Step-by-step expansion of $\|A - BC\|_F^2$

Andersen Ang
U.Southampton UK
angms.science

October 23, 2024

1st draft
December 23, 2018

$\|\boldsymbol{A}\|_F = \sqrt{\mathrm{Tr}\boldsymbol{A}^\top\boldsymbol{A}}$
$\langle\boldsymbol{A},\boldsymbol{B}\rangle = \mathrm{Tr}\boldsymbol{A}^\top\boldsymbol{B}$
Expand $\|\boldsymbol{A} - \boldsymbol{B}\boldsymbol{C}\|_F^2$

Content

Euclidean norm, Frobenius norm and trace

- We call $\|\boldsymbol{x}\|_2$ the Euclidean norm of a vector $\boldsymbol{x}$

$$\boldsymbol{x} = \begin{bmatrix} a \\ b \end{bmatrix} \quad \Longrightarrow \quad \|\boldsymbol{x}\|_2 = \sqrt{a^2 + b^2} = \sqrt{\sum_i x_i^2} = \sqrt{\text{Sum of (each element in } \boldsymbol{x})^2}.$$

- What about matrix: Frobenius norm

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \Longrightarrow \quad \|\boldsymbol{A}\|_F = \sqrt{\text{Sum of (element in } \boldsymbol{A})^2} = \sqrt{\sum_{ij} A_{ij}^2} = \sqrt{a^2 + b^2 + c^2 + d^2}.$$

- Trace of a matrix $\boldsymbol{B} = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$ is $\mathrm{Tr}\boldsymbol{B} = p + s =$ sum of diagonal.
- Fact: $\|\boldsymbol{A}\|_F = \sqrt{\mathrm{Tr}\boldsymbol{A}^\top \boldsymbol{A}}$

$$\begin{aligned} \|\boldsymbol{A}\|_F = \sqrt{\sum_{ij} A_{ij}^2} &= \sqrt{a^2 + b^2 + c^2 + d^2} \\ &= \sqrt{\mathrm{Tr} \begin{bmatrix} a^2 + c^2 & \text{we don't care} \\ \text{we don't care} & b^2 + d^2 \end{bmatrix}} \\ &= \sqrt{\mathrm{Tr}\left(\begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right)} = \sqrt{\mathrm{Tr}\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}^\top \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right)} = \sqrt{\mathrm{Tr}\boldsymbol{A}^\top \boldsymbol{A}} \end{aligned}$$

Inner product and trace

- $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \mathrm{Tr} \boldsymbol{A}^\top \boldsymbol{B}$
- $\langle \boldsymbol{A}, \boldsymbol{B} \rangle := \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} B_{ij}$
- Let $\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} w & x \\ y & z \end{bmatrix}$

$$
\begin{aligned}
\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \sum_{i,j} A_{ij} B_{ij} \quad &= \quad aw + bx + cy + dz \\
&= \quad \mathrm{Tr} \begin{bmatrix} aw + cy & \text{we don't care} \\ \text{we don't care} & bx + dz \end{bmatrix} \\
&= \quad \mathrm{Tr} \left(\begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} w & x \\ y & z \end{bmatrix} \right) = \mathrm{Tr} \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}^\top \begin{bmatrix} w & x \\ y & z \end{bmatrix} \right) = \mathrm{Tr} \boldsymbol{A}^\top \boldsymbol{B}
\end{aligned}
$$

Goal: expand $\|\boldsymbol{A} - \boldsymbol{B}\boldsymbol{C}\|_F^2$

- We usually consider $\|\cdot\|_F^2$ instead of $\|\cdot\|_F$
  - Reason: we hate to deal with $\sqrt{}$

- Given $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{B} \in \mathbb{R}^{m\times r}$, $\boldsymbol{C} \in \mathbb{R}^{r\times n}$

- We will see

$$\begin{aligned} f(\boldsymbol{B}) &= \langle \boldsymbol{Q}, \boldsymbol{B}^\top \boldsymbol{B}\rangle - 2\langle \boldsymbol{P}, \boldsymbol{B}\rangle, \qquad \boldsymbol{Q} = \boldsymbol{C}\boldsymbol{C}^\top, \boldsymbol{P} = \boldsymbol{A}\boldsymbol{C}^\top \\ f(\boldsymbol{C}) &= \langle \boldsymbol{Q}, \boldsymbol{C}\boldsymbol{C}^\top\rangle - 2\langle \boldsymbol{P}, \boldsymbol{C}\rangle, \qquad \boldsymbol{Q} = \boldsymbol{B}^\top\boldsymbol{B}, \boldsymbol{P} = \boldsymbol{B}^\top\boldsymbol{A}, \end{aligned}$$

Expand $\|\boldsymbol{A} - \boldsymbol{BC}\|_F^2$

$$
\begin{aligned}
&\|\boldsymbol{A} - \boldsymbol{BC}\|_F^2 \\
&= \operatorname{Tr}\Big((\boldsymbol{A} - \boldsymbol{BC})^\top (\boldsymbol{A} - \boldsymbol{BC})\Big) && \|\boldsymbol{X}\|_F^2 = \operatorname{Tr}\boldsymbol{X}^\top \boldsymbol{X} \\
&= \operatorname{Tr}\Big((\boldsymbol{A}^\top - (\boldsymbol{BC})^\top)(\boldsymbol{A} - \boldsymbol{BC})\Big) && \text{Tr distributive} \\
&= \operatorname{Tr}\Big((\boldsymbol{A}^\top - \boldsymbol{C}^\top \boldsymbol{B}^\top)(\boldsymbol{A} - \boldsymbol{BC})\Big) && (\boldsymbol{XY})^\top = \boldsymbol{Y}^\top \boldsymbol{X}^\top \\
&= \operatorname{Tr}\Big(\boldsymbol{A}^\top \boldsymbol{A} - \boldsymbol{A}^\top \boldsymbol{BC} - \boldsymbol{C}^\top \boldsymbol{B}^\top \boldsymbol{A} + \boldsymbol{C}^\top \boldsymbol{B}^\top \boldsymbol{BC}\Big) && \text{expand} \\
&= \operatorname{Tr}(\boldsymbol{A}^\top \boldsymbol{A}) - \operatorname{Tr}(\boldsymbol{A}^\top \boldsymbol{BC}) - \operatorname{Tr}(\boldsymbol{C}^\top \boldsymbol{B}^\top \boldsymbol{A}) + \operatorname{Tr}(\boldsymbol{C}^\top \boldsymbol{B}^\top \boldsymbol{BC}) && \text{Tr distributive} \\
&= c - \operatorname{Tr}(\boldsymbol{A}^\top \boldsymbol{BC}) - \operatorname{Tr}(\boldsymbol{C}^\top \boldsymbol{B}^\top \boldsymbol{A}) + \operatorname{Tr}(\boldsymbol{C}^\top \boldsymbol{B}^\top \boldsymbol{BC}) && c = \|\boldsymbol{A}\|_F^2 = \operatorname{Tr}\boldsymbol{A}^\top \boldsymbol{A} \\
&= c - \operatorname{Tr}(\boldsymbol{A}^\top \boldsymbol{BC}) - \operatorname{Tr}(\boldsymbol{C}^\top \boldsymbol{B}^\top \boldsymbol{A}) + \operatorname{Tr}(\boldsymbol{C}\boldsymbol{C}^\top \boldsymbol{B}^\top \boldsymbol{B}) && \text{cyclic invariant } \operatorname{Tr}(\boldsymbol{WXYZ}) = \operatorname{Tr}(\boldsymbol{ZWXY})
\end{aligned}
$$

Tr is transpose invariant: $\operatorname{Tr}\boldsymbol{D} = \operatorname{Tr}\boldsymbol{D}^\top$, so

$$
\operatorname{Tr}\Big(\boldsymbol{A}^\top \boldsymbol{BC}\Big) = \operatorname{Tr}\Big((\boldsymbol{A}^\top \boldsymbol{BC})^\top\Big) = \operatorname{Tr}\Big((\boldsymbol{BC})^\top \boldsymbol{A}\Big) = \operatorname{Tr}\Big(\boldsymbol{C}^\top \boldsymbol{B}^\top \boldsymbol{A}\Big).
$$

By cyclic invariant property of trace: $\mathrm{Tr}(\boldsymbol{W}\boldsymbol{X}\boldsymbol{Y}\boldsymbol{Z}) = \mathrm{Tr}(\boldsymbol{Z}\boldsymbol{W}\boldsymbol{X}\boldsymbol{Y})$

$$\|\boldsymbol{A} - \boldsymbol{B}\boldsymbol{C}\|_F^2 = \begin{cases} c - 2\mathrm{Tr}(\boldsymbol{A}^\top \boldsymbol{B}\boldsymbol{C}) + \mathrm{Tr}(\boldsymbol{C}\boldsymbol{C}^\top \boldsymbol{B}^\top \boldsymbol{B}) & (1) \\ c - 2\mathrm{Tr}(\boldsymbol{C}\boldsymbol{A}^\top \boldsymbol{B}) + \mathrm{Tr}(\boldsymbol{C}\boldsymbol{C}^\top \boldsymbol{B}^\top \boldsymbol{B}) & (2) \end{cases}$$

By definition of matrix inner product $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \mathrm{Tr}(\boldsymbol{A}^\top \boldsymbol{B})$

$$\begin{aligned} \|\boldsymbol{A} - \boldsymbol{B}\boldsymbol{C}\|_F^2 &\overset{(1)}{=} c - 2\mathrm{Tr}(\boldsymbol{A}^\top \boldsymbol{B}\boldsymbol{C}) + \mathrm{Tr}(\boldsymbol{C}\boldsymbol{C}^\top \boldsymbol{B}^\top \boldsymbol{B}) \\ &= c - 2\mathrm{Tr}((\boldsymbol{B}\boldsymbol{A}^\top)^\top \boldsymbol{C}) + \mathrm{Tr}((\boldsymbol{C}\boldsymbol{C}^\top)^\top \boldsymbol{B}^\top \boldsymbol{B}) \\ &= c - 2\langle \boldsymbol{B}^\top \boldsymbol{A}, \boldsymbol{C} \rangle + \langle \boldsymbol{C}\boldsymbol{C}^\top, \boldsymbol{B}^\top \boldsymbol{B} \rangle \\ &= c - 2\langle \boldsymbol{B}^\top \boldsymbol{A}, \boldsymbol{C} \rangle + \langle \boldsymbol{B}^\top \boldsymbol{B}, \boldsymbol{C}\boldsymbol{C}^\top \rangle \qquad \text{focus on } \boldsymbol{C} \end{aligned}$$

$$\begin{aligned} \|\boldsymbol{A} - \boldsymbol{B}\boldsymbol{C}\|_F^2 &\overset{(2)}{=} c - 2\mathrm{Tr}(\boldsymbol{C}\boldsymbol{A}^\top \boldsymbol{B}) + \mathrm{Tr}(\boldsymbol{C}\boldsymbol{C}^\top \boldsymbol{B}^\top \boldsymbol{B}) \\ &= c - 2\mathrm{Tr}((\boldsymbol{A}\boldsymbol{C}^\top)^\top \boldsymbol{B}) + \langle (\boldsymbol{C}\boldsymbol{C}^\top)^\top, \boldsymbol{B}^\top \boldsymbol{B} \rangle \\ &= c - 2\langle \boldsymbol{A}\boldsymbol{C}^\top, \boldsymbol{B} \rangle + \langle \boldsymbol{C}\boldsymbol{C}^\top, \boldsymbol{B}^\top \boldsymbol{B} \rangle \qquad \text{focus on } \boldsymbol{B} \end{aligned}$$

Last slide

- If we focus is on $\boldsymbol{C}$, we have
$$f(\boldsymbol{C}) = \|\boldsymbol{A} - \boldsymbol{B}\boldsymbol{C}\|_F^2 = \langle \boldsymbol{Q}, \boldsymbol{C}\boldsymbol{C}^\top \rangle - 2\langle \boldsymbol{P}, \boldsymbol{C} \rangle + c,$$
where $\boldsymbol{Q} = \boldsymbol{B}^\top \boldsymbol{B}$, $\boldsymbol{P} = \boldsymbol{B}^\top \boldsymbol{A}$, $c = \|\boldsymbol{A}\|_F^2$.

- If we focus is on $\boldsymbol{B}$, we have
$$f(\boldsymbol{B}) = \|\boldsymbol{A} - \boldsymbol{B}\boldsymbol{C}\|_F^2 = \langle \boldsymbol{Q}, \boldsymbol{B}^\top \boldsymbol{B} \rangle - 2\langle \boldsymbol{P}, \boldsymbol{B} \rangle + c,$$
where $\boldsymbol{Q} = \boldsymbol{C}\boldsymbol{C}^\top$, $\boldsymbol{P} = \boldsymbol{A}\boldsymbol{C}^\top$, $c = \|\boldsymbol{A}\|_F^2$.

- The minimization problems are then matrix quadratic programming:
$$\begin{aligned} &\operatorname*{argmin}_{\boldsymbol{B}} \quad \langle \boldsymbol{Q}, \boldsymbol{B}^\top \boldsymbol{B} \rangle - 2\langle \boldsymbol{P}, \boldsymbol{B} \rangle, \;\; \boldsymbol{Q} = \boldsymbol{C}\boldsymbol{C}^\top, \boldsymbol{P} = \boldsymbol{A}\boldsymbol{C}^\top \\ &\operatorname*{argmin}_{\boldsymbol{C}} \quad \langle \boldsymbol{Q}, \boldsymbol{C}\boldsymbol{C}^\top \rangle - 2\langle \boldsymbol{P}, \boldsymbol{C} \rangle, \;\; \boldsymbol{Q} = \boldsymbol{B}^\top \boldsymbol{B}, \boldsymbol{P} = \boldsymbol{B}^\top \boldsymbol{A}, \end{aligned}$$

- $\boldsymbol{Q}$ is always symmetric and positive semi-definite.
Furthermore, if $\boldsymbol{C}$ (respectively $\boldsymbol{B}$) is full rank, $\boldsymbol{Q}$ is positive definite and f is strongly convex.