Jae Hong Lee
Prof. C.E. Tsourakakis
Boston University

April 2, 2024

— **CAS CS 365 - Assignment 2**

1 Linear Algebra

1.1 Problem 1.1

Assume that A is a square n x n invertible matrix, $u, v \in \mathbb{R}^2$ and $1 + v^T A^{-1} u \neq 0$. Verify the Sherman-Morrison formula.

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1} u}$$

In order to make the left side I, multiply inverse of $(A + uv^T)$ both side.

$$(A + uv^T) \cdot (A + uv^T)^{-1} = (A + uv^T) \cdot (A^{-1} - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1} u})$$
$$I = (A + uv^T) \cdot A^{-1} - (A + uv^T) \cdot (\frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1} u})$$
$$I = I + uv^T A^{-1} - \frac{uv^T A^{-1} + uv^T A^{-1} uv^T A^{-1}}{1 + v^T A^{-1} u}$$
$$I = I + uv^T A^{-1} - \frac{u(1 + v^T A^{-1} u)v^T A^{-1}}{1 + v^T A^{-1} u}$$
$$I = I + uv^T A^{-1} - uv^T A^{-1} = I$$

Both sides are I so I verified the Sherman-Morrison formula.

1.2 Problem 1.2

Assume that $\lambda_i, i = 1, ..., n$ are the eigenvalues of A. What are the eigenvalues of 2A + 3I where I is the n x n identity matrix?

λ_i for $i = 1, ..., n$, the A's eigenvalue equation is $Ax = \lambda_i x$. I can modify both side of this equation to make A to $2A + 3I$.

$$Ax = \lambda_i x$$
$$2Ax = 2\lambda_i x$$
$$2Ax + 3Ix = 2\lambda_i x + 3Ix$$
$$(2A + 3I)x = (2\lambda_i + 3)x$$

Now I showed the eigenvalues of $(2A + 3I)$ is $(2\lambda_i + 3)$ for $i = 1, ..., n$

1.3 Problem 1.3

Check whethere the vectors $v_1 = \begin{bmatrix}1\\2\\3\end{bmatrix}, v_2 = \begin{bmatrix}4\\5\\6\end{bmatrix}$**, and** $v_3 = \begin{bmatrix}7\\8\\9\end{bmatrix}$ **are linearly independent. If they are not, write one of the vectors as a combination of the other two. Determine the rank of the matrix that has these vectors as its columns. Also, find the determinant of the matrix.**

In order to check whether these three vectors v_1, v_2, v_3 are linearly independent or not, Let's make row echelon form of whose columns are the three vectors.

$$\begin{bmatrix}1 & 4 & 7\\2 & 5 & 8\\3 & 6 & 9\end{bmatrix}$$

(1) multiply 2 to the first row and subtract second row. $\begin{bmatrix}1 & 4 & 7\\0 & 3 & 6\\3 & 6 & 9\end{bmatrix}$

(2) multiply 3 to the first row and subtract third row. $\begin{bmatrix}1 & 4 & 7\\0 & 3 & 6\\0 & 6 & 12\end{bmatrix}$

(3) multiply 2 to the second row and subtract third row. $\begin{bmatrix}1 & 4 & 7\\0 & 3 & 6\\0 & 0 & 0\end{bmatrix}$

Third columns of the row echelon form did not contain a leading 1, so the original vectors would not be linearly independent.

I can see $2 \cdot v_2 = v_1 + v_3$**.**

Since the first and second row of row echelon form has a leading 1, this matrix has the rank of 2.

The determinant equation of 3×3 matrix $\begin{bmatrix}a & b & c\\d & e & f\\g & h & i\end{bmatrix}$ is $a \cdot (ei - hf) - b \cdot (di - gf) + c \cdot (dh - ge)$. So the determinant of $\begin{bmatrix}1 & 4 & 7\\2 & 5 & 8\\3 & 6 & 9\end{bmatrix}$ is $1 \cdot (5 \cdot 9 - 6 \cdot 6) - 4 \cdot (2 \cdot 9 - 3 \cdot 8) + 7 \cdot (2 \cdot 6 - 3 \cdot 5) = 1 \cdot (45 - 36) - 4 \cdot (18 - 24) + 7 \cdot (12 - 15)$ $= 1 \cdot (-9) - 4 \cdot (-6) + 7 \cdot (-3) = -9 + 24 - 21 = 0$.

Determinant is 0

1.4 Problem 1.4

Recall that the range space or image of $A \in R^{m \times n}$ **is the set**

$$im(A) = \{y \in R^m : y = Ax \; for \; some \; x \in R^n\}.$$

The rank is defined as the dimension of this spaces, i.e.,

$$rank(A) = dim \; im(A).$$

(a) For $A \in R^{m \times n}$ **and** $B \in R^{n \times p}$**, show that**

$$im(AB) \subseteq im(A)$$

Using the given definition of range space and rank, I know the image of AB and image of B is the set

$$im(AB) = \{z \in R^m : z = ABt \; for \; some \; t \in R^p\}.$$

Since $t \in R^p$, t is $p \times 1$ vector. so Bt is dot product of $n \times p \cdot p \times 1$ which results in $n \times 1$ vector. From the definition of $im(A)$, $x \in R^n$ vector. so the dot product of B and tm Bt is subset of x, $Bt \in x$. **So** $im(AB) \subseteq im(A)$

(b) For $A, B \in R^{n \times n}$, show that

$$rank(AB) \leq min\{rank(A), rank(B)\},$$

$$rank(A + B) \leq rank(A) + rank(B).$$

(1) $rank(AB) \leq min\{rank(A), rank(B)\}$

Let's say the Range space spanned by the columns of AB is the space R of all vectors r that can be written as the linear combinations of the columns of AB.

$$r = (AB)x,\ for\ some\ x \in R^n$$

x is the $n \times 1$ vector. I can rewrite this to

$$r = A(Bx)$$

Bx is an $n \times 1$ vector. So any vector $r \in R$ can be written as a linear combination of the columns of A with coefficients taken from the vector Bx. Therefore, the range space R is not bigger than the span of the columns of A, which dimension is rank(A). This underlies that the dimension of R is less than or equal to rank(A). Since the dimension of R is the rank of AB, rank(AB) $\leq$ rank(A).

The other way works the same. now, the row range space spanned by the the rows of AB is the space S of all vectors s that can be written as linear combinations of the rows of AB.

$$s = x(AB),\ for\ some\ x \in R^{1 \times n}$$

x is the $1 \times n$ vector of coefficients of the linear combination. I can rewrite

$$s = (xA)B$$

where xA is a $1 \times n$ vector. So any vector $s \in S$ can be written as a linear combination of the rows of B with coefficient taken from the vector xA. Therefore the space S is not bigger than the span of the rows of B, which dimension is rank(B). This says the dimension of S is less than equal to rank(B). So the dimension of S is the rank of AB, I have $rank(AB) \leq rank(B)$

Combining two results: rank(AB) $\leq$ rank(A) and rank(AB) $\leq$ rank(B), the result is rank(AB) $\leq$ min(rank(A), rank(B))

(2) $rank(A + B) \leq rank(A) + rank(B)$

There are two situation with addition of $rank(A + B)$ of two matrices A and B.

First, when the range space of A and B is linearly independent and second, A and B are linearly dependent. When A and B is linearly independent, $A + B$ makes the whole plane. If there is the other axis, z, which is orthogonal to A and B plane, then z axis cannot be represented by $Rank(A + B)$. $Rank(A + B)$ cannot be greater than $Rank(A) + Rank(B)$. They are equal.

Second, When A and B are linearly dependent. Then A and B are spanned by each others. So A and B would make x and y axis. so the rank of $A + B$ would be smaller than $Rank(A) + Rank(B)$.

Therefore, overall $Rank(A + B) \leq Rank(A) + Rank(B)$

2 Optimization

2.1 Derivatives

Compute the derivative $\frac{df}{dx}$ for the following functions. It will be helpful to identify n, m where $f : \mathbb{R}^n \to \mathbb{R}^m$, and the dimensions of the derivative first.

2.1.1 (a) $f(x) = \frac{1}{1+e^{-x}}, x \in \mathbb{R}$

$f(x)^{'} = (\frac{1}{1+e^{-x}})^{'} = -\frac{1}{(1+e^{-x})^2} \cdot (1+e^{-x})^{'} = -\frac{1}{(1+e^{-x})^2} \cdot -1 \cdot e^{-x} = \frac{e^{-x}}{(1+e^{-x})^2}$

2.1.2 (b) $f(x) = exp(-\frac{1}{2\sigma^2}(x-\mu)^2), x \in \mathbb{R}$

$f(x)^{'} = (exp(-\frac{1}{2\sigma^2}(x-\mu)^2))^{'} = exp(-\frac{1}{2\sigma^2}(x-\mu)^2) \cdot (-\frac{1}{2\sigma^2}(x-\mu)^2)^{'} = exp(-\frac{1}{2\sigma^2}(x-\mu)^2) \cdot (-\frac{1}{2\sigma^2}(x-\mu)^2)^{'} = exp(-\frac{1}{2\sigma^2}(x-\mu)^2) \cdot (-\frac{(x-\mu)}{\sigma^2})$

2.1.3 (c) $f(x) = sin(x_1)cos(x_2), x \in \mathbb{R}^2$

x is 2×1 vector.

For $\frac{df}{dx_1}$, $\frac{df}{dx_1} = cos(x_1)cos(x_2)$

For $\frac{df}{dx_2}$, $\frac{df}{dx_2} = -sin(x_1)sin(x_2)$

So $f(x)^{'} = [cos(x_1)cos(x_2), -sin(x_1)sin(x_2)]$

2.1.4 (d) $f(x) = xx^T, x \in \mathbb{R}^n$

x is a $n \times 1$ matrix. So, xx^T is $n \times 1 \cdot 1 \times n = n \times n$ matrix. For each variable in $n \times 1$ vector, we need to differentiate respect to each variables.

First, let's dot product x and x^T.

$x = \begin{vmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{vmatrix}$, $x^T = \begin{vmatrix} x_1 & x_2 & x_3 & \dots & x_n \end{vmatrix}$.

So $xx^T = \begin{vmatrix} x_1^2 & x_1x_2 & x_1x_3 & \dots & x_1x_n \\ x_1x_2 & x_2^2 & x_2x_3 & \dots & x_2x_n \\ x_1x_3 & x_2x_3 & x_3^2 & \dots & x_3x_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1x_n & x_2x_n & x_3x_n & \dots & x_n^2 \end{vmatrix}$

Differentiate respect to x_1, $\frac{d(xx^T)}{dx_1} = \begin{vmatrix} 2x_1 & x_2 & x_3 & \dots & x_n \\ x_2 & 0 & 0 & \dots & 0 \\ x_3 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n & 0 & 0 & & 0 \end{vmatrix}$

Differentiate respect to x_2, $\frac{d(xx^T)}{dx_2} = \begin{vmatrix} 0 & x_1 & 0 & \dots & 0 \\ x_1 & 2x_2 & x_3 & \dots & x_n \\ 0 & x_3 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & x_n & 0 & \dots & 0 \end{vmatrix}$

Work this way all the way to differentiate respect to x_n.

Differentiate respect to x_n, $\frac{d(xx^T)}{dx_n} = \begin{vmatrix} 0 & 0 & 0 & \dots & x_1 \\ 0 & 0 & 0 & \dots & x_2 \\ 0 & 0 & 0 & \dots & x_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1 & x_2 & x_3 & \dots & 2x_n \end{vmatrix}$

so the derivative $\frac{df}{dx}$ for the xx^T is

$$\left[\begin{vmatrix} 2x_1 & x_2 & x_3 & \dots & x_n \\ x_2 & 0 & 0 & \dots & 0 \\ x_3 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n & 0 & 0 & & 0 \end{vmatrix}, \begin{vmatrix} 0 & x_1 & 0 & \dots & 0 \\ x_1 & 2x_2 & x_3 & \dots & x_n \\ 0 & x_3 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & x_n & 0 & \dots & 0 \end{vmatrix} \quad \dots \quad \begin{vmatrix} 0 & 0 & 0 & \dots & x_1 \\ 0 & 0 & 0 & \dots & x_2 \\ 0 & 0 & 0 & \dots & x_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1 & x_2 & x_3 & \dots & 2x_n \end{vmatrix} \right]$$

2.1.5 (e) $f(x) = sin(log(x^T x)), x \in \mathbb{R}^n$

x is a $n \times 1$ vector. So $(x^T \cdot x)$ is a scalar. $f(x)' = sin(log(x^T x))' = cos(log(x^T x)) \cdot log(x^T x)' = cos(log(x^T x)) \cdot \frac{1}{x^T x} \cdot (x^T x)' = cos(log(x^T x)) \cdot \frac{1}{x^T x} \cdot 2x$.

2.2 Convexity

2.2.1 Problem 2.1.1

Let x be a real-valued random variable which takes values in $a_1, ..., a_n$ where $a_1 < a_2 < \cdots < a_n$, with prob$(x = a_i) = p_i$, $i = 1, ..., n$. For each of the following functions of p (on the probability simplex $\mathbf{p} \in R^n_+ | 1^T \mathbf{p} = 1$), determine if the function is convex.

2.2.2 (a) $E(x)$.

In order to find the function is convex or not, I need to first differentiate twice and check if all eigenvalues of double differentiated function are 0 or above.

$E(x)$ is the expected value of the random variable vector, $a \in R^n$ with the probability vector $p \in R^n$. I can write $E(x) = a^T \cdot p$, since $E(x) = a_1 \cdot p_1 + a_2 \cdot p_2 + \ldots + a_n \cdot p_n$.

$E(x) = \sum_{i=1}^n a^T p = a_1 \cdot p_1 + a_2 \cdot p_2 + \ldots + a_n \cdot p_n$.

$E(x)' = \frac{df}{dp} \cdot E(x) = (\sum_{i=1}^n a^T p)' = \begin{bmatrix} a_1, & a_2, & \ldots, & a_n \end{bmatrix} = a^T$

$E(x)'' = \frac{df}{dp} \cdot E(x)' = (a^T)' = \begin{bmatrix} 0, & 0, & \dots & 0 \\ 0, & 0, & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}$

All zero matrix is a diagonal matrix. Diagonal matrix's eigenvalues are the values on the diagonal. In this case it is all zero. so it is positive semi definite. Which is convex.

2.2.3 (b) $\sum_{i=1}^n p_i log\ p_i$, the negative entropy of the distribution.

$f(p) = \sum_{i=1}^n p_i log\ p_i = p_1 \cdot log\ p_1 + p_2 \cdot log\ p_2 + \ldots + p_n \cdot log\ p_n$.

$$\begin{aligned} f(p)' &= \left[\tfrac{df}{dp_1} \cdot \textstyle\sum_{i=1}^n p_i log\ p_i, \quad \tfrac{df}{dp_2} \cdot \sum_{i=1}^n p_i log\ p_i, \quad \ldots, \quad \tfrac{df}{dp_n} \cdot \sum_{i=1}^n p_i log\ p_i \right] \\ &= \begin{bmatrix} 1 + log\ p_i, & 1 + log\ p_2, & \ldots, & 1 + log\ p_n \end{bmatrix} \end{aligned}$$

$$f(p)'' = \left[\frac{df}{dp_1} \cdot \left[1 + logp_i, \quad \ldots, \quad 1 + logp_n\right], \quad \frac{df}{dp_2} \cdot \left[1 + logp_i, \quad \ldots, \quad 1 + logp_n\right], \quad \ldots, \quad \frac{df}{dp_n} \cdot \left[1 + logp_i, \quad \ldots, \quad 1 + logp_n\right]\right.$$

$$= \begin{bmatrix} \frac{1}{p_1}, & 0, & \ldots & 0 \\ 0, & \frac{1}{p_2}, & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{1}{p_n} \end{bmatrix}$$

In this diagonal matrix, all eigenvalues are positive, so it is a positive semi definite. So the negative entropy of the distribution is convex.

2.2.4 (c) $V(x) = E((x - Ex)^2)$**.**

$$V(x) = E(x - (Ex)^2) = \sum_{i=1}^{n} a_i^2 p_i - (\sum_{i=1}^{n} a_1 p_1)^2$$

Let $a_{ij} = a_i \cdot a_j$.

$$V(x) = \sum_{i=1}^{n} a_i^2 p_i - \sum_{i=1, j=[1,n]}^{n} a_{ij} \cdot p_1 \cdot p_j$$

Let's separate two terms, left and right from V(x) function.

$$V(x)'' \; of \; the \; left \; term, (\sum_{i=1}^{n} a_i^2 p_i)'' = (\begin{bmatrix} a_1^2 \\ a_2^2 \\ \vdots \\ a_n^2 \end{bmatrix})' = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$V(x)'' \; of \; the \; right \; term, (\sum_{i=1, j=[1,n]}^{n} a_{ij} \cdot p_1 \cdot p_j)'' = (p^T A p)''$$
$$= (2Ap)' = 2A$$

A is $n \times n$ matrix. $(a_i \times a_j)$

So $V(x)''$ is $0 - 2A = -2A$. Since the -2 times square values is negative, $V(x)$ is not convex.

Submitted by Jae Hong Lee on April 2, 2024.