

# ST 661 Note

Frances Lin

1/4/2022

## Linearly Independence

A set of vectors  $a_1, a_2, \dots, a_n$  is said to be linearly independent if there exists  $c_1, c_2, \dots, c_n$  not all zero s.t.

$$c_1a_1 + c_2a_2 + \dots + c_na_n = 0.$$

If no such coefficients exist, then  $a_1, a_2, \dots, a_n$  is said to be linearly dependent.

The columns of  $A$  are linearly independently iff

$$Ac = 0 \text{ implies } c = 0.$$

The reason is that  $Ac$  involves linearly linear combination of column vectors of  $A$ .

e.g. If  $c_1$  is nonzero, then we can write that

$$a_1 = -\frac{c_2a_2}{c_1} - \frac{c_3a_3}{c_1} - \dots - \frac{c_na_n}{c_1}.$$

## Rank

Rank of  $A$  is the max number of columns in  $A$  that are linearly independent.

e.g.

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 3 & 0 \end{bmatrix}.$$

For column vector 1, 2, and 4, we can have nonzero  $c$  so they are linearly dependent.

As a result,

$$\text{rank}(A) = 1.$$

## Full Rank Matrix

Suppose  $A$  is  $n \times p$ , then

$$\text{rank}(A) \leq \min(n, p).$$

$A$  is said to be a full rank matrix if

$$\text{rank}(A) = \min(n, p).$$

e.g.

$$A = \begin{bmatrix} 1 & -2 & 3 \\ 5 & 2 & 4 \end{bmatrix}.$$

$$c_1 \begin{bmatrix} 1 \\ 5 \end{bmatrix} + c_2 \begin{bmatrix} -2 \\ 2 \end{bmatrix} + c_3 \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

One possible solution is that

$$c = \begin{bmatrix} 14 \\ -11 \\ -12 \end{bmatrix}.$$

In addition, because

$$\begin{bmatrix} 1 \\ 5 \end{bmatrix} \text{ and } \begin{bmatrix} -2 \\ 2 \end{bmatrix}$$

are linearly independent,

$$\text{rank}(A) = 2.$$

We can also define the row rank of a matrix. The row rank of  $A$  is the max number of rows that are linearly independent.

In addition, for all matrix  $A$ ,

$$\text{row rank}(A) = \text{column rank}(A).$$

As a result,

$$\text{rank}(A) = \text{rank}(A^T).$$

On the other hand, if  $A \neq 0$  and  $B \neq 0$ , it is possible that  $AB = 0$ .

e.g.

$$AB = CB \not\Rightarrow A = C,$$

unless  $B$  is a square full rank matrix.

## Thm

If  $A$  and  $B$  are conformal for multiplication, then

i)

$$\text{rank}(AB) \leq \text{rank}(A)$$

$$\text{rank}(AB) \leq \text{rank}(B).$$

ii)

$$\text{rank}(AA^T) = \text{rank}(A^TA) = \text{rank}(A).$$

e.g.

$$A : 2 \times 2, \text{rank}(A) = 2.$$

$$AA^T : 2 \times 2, \text{rank}(A) = 2.$$

$$A^TA : 10 \times 10, \text{rank}(A) = 2.$$

There is something special about rank. Rank is related to the information contained in the matrix.

iii) If  $B$  and  $C$  are full rank square matrices, then

$$\text{rank}(AB) = \text{rank}(CA) = \text{rank}(A).$$

Full rank square matrices preserve the information.

## Inverse

A full rank square matrix is said to be called nonsingular (or invertible). A nonsingular matrix  $A$  has an inverse, denoted by  $A^{-1}$  s.t.

$$AA^{-1} = I.$$

e.g. If  $AB = I$ , then  $B$  is called the inverse of  $A$ , where  $B = A^{-1}$ .

## Properties of Inverse

1.  $A^{-1}$  is unique.
2. For square (don't necessarily has to be nonsingular) matrices  $A$  and  $B$ , if

$$AB = I \iff BA = I.$$

Thus,

$$AA^{-1} = A^{-1}A = I.$$

Proof is left for HW.

- 3.

$$(A^{-1})^{-1} = A.$$

## Results

If  $B$  and  $C$  are square matrices, then

$$\text{rank}(AB) = \text{rank}(CA) = \text{rank}(A).$$

Proof. To show that  $\text{rank}(AB) = \text{rank}(A)$ , first note that

$$\text{rank}(AB) \leq \text{rank}(A). \quad \text{thm i)}$$

It remains to show  $\geq$ . Using the inverse trick,

$$A = A(BB^{-1}) = (AB)B^{-1}.$$

As a result,

$$\text{rank}(A) = \text{rank}((AB)B^{-1}) \leq \text{rank}(AB). \quad \text{thm i)}$$

$$\Rightarrow \text{rank}(A) = \text{rank}(AB). \quad \square$$

Note. In general, if  $B$  is not full rank (or nonsingular), there may not exist a matrix  $C$  s.t.  $A = (AB)C$ .

If a square matrix is not full rank (or nonsingular), then it is said to be a singular matrix.

Remark.

1. If  $B$  is nonsingular, then  $AB = CB \Rightarrow A = C$ .
2. If  $B$  is nonsingular, then the equation  $Bx = C$  has an unique solution, given by  $x = B^{-1}C$ .

## Thm

If  $A$  is nonsingular, then  $A^T$  is also nonsingular, and  $(A^T)^{-1} = (A^{-1})^T$ .

## Thm

If  $A$  and  $B$  are both nonsingular and of the same size, then  $AB$  is also nonsingular.

In addition,

$$(AB)^{-1} = B^{-1}A^{-1}.$$

## Computation

e.g. (updating a regression on the  $f(x)$ )

If a linear model is

$$y = XB + \varepsilon,$$

LSE (least square estimator) is given by

$$\hat{B} = (X^T X)^{-1} X^T y.$$

If we add new data points, a new design matrix

$$W = \begin{bmatrix} X \\ X^{(n+1)T} \end{bmatrix},$$

where  $X^{(n+1)T}$  is a row. To compute  $(W^T W)^{-1}$  more efficiently, notice that

$$W^T W = [X^T \quad X^{(n+1)}] \begin{bmatrix} X \\ X^{(n+1)T} \end{bmatrix} = X^T X + X^{n+1} X^{(n+1)T}.$$

Then taking the inverse,

$$(W^T W)^{-1} = (X^T X + X^{n+1} X^{(n+1)T})^{-1},$$

where  $X^{n+1} X^{(n+1)T}$  is a rank 1 item.

As it turns out,

$$(W^T W)^{-1} = (X^T X)^{-1} - \frac{(X^T X)^{-1} X^{n+1} X^{(n+1)T} (X^T X)^{-1}}{1 + X^{(n+1)T} (X^T X)^{-1} X^{n+1}}.$$

This takes advantages of the known matrix  $(X^T X)^{-1}$  and matrix multiplication, which is computationally faster than matrix inversion. This comes from the following thm.

## Thm

Suppose  $A$  is of the form  $A = B + CC^T$ , where  $B$  is a familiar matrix and  $A$  is a new matrix, then

$$A^{-1} = (B + CC^T)^{-1} = B^{-1} - \frac{B^{-1}CC^T B^{-1}}{1 + C^T B^{-1}C}.$$

Ques. How about for other item that is other than rank 1 item such as  $C$ ?

## Thm

In general, if  $A$ ,  $B$ , and  $A + PBQ$  are nonsingular, then

$$(A + PBQ)^{-1} = A^{-1} - A^{-1}PB(B + BQA^{-1}PB)^{-1}BQA^{-1},$$

where  $A$  has to be a square matrix and  $B$  is low rank and has to be a square matrix.

e.g. If  $A$  is 100x100,  $P$  is 100x5,  $B$  is 5x5, and  $Q$  is 5x100, then it turns out that it only involves inverting a 5x5 matrix  $(B + BQA^{-1}PB)^{-1}$ .

End. Double check the formulas. See page 39 of PDF (or 24 of BOOK) for more.