University of Toronto
Department of Electrical and Computer Engineering

ECE367 MATRIX ALGEBRA AND OPTIMIZATION

Problem Set #3

Autumn 2020

Prof. S. C. Draper **Due:** 5pm (Toronto time) Friday, 23 October 2020

Homework policy: Problem sets must be turned by the due date and time. Late problem sets will receive deductions for lateness. See the information sheet for futher details. The course text "Optimization Models" is abbreviated as "OptM". Also, see PS01 for details of the "Non-graded", "graded" and "optional" problem categories.

Note: In the categorization below Graded problems are highlighted in **red boldface**

Problem Set #3 problem categories: A quick categorization by topic of the problems in this problem set is as follows:

- Symmetric matrices: Problems 3.1, 3.2
- Quadratic constraints and ellipses: Problems 3.3, 3.4, **3.9**
- SVDs: Problems 3.4, 3.5, 3.6
- Applications of symmetric matrices and SVDs: Problems 3.7, 3.8, **3.10**, **3.11**

NON-GRADED PROBLEMS

Problem 3.1 (Eigenvectors of a symmetric 2×2 matrix)
OptM Problem 4.1.

Problem 3.2 (A lower bound on the rank)
OptM Problem 4.9 parts 1 and 2 only (not part 3).

Problem 3.3 (Quadratic constraints)
OptM Problem 4.2.

Problem 3.4 (SVD of an orthogonal matrix)
OptM Problem 5.1.

Problem 3.5 (Practice computing SVDs)
Compute by hand the singular value decomposition of

$$A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}.$$

In other words, express A as $A = U\tilde{\Sigma}V^T$ where $U \in \mathbb{R}^{2,2}$ and $V \in \mathbb{R}^{3,3}$ are orthogonal matrices and $\tilde{\Sigma} \in \mathbb{R}^{2,3}$, specify the values of

(a) Specify the singular values by specifying $\tilde{\Sigma}$.

(b) Specify the right singular vectors by specifying V.

(c) Specify the left singular vectors by specifying U.

(d) Reassemble your calculations of parts (a)-(c) to show that $U\tilde{\Sigma}V^T$ in fact does equal the A you started with. Also express A as a sum of rank-one matrices where each matrix is the outer product of a left singular vector and a right singular value scaled by a singular value.

Problem 3.6 (SVD of a score matrix)
OptM Problem 5.4.

Problem 3.7 (Connected graphs and the Laplacian)
OptM Problem 4.5.

Problem 3.8 (Fitting a hyperplane to data)
OptM problem 5.6.

(Hints: To get the quadratic function in the second part of the problem sum up the m distances—one per data point—and manipulate the resulting sum into a quadratic form, cf. Sec. 4.1.2. In the last part of the question recall our development of Rayleigh quotients in Ch. 4.)

GRADED PROBLEMS

Problem 3.9 (Ellipses, eigenvalues, eigenvectors, and volume)
Make neat and clearly-labelled *sketches* (i.e., draw by hand) of the ellipsoid $\mathcal{E} = \{x|(x-x_c)^T P^{-1}(x-x_c) = 1\}$ for the following sets of parameters:

(a) Center $x_c = [0\ 0]^T$ and $P = [1.5\ \ -0.5; -0.5\ 1.5]$.

(b) Center $x_c = [1\ \ -2]^T$ and $P = [3\ 0; 0\ 1]$.

(c) Center $x_c = [-2\ 1]^T$ and $P = [9\ \ -2; -2\ 6]$.

For each part (a)–(c) also compute each pair of eigenvalues and corresponding eigenvectors.

(d) Recall the geometrically meaningful property of the determinant of a square real matrix A: its magnitude $|\det A|$ is equal to the volume of the parallelepiped $\mathcal{P}$ formed by applying A to the unit cube $\mathcal{C} = \{x|\, 0 \le x_i \le 1, i \in [n]\}$. In other words, if $\mathcal{P} = \{Ax|x \in \mathcal{C}\}$ then $|\det(A)|$ is equal to the volume of $\mathcal{P}$. Furthermore, recall that the determinant of a matrix is zero if any of its eigenvalues are zero. Explain how to interpret this latter fact in terms of the interpretation of $|\det(A)|$ as the volume of $\mathcal{P}$. (This interpretation was mentioned in class so this is just a "I want to make sure you understand that comment" type of question.)

Problem 3.10 (Latent semantic indexing)
In this problem you build off the problem "Angles between word vectors" from PS01, making a connection to the singular value decomposition. First complete the following two parts of OptM problem 5.5:

(a) OptM problem 5.5 part 3.

(b) OptM problem 5.5 part 4.

In the following parts we connect OptM problem 5.5 to the "Angles between word vectors" (ABWV) problem in PS01, and apply the latent semantic indexing method to the Wikipedia article collection. We start by briefly re-describing the ABWV problem set-up below, but please feel free to go back and read the original problem statement.

In ABWV, we considered a set of documents $\mathcal{D}$ the where the number of documents is $|\mathcal{D}|$. The set $\mathcal{W}$ denotes the union of words in all articles, i.e., the lexicon of the set of documents where the cardinality of $\mathcal{W}$ is $|\mathcal{W}|$. We assume the lexicon is ordered "lexiographically" (e.g., alphabetically) so that there is a one-to-one mapping from each word $w \in \mathcal{W}$ to an element of the index set $t \in [|\mathcal{W}|]$. Let $f_{\text{term}}(t, d)$ denote the number of times the word $w \in \mathcal{W}$ that is indexed as $t \in [|\mathcal{W}|]$

appears in the dth article where $d \in [|\mathcal{D}|]$. For ABWV, you were provided with a pre-processed MATLAB data file `wordVecV.mat`. Please re-use the same data file for this problem. You can load the content in the second file into MATLAB by using command `load 'wordVecV.mat'`. After loading, you will see a variable `V` of dimensions 1651×10. We refer to this matrix as V. The value in the tth row and dth column of this matrix is $f_{\text{term}}(t, d)$. Note that in the given dataset $|\mathcal{D}| = 10$ and $|\mathcal{W}| = 1651$.

Now we connect OptM problem 5.5 to the ABWV problem set-up. You will immediately notice that $m = |\mathcal{D}|$ and $n = |\mathcal{W}|$. You can compute the $n \times m$ "(raw) term-by-document matrix" M by noting that $[M]_{i,j} = \mathbb{1}([V]_{i,j})$, where $\mathbb{1}(x)$ is 1 if $x > 0$ and 0 otherwise. The OptM problem 5.5 also describes how to obtain $\tilde{M}$, a normalized version of M.

(c) Use MATLAB `svd` command to compute the singular value decomposition of $\tilde{M}$. List the 10 largest singular values in sorted order.

(d) In part (b) you assumed a low-rank approximation of $\tilde{M}$ and found an expression for the document similarity. Let the distance between ith and jth documents be $d(i, j)$ as per your expression from part (b). Let the rank of your approximation be k where $0 < k \leq \min(m, n)$. Compute $d(i, j)$ for $i, j \in [m]$ by assuming $k = 9$. Write down the titles of two most similar documents.

(e) Repeat what you did in part (d) with $k = 8, 7, \ldots, 1$. What is the lowest k that does not change your answer for part (d)? If your answer for lowest k is greater than 1 what is the pair of most similar documents for $k - 1$?

Problem 3.11 (Eigenfaces and ℓ_2 projection)
In this problem you will familiarize with the concept of Eigenfaces and its uses. Download the dataset `yalefaces.mat` from the course website. This dataset consists of 32×32 gray scale images of faces taken from the *Extended Yale Face Database B* (`http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html`). Load the dataset into your MATLAB environment by executing the command `load('yalefaces.mat')`. You will see a new variable `M` of dimension $32 \times 32 \times 2414$ which consists of 2414 grayscale images, 32×32 pixels each. The pixel values of the images range from 0 to 255. You can view the loaded images by making use of MATLAB built-in functions `imshow` or `imagesc`. As an example, the first image of the dataset can be displayed by executing `imshow(M(:,:,1)/255)`.

Let N be the number of images in the dataset and let $d = 1024$, the total number of pixels in each image. An image can be thought of as a matrix with 32 columns and 32 rows consisting of entries in the range $[0, 255]$. In this exercise we consider the images in vector form. Let $\mathcal{J}_i$ be the ith image in the dataset where $i \in [N]$. We formulate a column vector $x^{(i)} \in \mathbb{R}^d$ by flattening the matrix that makes up $\mathcal{J}_i$. In our case we stack all 32 columns vertically to form a d-dimensional vector $x^{(i)}$. In MATLAB you can do this efficiently using the command `reshape`. The 'average face' vector of

the dataset $\bar{x}$ can be computed as $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x^{(i)}$. We can (roughly) visualize the $x^{(i)}$s as forming a cloud of data points where the cloud is centered at $\bar{x}$ in d-dimensional space. Translate the cloud center to the origin by subtracting $\bar{x}$ from each $x^{(i)}$. Let the resulting vectors be denoted as $\bar{x}^{(i)} = x^{(i)} - \bar{x}$, which we will refer to as centered image vectors. Construct a matrix $X \in \mathbb{R}^{d\times N}$ by concatenating all the centered vectors $\bar{x}^{(i)}$, i.e., $X = \left[\bar{x}^{(1)}, \bar{x}^{(2)}, \ldots, \bar{x}^{(N)}\right]$. The matrix $C = XX^T$ is N times the covariance matrix of the data samples $x^{(i)}, i \in [N]$.

(a) In class you learned about singular value decomposition (SVD) and eigendecomposition. What is the connection between the singular values of X and the eigenvalues of C? What is the connection between the left-singular vectors of X and the eigenvectors of C? Make sure to describe the reasoning behind your answers, e.g., by describing the singular or eigen decompositions of each matrix.

(b) Compute the eigenvalue/eigenvector pairs of C and arrange them in decreasing order of the magnitude of the eigenvalues. Let the jth pair in the ordered set be denoted as $\lambda_j \in \mathbb{R}, v^{(j)} \in \mathbb{R}^d$ for $j \in [d]$. You may find the MATLAB command `svd` or `eig` helpful. Comment whether the eigenvalues are real and briefly explain why. Plot $\log \lambda_j$ against $j \in [d]$. Include your plot in your solution.

(c) The set of eigenvectors you computed in the previous part form a basis for the centered image vectors. Reshape each eigenvector $v^{(j)}$ to obtain a 32×32 matrix. These matrices can be considered as images and they are known as *eigenfaces*. Include plots of two sets of eigenfaces, those corresponding to the largest 10, and those corresponding to the smallest 10, eigenvalues. Do you observe any difference between the two sets of eigenfaces you plotted? If so briefly explain the reason for this difference.

(d) In this part, consider the images $\mathcal{J}_i$ for $i \in \{1, 1076, 2043\}$. Let $\mathcal{B}_j = \{v^{(1)}, v^{(2)}, \ldots, v^{(j)}\}$ where $j = \{2^1, 2^2, 2^3, \ldots, 2^{10}\}$, i.e., j indexes sets each consisting of a different number of the eigenvectors. Let us denote the ℓ_2 projection of $\bar{x}^{(i)}$ onto the basis vector set $\mathcal{B}_j$ as $\bar{y}^{(i,j)}$. Compute $\bar{y}^{(i,j)}$ for the given i, j pairs using MATLAB. (I.e., do this using your numerical tool and not by hand.) Note that $\bar{y}^{(i,j)}$ vectors are computed using the centered image vectors. Translate these vectors back (un-center them) by the cloud center shift $\bar{x}$ to get the image vectors $y^{(i,j)} = \bar{y}^{(i,j)} + \bar{x}$. Reshape the $y^{(i,j)}$ vectors into 32×32 matrices and plot them as images. Note that you will need to plot 30 images and these can be compactly plotted using `subplot` command in MATLAB.

(e) In this part you will learn how the eigenfaces can be used for the task of *face recognition*. Consider the two sets of indices $\mathcal{I}_1 = \{1, 2, 7\}$ and $\mathcal{I}_2 = \{2043, 2044, 2045\}$. The faces in the set $\mathcal{J}_i$ for $i \in \mathcal{I}_1$ belong to one person and those for $i \in \mathcal{I}_2$ belong to a second person. We have carefully picked the image indices so that the corresponding images are taken under similar lighting conditions. Consider $\mathcal{B}_{25}$ where $\mathcal{B}_j$ is defined as in part (d). Compute the projection coefficients obtained by projecting $\bar{x}^{(i)}, i \in \mathcal{I}_1 \cup \mathcal{I}_2$ onto the eigenvectors in $\mathcal{B}_{25}$. Let $c^{(i)} \in \mathbb{R}^{25}$ be the vector that consists of coefficients obtained for ith image. The vector

$c^{(i)}$ can be thought of as a representation of the corresponding original image $\mathcal{J}_i$. Intuitively, images that belong to same person should have similar coefficient vectors. To verify this, compute the pairwise Euclidean distances between the $c^{(i)}$ vectors. Tabulate the values and comment whether the distance between any two $c^{(i)}$ vectors that belong to the same person is smaller than those belonging to the other person. Briefly explain how you can use this to build a simple face recognition scheme.

