with all entries equal to 1. Let us denote the mean value of z by $\mu_z = \frac{1}{n}\mathbb{1}^\top z$. Let $\mu_X^\top$, where $\mu_X \in \mathbb{R}^m$, denote the row vector containing the mean of each column of X. Then

$$\begin{aligned}\mu_z &= \frac{1}{n}\sum_{i=1}^{n} f(x_i, a, b) \\ &= a^\top \mu_X + b\end{aligned}$$

The empirical variance of the scores can then be obtained as

$$\begin{aligned}\text{Var}(f(X, a, b)) &= \text{Var}(z) \\ &= \frac{1}{n}(z - \mu_z\mathbb{1})^\top (z - \mu_z\mathbb{1}) \\ &= \frac{1}{n}(Xa + b\mathbb{1} - a^\top \mu_X \mathbb{1} - b\mathbb{1})^\top (Xa + b\mathbb{1} - a^\top \mu_X \mathbb{1} - b\mathbb{1}) \\ &= \frac{1}{n}(Xa - \mathbb{1}\mu_X^\top a)^\top (Xa - \mathbb{1}\mu_X^\top a) \\ &= \frac{1}{n}a^\top (X - \mathbb{1}\mu_X^\top)^\top (X - \mathbb{1}\mu_X^\top) a\end{aligned}$$

Note that this variance is therefore a function of the "centered" data matrix $X - \mathbb{1}\mu_X^\top$ in which the mean of each column is zero. It also does not depend on b.

For the remainder of this problem, we assume that the data has been pre-centered (i.e., $\mu_X = 0$); note that this has been pre-computed for you in the code Notebook. Assume also that $b = 0$, so that $\mu_z = 0$. Defining $f(X, a) := f(X, a, 0)$, we can then write simpler variance formula

$$\text{Var}(f(X, a)) = \frac{1}{n}a^\top X^\top X a$$

Suppose we restrict a to have unit-norm. In the provided code, find a that maximizes $\text{Var}(f(X, a))$. What is the value of the maximum variance?

2. We next consider party affiliation as a predictor for how a senator will vote. Follow the instructions in the Notebook to compute the mean voting vector for each party and relate it to the direction of maximum variance.

3. Given a vector $z = Xu$ (i.e., the vector of scalar projections of each row of X along u), assuming it has mean value zero, we can compute its variance as

$$\text{Var}(z) = u^\top C u,$$

   where $C := \frac{X^\top X}{n}$. We will show in a future homework problem that the variance along each principal component a_i is precisely its corresponding eigenvalue of C, $\lambda_i(C)$. (For now, just note that this fact should make intuitive sense, since PCA is searching for directions of maximum variance of the data, and these occur along the covariance matrix's eigenvectors.)

   In the Notebook, compute the sum of the variance along a_1 and a_2 and plot the data projected on the a_1–a_2 plane.

4. Suppose we want to find the bills that are most and least contentious — i.e., those that have high variability in senators' votes, and those for which voting was almost unanimous. Follow the instructions in the Jupyter Notebook to compute several possible measures of "contentiousness" for each bill, plot the vote counts for exemplar bills, and comment on the metrics' relationship to each other.

5. Finally, we can use the defined score $f(X, a, b)$, computed along first principal component a_1, to classify the most and least "extreme" senators based on their voting record. Follow the instructions in the Jupyter Notebook to compute these scores and comment on their relationship to party affliation.