

CSE 392: Matrix and Tensor Algorithms for Data

Instructor: Shashanka Ubaru

University of Texas, Austin
Spring 2024

Lecture 16: Canonical Polyadic (CP) decomposition

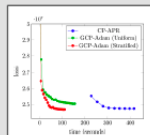
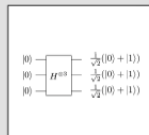
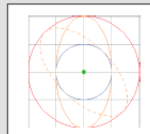
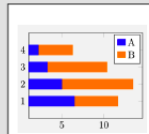
Outline

- 1 Introduction to CP
- 2 Khatri Rao Product (KRP)
- 3 CP-ALS

Books in preparation by Dr. Tamara Kolda



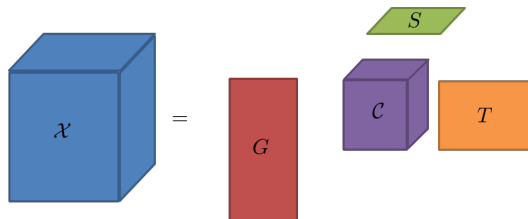
Learning L^AT_EX Graphics: TikZ/PGF and PGFPLOTS



Tamara G. Kolda

Tensor Decomposition

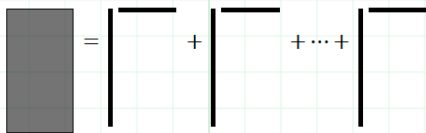
- Datasets (tensors) can be typically large in size.
- **Tensor Decomposition:**
 - ▶ Compression and storage.
 - ▶ Denoising.
 - ▶ *Hidden* multi-dimensional correlations and patterns.
- Different types of tensor decompositions.
 - ▶ Pros and cons.
 - ▶ *Application dependent.*



Tensor Factorization

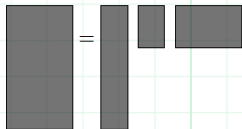
Two views of Matrix Factorization

1) Sum of rank-1 matrices



Ex: SVD, EVD, PCA, Sparse SVD, NMF

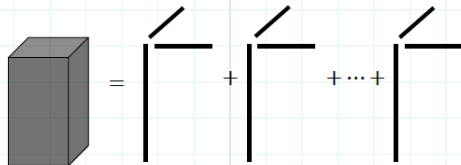
2) Compression via subspaces



Ex: SVD, CUR

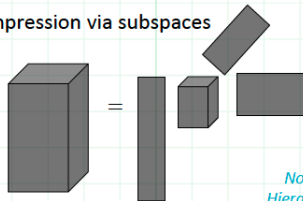
Two views of Tensor Factorization

1) Sum of rank-1 tensors



Ex: CANDECOMP/PARAFAC (CP)

2) Compression via subspaces

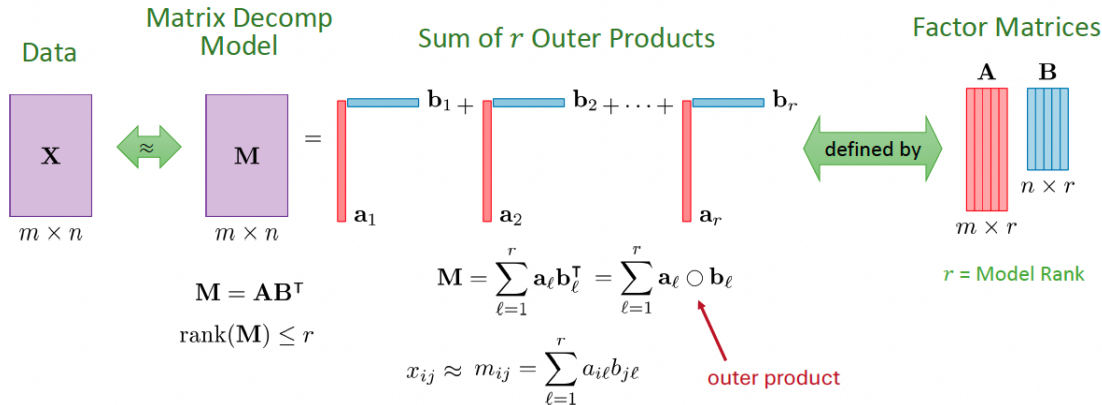


Ex: HOSVD, Tucker

*Not discussed:
Hierarchical Tucker,
Tensor Train,
Tensor Ring Decompositions*

Matrix Factorization

Examples include singular value decomposition, nonnegative matrix factorization, etc.



Step Back to the Matrix SVD

Traditional workhorse, dim reduction/feature extraction: matrix SVD

- PCA - directions of most variability; projections in ‘dominant’ directions allows for dim reduction/relative comparison
- Compression (reduce near redundancies) via truncated SVD expansion is **optimal** (Eckart-Young Theorem)

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^\top = \sum_{i=1}^r \sigma_i(\mathbf{u}_i \circ \mathbf{v}_i), \sigma_1 \geq \sigma_2 \geq \dots \geq 0$$

$$\mathbf{B} = \sum_{i=1}^k \sigma_i(\mathbf{u}_i \circ \mathbf{v}_i) \quad \text{solves}$$

$$\min \|\mathbf{A} - \mathbf{B}\|_F \quad \text{s.t. } \mathbf{B} \text{ has rank } k \leq r$$

Step Back to the Matrix SVD

Traditional workhorse, dim reduction/feature extraction: matrix SVD

- PCA - directions of most variability; projections in ‘dominant’ directions allows for dim reduction/relative comparison
- Compression (reduce near redundancies) via truncated SVD expansion is **optimal** (Eckart-Young Theorem)

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^\top = \sum_{i=1}^r \sigma_i(\mathbf{u}_i \circ \mathbf{v}_i), \sigma_1 \geq \sigma_2 \geq \dots \geq 0$$

$$\mathbf{B} = \sum_{i=1}^k \sigma_i(\mathbf{u}_i \circ \mathbf{v}_i) \quad \text{solves}$$

$$\min \|\mathbf{A} - \mathbf{B}\|_F \quad \text{s.t. } \mathbf{B} \text{ has rank } k \leq r$$

Implicit storage: for an $m \times n$, $k(n + m)$ numbers stored, vs mn .

Step Back to the Matrix SVD

Traditional workhorse, dim reduction/feature extraction: matrix SVD

- PCA - directions of most variability; projections in ‘dominant’ directions allows for dim reduction/relative comparison
- Compression (reduce near redundancies) via truncated SVD expansion is **optimal** (Eckart-Young Theorem)

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^\top = \sum_{i=1}^r \sigma_i(\mathbf{u}_i \circ \mathbf{v}_i), \sigma_1 \geq \sigma_2 \geq \dots \geq 0$$

$$\mathbf{B} = \sum_{i=1}^k \sigma_i(\mathbf{u}_i \circ \mathbf{v}_i) \quad \text{solves}$$

$$\min \|\mathbf{A} - \mathbf{B}\|_F \quad \text{s.t. } \mathbf{B} \text{ has rank } k \leq r$$

Question: What’s the right high-dimensional analogue? (history, see Kolda & Bader)

Idea 1 (Hitchcock, 1927): Like SVD, try to decompose as a sum of **rank-1** tensors.

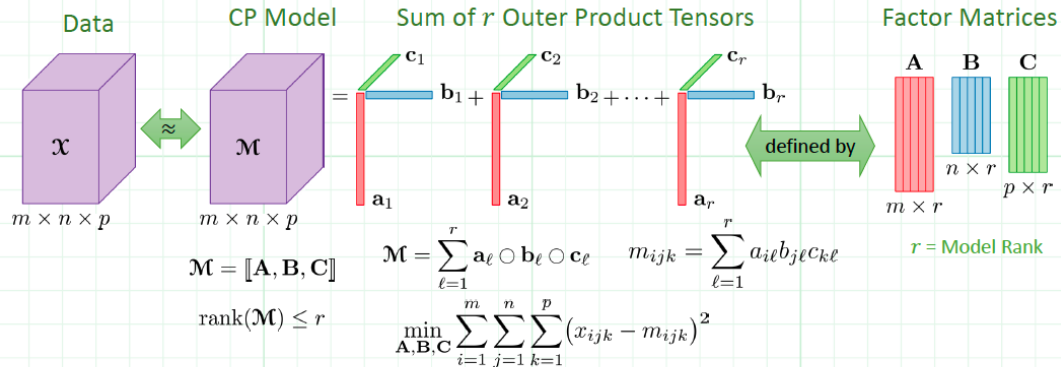
$$\mathcal{X} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \Rightarrow \mathcal{X}_{i,j,k} = a_i b_j c_k$$

Note that $\text{vec}(\mathcal{X}) = \mathbf{c} \otimes \mathbf{b} \otimes \mathbf{a}$.

Thus, some papers use Kronecker in place of outer-product notation.

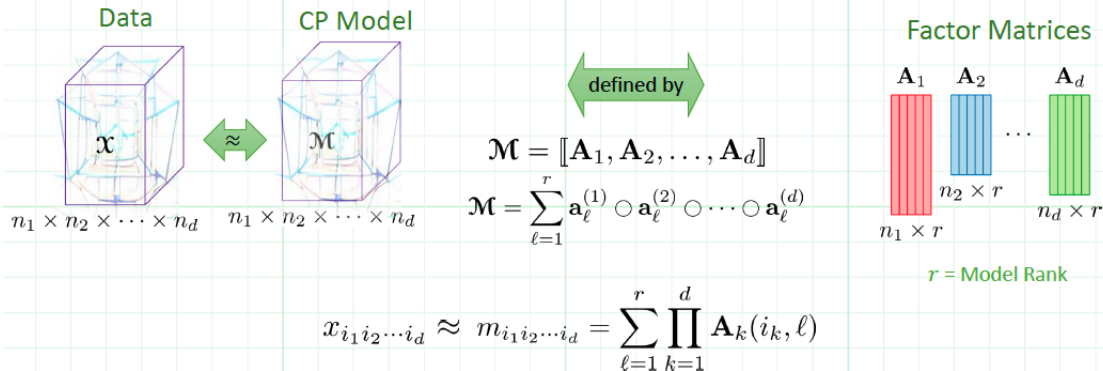
Canonical Polyadic (CP) Tensor Decomposition

Also known as Parallel Factors (PARAFAC) or Canonical Decomposition (CANDECOMP).



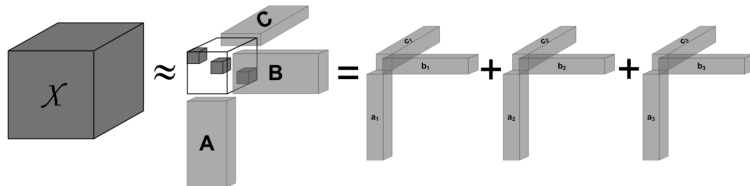
Hitchcock (1927), Carroll & Chang (1970), Harshman (1970)

CP Tensor Decomposition (d way)



Tensor image source: DeepAI txt2img

Kruskal Notation

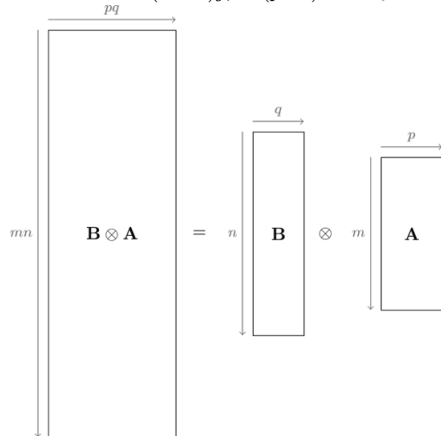


$$\mathcal{X} \approx \mathcal{M} = \sum_{\ell=1}^r \mathbf{a}_{\ell} \circ \mathbf{b}_{\ell} \circ \mathbf{c}_{\ell}$$

Kruskal notation: $[[\mathbf{A}, \mathbf{B}, \mathbf{C}]]$ or, if unit-normalized $[[\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]]$.

Recall : Matrix Kronecker Product

$$(\mathbf{B} \otimes \mathbf{A})_{i+(m-1)j, k+(p-1)\ell} = b_{j\ell} a_{ik}$$

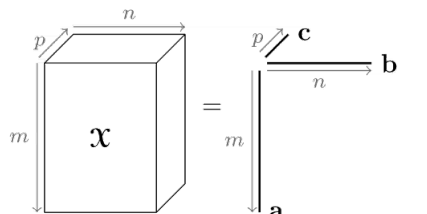


Key properties

- $(\mathbf{C} \otimes \mathbf{B}) \otimes \mathbf{A} = \mathbf{C} \otimes (\mathbf{B} \otimes \mathbf{A})$
- $(\mathbf{B} \otimes \mathbf{A})^\top = \mathbf{B}^\top \otimes \mathbf{A}^\top$
- $(\mathbf{B} \otimes \mathbf{A})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{BD}) \otimes (\mathbf{AC})$
- $\text{vec}(\mathbf{AXB}^\top) = (\mathbf{B} \otimes \mathbf{A})\text{vec}(\mathbf{X})$
- $(\mathbf{B} \otimes \mathbf{A})^{-1} = \mathbf{B}^{-1} \otimes \mathbf{A}^{-1}$

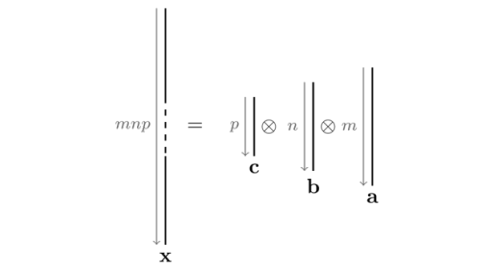
Vector Outer & Kronecker Products

A **vector outer product** for vectors $\mathbf{a} \in \mathbb{R}^m$, $\mathbf{b} \in \mathbb{R}^n$, $\mathbf{c} \in \mathbb{R}^p$ is denoted $\mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$ produces an $m \times n \times p$ tensor such that element (i, j, k) equals $a_i b_j c_k$, i.e.,



$$\mathbf{X} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \quad \Leftrightarrow \quad x_{ijk} = a_i b_j c_k$$

A **vector Kronecker product** for vectors $\mathbf{a} \in \mathbb{R}^m$, $\mathbf{b} \in \mathbb{R}^n$, $\mathbf{c} \in \mathbb{R}^p$ is denoted $\mathbf{c} \otimes \mathbf{b} \otimes \mathbf{a}$ produces a vector or length mnp such that element $\ell = \mathbb{L}(i, j, k)$ equals $a_i b_j c_k$, i.e.,

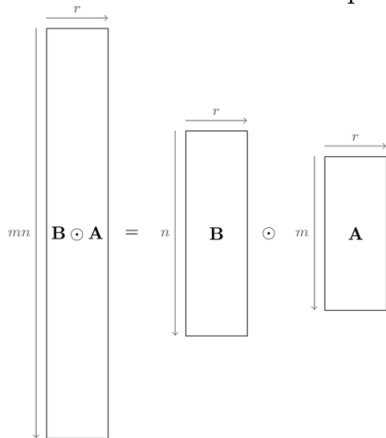


$$\mathbf{x} = \mathbf{c} \otimes \mathbf{b} \otimes \mathbf{a} \quad \Leftrightarrow \quad x_{\mathbb{L}(i,j,k)} = a_i b_j c_k$$

$$\mathbb{L}(i, j, k) = i + (j - 1)m + (k - 1)mn$$

Matrix KhatriRao Product (KRP)

KRP = columnwise Kronecker product

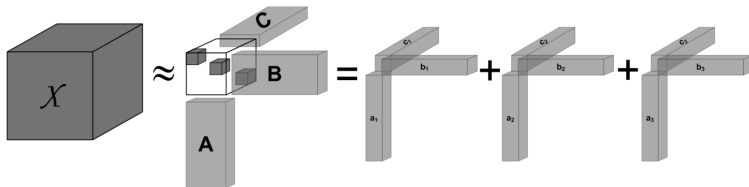


$$(B \odot A)_{*j} = B_{*j} \otimes A_{*j}$$

Key properties

- $C \odot (B \odot A) = (C \odot B) \odot A$
- $(B \odot A)^\top (B \odot A) = B^\top B * A^\top A$
- $(B \otimes A)(D \odot C) = (BD) \odot (AC)$

Kruskal Tensor



$$\mathcal{X} \approx \mathcal{M} = [\mathbf{A}, \mathbf{B}, \mathbf{C}] = \sum_{\ell=1}^r \mathbf{a}_{\ell} \circ \mathbf{b}_{\ell} \circ \mathbf{c}_{\ell}$$

Sum of r Outer Product Tensors.

$$\begin{aligned} \mathbf{M}_{(1)} &= \sum_{\ell=1}^r \mathbf{a}_{\ell} (\mathbf{c}_{\ell} \otimes \mathbf{b}_{\ell})^{\top} \\ &= [\mathbf{a}_1, \dots, \mathbf{a}_r] [\mathbf{c}_1 \otimes \mathbf{b}_1, \dots, \mathbf{c}_r \otimes \mathbf{b}_r]^{\top} \\ &= \mathbf{A} (\mathbf{C} \odot \mathbf{B})^{\top} \end{aligned}$$

Matricized Tensor Times KRP (MTTKRP)

Three way tensor:

$$\mathcal{X} \in \mathbb{R}^{m \times n \times p}, \mathbf{A} \in \mathbb{R}^{m \times r}, \mathbf{B} \in \mathbb{R}^{n \times r}, \mathbf{C} \in \mathbb{R}^{p \times r}$$

Then, we can define

$$\mathbf{X}_{(1)}(\mathbf{C} \odot \mathbf{B}), \quad \mathbf{X}_{(2)}(\mathbf{C} \odot \mathbf{A}), \quad \mathbf{X}_{(3)}(\mathbf{B} \odot \mathbf{A})$$

Matricized Tensor Times KRP (MTTKRP)

Three way tensor:

$$\mathcal{X} \in \mathbb{R}^{m \times n \times p}, \mathbf{A} \in \mathbb{R}^{m \times r}, \mathbf{B} \in \mathbb{R}^{n \times r}, \mathbf{C} \in \mathbb{R}^{p \times r}$$

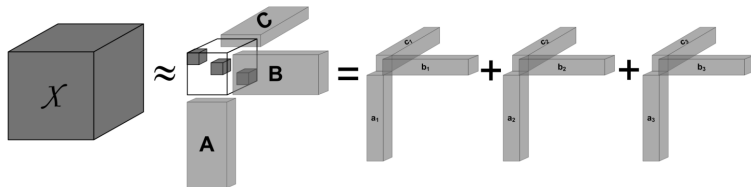
Then, we can define

$$\mathbf{X}_{(1)}(\mathbf{C} \odot \mathbf{B}), \quad \mathbf{X}_{(2)}(\mathbf{C} \odot \mathbf{A}), \quad \mathbf{X}_{(3)}(\mathbf{B} \odot \mathbf{A})$$

For general d -way tensor, say $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$ and $\mathbf{A}_k \in \mathbb{R}^{n_k \times r}$ for all $k \in [d]$, the mode- k matricized tensor times KRP (MTTKRP) is

$$\mathbf{V} = \mathbf{X}_{(k)}(\mathbf{A}_d \odot \dots \odot \mathbf{A}_{k+1} \odot \mathbf{A}_{k-1} \odot \dots \odot \mathbf{A}_1) \in \mathbb{R}^{n_k \times r}$$

CP Tensor Decomposition

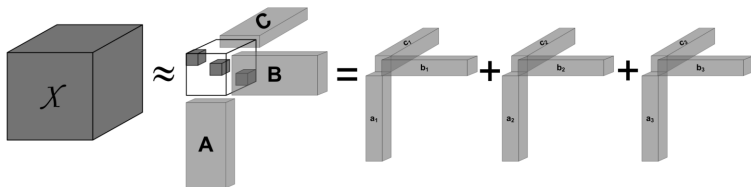


$$\mathcal{X} \approx \mathcal{M} = [\mathbf{A}, \mathbf{B}, \mathbf{C}] = \sum_{\ell=1}^r \mathbf{a}_\ell \circ \mathbf{b}_\ell \circ \mathbf{c}_\ell$$

- Find the best **tensor rank- r** fit:

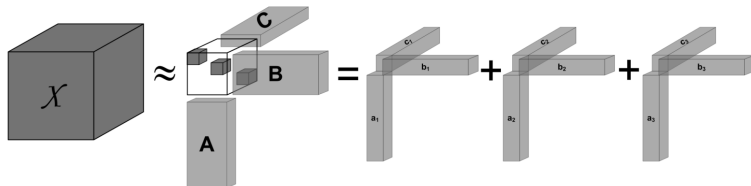
$$\min_{\mathbf{a}_\ell, \mathbf{b}_\ell, \mathbf{c}_\ell} \left\| \mathcal{X} - \sum_{\ell=1}^r \sigma_\ell \cdot \mathbf{a}_\ell \circ \mathbf{b}_\ell \circ \mathbf{c}_\ell \right\|_F$$

CP Properties



$$\mathcal{X} \approx \sum_{\ell=1}^r \mathbf{a}_\ell \circ \mathbf{b}_\ell \circ \mathbf{c}_\ell$$

- If equality & r **minimal**, then r is called the **rank** of the tensor
- Not generally orthogonal
- Not based on a ‘product based factorization’
- Finding the rank is NP hard!
- No perfect procedure for fitting CP model to k terms

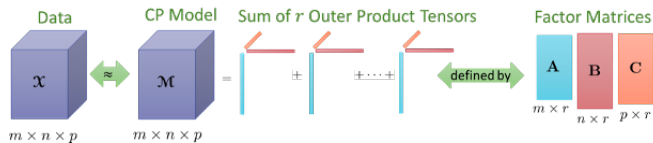


- $\mathcal{M} = \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket$ is essentially unique if

$$\text{rank}_k(\mathbf{A}) + \text{rank}_k(\mathbf{B}) + \text{rank}_k(\mathbf{C}) \geq 2r + 2$$

- $\text{rank}_k(\mathbf{A})$ = maximum value of k such that any k columns of \mathbf{A} are linearly independent.
- Matrix factorization does not share this property! Usually need orthogonality constraint.

Alternating Least Squares (CP-ALS)



$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \|\mathcal{X} - \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket\|_F$$

General Idea: solve for ONE matrix, holding the others fixed.

- **CP-ALS:** Repeat until converged...

- ▶ Solve for \mathbf{A} (with \mathbf{B} and \mathbf{C} fixed)
- ▶ Solve for \mathbf{B} (with \mathbf{A} and \mathbf{C} fixed)
- ▶ Solve for \mathbf{C} (with \mathbf{A} and \mathbf{B} fixed)

Special Structure of Least Squares Problem

$$\min_{\mathbf{A}} \|\mathbf{X}_{(1)} - \mathbf{A}(\mathbf{C} \odot \mathbf{B})^\top\|_F^2$$

Special Structure of Least Squares Problem

$$\min_{\mathbf{A}} \|\mathbf{X}_{(1)} - \mathbf{A}(\mathbf{C} \odot \mathbf{B})^\top\|_F^2$$

$$\min_{\mathbf{A}} \|(\mathbf{C} \odot \mathbf{B})\mathbf{A}^\top - \mathbf{X}_{(1)}^\top\|_F^2$$

Special Structure of Least Squares Problem

$$\min_{\mathbf{A}} \|\mathbf{X}_{(1)} - \mathbf{A}(\mathbf{C} \odot \mathbf{B})^\top\|_F^2$$

$$\min_{\mathbf{A}} \|(\mathbf{C} \odot \mathbf{B})\mathbf{A}^\top - \mathbf{X}_{(1)}^\top\|_F^2$$

By normal equations:

$$(\mathbf{C} \odot \mathbf{B})^\top (\mathbf{C} \odot \mathbf{B})\mathbf{A}^\top = (\mathbf{C} \odot \mathbf{B})^\top \mathbf{X}_{(1)}^\top$$

$$(\mathbf{C}^\top \mathbf{C} * \mathbf{B}^\top \mathbf{B})\mathbf{A}^\top = (\mathbf{C} \odot \mathbf{B})^\top \mathbf{X}_{(1)}^\top$$

Special Structure of Least Squares Problem

$$\min_{\mathbf{A}} \|\mathbf{X}_{(1)} - \mathbf{A}(\mathbf{C} \odot \mathbf{B})^\top\|_F^2$$

$$\min_{\mathbf{A}} \|(\mathbf{C} \odot \mathbf{B})\mathbf{A}^\top - \mathbf{X}_{(1)}^\top\|_F^2$$

By normal equations:

$$(\mathbf{C} \odot \mathbf{B})^\top (\mathbf{C} \odot \mathbf{B})\mathbf{A}^\top = (\mathbf{C} \odot \mathbf{B})^\top \mathbf{X}_{(1)}^\top$$

$$(\mathbf{C}^\top \mathbf{C} * \mathbf{B}^\top \mathbf{B})\mathbf{A}^\top = (\mathbf{C} \odot \mathbf{B})^\top \mathbf{X}_{(1)}^\top$$

$$\mathbf{A}^\top = (\mathbf{C}^\top \mathbf{C} * \mathbf{B}^\top \mathbf{B})^{-1} (\mathbf{C} \odot \mathbf{B})^\top \mathbf{X}_{(1)}^\top$$

$$\mathbf{A} = \mathbf{X}_{(1)} (\mathbf{C} \odot \mathbf{B}) (\mathbf{C}^\top \mathbf{C} * \mathbf{B}^\top \mathbf{B})^{-1}$$

Special Structure of Least Squares Problem (d -way)

$$\min_{\mathbf{A}_k} \|\mathbf{X}_{(k)} - \mathbf{A}_k \underbrace{(\mathbf{A}_d \odot \cdots \odot \mathbf{A}_{k+1} \odot \mathbf{A}_{k-1} \odot \cdots \odot \mathbf{A}_1)}_{\mathbf{Z}_k}\|_F^2$$

$$\min_{\mathbf{A}_k} \|\mathbf{Z}_k \mathbf{A}_k^\top - \mathbf{X}_{(k)}^\top\|_F^2$$

$$\mathbf{Z}_k^\top \mathbf{Z}_k \mathbf{A}_k^\top = \mathbf{Z}_k^\top \mathbf{X}_{(k)}^\top$$

$$\underbrace{(\mathbf{A}_d^\top \mathbf{A}_d * \cdots * \mathbf{A}_{k+1}^\top \mathbf{A}_{k+1} * \mathbf{A}_{k-1}^\top \mathbf{A}_{k-1} \cdots * \mathbf{A}_1^\top \mathbf{A}_1)}_{\mathbf{V}_k} \mathbf{A}_k^\top = \mathbf{Z}_k^\top \mathbf{X}_{(k)}^\top$$

$$\mathbf{A}_k = \mathbf{X}_{(k)} \mathbf{Z}_k \mathbf{V}_k^{-1}$$

CP-ALS Full Algorithm

Inputs: Tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$, desired rank $r \in \mathbb{N}$.

- ➊ Initialize $\mathbf{A}_k \in \mathbb{R}^{n_k \times r}$ for all $k \in [d]$
- ➋ repeat
- ➌ for $k = 1, \dots, d$ do
- ➍ $\mathbf{Z}_k \leftarrow \mathbf{A}_d \odot \dots \odot \mathbf{A}_{k+1} \odot \mathbf{A}_{k-1} \odot \dots \odot \mathbf{A}_1$
- ➎ $\mathbf{A}_k \leftarrow \arg \min_{\mathbf{B}} \|\mathbf{Z}_k \mathbf{B}^\top - \mathbf{X}_{(k)}^\top\|_F^2$
- ➏ end
- ➐ until $\|\mathcal{X} - \llbracket \mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_d \rrbracket\|_F^2$ ceases to decrease

Matlab Demo