Roadmap

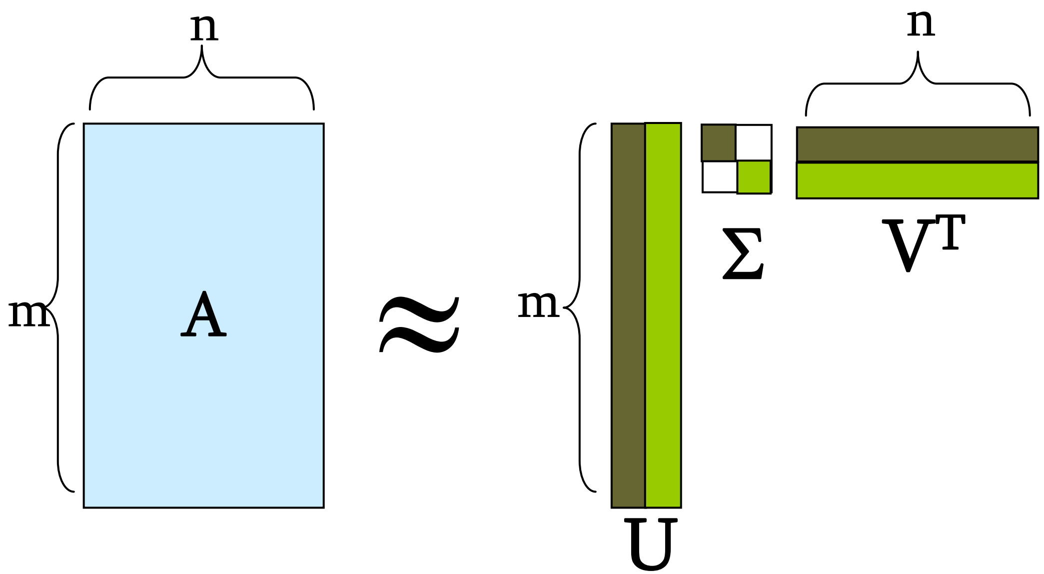
- Motivation
- Matrix tools
- Tensor tools
- Case studies

- Tensor Basics
- Tucker
  - Tucker 1
  - Tucker 2
  - Tucker 3
- PARAFAC
- Incrementalization

Tensor Basics

Reminder: SVD

$$\mathbf{A} \approx \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_i \sigma_i \mathbf{u}_i \circ \mathbf{v}_i$$

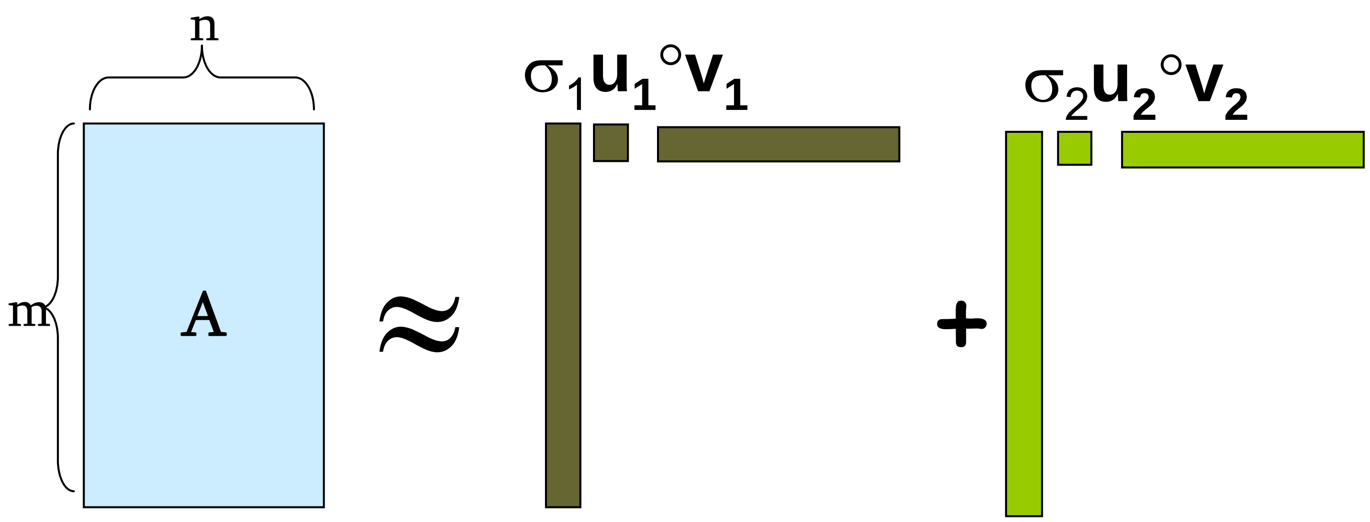


– Best rank-k approximation in L2

See also PARAFAC

Reminder: SVD

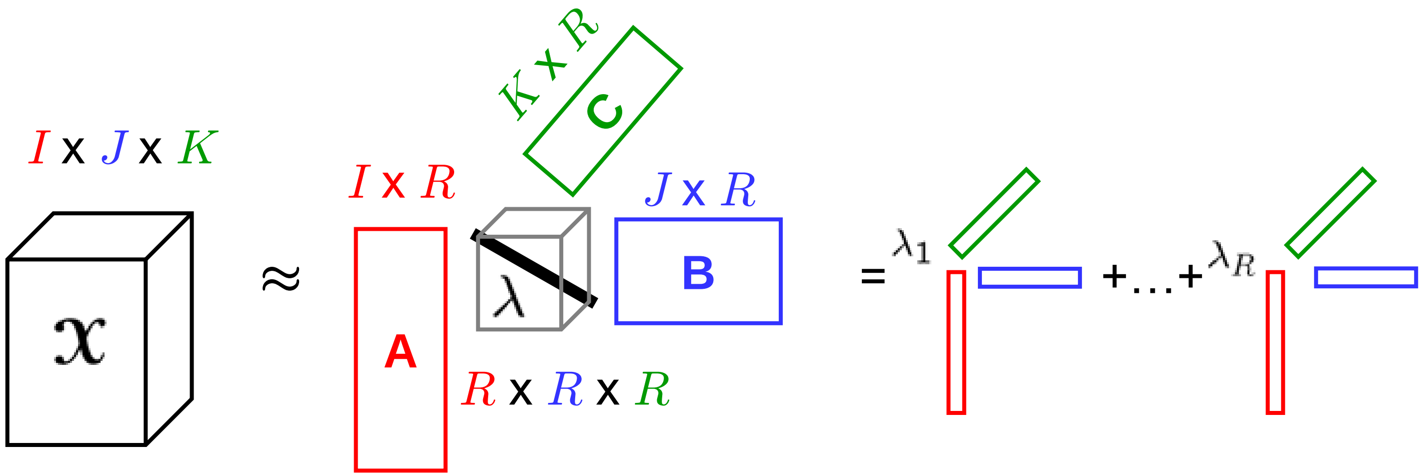
$$\mathbf{A} \approx \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_i \sigma_i \mathbf{u}_i \circ \mathbf{v}_i$$



– Best rank-k approximation in L2

See also PARAFAC

Goal: extension to >=3 modes

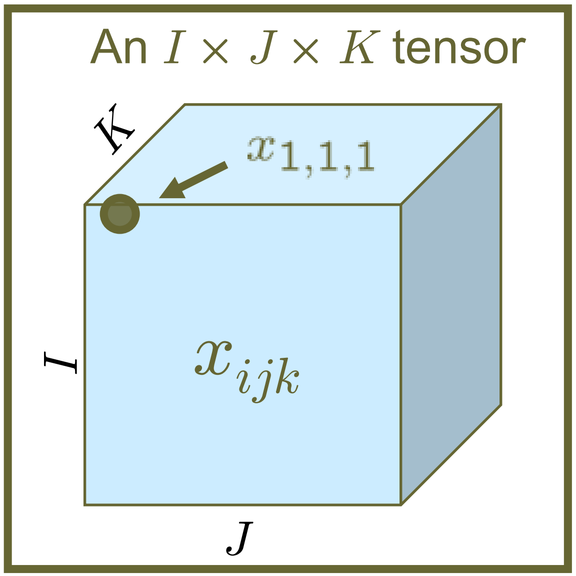


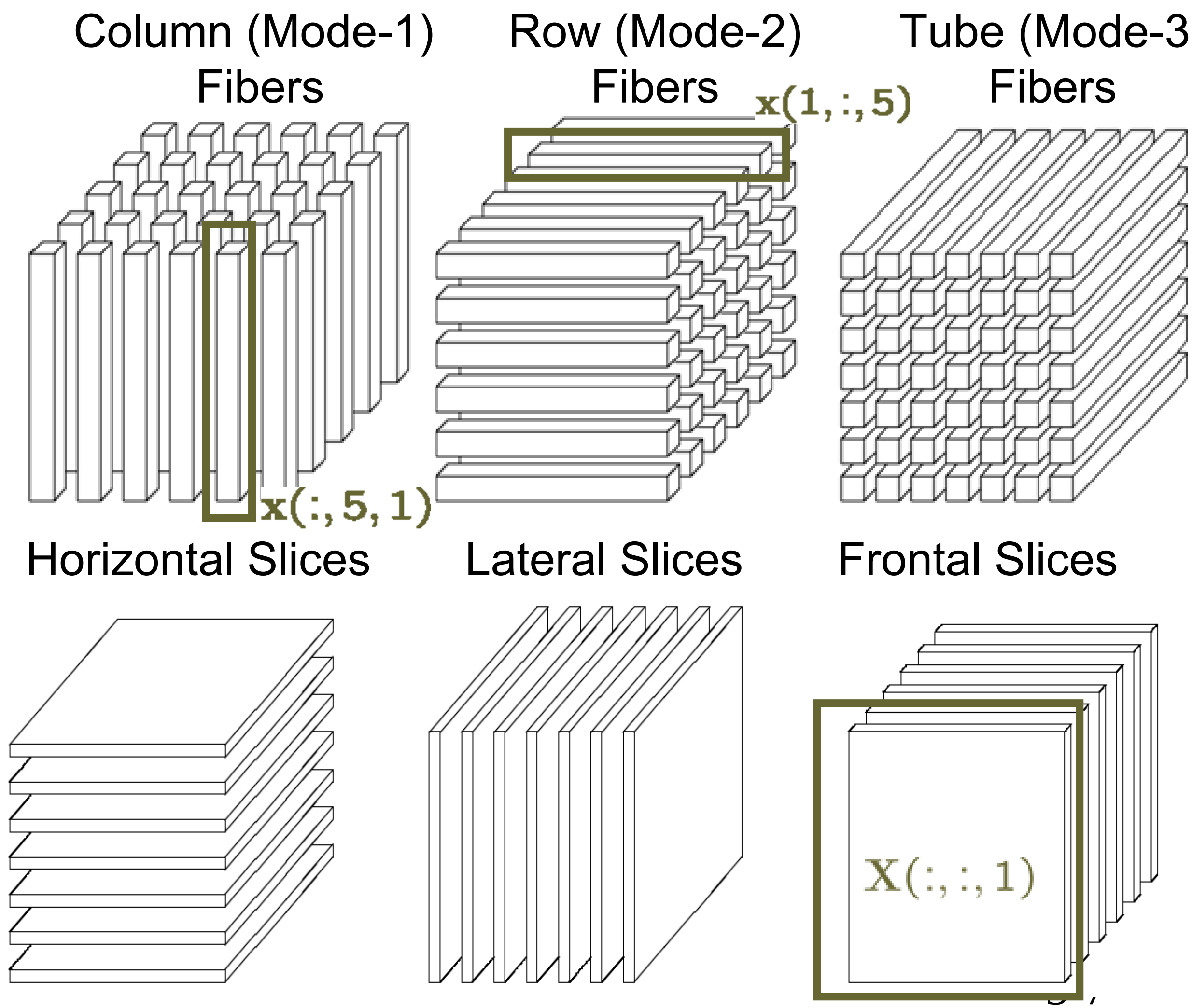
$$\mathcal{X} \approx [\![\lambda \,;\, \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] = \sum_r \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

Main points:

- 2 major types of tensor decompositions: PARAFAC and Tucker
- both can be solved with ``alternating least squares'' (ALS)
- Details follow – we start with terminology:

A tensor is a multidimensional array

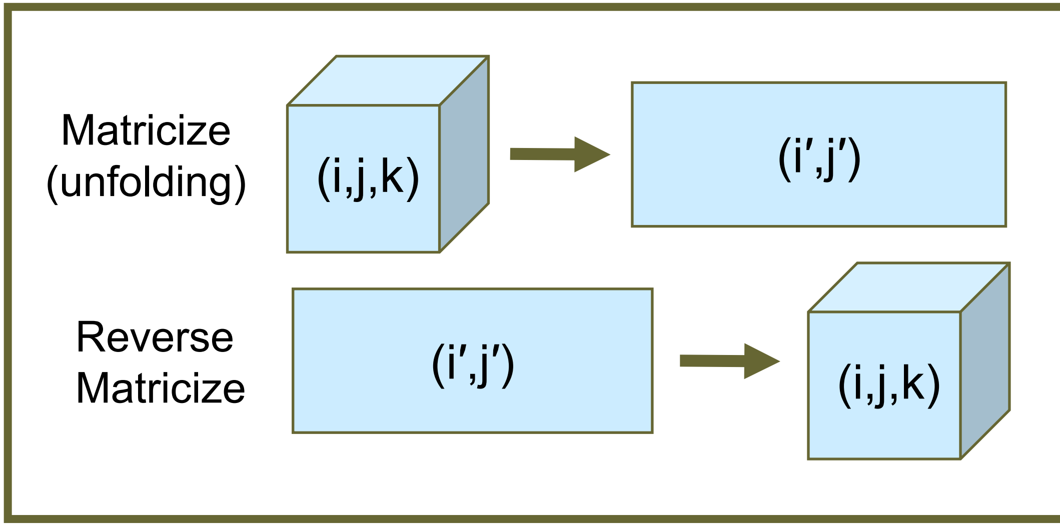




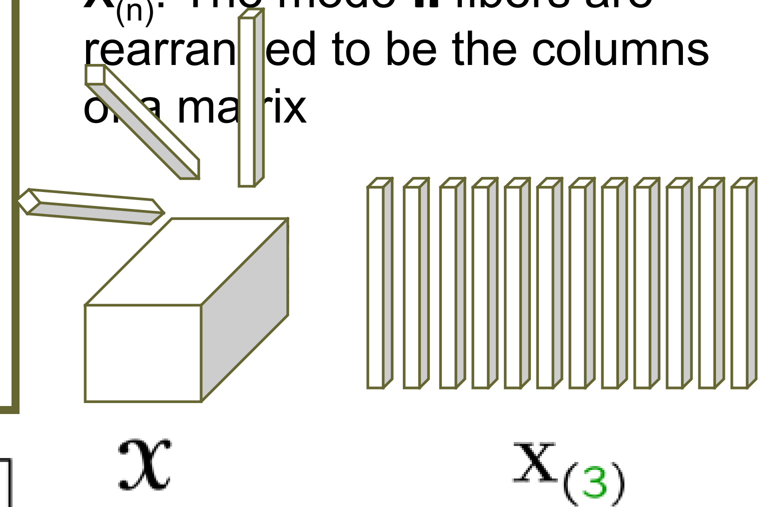
3rd order tensor
mode 1 has dimension I
mode 2 has dimension J
mode 3 has dimension K

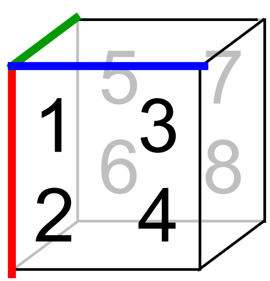
Note: Tutorial focus is on 3rd order, but everything can be extended to higher orders.

Matrization: Converting a Tensor to a Matrix



$\mathbf{X}_{(n)}$: The mode-**n** fibers are rearranged to be the columns of a matrix



$\mathcal{X} =$ 

$$\mathbf{X}_{(1)} = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{bmatrix}$$

$$\mathbf{X}_{(2)} = \begin{bmatrix} 1 & 2 & 5 & 6 \\ 3 & 4 & 7 & 8 \end{bmatrix}$$

$$\mathbf{X}_{(3)} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}$$

Vectorization

$$\mathrm{vec}(\mathcal{X}) = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{bmatrix}$$

Tensor Mode-n Multiplication

$$\mathcal{X} \in \mathbb{R}^{I \times J \times K}, \ \mathbf{B} \in \mathbb{R}^{M \times J}, \ \mathbf{a} \in \mathbb{R}^{I}$$

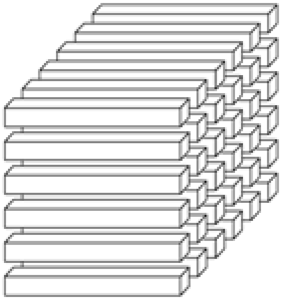
- Tensor Times Matrix

$$\mathcal{Y} = \mathcal{X} \times_2 \mathbf{B} \in \mathbb{R}^{I \times M \times K}$$

$$y_{imk} = \sum_j x_{ijk} \, b_{mj}$$

$$\mathbf{Y}_{(2)} = \mathbf{B}\mathbf{X}_{(2)}$$

Multiply each row (mode-2) fiber by **B**

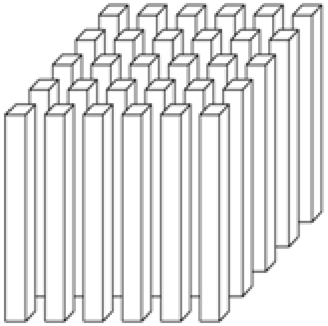


- Tensor Times Vector

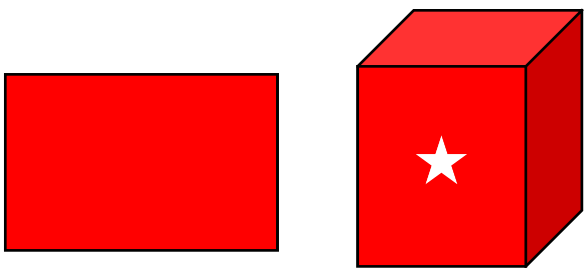
$$\mathbf{Y} = \mathcal{X} \ \bar{\times}_1 \ \mathbf{a} \in \mathbb{R}^{J \times K}$$

$$y_{jk} = \sum_i x_{ijk} \, a_i$$

Compute the dot product of a and each column (mode-1) fiber

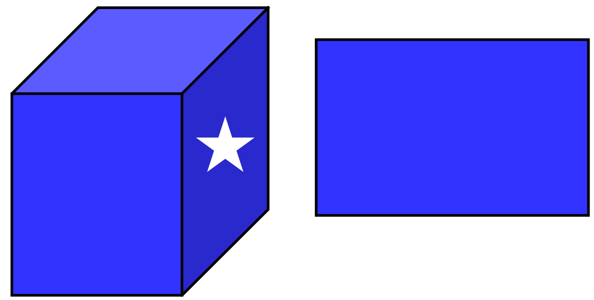


Pictorial View of Mode-n Matrix Multiplication



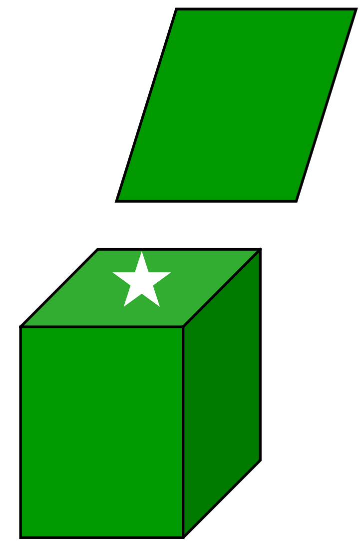
Mode-1 multiplication
(frontal slices)

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{A}$$
$$\mathbf{Y}_{::k} = \mathbf{X}_{::k}\mathbf{A}^{\mathsf{T}}$$



Mode-2 multiplication
(lateral slices)

$$\mathcal{Y} = \mathcal{X} \times_2 \mathbf{B}$$
$$\mathbf{Y}_{:j:} = \mathbf{X}_{:j:}\mathbf{B}^{\mathsf{T}}$$

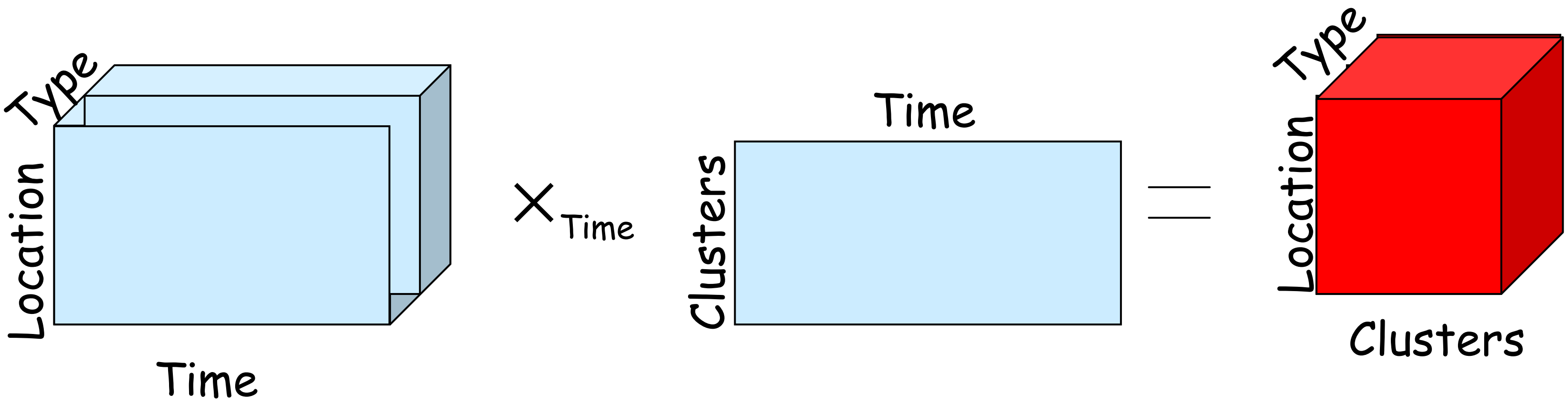


Mode-3 multiplication
(horizontal slices)

$$\mathcal{Y} = \mathcal{X} \times_3 \mathbf{C}$$
$$\mathbf{Y}_{i::} = \mathbf{X}_{i::}\mathbf{C}^{\mathsf{T}}$$

Mode-n product Example

- Tensor times a matrix

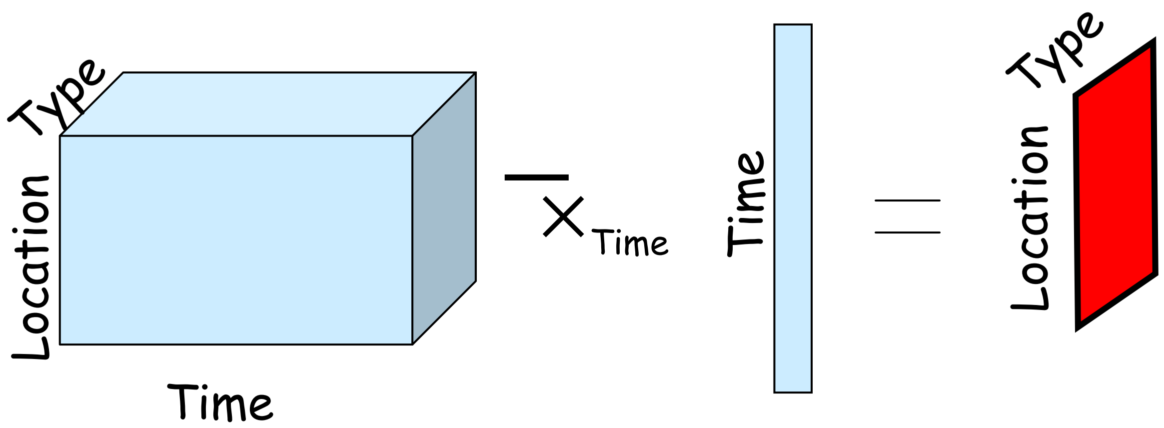


details

Sandia National Laboratories

Mode-n product Example

- Tensor times a vector



details

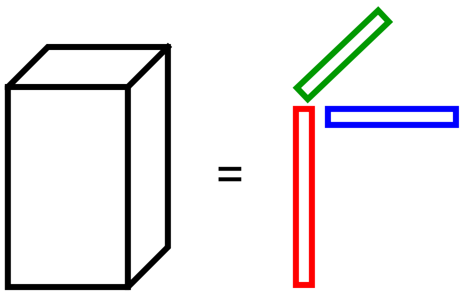
Sandia National Laboratories

Outer, Kronecker, & Khatri-Rao Products

3-Way Outer Product

$$\mathcal{X} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$$

$$x_{ijk} = a_i b_j c_k$$



Rank-1 Tensor

Review: Matrix Kronecker Product

$$\underset{M \times N}{\mathbf{A}} \otimes \underset{P \times Q}{\mathbf{B}} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1N}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2N}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1}\mathbf{B} & a_{M2}\mathbf{B} & \cdots & a_{MN}\mathbf{B} \end{bmatrix}$$

MP x NQ

$$= \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_1 \otimes \mathbf{b}_2 & \cdots & \mathbf{a}_N \otimes \mathbf{b}_Q \end{bmatrix}$$

Matrix Khatri-Rao Product

$$\underset{M \times R}{\mathbf{A}} \odot \underset{N \times R}{\mathbf{B}} = \underset{MN \times R}{\begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_2 \otimes \mathbf{b}_2 & \cdots & \mathbf{a}_R \otimes \mathbf{b}_R \end{bmatrix}}$$

<u>Observe</u>: For two vectors **a** and **b**, **a** $\circ$ **b** and **a** $\otimes$ **b** have the same elements, but one is shaped into a matrix and the other into a vector.

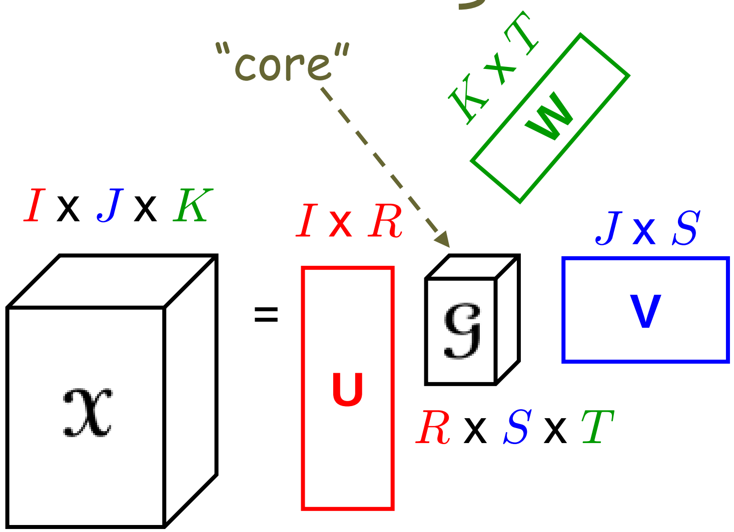
Specially Structured Tensors

Specially Structured Tensors

- Tucker Tensor

$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}$$
$$= \sum_r \sum_s \sum_t g_{rst}\, \mathbf{u}_r \circ \mathbf{v}_s \circ \mathbf{w}_t$$
$$\equiv [\![\mathcal{G} \,; \mathbf{U}, \mathbf{V}, \mathbf{W}]\!]$$

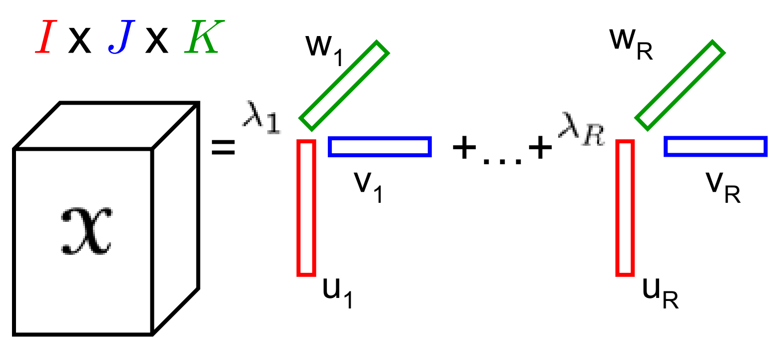
} Our Notation



- Kruskal Tensor

$$\mathcal{X} = \sum_r \lambda_r\, \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r$$
$$\equiv [\![\lambda \,; \mathbf{U}, \mathbf{V}, \mathbf{W}]\!]$$

} Our Notation



details

Sandia National Laboratories

Specially Structured Tensors

- Tucker Tensor

$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}$$
$$= \sum_r \sum_s \sum_t g_{rst}\, \mathbf{u}_r \circ \mathbf{v}_s \circ \mathbf{w}_t$$
$$\equiv [\![\mathcal{G} \,;\, \mathbf{U}, \mathbf{V}, \mathbf{W}]\!]$$

In matrix form:

$$\mathbf{X}_{(1)} = \mathbf{U}\mathbf{G}_{(1)}(\mathbf{W} \otimes \mathbf{V})^{\mathsf{T}}$$
$$\mathbf{X}_{(2)} = \mathbf{V}\mathbf{G}_{(2)}(\mathbf{W} \otimes \mathbf{U})^{\mathsf{T}}$$
$$\mathbf{X}_{(3)} = \mathbf{W}\mathbf{G}_{(3)}(\mathbf{V} \otimes \mathbf{U})^{\mathsf{T}}$$

$$\mathrm{vec}(\mathcal{X}) = (\mathbf{W} \otimes \mathbf{V} \otimes \mathbf{U})\mathrm{vec}(\mathcal{G})$$

- Kruskal Tensor

$$\mathcal{X} = \sum_r \lambda_r\, \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r$$
$$\equiv [\![\lambda \,;\, \mathbf{U}, \mathbf{V}, \mathbf{W}]\!]$$

In matrix form:

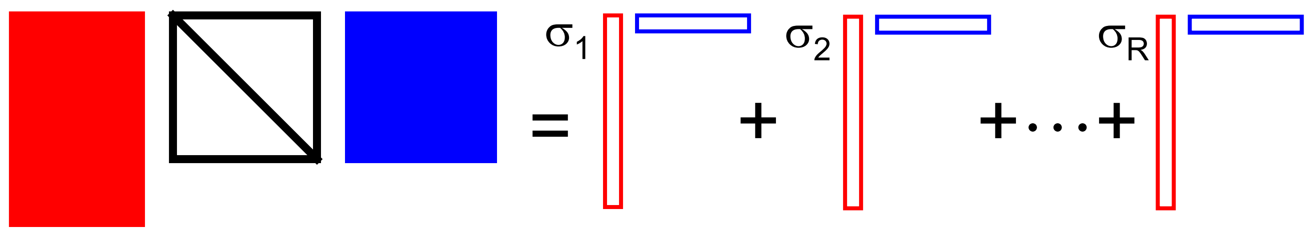
$$\text{Let } \mathbf{\Lambda} = \mathrm{diag}(\lambda)$$
$$\mathbf{X}_{(1)} = \mathbf{U}\,\mathbf{\Lambda}\,(\mathbf{W} \odot \mathbf{V})^{\mathsf{T}}$$
$$\mathbf{X}_{(2)} = \mathbf{V}\,\mathbf{\Lambda}\,(\mathbf{W} \odot \mathbf{U})^{\mathsf{T}}$$
$$\mathbf{X}_{(3)} = \mathbf{W}\,\mathbf{\Lambda}\,(\mathbf{V} \odot \mathbf{U})^{\mathsf{T}}$$

$$\mathrm{vec}(\mathcal{X}) = (\mathbf{W} \odot \mathbf{V} \odot \mathbf{U})\,\lambda$$

What is the HO Analogue of the Matrix SVD?

Matrix SVD:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathsf{T}} =$$



Tucker Tensor (finding bases for each subspace):

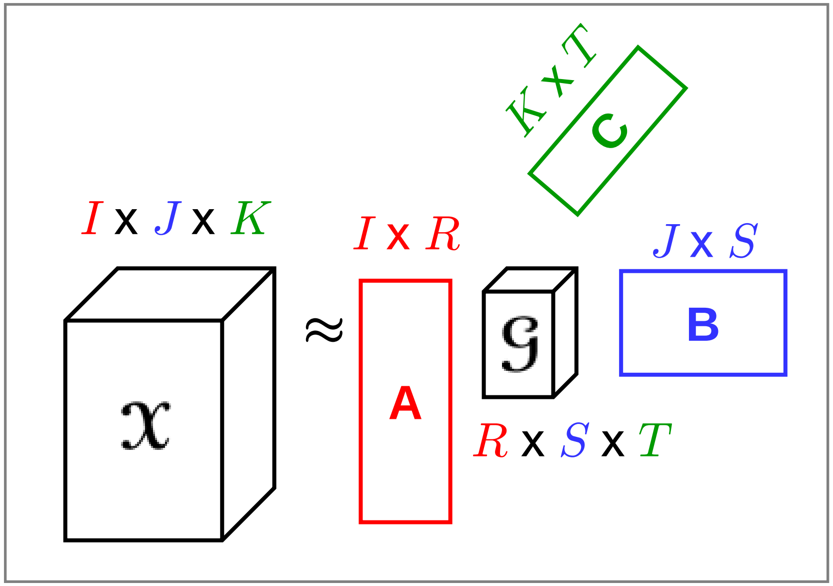
$$\mathbf{X} = \mathbf{\Sigma} \times_1 \mathbf{U} \times_2 \mathbf{V} = [\![\mathbf{\Sigma} \,; \mathbf{U}, \mathbf{V}]\!]$$

Kruskal Tensor (sum of rank-1 components):

$$\mathbf{X} = \sum_{r=1}^{R} \sigma_r \, \mathbf{u}_r \circ \mathbf{v}_r = [\![\sigma \,; \mathbf{U}, \mathbf{V}]\!]$$

Tensor Decompositions

Tucker Decomposition - intuition



- author x keyword x conference
- A: author x author-group
- B: keyword x keyword-group
- C: conf. x conf-group
- $\mathcal{G}$: how groups relate to each other

Reminder

$$\begin{bmatrix} .05 & .05 & .05 & 0 & 0 & 0 \\ .05 & .05 & .05 & 0 & 0 & 0 \\ 0 & 0 & 0 & .05 & .05 & .05 \\ 0 & 0 & 0 & .05 & .05 & .05 \\ .04 & .04 & 0 & .04 & .04 & .04 \\ .04 & .04 & .04 & 0 & .04 & .04 \end{bmatrix}$$

med. terms

cs terms

common terms

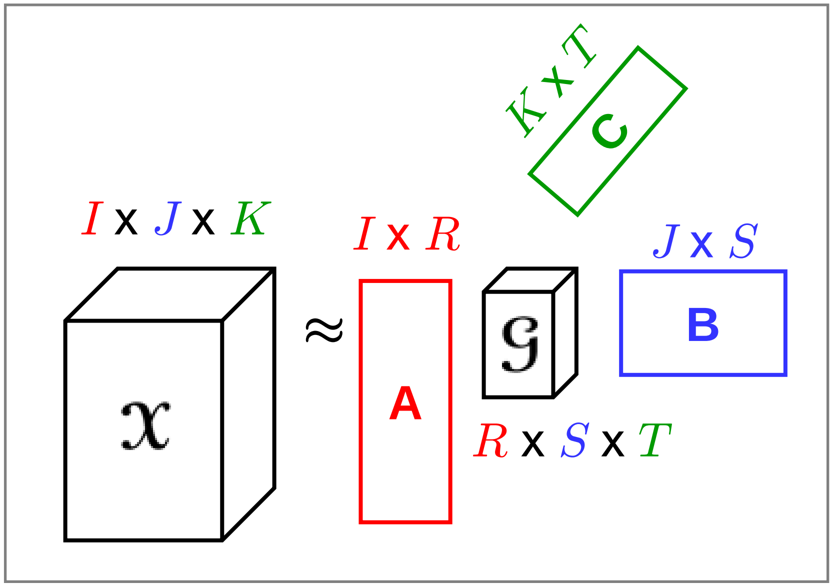
term group x doc. group

$$\begin{bmatrix} .5 & 0 & 0 \\ .5 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & .5 \\ 0 & 0 & .5 \end{bmatrix} \begin{bmatrix} .3 & 0 \\ 0 & .3 \\ .2 & .2 \end{bmatrix} \begin{bmatrix} .36 & .36 & .28 & 0 & 0 & 0 \\ 0 & 0 & 0 & .28 & .36 & .36 \end{bmatrix} = \begin{bmatrix} .054 & .054 & .042 & 0 & 0 & 0 \\ .054 & .054 & .042 & 0 & 0 & 0 \\ 0 & 0 & 0 & .042 & .054 & .054 \\ 0 & 0 & 0 & .042 & .054 & .054 \\ .036 & .036 & 028 & .028 & .036 & .036 \\ .036 & .036 & .028 & .028 & .036 & .036 \end{bmatrix}$$

term x term-group

doc x doc group

Tucker Decomposition



$$\mathcal{X} \approx [\![\mathcal{G} \, ; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

Given A, B, C, the optimal core is:

$$\mathcal{G} = [\![\mathcal{X} \, ; \mathbf{A}^{\dagger}, \mathbf{B}^{\dagger}, \mathbf{C}^{\dagger}]\!]$$

- Proposed by Tucker (1966)
- AKA: Three-mode factor analysis, three-mode PCA, orthogonal array decomposition
- **A**, **B**, and **C** generally assumed to be orthonormal (generally assume they have full column rank)
- $\mathcal{G}$ is <u>not</u> diagonal
- Not unique

Recall the equations for converting a tensor to a matrix

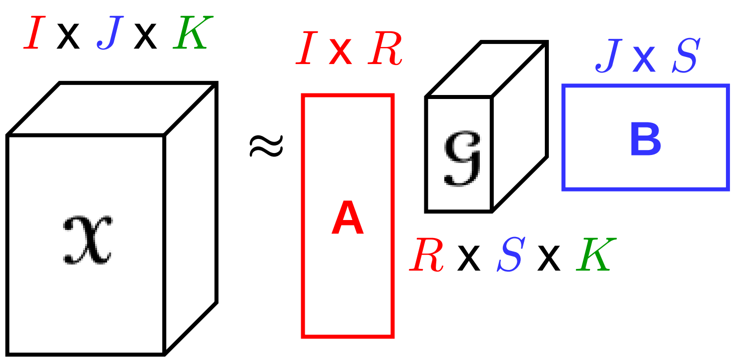
$$\mathbf{X}_{(1)} = \mathbf{A}\mathbf{G}_{(1)}(\mathbf{C} \otimes \mathbf{B})^{\mathsf{T}}$$
$$\mathbf{X}_{(2)} = \mathbf{B}\mathbf{G}_{(2)}(\mathbf{C} \otimes \mathbf{A})^{\mathsf{T}}$$
$$\mathbf{X}_{(3)} = \mathbf{C}\mathbf{G}_{(3)}(\mathbf{B} \otimes \mathbf{A})^{\mathsf{T}}$$
$$\text{vec}(\mathcal{X}) = (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})\text{vec}(\mathcal{G})$$

details

Tucker Variations

See Kroonenberg & De Leeuw, Psychometrika,1980 for discussion.

- Tucker2

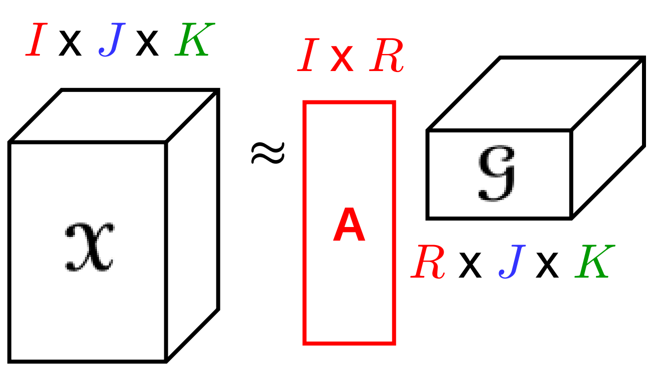


Identity Matrix

$$\mathcal{X} \approx [\![\mathcal{G} \,; \mathbf{A}, \mathbf{B}, \mathbf{I}]\!]$$

$$\mathbf{X}_{(3)} \approx \mathbf{G}_{(3)}(\mathbf{B} \otimes \mathbf{A})^{\mathsf{T}}$$

- Tucker1



$$\mathcal{X} \approx [\![\mathcal{G} \,; \mathbf{A}, \mathbf{I}, \mathbf{I}]\!]$$

$$\mathbf{X}_{(1)} \approx \mathbf{A}\mathbf{G}_{(1)}$$

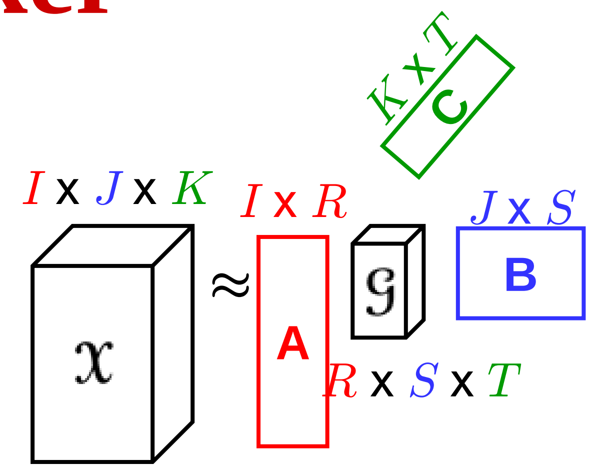
Finding principal components in only mode 1 can be solved via rank-R matrix SVD

Solving for Tucker

$$\mathcal{X} \approx [\![\mathcal{G} \,;\, \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

Given A, B, C orthonormal, the optimal core is:

$$\mathcal{G} = [\![\mathcal{X} \,;\, \mathbf{A}^\mathsf{T}, \mathbf{B}^\mathsf{T}, \mathbf{C}^\mathsf{T}]\!]$$



Tensor norm is the square root of the sum of all the elements squared

Eliminate the core to get:

$$\| \mathcal{X} - [\![\mathcal{G} \,;\, \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2 = \| \mathcal{X} \|^2 - 2 \langle \mathcal{X}, [\![\mathcal{G} \,;\, \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \rangle + \| \mathcal{G} \|^2$$

$$= \| \mathcal{X} \|^2 - \left\| [\![\mathcal{X} \,;\, \mathbf{A}^\mathsf{T}, \mathbf{B}^\mathsf{T}, \mathbf{C}^\mathsf{T}]\!] \right\|^2$$

Minimize s.t. **A,B,C** orthonormal

fixed

maximize this

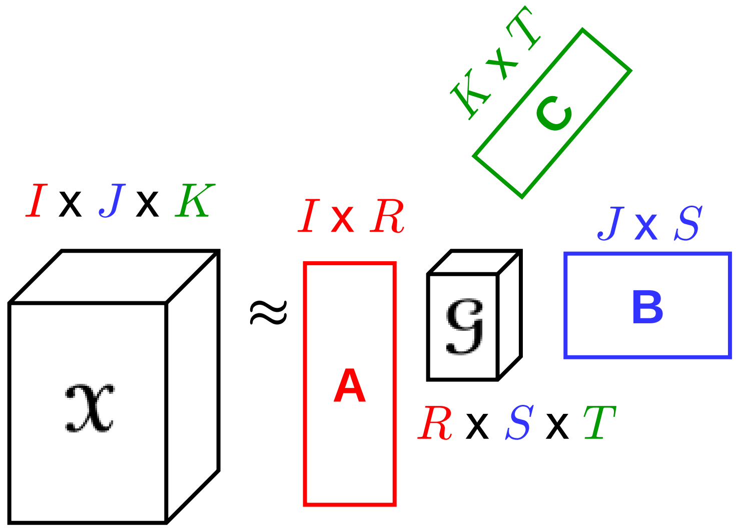
If B & C are fixed, then we can solve for A as follows:

$$\left\| [\![\mathcal{X} \,;\, \mathbf{A}^\mathsf{T}, \mathbf{B}^\mathsf{T}, \mathbf{C}^\mathsf{T}]\!] \right\| = \left\| \mathbf{A}^\mathsf{T} \mathbf{X}_{(1)} (\mathbf{C} \otimes \mathbf{B}) \right\|$$

Optimal **A** is R left leading singular vectors for $\mathbf{X}_{(1)} (\mathbf{C} \otimes \mathbf{B})$

details

Higher Order SVD (HO-SVD)



Not optimal, but often used to initialize Tucker-ALS algorithm.

(Observe connection to Tucker1)

$\mathbf{A}$ = leading R left singular vectors of $\mathbf{X}_{(1)}$

$\mathbf{B}$ = leading S left singular vectors of $\mathbf{X}_{(2)}$

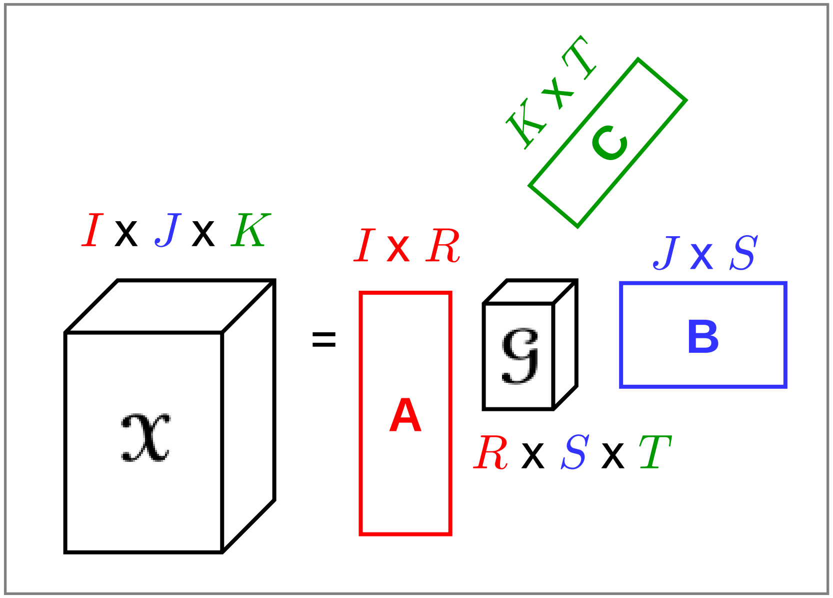
$\mathbf{C}$ = leading T left singular vectors of $\mathbf{X}_{(3)}$

$$\mathcal{G} = [\![\mathcal{X} \, ; \mathbf{A}^{\mathsf{T}}, \mathbf{B}^{\mathsf{T}}, \mathbf{C}^{\mathsf{T}}]\!]$$

De Lathauwer, De Moor, & Vandewalle, SIMAX, 1980

Tucker-Alternating Least Squares (ALS)

*Successively solve for each component (**A**,**B**,**C**).*

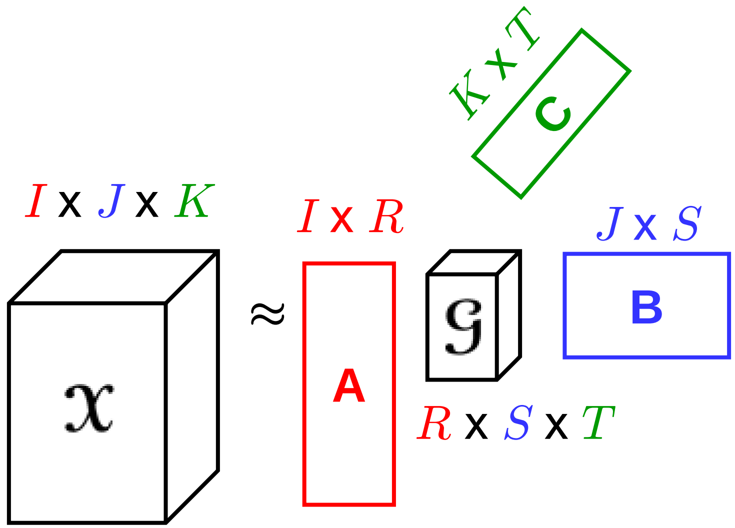


- Initialize
  - Choose R, S, T
  - Calculate **A**, **B**, **C** via HO-SVD
- Until converged do…
  - **A** = R leading left singular vectors of $\mathbf{X}_{(1)}(\mathbf{C}\otimes\mathbf{B})$
  - **B** = S leading left singular vectors of $\mathbf{X}_{(2)}(\mathbf{C}\otimes\mathbf{A})$
  - **C** = T leading left singular vectors of $\mathbf{X}_{(3)}(\mathbf{B}\otimes\mathbf{A})$
- Solve for core:

$$\mathcal{G} = [\![\mathcal{X}\,;\mathrm{A}^{\mathsf{T}},\mathrm{B}^{\mathsf{T}},\mathrm{C}^{\mathsf{T}}]\!]$$

Kroonenberg & De Leeuw, Psychometrika, 1980

Tucker in Not Unique



Tucker decomposition is *not* unique. Let Y be an RxR orthogonal matrix. Then…

$$\mathcal{X} \approx \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = \left(\mathcal{G} \times_1 \mathbf{Y}^\mathsf{T}\right) \times_1 (\mathbf{A}\mathbf{Y}) \times_2 \mathbf{B} \times_3 \mathbf{C}$$

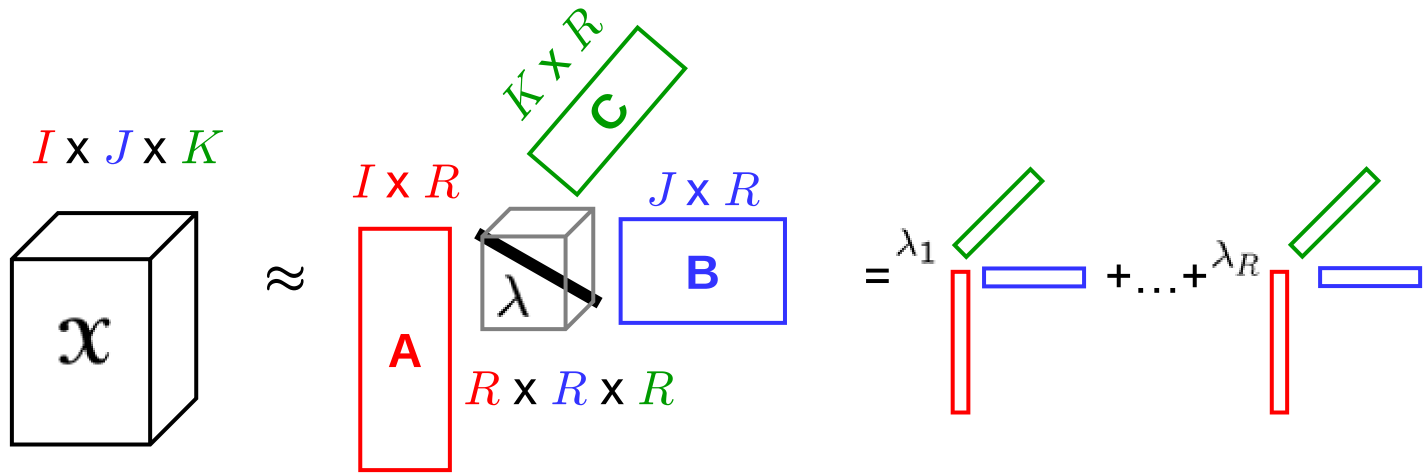
$$\mathbf{X}_{(1)} \approx \mathbf{A}\mathbf{G}_{(1)}(\mathbf{C} \otimes \mathbf{B})^\mathsf{T} = \mathbf{A}\mathbf{Y}\mathbf{Y}^\mathsf{T}\mathbf{G}_{(1)}(\mathbf{C} \otimes \mathbf{B})^\mathsf{T}$$

Roadmap

- Motivation
- Matrix tools
- Tensor tools
- Case studies

- Tensor Basics
- Tucker
  - Tucker 1
  - Tucker 2
  - Tucker 3
- PARAFAC
- Incrementalization

CANDECOMP/PARAFAC Decomposition



$$\mathfrak{X} \approx [\![\lambda \,;\, \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] = \sum_r \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

- CANDECOMP = Canonical Decomposition (Carroll & Chang, 1970)
- PARAFAC = Parallel Factors (Harshman, 1970)
- Core is <u>diagonal</u> (specified by the vector λ)
- Columns of **A**, **B**, and **C** are <u>not</u> orthonormal
- If R is *minimal*, then R is called the **rank** of the tensor (Kruskal 1977)
- Can have rank($\mathfrak{X}$) > min{I,J,K}

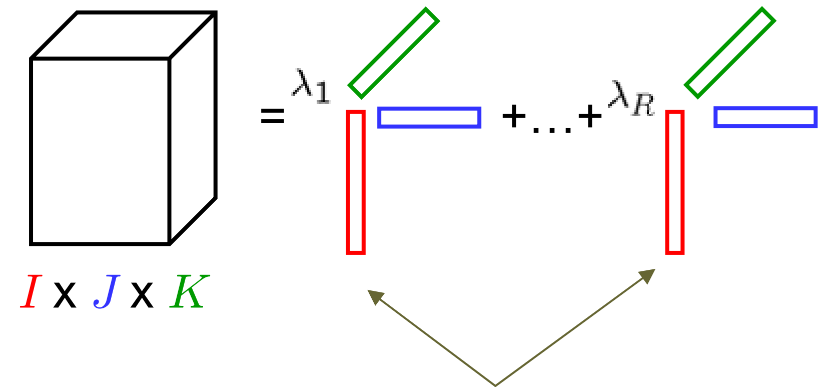
details

PARAFAC-Alternating Least Squares (ALS)

*Successively solve for each component (**A**,**B**,**C**).*

$$\mathcal{X} \approx [\![\lambda \,;\, \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

$$\mathbf{X}_{(1)} \approx \mathbf{A}\Lambda(\mathbf{C} \odot \mathbf{B})^{\mathsf{T}}$$



Find all the vectors in one mode at a time

KHATRI-RAO PRODUCT

(column-wise Kronecker product)

$$\mathbf{C} \odot \mathbf{B} \equiv \left[\mathbf{c}_1 \otimes \mathbf{b}_1 \quad \mathbf{c}_2 \otimes \mathbf{b}_2 \quad \cdots \mathbf{c}_R \otimes \mathbf{b}_R \right]$$

$$(\mathbf{C} \odot \mathbf{B})^{\dagger} \equiv (\mathbf{C}^{\mathsf{T}}\mathbf{C} * \mathbf{B}^{\mathsf{T}}\mathbf{B})^{\dagger}(\mathbf{C} \odot \mathbf{B})^{\mathsf{T}}$$

Hadamard Product

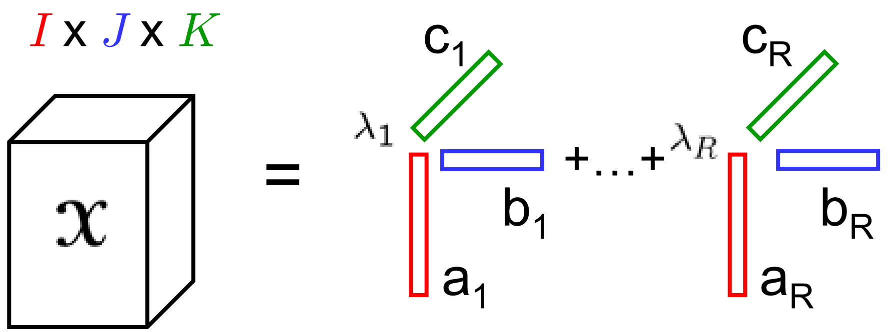
If **C**, **B**, and Λ are fixed, the optimal A is given by:

$$\mathbf{A} = \mathbf{X}_{(1)}(\mathbf{C} \odot \mathbf{B})(\mathbf{C}^{\mathsf{T}}\mathbf{C} * \mathbf{B}^{\mathsf{T}}\mathbf{B})^{\dagger}\Lambda^{-1}$$

*Repeat for **B**,**C**, etc.*

details

PARAFAC is often unique



Assume PARAFAC decomposition is exact.

$$\mathcal{X} = [\![\lambda \, ; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] = \sum_r \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

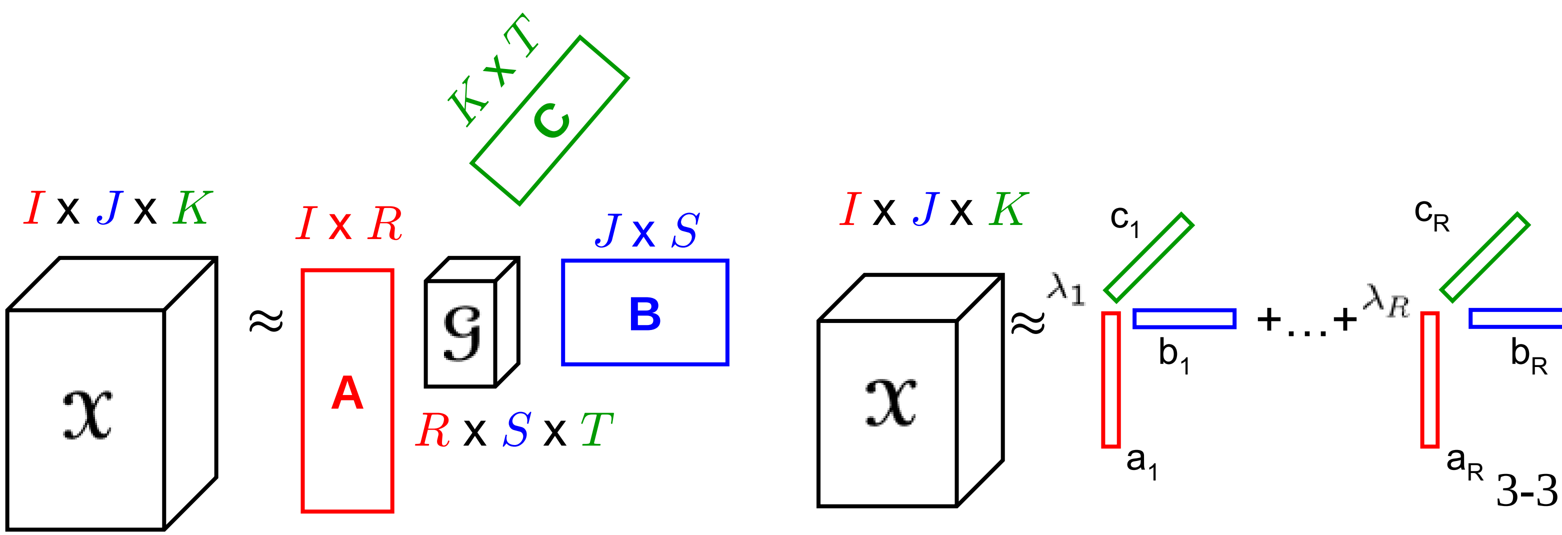
Sufficient condition for uniqueness (Kruskal, 1977):

$$2R + 2 \leq k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}}$$

$k_{\mathbf{A}}$ = k-rank of **A** = max number k such that every set of k columns of **A** is linearly independent

Tucker vs. PARAFAC Decompositions

- Tucker
  - Variable transformation in each mode
  - Core G may be dense
  - A, B, C generally orthonormal
  - Not unique
- PARAFAC
  - Sum of rank-1 components
  - No core, i.e., superdiagonal core
  - A, B, C may have linearly dependent columns
  - Generally unique



Tensor tools - summary

- Two main tools
  - PARAFAC
  - Tucker
- Both find row-, column-, tube-groups
  - but in PARAFAC the three groups are identical
- To solve: Alternating Least Squares

- Toolbox: from Tamara Kolda:

http://csmr.ca.sandia.gov/~tgkolda/TensorToolbox/