Tensor decomposition

Jean Kossaifi

@JeanKossaifi

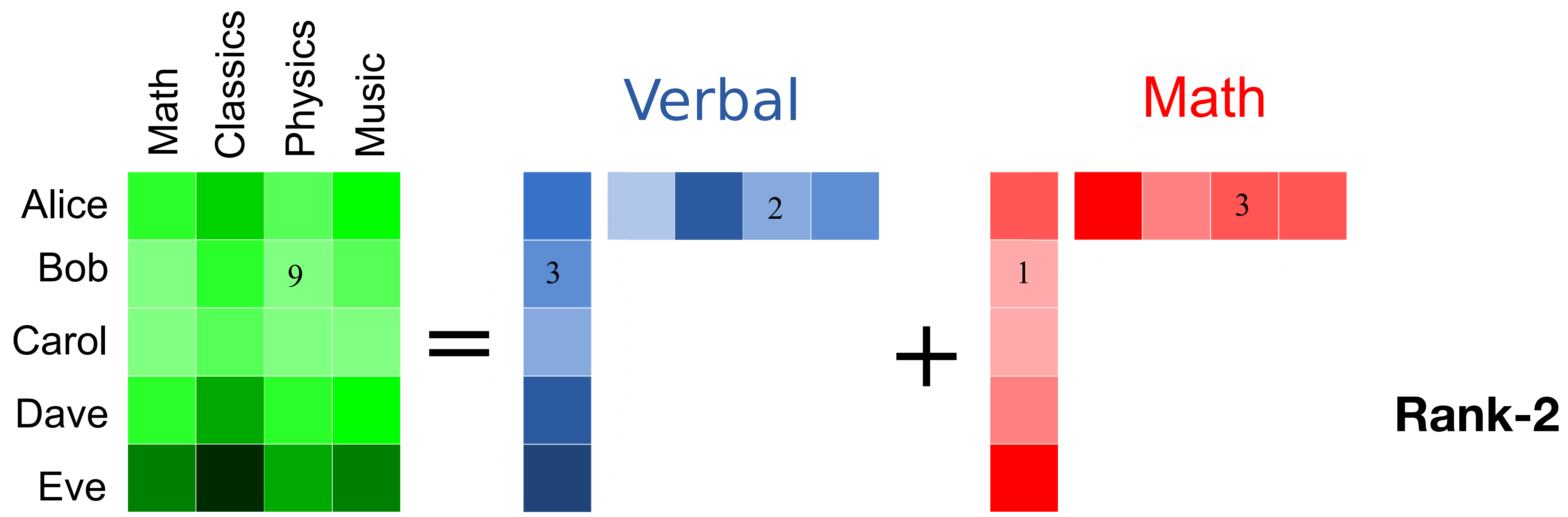
jean.kossaifi@gmail.com

Example: factor analysis

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

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

$$\mathbf{A}\mathbf{B} = \sum_{k=1}^{R} \mathbf{a}_{:,k}\mathbf{b}_{k,:}^{\top} = \sum_{k=1}^{R} \underbrace{\mathbf{a}_{:,k} \circ \mathbf{b}_{k,:}}_{\textbf{rank-1 components}}$$

$$\begin{pmatrix} b_{0,0}, & b_{0,1}, & \cdots, & b_{0,n} \end{pmatrix} \qquad \begin{pmatrix} b_{p,0}, & b_{p,1}, & \cdots, & b_{p,n} \end{pmatrix}$$

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

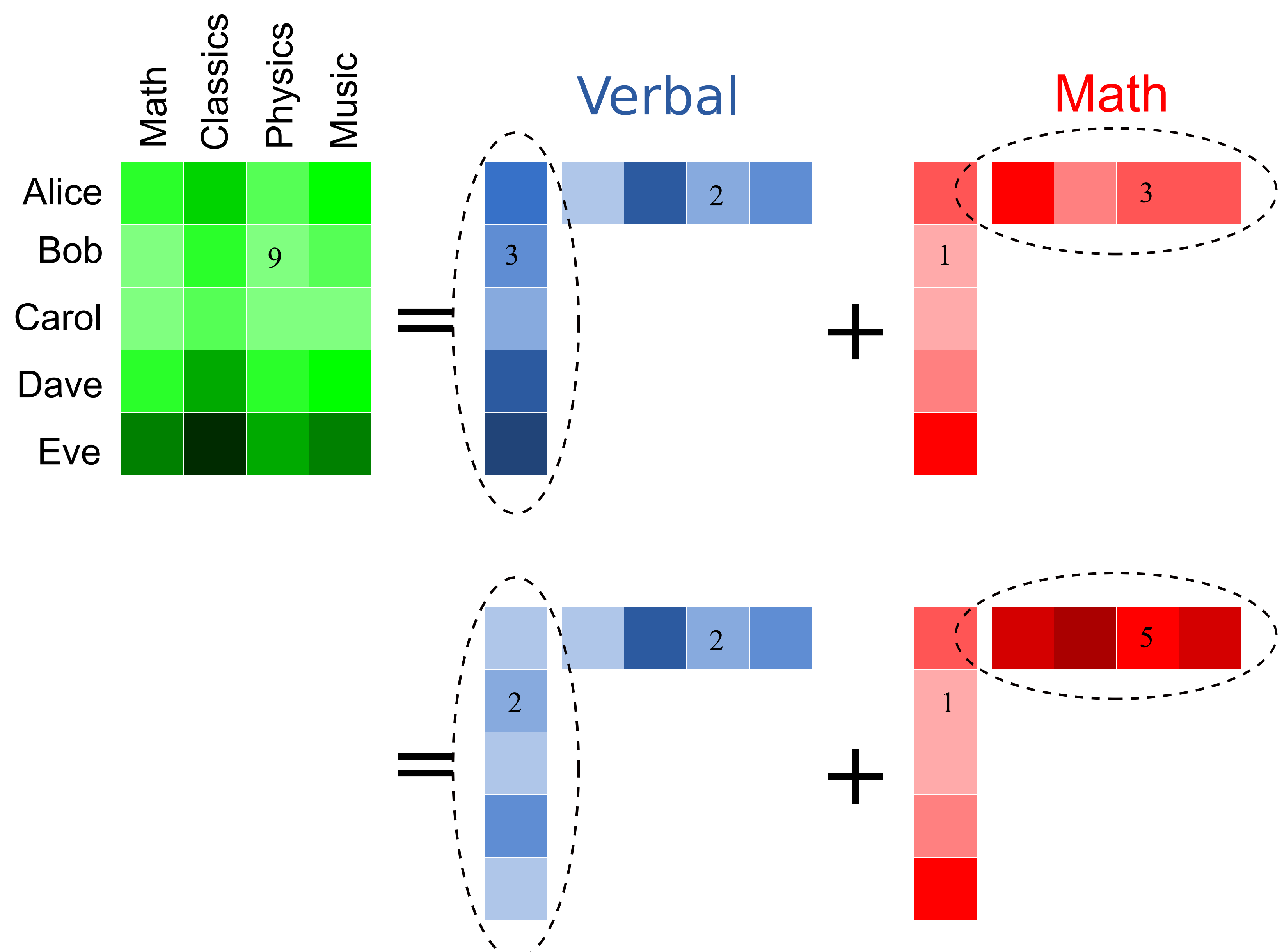
Ambiguity of Matrix Decomposition

- Low rank decomposition of matrix: $\mathbf{M} = \sum_{k=1}^{R} \mathbf{a}_k \mathbf{b}_k^\top = \mathbf{A}\mathbf{B}^\top$

- One possible choice (SVD on M): $\mathbf{M} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$

$$\mathbf{A} = \mathbf{U}\mathbf{S} \qquad \mathbf{B} = \mathbf{V}^\top$$

Ambiguity of Matrix Decomposition



Ambiguity of Matrix Decomposition

- Low rank decomposition of matrix: $\mathbf{M} = \sum_{k=1}^{R} \mathbf{a}_k \mathbf{b}_k^\top = \mathbf{A}\mathbf{B}^\top$

- One possible choice (truncated SVD on M): $\mathbf{M} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$

$$\mathbf{A} = \mathbf{U}[: , : R]\mathbf{S}[: R, : R] \qquad \mathbf{B} = \mathbf{V}^\top[: R\, ., :]$$

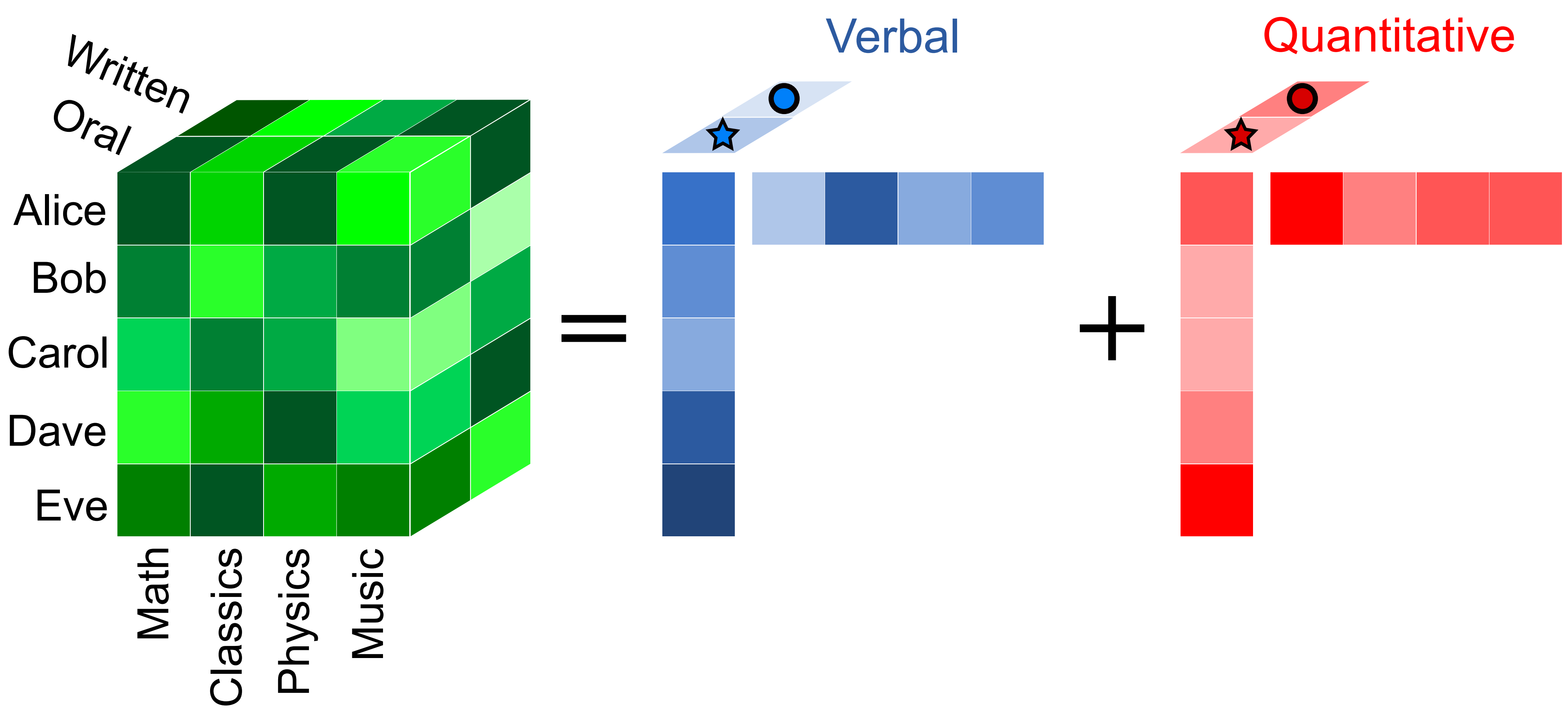
- For any orthonormal matrix $\mathbf{R} \in \mathbb{R}^{R \times R}$ such that $\mathbf{R}\mathbf{R}^\top = \mathbf{R}^\top\mathbf{R} = \mathbf{I}_R$

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

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

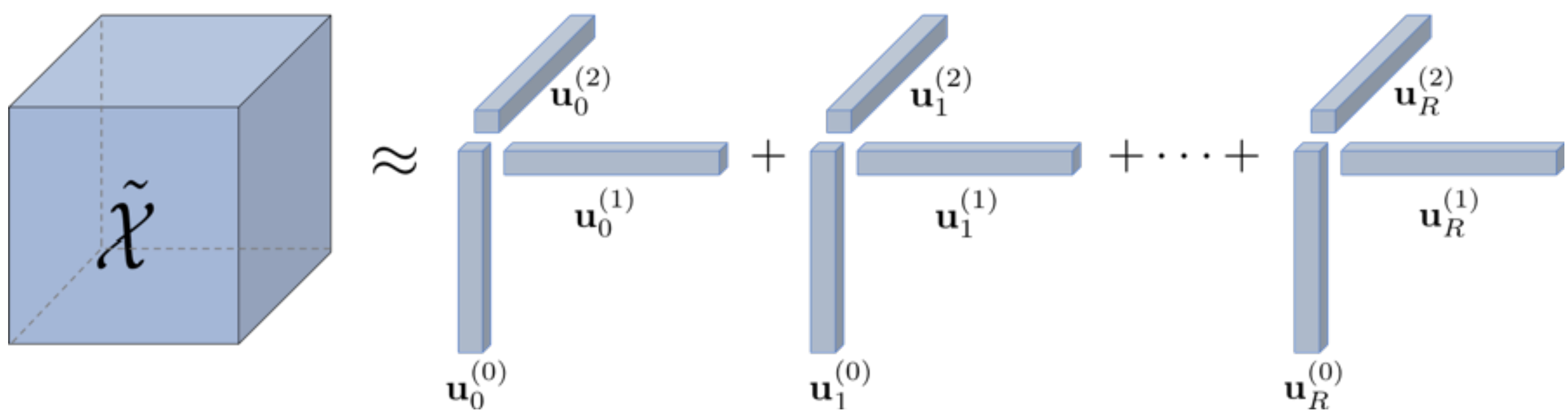
- Matrix decomposition is not *unique*
- Tensor decomposition unique under mild conditions

  ➡ Identifiability
- Almost all tensor problems are NP-hard…
- Good algorithms to solve them heuristically

CP decomposition

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

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



CP decomposition

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

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

- 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

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

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

- Collect factors in matrices:

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

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

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

CP decomposition

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

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

- Collect factors in matrices:

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

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

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

- Problem can be rewritten as:

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

Khatri-Rao product

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

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

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix}\mathbf{A}_{:,0} \otimes \mathbf{B}_{:,0}, & \mathbf{A}_{:,1} \otimes \mathbf{B}_{:,1}, & \cdots, & \mathbf{A}_{:,R} \otimes \mathbf{B}_{:,R}, & \end{bmatrix} \in \mathbb{R}^{mn,R}$$

Khatri-Rao product

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

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

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

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

Khatri-Rao product

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

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

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} \mathbf{A}_{:,0} \otimes \mathbf{B}_{:,0}, & \mathbf{A}_{:,1} \otimes \mathbf{B}_{:,1}, & \cdots, & \mathbf{A}_{:,R} \otimes \mathbf{B}_{:,R}, & \end{bmatrix} \in \mathbb{R}^{mn,R}$$

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

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

CP decomposition

$$\hat{\mathcal{X}} \in \mathbb{R}^{I,J,K} \qquad \hat{\mathcal{X}} = [\,|\,\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}\,|\,]$$

- Collect factors in matrices:

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

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

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

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

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

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

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

CP decomposition: ALS

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

- Frobenius norm:

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

CP decomposition: ALS

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

- 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

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

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

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

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

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

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

Refresher on Linear Least Squares

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

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

$$\Rightarrow \quad \beta = \mathbf{Y}^\top \underbrace{\mathbf{X}\left(\mathbf{X}^\top \mathbf{X}\right)^{-1}}_{\left(\mathbf{X}^\top\right)^\dagger}$$

CP decomposition: ALS

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

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

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

CP decomposition: ALS

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

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

Khatri-Rao product: properties

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

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} \mathbf{A}_{:,0} \otimes \mathbf{B}_{:,0}, & \mathbf{A}_{:,1} \otimes \mathbf{B}_{:,1}, & \cdots, & \mathbf{A}_{:,R} \otimes \mathbf{B}_{:,R}, & \end{bmatrix} \in \mathbb{R}^{mn,R}$$

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

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

CP decomposition: ALS

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

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

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

CP decomposition: ALS

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

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

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

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

CP decomposition: ALS

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

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

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

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

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 x J x K:

  rank <= 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:

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

- Mild condition if rank not too high: in general (non-degenerate case):
  krank(A) = min(n_rows, n_columns)…

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:

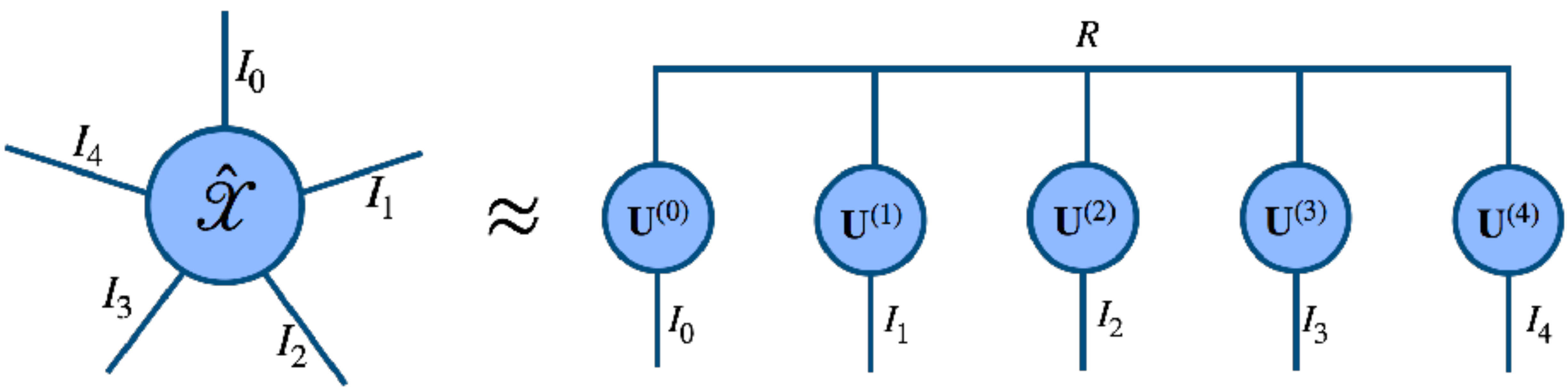
$$krank(A) + krank(B) + krank(C) \geq 2R + 2$$

- Mild condition if rank not too high

- Sufficient condition in general:

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

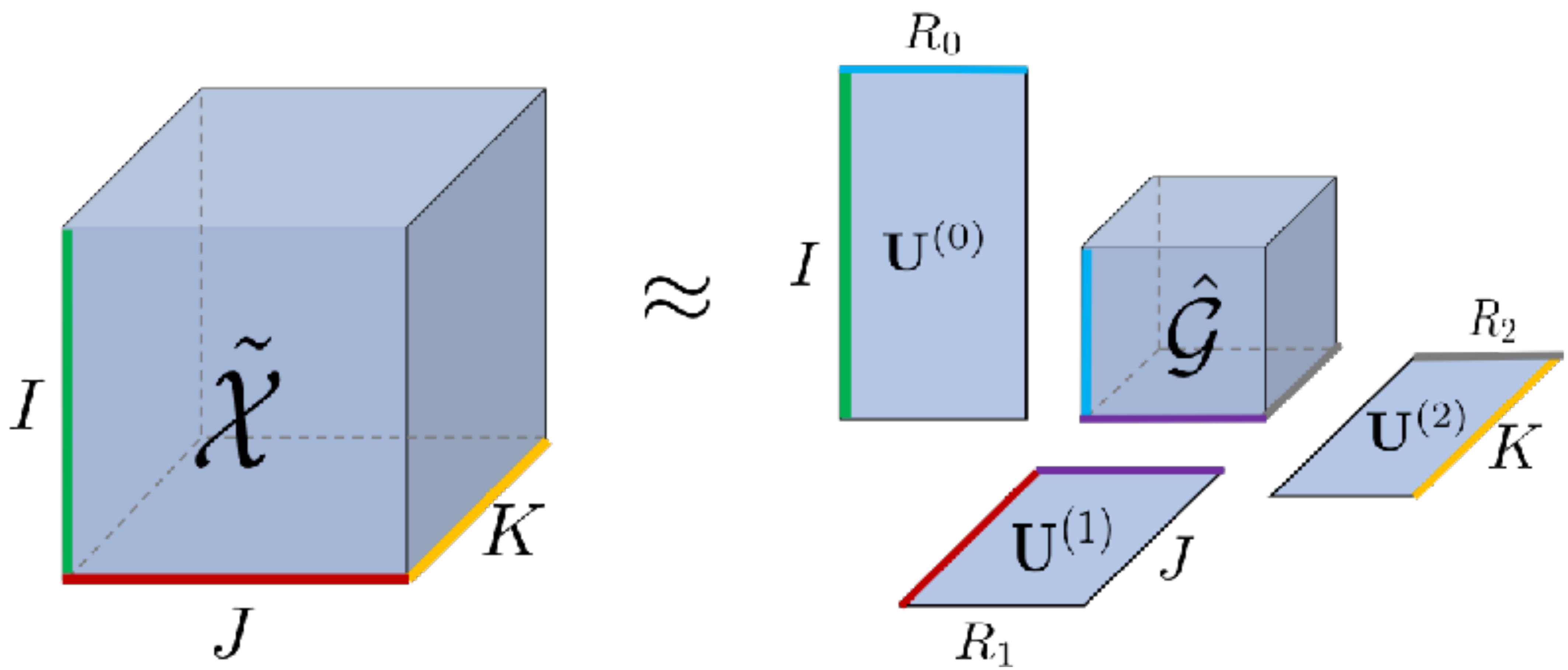
CP decomposition



Tucker decomposition

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

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



Tucker decomposition

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

$$\hat{\mathscr{X}} \in \mathbb{R}^{I_0, I_1, \cdots, I_N}$$

$$\hat{\mathscr{G}} \in \mathbb{R}^{R_0, R_1, \cdots, R_N}$$

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

Tucker decomposition

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

$$\hat{\mathscr{X}} \in \mathbb{R}^{I_0, I_1, \cdots, I_N}$$

$$\hat{\mathscr{G}} \in \mathbb{R}^{R_0, R_1, \cdots, R_N}$$

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

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

Tucker decomposition

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

$$\hat{\mathscr{X}} \in \mathbb{R}^{I_0, I_1, \cdots, I_N}$$

$$\hat{\mathscr{G}} \in \mathbb{R}^{R_0, R_1, \cdots, R_N}$$

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

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

Tucker decomposition

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

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

$$vec(\hat{\mathscr{X}}) = \left(\mathbf{U}^{(0)} \otimes \cdots \otimes \mathbf{U}^{(N)} \right) vec(\hat{\mathscr{G}})$$

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

Higher Order SVD

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

- For n=0, …N:

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

  Truncated SVD

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

Tucker via ALS (Higher-Order Orthogonal Iteration)

- Use SVD to compute orthonormal factors

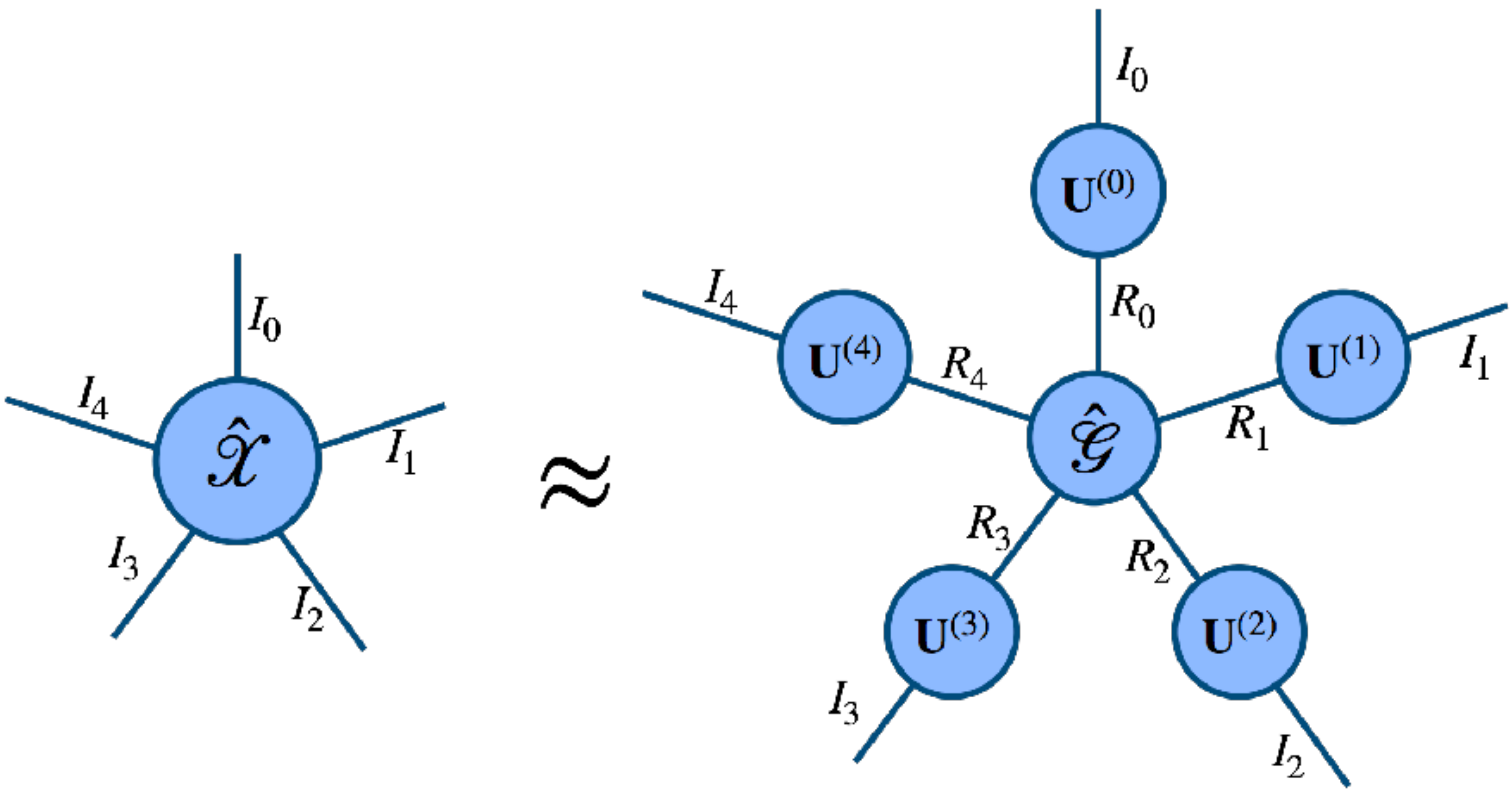
$$\hat{\mathcal{X}} = \hat{\mathcal{G}} \times_0 \mathbf{U}^{(0)} \times_1 \mathbf{U}^{(1)} \times \cdots \times_N \mathbf{U}^{(N)}$$

- While not converged:
  - For n=0, …N:

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

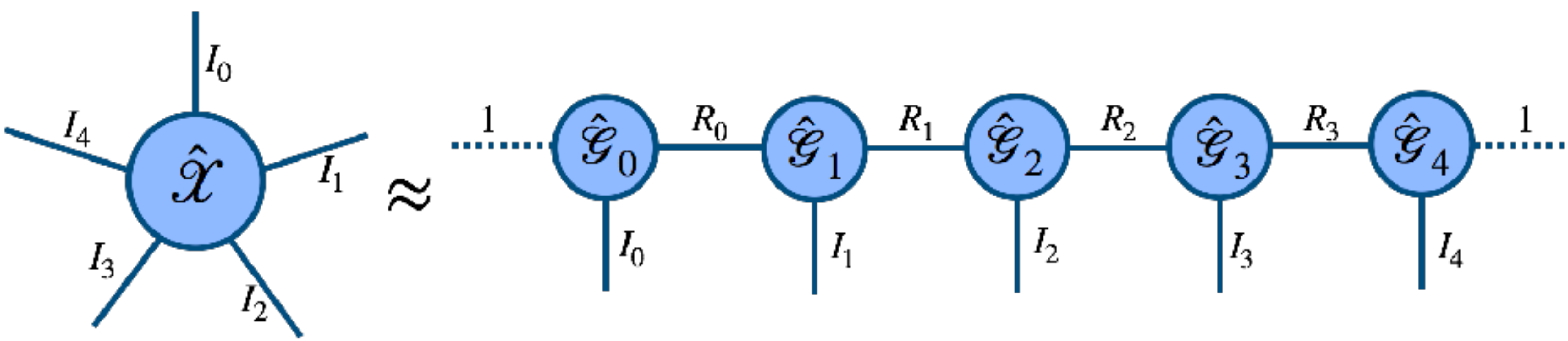
    $\mathbf{U}^{(n)} = R_n$ **leading left singular vectors of** $\hat{\mathcal{Y}}_{[n]}$
  - End for
- $\hat{\mathcal{G}} = \hat{\mathcal{X}} \times_0 \mathbf{U}^{(0)^T} \times_1 \mathbf{U}^{(1)^T} \times \cdots \times_N \mathbf{U}^{(N)^T}$

Tucker decomposition



MPS / T-Train

- Physics: Matrix Product State
- Machine Learning: Tensor-Train



MPS / T-Train

$\hat{\mathscr{X}} \in \mathbb{R}^{I_0, I_1, \cdots, I_N}$, **Ranks** $R_0, \cdots, R_N$, $, R_0 = R_N = 1$

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

- For n=0, …N:

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

  $\mathbf{C} = \mathbf{USV}$, **keep first** R_k**components of** $\mathbf{U}$

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

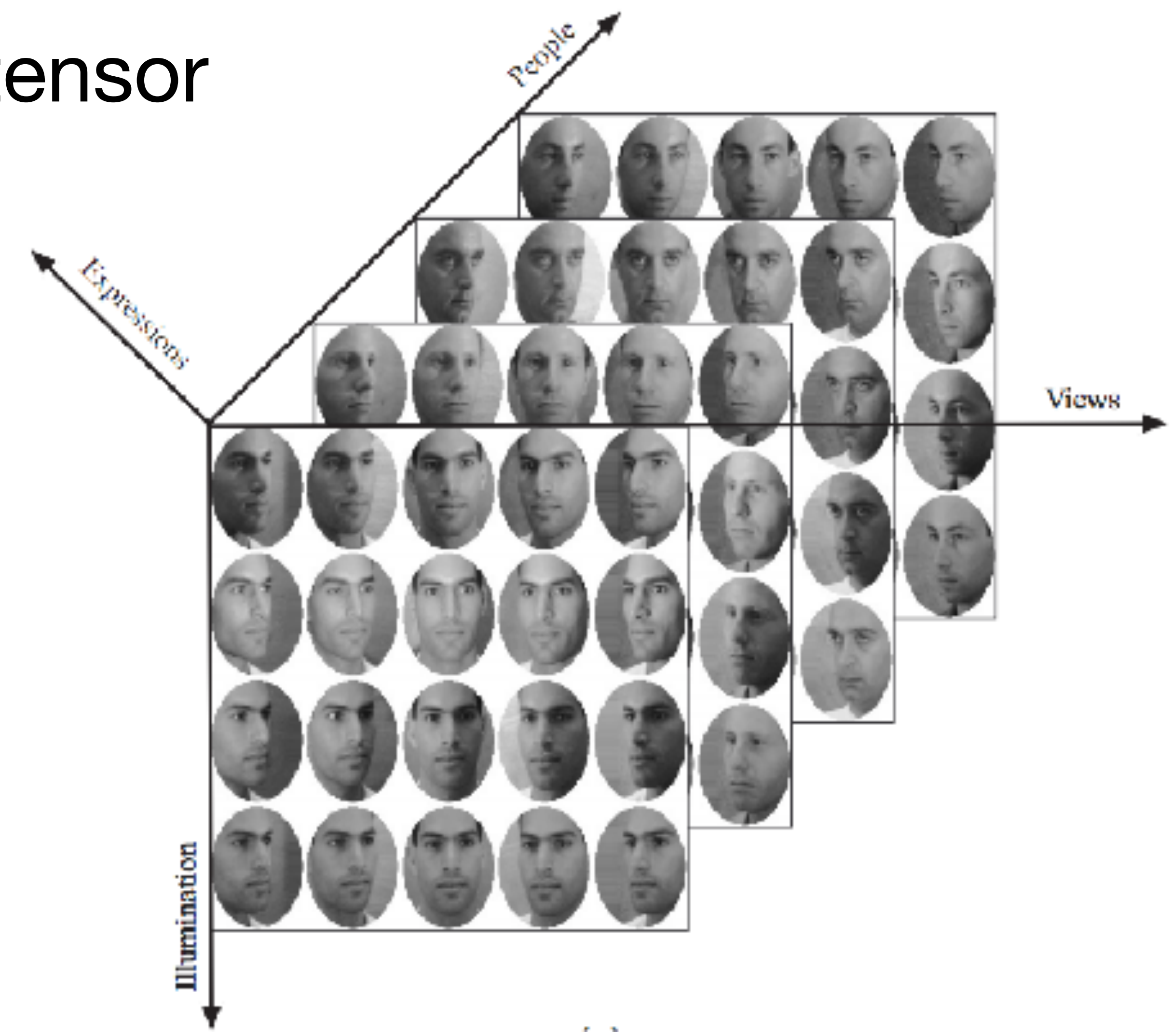
  $\mathbf{C} = \mathbf{SV}^\top$

- End for

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

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: $\mathbb{E}[X], \mathbb{E}[X^2], \mathbb{E}[X^3], \cdots$
- Topic modelling: moments represent probabilities of occurrence of words, co-occurence, 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 <-> 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