Midterm

Last Name: Chakraborty

First Name: Dhruv

Student ID: 204-962-098

Instructions: This is an open book exam, so you may use your notes, lecture slides and videos, and content on Piazza. You may also use any third part resources you find on the internet or in books. The only thing you are not allowed to do is ask someone else for help. This includes asking your classmates questions over Piazza, Group.me, email etc; consulting with other friends or family members; and using online Q& A services. If you need clarification on any of the problems you can contact Daniel McKenzie via private message on Piazza (preferred) or via email.

Please do not write below this line.

Question	Points	Score
1	25	
2	15	
3	10	
4	10	
5	10	
6	15	
Total:	85	

1. (10 points) Consider the least squares problem $\arg\min_{x\in\mathbb{R}^3} \|A\boldsymbol{x} - \boldsymbol{b}\|_2$. The matrix A, together with its (thick) QR decomposition, is given below:

$$A = \begin{bmatrix} 0.8147 & 0.6324 & 0.9575 \\ 0.9058 & 0.0975 & 0.9649 \\ 0.1270 & 0.2785 & 0.1576 \\ 0.9134 & 0.5469 & 0.9706 \end{bmatrix}$$

$$Q = \begin{bmatrix} -0.5332 & 0.4892 & 0.6519 & 0.2267 \\ -0.5928 & -0.7162 & 0.1668 & -0.3284 \\ -0.0831 & 0.4507 & -0.0991 & -0.8833 \\ -0.5978 & 0.2112 & -0.7331 & 0.2462 \end{bmatrix} \quad R = \begin{bmatrix} -1.5279 & -0.7451 & -1.6759 \\ 0 & 0.4805 & 0.0534 \\ 0 & 0 & 0.0580 \\ 0 & 0 & 0 \end{bmatrix}$$

and $\boldsymbol{b} = \begin{bmatrix} 0.9572 & 0.4854 & 0.8003 & 0.1419 \end{bmatrix}^\top$.

(a) (10 points) Use the QR decomposition to solve the least squares problem. Show all your work. As in the homework, you may use software or a calculator to perform all the necessary matrix-vector products.

We know from lecture that $\|An-b\|_2 = \|Rn - Q^Tb\|_2$. Taking $R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$ with $R_1 \in \mathbb{R}^{3\times 3}$, and $Q^Tb = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ for $b_1 \in \mathbb{R}^3$ and $b_2 \in \mathbb{R}^1$, we get $\|Rn - Q^Tb\|_2^2 = \|R_1 n - b_1\|_2^2 + \|b_2\|_2^2$. $\therefore \arg\min_{n\in\mathbb{R}^3} \|An - b\|_2 = \arg\min_{n\in\mathbb{R}^3} \|R_1 n - b_1\|_2$.

We can minimize this by taking n s.t. $R_1 n = b_1$.

Hence, we first find $Q^Tb = \begin{pmatrix} -0.9495 \\ 0.5113 \\ 0.5216 \\ -0.6144 \end{pmatrix} \Rightarrow b_1 = \begin{pmatrix} -0.9495 \\ 0.5113 \\ 0.5216 \end{pmatrix}$. Since $R_1 = \begin{pmatrix} -1.5279 & -0.7451 & -1.6759 \\ 0 & 0.4805 & 0.0534 \\ 0 & 0 & 0.0580 \end{pmatrix}$

we solve $R_1 n = b_1$ to yield $0.0580 n_3 = 0.5216$, $0.0534 n_3 + 0.4805 n_2 = 0.5113$ and $-1.5279 n_1 - 0.7451 n_2 - 1.6759 n_3 = -0.9495$. Solving using backward substitution, we get $n = \begin{pmatrix} -9.2743 \\ 0.0647 \\ 8.9931 \end{pmatrix}$.

(b) (5 points) Write down the *thin* QR decomposition for this matrix A.

The thin QR decomposition exploits the fact that we have $A = QR = [Q_1, Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1 + Q_2 0 = Q_1 R_1$. In our case we have $Q_1 \in \mathbb{R}^{4\times 3}$ and $R_1 \in \mathbb{R}^{3\times 3}$. In particular,

$$A = Q_1 \cdot R_1$$

$$Q_1 = \begin{pmatrix} -0.5332 & 0.4892 & 0.6519 \\ -0.5928 & -0.7162 & 0.1668 \\ -0.0831 & 0.4507 & -0.0991 \\ -0.5978 & 0.2112 & -0.7331 \end{pmatrix} \qquad R_1 = \begin{pmatrix} -1.5279 & -0.7451 & -1.6759 \\ 0 & 0.4805 & 0.0534 \\ 0 & 0 & 0.0580 \end{pmatrix}$$

UID: 204-962-098

2. Consider the matrix $A \in \mathbb{R}^{3\times 4}$, given below together with its (thin) singular value decomposition:

$$A = \begin{bmatrix} 0.8363 & 0.4474 & 0.9212 \\ 0.3587 & 0.5217 & 0.0485 \\ 0.3074 & 0.5509 & 0.7195 \\ 0.0236 & 0.7245 & 0.7871 \end{bmatrix}$$
$$= \begin{bmatrix} -0.6574 & -0.6498 & -0.2559 \\ -0.2479 & -0.1108 & 0.9616 \\ -0.4963 & 0.1615 & -0.0746 \\ -0.5099 & 0.7344 & -0.0649 \end{bmatrix} \begin{bmatrix} 1.9173 & 0 & 0 \\ 0 & 0.5906 & 0 \\ 0 & 0 & 0.4326 \end{bmatrix} \begin{bmatrix} -0.4190 & -0.8740 & 0.2461 \\ -0.5562 & 0.4613 & 0.6913 \\ -0.7177 & 0.1528 & -0.6794 \end{bmatrix}$$

(a) (5 points) Give the definition of the matrix 2-norm, $\|A\|_2$.

The matrix 2-norm is the induced norm of the vector Euclidean norm such that it is defined as

$$\|A\|_2 = \sup_{v \neq 0} \frac{\|Av\|_2}{\|v\|_2} = \max_{\|v\|_2 = 1} \|Av\|_2 .$$

It turns out that this is equivalent to the largest singular value of A i.e. $\|A\|_2 = \lambda_{max}(A^TA)^{1/2} = \sigma_{max}(A)$.

(b) (5 points) Using the information given, compute $\|A\|_2$.

We know that the matrix 2-norm is the special case of induced vector p-norms s.t. $\|A\|_2 = \sigma_{max}(A)$.
Since we have the Singular Value Decomposition of $A = U\Sigma V^T$, we have the singular values of A in Σ and

$\therefore \|A\|_2 = 1.9173.$

Question 2 continues on the next page...

Question 2 continued…

(c) Compute the matrix ∞-norm, $\|A\|_\infty$.

We know that $\|A\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^{m} |a_{ij}|$ for $A \in \mathbb{R}^{m \times n}$ i.e. it is the max absolute row sum in A.

$\therefore \|A\|_\infty = 2.2049$

(d) (5 points) Compute, using the information given, the best rank 2 approximation to A .

We know that the best k rank approximation to A, $A^{(k)} = \sum_{i=1}^{k} \sigma_i u_i v_i^T$ where $\Sigma = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_3 \end{pmatrix}$.

So $A^{(2)} = 1.9173 \begin{pmatrix} -0.6574 \\ -0.2479 \\ -0.4963 \\ -0.5099 \end{pmatrix} \otimes \begin{pmatrix} -0.4190 \\ -0.5562 \\ -0.7177 \end{pmatrix} +$

$0.5906 \begin{pmatrix} -0.6498 \\ -0.1108 \\ 0.1615 \\ 0.7344 \end{pmatrix} \otimes \begin{pmatrix} -0.8740 \\ 0.4613 \\ 0.1528 \end{pmatrix}$

$\otimes$ is outer product

$$\therefore A^{(2)} = \begin{pmatrix} 0.8635 & 0.5240 & 0.8460 \\ 0.2563 & 0.2342 & 0.3311 \\ 0.3153 & 0.5733 & 0.6975 \\ 0.0305 & 0.7438 & 0.7679 \end{pmatrix}$$

UID: 204-962-098

3. Let $\mathcal{G}$ be a $2 \times 2 \times 3$ tensor with slices:

$$\mathcal{G}(:,:,1) = \begin{bmatrix} 0 & 4 \\ -7 & 2 \end{bmatrix} \quad \mathcal{G}(:,:,2) = \begin{bmatrix} 3 & 1 \\ 12 & -1 \end{bmatrix} \quad \mathcal{G}(:,:,3) = \begin{bmatrix} 3 & 0 \\ 2 & 31 \end{bmatrix}$$

(a) (5 points) Compute $\text{unfold}_1(\mathcal{G}), \text{unfold}_2(\mathcal{G})$ and $\text{unfold}_3(\mathcal{G})$.

$$\text{unfold}_1(G) = \begin{pmatrix} 0 & 4 & 3 & 1 & 3 & 0 \\ -7 & 2 & 12 & -1 & 2 & 31 \end{pmatrix} \in \mathbb{R}^{2 \times (2 \times 3)}$$

$$\text{unfold}_2(G) = \begin{pmatrix} 0 & -7 & 3 & 12 & 3 & 2 \\ 4 & 2 & 1 & -1 & 0 & 31 \end{pmatrix} \in \mathbb{R}^{2 \times (2 \times 3)}$$

$$\text{unfold}_3(G) = \begin{pmatrix} 0 & 4 & -7 & 2 \\ 3 & 1 & 12 & -1 \\ 3 & 0 & 2 & 31 \end{pmatrix} \in \mathbb{R}^{3 \times (2 \times 2)}$$

(b) (5 points) How would you compute the rank 1 CP decomposition of $\mathcal{G}$? Write down the appropriate optimization problem, taking care to correctly specify the dimensions of each variable. *You don't need to write pseudocode for an algorithm to solve this optimization problem, or anything like that.*

$G \in \mathbb{R}^{2 \times 2 \times 3}$

We want $B = u_1 \otimes v_1 \otimes w_1$ where $u_1 \in \mathbb{R}^2, v_1 \in \mathbb{R}^2$ and $w_1 \in \mathbb{R}^3$.

such that the rank 1 CP decomposition of G

$$G^{(1)} = \underset{B}{\text{argmin}} \, \| G - B \|_F = \underset{B}{\text{argmin}} \, \| G - u_1 \otimes v_1 \otimes w_1 \|_F .$$

We can then set up an alternating least squares algorithm that iteratively finds better approximations to u_1, v_1 and w_1. (can get unfoldings in terms of the Khatri-Rao product $\odot$)

This ends up looking like

$$u_{1_k} = \underset{u_1}{\text{argmin}} \, \| \text{unfold}_1(G) - u_1 (w_{1_{k-1}} \odot v_{1_{k-1}})^T \|_F^2$$

for each of u_1, v_1 and w_1 in an alternating fashion.

4. Consider the linear system $A\boldsymbol{x} = \boldsymbol{b}$. Suppose that A is invertible and symmetric, and let $\kappa(A)$ denote the condition number of the matrix A.

(a) (5 points) Show that, for all A, we have that $\kappa(A) \geq 1$. *Hint: Use the definition of $\kappa(A)$ in terms of eigenvalues of A given in Question 7 of Homework 1.*

We first note that A is invertible so it cannot have an eigenvalue of 0. Moreover, since A is symmetric, we know that it only has real eigenvalues. In Problem 7 of HW 1, we showed that the following 2 defns. of $K(A)$ are equivalent: $K(A) = \|A\|_2 \|A^{-1}\|_2 = \dfrac{\lambda_{max}(A)}{\lambda_{min}(A)}$. Recall that norms are always non-negative and thus note that $K(A) = \dfrac{\lambda_{max}(A)}{\lambda_{min}(A)} > 0$. In particular then, $\lambda_{max}(A) \geq \lambda_{min}(A)$ so that $\dfrac{\lambda_{max}(A)}{\lambda_{min}(A)} \geq 1$ giving us $K(A) \geq 1$.

(b) (5 points) Which is preferable, $\kappa(A) = 1.01$ or $\kappa(A) = 900$? Justify your answer.

By the condition number theorem, we know that

$$\frac{\|x - \hat{x}\|_2}{\|x\|_2} \leq \frac{K(A)}{1-r}\left(\frac{\|\delta A\|_2}{\|A\|_2} + \frac{\|\delta b\|_2}{\|b\|_2}\right)$$

where $\hat{x}$ is the least squares approx. to x, $r = \|\delta A\| \|A^{-1}\|_2 < 1$, and δA and δb are the noise in the data.

In particular then, note that a higher condition number seems to suggest a larger amplification of the noise which would make $\hat{x}$ a worse and more unstable approximation of x. $\therefore K(A) = 1.01$ is preferable.

UID: 204-962-098

5. Consider the matrix $B = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$. Suppose that the rows of B represent Netflix users, while the columns represent movies on Netflix, *i.e.* $B_{ij} = \begin{cases} 1 & \text{if user } i \text{ has watched movie } j \\ 0 & \text{otherwise} \end{cases}$

(a) (5 points) What is a non-negative matrix factorization of B? Is it unique?

A non-negative matrix factorization of $B = WH$ where $W \in \mathbb{R}^{4\times r}$ and $H \in \mathbb{R}^{r \times 4}$ is an approximation of B i.e. $W, H = \underset{U,V}{\arg\min} \|A - UV\|_F$. We typically have $B \approx WH$, there are no unique non-negative factorizations. NMFs are introduced more as an interpretable way to factor matrices, which we discuss more in part (b). An example factorization of $B \approx WH$ with $W \in {}^{4\times 2}$ and $H \in \mathbb{R}^{2\times 4}$ is shown: $W = \begin{pmatrix} 2.03\times10^{-4} & 9.05 \times 10^{-1} \\ 8.21\times10^{-1} & 8.95\times10^{7} \\ 4.56\times10^{-1} & 0 \\ 1.02 & 0 \end{pmatrix}$

$\& \; H = \begin{pmatrix} 0.5303 & 0.9556 & 1.1916 & 0.0130 \\ 0 & 0 & 0 & 1.1047 \end{pmatrix}$

(b) (5 points) How might we use a non-negative matrix factorization to better understand this data set?

The idea behind practical applications of NMF is typically topic modelling. For example, in this case we hypothesize that these users typically like movies within the same genre and get split B into W and H in a way that represents splitting a users affinity to a movie by considering the product of the users affinity to the given genres and the movie relation to each genre.
The idea behind making these matrices positive is that it is hard to understand what a negative number would represent in the context of a movie "not belong" to a genre or a user "disliking" genres.

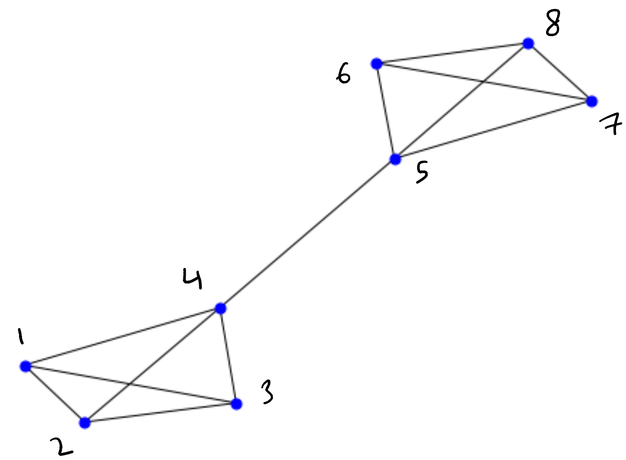

Figure 1: Just a graph

6. Consider the graph shown in Figure 1.

   (a) (5 points) Write down the adjacency matrix and Laplacian for this graph. Note that writing down the adjacency matrix requires you to choose a labelling of the vertices (*i.e.* vertex 1, vertex 2 and so on). As part of your answer, indicate the ordering you have chosen on the pictured graph.

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

$$L = D - A = \begin{pmatrix} 3 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & -1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 3 & -1 & 0 & 0 & 0 & 0 \\ -1 & -1 & -1 & 4 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 4 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 & -1 & 3 & -1 \\ 0 & 0 & 0 & 0 & -1 & -1 & -1 & 3 \end{pmatrix}$$

   (b) (10 points) Use spectral clustering to split this graph into two clusters. Show all your work. Again, you may use software or a calculator to do linear algebra computations such as determining eigenvectors.

We compute the second eigenpair of L, (λ_2, v_2) i.e. second smallest eigenvalue $(\lambda_1 \leq \lambda_2 \leq \dots)$.

Here, $\lambda_2 = 0.3543$ and $v_2 = \begin{pmatrix} -0.3825 \\ -0.3825 \\ -0.3825 \\ -0.2470 \\ 0.2470 \\ 0.3825 \\ 0.3825 \\ 0.3825 \end{pmatrix}$. Recall that

by spectral clustering algorithm we assign vertices to clusters by $v_i \in C_1$ if $(v_2)_i > 0$ or $v_i \in C_2$ if $(v_2)_i < 0$.

$\therefore$ $C_2 = \{v_1, v_2, v_3, v_4\}$ & $C_1 = \{v_5, v_6, v_7, v_8\}$.