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

for moneky

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be Homepage: angms.science

First draft : December 23, 2018
Last update : July 25, 2019

Given $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{B} \in \mathbb{R}^{m \times r}$, $\mathbf{C} \in \mathbb{R}^{r \times n}$. Simplify $\|\mathbf{A} - \mathbf{BC}\|_F^2$.

1. By definition of $\| \cdot \|_F^2$, we have $\|\mathbf{A}\|_F^2 = \mathrm{Tr}\,\mathbf{A}^\top \mathbf{A}$, so

$$\|\mathbf{A} - \mathbf{BC}\|_F^2 \;=\; \mathrm{Tr}\Big((\mathbf{A} - \mathbf{BC})^\top (\mathbf{A} - \mathbf{BC})\Big)$$

2a. Consider the product inside the big bracket

$$(\mathbf{A} - \mathbf{BC})^\top (\mathbf{A} - \mathbf{BC})$$

2b. Transpose is distributive, so

$$(\mathbf{A} - \mathbf{BC})^\top (\mathbf{A} - \mathbf{BC}) \;=\; (\mathbf{A}^\top - (\mathbf{BC})^\top)(\mathbf{A} - \mathbf{BC})$$

2c. By $(\mathbf{XY})^\top = \mathbf{Y}^\top \mathbf{X}^\top$

$$(\mathbf{A}^\top - (\mathbf{BC})^\top)(\mathbf{A} - \mathbf{BC}) \;=\; (\mathbf{A}^\top - \mathbf{C}^\top \mathbf{B}^\top)(\mathbf{A} - \mathbf{BC})$$

2d. Expand the product

$$(\mathbf{A}^\top - \mathbf{C}^\top \mathbf{B}^\top)(\mathbf{A} - \mathbf{BC}) \;=\; \mathbf{A}^\top \mathbf{A} - \mathbf{A}^\top \mathbf{BC} - \mathbf{C}^\top \mathbf{B}^\top \mathbf{A} + \mathbf{C}^\top \mathbf{B}^\top \mathbf{BC}$$

Therefore

$$\|\mathbf{A} - \mathbf{BC}\|_F^2 \;=\; \mathrm{Tr}\Big(\mathbf{A}^\top \mathbf{A} - \mathbf{A}^\top \mathbf{BC} - \mathbf{C}^\top \mathbf{B}^\top \mathbf{A} + \mathbf{C}^\top \mathbf{B}^\top \mathbf{BC}\Big)$$

We have

$$\|\mathbf{A}-\mathbf{B}\mathbf{C}\|_F^2 \;=\; \mathrm{Tr}\Big(\mathbf{A}^\top\mathbf{A}-\mathbf{A}^\top\mathbf{B}\mathbf{C}-\mathbf{C}^\top\mathbf{B}^\top\mathbf{A}+\mathbf{C}^\top\mathbf{B}^\top\mathbf{B}\mathbf{C}\Big)$$

3a. Trace is distributive, so

$$\begin{aligned}&\mathrm{Tr}\Big(\mathbf{A}^\top\mathbf{A}-\mathbf{A}^\top\mathbf{B}\mathbf{C}-\mathbf{C}^\top\mathbf{B}^\top\mathbf{A}+\mathbf{C}^\top\mathbf{B}^\top\mathbf{B}\mathbf{C}\Big)\\ =\;&\mathrm{Tr}(\mathbf{A}^\top\mathbf{A})-\mathrm{Tr}(\mathbf{A}^\top\mathbf{B}\mathbf{C})-\mathrm{Tr}(\mathbf{C}^\top\mathbf{B}^\top\mathbf{A})+\mathrm{Tr}(\mathbf{C}^\top\mathbf{B}^\top\mathbf{B}\mathbf{C})\end{aligned}$$

3b. Use $\|\mathbf{A}\|_F^2=\mathrm{Tr}\mathbf{A}^\top\mathbf{A}$ again

$$\begin{aligned}&\mathrm{Tr}\Big(\mathbf{A}^\top\mathbf{A}-\mathbf{A}^\top\mathbf{B}\mathbf{C}-\mathbf{C}^\top\mathbf{B}^\top\mathbf{A}+\mathbf{C}^\top\mathbf{B}^\top\mathbf{B}\mathbf{C}\Big)\\ =\;&\|\mathbf{A}\|_F^2-\mathrm{Tr}(\mathbf{A}^\top\mathbf{B}\mathbf{C})-\mathrm{Tr}(\mathbf{C}^\top\mathbf{B}^\top\mathbf{A})+\mathrm{Tr}(\mathbf{C}^\top\mathbf{B}^\top\mathbf{B}\mathbf{C})\end{aligned}$$

3c. Trace is cyclic invariant : $\mathrm{Tr}(\mathbf{W}\mathbf{X}\mathbf{Y}\mathbf{Z})=\mathrm{Tr}(\mathbf{Z}\mathbf{W}\mathbf{X}\mathbf{Y})$

$$\begin{aligned}&\mathrm{Tr}\Big(\mathbf{A}^\top\mathbf{A}-\mathbf{A}^\top\mathbf{B}\mathbf{C}-\mathbf{C}^\top\mathbf{B}^\top\mathbf{A}+\mathbf{C}^\top\mathbf{B}^\top\mathbf{B}\mathbf{C}\Big)\\ =\;&\|\mathbf{A}\|_F^2-\mathrm{Tr}(\mathbf{A}^\top\mathbf{B}\mathbf{C})-\mathrm{Tr}(\mathbf{C}^\top\mathbf{B}^\top\mathbf{A})+\mathrm{Tr}(\mathbf{C}\mathbf{C}^\top\mathbf{B}^\top\mathbf{B})\end{aligned}$$

Hence

$$\|\mathbf{A}-\mathbf{B}\mathbf{C}\|_F^2 \;=\; \|\mathbf{A}\|_F^2-\mathrm{Tr}(\mathbf{A}^\top\mathbf{B}\mathbf{C})-\mathrm{Tr}(\mathbf{C}^\top\mathbf{B}^\top\mathbf{A})+\mathrm{Tr}(\mathbf{C}\mathbf{C}^\top\mathbf{B}^\top\mathbf{B})$$

4a. Trace is transpose invariant : $\mathrm{Tr}\mathbf{D} = \mathrm{Tr}\mathbf{D}^\top$, so

$$\mathrm{Tr}\Big(\mathbf{A}^\top\mathbf{B}\mathbf{C}\Big) = \mathrm{Tr}\Big((\mathbf{A}^\top\mathbf{B}\mathbf{C})^\top\Big) = \mathrm{Tr}\Big((\mathbf{B}\mathbf{C})^\top\mathbf{A}\Big) = \mathrm{Tr}\Big(\mathbf{C}^\top\mathbf{B}^\top\mathbf{A}\Big)$$

therefore

$$\|\mathbf{A}-\mathbf{B}\mathbf{C}\|_F^2 \quad = \quad \|\mathbf{A}\|_F^2 - 2\mathrm{Tr}(\mathbf{A}^\top\mathbf{B}\mathbf{C}) + \mathrm{Tr}(\mathbf{C}\mathbf{C}^\top\mathbf{B}^\top\mathbf{B})$$

4b. As trace is cyclic invariant, we also have

$$\|\mathbf{A}-\mathbf{B}\mathbf{C}\|_F^2 \quad = \quad \|\mathbf{A}\|_F^2 - 2\mathrm{Tr}(\mathbf{C}\mathbf{A}^\top\mathbf{B}) + \mathrm{Tr}(\mathbf{C}\mathbf{C}^\top\mathbf{B}^\top\mathbf{B})$$

4c. By definition of matrix inner product $\langle\mathbf{A},\mathbf{B}\rangle = \mathrm{Tr}(\mathbf{A}^\top\mathbf{B})$

$$\begin{aligned}
\|\mathbf{A}-\mathbf{B}\mathbf{C}\|_F^2 &= \|\mathbf{A}\|_F^2 - 2\mathrm{Tr}(\mathbf{A}^\top\mathbf{B}\mathbf{C}) + \mathrm{Tr}(\mathbf{C}\mathbf{C}^\top\mathbf{B}^\top\mathbf{B}) \\
&= \|\mathbf{A}\|_F^2 - 2\mathrm{Tr}((\mathbf{B}\mathbf{A}^\top)^\top\mathbf{C}) + \mathrm{Tr}((\mathbf{C}\mathbf{C}^\top)^\top\mathbf{B}^\top\mathbf{B}) \\
&= \|\mathbf{A}\|_F^2 - 2\langle\mathbf{B}^\top\mathbf{A},\mathbf{C}\rangle + \langle\mathbf{C}\mathbf{C}^\top,\mathbf{B}^\top\mathbf{B}\rangle
\end{aligned}$$

$$\begin{aligned}
\|\mathbf{A}-\mathbf{B}\mathbf{C}\|_F^2 &= \|\mathbf{A}\|_F^2 - 2\mathrm{Tr}(\mathbf{C}\mathbf{A}^\top\mathbf{B}) + \mathrm{Tr}(\mathbf{C}\mathbf{C}^\top\mathbf{B}^\top\mathbf{B}) \\
&= \|\mathbf{A}\|_F^2 - 2\mathrm{Tr}((\mathbf{A}\mathbf{C}^\top)^\top\mathbf{B}) + \langle(\mathbf{C}\mathbf{C}^\top)^\top,\mathbf{B}^\top\mathbf{B}\rangle \\
&= \|\mathbf{A}\|_F^2 - 2\langle\mathbf{A}\mathbf{C}^\top,\mathbf{B}\rangle + \langle\mathbf{C}\mathbf{C}^\top,\mathbf{B}^\top\mathbf{B}\rangle
\end{aligned}$$

If our focus is on $\mathbf{B}$, we have

$$f(\mathbf{B}) = \|\mathbf{A} - \mathbf{BC}\|_F^2 = \langle \mathbf{Q}, \mathbf{B}^\top \mathbf{B} \rangle - 2\langle \mathbf{P}, \mathbf{B} \rangle + c,$$

where $\mathbf{Q} = \mathbf{CC}^\top$, $\mathbf{P} = \mathbf{AC}^\top$, $c = \|\mathbf{A}\|_F^2$.

If our focus is on $\mathbf{C}$, we have

$$f(\mathbf{C}) = \|\mathbf{A} - \mathbf{BC}\|_F^2 = \langle \mathbf{Q}, \mathbf{CC}^\top \rangle - 2\langle \mathbf{P}, \mathbf{C} \rangle + c,$$

where $\mathbf{Q} = \mathbf{B}^\mathbf{t}\mathbf{B}$, $\mathbf{P} = \mathbf{B}^\top \mathbf{A}$, $c = \|\mathbf{A}\|_F^2$.

The minimization problems are then matrix quadratic programming :

$$\begin{aligned} \min_{\mathbf{B}} f(\mathbf{B}) &= \min_{\mathbf{B}} \langle \mathbf{Q}, \mathbf{B}^\top \mathbf{B} \rangle - 2\langle \mathbf{P}, \mathbf{B} \rangle, \quad \mathbf{Q} = \mathbf{CC}^\top, \mathbf{P} = \mathbf{AC}^\top \\ \min_{\mathbf{C}} f(\mathbf{C}) &= \min_{\mathbf{C}} \langle \mathbf{Q}, \mathbf{CC}^\top \rangle - 2\langle \mathbf{P}, \mathbf{C} \rangle, \quad \mathbf{Q} = \mathbf{B}^\top \mathbf{B}, \mathbf{P} = \mathbf{B}^\top \mathbf{A}, \end{aligned}$$

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