

# ST 661 Note

Frances Lin

1/20/2022

## Outline

- Idempotent matrix
- Matrix calculus

## Last time

Let  $A_1$  be the 1st subspace and  $A_2$  be the second subspace, then we can write  $x$  s.t.

$$x = A_1x + A_2x \quad \forall x \in \mathbb{R}^n$$

$$\begin{aligned} I &= A_1 + A_2 \quad , \quad \text{rank}(I) = 2 \\ \text{rank}(A_1) &= 1 \quad , \quad \text{rank}(A_2) = 2 \end{aligned}$$

Both  $A_1$  and  $A_2$  are idempotent matrices.

Now, we pick  $x \in \mathbb{R}^3$  and decompose it s.t.

$$x = A_1x + A_2x,$$

where

$$\text{rank}(A_1) = 2 \quad \text{a plane} \quad , \quad \text{rank}(A_2) = 1 \quad \text{a line},$$

so

$$\text{rank}(I) = 3 = 2 + 1 = \text{rank}(A_1) + \text{rank}(A_2).$$

## Thm

$I : nxn$ ,  $I = A_1 + A_2 + \dots + A_k$  ( $k$  can be anything  $\leq n$ ), where each  $A_i$  is  $n \times n$  symmetric of rank  $r_i$ .

If  $\sum_{i=1}^k r_i = n$  (no gain or loss of rank), then

- $A_i$  (each  $A_i$ ) is idempotent  $i = 1, 2, \dots, k$
- $A_i A_j = 0$   $i \neq j$  (complementary)

Comment. In general,  $\text{rank}(A + B) \neq \text{rank}(A) + \text{rank}(B)$ . So when it is equal, then all  $A_i$  are idempotent.

e.g.  $n = 2$

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = A_1 + A_2$$

So,  $\text{rank}(A_1) + \text{rank}(A_2) = 2 = \text{rank}(I)$ .

=>

- i)  $A_1$   $A_2$  are both idempotent
- ii)  $A_1 A_2 = 0$  Info does not overlap.

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$A_1 x = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

projecting to x axis.

$$A_2 x = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

projecting to y axis.

e.g.

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} + \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

In this case,

$$\text{rank}(I) = 2 \neq 2 + 2 = \text{rank}(A_1) + \text{rank}(A_2)$$

Condition is not met. Info overlap.

ii)

$$A_1 A_2 = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \neq 0,$$

so info overlap.

## Vector or matrix calculus (derivatives mostly)

1. vector  $\rightarrow$  scalar

$u = f(x)$ , where  $u$  is a scalar,  $x$  is a column vector and

$$x = [x_1 \ x_2 \ \cdots \ x_p]^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix},$$

then

$$\frac{\partial u}{\partial x} = \begin{bmatrix} \frac{\partial u}{\partial x_1} \\ \frac{\partial u}{\partial x_2} \\ \vdots \\ \frac{\partial u}{\partial x_p} \end{bmatrix}.$$

### Thm

Let  $u = a^T x$  and  $a = (a_1, a_2, \dots, a_p)^T$  is a constant vector, then

$$\frac{\partial u}{\partial x} = a,$$

i.e.

$$\frac{\partial u}{\partial x_1} = a_1.$$

## Thm

$u = X^T AX$  (quadratic form). A is symmetric matrix of constants.

$$u = X^T AX = \sum_{i=1}^n \sum_{i=j}^n a_{ij} x_i x_j,$$

then

$$\frac{\partial u}{\partial x} = \frac{\partial X^T AX}{\partial x} = 2AX,$$

which is analogous to taking

$$\frac{\partial}{\partial x}(ax^2) = 2ax.$$

Proof.

$$\frac{\partial u}{\partial x_i} = \frac{\partial X^T AX}{\partial x_i} = \dots \quad \square$$

2. matrix  $\rightarrow$  scalar

$\mu = f(X)$   $X$  is a  $p \times p$  square matrix. e.g.

$$\frac{\partial u}{\partial X} = \begin{bmatrix} \frac{\partial u}{\partial x_{11}} & \frac{\partial u}{\partial x_{12}} & \cdots & \frac{\partial u}{\partial x_{1p}} \\ \vdots & & & \\ \frac{\partial u}{\partial x_{p1}} & \cdots & \cdots & \frac{\partial u}{\partial x_{pp}} \end{bmatrix}.$$

Notation.  $x$  is vector.  $X$  is matrix.

## Thm

$u = \text{tr}(XA)$  where  $A$  is a  $p \times p$  constant matrix (This is more general than  $u = \text{tr}(X)$ .), then

$$\frac{\partial u}{\partial X} = \frac{\partial}{\partial X} \text{tr}(XA)$$

$$= A + A^T - \text{diag}(A).$$

Proof.

$$\text{tr}(XA) = \sum_{i=1}^p \sum_{j=1}^p x_{ij} a_{ji}$$

Verify the above.

$$\frac{\partial}{\partial x_{ij}} = \begin{cases} a_{ij} + a_{ji} & i \neq j \\ a_{ii} & i = j \end{cases}$$

The idea is that  $\text{tr}()$  is a linear function.

## Thm

$u = \log(\det(X))$ ,  $X$  is  $p \times p$  p.d., then

$$\frac{\partial}{\partial X}(\det(X)) = 2X^{-1} - \text{diag}(X^{-1}).$$

Proof. Skip here.

E.g. log likelihood of MVN's sigma term.

3. scalar  $\rightarrow$  vector or matrix

$A : A(x)$  where  $x$  is scalar.

$A : (a_{ij})$  is matrix.

## Properties

$A = A(x)$  and  $B = B(x)$

i) If  $A$  and  $B$  have the same dim, then

$$\frac{\partial}{\partial X}(A + B) = \frac{\partial A}{\partial X} + \frac{\partial B}{\partial X}.$$

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

$$\frac{\partial}{\partial X}(AB) = \frac{\partial A}{\partial X}B + A\frac{\partial B}{\partial X}$$

$$\frac{\partial}{\partial X}(ABC) = \frac{\partial A}{\partial X}BC + A\frac{\partial B}{\partial X}C + AB\frac{\partial C}{\partial X}.$$

## Thm

$A = A(x)$  is nonsingular,  $A^{-1} = A^{-1}(x)$  is another matrix function of  $x$ , then

$$\frac{\partial}{\partial X}A^{-1} = -A^{-1} + \frac{\partial A}{\partial X}A^{-1}$$

Proof. Start with  $AA^{-1} = I$ .  $\frac{\partial A}{\partial X}$ ,  $A^{-1}$ , and  $A$  are known, then

$$0 = \frac{\partial AA^{-1}}{\partial X} = \frac{\partial A}{\partial X}A^{-1} + A\frac{\partial A^{-1}}{\partial X},$$

so

$$A\frac{\partial A^{-1}}{\partial X} = -\frac{\partial A}{\partial X}A^{-1}$$

$$\frac{\partial A^{-1}}{\partial X} = -A^{-1}\frac{\partial A}{\partial X}A^{-1} \quad \square$$

## Thm

$A = A(x)$  Now  $A$  is indexed by scalar.  $A$  is  $n \times n$ , symmetric, p.d.  $\log(\det(A))$  is a scalar function of  $x$ , then

$$\frac{\partial}{\partial X} \log(\det(A)) = \text{tr}(A^{-1} \frac{\partial A}{\partial X})$$

Proof.  $A = CDC^T$

First, show that

$$\begin{aligned} \frac{\partial}{\partial X} \log(\det(D)) &= \frac{\partial}{\partial X} \log(\lambda_1, \dots, \lambda_n) = \frac{\partial}{\partial X} (\log(\lambda_1), \dots, \log(\lambda_n)) = \sum_{i=1}^n \frac{1}{\lambda_i} \frac{\partial \lambda_i}{\partial X} \\ &\stackrel{?}{=} \text{tr}(D^{-1} \frac{\partial D}{\partial X}) \\ &= \text{tr}\left(\begin{bmatrix} \frac{1}{\lambda_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{\lambda_n} \end{bmatrix} \begin{bmatrix} \frac{\partial \lambda_1}{\partial X} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{\partial \lambda_n}{\partial X} \end{bmatrix}\right) \\ &= \text{tr}\left(\begin{bmatrix} \frac{1}{\lambda_1} \frac{\partial \lambda_1}{\partial X} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{\lambda_n} \frac{\partial \lambda_n}{\partial X} \end{bmatrix}\right) \\ &= \sum_{i=1}^n \frac{1}{\lambda_i} \frac{\partial \lambda_i}{\partial X}. \end{aligned}$$

Next, for general  $A$ , we want to diagonalize it:  $A = CDC^T$ .

Recall that  $\det(CDC^T) = \det(D)$  since  $C \dots$ , so

$$\frac{\partial}{\partial X} \log(\det(A)) = \frac{\partial}{\partial X} \log(\det(D)) = \text{tr}(D^{-1} \frac{\partial D}{\partial X}) \stackrel{?}{=} \text{tr}(A^{-1} \frac{\partial A}{\partial X})$$

? remains to be shown.

Start with RHS, we have that

$$\begin{aligned} \text{tr}(A^{-1} \frac{\partial A}{\partial X}) &= \text{tr}((CDC^T)^{-1} \frac{\partial (CDC^T)}{\partial X}) \\ &= \text{tr}((CD^{-1}C^T)((\frac{\partial C}{\partial X})DC^T + C(\frac{\partial D}{\partial X})C^T + CD(\frac{\partial C^T}{\partial X}))) \\ &= \text{tr}(CD^{-1}C^T(\frac{\partial C}{\partial X})DC^T) + \text{tr}(CD^{-1}C^T C(\frac{\partial D}{\partial X})C^T) + \text{tr}(CD^{-1}C^T CD(\frac{\partial C^T}{\partial X})) \end{aligned}$$

Recall that  $\text{tr}(AB) = \text{tr}(BA)$

$$= \text{tr}(D^{-1}C^T \frac{\partial C}{\partial X} DI) + \dots + \dots = \text{tr}(C^T \frac{\partial C}{\partial X}) + \text{tr}(D^{-1} \frac{\partial D}{\partial X}) + \text{tr}(C \frac{\partial C^T}{\partial X})$$

To show that

$$\text{tr}(A^{-1} \frac{\partial A}{\partial X}) = \text{tr}(D^{-1} \frac{\partial D}{\partial X}),$$

need to show

$$tr(C^T \frac{\partial C}{\partial X}) + tr(C \frac{\partial C^T}{\partial X}) = 0$$

$$tr(C^T \frac{\partial C}{\partial X}) + tr(C \frac{\partial C^T}{\partial X}) = tr(C^T \frac{\partial C}{\partial X} + C \frac{\partial C^T}{\partial X}),$$

but this is not quite the same.

However, notice that

$$\frac{\partial I}{\partial X} = \frac{\partial(CC^T)}{\partial X} = \frac{\partial C}{\partial X}C^T + C\frac{\partial C^T}{\partial X} = 0,$$

since  $I$  does not depend on  $X$ .

Reverse the order of the 1st term s.t.

$$tr(\frac{\partial C}{\partial X}C^T + C\frac{\partial C^T}{\partial X}),$$

then

$$= \frac{\partial(CC^T)}{\partial X} = \frac{\partial I}{\partial X} = 0 \quad \square$$