18.06 Professor Edelman Quiz 2 April 6, 2018

Your PRINTED name is: ____________

Please circle your recitation:

(1)	T 10	26-328	D. Kubrak
(2)	T 11	26-328	D. Kubrak
(3)	T 12	4-159	P.B. Alvarez
(7)	T 12	4-153	E. Belmont
(4)	T 1	4-149	P.B. Alvarez
(5)	T 2	4-149	E. Belmont
(6)	T 3	4-261	J. Wang

Note: We are not planning to use gradescope for this exam.

Grading

1	
2	
3	
4	
Total:	

Your Initials: ____________

1 (25 pts.) Find the QR decomposition (Q is 4×2, R is 2×2 upper triangular) of

$$A = \begin{pmatrix} 1 & a \\ 1 & b \\ 1 & c \\ 1 & d \end{pmatrix} \text{ in terms of } \mu = \tfrac{a+b+c+d}{4},$$

the mean of the second column and the elements of A.

Lets call the 2 column vectors v_1 and v_2. We apply Gram-Schmidt. To start we normalize the first vector v_1 to get

$$q_1 = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

For the next step we need to first take away the projection of v_2 onto the space spanned by q_1 to get

$$q_2' = v_2 - (q_1 \cdot v_2)q_1 = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} - \frac{a+b+c+d}{4}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} a-\mu \\ b-\mu \\ c-\mu \\ d-\mu \end{pmatrix}$$

To get the second orthonormal element we need to normalize this to get

$$q_2 = \frac{1}{x}\begin{pmatrix} a-\mu \\ b-\mu \\ c-\mu \\ d-\mu \end{pmatrix}$$

where $x = \sqrt{a^2 + b^2 + c^2 + d^2 - 4\mu^2}$. Thus

$$Q = \begin{pmatrix} \frac{1}{2} & \frac{a-\mu}{x} \\ \frac{1}{2} & \frac{b-\mu}{x} \\ \frac{1}{2} & \frac{c-\mu}{x} \\ \frac{1}{2} & \frac{d-\mu}{x} \end{pmatrix}$$

To find R we can either track the previous operations or compute $R = Q^T A$ to get

$$R = \begin{pmatrix} 2 & 2\mu \\ 0 & x \end{pmatrix}$$

Your Initials: ____________

2 (20 pts.) An experimenter has data in the form of pairs (x_i, y_i) for $i = 1, \ldots, n$ where the x_i are distinct and positive. Given the matrix

$$A = \begin{pmatrix} \sin(x_1) & e^{x_1} & \sqrt{x_1} \\ \sin(x_2) & e^{x_2} & \sqrt{x_2} \\ \vdots & \vdots & \vdots \\ \sin(x_n) & e^{x_n} & \sqrt{x_n} \end{pmatrix},$$

suggest a method for computing the best fit function of the form $f(x) = C\sin(x) + De^x + E\sqrt{x}$ through the n points. In what precise sense is your answer a best fit?

Solution: Let $v = \begin{pmatrix} C \\ D \\ E \end{pmatrix}$ and $y = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix}$. The vector Av will then be nothing but $\begin{pmatrix} f(x_1) \\ f(x_2) \\ \cdots \\ f(x_n) \end{pmatrix}$ where $f(x) = C\sin(x) + De^x + E\sqrt{x}$. So to find f such that $f(x_i) = y_i$ for all i we need to solve $Av = y$.

If there is no such solution (which is very possible since n is arbitrary and probably greater than 3) we can do the method of least squares, namely try to find v such that $||Av - y||$ is minimal possible. Assuming columns of A are linearly independent, this is done by solving the equation $A^TAv = A^Ty$, or in other words $v = (A^TA)^{-1}A^Ty$ (indeed, then $Av = A(A^TA)^{-1}A^Ty$ will be the projection of y on the column space of A and so for this v the distance $||Av - y||$ will be minimal). Also to minimize $||Av - y||$ is the same as to minimize $||Av - y||^2$ and $||Av - y||^2 = \sum_{i=1}^n (C\sin(x_i) + De^{x_i} + E\sqrt{x_i} - y_i)^2$, so such $f(x) = C\sin(x) + De^x + E\sqrt{x}$ (where $v = \begin{pmatrix} C \\ D \\ E \end{pmatrix}$ is $(A^TA)^{-1}A^Ty$) will be the best fit in a sense that the sum of squares of differences $\sum_{i=1}^n (C\sin(x_i) + De^{x_i} + E\sqrt{x_i} - y_i)^2$ will be minimal.

Your Initials: ____________

3 (20 pts.)

A form of the singular value decomposition of a rank r, $m \times n$ matrix A is $U\Sigma V^T$ where Σ is square r by r with positive diagonal entries, U is $m \times r$ and V is $n \times r$. Write down projection matrices for the four fundamental subspaces of A, in terms of one of U, Σ, or V in each expression. Be sure to clearly identify which fundamental subspace of A goes with which projection matrix.

The four fundamental subspaces are $C(A)$, $C(A^T)$, $N(A)$, and $N(A^T)$.

In problem set 4 we showed that $C(A) = C(U)$ in this case. Since U has orthonormal columns, the projection matrix to $C(U)$ is UU^T.

Applying transpose to the decomposition $A = U\Sigma V^T$ gives $A^T = (U\Sigma V^T) = V\Sigma^T U^T = V\Sigma U^T$. By the same reasoning as for $C(A)$, we have $C(A^T) = C(V)$, and the projection matrix to $C(V)$ is VV^T.

If P is a projection matrix to a subspace V, the projection matrix to $V^\perp$ is $I - P$. By the main facts about the fundamental subspaces, $C(A)^\perp = N(A^T)$ and $C(A^T)^\perp = N(A)$, so the projection matrix to $N(A^T)$ is $I - UU^T$ and the projection matrix to $N(A)$ is $I - VV^T$. To summarize:

Subspace	Projection matrix
$C(A)$	UU^T
$C(A^T)$	VV^T
$N(A^T)$	$I - UU^T$
$N(A)$	$I - VV^T$

Warning: A lot of people tried solving this using the projection formula $A(A^TA)^{-1}A^T$ (for projection to $C(A)$) directly. This formula assumes that the columns of A are linearly independent! In particular, the middle bit A^TA won't be invertible otherwise. We don't run into this problem when applying this to U, because U has orthonormal columns.

Your Initials: ____________

4 (35 pts.)

Let $d(A)$ be a scalar function of 3×2 matrices A with the following properties:

α) If you interchange the two columns of A, $d(A)$ flips sign.

β) $d(A)$ is linear in each of the columns of A.

γ) $d(A)$ is non-zero for at least one 3×2 A.

a. (5 pts.) What is $d(2A)$ in terms of $d(A)$?

If the columns of A are denoted A_1 and A_2, then

$$d(2A) = d\left(\begin{bmatrix} 2A_1 & 2A_2 \end{bmatrix}\right) = 2d\left(\begin{bmatrix} A_1 & 2A_2 \end{bmatrix}\right) = 4d\left(\begin{bmatrix} A_1 & A_2 \end{bmatrix}\right) = 4d(A),$$

by the linearity in columns.

b. (10 pts.) Give an example $d(A)$ that satisfies the three requirements of this question.

Let the matrix A be denoted

$$A = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}.$$

Then, one possible function is $d(A) = ad - bc$, the determinant of the top 2×2 block. The three conditions hold essentially because they hold for 2×2 matrices. We can also check explicitly that the three conditions hold:

α) $d\left(\begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}\right) = ad - bc, \qquad \left(\begin{bmatrix} b & a \\ d & c \\ f & e \end{bmatrix}\right) = bc - ad = -(ad - bc).$

β)

$$\begin{aligned} d\left(\begin{bmatrix} a_1 + \lambda a_2 & b \\ c_1 + \lambda c_2 & d \\ e_1 + \lambda e_2 & f \end{bmatrix}\right) &= (a_1 + \lambda a_2)d - b(c_1 + \lambda c_2) \\ &= (a_1 d - bc_1) + \lambda(a_2 d - bc_2) \\ &= d\left(\begin{bmatrix} a_1 & b \\ c_1 & d \\ e_1 & f \end{bmatrix}\right) + \lambda d\left(\begin{bmatrix} a_2 & b \\ c_2 & d \\ e_2 & f \end{bmatrix}\right). \end{aligned}$$

Similarly, linearity in the second column holds.

γ) For example,

$$d\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}\right) = 1.$$

Your Initials: ____________

c. (10 pts.) We recall that the determinant of square matrices is linear in each column and each row of the square matrix. Can property β be extended to rows and columns of 3×2 matrices A to create a $d(A)$ with the three requirements of this question? If yes, give an example, if not, why not?

Linearity in columns gave us that $d(2A) = 4d(A)$ for all matrices A. By similar reasoning, linearity in rows would give us $d(2A) = 8d(A)$, since there are three rows. But then, for every matrix A, we would have $8d(A) = 4d(A)$, which implies that $d(A) = 0$ for all matrices. But this contradicts condition γ. Therefore, property β cannot be extended to the rows of A.

d. (10 pts.) If we discard property γ to allow the "zero" function, the set of all functions $d(A)$ satisfying α and β form a three dimensional vector space. Describe explicitly this vector space of functions in terms of the elements of A.

Remark: There are a number of thoughts that would guide students to a general understanding of this problem. One thought is that augmenting the matrix with a column of three variables c_1, c_2, c_3 gives rise to a 3x3 determinant which one can expand in cofactors (which gives the entire three dimensional space!). Essentially the same thought is behind the cross product in three dimensions. One can more simply delete any of the three rows and notice that we would have all the requirements for the 2x2 determinant.

If we let our matrix A be denoted as

$$A = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix},$$

then the three dimensional vector space of functions $d(A)$ satisfying the conditions α and β have a basis

$$\begin{cases} d_1(A) = ad - bc \\ d_2(A) = af - be \\ d_3(A) = cf - de. \end{cases}$$

Each of these satisfies conditions α and β by similar reasoning as in part b. Furthermore, they are linearly independent.

One way to prove explicitly that these three functions span the whole vector space of possible functions $d(A)$ is to use linearity in the columns to split apart $d(A)$ as

$$d\left(\begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}\right) = d\left(\begin{bmatrix} a & b \\ 0 & d \\ 0 & f \end{bmatrix}\right) + d\left(\begin{bmatrix} 0 & b \\ c & d \\ 0 & f \end{bmatrix}\right) + d\left(\begin{bmatrix} 0 & b \\ 0 & d \\ e & f \end{bmatrix}\right)$$
$$= d\left(\begin{bmatrix} a & b \\ 0 & 0 \\ 0 & 0 \end{bmatrix}\right) + d\left(\begin{bmatrix} a & 0 \\ 0 & d \\ 0 & 0 \end{bmatrix}\right) + d\left(\begin{bmatrix} a & 0 \\ 0 & 0 \\ 0 & f \end{bmatrix}\right) + \ldots + d\left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ e & f \end{bmatrix}\right),$$

where in the last line, we have nine terms corresponding to the nine possible ways to choose one entry in the first column to be nonzero and one entry in the second column to be nonzero. For any entry that has two nonzero elements in the same row, we can conclude that d of that matrix must be zero. For example,

$$d\left(\begin{bmatrix} a & b \\ 0 & 0 \\ 0 & 0 \end{bmatrix}\right) = ab \times d\left(\begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}\right) = -ab \times d\left(\begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}\right),$$

by property α. But this implies that

$$d\left(\begin{bmatrix} a & b \\ 0 & 0 \\ 0 & 0 \end{bmatrix}\right) = 0.$$

By the row swap property, we also have that

$$d\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}\right) = -d\left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}\right),$$

and similarly for any matrix where there are two ones in different columns and different rows.

This implies that

$$d\left(\begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}\right) = (ad-bc)\times d\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}\right) + (af-be)\times d\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}\right) + (cf-de)\times d\left(\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}\right).$$

Choosing the three determinants on the right hand side to be arbitrary real numbers exactly tells us that the determinant function $d(A)$ must be a linear combination of our three functions $d_1(A), d_2(A)$, and $d_3(A)$.