

# Definition, Properties, and Derivatives of Matrix Traces

A Class for Graduate Students

**Xinyu Chen**

March 31, 2024

(release)

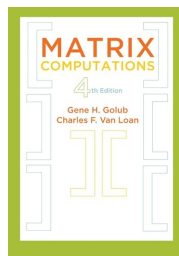
February 21, 2025

(latest update)

## Class Targets

Throughout this class, you will:

- Understand some basic concepts (e.g., norms, traces, and derivatives)
- Connect them with linear algebra and machine learning
- Use matrix norms and traces in matrix computations (very useful!)



## Vector & Matrix

### Notation:

- On the vector  $\boldsymbol{x} \in \mathbb{R}^n$  of length  $n$

$$\boldsymbol{x} = (x_1, x_2, \dots, x_n)^\top \quad \text{or} \quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

whose  $i$ -th entry is  $x_i$ ,  $i \in [n]$ .<sup>1</sup>

---

<sup>1</sup>The set of integers,  $\{1, 2, \dots, n\}$ , is represented by  $[n]$ ,  $n \in \mathbb{Z}^+$ .

## Vector & Matrix

### Notation:

- On the vector  $\mathbf{x} \in \mathbb{R}^n$  of length  $n$

$$\mathbf{x} = (x_1, x_2, \dots, x_n)^\top \quad \text{or} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

whose  $i$ -th entry is  $x_i$ ,  $i \in [n]$ .<sup>1</sup>

- On the matrix  $\mathbf{X} \in \mathbb{R}^{m \times n}$  with  $m$  rows and  $n$  columns

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

whose  $(i, j)$ -th entry is  $x_{ij}$ ,  $i \in [m], j \in [n]$ .

---

<sup>1</sup>The set of integers,  $\{1, 2, \dots, n\}$ , is represented by  $[n]$ ,  $n \in \mathbb{Z}^+$ .

## Vector Norms

A number of concepts to mention, e.g.,  $\ell_0$ -norm,  $\ell_1$ -norm, and  $\ell_2$ -norm.

- **Definition.** For any vector  $\mathbf{x} \in \mathbb{R}^n$ , the  $\ell_2$ -norm of  $\mathbf{x}$  is given by

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\sum_{i=1}^n x_i^2}$$

where  $x_i$ ,  $\forall i \in [n]$  is the  $i$ -th entry of  $\mathbf{x}$ .

## Vector Norms

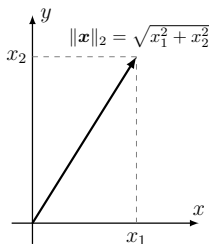
A number of concepts to mention, e.g.,  $\ell_0$ -norm,  $\ell_1$ -norm, and  $\ell_2$ -norm.

- **Definition.** For any vector  $\mathbf{x} \in \mathbb{R}^n$ , the  $\ell_2$ -norm of  $\mathbf{x}$  is given by

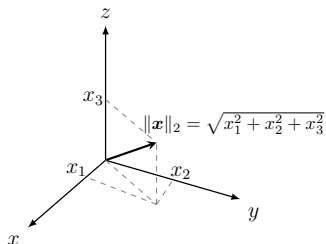
$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\sum_{i=1}^n x_i^2}$$

where  $x_i, \forall i \in [n]$  is the  $i$ -th entry of  $\mathbf{x}$ .

- Intuitive examples:



On  $\mathbf{x} = (x_1, x_2)^\top$

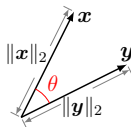


On  $\mathbf{x} = (x_1, x_2, x_3)^\top$

# Inner Product

- Basics: For any  $\mathbf{x} = (x_1, x_2)^\top$  and  $\mathbf{y} = (y_1, y_2)^\top$ , the angle  $\theta$  can be computed by

$$\cos \theta = \frac{x_1 y_1 + x_2 y_2}{\sqrt{x_1^2 + x_2^2} \sqrt{y_1^2 + y_2^2}}$$



- in which
  - $\ell_2$ -norm:

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2} \quad \|\mathbf{y}\|_2 = \sqrt{y_1^2 + y_2^2}$$

- inner product:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y} = x_1 y_1 + x_2 y_2$$

- It leads to

$$\cos \theta = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|_2 \cdot \|\mathbf{y}\|_2}$$

## Inner Product

- **Definition.** For any vectors  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ , the inner product is given by

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y} = \sum_{i=1}^n x_i y_i$$

**Example.** Given  $\mathbf{x} = (1, 2, 3, 4)^\top$  and  $\mathbf{y} = (2, -1, 3, 0)^\top$ , write down the inner product  $\langle \mathbf{x}, \mathbf{y} \rangle$ .

In this case,

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y} = 1 \times 2 + 2 \times (-1) + 3 \times 3 + 4 \times 0 = 9$$



## Inner Product

- **Definition.** For any matrices  $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m \times n}$ , the inner product is

$$\langle \mathbf{X}, \mathbf{Y} \rangle = \sum_{i=1}^m \sum_{j=1}^n x_{ij} y_{ij}$$

**Example.** Given  $\mathbf{X} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}$  and  $\mathbf{Y} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ , write down the inner product  $\langle \mathbf{X}, \mathbf{Y} \rangle$ .

In this case,

$$\langle \mathbf{X}, \mathbf{Y} \rangle = 2 \times 2 + 1 \times (-1) + 1 \times (-1) + 2 \times 2 + 1 \times (-1) + 3 \times 2 = 11$$

## Frobenius Norm

- Definition.** For any matrix  $\mathbf{X} \in \mathbb{R}^{m \times n}$ , the Frobenius norm of  $\mathbf{X}$  is given by

$$\|\mathbf{X}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n x_{ij}^2}$$

where  $x_{ij}$ ,  $\forall i \in [m], j \in [n]$  is the  $(i, j)$ -th entry of  $\mathbf{X}$ .

**Example.** Given  $\mathbf{X} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}$ , write down the Frobenius norm of  $\mathbf{X}$ .

$$\|\mathbf{X}\|_F = \sqrt{2^2 + 1^2 + 1^2 + 1^2 + 2^2 + 1^2 + 3^2} = \sqrt{21}$$

# Frobenius Norm

- Connection with  $\ell_2$ -norm:

$$\|\mathbf{X}\|_F = \sqrt{\sum_{j=1}^n \sum_{i=1}^m x_{ij}^2} = \sqrt{\sum_{j=1}^n \|\mathbf{x}_j\|_2^2}$$

with the column vectors  $\mathbf{x}_j \in \mathbb{R}^m$ ,  $j \in [n]$  such that

$$\mathbf{X} = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{m \times n}$$

## Definition of Matrix Trace

- **Definition.** For any **square matrix**  $\mathbf{X} \in \mathbb{R}^{n \times n}$ , the matrix trace (denoted by  $\text{tr}(\cdot)$ ) is the sum of diagonal entries, i.e.,

$$\text{tr}(\mathbf{X}) = \sum_{i=1}^n \underbrace{x_{ii}}_{\text{diagonal}}$$

where  $x_{ii}, \forall i \in [n]$  is the  $(i, i)$ -th entry of  $\mathbf{X}$ . Thus,  $\text{tr}(\mathbf{X}) = \text{tr}(\mathbf{X}^\top)$ .

**Example.** Given  $\mathbf{X} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}$ , write down the matrix trace of  $\mathbf{X}$ .

$$\text{tr}(\mathbf{X}) = 2 + 2 + 3 = 7$$

## Property: $\text{tr}(\mathbf{X} + \mathbf{Y}) = \text{tr}(\mathbf{X}) + \text{tr}(\mathbf{Y})$

- **Property.** For any square matrices  $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n \times n}$ , it always holds that

$$\text{tr}(\mathbf{X} + \mathbf{Y}) = \text{tr}(\mathbf{X}) + \text{tr}(\mathbf{Y})$$

**Example.** Given  $\mathbf{X} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}$  and  $\mathbf{Y} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ , write down  $\text{tr}(\mathbf{X} + \mathbf{Y})$ .

In this case,

$$\mathbf{X} + \mathbf{Y} = \begin{bmatrix} 2+2 & 1-1 & 1+0 \\ 1-1 & 2+2 & 1-1 \\ 0+0 & 0-1 & 3+2 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 4 & 0 \\ 0 & -1 & 5 \end{bmatrix}$$

Thus,  $\text{tr}(\mathbf{X} + \mathbf{Y}) = 4 + 4 + 5 = 13$ . Note that  $\text{tr}(\mathbf{X}) = 7$  and  $\text{tr}(\mathbf{Y}) = 6$ , it shows that  $\text{tr}(\mathbf{X} + \mathbf{Y}) = \text{tr}(\mathbf{X}) + \text{tr}(\mathbf{Y}) = 13$ .

- **Variant.** For any  $\alpha, \beta \in \mathbb{R}$ , we have

$$\text{tr}(\alpha \mathbf{X} + \beta \mathbf{Y}) = \alpha \text{tr}(\mathbf{X}) + \beta \text{tr}(\mathbf{Y})$$

## Property: $\text{tr}(\mathbf{XY}) = \text{tr}(\mathbf{YX})$

- **Property.** For any matrices  $\mathbf{X} \in \mathbb{R}^{m \times n}$  and  $\mathbf{Y} \in \mathbb{R}^{n \times m}$ , it always holds that

$$\text{tr}(\mathbf{XY}) = \text{tr}(\mathbf{YX})$$

- **Proof.**

$$\begin{aligned}
 \text{tr}(\mathbf{XY}) &= [\mathbf{XY}]_{11} + [\mathbf{XY}]_{22} + \cdots + [\mathbf{XY}]_{mm} \\
 &= \underbrace{x_{11}y_{11} + x_{12}y_{21} + \cdots + x_{1n}y_{n1}}_{\text{the first row of } \mathbf{X} \text{ times the first column of } \mathbf{Y}} \\
 &\quad + x_{21}y_{12} + x_{22}y_{22} + \cdots + x_{2n}y_{n2} \\
 &\quad + \cdots + x_{m1}y_{1m} + x_{m2}y_{2m} + \cdots + x_{mn}y_{nm}
 \end{aligned}$$

- **Proof.**

$$\begin{aligned}
 \text{tr}(\mathbf{XY}) &= [\mathbf{XY}]_{11} + [\mathbf{XY}]_{22} + \cdots + [\mathbf{XY}]_{mm} \\
 &= \underbrace{x_{11}y_{11} + x_{12}y_{21} + \cdots + x_{1n}y_{n1}}_{\text{the first row of } \mathbf{X} \text{ times the first column of } \mathbf{Y}} \\
 &\quad + x_{21}y_{12} + x_{22}y_{22} + \cdots + x_{2n}y_{n2} \\
 &\quad + \cdots + x_{m1}y_{1m} + x_{m2}y_{2m} + \cdots + x_{mn}y_{nm} \\
 &= x_{11}y_{11} + x_{12}y_{21} + \cdots + x_{1n}y_{n1} \\
 &\quad + x_{21}y_{12} + x_{22}y_{22} + \cdots + x_{2n}y_{n2} \\
 &\quad + \cdots + x_{m1}y_{1m} + x_{m2}y_{2m} + \cdots + x_{mn}y_{nm}
 \end{aligned}$$



- **Proof.**

$$\begin{aligned}
 \text{tr}(\mathbf{XY}) &= [\mathbf{XY}]_{11} + [\mathbf{XY}]_{22} + \cdots + [\mathbf{XY}]_{nn} \\
 &= \underbrace{x_{11}y_{11} + x_{12}y_{21} + \cdots + x_{1n}y_{n1}}_{\text{the first row of } \mathbf{X} \text{ times the first column of } \mathbf{Y}} \\
 &\quad + x_{21}y_{12} + x_{22}y_{22} + \cdots + x_{2n}y_{n2} \\
 &\quad + \cdots + x_{m1}y_{1m} + x_{m2}y_{2m} + \cdots + x_{mn}y_{nm} \\
 &= x_{11}y_{11} + x_{12}y_{21} + \cdots + x_{1n}y_{n1} \\
 &\quad + x_{21}y_{12} + x_{22}y_{22} + \cdots + x_{2n}y_{n2} \\
 &\quad + \cdots + x_{m1}y_{1m} + x_{m2}y_{2m} + \cdots + x_{mn}y_{nm} \\
 &= \underbrace{y_{11}x_{11} + y_{12}x_{21} + \cdots + y_{1m}x_{m1}}_{\text{the first row of } \mathbf{Y} \text{ times the first column of } \mathbf{X}} \\
 &\quad + y_{21}x_{12} + y_{22}x_{22} + \cdots + y_{2m}x_{m2} \\
 &\quad + \cdots + y_{n1}x_{1n} + \cdots + y_{n2}x_{2n} + \cdots + y_{nm}x_{mn} \\
 &= [\mathbf{YX}]_{11} + [\mathbf{YX}]_{22} + \cdots + [\mathbf{YX}]_{nn} = \text{tr}(\mathbf{YX})
 \end{aligned}$$

## Property: $\text{tr}(\mathbf{XY}) = \text{tr}(\mathbf{YX})$

**Example.** Given  $\mathbf{X} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}$  and  $\mathbf{Y} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ , write down  $\text{tr}(\mathbf{XY})$  and  $\text{tr}(\mathbf{YX})$ , respectively.

In this case,

$$\mathbf{XY} = \begin{bmatrix} 3 & -1 & 1 \\ 0 & 2 & 0 \\ 0 & -3 & 6 \end{bmatrix} \quad \mathbf{YX} = \begin{bmatrix} 3 & 0 & 1 \\ 0 & 3 & -2 \\ -1 & -2 & 5 \end{bmatrix}$$

Thus,

$$\text{tr}(\mathbf{XY}) = 3 + 2 + 6 = 11 \quad \text{tr}(\mathbf{YX}) = 3 + 3 + 5 = 11$$

**Property:**  $\text{tr}(\mathbf{AXB}) = \text{tr}(\mathbf{BAX}) = \text{tr}(\mathbf{XBA})$

- Property: For any matrices  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{X} \in \mathbb{R}^{n \times d}$ , and  $\mathbf{B} \in \mathbb{R}^{d \times m}$ , it always holds that

$$\begin{cases} \text{tr}(\mathbf{AXB}) = \text{tr}((\mathbf{AX})\mathbf{B}) = \text{tr}(\mathbf{B}(\mathbf{AX})) = \text{tr}(\mathbf{BAX}) \\ \text{tr}(\mathbf{AXB}) = \text{tr}(\mathbf{A}(\mathbf{XB})) = \text{tr}((\mathbf{XB})\mathbf{A}) = \text{tr}(\mathbf{XBA}) \end{cases}$$

## Property: $\|X\|_F^2 = \text{tr}(X^\top X)$

- **Property.** For any matrix  $X \in \mathbb{R}^{m \times n}$ , it always holds that

$$\|X\|_F^2 = \text{tr}(X^\top X)$$

- **Proof.**

$$\begin{aligned} \text{tr}(X^\top X) &= [X^\top X]_{11} + [X^\top X]_{22} + \cdots + [X^\top X]_{nn} \\ &= x_{11}^2 + x_{21}^2 + \cdots + x_{m1}^2 \\ &\quad + x_{12}^2 + x_{22}^2 + \cdots + x_{m2}^2 \\ &\quad + \cdots + x_{1n}^2 + x_{2n}^2 + \cdots + x_{mn}^2 \\ &= \sum_{i=1}^m x_{i1}^2 + \sum_{i=1}^m x_{i2}^2 + \cdots + \sum_{i=1}^m x_{in}^2 \\ &= \sum_{i=1}^m \sum_{j=1}^n x_{ij}^2 \\ &= \|X\|_F^2 \end{aligned}$$

## Property: $\langle X, Y \rangle = \text{tr}(X^\top Y)$

- **Property.** For any matrices  $X, Y \in \mathbb{R}^{m \times n}$ , it always holds that

$$\langle X, Y \rangle = \text{tr}(X^\top Y)$$

- **Proof.**

$$\begin{aligned} \text{tr}(X^\top Y) &= [X^\top Y]_{11} + [X^\top Y]_{22} + \cdots + [X^\top Y]_{nn} \\ &= x_{11}y_{11} + x_{21}y_{21} + \cdots + x_{m1}y_{m1} \\ &\quad + x_{12}y_{12} + x_{22}y_{22} + \cdots + x_{m2}y_{m2} \\ &\quad + \cdots + x_{1n}y_{1n} + x_{2n}y_{2n} + \cdots + x_{mn}y_{mn} \\ &= \langle x_1, y_1 \rangle + \langle x_2, y_2 \rangle + \cdots + \langle x_n, y_n \rangle \\ &= \langle X, Y \rangle \end{aligned}$$

where  $x_i, y_i \in \mathbb{R}^m$ ,  $\forall i \in [n]$  are the  $i$ -th column vectors of  $X$  and  $Y$ , respectively.

**Property:**  $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{tr}(\mathbf{X}^\top \mathbf{Y})$

**Example.** Given  $\mathbf{X} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}$  and  $\mathbf{Y} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ , write down  $\langle \mathbf{X}, \mathbf{Y} \rangle$  and  $\text{tr}(\mathbf{X}^\top \mathbf{Y})$ , respectively.

Recall that

$$\langle \mathbf{X}, \mathbf{Y} \rangle = 2 \times 2 + 1 \times (-1) + 1 \times (-1) + 2 \times 2 + 1 \times (-1) + 3 \times 2 = 11$$

For the matrix,

$$\mathbf{X}^\top \mathbf{Y} = \begin{bmatrix} 3 & 0 & -1 \\ 0 & 3 & -2 \\ 1 & -2 & 5 \end{bmatrix}$$

we have  $\text{tr}(\mathbf{X}^\top \mathbf{Y}) = 3 + 3 + 5 = 11$ .

# Derivatives

A quick revisit!

- **Derivative.** Given a scalar function  $f(x)$  of the single variable  $x$ , the derivative is defined by

$$\frac{d f(x)}{d x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\Delta f}{\Delta x}$$

# Derivatives

A quick revisit!

- **Derivative.** Given a scalar function  $f(x)$  of the single variable  $x$ , the derivative is defined by

$$\frac{d f(x)}{d x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\Delta f}{\Delta x}$$

- **Partial derivatives.** Given a scalar function  $f(x, y)$  of two variables  $x, y$ , the partial derivatives are defined by

$$\begin{cases} \frac{\partial f(x, y)}{\partial x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x} \\ \frac{\partial f(x, y)}{\partial y} = \lim_{\Delta y \rightarrow 0} \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y} \end{cases}$$



# Derivatives

**Example.** Given  $f(\mathbf{x}) = \|\mathbf{x}\|_2^2$ , write down the derivative  $\frac{d f(\mathbf{x})}{d \mathbf{x}}$ .

First, notice that the function  $f(\mathbf{x})$  can be written as

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_n) = x_1^2 + x_2^2 + \dots + x_n^2$$

Hence, the partial derivatives of  $f(x_1, x_2, \dots, x_n)$  with respect to  $x_1, x_2, \dots, x_n$  are

$$\begin{aligned} \frac{\partial f(x_1, x_2, \dots, x_n)}{\partial x_1} &= 2x_1 \\ \frac{\partial f(x_1, x_2, \dots, x_n)}{\partial x_2} &= 2x_2 \\ &\vdots \\ \frac{\partial f(x_1, x_2, \dots, x_n)}{\partial x_n} &= 2x_n \end{aligned} \quad \Rightarrow \quad \frac{d f(\mathbf{x})}{d \mathbf{x}} = \begin{bmatrix} \frac{\partial f(x_1, x_2, \dots, x_n)}{\partial x_1} \\ \frac{\partial f(x_1, x_2, \dots, x_n)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x_1, x_2, \dots, x_n)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{bmatrix} = 2\mathbf{x}$$

## Derivative of $f(\mathbf{X}) = \text{tr}(\mathbf{X})$

- **Function.** For any square matrix  $\mathbf{X} \in \mathbb{R}^{n \times n}$ , what is the derivative of  $f(\mathbf{X}) = \text{tr}(\mathbf{X})$ ?
- **Derivative.** Since  $f(\mathbf{X}) = \sum_{i=1}^n x_{ii}$ , we have

$$\begin{aligned} \frac{d f(\mathbf{X})}{d \mathbf{X}} &= \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \frac{\partial f(\mathbf{X})}{\partial x_{12}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{1n}} \\ \frac{\partial f(\mathbf{X})}{\partial x_{21}} & \frac{\partial f(\mathbf{X})}{\partial x_{22}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{n1}} & \frac{\partial f(\mathbf{X})}{\partial x_{n2}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{nn}} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \mathbf{I}_n \end{aligned}$$

## Derivative of $f(\mathbf{X}) = \text{tr}(\mathbf{A}\mathbf{X})$

- **Function.** For any matrices  $\mathbf{A} \in \mathbb{R}^{m \times n}$  and  $\mathbf{X} \in \mathbb{R}^{n \times m}$ , what is the derivative of  $f(\mathbf{X}) = \text{tr}(\mathbf{A}\mathbf{X})$ ?
- **Derivative.** Since

$$f(\mathbf{X}) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_{ji}$$

Hence, the partial derivative of  $f(\mathbf{X})$  with respect to the entry  $x_{ji}$  is given by

$$\frac{\partial f(\mathbf{X})}{\partial x_{ji}} = a_{ij}$$

As a result, we have

$$\frac{d f(\mathbf{X})}{d \mathbf{X}} = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \frac{\partial f(\mathbf{X})}{\partial x_{12}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{1m}} \\ \frac{\partial f(\mathbf{X})}{\partial x_{21}} & \frac{\partial f(\mathbf{X})}{\partial x_{22}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{2m}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{n1}} & \frac{\partial f(\mathbf{X})}{\partial x_{n2}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{nm}} \end{bmatrix} = \mathbf{A}^\top$$

- By the way, what is the derivative of  $f(\mathbf{X}) = \text{tr}(\mathbf{X}\mathbf{A})$ ? How about  $f(\mathbf{X}) = \text{tr}(\mathbf{A}^\top \mathbf{X}^\top)$ ?

## Derivative of $f(\mathbf{X}) = \text{tr}(\mathbf{AXB})$

- **Function.** For any matrices  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{X} \in \mathbb{R}^{n \times d}$ , and  $\mathbf{B} \in \mathbb{R}^{d \times m}$ , what is the derivative of  $f(\mathbf{X}) = \text{tr}(\mathbf{AXB})$ ?
- **Derivative.** Since

$$f(\mathbf{X}) = \sum_{i=1}^m [\mathbf{AXB}]_{i,i} = \sum_{i=1}^m \sum_{j=1}^n a_{ij} [\mathbf{XB}]_{j,i} = \sum_{i=1}^m \sum_{j=1}^n a_{ij} \sum_{k=1}^d x_{jk} b_{ki}$$

Hence, the partial derivative of  $f(\mathbf{X})$  with respect to the entry  $x_{jk}$  is given by

$$\frac{\partial f(\mathbf{X})}{\partial x_{jk}} = \sum_{i=1}^m a_{ij} b_{ki} = [\mathbf{A}^\top \mathbf{B}^\top]_{j,k}$$

As a result, we have

$$\frac{d f(\mathbf{X})}{d \mathbf{X}} = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \frac{\partial f(\mathbf{X})}{\partial x_{12}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{1d}} \\ \frac{\partial f(\mathbf{X})}{\partial x_{21}} & \frac{\partial f(\mathbf{X})}{\partial x_{22}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{2d}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{n1}} & \frac{\partial f(\mathbf{X})}{\partial x_{n2}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{nd}} \end{bmatrix} = \mathbf{A}^\top \mathbf{B}^\top$$

# Derivative of $f(\mathbf{X}) = \text{tr}(\mathbf{X}^\top \mathbf{A} \mathbf{X})$

- **Function.** For any matrices  $\mathbf{A}, \mathbf{X} \in \mathbb{R}^{n \times n}$ , what is the derivative of  $f(\mathbf{X}) = \text{tr}(\mathbf{X}^\top \mathbf{A} \mathbf{X})$ ?
- **Derivative.**

$$\begin{aligned} \frac{d f(\mathbf{X})}{d \mathbf{X}} &= \frac{d \text{tr}(\mathbf{X}^\top \mathbf{B})}{d \mathbf{X}} + \frac{d \text{tr}(\mathbf{C} \mathbf{X})}{d \mathbf{X}} \\ &= \frac{d \text{tr}(\mathbf{B}^\top \mathbf{X})}{d \mathbf{X}} + \frac{d \text{tr}(\mathbf{C} \mathbf{X})}{d \mathbf{X}} \\ &= \mathbf{B} + \mathbf{C}^\top \\ &= \mathbf{A} \mathbf{X} + \mathbf{A}^\top \mathbf{X} \end{aligned}$$

where  $\mathbf{B} \triangleq \mathbf{A} \mathbf{X}$  and  $\mathbf{C} \triangleq \mathbf{X}^\top \mathbf{A}$ .

## Derivative of $f(X) = \text{tr}(AXBXC)$

- **Function.** For any matrices  $A \in \mathbb{R}^{m \times n}$ ,  $X \in \mathbb{R}^{n \times d}$ ,  $B \in \mathbb{R}^{d \times n}$ , and  $C \in \mathbb{R}^{d \times m}$ , what is the derivative of  $f(X) = \text{tr}(AXBXC)$ ?
- **Derivative.**

$$\begin{aligned} \frac{d f(X)}{d X} &= \frac{d \text{tr}(AXD)}{d X} + \frac{d \text{tr}(EXC)}{d X} \\ &= A^\top D^\top + E^\top C^\top \\ &= A^\top C^\top X^\top B^\top + B^\top X^\top A^\top C^\top \end{aligned}$$

where  $D \triangleq BXC$  and  $E \triangleq AXB$ .

## Derivative of $f(\mathbf{X}) = \|\mathbf{A}\mathbf{X}\|_F^2$

- **Function.** For any matrices  $\mathbf{A} \in \mathbb{R}^{m \times n}$  and  $\mathbf{X} \in \mathbb{R}^{n \times d}$ , what is the derivative of  $f(\mathbf{X}) = \|\mathbf{A}\mathbf{X}\|_F^2$ ?
- **Derivative.** Since

$$f(\mathbf{X}) = \text{tr}(\mathbf{X}^\top \mathbf{A}^\top \mathbf{A} \mathbf{X})$$

Hence, we have

$$\begin{aligned} \frac{d f(\mathbf{X})}{d \mathbf{X}} &= \frac{d \text{tr}(\mathbf{B} \mathbf{X})}{d \mathbf{X}} + \frac{d \text{tr}(\mathbf{X}^\top \mathbf{B}^\top)}{d \mathbf{X}} \\ &= \frac{d \text{tr}(\mathbf{B} \mathbf{X})}{d \mathbf{X}} + \frac{d \text{tr}(\mathbf{B} \mathbf{X})}{d \mathbf{X}} \\ &= 2\mathbf{B}^\top \\ &= 2\mathbf{A}^\top \mathbf{A} \mathbf{X} \end{aligned}$$

where  $\mathbf{B} \triangleq \mathbf{X}^\top \mathbf{A}^\top \mathbf{A}$ .

## Orthogonal Procrustes Problem (Optional)

- **Orthogonal Procrustes problem:**

For any  $Q \in \mathbb{R}^{m \times r}$ ,  $m \geq r$ , the solution to

$$\begin{aligned} \min_F \quad & \|F - Q\|_F^2 \\ \text{s. t.} \quad & \underbrace{F^\top F = I_r}_{\text{orthogonal}} \end{aligned}$$

is

$$F := UV^\top$$

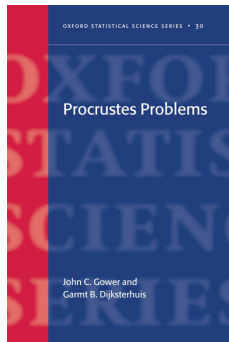
where

$$Q = \underbrace{U \Sigma V^\top}_{\text{singular value decomposition}}$$

- Equivalent form:

$$\|F - Q\|_F^2 = \text{tr}(\underbrace{F^\top F}_{=I_r} - F^\top Q - Q^\top F + \underbrace{Q^\top Q}_{\text{const.}}) = -2 \text{tr}(F^\top Q) + \text{const.}$$

$$\implies F =: \arg \min_{F^\top F = I_r} \|F - Q\|_F^2 = \arg \max_{F^\top F = I_r} \text{tr}(F^\top Q)$$





## Orthogonal Procrustes Problem (Optional)

**Example.** For any matrices  $A \in \mathbb{R}^{m \times n}$ ,  $X \in \mathbb{R}^{n \times d}$ , and  $B \in \mathbb{R}^{d \times m}$ , write down the closed-form solution to the optimization problem such that

$$\begin{aligned} \max_X \quad & \sum_{k=1}^{\tau} \text{tr}(A_k X B_k) \\ \text{s.t.} \quad & X^\top X = I_d \end{aligned}$$

where  $n \geq d$  and  $\tau \in \mathbb{Z}^+$ .

The closed-form solution to  $X$  is

$$X := UV^\top$$

where

$$\underbrace{\sum_{k=1}^{\tau} A_k^\top B_k^\top}_{\text{singular value decomposition}} = U \Sigma V^\top$$

## A Quick Look

### Content:

- Vector structure,  $\ell_2$ -norm
- Matrix structure, Frobenius norm
- Inner product
- Definition, properties, and derivatives of matrix trace (including a lot of examples)

### For your need!

- Slides: [https://xinychen.github.io/slides/matrix\\_trace.pdf](https://xinychen.github.io/slides/matrix_trace.pdf)
- E-book: [https://xinychen.github.io/books/spatiotemporal\\_low\\_rank\\_models.pdf](https://xinychen.github.io/books/spatiotemporal_low_rank_models.pdf)

### Reference material:

- The matrix cookbook: <https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf>

### **Special thanks to**

- Zhanhong Cheng (McGill University)
- Benzhenh Li (UESTC)
- Fuqiang Liu (McGill University)
- Sheng Liu (UESTC)

for providing feedback and suggestions.

# Thanks for your attention!

## Any Questions?

### About me:

- Homepage: <https://xinychen.github.io>
- How to reach me: [chenxy346@gmail.com](mailto:chenxy346@gmail.com)