

Tensor decomposition

Jean Kossaifi



@JeanKossaifi

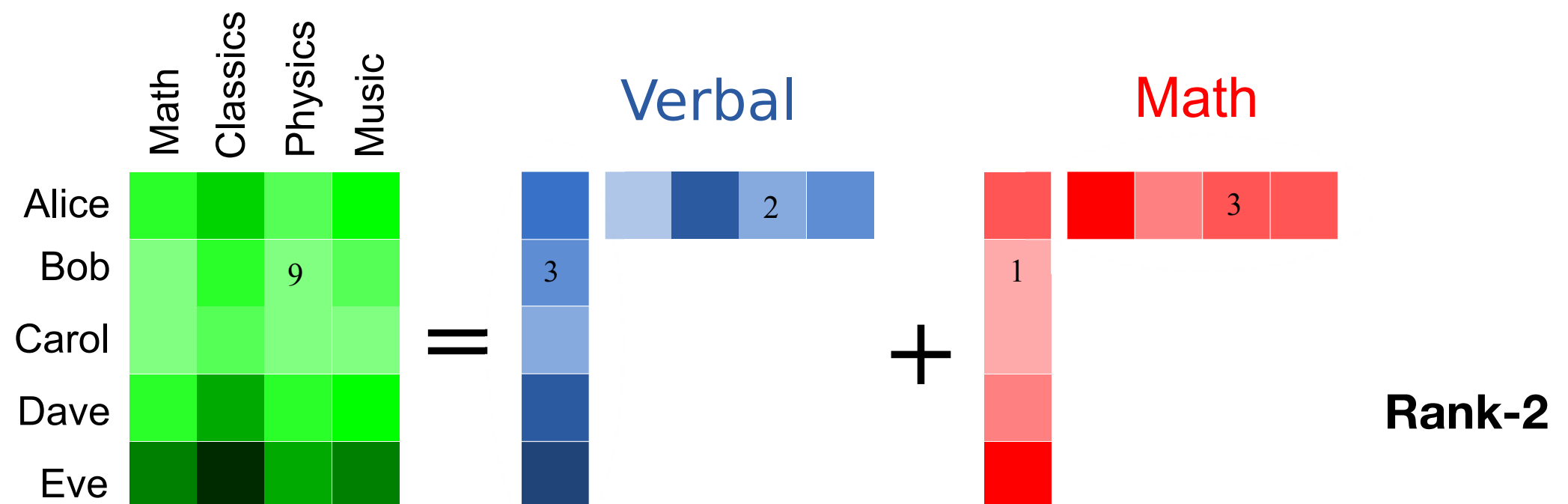
jean.kossaifi@gmail.com

Example: factor analysis

- Spearman worked on understanding whether intelligence is a composite of measurable intelligences
- Intelligence = quantitative and verbal ?
- Score matrix of n students \times m subjects

Example: factor analysis

- Spearman worked on understanding whether intelligence is a composite of measurable intelligences
- Intelligence = quantitative and verbal ?
- Score matrix of n students x m subjects



Matrix decomposition

$$\mathbf{X} \in \mathbb{R}^{m,n}$$

$$\mathbf{X} = \mathbf{AB}, \quad \mathbf{A} \in \mathbb{R}^{m,p+1}, \mathbf{B} \in \mathbb{R}^{p+1,n}, p+1 \leq \min(m, n)$$

$$\mathbf{AB} = \sum_{k=0}^p \mathbf{a}_{:,k} \mathbf{b}_{k,:}^\top = \sum_{k=0}^p \underbrace{\mathbf{a}_{:,k} \circ \mathbf{b}_{k,:}}_{\text{rank-1 components}}$$

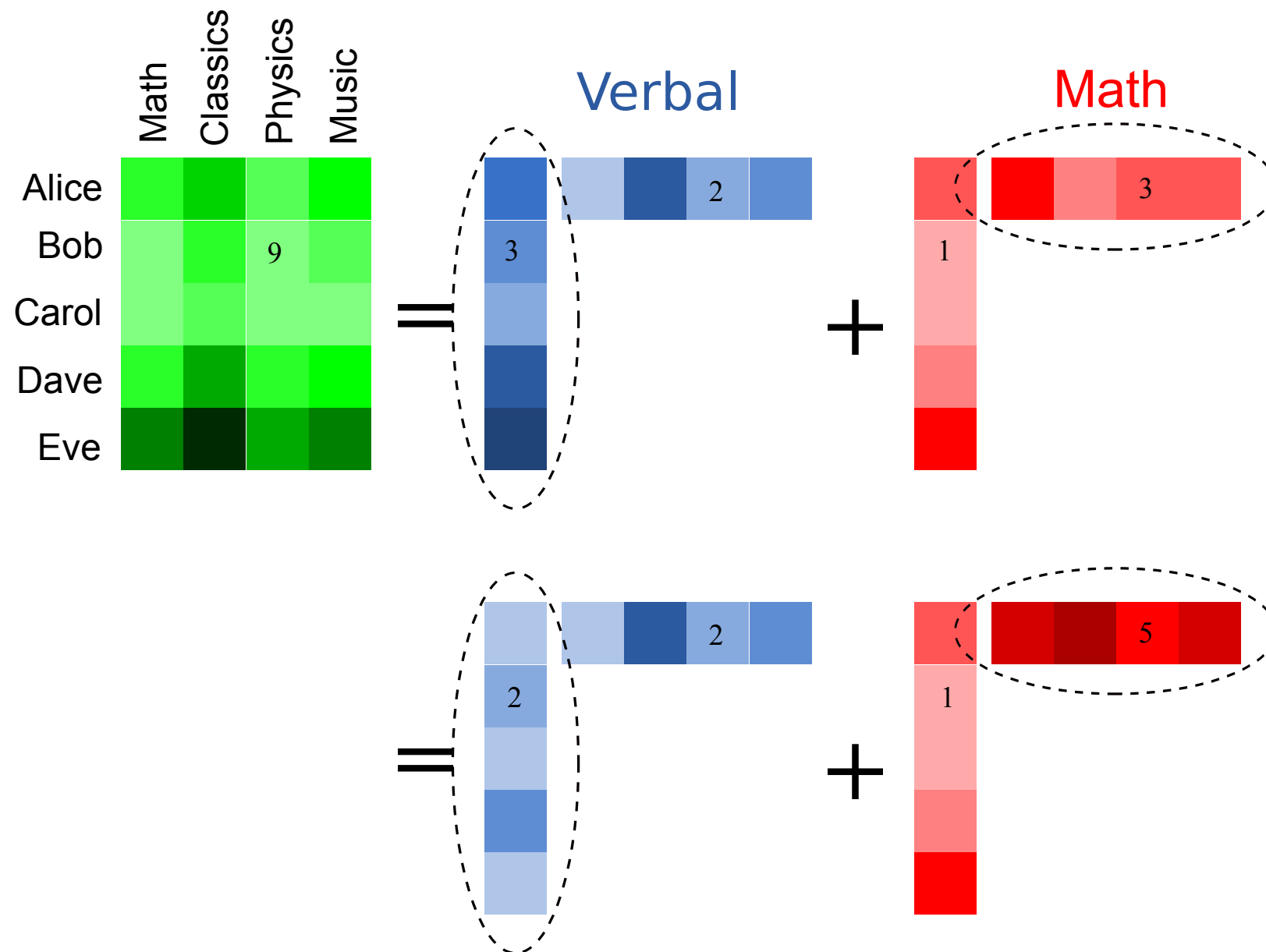
$$\mathbf{C} = \begin{pmatrix} a_{0,0} \\ a_{1,0} \\ \vdots \\ a_{m,0} \end{pmatrix} \begin{pmatrix} a_{0,0}b_{0,0} & a_{0,0}b_{0,1} & \cdots & a_{0,0}b_{0,n} \\ a_{1,0}b_{0,0} & a_{1,0}b_{0,1} & \cdots & a_{1,0}b_{0,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,0}b_{0,0} & a_{m,0}b_{0,1} & \cdots & a_{m,0}b_{0,n} \end{pmatrix} + \cdots + \begin{pmatrix} a_{0,p} \\ a_{1,p} \\ \vdots \\ a_{m,p} \end{pmatrix} \begin{pmatrix} a_{0,p}b_{p,0} & a_{0,p}b_{p,1} & \cdots & a_{0,p}b_{p,n} \\ a_{1,p}b_{p,0} & a_{1,p}b_{p,1} & \cdots & a_{1,p}b_{p,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,p}b_{p,0} & a_{m,p}b_{p,1} & \cdots & a_{m,p}b_{p,n} \end{pmatrix}$$

Diagram illustrating the rank-1 components of the matrix decomposition. A red circle with an 'x' is shown above the first term, with arrows pointing to the column vector $\begin{pmatrix} a_{0,0} \\ a_{1,0} \\ \vdots \\ a_{m,0} \end{pmatrix}$ and the row vector $(b_{0,0}, b_{0,1}, \dots, b_{0,n})$. Similarly, another red circle with an 'x' is shown above the last term, with arrows pointing to the column vector $\begin{pmatrix} a_{0,p} \\ a_{1,p} \\ \vdots \\ a_{m,p} \end{pmatrix}$ and the row vector $(b_{p,0}, b_{p,1}, \dots, b_{p,n})$.

Ambiguity of Matrix Decomposition

- Low rank decomposition of matrix: $\mathbf{M} = \sum_{k=0}^{R-1} \mathbf{a}_k \mathbf{b}_k^\top = \mathbf{A} \mathbf{B}^\top$
- One possible choice (SVD on M): $\mathbf{M} = \mathbf{U} \mathbf{S} \mathbf{V}^\top$
 $\mathbf{A} = \mathbf{U} \mathbf{S}$ $\mathbf{B} = \mathbf{V}^\top$

Ambiguity of Matrix Decomposition



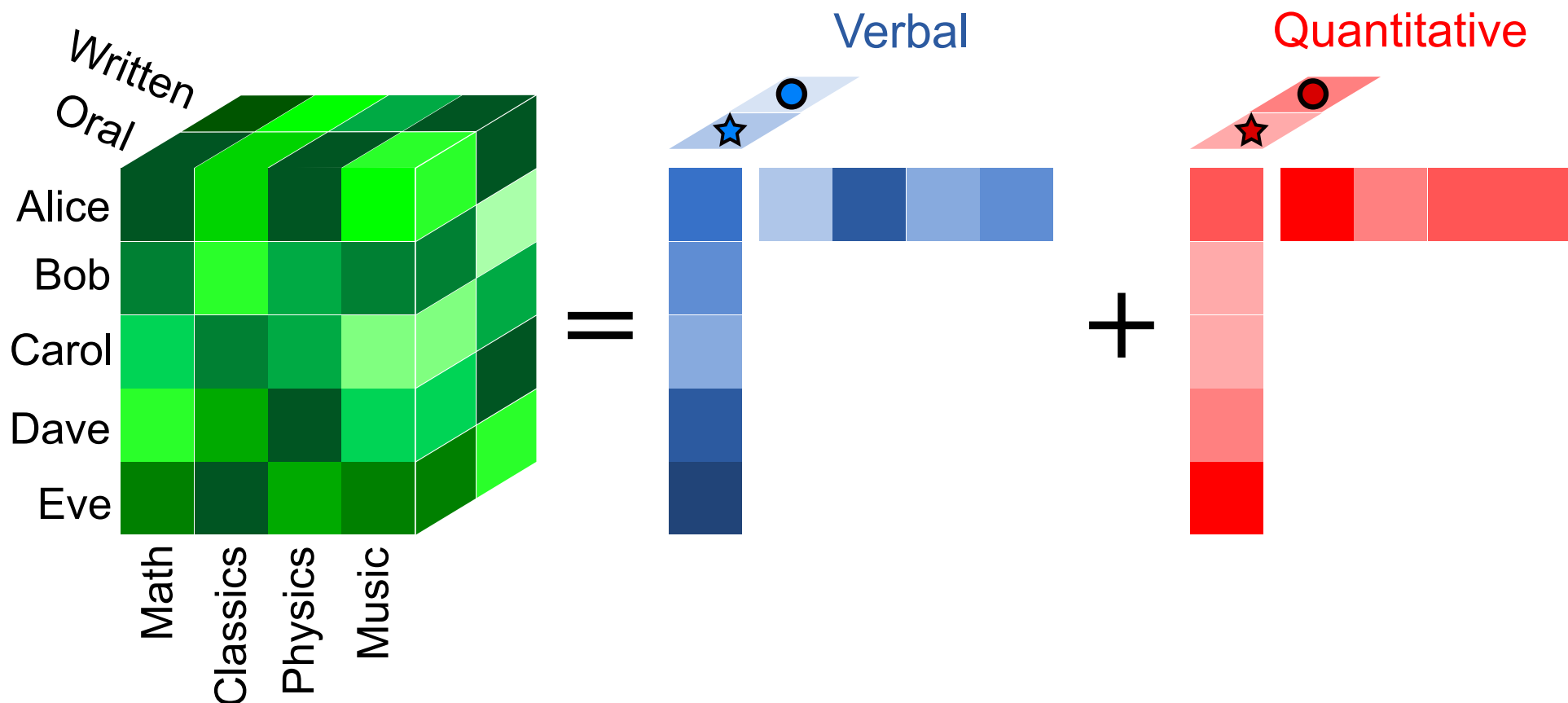
Ambiguity of Matrix Decomposition

- Low rank decomposition of matrix: $\mathbf{M} = \sum_{k=0}^{R-1} \mathbf{a}_k \mathbf{b}_k^\top = \mathbf{A} \mathbf{B}^\top$
- One possible choice (truncated SVD on \mathbf{M}): $\mathbf{M} = \mathbf{U} \mathbf{S} \mathbf{V}^\top$
 $\mathbf{A} = \mathbf{U}[:, : R] \mathbf{S}[:, R, : R] \quad \mathbf{B} = \mathbf{V}^\top[:, R, :]$
- For any orthonormal matrix $\mathbf{R} \in \mathbb{R}^{R \times R}$ such that $\mathbf{R} \mathbf{R}^\top = \mathbf{R}^\top \mathbf{R} = \mathbf{I}_R$

$$\mathbf{M} = \mathbf{A} \mathbf{R} \mathbf{R}^\top \mathbf{B}^\top = (\mathbf{A} \mathbf{R}) (\mathbf{R} \mathbf{B})^\top$$

From matrix to tensor

- Imagine exam having two parts: written and oral
-> 2 score matrices (frontal slices of a $n*m*2$ tensor)
- Verbal skills might matter more in oral exams



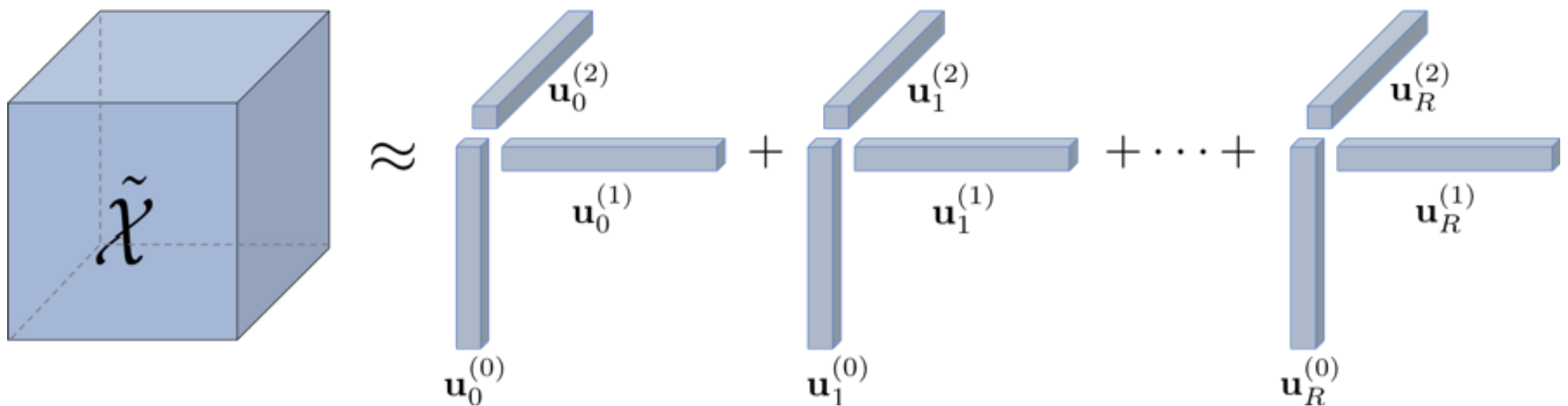
From matrix to tensor

- Matrix decomposition is not *unique*
- Tensor decomposition unique under mild conditions
 - ➡ Identifiability
- Almost all tensor problems are NP-hard...
- Good algorithms to solve them heuristically

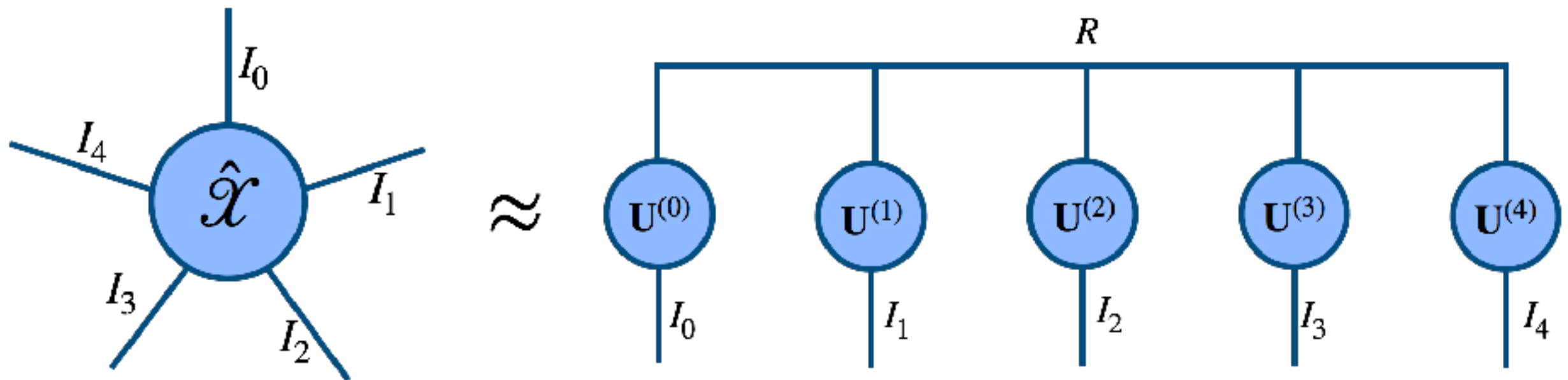
CP decomposition

$$\hat{\mathcal{X}} \in \mathbb{R}^{I,J,K}$$

$$\hat{\mathcal{X}} = \sum_{k=0}^{R-1} \underbrace{\mathbf{u}_k^{(0)} \circ \mathbf{u}_k^{(1)} \circ \mathbf{u}_k^{(2)}}_{\text{rank-1 components}}$$



CP decomposition



CP decomposition

$$\hat{\mathcal{X}} \in \mathbb{R}^{I,J,K}$$

$$\hat{\mathcal{X}} = \underbrace{\sum_{k=0}^{R-1} \lambda \underbrace{\mathbf{u}_k^{(0)} \circ \mathbf{u}_k^{(1)} \circ \mathbf{u}_k^{(2)}}_{\text{rank-1 components}}}_{\text{rank-R tensor}}$$

- Scaling ambiguity: equivalent to choosing units (e.g. for student, scale grade between 1 and 100)
- Permutation between rank-1 components gives same result

CP decomposition

$$\hat{\mathcal{X}} \in \mathbb{R}^{I,J,K}$$

$$\hat{\mathcal{X}} = \underbrace{\sum_{k=0}^{R-1} \underbrace{\mathbf{u}_k^{(0)} \circ \mathbf{u}_k^{(1)} \circ \mathbf{u}_k^{(2)}}_{\text{rank-1 components}}}_{\text{rank-R tensor}}$$

- Collect factors in matrices:

$$\mathbf{U}^{(0)} = [\mathbf{u}_0^{(0)}, \mathbf{u}_1^{(0)}, \dots, \mathbf{u}_{R-1}^{(0)}] \in \mathbb{R}^{I,R}$$

$$\mathbf{U}^{(1)} = [\mathbf{u}_0^{(1)}, \mathbf{u}_1^{(1)}, \dots, \mathbf{u}_{R-1}^{(1)}] \in \mathbb{R}^{J,R}$$

$$\mathbf{U}^{(2)} = [\mathbf{u}_0^{(2)}, \mathbf{u}_1^{(2)}, \dots, \mathbf{u}_{R-1}^{(2)}] \in \mathbb{R}^{K,R}$$

CP decomposition

$$\hat{\mathcal{X}} \in \mathbb{R}^{I,J,K}$$

$$\hat{\mathcal{X}} = \underbrace{\sum_{k=0}^{R-1} \underbrace{\mathbf{u}_k^{(0)} \circ \mathbf{u}_k^{(1)} \circ \mathbf{u}_k^{(2)}}_{\text{rank-1 components}}}_{\text{rank-R tensor}}$$

- Collect factors in matrices:

$$\mathbf{U}^{(0)} = [\mathbf{u}_0^{(0)}, \mathbf{u}_1^{(0)}, \dots, \mathbf{u}_{R-1}^{(0)}] \in \mathbb{R}^{I,R}$$

$$\mathbf{U}^{(1)} = [\mathbf{u}_0^{(1)}, \mathbf{u}_1^{(1)}, \dots, \mathbf{u}_{R-1}^{(1)}] \in \mathbb{R}^{J,R}$$

$$\mathbf{U}^{(2)} = [\mathbf{u}_0^{(2)}, \mathbf{u}_1^{(2)}, \dots, \mathbf{u}_{R-1}^{(2)}] \in \mathbb{R}^{K,R}$$

- Problem can be rewritten as:

$$\hat{\mathcal{X}} = [|\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}|]$$

Khatri-Rao product

- Khatri-Rao = column-wise Kronecker product
=> defined for matrices with **same number of columns**

$$A \in \mathbb{R}^{m,R}, B \in \mathbb{R}^{n,R}$$

$$\mathbf{A} \odot \mathbf{B} = \left[\mathbf{A}_{:,0} \otimes \mathbf{B}_{:,0}, \quad \mathbf{A}_{:,1} \otimes \mathbf{B}_{:,1}, \quad \dots, \quad \mathbf{A}_{:,R} \otimes \mathbf{B}_{:,R}, \right] \in \mathbb{R}^{mn,R}$$

Khatri-Rao product

- Khatri-Rao = column-wise Kronecker product
=> defined for matrices with **same number of columns**

$$\mathbf{A} \in \mathbb{R}^{m,R}, \mathbf{B} \in \mathbb{R}^{n,R}$$

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{A}_{:,0} \otimes \mathbf{B}_{:,0}, \quad \mathbf{A}_{:,1} \otimes \mathbf{B}_{:,1}, \quad \dots, \quad \mathbf{A}_{:,R} \otimes \mathbf{B}_{:,R}, \quad] \in \mathbb{R}^{mn,R}$$

$$(1) \quad (\mathbf{A} \odot \mathbf{B})^\top (\mathbf{A} \odot \mathbf{B}) = \mathbf{A}^\top \mathbf{A} * \mathbf{B}^\top \mathbf{B}$$

Khatri-Rao product

- Khatri-Rao = column-wise Kronecker product
=> defined for matrices with **same number of columns**

$$\mathbf{A} \in \mathbb{R}^{m,R}, \mathbf{B} \in \mathbb{R}^{n,R}$$

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{A}_{:,0} \otimes \mathbf{B}_{:,0}, \quad \mathbf{A}_{:,1} \otimes \mathbf{B}_{:,1}, \quad \dots, \quad \mathbf{A}_{:,R} \otimes \mathbf{B}_{:,R}, \quad] \in \mathbb{R}^{mn,R}$$

$$(1) \quad (\mathbf{A} \odot \mathbf{B})^\top (\mathbf{A} \odot \mathbf{B}) = \mathbf{A}^\top \mathbf{A} * \mathbf{B}^\top \mathbf{B}$$

$$(2) \quad (\mathbf{A} \odot \mathbf{B})^\dagger = ((\mathbf{A}^\top \mathbf{A}) * (\mathbf{B}^\top \mathbf{B}))^{-1} (\mathbf{A} \odot \mathbf{B})^\top$$

CP decomposition

$$\hat{\mathcal{X}} \in \mathbb{R}^{I,J,K}$$

$$\hat{\mathcal{X}} = [| \mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)} |]$$

- Collect factors in matrices: $\mathbf{U}^{(0)} = [\mathbf{u}_0^{(0)}, \mathbf{u}_1^{(0)}, \dots, \mathbf{u}_{R-1}^{(0)}] \in \mathbb{R}^{I,R}$

$$\mathbf{U}^{(1)} = [\mathbf{u}_0^{(1)}, \mathbf{u}_1^{(1)}, \dots, \mathbf{u}_{R-1}^{(1)}] \in \mathbb{R}^{J,R}$$

$$\mathbf{U}^{(2)} = [\mathbf{u}_0^{(2)}, \mathbf{u}_1^{(2)}, \dots, \mathbf{u}_{R-1}^{(2)}] \in \mathbb{R}^{K,R}$$

- Unfoldings can be expressed as khatri-rao product of the factors:

$$\hat{\mathcal{X}}_{[0]} = \mathbf{U}^{(0)} (\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)})^\top$$

$$\hat{\mathcal{X}}_{[1]} = \mathbf{U}^{(1)} (\mathbf{U}^{(0)} \odot \mathbf{U}^{(2)})^\top$$

$$\hat{\mathcal{X}}_{[2]} = \mathbf{U}^{(2)} (\mathbf{U}^{(0)} \odot \mathbf{U}^{(1)})^\top$$

CP decomposition: ALS

$$\begin{aligned} & \min_{\hat{\mathcal{X}}'} \|\hat{\mathcal{X}} - \hat{\mathcal{X}}'\|_F \\ = & \min_{\hat{\mathcal{X}}'} \|\hat{\mathcal{X}} - [\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}]\|_F \end{aligned}$$

Frobenius norm:

$$\|\mathcal{X}\|_F = \sqrt{\sum_{i_0=0}^{I_0} \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=0}^{I_N} x_{i_0 i_1 \dots i_N}^2}$$

CP decomposition: ALS

$$\begin{aligned} & \min_{\hat{\mathcal{X}}'} \|\hat{\mathcal{X}} - \hat{\mathcal{X}}'\|_F \\ = & \min_{\hat{\mathcal{X}}'} \|\hat{\mathcal{X}} - [|\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}|]\|_F \end{aligned}$$

- Alternating Least-Square: update one factors at a time, fix all others
- Rewrite the problem in term of the unfolding
- problem becomes *linear* least squares

Refresher on Linear Least Squares

$$\mathbf{Y} \in \mathbb{R}^{N,O}, \mathbf{X} \in \mathbb{R}^{N,F}, \beta \in \mathbb{R}^{F,O}$$

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{X}\beta\|_F$$

$$= (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta)$$

$$= \mathbf{Y}^\top \mathbf{Y} - \underbrace{\beta^\top \mathbf{X}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X} \beta}_{\text{symmetric terms}} + \beta^\top \mathbf{X}^\top \mathbf{X} \beta$$

$$= \mathbf{Y}^\top \mathbf{Y} - 2\beta^\top \mathbf{X}^\top \mathbf{Y} + \beta^\top \mathbf{X}^\top \mathbf{X} \beta$$

$$\rightarrow -\mathbf{X}^\top \mathbf{Y} + \mathbf{X}^\top \mathbf{X} \beta = 0 \rightarrow \beta = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$$

Refresher on Linear Least Squares

$$\mathbf{Y} \in \mathbb{R}^{N,O}, \mathbf{X} \in \mathbb{R}^{N,F}, \beta \in \mathbb{R}^{F,O}$$

$$\min_{\beta} \|\mathbf{Y}^\top - \beta^\top \mathbf{X}^\top\|_F$$

$$\rightarrow \beta = \mathbf{Y}^\top \underbrace{\mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1}}_{(\mathbf{X}^\top)^\dagger}$$

CP decomposition: ALS

$$\min_{\mathbf{U}^{(0)}} \left\| \underbrace{\hat{\mathcal{X}}_{[0]}}_{\mathbf{Y}^\top} - \underbrace{\mathbf{U}^{(0)}}_{\boldsymbol{\beta}^\top} \underbrace{(\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)})^\top}_{\mathbf{X}^\top} \right\|_F$$

$$\boldsymbol{\beta} = \mathbf{Y}^\top \mathbf{X} \underbrace{(\mathbf{X}^\top \mathbf{X})^{-1}}_{(\mathbf{X}^\top)^\dagger}$$

$$\mathbf{U}^{(0)} = \hat{\mathcal{X}}_{[0]} \left[(\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)})^\top \right]^\dagger$$

CP decomposition: ALS

$$\min_{\mathbf{U}^{(0)}} \|\hat{\mathcal{X}}_{[0]} - \mathbf{U}^{(0)} (\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)})^\top\|_F$$

$$\mathbf{U}^{(0)} = \hat{\mathcal{X}}_{[0]} \left[(\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)})^\top \right]^\dagger$$

CP decomposition: ALS

$$\min_{\mathbf{U}^{(0)}} \|\hat{\mathcal{X}}_{[0]} - \mathbf{U}^{(0)} (\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)})^\top\|_F$$

$$\mathbf{U}^{(0)} = \hat{\mathcal{X}}_{[0]} \left[(\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)})^\top \right]^\dagger \quad \rightarrow \quad \text{We can do better!}$$

Khatri-Rao product properties

$$\mathbf{A} \in \mathbb{R}^{m,R}, \mathbf{B} \in \mathbb{R}^{n,R}$$

$$\mathbf{A} \odot \mathbf{B} = \left[\mathbf{A}_{:,0} \otimes \mathbf{B}_{:,0}, \quad \mathbf{A}_{:,1} \otimes \mathbf{B}_{:,1}, \quad \dots, \quad \mathbf{A}_{:,R} \otimes \mathbf{B}_{:,R}, \right] \in \mathbb{R}^{mn,R}$$

$$(1) \quad (\mathbf{A} \odot \mathbf{B})^\top (\mathbf{A} \odot \mathbf{B}) = \mathbf{A}^\top \mathbf{A} * \mathbf{B}^\top \mathbf{B}$$

$$(2) \quad (\mathbf{A} \odot \mathbf{B})^\dagger = ((\mathbf{A}^\top \mathbf{A}) * (\mathbf{B}^\top \mathbf{B}))^{-1} (\mathbf{A} \odot \mathbf{B})^\top$$

CP decomposition: ALS

$$\min_{\mathbf{U}^{(0)}} \|\hat{\mathcal{X}}_{[0]} - \mathbf{U}^{(0)} (\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)})^\top\|_F$$

$$\mathbf{U}^{(0)} = \hat{\mathcal{X}}_{[0]} \left[(\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)})^\top \right]^\dagger$$

$$\mathbf{U}^{(0)} = \hat{\mathcal{X}}_{[0]} (\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)}) \left[(\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)})^\top (\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)}) \right]^{-1}$$

CP decomposition: ALS

$$\min_{\mathbf{U}^{(0)}} \|\hat{\mathcal{X}}_{[0]} - \mathbf{U}^{(0)} (\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)})^\top\|_F$$

$$\mathbf{U}^{(0)} = \hat{\mathcal{X}}_{[0]} \left[(\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)})^\top \right]^\dagger$$

$$\mathbf{U}^{(0)} = \hat{\mathcal{X}}_{[0]} (\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)}) \left[(\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)})^\top (\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)}) \right]^{-1}$$

$$\mathbf{U}^{(0)} = \hat{\mathcal{X}}_{[0]} (\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)}) \underbrace{\left[(\mathbf{U}^{(1)})^\top (\mathbf{U}^{(1)}) * (\mathbf{U}^{(2)})^\top (\mathbf{U}^{(2)}) \right]^{-1}}_{R \times R}$$

CP decomposition: ALS

$$\min_{\mathbf{U}^{(0)}} \|\hat{\mathcal{X}}_{[0]} - \mathbf{U}^{(0)} (\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)})^\top\|_F$$

$$\mathbf{U}^{(0)} = \hat{\mathcal{X}}_{[0]} (\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)}) \left[(\mathbf{U}^{(1)})^\top (\mathbf{U}^{(1)}) * (\mathbf{U}^{(2)})^\top (\mathbf{U}^{(2)}) \right]^{-1}$$

$$\mathbf{U}^{(1)} = \hat{\mathcal{X}}_{[1]} (\mathbf{U}^{(0)} \odot \mathbf{U}^{(2)}) \left[(\mathbf{U}^{(0)})^\top (\mathbf{U}^{(0)}) * (\mathbf{U}^{(2)})^\top (\mathbf{U}^{(2)}) \right]^{-1}$$

$$\mathbf{U}^{(2)} = \hat{\mathcal{X}}_{[2]} (\mathbf{U}^{(0)} \odot \mathbf{U}^{(1)}) \left[(\mathbf{U}^{(0)})^\top (\mathbf{U}^{(0)}) * (\mathbf{U}^{(1)})^\top (\mathbf{U}^{(1)}) \right]^{-1}$$

CP decomposition

- Rank of a tensor: smallest number of rank 1 tensors that generate the tensor as their sum
- No algorithm to determine rank (NP-hard)
- For third order tensor of size $I \times J \times K$:
 $\text{rank} \leq \min(IJ, IK, JK)$

CP decomposition

- Kruskal rank: maximum r such that every subset of r columns of A is linearly independent.
- For 3rd order decomposition, with factor matrices, A , B and C , sufficient condition for unicity of rank- R CP:

$$\text{krank}(A) + \text{krank}(B) + \text{krank}(C) \geq 2R + 2$$

- Mild condition if rank not too high: true in general (non-degenerate case)

CP decomposition

- Kruskal rank: maximum r such that every subset of r columns of A is linearly independent.
- For 3rd order decomposition, with factor matrices, A , B and C , sufficient condition for unicity of rank- R CP:

$$\text{krank}(A) + \text{krank}(B) + \text{krank}(C) \geq 2R + 2$$

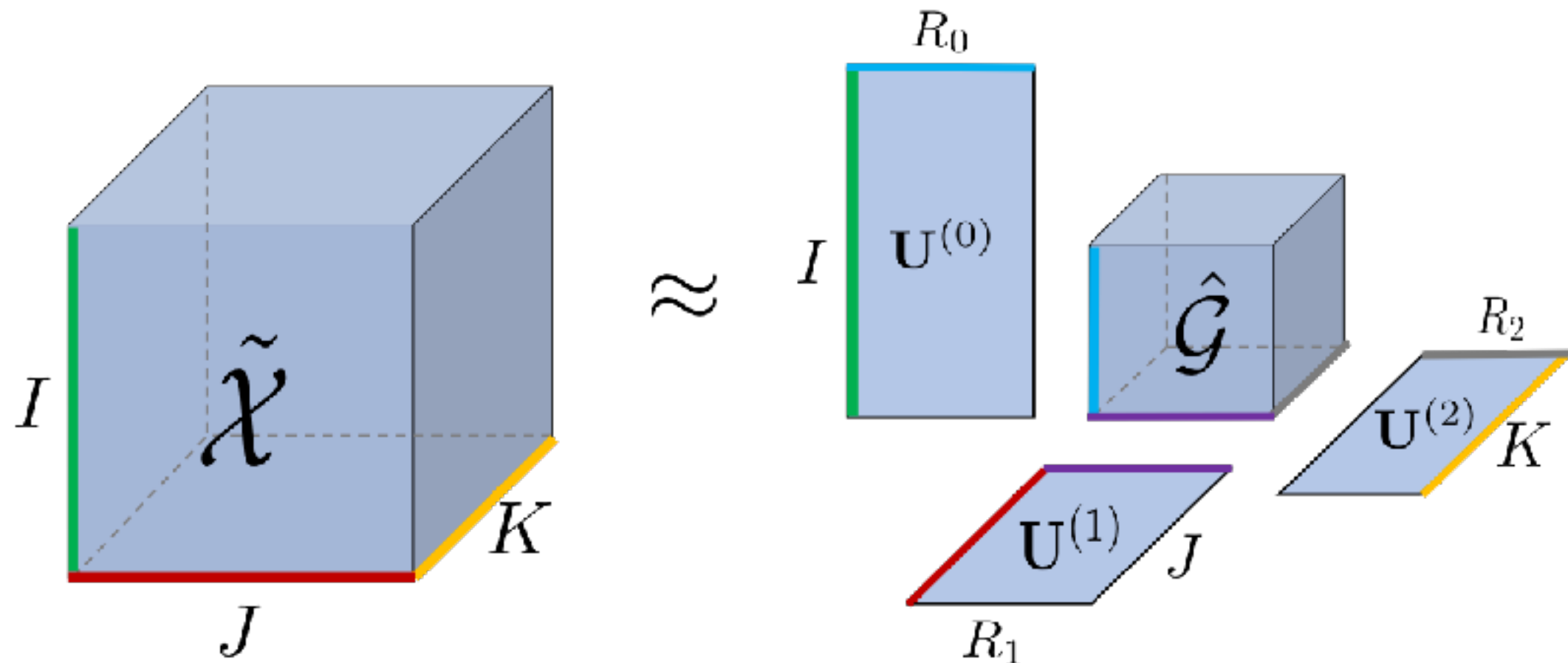
- Mild condition if rank not too high: true in general (non-degenerate case)
- Sufficient condition in general:

$$\sum_{k=0}^N \text{krank}(\mathbf{U}^{(k)}) \geq 2R + N$$

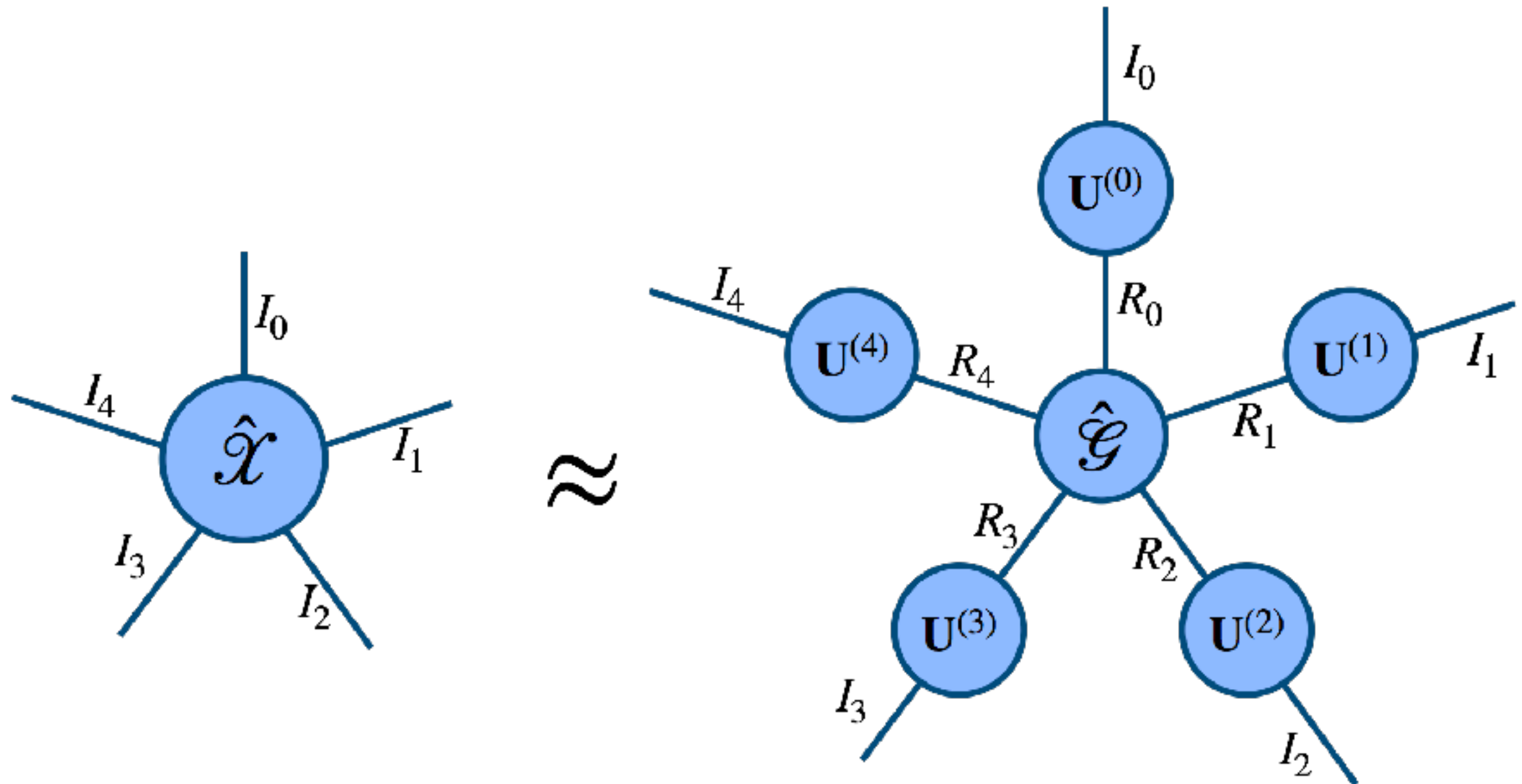
Tucker decomposition

$$\hat{\mathcal{X}} \in \mathbb{R}^{I,J,K}$$

$$\hat{\mathcal{X}} = \hat{\mathcal{G}} \times_0 \mathbf{U}^{(0)} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)}$$



Tucker decomposition



Tucker decomposition

$$\hat{\mathcal{X}} = \hat{\mathcal{G}} \times_0 \mathbf{U}^{(0)} \times_1 \mathbf{U}^{(1)} \times \cdots \times_N \mathbf{U}^{(N)}$$

$$\hat{\mathcal{X}} \in \mathbb{R}^{I_0, I_1, \dots, I_N}$$

$$\hat{\mathcal{G}} \in \mathbb{R}^{R_0, R_1, \dots, R_N}$$

$$\mathbf{U}^{(0)} \in \mathbb{R}^{I_0, R_0}, \mathbf{U}^{(1)} \in \mathbb{R}^{I_1, R_1}, \dots, \mathbf{U}^{(N)} \in \mathbb{R}^{I_N, R_N},$$

Tucker decomposition

$$\hat{\mathcal{X}} = \hat{\mathcal{G}} \times_0 \mathbf{U}^{(0)} \times_1 \mathbf{U}^{(1)} \times \cdots \times_N \mathbf{U}^{(N)}$$

$$\hat{\mathcal{X}} \in \mathbb{R}^{I_0, I_1, \dots, I_N}$$

$$\hat{\mathcal{G}} \in \mathbb{R}^{R_0, R_1, \dots, R_N}$$

$$\mathbf{U}^{(0)} \in \mathbb{R}^{I_0, R_0}, \mathbf{U}^{(1)} \in \mathbb{R}^{I_1, R_1}, \dots, \mathbf{U}^{(N)} \in \mathbb{R}^{I_N, R_N},$$

- Short form: $\hat{\mathcal{G}} \times_0 \mathbf{U}^{(0)} \times_1 \mathbf{U}^{(1)} \times \cdots \times_N \mathbf{U}^{(N)} = [| \hat{\mathcal{G}}; \mathbf{U}^{(0)}, \dots, \mathbf{U}^{(N)} |]$

Tucker decomposition

$$\hat{\mathcal{X}} = \hat{\mathcal{G}} \times_0 \mathbf{U}^{(0)} \times_1 \mathbf{U}^{(1)} \times \cdots \times_N \mathbf{U}^{(N)}$$

$$\hat{\mathcal{X}} \in \mathbb{R}^{I_0, I_1, \dots, I_N}$$

$$\hat{\mathcal{G}} \in \mathbb{R}^{R_0, R_1, \dots, R_N}$$

$$\mathbf{U}^{(0)} \in \mathbb{R}^{I_0, R_0}, \mathbf{U}^{(1)} \in \mathbb{R}^{I_1, R_1}, \dots, \mathbf{U}^{(N)} \in \mathbb{R}^{I_N, R_N},$$

- Short form: $\hat{\mathcal{G}} \times_0 \mathbf{U}^{(0)} \times_1 \mathbf{U}^{(1)} \times \cdots \times_N \mathbf{U}^{(N)} = [| \hat{\mathcal{G}}; \mathbf{U}^{(0)}, \dots, \mathbf{U}^{(N)} |]$
- Equivalent: $\hat{\mathcal{X}} = \hat{\mathcal{G}} (\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \dots, \mathbf{U}^{(N)})$

Tucker decomposition

$$\hat{\mathcal{X}} = \hat{\mathcal{G}} \times_0 \mathbf{U}^{(0)} \times_1 \mathbf{U}^{(1)} \times \dots \times_N \mathbf{U}^{(N)}$$

- Short form: $\hat{\mathcal{G}} \times_0 \mathbf{U}^{(0)} \times_1 \mathbf{U}^{(1)} \times \dots \times_N \mathbf{U}^{(N)} = [| \hat{\mathcal{G}}; \mathbf{U}^{(0)}, \dots, \mathbf{U}^{(N)} |]$
- Equivalent: $\hat{\mathcal{X}} = \hat{\mathcal{G}} (\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \dots, \mathbf{U}^{(N)})$
- We can isolate the core and the factors by considering vectorized expression and unfoldings:

Tucker decomposition

$$\hat{\mathcal{X}} = \hat{\mathcal{G}} \times_0 \mathbf{U}^{(0)} \times_1 \mathbf{U}^{(1)} \times \dots \times_N \mathbf{U}^{(N)}$$

- Short form: $\hat{\mathcal{G}} \times_0 \mathbf{U}^{(0)} \times_1 \mathbf{U}^{(1)} \times \dots \times_N \mathbf{U}^{(N)} = [|\hat{\mathcal{G}}; \mathbf{U}^{(0)}, \dots, \mathbf{U}^{(N)}|]$
- Equivalent: $\hat{\mathcal{X}} = \hat{\mathcal{G}} (\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \dots, \mathbf{U}^{(N)})$
- We can isolate the core and the factors by considering vectorized expression and unfoldings:

$$\text{vec}(\hat{\mathcal{X}}) = (\mathbf{U}^{(0)} \otimes \dots \otimes \mathbf{U}^{(N)}) \text{vec}(\hat{\mathcal{G}})$$

$$\hat{\mathcal{X}}_{[n]} = \mathbf{U}^{(n)} \hat{\mathcal{G}}_{[n]} (\mathbf{U}^{(0)} \otimes \dots \otimes \mathbf{U}^{(n-1)} \otimes \mathbf{U}^{(n+1)} \otimes \dots \otimes \mathbf{U}^{(N)})^\top$$

Higher Order SVD

$$\hat{\mathcal{X}} = \hat{\mathcal{G}} \times_0 \mathbf{U}^{(0)} \times_1 \mathbf{U}^{(1)} \times \cdots \times_N \mathbf{U}^{(N)}$$

- For $n=0, \dots, N$:

$$\mathbf{U}^{(n)} = R_n \text{ leading left singular vectors of } \hat{\mathcal{X}}_{[n]}$$

Truncated SVD

- End for

- $\hat{\mathcal{G}} = \hat{\mathcal{X}} \times_0 \mathbf{U}^{(0)T} \times_1 \mathbf{U}^{(1)T} \times \cdots \times_N \mathbf{U}^{(N)T}$

Higher Order SVD

$$\hat{\mathcal{X}} = \hat{\mathcal{G}} \times_0 \mathbf{U}^{(0)} \times_1 \mathbf{U}^{(1)} \times \dots \times_N \mathbf{U}^{(N)}$$

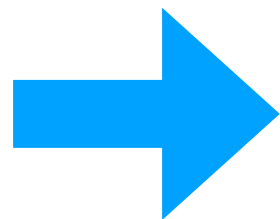
- For $n=0, \dots, N$:

$$\mathbf{U}^{(n)} = R_n \text{ leading left singular vectors of } \hat{\mathcal{X}}_{[n]}$$

Truncated SVD

- End for

- $\hat{\mathcal{G}} = \hat{\mathcal{X}} \times_0 \mathbf{U}^{(0)T} \times_1 \mathbf{U}^{(1)T} \times \dots \times_N \mathbf{U}^{(N)T}$



Good initialisation

Tucker via ALS

(Higher-Order Orthogonal Iteration)

- Use SVD to compute orthonormal factors

$$\hat{\mathcal{X}} = \hat{\mathcal{G}} \times_0 \mathbf{U}^{(0)} \times_1 \mathbf{U}^{(1)} \times \dots \times_N \mathbf{U}^{(N)}$$

- While not converged:

- For $n=0, \dots, N$:

$$\hat{\mathcal{Y}} = \hat{\mathcal{X}} \times_0 \mathbf{U}^{(0)T} \times_1 \mathbf{U}^{(1)T} \times \dots \times_{n-1} \mathbf{U}^{(n-1)T} \times_{n+1} \mathbf{U}^{(n+1)T} \times \dots \times_N \mathbf{U}^{(N)T}$$

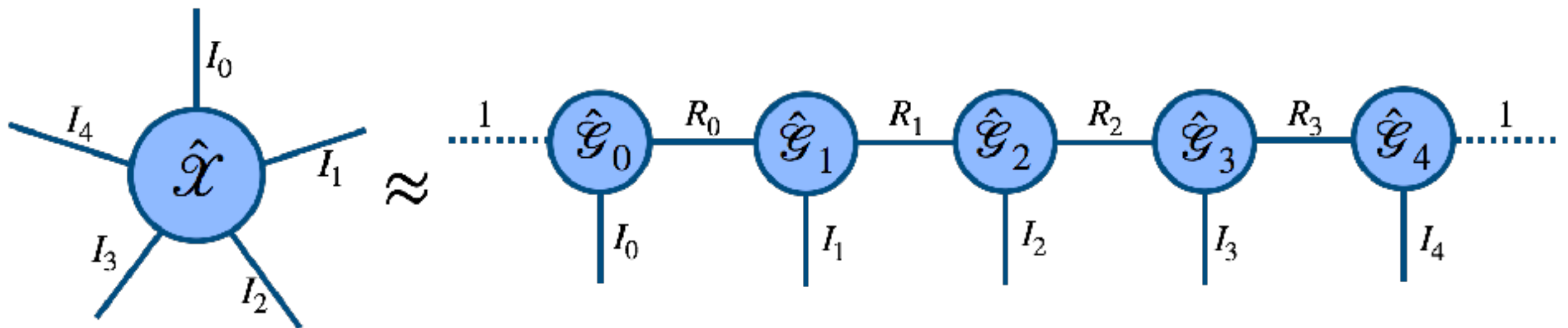
$$\mathbf{U}^{(n)} = R_n \text{ leading left singular vectors of } \hat{\mathcal{Y}}_{[n]}$$

- End for

$$\hat{\mathcal{G}} = \hat{\mathcal{X}} \times_0 \mathbf{U}^{(0)T} \times_1 \mathbf{U}^{(1)T} \times \dots \times_N \mathbf{U}^{(N)T}$$

MPS / T-Train

- Physics: Matrix Product State
- Machine Learning: Tensor-Train (I. V. Oseledets)



MPS / T-Train

$$\hat{\mathcal{X}} \in \mathbb{R}^{I_0, I_1, \dots, I_N}, \quad \textbf{Ranks } R_0, \dots, R_N, \quad , R_0 = R_N = 1$$

$$\hat{\mathcal{C}} = \hat{\mathcal{X}}$$

- For $n=0, \dots, N$:

$$\mathbf{C} = \textbf{reshape}(\hat{\mathcal{C}}, R_{k-1} \times I_k, -1)$$

$$\mathbf{C} = \mathbf{U}\mathbf{S}\mathbf{V}, \quad \textbf{keep first } R_k \textbf{ components of } \mathbf{U}$$

$$\hat{\mathcal{G}}_k = \textbf{reshape}(\mathbf{U}[:, : k], [R_{k-1}, I_k, R_k])$$

$$\mathbf{C} = \mathbf{S}\mathbf{V}^\top$$

- End for

$$\hat{\mathcal{G}}_N = \mathbf{C}$$

Tensor decompositions

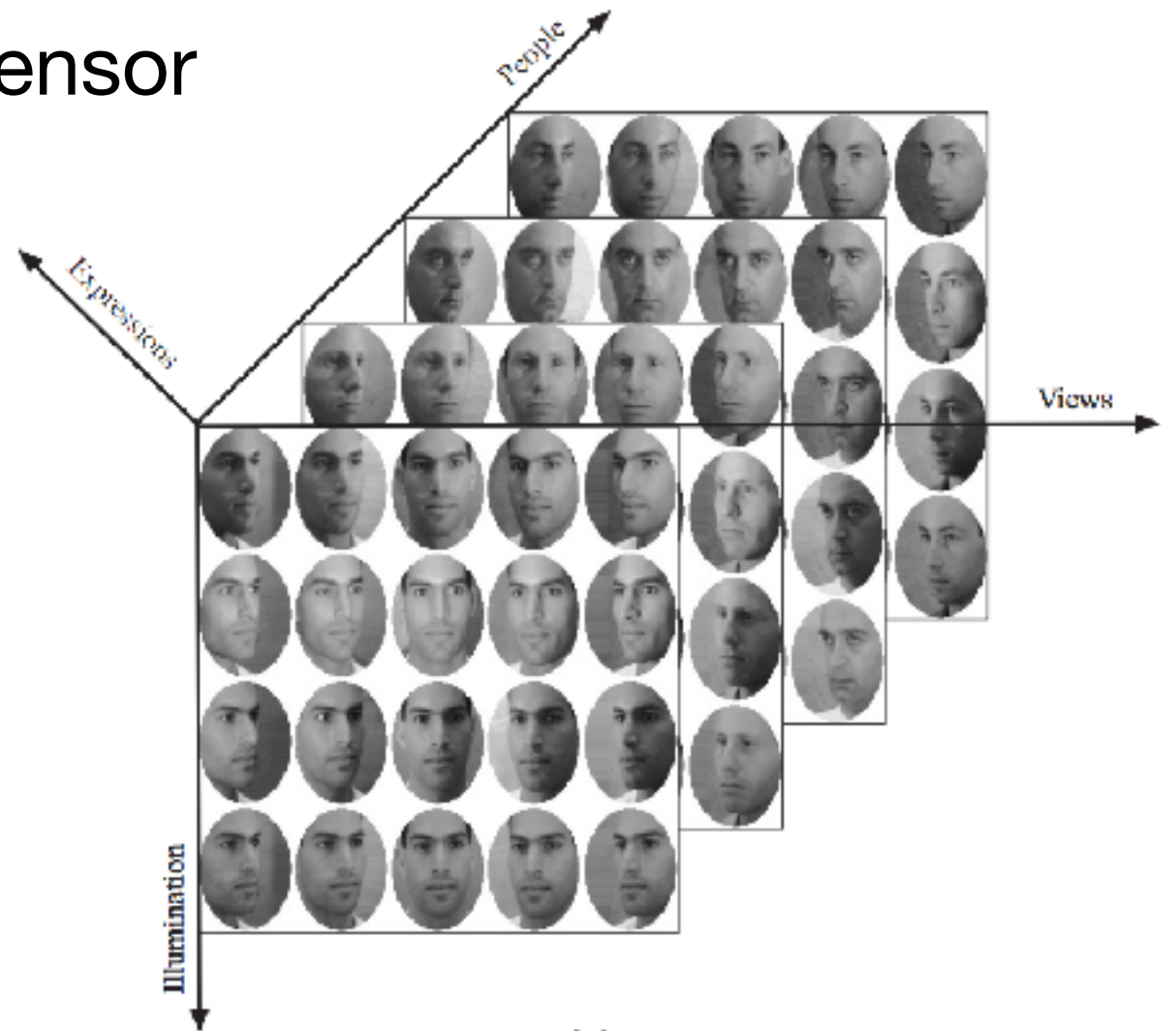
- Many others (Hierarchical Tucker, DEDICOM, PARAFAC2, ...)
- Excellent surveys:
 - Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. SIAM REVIEW, 51(3):455–500, 2009.
 - Sidiropoulos, N.D., De Lathauwer, L., Fu, X., Huang, K., Papalexakis, E.E. and Faloutsos, C., 2017. Tensor decomposition for signal processing and machine learning. IEEE Transactions on Signal Processing, 65(13), pp.3551-3582.

Tensor decompositions

- Many others (Hierarchical Tucker, DEDICOM, PARAFAC2, ...)
- Excellent surveys:
 - Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. SIAM REVIEW, 51(3):455–500, 2009.
 - Sidiropoulos, N.D., De Lathauwer, L., Fu, X., Huang, K., Papalexakis, E.E. and Faloutsos, C., 2017. Tensor decomposition for signal processing and machine learning. IEEE Transactions on Signal Processing, 65(13), pp.3551-3582.
- Some notable examples:

TensorFaces

- Arrange image of faces in a tensor (e.g. subject x illumination x expression x pixels)
- Can be used for recognition
- Can remove effect (e.g. illumination from faces)



M. Alex O. Vasilescu and Demetri Terzopoulos, Multilinear Analysis of Image Ensembles: TensorFaces, ECCV'02

M. Alex O. Vasilescu and Demetri Terzopoulos, Multilinear Subspace Analysis of Image Ensembles, CVPR'03

Method of Moments

- Pearson: find parameters (of models) consistent with observed moments: $\mathbb{E}[X], \mathbb{E}[X^2], \mathbb{E}[X^3], \dots$
- Topic modelling: moments represent probabilities of occurrence of words, co-occurrence, etc..
- By decomposing the moment tensors, we can recover the parameters (here, the rank of the tensor corresponds to the number of topics)

A Anandkumar, R Ge, D Hsu, SM Kakade, M Telgarsky,
Tensor Decompositions for Learning Latent Variable Models, JMLR, 2014

Expressiveness of deep nets

- Equivalence between networks and tensor factorisation
- Shallow networks \leftrightarrow CP decomposition
- Deeper networks correspond to hierarchical Tucker

Nadav Cohen, Or Sharir, Amnon Shashua,
On the Expressive Power of Deep Learning: A Tensor Analysis, 2016



Any questions?



@JeanKossaifi
jean.kossaifi@gmail.com