University of Toronto
Department of Electrical and Computer Engineering

ECE367 MATRIX ALGEBRA AND OPTIMIZATION

Problem Set #3
Autumn 2021

Prof. S. C. Draper **Due:** 8pm (Toronto time) Saturday, 16 October 2021

Homework policy: Problem sets must be turned by the due date and time. Late problem sets will not be accepted. See the information sheet for further details. The course text "Optimization Models" is abbreviated as "OptM" and "Introduction to Applied Linear Algebra" as "IALA".

Problems are categorized as:

- **"Theory" problems:** These are mostly mathematical questions designed to give you deeper insight into the fundamentals of the ideas introduced in this class.
- **"Application" problems:** These questions are designed to expose you to the breadth of application of the ideas developed in class and to introduce you to useful numerical toolboxes. Problems of this sort often ask you to produce plots and discuss your results; said plots and discussions should be included in and form part of your submission – think of your submitted solution like a lab book. Your attached code simply provides back-up evidence.
- **"Optional" problems:** Optional problems provide extra practice or introduce interesting connections or extensions. They need not be turned in. I will assume you have reviewed and understood the solutions to the optional problems when designing the exams.

Hand-in procedure:

- **Initial submission:** Your initial submission of the "Theory" and "Application" questions must be submitted via Quercus upload by the due date. Click on the `Assignments` tab, then look for the `Initial submission` tab and upload under the correct problem set number.
- **Self-assessment:** After the problem set is due we will post solutions. You will have one week from the initial due date to submit a commented version of your assignment in which, using as a reference the posted solutions, you highlight your errors or omissions in red. Annotate the PDF you initially submitted. If you have not submitted a solution you cannot submit the self-assessment. To submit the self-assessment on Quercus, click on the `Assignments` tab, then look for the `Self-assessment` tab and upload under the correct problem set number.
- **Late problem sets are not accepted**
- **Grading:** Per the course handout problem sets are graded for completion only. Points are assigned to (i) Initial submission of theory part, (ii) Submission of application part, (iii) Self-assessment of theory part. The relative points breakdown is detailed in each problem set.

Points allocation

- Theory parts (initial assessment): 1 pt
- Application parts (initial assessment): 1 pt
- Theory parts (self-assessment): 1 pt

Problem categorization and main concepts covered

Theory

- Eigenvalue and eigenvectors: Problems 3.1, 3.2
- Symmetric matrices: Problems 3.3-3.5
- Ellipses: Problem 3.6

Application

- Second-order approximations: Problem 3.7
- PageRank: Problem 3.8

Optional

- Diagonalizing circulant matrices: Problem 3.9

THEORY PROBLEMS

Problem 3.1 (Practice computing eigenvalues and eigenvectors)
In this problem you consider the eigenvalues and eigenvectors of each of the following four matrices:

$$\text{(a)}\quad A = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}, \qquad \text{(b)}\quad A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \text{(c)}\quad A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad \text{(d)}\quad A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

For each of the matrices in parts (a)-(d) do the following:

(i) compute the eigenvalues of the matrix,

(ii) compute the eigenvectors of the matrix,

(iii) specify both the algebraic and geometric multiplicity of each distinct eigenvalue, and

(iv) *if* the matrix is diagonalizable express the matrix in its diagonalized form. In other words, if A is diagonalizable express it as $A = V\Lambda V^{-1}$ where V is the matrix of eigenvectors and Λ is a diagonal matrix of eigenvalues. Specify V and Λ for each matrix that is diagonalizable.

Problem 3.2 (Eigenvectors of a symmetric 2×2 matrix)
OptM Problem 4.1.

Problem 3.3 (Function approximation), from a previous exam
In this problem you consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined pointwise as

$$f(x) = -\sqrt{x_1 x_2}$$

where $\text{dom} f = \{(x_1, x_2) | x_1 > 0, x_2 > 0\} = \mathbb{R}^2_{++}$, i.e., the interior of the first quadrant. This problem concerns first-order and second-order approximations of the function f.

(a) Denote by $\hat{f}_1$ the first-order (affine) approximation of f when the approximation is centered at some point $x^{(0)} \in \text{dom} f$. Say you center the approximation at $x^{(0)} = (4, 1)$ and you want $\hat{f}_1((6, 3))$, the first-order approximation at the point $x = (6, 3)$. What is $\hat{f}_1((6, 3))$? (Make sure to show your work developing the approximation and then evaluate it for the specified parameters.)

(b) Denote by $\hat{f}_2$ the second-order approximation of f when the approximation is centered at some point $x^{(0)} \in \text{dom} f$. Say you center the approximation at $x^{(0)} = (4, 1)$ and you want $\hat{f}_2((6, 3))$, the second-order approximation at the point $x = (6, 3)$. What is $\hat{f}_2((6, 3))$? (Make sure to show your work developing the approximation and then evaluate it for the specified parameters.)

(c) **True or false:** The Hessian $\nabla^2 f(x^{(0)})$ is positive semi definite (PSD) at all points $x^{(0)} \in \mathrm{dom} f$.
Please note: If you answer "true" you must provide a proof. If you answer "false" you must provide a counterexample. Providing a counterexample means you specific a point $x^{(0)} \in \mathbb{R}^2$ and show that $\nabla^2 f(x^{(0)})$ is not PSD. No credit is given for guessing. A "guess" is an answer provided without justification.
Hint: In class we have seen multiple ways to test for positive definiteness. For this problem some tests are easier to use than others.

Part (d) concerns lines. Recall the definition of a line. A line $\mathcal{L} \subseteq \mathbb{R}^n$ can be specified by a point $x^{(0)} \in \mathbb{R}^n$ and a direction $v \in \mathbb{R}^n$, which we assume is normalized so $\|v\|_2 = 1$. The line is the set $\mathcal{L} = \{x | x = x^{(0)} + tv, t \in \mathbb{R}\}$. In this problem $n = 2$ so $x^{(0)} \in \mathbb{R}^2$ and $v \in \mathbb{R}^2$.

(d) Find a point $x^{(0)} \in \mathrm{dom} f$ and a direction v, $\|v\|_2 = 1$, such that when both approximations, $\hat{f}_1$ and $\hat{f}_2$, are centered at $x^{(0)}$ the two approximations agree along the line $\mathcal{L} = \{x | x = x^{(0)} + tv, t \in \mathbb{R}\}$. What is your pair $(x^{(0)}, v)$?

Problem 3.4 (Symmetry across lines), from a previous exam
In this problem we consider the operation of finding the point symmetric to a given point in $\mathbb{R}^n$ about some line $\mathcal{L}$. Recall the definition of a line. A line $\mathcal{L} \subseteq \mathbb{R}^n$ can be specified by a point $x^{(0)} \in \mathbb{R}^n$ and a direction $v \in \mathbb{R}^n$, which we assume is normalized so $\|v\|_2 = 1$. The line is the set $\mathcal{L} = \{x | x = x^{(0)} + tv, t \in \mathbb{R}\}$.

For any point $x \in \mathbb{R}^n$ there is a point $f(x)$ symmetric to x about the line. This point is defined as $f(x) = x + 2(p(x) - x) = 2p(x) - x$ where $p(x)$ is the projection of x onto $\mathcal{L}$, i.e.,

$$p(x) = \arg\min_{p \in \mathcal{L}} \|p - x\|_2.$$

See Figure 1 for an illustration. Note almost all parts of this problem can be solved independently of the other parts.

(a) Show that, for any x, f can be expressed as the map $f(x) = Ax + b$ where $A = 2P - I$, $b = 2(I - P)x^{(0)}$, and $P = vv^T$, where $(x^{(0)}, v)$ are given and define the line $\mathcal{L}$.

(b) What is the geometric interpretation of the vector b?

(c) Show the mapping f is linear *if and only if* the line passes through the origin $0 \in \mathbb{R}^n$.

(d) Show that $f(f(x)) = x$ for every $x \in \mathbb{R}^n$. What is the geometric meaning of this?

(e) Show that A is symmetric and find its spectral decomposition. In particular, define $\{u^{(2)}, u^{(3)}, \ldots, u^{(n)}\}$ to be an orthonormal basis for the subspace orthogonal to the v vector defined in part (a). Show that the orthogonal matrix $U = [v \; u^{(2)} \ldots u^{(n)}]$ contains the eigenvectors of A where, again, v is as per part (a). State your spectral decomposition in terms of U.

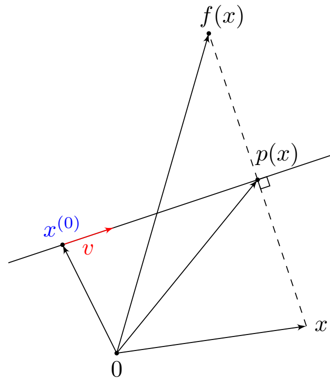

Figure 1: The point $f(x)$ is symmetric to the point x about the line $\mathcal{L}$ defined by $(x^{(0)}, v)$.

(f) Consider the spectral decomposition of A you found in part (e). Explain the structure you observe in the eigenvalue/vector pairs in terms of the geometry of the problem of finding the point symmetric to a given point about the line $\mathcal{L}$. Your explanation must connect the structure to the problem. This is *not* a generic question about the interpretation of eigenvalues/vectors.

Problem 3.5 (A lower bound on the rank)
OptM Problem 4.9 parts 1 and 2 only (not part 3).

Problem 3.6 (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.)

APPLICATION PROBLEMS

Problem 3.7 (Second-order approximation of functions)
In this exercise you extend your code from the problem "First-order approximation of functions" to second order. (As before you are welcome to use Matlab or Python or the software of your choice.) The objective is, as before, to write code to plot approximations each of (the same) three functions $f_i : \mathbb{R}^2 \to \mathbb{R}$, $i \in [3]$, but now the approximation is quadratic rather than linear. To recall, the three functions are defined pointwise as

$$\begin{aligned} f_1(x, y) &= 2x + 3y + 1, \\ f_2(x, y) &= x^2 + y^2 - xy - 5, \\ f_3(x, y) &= (x - 5)\cos(y - 5) - (y - 5)\sin(x - 5). \end{aligned}$$

For each of the above function, do the following.

(a) Write down the gradient and Hessian with respect to x and y in closed form. To recall the gradient and Hessian are compactly denoted as ∇f_i and $\nabla^2 f_i$ for $i \in [3]$ where

$$\nabla f_i = \begin{bmatrix} \frac{\partial f_i}{\partial x} \\ \frac{\partial f_i}{\partial y} \end{bmatrix} \quad \nabla^2 f_i = \begin{bmatrix} \frac{\partial^2 f_i}{\partial^2 x} & \frac{\partial^2 f_i}{\partial x \partial y} \\ \frac{\partial^2 f_i}{\partial y \partial x} & \frac{\partial^2 f_i}{\partial^2 y} \end{bmatrix}.$$

(b) For each function produce do the following two things. First, produce a 2-D contour plot indicating the level sets of each function in the range $-2 \leq x, y \leq 3.5$ and plot the direction of the gradient and tanget to the level set at the point (x,y) = 1,0). (Note, you have already produced these plots in the previous problem, we are just reproducing them here to help with visualization). Second, For the same point $(x, y) = (1, 0)$ plot the 3-D quadratic approximation of the function.

(c) Now, repeat the previous plot for point $(x, y) = (-0.7, 2)$ and $(x, y) = (2.5, -1)$.

(d) Comment on where your approximations are accurate and where they are not (if anywhere) for the three functions. Discuss what the reason is behind your observations.

Note: We recommend you design these plotting scripts as Matlab functions so that you can reuse them for to plot approximations for different non-linear functions (or for theses functions at different points). In either case make sure to attach your code.

Problem 3.8 (Google's PageRank algorithm)
In this problem you will implement and analyse approaches to the eigenvector computation that is at the heart of Google's PageRank algorithm (cf. discussion in Example 7.1 of OptM). Download `pagerank_urls.txt` and `pagerank_adj.mat` files from the course website. The first is a plain text file in which each line consists of a URL of a web page. We refer to the web pages by their

respective URL-line numbers starting from 1. For example, web page 9 is the page that the URL in line 9 points to. The URLs are of the internal web pages of Hollins University in Roanoke, Virginia. The provided data files consist of a modified version of web crawling data downloaded from `https://www.limfinity.com/ir`. The original dataset has been created in January, 2004 therefore you might notice that some of the links are inactive.

Let N be the number of URLs (therefore the number of web pages) in the first file and $i, j \in [N]$. Execute the command `load 'pagerank_adj.mat'` in MATLAB to load the content of the second file into your MATLAB workspace. You will see that a new $N \times N$ variable `J` has been imported. This variable represents an *adjacency matrix* $J \in \{0,1\}^{N \times N}$ which describes the relationships between the web pages. Specifically, the element in ith row and jth column $J_{i,j} = 1$ if there exists a link from web page j to web page i, and $J_{i,j} = 0$ otherwise. Data have been carefully filtered so that $J_{j,j} = 0$ and $\sum_{i=1}^{N} J_{i,j} > 0$ for all $j \in [N]$. In other words, we do not allow links from a web page to itself, and we do not allow for dangling pages, that is pages with no outgoing links.

Use J to obtain the *link matrix* A where

$$A_{i,j} = \frac{J_{i,j}}{\sum_{k=1}^{N} J_{k,j}}.$$

Also, let $x \in \mathbb{R}^N$ be a vector with all entries equal to 1. Use the matrix A and the vector x the same way as described in Example 7.1 to solve the following.

(a) Verify that each column in the provided matrix A sum to 1. What is the importance of this property for the Google PageRank algorithm?

(b) In the following we are consistent with the notation used in OptM. Let $x(k+1)$ be the approximation of the eigenvector in the $k+1$th iteration. We define the approximation error in the $k+1$th iteration as $e(k+1) = \|Ax(k+1) - x(k+1)\|_2$. Using MATLAB, implement the power iteration algorithm described in Section 7.1.1 in OptM. Run the algorithm for 10 iterations and plot $\log(e(k+1))$ versus k.

(c) Implement the shift-invert power iteration and Rayleigh quotient iteration algorithms presented in Sections 7.1.2 and 7.1.3 of OptM. For the shift-invert power method use $\sigma = 0.99$. In the Rayleigh quotient iteration method, use $\sigma_1 = \sigma_2 = 0.99$ for the first two iterations and $\sigma_k = \frac{x^*(k)Ax(k)}{x^*(k)x(k)}$ for $k > 2$, in a similar manner as the discussion in Example 7.1. Repeat your experiment of part (b) for these two algorithms. Plot $\log(e(k+1))$ for each of your three algorithms on a single plot. Check whether your results are consistent with Example 7.1. Include your MATLAB code with the answers.

(d) List the (page index, PageRank score) pairs of the top 5 and bottom 5 pages according to your PageRank scores. Compare them with the provided web page links and briefly explain whether the ranking seem intuitively correct.

(Note: While we write "MATLAB" in the above you are, as usual, welcome to use any software you would like to, just not to use built-in functions that accomplish the objective.)

OPTIONAL PROBLEMS

Problem 3.9 (Eigenvectors and diagonalizing circulant matrices)
A square matrix $\mathbf{A}$ is "circulant" if

$$\mathbf{A} = \begin{bmatrix} a_0 & a_{N-1} & a_{N-2} & a_{N-3} & \dots & a_1 \\ a_1 & a_0 & a_{N-1} & a_{N-2} & \dots & a_2 \\ a_2 & a_1 & a_0 & a_{N-1} & \dots & a_3 \\ a_3 & a_2 & a_1 & a_0 & \dots & a_4 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{N-1} & a_{N-2} & a_{N-3} & a_{N-4} & \dots & a_0 \end{bmatrix}. \tag{1}$$

In other words each column is a circularly-shifted version of the column to its left, where the shift is downwards by one.

The (ij)th entry of the square N-dimensional Fourier matrix is defined as

$$[\mathbf{F}]_{ij} = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N}(i-1)(j-1)}.$$

The relationship between a length-N vector x and its length-N discrete-time Fourier series (DTFS) $\tilde{x}$ is given by

$$\tilde{x} = \mathbf{F}^H x \qquad \text{and} \qquad x = \mathbf{F}\tilde{x},$$

where $\mathbf{F}^H$ is the Hermitian-transpose of $\mathbf{F}$. That is, $[\mathbf{F}^H]_{ij} = [\mathbf{F}]^*_{ji}$ where $(\cdot)^*$ denotes the complex-conjugate of the scalar argument.

In this problem you prove that the Fourier matrix diagonalizes all circulant matrices.

(a) Denote the DTFS of the vector $a = [a_0\ a_1 \dots a_{N-1}]^T$ as $\tilde{a} = [\tilde{a}_0\ \tilde{a}_1\ \dots\ \tilde{a}_{N-1}]^T$. Define the matrix $\tilde{\mathbf{A}}$ as

$$\tilde{\mathbf{A}} = \mathbf{F}^H \mathbf{A}.$$

Write out the matrix $\tilde{\mathbf{A}}$ in terms of $\{\tilde{a}_0, \tilde{a}_1, \dots \tilde{a}_{N-1}\}$ (or in terms of $\tilde{a}$).

Hint: The first column of $\tilde{\mathbf{A}}$ is simply given by $\tilde{a} = \mathbf{F}^H a$. To find the second (and other) column(s) it helps (though is not necessary) to recall some Fourier properties.

(b) Using the matrix $\tilde{\mathbf{A}}$ from above, compute the matrix product $\mathbf{B}$ where

$$\mathbf{B} = \mathbf{F}^H \mathbf{A} \mathbf{F} = \tilde{\mathbf{A}} \mathbf{F}.$$

(c) What are the eigenvalues of $\mathbf{A}$? What are its eigenvectors?

