Definition, Properties, and Derivatives of Matrix Traces

A Class for Graduate Students

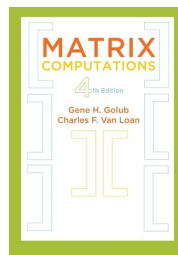**Xinyu Chen**

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

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

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

---

[1] The set of integers, $\{1, 2, \ldots, n\}$, is represented by $[n]$, $n \in \mathbb{Z}^+$.

Vector & Matrix

Notation:

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

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

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

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

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

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

---

[1] The set of integers, $\{1, 2, \ldots, n\}$, is represented by $[n]$, $n \in \mathbb{Z}^+$.

Vector Norms

A number of concepts to mention, e.g., ℓ_0-norm, ℓ_1-norm, and ℓ_2-norm.

- **Definition.** For any vector $\boldsymbol{x} \in \mathbb{R}^n$, the ℓ_2-norm of $\boldsymbol{x}$ is given by

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

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

Vector Norms

A number of concepts to mention, e.g., ℓ_0-norm, ℓ_1-norm, and ℓ_2-norm.

- **Definition.** For any vector $\boldsymbol{x} \in \mathbb{R}^n$, the ℓ_2-norm of $\boldsymbol{x}$ is given by

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

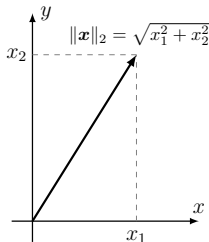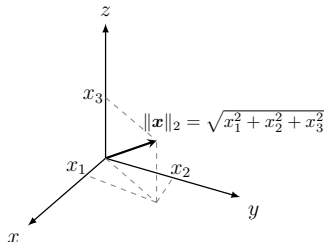where x_i, $\forall i \in [n]$ is the i-th entry of $\boldsymbol{x}$.

- Intuitive examples:

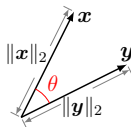On $\boldsymbol{x} = (x_1, x_2)^\top$ On $\boldsymbol{x} = (x_1, x_2, x_3)^\top$

Inner Product

- Basics: For any $\boldsymbol{x} = (x_1, x_2)^\top$ and $\boldsymbol{y} = (y_1, y_2)^\top$, the angle θ 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

  - ℓ_2-norm:

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

  - inner product:
  $$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^\top \boldsymbol{y} = x_1 y_1 + x_2 y_2$$

- It leads to

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

Inner Product

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

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

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

In this case,

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

Inner Product

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

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

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

In this case,

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

Frobenius Norm

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

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

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

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

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

Frobenius Norm

- Connection with ℓ_2-norm:

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

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

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

Basics
ooooooo

Definition
●

Properties
ooooooooo

Derivatives
ooooooooo

Applications
oo

Summary
ooo

Definition of Matrix Trace

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

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

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

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

$$\mathrm{tr}(\boldsymbol{X}) = 2 + 2 + 3 = 7$$

Property: $\operatorname{tr}(\boldsymbol{X} + \boldsymbol{Y}) = \operatorname{tr}(\boldsymbol{X}) + \operatorname{tr}(\boldsymbol{Y})$

- **Property.** For any square matrices $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{n \times n}$, it always holds that
$$\operatorname{tr}(\boldsymbol{X} + \boldsymbol{Y}) = \operatorname{tr}(\boldsymbol{X}) + \operatorname{tr}(\boldsymbol{Y})$$

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

In this case,

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

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

- **Variant.** For any $\alpha, \beta \in \mathbb{R}$, we have
$$\operatorname{tr}(\alpha \boldsymbol{X} + \beta \boldsymbol{Y}) = \alpha \operatorname{tr}(\boldsymbol{X}) + \beta \operatorname{tr}(\boldsymbol{Y})$$

Property: $\operatorname{tr}(\boldsymbol{XY}) = \operatorname{tr}(\boldsymbol{YX})$

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

$$\operatorname{tr}(\boldsymbol{XY}) = \operatorname{tr}(\boldsymbol{YX})$$

- **Proof.**

$$\mathrm{tr}(\boldsymbol{XY}) = [\boldsymbol{XY}]_{11} + [\boldsymbol{XY}]_{22} + \cdots + [\boldsymbol{XY}]_{mm}$$

$$= \underbrace{x_{11}y_{11} + x_{12}y_{21} + \cdots + x_{1n}y_{n1}}_{\text{the first row of } \boldsymbol{X} \text{ times the first column of } \boldsymbol{Y}}$$

$$+ x_{21}y_{12} + x_{22}y_{22} + \cdots + x_{2n}y_{n2}$$

$$+ \cdots + x_{m1}y_{1m} + x_{m2}y_{2m} + \cdots + x_{mn}y_{nm}$$

- **Proof.**

$$\text{tr}(\boldsymbol{XY}) = [\boldsymbol{XY}]_{11} + [\boldsymbol{XY}]_{22} + \cdots + [\boldsymbol{XY}]_{mm}$$

$$= \underbrace{x_{11}y_{11} + x_{12}y_{21} + \cdots + x_{1n}y_{n1}}_{\text{the first row of } \boldsymbol{X} \text{ times the first column of } \boldsymbol{Y}}$$

$$+ x_{21}y_{12} + x_{22}y_{22} + \cdots + x_{2n}y_{n2}$$

$$+ \cdots + x_{m1}y_{1m} + x_{m2}y_{2m} + \cdots + x_{mn}y_{nm}$$

$$= x_{11}y_{11} + x_{12}y_{21} + \cdots + x_{1n}y_{n1}$$

$$+ x_{21}y_{12} + x_{22}y_{22} + \cdots + x_{2n}y_{n2}$$

$$+ \cdots + x_{m1}y_{1m} + x_{m2}y_{2m} + \cdots + x_{mn}y_{nm}$$

- **Proof.**

$$\mathrm{tr}(\boldsymbol{XY}) = [\boldsymbol{XY}]_{11} + [\boldsymbol{XY}]_{22} + \cdots + [\boldsymbol{XY}]_{mm}$$

$$= \underbrace{x_{11}y_{11} + x_{12}y_{21} + \cdots + x_{1n}y_{n1}}$$

the first row of $\boldsymbol{X}$ times the first column of $\boldsymbol{Y}$

$$+ x_{21}y_{12} + x_{22}y_{22} + \cdots + x_{2n}y_{n2}$$

$$+ \cdots + x_{m1}y_{1m} + x_{m2}y_{2m} + \cdots + x_{mn}y_{nm}$$

$$= x_{11}y_{11} + x_{12}y_{21} + \cdots + x_{1n}y_{n1}$$

$$+ x_{21}y_{12} + x_{22}y_{22} + \cdots + x_{2n}y_{n2}$$

$$+ \cdots + x_{m1}y_{1m} + x_{m2}y_{2m} + \cdots + x_{mn}y_{nm}$$

$$= \underbrace{y_{11}x_{11} + y_{12}x_{21} + \cdots + y_{1m}x_{m1}}$$

the first row of $\boldsymbol{Y}$ times the first column of $\boldsymbol{X}$

$$+ y_{21}x_{12} + y_{22}x_{22} + \cdots + y_{2m}x_{m2}$$

$$+ \cdots + y_{n1}x_{1n} + \cdots + y_{n2}x_{2n} + \cdots + y_{nm}x_{mn}$$

$$= [\boldsymbol{YX}]_{11} + [\boldsymbol{YX}]_{22} + \cdots + [\boldsymbol{YX}]_{nn} = \mathrm{tr}(\boldsymbol{YX})$$

Property: $\mathrm{tr}(\boldsymbol{XY}) = \mathrm{tr}(\boldsymbol{YX})$

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

In this case,

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

Thus,

$$\mathrm{tr}(\boldsymbol{XY}) = 3 + 2 + 6 = 11 \qquad \mathrm{tr}(\boldsymbol{YX}) = 3 + 3 + 5 = 11$$

Property: $\mathrm{tr}(\boldsymbol{AXB}) = \mathrm{tr}(\boldsymbol{BAX}) = \mathrm{tr}(\boldsymbol{XBA})$

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

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

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

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

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

- **Proof.**

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

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

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

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

- **Proof.**

$$\begin{aligned}
\mathrm{tr}(X^\top Y) =& [X^\top Y]_{11} + [X^\top Y]_{22} + \cdots + [X^\top Y]_{nn} \\
=& x_{11}y_{11} + x_{21}y_{21} + \cdots + x_{m1}y_{m1} \\
& + x_{12}y_{12} + x_{22}y_{22} + \cdots + x_{m2}y_{m2} \\
& + \cdots + x_{1n}y_{1n} + x_{2n}y_{2n} + \cdots + x_{mn}y_{mn} \\
=& \langle \boldsymbol{x}_1, \boldsymbol{y}_1 \rangle + \langle \boldsymbol{x}_2, \boldsymbol{y}_2 \rangle + \cdots + \langle \boldsymbol{x}_n, \boldsymbol{y}_n \rangle \\
=& \langle X, Y \rangle
\end{aligned}$$

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

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

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

Recall that

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

For the matrix,

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

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

Derivatives

A quick revisit!

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

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

Derivatives

A quick revisit!

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

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

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

$$\begin{cases} \dfrac{\partial f(x, y)}{\partial x} = \lim_{\Delta x \to 0} \dfrac{f(x + \Delta x, y)}{\Delta x} \\ \dfrac{\partial f(x, y)}{\partial y} = \lim_{\Delta y \to 0} \dfrac{f(x, y + \Delta y)}{\Delta y} \end{cases}$$

Derivatives

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

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

$$f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_n) = x_1^2 + x_2^2 + \cdots + x_n^2$$

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

$$\frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_1} = 2x_1$$
$$\frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_2} = 2x_2$$
$$\vdots$$
$$\frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_n} = 2x_n$$

$$\implies \quad \frac{\mathrm{d}\, f(\boldsymbol{x})}{\mathrm{d}\, \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_1} \\ \frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{bmatrix} = 2\boldsymbol{x}$$

Derivative of $f(\boldsymbol{X}) = \operatorname{tr}(\boldsymbol{X})$

- **Function.** For any square matrix $\boldsymbol{X} \in \mathbb{R}^{n \times n}$, what is the derivative of $f(\boldsymbol{X}) = \operatorname{tr}(\boldsymbol{X})$?

- **Derivative.** Since $f(\boldsymbol{X}) = \displaystyle\sum_{i=1}^{n} x_{ii}$, we have

$$
\frac{\mathrm{d}\, f(\boldsymbol{X})}{\mathrm{d}\, \boldsymbol{X}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{X})}{\partial x_{11}} & \frac{\partial f(\boldsymbol{X})}{\partial x_{12}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial x_{1n}} \\ \frac{\partial f(\boldsymbol{X})}{\partial x_{21}} & \frac{\partial f(\boldsymbol{X})}{\partial x_{22}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial x_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(\boldsymbol{X})}{\partial x_{n1}} & \frac{\partial f(\boldsymbol{X})}{\partial x_{n2}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial x_{nn}} \end{bmatrix}
$$

$$
= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \boldsymbol{I}_n
$$

Derivative of $f(\boldsymbol{X}) = \operatorname{tr}(\boldsymbol{A}\boldsymbol{X})$

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

- **Derivative.** Since

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

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

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

As a result, we have

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

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

Derivative of $f(\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{AXB})$

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

- **Derivative.** Since

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

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

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

As a result, we have

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

Derivative of $f(\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{A} \boldsymbol{X})$

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

- **Derivative.**
$$\begin{aligned}
\frac{\mathrm{d}\, f(\boldsymbol{X})}{\mathrm{d}\, \boldsymbol{X}} &= \frac{\mathrm{d}\, \mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{B})}{\mathrm{d}\, \boldsymbol{X}} + \frac{\mathrm{d}\, \mathrm{tr}(\boldsymbol{C} \boldsymbol{X})}{\mathrm{d}\, \boldsymbol{X}} \\
&= \frac{\mathrm{d}\, \mathrm{tr}(\boldsymbol{B}^\top \boldsymbol{X})}{\mathrm{d}\, \boldsymbol{X}} + \frac{\mathrm{d}\, \mathrm{tr}(\boldsymbol{C} \boldsymbol{X})}{\mathrm{d}\, \boldsymbol{X}} \\
&= \boldsymbol{B} + \boldsymbol{C}^\top \\
&= \boldsymbol{A} \boldsymbol{X} + \boldsymbol{A}^\top \boldsymbol{X}
\end{aligned}$$

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

Derivative of $f(\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}\boldsymbol{C})$

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

- **Derivative.**

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

where $\boldsymbol{D} \triangleq \boldsymbol{B}\boldsymbol{X}\boldsymbol{C}$ and $\boldsymbol{E} \triangleq \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}$.

Derivative of $f(\boldsymbol{X}) = \|\boldsymbol{AX}\|_F^2$

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

- **Derivative.** Since

$$f(\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{A}^\top \boldsymbol{AX})$$

Hence, we have

$$
\begin{aligned}
\frac{\mathrm{d}\, f(\boldsymbol{X})}{\mathrm{d}\, \boldsymbol{X}} &= \frac{\mathrm{d}\, \mathrm{tr}(\boldsymbol{BX})}{\mathrm{d}\, \boldsymbol{X}} + \frac{\mathrm{d}\, \mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{B}^\top)}{\boldsymbol{X}} \\
&= \frac{\mathrm{d}\, \mathrm{tr}(\boldsymbol{BX})}{\mathrm{d}\, \boldsymbol{X}} + \frac{\mathrm{d}\, \mathrm{tr}(\boldsymbol{BX})}{\mathrm{d}\, \boldsymbol{X}} \\
&= 2\boldsymbol{B}^\top \\
&= 2\boldsymbol{A}^\top \boldsymbol{AX}
\end{aligned}
$$

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

Orthogonal Procrustes Problem (Optional)

- **Orthogonal Procrustes problem**: For any $\boldsymbol{Q} \in \mathbb{R}^{m \times r}$, $m \geq r$, the solution to

$$\min_{\boldsymbol{F}} \ \|\boldsymbol{F} - \boldsymbol{Q}\|_F^2$$

$$\text{s. t.} \ \underbrace{\boldsymbol{F}^\top \boldsymbol{F} = \boldsymbol{I}_r}_{\text{orthogonal}}$$

is

$$\boldsymbol{F} := \boldsymbol{U}\boldsymbol{V}^\top$$

where

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

OXFORD STATISTICAL SCIENCE SERIES · 30

Procrustes Problems

John C. Gower and
Garmt B. Dijksterhuis

- Equivalent form:

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

$$\Longrightarrow \boldsymbol{F} =: \underset{\boldsymbol{F}^\top \boldsymbol{F} = \boldsymbol{I}_r}{\arg\min} \ \|\boldsymbol{F} - \boldsymbol{Q}\|_F^2 = \underset{\boldsymbol{F}^\top \boldsymbol{F} = \boldsymbol{I}_r}{\arg\max} \ \operatorname{tr}(\boldsymbol{F}^\top \boldsymbol{Q})$$

Basics
ooooooo

Definition
o

Properties
ooooooo

Derivatives
ooooooo

Applications
oo

Summary
ooo
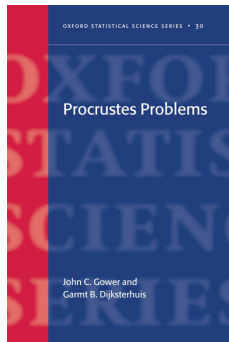
Orthogonal Procrustes Problem (Optional)

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

$$\max_{\boldsymbol{X}} \ \sum_{k=1}^{\tau} \mathrm{tr}(\boldsymbol{A}_k \boldsymbol{X} \boldsymbol{B}_k)$$

$$\text{s.t. } \boldsymbol{X}^{\top} \boldsymbol{X} = \boldsymbol{I}_d$$

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

The closed-form solution to $\boldsymbol{X}$ is

$$\boldsymbol{X} := \boldsymbol{U} \boldsymbol{V}^{\top}$$

where

$$\underbrace{\sum_{k=1}^{\tau} \boldsymbol{A}_k^{\top} \boldsymbol{B}_k^{\top}}_{\text{singular value decomposition}} = \boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{V}^{\top}$$

A Quick Look

Content:

- Vector structure, ℓ_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
- E-book: https://xinychen.github.io/books/spatiotemporal_low_rank_models.pdf

Reference material:

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

Thanks for your attention!

Any Questions?

About me:
- Homepage: https://xinychen.github.io
- How to reach me: chenxy346@gmail.com