Matrix Algebra Marathon

J2200071 Ryuto Saito

May 3, 2024

1 Statistics

1.1 Exercise

Exercise 2.6.2.

Given ℓ vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\ell \in \mathbb{R}^n$, the covariance matrix $\boldsymbol{C} \in \mathbb{S}^n_+$ is defined by

$$\boldsymbol{C} \coloneqq \frac{1}{\ell} \sum_{i=1}^{\ell} (\boldsymbol{x}_i - \boldsymbol{m})(\boldsymbol{x}_i - \boldsymbol{m})^{\mathrm{T}}. \tag{1}$$

where

$$\boldsymbol{m} \coloneqq \frac{1}{\ell} \sum_{i=1}^{\ell} \boldsymbol{x}_i. \tag{2}$$

Show that the $n \times \ell$ matrix $\boldsymbol{X} \coloneqq [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\ell]$ satisfies

$$\boldsymbol{C} + \boldsymbol{m}\boldsymbol{m}^{\mathrm{T}} = \frac{1}{\ell} \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}. \tag{3}$$

Proof. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\ell \in \mathbb{R}^n$.

$$
\begin{aligned}
\boldsymbol{C} + \boldsymbol{m}\boldsymbol{m}^{\mathrm{T}} &= \frac{1}{\ell}\sum_{i=1}^{\ell}(\boldsymbol{x}_i - \boldsymbol{m})(\boldsymbol{x}_i - \boldsymbol{m})^{\mathrm{T}} + \boldsymbol{m}\boldsymbol{m}^{\mathrm{T}} \\
&= \frac{1}{\ell}\sum_{i=1}^{\ell}(\boldsymbol{x}_i - \boldsymbol{m})(\boldsymbol{x}_i^{\mathrm{T}} - \boldsymbol{m}^{\mathrm{T}}) + \boldsymbol{m}\boldsymbol{m}^{\mathrm{T}} \\
&= \frac{1}{\ell}\sum_{i=1}^{\ell}(\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}} - \boldsymbol{x_i}\boldsymbol{m}^{\mathrm{T}} - \boldsymbol{m}\boldsymbol{x}_i^{\mathrm{T}} + \boldsymbol{m}\boldsymbol{m}^{\mathrm{T}}) + \boldsymbol{m}\boldsymbol{m}^{\mathrm{T}} \\
&= \frac{1}{\ell}\left(\sum_{i=1}^{\ell}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}} - \sum_{i=1}^{\ell}\boldsymbol{x}_i\boldsymbol{m}^{\mathrm{T}} - \sum_{i=1}^{\ell}\boldsymbol{m}\boldsymbol{x}_i^{\mathrm{T}} + \sum_{i=1}^{\ell}\boldsymbol{m}\boldsymbol{m}^{\mathrm{T}}\right) + \boldsymbol{m}\boldsymbol{m}^{\mathrm{T}} \\
&= \frac{1}{\ell}\sum_{i=1}^{\ell}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}} - \left(\frac{1}{\ell}\sum_{i=1}^{\ell}\boldsymbol{x}_i\right)\boldsymbol{m}^{\mathrm{T}} - \boldsymbol{m}\left(\frac{1}{\ell}\sum_{i=1}^{\ell}\boldsymbol{x}_i^{\mathrm{T}}\right) + \frac{1}{\ell}\sum_{i=1}^{\ell}\boldsymbol{m}\boldsymbol{m}^{\mathrm{T}} + \boldsymbol{m}\boldsymbol{m}^{\mathrm{T}} \\
&= \frac{1}{\ell}\sum_{i=1}^{\ell}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}} - \boldsymbol{m}\boldsymbol{m}^{\mathrm{T}} - \boldsymbol{m}\left(\frac{1}{\ell}\sum_{i=1}^{\ell}\boldsymbol{x}_i^{\mathrm{T}}\right) + \boldsymbol{m}\boldsymbol{m}^{\mathrm{T}} + \boldsymbol{m}\boldsymbol{m}^{\mathrm{T}} \\
&= \frac{1}{\ell}\sum_{i=1}^{\ell}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}} - \boldsymbol{m}\left(\frac{1}{\ell}\sum_{i=1}^{\ell}\boldsymbol{x}_i\right)^{\mathrm{T}} + \boldsymbol{m}\boldsymbol{m}^{\mathrm{T}} \\
&= \frac{1}{\ell}\sum_{i=1}^{\ell}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}} - \boldsymbol{m}\boldsymbol{m}^{\mathrm{T}} + \boldsymbol{m}\boldsymbol{m}^{\mathrm{T}} \\
&= \frac{1}{\ell}\sum_{i=1}^{\ell}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{\ell}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} &= \frac{1}{\ell}[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\ell]\begin{bmatrix}\boldsymbol{x}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_\ell^{\mathrm{T}}\end{bmatrix} \\
&= \frac{1}{\ell}\sum_{i=1}^{\ell}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}
\end{aligned}
$$

Thus, we get (3). $\square$

2 Idempotent Matrices

2.1 Exercise

Exercise 2.7.2.

Let $\boldsymbol{v} \in \boldsymbol{\Delta}_\ell$ where $\boldsymbol{\Delta}_\ell$ denotes the ℓ-dimensional probablistic simplex: $\boldsymbol{\Delta}_\ell \coloneqq \{\boldsymbol{x} \in \mathbb{R}_+^\ell \mid \boldsymbol{x}^\mathrm{T}\mathbf{1}_\ell = 1\}$. Show that the $\ell \times \ell$ matrix

$$\boldsymbol{K} \coloneqq \boldsymbol{I} - \mathbf{1}_\ell \boldsymbol{v}^\mathrm{T}$$

satisfies $\boldsymbol{K}^2 = \boldsymbol{K}$.

Proof. Let $\boldsymbol{v} \in \boldsymbol{\Delta}_\ell$.

$$\begin{aligned}
\boldsymbol{K}^2 &= (\boldsymbol{I} - \mathbf{1}_\ell \boldsymbol{v}^\mathrm{T})^2 \\
&= \boldsymbol{I} - \mathbf{1}_\ell \boldsymbol{v}^\mathrm{T} - \mathbf{1}_\ell \boldsymbol{v}^\mathrm{T} + \mathbf{1}_\ell \boldsymbol{v}^\mathrm{T} \mathbf{1}_\ell \boldsymbol{v}^\mathrm{T} \\
&= \boldsymbol{I} - 2\mathbf{1}_\ell \boldsymbol{v}^\mathrm{T} + \mathbf{1}_\ell (\boldsymbol{v}^\mathrm{T} \mathbf{1}_\ell) \boldsymbol{v}^\mathrm{T} \\
&= \boldsymbol{I} - 2\mathbf{1}_\ell \boldsymbol{v}^\mathrm{T} + \mathbf{1}_\ell \boldsymbol{v}^\mathrm{T} \\
&= \boldsymbol{I} - \mathbf{1}_\ell \boldsymbol{v}^\mathrm{T} \\
&= \boldsymbol{K}
\end{aligned}$$

□

3 Trace

3.1 Definition

Definition 3.1.1. *Let $\boldsymbol{A}$ be an $n \times n$ square matrix. The trace of $\boldsymbol{A}$ is defined as*

$$tr(\boldsymbol{A}) = \sum_{i=1}^{n} A_{i,i}. \tag{4}$$

3.2 Exercise

Exercise 2.8.3.

For $\forall a, \forall b \in \mathbb{R}$ and $\forall \boldsymbol{X}, \forall \boldsymbol{Y} \in \mathbb{R}^{n \times n}$, show that

$$tr(a\boldsymbol{X} + b\boldsymbol{Y}) = a tr(\boldsymbol{X}) + b tr(\boldsymbol{Y}). \tag{5}$$

Proof. For all $a, b \in \mathbb{R}$ and $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{n \times n}$,

$$\begin{aligned}
tr(a\boldsymbol{X} + b\boldsymbol{Y}) &= \sum_{i=1}^{n} (a\boldsymbol{X}_{i,i} + b\boldsymbol{Y}_{i,i}) \\
&= a \sum_{i=1}^{n} \boldsymbol{X}_{i,i} + b \sum_{i=1}^{n} \boldsymbol{Y}_{i,i} \\
&= atr(\boldsymbol{X}) + btr(\boldsymbol{Y}).
\end{aligned}$$

□

Exercise 2.8.5.

Show that $tr(\boldsymbol{AB}) = tr(\boldsymbol{BA})$ for $\boldsymbol{A} \in \mathbb{R}^{m \times n}, \boldsymbol{B} \in \mathbb{R}^{n \times m}$. Then, show that $tr(\boldsymbol{ABC}) = tr(\boldsymbol{CAB}) = tr(\boldsymbol{BCA})$ for $\boldsymbol{A} \in \mathbb{R}^{m \times n}, \boldsymbol{B} \in \mathbb{R}^{n \times p}, \boldsymbol{C} \in \mathbb{R}^{p \times m}$.

Proof. Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}, \boldsymbol{B} \in \mathbb{R}^{n \times m}$.

$$\begin{aligned}
tr(\boldsymbol{AB}) &= \sum_{i=1}^{m} (\boldsymbol{AB})_{i,i} \\
&= \sum_{i=1}^{m} \sum_{j=1}^{n} A_{i,j} B_{j,i} \\
&= \sum_{j=1}^{n} \sum_{i=1}^{m} B_{j,i} A_{i,j} \\
&= \sum_{j=1}^{n} (\boldsymbol{BA})_{j,j} \\
&= tr(\boldsymbol{BA})
\end{aligned}$$

Thus, we get

$$tr(\boldsymbol{AB}) = tr(\boldsymbol{BA}). \tag{6}$$

Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}, \boldsymbol{B} \in \mathbb{R}^{n \times p}, \boldsymbol{C} \in \mathbb{R}^{p \times m}$. Since (6),

$$\begin{aligned}
tr(\boldsymbol{ABC}) &= tr((\boldsymbol{AB})\boldsymbol{C}) \\
&= tr(\boldsymbol{C}(\boldsymbol{AB})) \\
&= tr(\boldsymbol{CAB}).
\end{aligned}$$

Moreover

$$\begin{aligned}
tr(\boldsymbol{ABC}) &= tr(\boldsymbol{A}(\boldsymbol{BC})) \\
&= tr((\boldsymbol{BC})\boldsymbol{A}) \\
&= tr(\boldsymbol{BCA}).
\end{aligned}$$

Therefore, it follows that

$$tr(\boldsymbol{ABC}) = tr(\boldsymbol{CAB}) = tr(\boldsymbol{BCA}).$$

$\square$

Exercise 2.8.8.

For $\forall \boldsymbol{x}, \forall \boldsymbol{y} \in \mathbb{R}^n$, show that

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = tr(\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}) \tag{7}$$

Proof. Since (6),

$$\begin{aligned} tr(\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}) &= tr(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}) \\ &= tr(\langle \boldsymbol{y}, \boldsymbol{x} \rangle) \\ &= \langle \boldsymbol{y}, \boldsymbol{x} \rangle \\ &= \langle \boldsymbol{x}, \boldsymbol{y} \rangle. \end{aligned}$$

$\square$

4 Inner-Product of Matrices

4.1 Definition

Definition 4.1.1. *The inner-product of matrices is defined as*

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle := \sum_{i=1}^{m} \sum_{j=1}^{n} X_{i,j} Y_{i,j} \tag{8}$$

where $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{m \times n}$.

4.2 Exercise

Exercise 2.10.2.

For $\forall \boldsymbol{X}, \forall \boldsymbol{Y} \in \mathbb{R}^{m \times n}$, derive the equalities:

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \langle \boldsymbol{Y}, \boldsymbol{X} \rangle = tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}) = tr(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}) \tag{9}$$

Proof. Let $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{m \times n}$.

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} X_{i,j} Y_{i,j} = \sum_{i=1}^{m} \sum_{j=1}^{n} Y_{i,j} X_{i,j} = \langle \boldsymbol{Y}, \boldsymbol{X} \rangle$$

$$
\begin{aligned}
tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}) &= \sum_{j=1}^{n}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y})_{j,j} \\
&= \sum_{j=1}^{n}\sum_{i=1}^{m}(\boldsymbol{X}^{\mathrm{T}})_{j,i}Y_{i,j} \\
&= \sum_{j=1}^{n}\sum_{i=1}^{m}X_{i,j}Y_{i,j} \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}X_{i,j}Y_{i,j} \\
&= \langle \boldsymbol{X}, \boldsymbol{Y} \rangle \\
tr(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}) &= \sum_{j=1}^{n}(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X})_{j,j} \\
&= \sum_{j=1}^{n}\sum_{i=1}^{m}(\boldsymbol{Y}^{\mathrm{T}})_{j,i}X_{i,j} \\
&= \sum_{j=1}^{n}\sum_{i=1}^{m}Y_{i,j}X_{i,j} \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}Y_{i,j}X_{i,j} \\
&= \langle \boldsymbol{Y}, \boldsymbol{X} \rangle
\end{aligned}
$$

Therefore,

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \langle \boldsymbol{Y}, \boldsymbol{X} \rangle = tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}) = tr(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}).$$

□

Exercise 2.10.5.

For $\forall a, \forall b \in \mathbb{R}, \forall \boldsymbol{X}, \forall \boldsymbol{Y} \in \mathbb{R}^{m\times n}$, show that

$$\langle a\boldsymbol{X}, b\boldsymbol{Y} \rangle = ab\langle \boldsymbol{X}, \boldsymbol{Y} \rangle. \tag{10}$$

Proof. Let $a, b \in \mathbb{R}$ and $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{m\times n}$.

$$
\begin{aligned}
\langle a\boldsymbol{X}, b\boldsymbol{Y} \rangle &= \sum_{i=1}^{m}\sum_{j=1}^{n}aX_{i,j}bY_{i,j} \\
&= ab\sum_{i=1}^{m}\sum_{j=1}^{n}X_{i,j}Y_{i,j} \\
&= ab\langle \boldsymbol{X}, \boldsymbol{Y} \rangle
\end{aligned}
$$

□

Exercise 2.10.7.

For $\forall \boldsymbol{X} \in \mathbb{R}^{m\times n}$, show that

$$\langle \boldsymbol{X}, \boldsymbol{E}_{i,j} \rangle = tr(\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{E}_{i,j}) = X_{i,j} \tag{11}$$

where $\boldsymbol{E}_{i,j}$ denotes an $m \times n$ matrix in which (i,j)th entry is one and all the others are zero.

Proof. Let $\boldsymbol{X} \in \mathbb{R}^{m\times n}$. Recall (9). It leads

$$\langle \boldsymbol{X}, \boldsymbol{E}_{i,j} \rangle = tr(\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{E}_{i,j}).$$

Furthermore, since $(\boldsymbol{E}_{i,j})_{k,l} = \begin{cases} 1 & (k = i, l = j) \\ 0 & (k \neq i \ or \ l \neq j), \end{cases}$

$$\begin{aligned} \langle \boldsymbol{X}, \boldsymbol{E}_{i,j} \rangle &= \sum_{k=1}^{m} \sum_{l=1}^{n} X_{k,l} (\boldsymbol{E}_{i,j})_{k,l} \\ &= X_{i,j}. \end{aligned}$$

Therefore, we get (11). □

5 Frobenius Norm

5.1 Definition

Definition 5.1.1. *The Frobenius norm of a matrix $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ is defined as*

$$\|\boldsymbol{X}\|_{\mathrm{F}} \coloneqq \sqrt{\langle \boldsymbol{X}, \boldsymbol{X} \rangle}.$$

5.2 Exercise

Exercise 2.11.3.

For $\forall \boldsymbol{X} \in \mathbb{R}^{m\times n}$, show that

$$\|\boldsymbol{X}\|_{\mathrm{F}} \geq 0.$$

Proof. Let $\boldsymbol{X} \in \mathbb{R}^{m\times n}$, and $k \in \mathbb{N}_m, l \in \mathbb{N}_n$.

$$\|\boldsymbol{X}\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} X_{i,j}^2} \geq \sqrt{X_{k,l}^2} = |X_{k,l}| \geq 0$$

Thus, it follows that

$$\|\boldsymbol{X}\|_{\mathrm{F}} \geq 0.$$

□

Exercise 2.11.5.

Let $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{m\times n}$. Derive the equality:

$$\|\boldsymbol{X}+\boldsymbol{Y}\|_{\mathrm{F}}^2 = \|\boldsymbol{X}\|_{\mathrm{F}}^2 + \|\boldsymbol{Y}\|_{\mathrm{F}}^2 + 2\langle \boldsymbol{X}, \boldsymbol{Y}\rangle. \tag{12}$$

Proof. Let $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{m\times n}$.

$$\begin{aligned}
\|\boldsymbol{X}+\boldsymbol{Y}\|_{\mathrm{F}}^2 &= \langle \boldsymbol{X}+\boldsymbol{Y}, \boldsymbol{X}+\boldsymbol{Y}\rangle \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}(X_{i,j}+Y_{i,j})^2 \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}(X_{i,j}^2+Y_{i,j}^2+2X_{i,j}Y_{i,j}) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}X_{i,j}^2 + \sum_{i=1}^{m}\sum_{j=1}^{n}Y_{i,j}^2 + 2\sum_{i=1}^{m}\sum_{j=1}^{n}X_{i,j}Y_{i,j} \\
&= \langle \boldsymbol{X}, \boldsymbol{X}\rangle + \langle \boldsymbol{Y}, \boldsymbol{Y}\rangle + 2\langle \boldsymbol{X}, \boldsymbol{Y}\rangle \\
&= \sqrt{\langle \boldsymbol{X}, \boldsymbol{X}\rangle}^2 + \sqrt{\langle \boldsymbol{Y}, \boldsymbol{Y}\rangle}^2 + 2\langle \boldsymbol{X}, \boldsymbol{Y}\rangle \\
&= \|\boldsymbol{X}\|_{\mathrm{F}}^2 + \|\boldsymbol{Y}\|_{\mathrm{F}}^2 + 2\langle \boldsymbol{X}, \boldsymbol{Y}\rangle
\end{aligned} \tag{13}$$

□

Exercise 2.11.7.

For $\forall\boldsymbol{X}, \forall\boldsymbol{Y} \in \mathbb{R}^{m\times n}$, show that

$$\|\boldsymbol{X}+\boldsymbol{Y}\|_{\mathrm{F}} \le \|\boldsymbol{X}\|_{\mathrm{F}} + \|\boldsymbol{Y}\|_{\mathrm{F}}. \tag{14}$$

Proof. Let $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{m\times n}, t \in \mathbb{R}$. Consider the function

$$\begin{aligned}
f(t) &= \|\boldsymbol{X}+t\boldsymbol{Y}\|_{\mathrm{F}} \\
&= \langle \boldsymbol{X}+t\boldsymbol{Y}, \boldsymbol{X}+t\boldsymbol{Y}\rangle \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}(X_{i,j}+tY_{i,j})^2 \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}(X_{i,j}^2+t^2Y_{i,j}^2+2tX_{i,j}Y_{i,j}) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}X_{i,j}^2 + t^2\sum_{i=1}^{m}\sum_{j=1}^{n}Y_{i,j}^2 + 2tX_{i,j}Y_{i,j} \\
&= \langle \boldsymbol{X}, \boldsymbol{X}\rangle + t^2\langle \boldsymbol{Y}, \boldsymbol{Y}\rangle + 2t\langle \boldsymbol{X}, \boldsymbol{Y}\rangle \\
&= \|\boldsymbol{Y}\|_{\mathrm{F}}^2 t^2 + 2\langle \boldsymbol{X}, \boldsymbol{Y}\rangle t + \|\boldsymbol{X}\|_{\mathrm{F}}^2 \ge 0.
\end{aligned}$$

Then, the quadratic equation

$$\|\boldsymbol{Y}\|_{\mathrm{F}}^2 t^2 + 2\langle \boldsymbol{X}, \boldsymbol{Y}\rangle t + \|\boldsymbol{X}\|_{\mathrm{F}}^2 = 0$$

has at most one solution. This implies that its discriminant must be less or zero, that is

$$(2\langle \boldsymbol{X}, \boldsymbol{Y}\rangle)^2 - 4\|\boldsymbol{Y}\|_{\mathrm{F}}^2\|\boldsymbol{X}\|_{\mathrm{F}}^2 \leq 0.$$

Hence

$$\begin{aligned} 4\langle \boldsymbol{X}, \boldsymbol{Y}\rangle^2 &\leq 4\|\boldsymbol{X}\|_{\mathrm{F}}^2\|\boldsymbol{Y}\|_{\mathrm{F}}^2 \\ \langle \boldsymbol{X}, \boldsymbol{Y}\rangle^2 &\leq \|\boldsymbol{X}\|_{\mathrm{F}}^2\|\boldsymbol{Y}\|_{\mathrm{F}}^2. \end{aligned}$$

It follows that

$$-\|\boldsymbol{X}\|_{\mathrm{F}}\|\boldsymbol{Y}\|_{\mathrm{F}} \leq \langle \boldsymbol{X}, \boldsymbol{Y}\rangle \leq \|\boldsymbol{X}\|_{\mathrm{F}}\|\boldsymbol{Y}\|_{\mathrm{F}}.$$

This also implies

$$\langle \boldsymbol{X}, \boldsymbol{Y}\rangle \leq \|\boldsymbol{X}\|_{\mathrm{F}}\|\boldsymbol{Y}\|_{\mathrm{F}}.$$

Therefore, recall (12),

$$\begin{aligned} \|\boldsymbol{X} + \boldsymbol{Y}\|_{\mathrm{F}} &= \sqrt{\|\boldsymbol{X}\|_{\mathrm{F}}^2 + \|\boldsymbol{Y}\|_{\mathrm{F}}^2 + 2\langle \boldsymbol{X}, \boldsymbol{Y}\rangle} \\ &\leq \sqrt{\|\boldsymbol{X}\|_{\mathrm{F}}^2 + \|\boldsymbol{Y}\|_{\mathrm{F}}^2 + 2\|\boldsymbol{X}\|_{\mathrm{F}}\|\boldsymbol{Y}\|_{\mathrm{F}}} \\ &= \sqrt{(\|\boldsymbol{X}\|_{\mathrm{F}} + \|\boldsymbol{Y}\|_{\mathrm{F}})^2} \\ &= \|\boldsymbol{X}\|_{\mathrm{F}} + \|\boldsymbol{Y}\|_{\mathrm{F}}. \end{aligned}$$

□