EECS 127/227AT: Optimization Models in Engineering 10/03/2019
A. Bayen

Midterm 1

NAME:

SID:

Instruction:

1. The exam lasts for 1h 20mins.
2. The maximum score is 40.
3. Notes are *not* allowed, except for a one-page, two sided cheat sheet.
4. Do not open the exam until you are told to do so.
5. TO HAVE A CLEAR AND VISIBLE SCAN OF THE EXAM, DO NOT WRITE ON THE BACK OF THE EXAM PAGES.
6. **Only answers written inside the box provided after each question will be graded**.

The breakdown of points is as follows.

Question	a		b		c		d	Total
Number	i	ii	i	ii	i	ii		
1	2	-	2	3	2	2	4	15
2	3	2	3	2	-	-	-	10
3	2	-	2	-	2	-	-	6
4	3	-	4	-	2	-	-	9

1. *Properties of Matrices.* In this problem we consider some interesting matrix properties.

(a) Show that any square matrix P can be written as $P = A + B$, where A is symmetric (i.e. $A^\top = A$) and B is skew-symmetric (i.e. $B^\top = -B$). Explicitly state A and B. [2 pts]

Solution: $A = \frac{1}{2}\left(P + P^\top\right)$ and $B = \frac{1}{2}\left(P - P^\top\right)$

(b) i. For a square matrix P, the characteristic polynomial, denoted by $f(t)$, is defined by $f(t) = \det(tI - P)$, where I is the identity matrix of the same size as P. Write the characteristic polynomial for $P = \begin{bmatrix} 2 & -3 \\ 1 & 4 \end{bmatrix}$ explicitly. [2 pts]

Solution: We have the characteristic polynomial as,

$$\begin{aligned} f(t) &= \det\left(\begin{bmatrix} t & 0 \\ 0 & t \end{bmatrix} - \begin{bmatrix} 2 & -3 \\ 1 & 4 \end{bmatrix}\right) \\ &= \det\left(\begin{bmatrix} t-2 & 3 \\ -1 & t-4 \end{bmatrix}\right) \\ &= t^2 - 6t + 11 \end{aligned}$$

ii. The Cayley-Hamilton theorem states that every square matrix (P) satisfies its own characteristic equation. I.e. if the characteristic polynomial is $f(t) = a_n t^n + a_{n-1} t^{n-1} + \ldots + a_0$, then the equation $f(P) = a_n P^n + a_{n-1} P^{n-1} + \ldots + a_0 I = 0$ holds. Use the above theorem to find the inverse of the matrix P as given in Part (i). [3 pts]

Solution: The determinant of P is non-zero, so P is invertible. As per CHT, P satisfies its own characteristic equation. So,

$$f(P) = P^2 - 6P + 11I = 0$$
$$P^2 = 6P - 11I$$

postmultiplying P^{-1} on both sides,

$$P = 6I - 11P^{-1}$$
$$P^{-1} = \frac{-1}{11}(P - 6I)$$
$$P^{-1} = \frac{-1}{11}\begin{bmatrix} -4 & -3 \\ 1 & -2 \end{bmatrix}$$

(c) A square matrix P is called *Nilpotent* if $P^k = 0$ for some positive integer k. Note that the lower powers of P, such as $P, P^2, \cdots, P^{k-1}$ might not be 0.

i. Prove that all the eigenvalues of a *nilpotent* matrix P are 0. [2 pts]

Solution: Consider any eigenvalue λ of P. Then P^k has an eigenvalue λ^k. Let v be the corresponding eigenvector of P^k, then

$$P^k v = \lambda^k v = 0 \quad (since\ P^k = 0)$$
$$\Rightarrow \lambda^k = 0$$
$$\Rightarrow \lambda = 0$$

So, all the eigenvalues of P are 0.

ii. Using the result in Part (i) above, prove that $\det(P + I) = 1$, where P is a *nilpotent* matrix and I is an identity matrix of the same dimensions as P. [2 pts]

Solution: If u is the eigenvector of $P + I$ corresponding to an eigenvalue λ, then

$$\begin{aligned}(P + I)u &= \lambda u \\ Pu + u &= \lambda u \\ Pu &= (\lambda - 1)u\end{aligned}$$

So, if λ is an eigenvalue of $P + I$ then $\lambda - 1$ is an eigenvalue of P. The above result holds for any square matrix. Since all the eigen values of a *nilpotent* matrix P is 0, all the eigen values of $P + I$ are equal to 1. Determinant is the product of eigenvalues, so the determinant of $P + I$ is 1.

(d) State if the following statement is TRUE or FALSE with proper reasoning, i.e. if your answer is "TRUE ", provide a proof for the statement and if your answer is "FALSE ", provide a counter-example.
Statement: If two matrices P and Q are Positive Semi-Definite (PSD), then $Trace(PQ) = 0$ **if and only if** $PQ = 0$. [4 pts]
Hint: For every PSD matrix A, there exists a matrix B such that $A = BB^\top$

Solution: TRUE

Proof for $Trace(PQ) = 0$ if $PQ = 0$:

Since P and Q are positive semi-definite, they can be written as:

$$P = A^T A \quad and \quad Q = BB^T, \text{ for some matrices } A \text{ and } B$$

$$\begin{aligned} Trace(PQ) = Trace(A^T ABB^T) = Trace(B^T A^T AB) &= Trace((AB)^T(AB)) \\ &= \|AB\|_F^2 \end{aligned}$$

where $\|AB\|_F$ is the frobenius norm of AB.

To have $Trace(PQ) = 0$, $\|AB\|_F^2$ has to be 0, i.e. $AB = 0$ and so $PQ = 0$

Proof for $Trace(PQ) = 0$ only if $PQ = 0$:

If $PQ = 0$, then the trace, which is sum of diagonal elements of PQ is 0.

2. *Singular Value and Eigen Value Decomposition*

(a) i. Consider the 2×2 matrix

$$P = \frac{1}{5}\begin{pmatrix} 8 \\ 6 \end{pmatrix}\begin{pmatrix} 1 & 0 \end{pmatrix} + \frac{2}{5}\begin{pmatrix} -6 \\ 8 \end{pmatrix}\begin{pmatrix} 0 & 1 \end{pmatrix}.$$

Express the SVD of P as $U\Sigma V^\top$, with Σ the diagonal matrix of singular values ordered in decreasing fashion. Make sure to check all the properties required for U, Σ, V. [3 pts]

Solution: We have

$$P = \sigma_1 u_1 v_1^\top + \sigma_2 u_2 v_2^\top = USV^\top,$$

where $U = [u_1, u_2]$, $V = [v_1, v_2]$ and $\Sigma = \text{diag}(\sigma_1, \sigma_2)$, with $\sigma_1 = 4$, $\sigma_2 = 2$, and

$$u_1 = \frac{1}{10}\begin{pmatrix} -6 \\ 8 \end{pmatrix}, \quad u_2 = \frac{1}{10}\begin{pmatrix} 8 \\ 6 \end{pmatrix}, \quad v_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The triplet (U, Σ, V) is an SVD of A, since Σ is diagonal with non-negative elements on the diagonal, and U, V are orthogonal matrices ($U^\top U = V^\top V = I_2$). To check this, we first check that the Euclidean norm of u_1, u_2, v_1, v_2 is one. In addition, $u_1^\top u_2 = v_1^\top v_2 = 0$. Thus, U, V are orthogonal, as claimed.

ii. Given a matrix P and its Singular Value Decomposition as:

$$P = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}}_{\text{U}} \underbrace{\begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{18}} & \frac{-1}{\sqrt{18}} & \frac{4}{\sqrt{18}} \\ \frac{2}{3} & \frac{-2}{3} & \frac{-1}{3} \end{bmatrix}}_{\text{V}^\top}$$

A. What is the dimension of the null-space of P? [1 pt]

Solution: As one of the columns in Σ is all 0, Dimension of the null-space of P = 1.

B. Provide an orthonormal basis for the range of P [1 pt]

Solution: The columns of U constitute an orthonormal basis for the range of P.

(b) i. Prove that in the case of matrices for which both singular values and eigenvalues are computable, the largest singular value is greater than or equal to all the eigenvalues.[3 pts]

Solution: Both values are computable for a square matrix. Let us have λ_1 as the largest eigen value of P (a square matrix) and corresponding eigen vector be v_1. Let the largest singular value of P be σ_1. We have,

$$Pv_1 = \lambda_1 v_1 \tag{1}$$

taking transpose on both sides,

$$v_1^\top P^\top = \lambda_1 v_1^\top \tag{2}$$

Pre-multiplying equation 2 to equation 1, we have,

$$\begin{aligned}
&v_1^\top P^\top P v_1 = \lambda_1^2 v_1^\top v_1 \\
&\Rightarrow \lambda_1^2 \|v_1\|_2^2 = v_1^\top P^\top P v_1 \leq \lambda_{max}\left(P^\top P\right)\|v_1\|_2^2 \\
&\Rightarrow \lambda_1^2 \leq \lambda_{max}\left(P^\top P\right) = \sigma_1^2 \\
&\Rightarrow |\lambda_1| \leq \sigma_1
\end{aligned}$$

Since the largest singular value is greater than the largest eigen value, it is proved that it is greater than all the eigen values.

ii. Provide a geometrical interpretation of the proposition given in part (i) and justify it. [2 pts]

Solution: Eigen values of any matrix are the amounts of scaling that is achieved when the matrix as a linear map acts on particular vectors (eigen vector) and no change in direction (except 180° rotation in case of negative eigen values) of the vectors occur. Whereas, the maximum singular value is the maximum amount of stretch that can happen when the matrix as a linear map acts on any vector. As maximum singular value is the maximum amount of stretch that can occur, it is greater than all the eigen values.

3. *Solution of Linear Equations.* Consider the set $\mathcal{P} = \{x : Px = q\}$
Where $P \in \mathbb{R}^{n \times m}$ and $q \in \mathbb{R}^n$ are given.

(a) Give a criterion for the set $\mathcal{P}$ to be a non empty-set, i.e. a solution exists for $Px = q$? [2 pts]

Solution: q lies in the range of P.

(b) Assuming a solution exists, give a criterion for existence of an unique solution for $Px = q$? [2 pts]

Solution: The columns of P should be linearly independent of each other.

(c) Give a criterion for existence of an unique solution for $Px = q$ in terms of the matrices U,Σ and V in the full singular value decomposition of $P = U\Sigma V^\top$? [2 pts]

Solution: Σ has full-column rank.

4. *Polynomial Regression.* You are the admin of the popular Facebook group "UC Berkeley Memes For Edgy Teens". To be well-prepared for the high load of memes expected to be posted during the upcoming finals season, you plan to hire few interns. You are confused on how many interns to hire, so you first want to get an estimate of work load by studying the trend in number of memes posted over time, using data from last year's fall semester. The number of memes posted is estimated to vary with time (from beginning of semester) as per the following polynomial:

$$\begin{aligned} \hat{y}_i &= \alpha_0 + \alpha_1 t_i + \alpha_2 t_i^2 + \cdots + \alpha_{m-1} t_i^{m-1} \quad (i = 1, 2, \cdots, n \text{ and } n > m) \\ \text{where } \hat{y}_i &= \text{the estimate of the actual total number of memes } y_i\text{, posted till } i^{th} \text{ observation} \\ t_i &= \text{the time at } i^{th} \text{ observation} \\ \alpha_0, \alpha_1, \cdots, \alpha_{m-1} &= \text{the regression coefficients} \end{aligned}$$

The aim is to minimize the squared norm of the residual ($\|y - \hat{y}\|_2^2$) using least square method to determine the best regression coefficients $\alpha_0, \alpha_1, \alpha_2, \cdots, \alpha_{m-1}$.

(a) Model the problem into the least squares residual equation,

$$r \ \text{(residual)} = b - Ax$$

Specifically, write the matrices/vectors b, A and x explicitly using the notations provided in the problem.[3 pts]

Solution:

$$A = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{m-1} \\ 1 & t_2 & t_2^2 & \cdots & t_2^{m-1} \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ 1 & t_n & t_n^2 & \cdots & t_n^{m-1} \end{bmatrix}, b = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{bmatrix}, x = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ . \\ . \\ \alpha_{m-1} \end{bmatrix}$$

Matrices such as A which have the terms of a geometric progression in each row are called Vandermonde matrix, named after French mathematician Alexandre-Théophile Vandermonde (c. 1735 – 1796) (Refer to the Figure below)

(b) To solve the least squares equation, we will use Gram-Schmidt Orthonormalization (GSO). Using GSO, the matrix can be represented as,

$$A = QR = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

where, Q is an orthogonal matrix and R is an upper-triangular matrix. Let Q and R be conformably partitioned (i.e. $Q_1 \in \mathbb{R}^{n\times m}$, $Q_2 \in \mathbb{R}^{n\times(n-m)}$ and $R_1 \in \mathbb{R}^{m\times m}$), such that R_1 is upper triangular and invertible. Show that the squared norm of the residual $\|r\|_2^2$ can be expressed as,

$$\|r\|_2^2 = \|c_1 - R_1x\|_2^2 + \|c_2\|_2^2$$

where you will find the expression for c_1 and c_2 in terms of Q_1,Q_2 and b.[4 pts]

Solution: We have,

$$\begin{aligned} r &= b - Ax \\ r &= b - QRx \\ \|r\|_2^2 &= \left\| b - Q \begin{bmatrix} R_1 \\ 0 \end{bmatrix} x \right\|_2^2 \end{aligned}$$

Since multiplying by an orthogonal matrix does not change the l_2-norm of a vector,

$$\begin{aligned} &= \left\| Q^\top \left(b - Q \begin{bmatrix} R_1 \\ 0 \end{bmatrix} x \right) \right\|_2^2 \\ &= \left\| \begin{bmatrix} Q_1^\top b \\ Q_2^\top b \end{bmatrix} - \begin{bmatrix} R_1 x \\ 0 \end{bmatrix} \right\|_2^2 \\ &= \left\| \begin{bmatrix} Q_1^\top b - R_1 x \\ Q_2^\top b \end{bmatrix} \right\|_2^2 \\ &= \left\| \left(Q_1^\top b - R_1 x \right) \right\|_2^2 + \left\| Q_2^\top b \right\|_2^2 \end{aligned}$$

If we take, $c_1 = Q_1^\top b$ and $c_2 = Q_2^\top b$, $\|r\|_2^2 = \|c_1 - R_1x\|_2^2 + \|c_2\|_2^2$

(c) Find an expression of x such that the squared norm of the residual as given in Part (ii) is minimized? Express it in terms of Q_1,Q_2, R_1 and b. The expression you find for x is the solution to the polynomial regression problem. [2 pts]

Solution: To solve for x, we need to minimize $\|r\|_2^2 = \|c_1 - R_1 x\|_2^2 + \|c_2\|_2^2$.
$\|r\|_2^2$ is minimized when $c_1 - R_1 x = 0$,
so $x = R_1^{-1} c_1 = R_1^{-1} Q_1^\top b$

The partitioning of A in the pattern as mentioned carries geometrical meaning. Q_1 constitutes an orthonormal basis for $\mathcal{R}(A)$ and Q_2 constitutes an orthonormal basis for $\mathcal{R}(A)^\perp$. After optimization, the residual obtained is $\|c_2\|_2^2$ which is $\|Q_2^\top b\|_2^2$, which corresponds to the distance between b and the orthogonal projection of b on $\mathcal{R}(A)$.