Lecture on ALS. continued.

$$T \rightsquigarrow \boxed{\sum_{r=1}^{R} \underline{a}_r \otimes \underline{b}_r \otimes \underline{c}_r \equiv S}.$$

we should first see how to write down this sum in terms of three matricizations.

- $$S_{(1)} = A\,(C \otimes_{Khr} B)^T \qquad I_1 \times I_2 I_3$$

$$S_{(2)} = B\,(A \otimes_{Khr} C)^T \qquad I_2 \times I_1 I_3$$

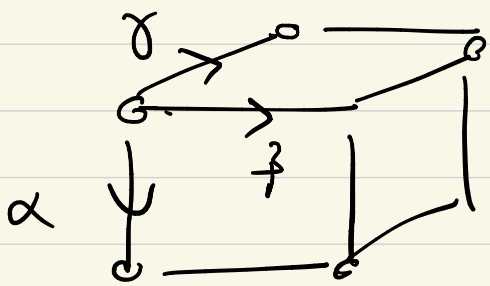
$$S_{(3)} = C\,(B \otimes_{Khr} A)^T \qquad I_3 \times I_2 I_1.$$

$$A = [\underline{a}_1, \dots, \underline{a}_R]$$

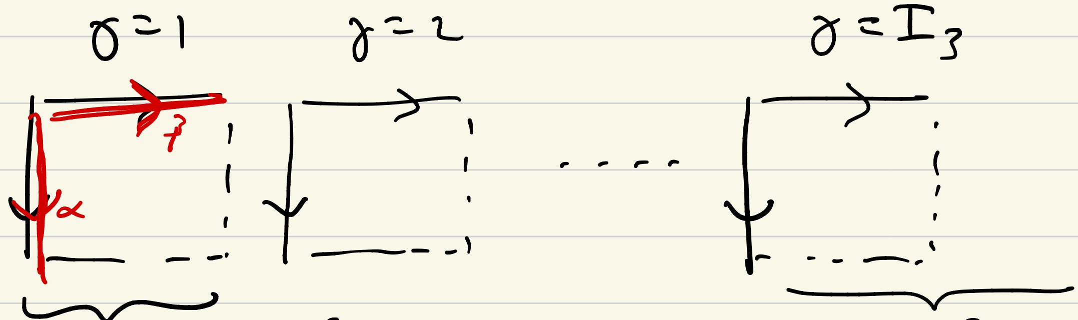
$$B = [\underline{b}_1, \dots, \underline{b}_R]$$

$$C = [\underline{c}_1, \dots, \underline{c}_R].$$

$$S = \sum_{r=1}^{R} \underline{a}_r \otimes \underline{b}_r \otimes \underline{c}_r \quad \text{cube with components} \quad \sum_{r=1}^{R} a_r^\alpha b_r^\beta c_r^\gamma$$



$S_{(1)} =$



$$= \left[\sum_{r=1}^{R} a_r^\alpha b_r^\beta c_r^1 \;\middle|\; \sum_{r=1}^{R} a_r^\alpha b_r^\beta c_r^2 \;\middle|\; \cdots \;\middle|\; \sum_{r=1}^{R} a_r^\alpha b_r^\beta c_r^{I_3} \right]$$

$$= \begin{array}{c} \\ \alpha \end{array}\!\!\left[a_1^\alpha \cdots a_R^\alpha \right] \left[\begin{matrix} b_1^\beta c_1^1 \\ \vdots \\ b_R^\beta c_R^1 \end{matrix} \;\middle|\; \begin{matrix} b_1^\beta c_1^2 \\ \vdots \\ b_R^\beta c_R^2 \end{matrix} \;\middle|\; \cdots \;\middle|\; \begin{matrix} b_1^\beta c_1^{I_3} \\ \vdots \\ b_R^\beta c_R^{I_3} \end{matrix} \right]$$

$$= A \left[\begin{matrix} b_1^T c_1^1 \\ \vdots \\ b_R^T c_R^1 \end{matrix} \;\middle|\; \cdots \;\middle|\; \begin{matrix} b_1^T c_1^{I_3} \\ \vdots \\ b_R^T c_R^{I_3} \end{matrix} \right] \doteq A \left[\begin{matrix} \underline{c}_1^T \otimes_{Kro} \underline{b}_1^T \\ \vdots \\ \underline{c}_R^T \otimes_{Kr} \underline{b}_R^T \end{matrix} \right]$$

Go back to previous video and defs.

$$\doteq A \left(C \otimes_{KhR} B \right)^T$$

3) Formulate the ALS algorithm:

- Performance is not so well known.
- empirical method popular
- Many extensions exist.

- I take A, B, C and I construct

$$T = \sum_{r=1}^{R} \underline{a}_r \otimes \underline{b}_r \otimes \underline{c}_r \qquad (*)\ (*)$$

assume that A, B, C are all three full column rank.
By Jenrich's thm you know that $(*)$ is a unique decomposition (up to scalings & permutation) into a sum of rank-one term. And rank is R.

- I give you $T^{\alpha\beta\gamma}$ and you have to find A, B, C.

- An idea for an algorithm would be to minimize the following sum over a's, b's, c's

$$\sum_{\alpha,\beta,\gamma} \left(T^{\alpha\beta\gamma} - \sum_{r=1}^{R} a_r^\alpha b_r^\beta c_r^\gamma \right)^2 \leftarrow$$

Obviously the minimum is <u>zero</u>. and the sol is the correct one and also is unique. not obvious

- How do you perform this minimization? highly Non convex probl.

Definition Frobenius Norm.

$$\| T \|_F^2 \equiv \sum_{\alpha=1}^{I_1} \sum_{\beta=1}^{I_2} \sum_{\gamma=1}^{I_3} \left(T^{\alpha\beta\gamma} \right)^2$$

for real valued tensors.

Idea of ALS is to minimize Frobenius Norm:

$$\| \underset{\cdot}{T} - \sum_{r=1}^{R} \underline{a}_r \otimes \underline{b}_r \otimes \underline{c}_r \|_F^2 . \quad \leftarrow$$

over $\underline{a}_r, \underline{b}_r, \underline{c}_r \quad r = 1 \ldots R$.

Remark: Frob Norm is sum of the list of all components squared.

equal to $\| T_{(1)} - \underset{\approx}{A} (C \otimes_{khr} B)^T \|_F^2 \leftarrow$

equal to $\| T_{(2)} - B (A \otimes_{khr} C)^T \|_F^2 \leftarrow$

equal to $\| T_{(3)} - C (B \otimes_{khr} A)^T \|_F^2 \leftarrow$

, Idea is that if e.g. B & C where known then you could solve

$$\underset{A}{\operatorname{argmin}} \| T_{(1)} - A (B \otimes_{khr} C)^T \|_F^2$$

(by the least square method where A is unknown.).

• **ALS main steps are**

$$A^{m+1} \leftarrow \underset{A}{\operatorname{argmin}} \| T_{(1)} - A (B^{m} \otimes_{khr} C^{m})^T \|_F^2 \checkmark$$

$$B^{m+1} \leftarrow \underset{B}{\operatorname{argmin}} \| T_{(2)} - B (A^{m+1} \otimes_{khr} C^{m})^T \|_F^2$$

$$C^{m+1} \leftarrow \underset{C}{\operatorname{argmin}} \| T_{(3)} - C (B^{m+1} \otimes_{khr} A^{m+1})^T \|_F^2$$

with initial condition $(B^{(0)}, C^{(0)})$ say.

Using least square solution you find for the algo

$$
\begin{cases}
\checkmark A^{m+1} = T_{(1)} \left(C^{(m)} \otimes_{Khr} B^{(m)} \right) \left(C^{mT} C^{m} * B^{mT} B \right)^{-1} \\
\checkmark B^{m+1} = T_{(2)} \left(A^{m+1} \otimes_{Khr} B^{m} \right) \left(A^{m+1\,T} A^{m+1} * C^{mT} C^{m} \right)^{-1} \\
\checkmark C^{m+1} = T_{(3)} \left(B^{m+1} \otimes_{Khr} A^{m+1} \right) \left(B^{m+1\,T} B^{m+1} * A^{m+1} A^{m+1} \right)^{-1}
\end{cases}
$$

choosing initial cond $\left(C^{(0)}, B^{(0)} \right)$.

#.

Least square solution of :

$$\operatorname*{argmin}_{\hat{A}} \| T_{(1)} - \hat{A}\,\underbrace{(C \otimes_{khr} B)^{T}} \|_F^2$$

assuming C & B are known.

Use the usual least square soln :

$$\operatorname*{argmin}_{\hat{A}} \| \underbrace{Y}_{\text{matrix}} - \underbrace{\hat{A}}_{\text{matrix}} \underbrace{\Phi}_{\text{matrix.}} \|_F^2$$ with Φ full row rank

The unique solution for $\hat{A} = Y \underbrace{\Phi^T (\Phi \Phi^T)^{-1}}_{\Phi^{+} \text{ MP pseudoinverse}}$

read the $\boxed{\text{notes}}$ for the proof recap.

usualy $\vec{y}^{\,T}$ and $\vec{a}^{\,T}$ solve $\operatorname*{argmin}_{a} \| \vec{y}^{\,T} - \vec{a}^{\,T} \phi \|_2^2$

$= \operatorname*{argmin} \| \vec{y} - \phi^T \vec{a} \|_2^2$

with ϕ^T full column rank.

$\longrightarrow$

apply this result.

$$\hat{A} = T_{(1)} \left((C \otimes_{khr} B)^T \right)^+$$

remember that because C & B are full col rank

$C \otimes_{khr} B$ is full col rank and so

$(C \otimes_{khr} B)^T$ is full row rank. and

$$\left((C \otimes_{khr} B)^T \right)^+ = (C \otimes_{khr} B) \Big[\underbrace{(C \otimes_{khr} B)^T (C \otimes_{khr} B)}_{\text{from previous video property } (C^T C) * (B^T B)} \Big]^{-1}$$

here $C \otimes_{khr} B = \phi$

$\Rightarrow$

$$\hat{A} = T_{(1)} (C \otimes_{khr} B) \left((C^T C) * (B^T B) \right)^{-1}$$

$\blacksquare$