CSE 392: Matrix and Tensor Algorithms for Data

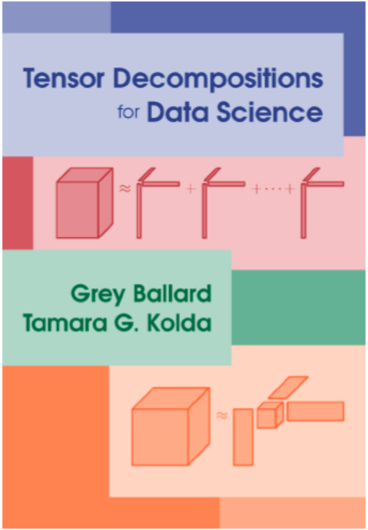
Instructor: Shashanka Ubaru

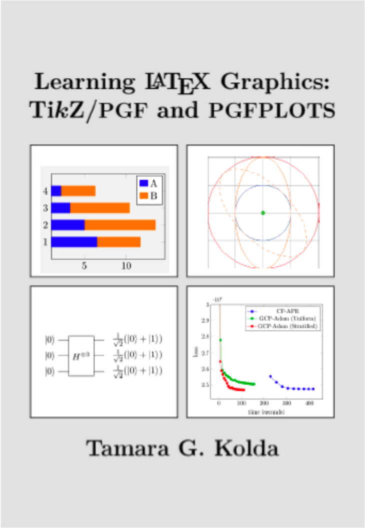
University of Texas, Austin
Spring 2024

Lecture 16: Canonical Polyadic (CP) decomposition

Outline

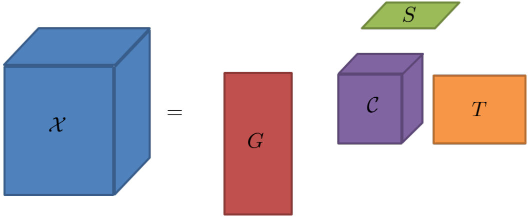
Books in preparation by Dr. Tamara Kolda



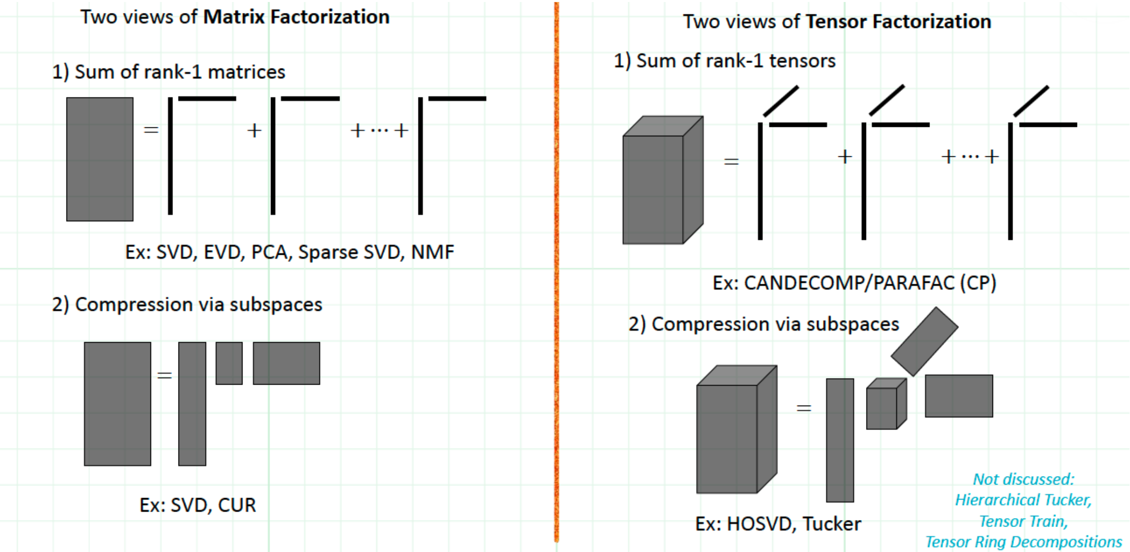


Tensor Decomposition

- 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.*

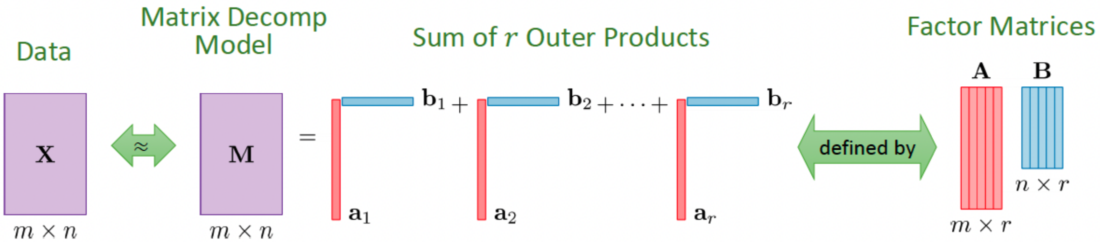


Tensor Factorization



Matrix Factorization

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



$$\mathbf{M} = \mathbf{A}\mathbf{B}^{\intercal}$$

$$\operatorname{rank}(\mathbf{M}) \leq r$$

$$\mathbf{M} = \sum_{\ell=1}^{r} \mathbf{a}_\ell \mathbf{b}_\ell^{\intercal} = \sum_{\ell=1}^{r} \mathbf{a}_\ell \circ \mathbf{b}_\ell$$

r = Model Rank

$$x_{ij} \approx m_{ij} = \sum_{\ell=1}^{r} a_{i\ell} b_{j\ell}$$

outer product

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)

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

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

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

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)

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

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

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

Implicit storage: for an $m \times n$, $p(n+m)$ 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)

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

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

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

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

Rank-1 Tensor

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

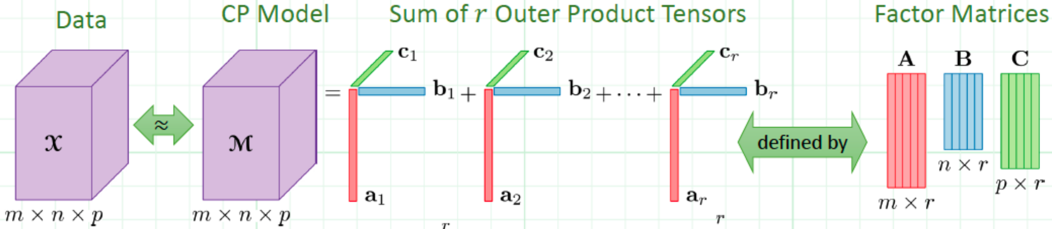
$$\mathcal{X} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \Rightarrow \mathcal{X}_{\ell,j,k} = a_\ell b_j c_k$$

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

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).



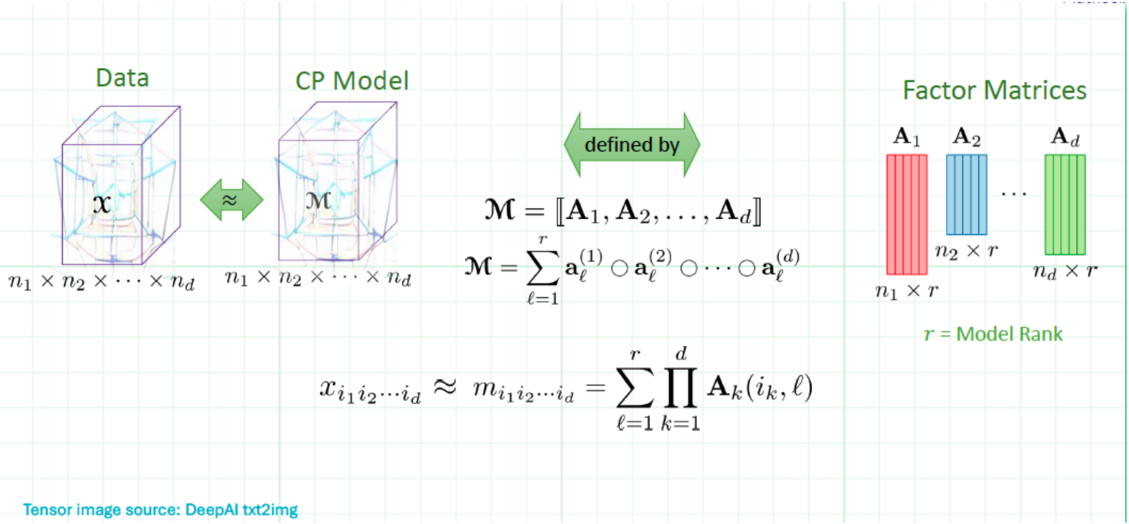
$$\mathcal{M} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \qquad \mathcal{M} = \sum_{\ell=1}^{r} \mathbf{a}_\ell \circ \mathbf{b}_\ell \circ \mathbf{c}_\ell \qquad m_{ijk} = \sum_{\ell=1}^{r} a_{i\ell} b_{j\ell} c_{k\ell}$$

r = Model Rank

$$\mathrm{rank}(\mathcal{M}) \leq r \qquad \min_{\mathbf{A},\mathbf{B},\mathbf{C}} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} \left(x_{ijk} - m_{ijk}\right)^2$$

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

CP Tensor Decomposition (d way)

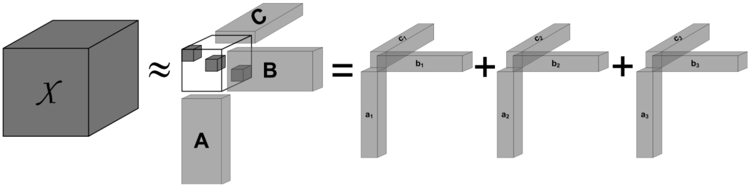


$$\mathcal{M} = [\![\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_d]\!]$$

$$\mathcal{M} = \sum_{\ell=1}^{r} \mathbf{a}_\ell^{(1)} \circ \mathbf{a}_\ell^{(2)} \circ \cdots \circ \mathbf{a}_\ell^{(d)}$$

$$x_{i_1 i_2 \cdots i_d} \approx m_{i_1 i_2 \cdots i_d} = \sum_{\ell=1}^{r} \prod_{k=1}^{d} \mathbf{A}_k(i_k, \ell)$$

Kruskal Notation

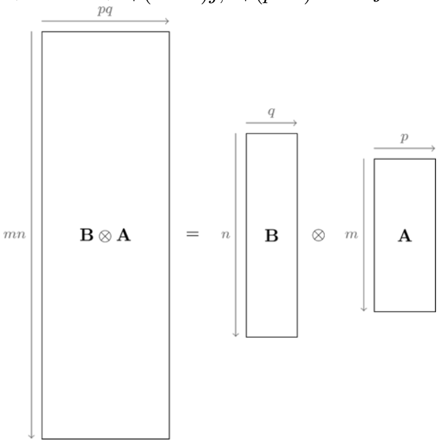


$$\mathcal{X} \approx \mathcal{M} = \sum_{\ell=1}^{r} \mathbf{a}_\ell \circ \mathbf{b}_\ell \circ \mathbf{c}_\ell$$

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

Recall : Matrix Kronecker Product

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

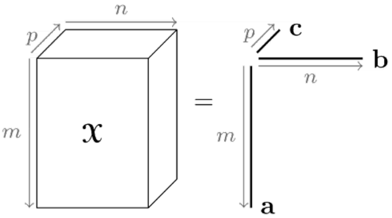


Key properties

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

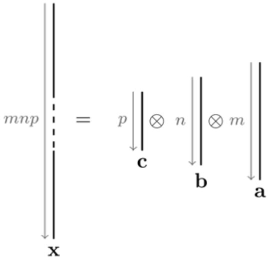
Vector Outer & Kronecker Products

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 $\mathbf{a} \bigcirc \mathbf{b} \bigcirc \mathbf{c}$ produces an $m \times n \times p$ tensor such that element (i, j, k) equals $a_i b_j c_k$, i.e.,



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

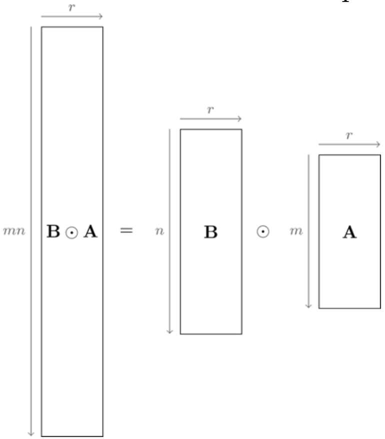
A **vector Kronecker product** for vectors $\mathbf{a} \in \mathbb{R}^m$, $\mathbf{b} \in \mathbb{R}^n$, $\mathbf{c} \in \mathbb{R}^p$ is denoted $\mathbf{c} \otimes \mathbf{b} \otimes \mathbf{a}$ produces a vector or length mnp such that element $\ell = \mathbb{L}(i, j, k)$ equals $a_i b_j c_k$, i.e.,



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

Matrix KhatriRao Product (KRP)

KRP = columnwise Kronecker product

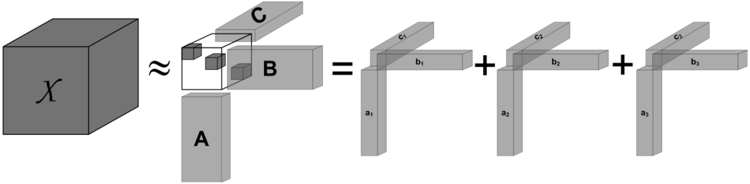


$$(\boldsymbol{B} \odot \boldsymbol{A})_{*j} = \boldsymbol{B}_{*j} \otimes \boldsymbol{A}_{*j}$$

Key properties

- $\boldsymbol{C} \odot (\boldsymbol{B} \odot \boldsymbol{A}) = (\boldsymbol{C} \odot \boldsymbol{B}) \odot \boldsymbol{A}$
- $(\boldsymbol{B} \odot \boldsymbol{A})^\top (\boldsymbol{B} \odot \boldsymbol{A}) = \boldsymbol{B}^\top \boldsymbol{B} * \boldsymbol{A}^\top \boldsymbol{A}$
- $(\boldsymbol{B} \otimes \boldsymbol{A})(\boldsymbol{D} \odot \boldsymbol{C}) = (\boldsymbol{B}\boldsymbol{D}) \odot (\boldsymbol{A}\boldsymbol{C})$

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.

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

Matricized Tensor Times KRP (MTTKRP)

Three way tensor:

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

Then, we can define

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

Matricized Tensor Times KRP (MTTKRP)

Three way tensor:

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

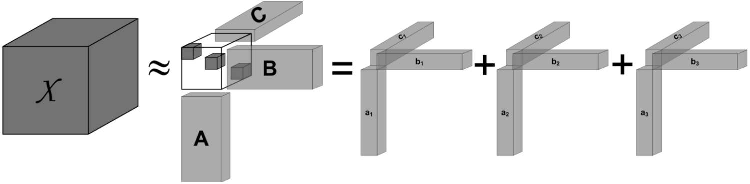
Then, we can define

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

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

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

CP Tensor Decomposition

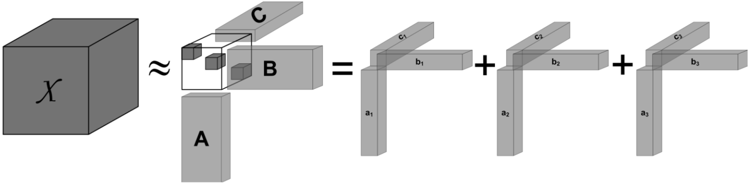


$$\mathcal{X} \approx \mathcal{M} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] = \sum_{\ell=1}^{r} \mathbf{a}_\ell \circ \mathbf{b}_\ell \circ \mathbf{c}_\ell$$

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

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

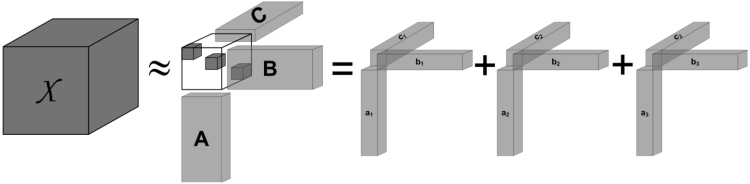
CP Properties



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

- 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

CP - Uniqueness

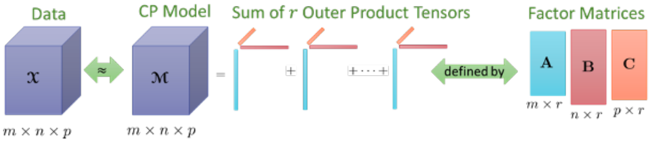


- $\mathcal{M} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$ is essentially unique if

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

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

Alternating Least Squares (CP-ALS)



$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \|\mathcal{X} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]\|_F$$

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

- **CP-ALS:** Repeat until converged…
  - Solve for $\boldsymbol{A}$ (with $\boldsymbol{B}$ and $\boldsymbol{C}$ fixed)
  - Solve for $\boldsymbol{B}$ (with $\boldsymbol{A}$ and $\boldsymbol{C}$ fixed)
  - Solve for $\boldsymbol{C}$ (with $\boldsymbol{A}$ and $\boldsymbol{B}$ fixed)

Special Structure of Least Squares Problem

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

Special Structure of Least Squares Problem

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

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

Special Structure of Least Squares Problem

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

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

By normal equations:

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

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

Special Structure of Least Squares Problem

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

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

By normal equations:

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

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

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

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

Special Structure of Least Squares Problem (d-way)

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

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

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

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

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

CP-ALS Full Algorithm

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

1. Initialize $\boldsymbol{A}_k \in \mathbb{R}^{n_k \times r}$ for all $k \in [d]$
2. **repeat**
3.   **for** $k = 1, \ldots, d$ **do**
4. $\boldsymbol{Z}_k \leftarrow \boldsymbol{A}_d \odot \cdots \odot \boldsymbol{A}_{k+1} \odot \boldsymbol{A}_{k-1} \odot \cdots \odot \boldsymbol{A}_1$
5. $\boldsymbol{A}_k \leftarrow \arg\min_{\boldsymbol{B}} \|\boldsymbol{Z}_k \boldsymbol{B}^\top - \boldsymbol{X}_{(k)}^\top\|_F^2$
6.   **end**
7. **until** $\|\mathcal{X} - [\![\mathbf{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_d]\!]\|_F^2$ ceases to decrease

Matlab Demo