Tensor Factorization

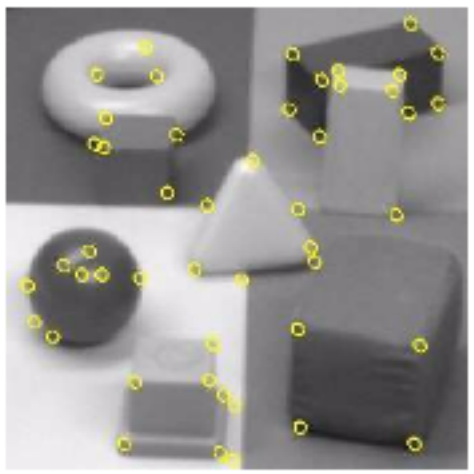
CS 584: Big Data Analytics

Tensor

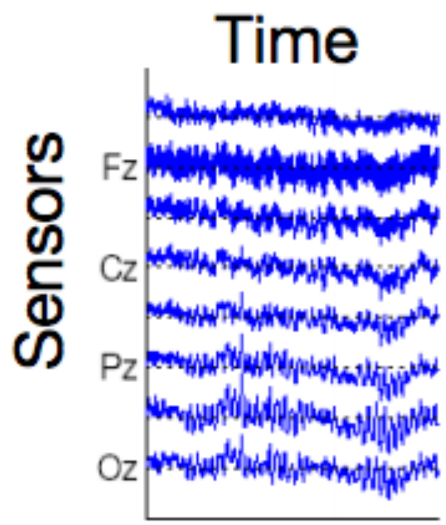
- *Definition: An element of the tensor product of N vector spaces*
- Generalization of scalars, vectors, and matrices to multidimensional array
- Representation of an n-way interaction
    - Matrix represents a binary relation
- Difficult to visualize beyond 3rd order tensors

Tensors are Everywhere

Matrices



Black and White



Multivariate time series

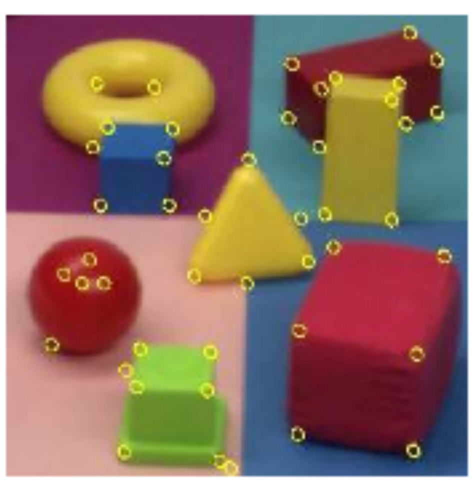
Movies

Users

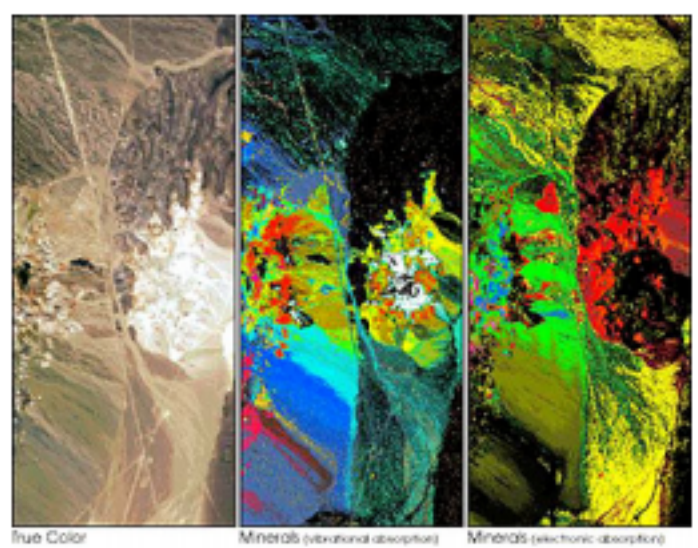
	Star Wars	Titanic	Blade Runner
User 1	5	2	4
User 2	1	4	2
User 3	5	?	?

Movie recommendation

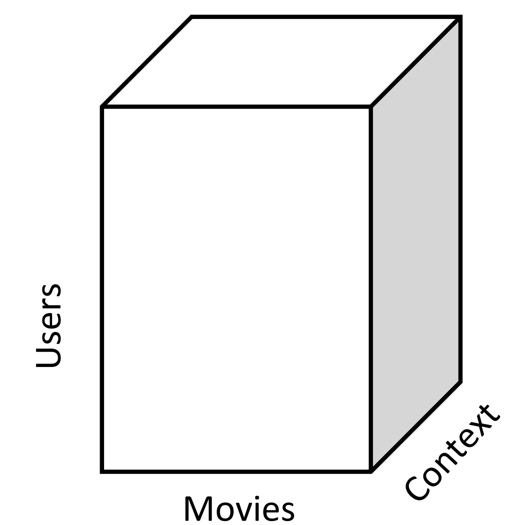
Tensors



Color



Spatio-temporal data



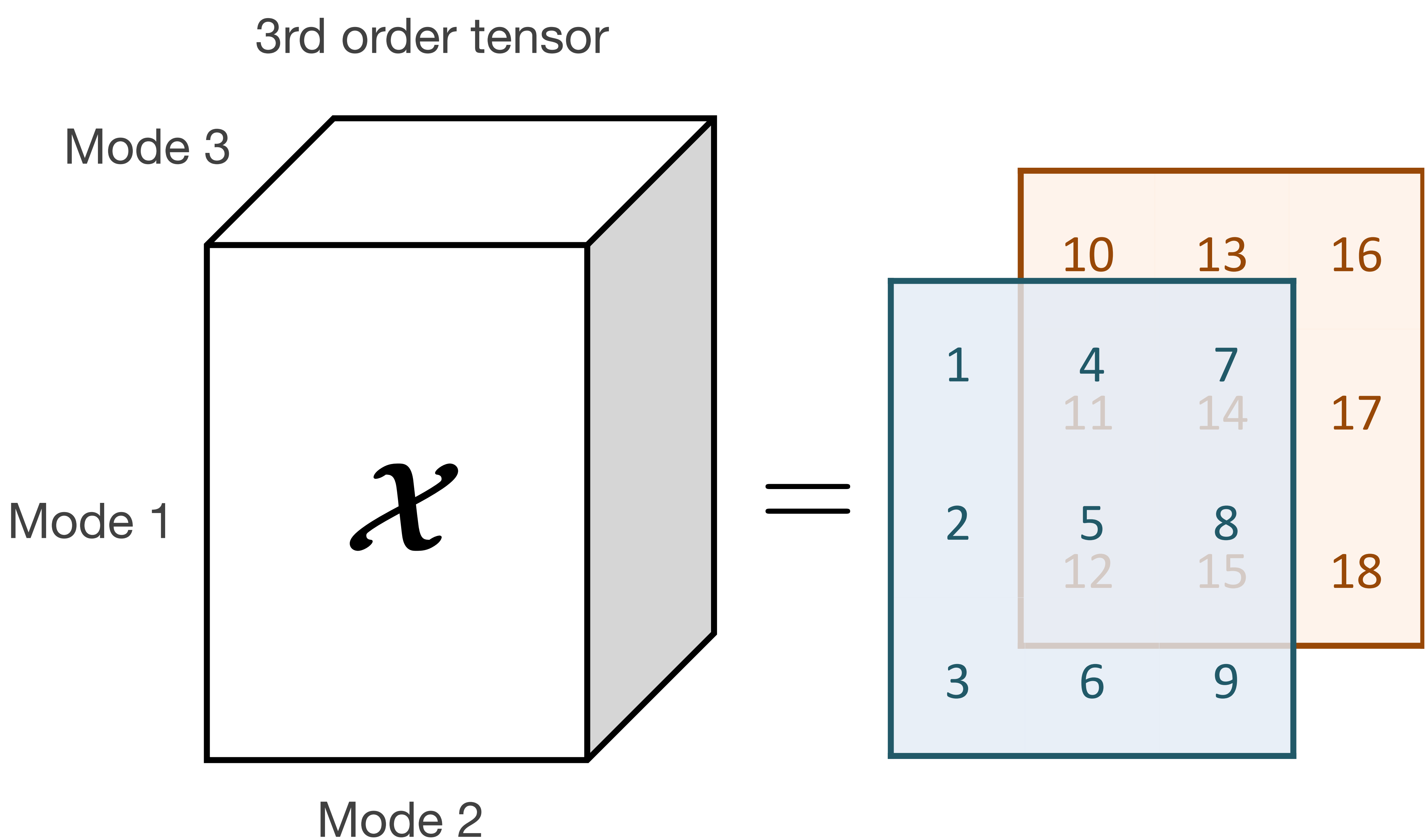
Multiple relations

Adapted from Ryoto Tomioka's Atalk @ MLSS 2015 talk on Tensor decompositions: old, new, and beyond

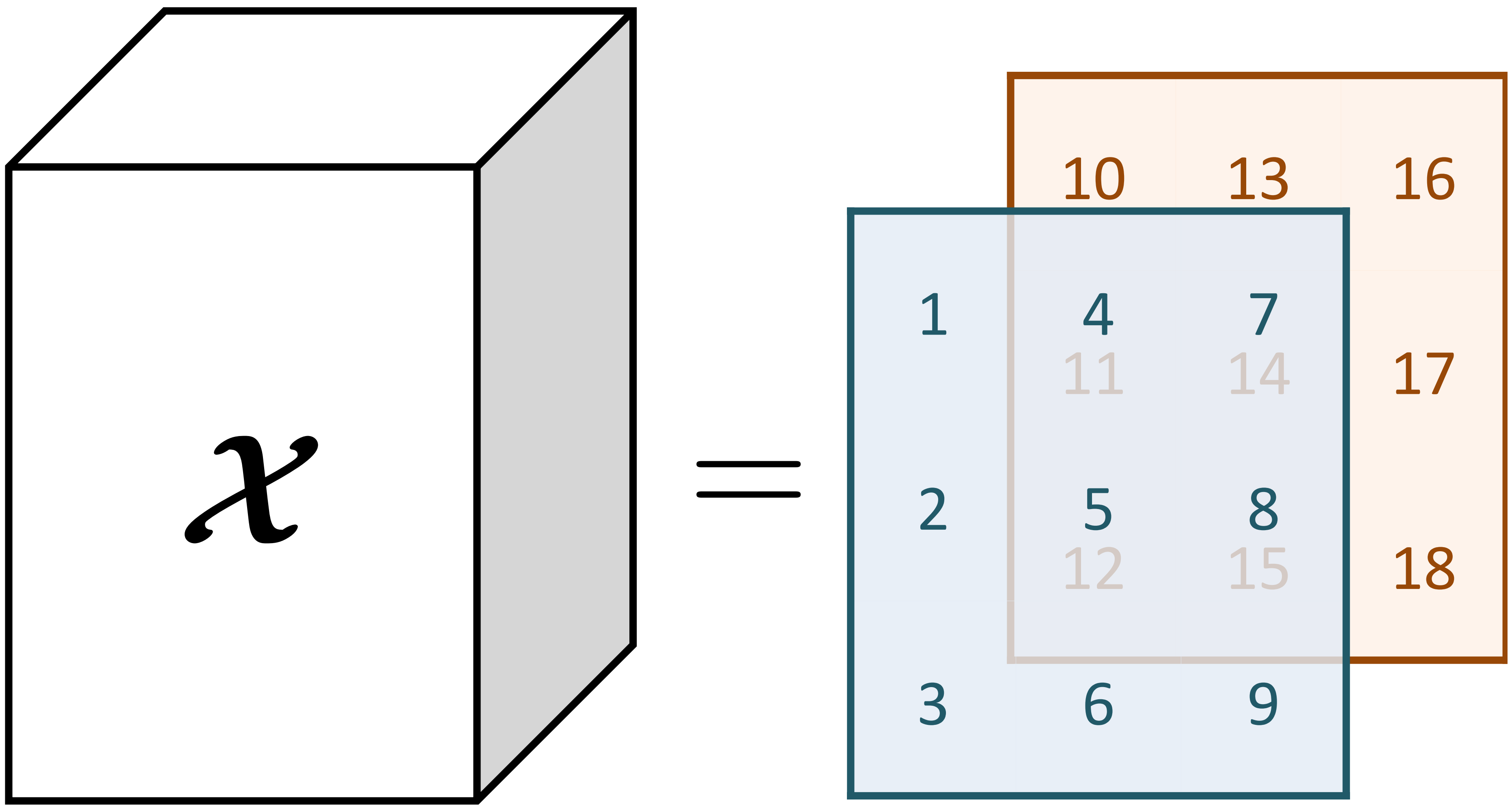
Tensor Terminology

Term	Meaning
Way, Mode	dimension, axis
Order	number of modes
Fiber	fix every index but one
Slice	fix all but two indices

Tensor Terminology



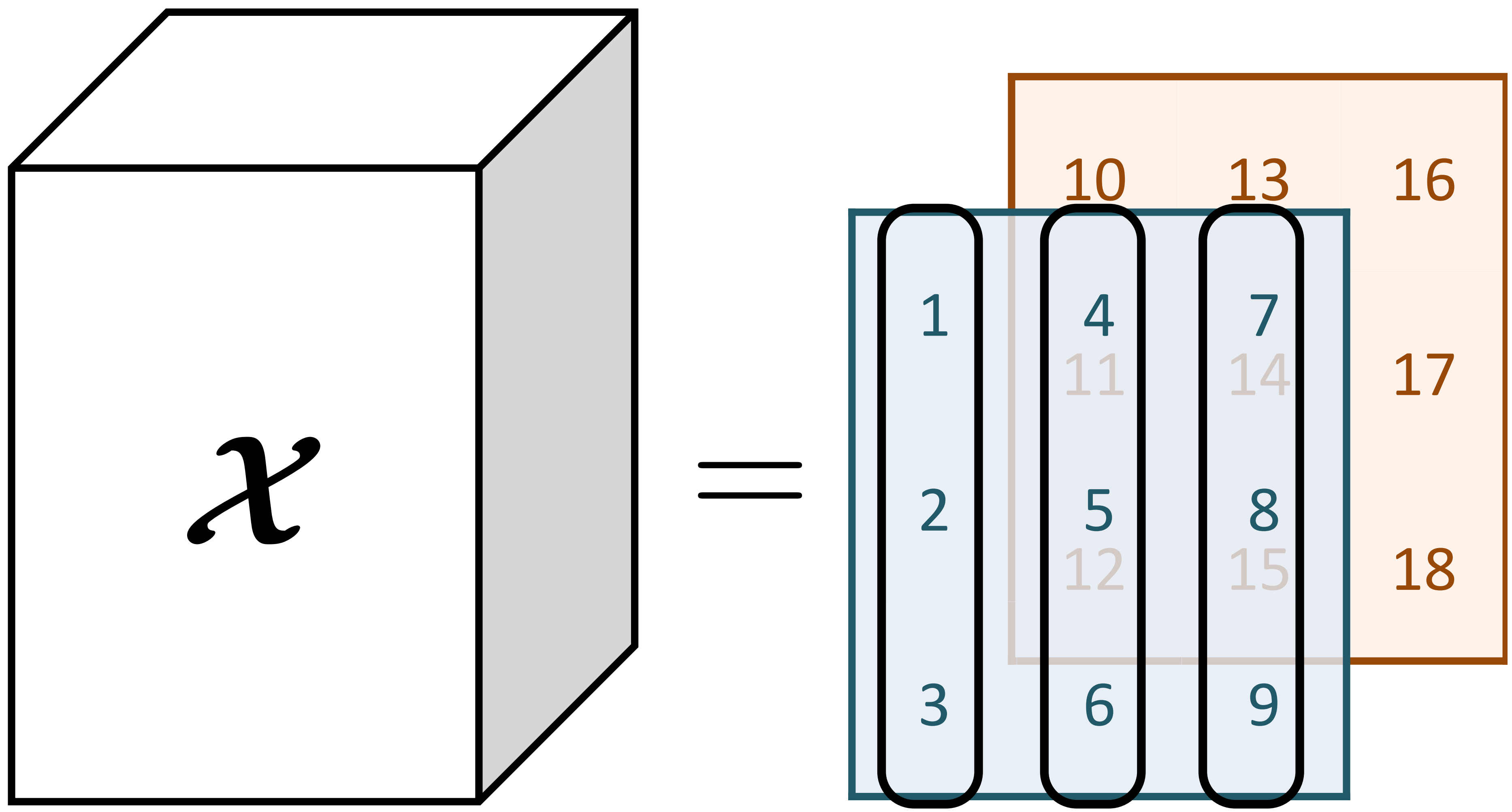
Tensor Terminology



Frontal Slices

$\mathbf{X}_{::k}$

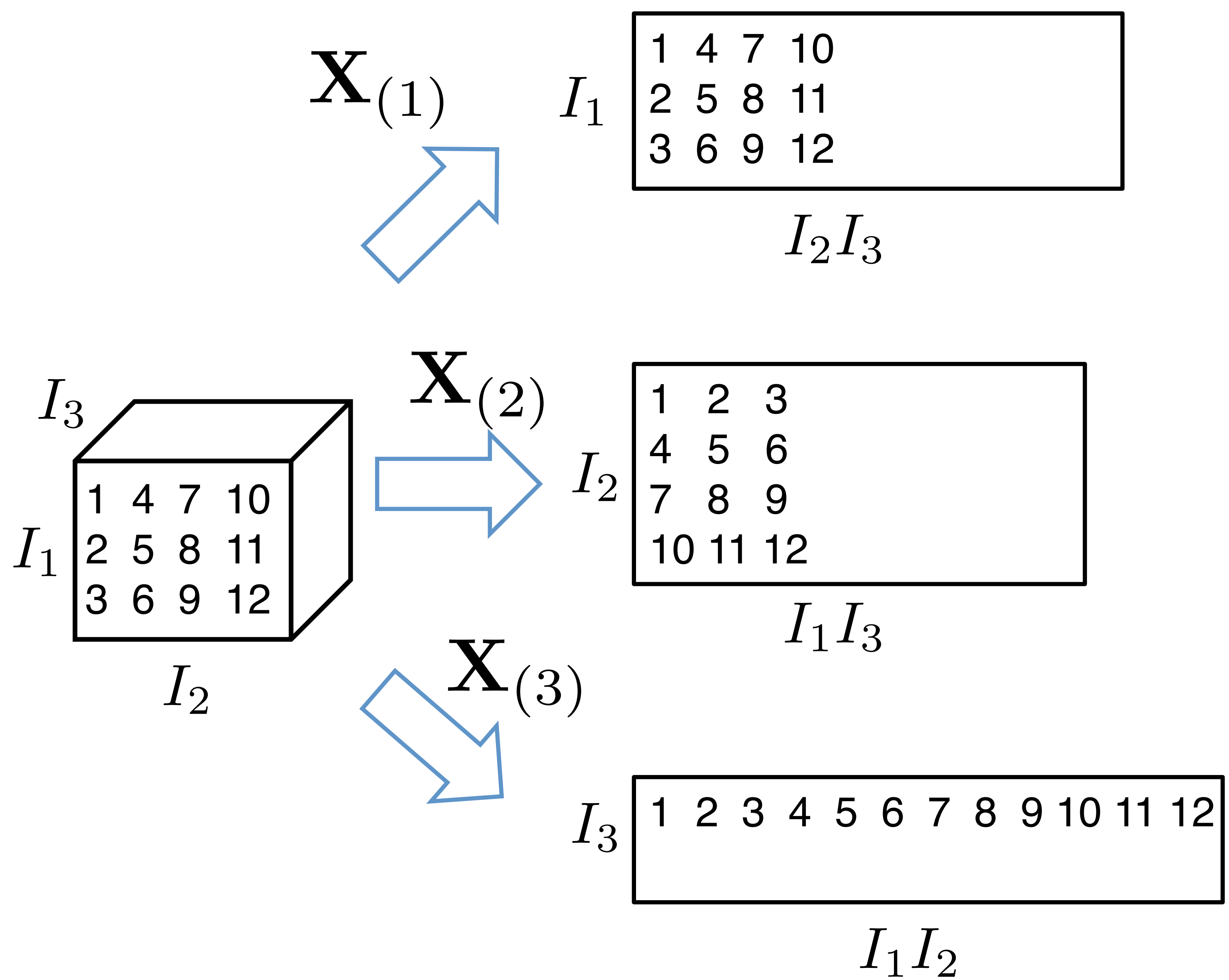
Tensor Terminology



Mode-1 fibers

$\mathbf{x}_{:jk}$

Matricization: Tensor Unfolding



Column ordering can be different across papers!

Useful Matrix Operations

- Kronecker product: generalization of outer product

$$\underbrace{P}_{\in \mathbf{R}^{m \times n}} \otimes \underbrace{Q}_{\in \mathbf{R}^{k \times l}} = \begin{bmatrix} p_{11}Q & \cdots & p_{1n}Q \\ \vdots & \ddots & \vdots \\ p_{m1}Q & \cdots & p_{mn}Q \end{bmatrix} \in \mathbf{R}^{mk \times nl}$$

- Khatri-Rao product: column-wise Kronecker product

$$P = \begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix}, Q = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix}$$

$$P \odot Q = \begin{bmatrix} p_1 \otimes q_1 & \cdots & p_n \otimes q_n \end{bmatrix} \in \mathbf{R}^{mk \times n}$$

Useful Matrix Operations (2)

- Hadamard product: element-wise multiplication

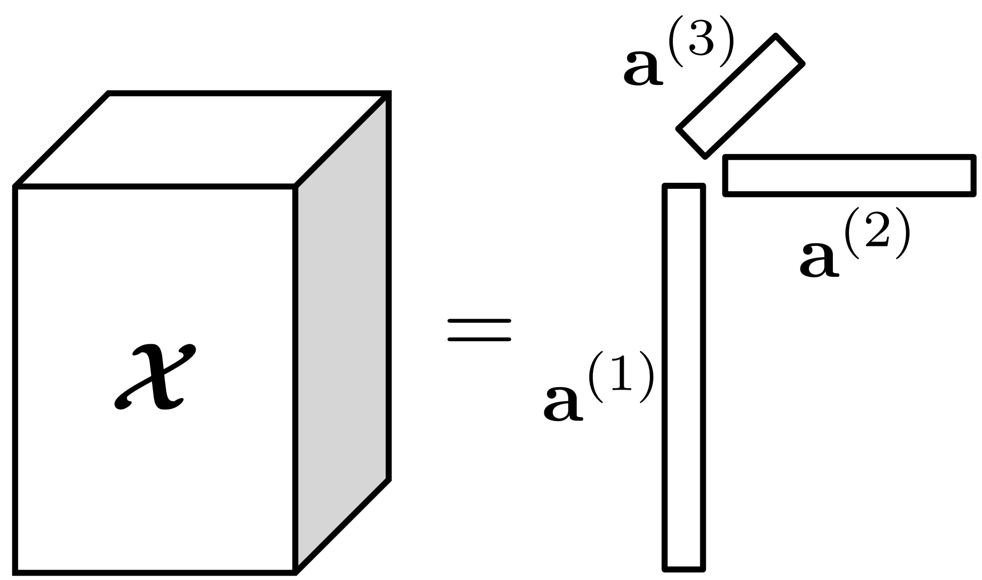
$$\underbrace{\mathbf{A}}_{\mathbf{R}^{I \times J}} * \underbrace{\mathbf{B}}_{\mathbf{R}^{I \times J}} = \begin{bmatrix} a_{11}b_{11} & \cdots & a_{1J}b_{1J} \\ \vdots & \ddots & \vdots \\ a_{I1}b_{I1} & \cdots & a_{IJ}b_{IJ} \end{bmatrix}$$

Rank-1 Tensor: Outer Product of N Vectors

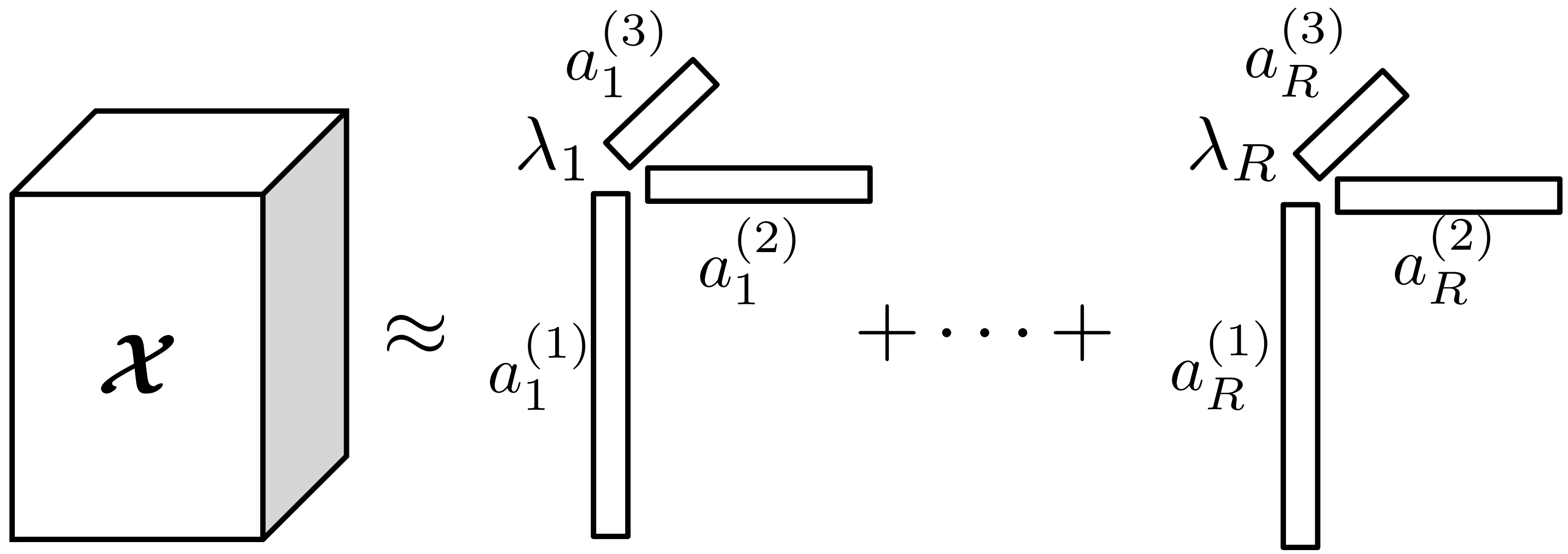
$$\boldsymbol{\mathcal{X}} = \mathbf{a}^{(1)} \circ \mathbf{a}^{(2)} \circ \cdots \circ \mathbf{a}^{(N)}$$

$$\Updownarrow$$

$$x_{i_1 i_2 \cdots i_N} = a^{(1)}_{i_1} a^{(2)}_{i_2} \cdots a^{(N)}_{i_N}$$



CP Decomposition



$$\mathcal{X} \approx \sum_{r=1}^{R} \lambda_r \mathbf{a}_r^{(1)} \circ \ldots \circ \mathbf{a}_r^{(N)}$$
$$= [\![\boldsymbol{\lambda}; \mathbf{A}^{(1)}; \ldots; \mathbf{A}^{(N)}]\!]$$

- Polyadic form (Hitchcock, 1927)
- CANDECOMP = Canonical Decomposition (Carroll and Chang, 1970)
- PARAFAC = Parallel Factors (Harshman, 1970)

Matricization of CP

$$\boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} a_r \circ b_r \circ c_r$$

$$\Updownarrow$$

$$\mathbf{X}_{(1)} = A\,(C \odot B)^{\mathsf{T}}$$
$$\mathbf{X}_{(2)} = B\,(C \odot A)^{\mathsf{T}}$$
$$\mathbf{X}_{(3)} = C\,(B \odot A)^{\mathsf{T}}$$

CP Property: Uniqueness

- Unique representation (unlike matrix)
  - Up to permutation of terms
  - Up to scaling of factors
  - Sufficient condition

Kruskal rank
max k such that k columns
are linearly independent

$$k_A + k_B + k_C \geq 2R + 2$$

- Problems:
  - Degenerate solutions (component loadings highly correlated in all modes)
  - Unstable and slow to converge

CP: Tensor Rank

- Smallest number of rank-one tensors that generates tensor X as their sum

- Not bounded by dimensions of tensor

- Computing rank is NP-hard problem: determined numerically by trying many rank-R models

- Weak upper-bound for general third order tensor

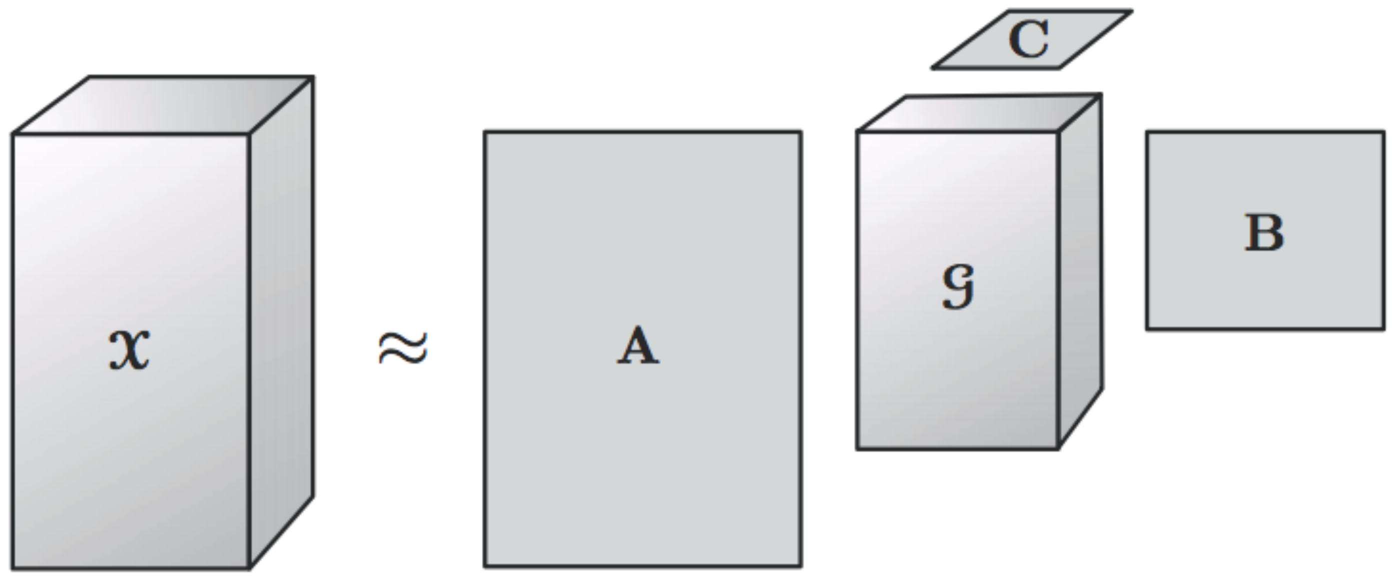
$$\text{rank}(\underbrace{\boldsymbol{\mathcal{X}}}_{\mathbf{R}^{I \times J \times K}}) \leq \min\{IJ, IK, JK\}$$

CP Decomposition Algorithm

$$\min ||\boldsymbol{\mathcal{X}} - [\![\boldsymbol{\lambda}; \mathbf{A}^{(1)}; \mathbf{A}^{(2)}; \cdots; \mathbf{A}^{(N)}]\!]||$$

- Many algorithms
  - Alternating least squares (Carroll and Chang, 1970; Harshman, 1970)
  - Nonlinear least squares (Paatero, 1997; Tomasi and Bro, 2005)
  - Nonlinear conjugate gradient (Acar, Kolda and Dunlavy)

Tucker Decomposition



$$\mathcal{X} \approx \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} \, \mathbf{a}_p \circ \mathbf{b}_q \circ \mathbf{c}_r = [\![\mathcal{G} \, ; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!].$$

- Tucker3 (Tucker, 1996)
- N-mode PCA (Kapteyn et al., 1986)
- Higher-order SVD (De Lathauwer et al., 2000)

Tucker Decomposition (2)

- Many degrees of freedom: A, B, and C are often required to be orthogonal
- CP decomposition is a special case of Tucker when P = Q = R and core tensor G is superdiagonal
- Decomposition is not unique
- n-rank of tensor X is the column rank of $X_{(n)}$ => X is a rank-($R1, R2, \ldots, R_N$) tensor
  - Different than idea of rank (minimum number of rank-one components)

Matricization of Tucker

$$\mathcal{X} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} a_p \circ b_q \circ c_r$$

$$\Updownarrow$$

$$\mathbf{X}_{(1)} \approx \mathbf{A}\mathbf{G}_{(1)}(\mathbf{C} \otimes \mathbf{B})^\top$$

$$\mathbf{X}_{(2)} \approx \mathbf{B}\mathbf{G}_{(2)}(\mathbf{C} \otimes \mathbf{A})^\top$$

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

Tucker Decomposition Algorithms

- Alternating least squares
  - Without orthogonality
  - With orthogonality
    - Higher-order SVD (De Lathauwer et al., 2000)
- Newton-Grassman optimization approach (Elden and Savas, 2009)

Other Decompositions

Model	Decomposition	Unique
INDSCAL	Imposing symmetry on two modes of the CP model	Yes
CANDELINC	CP with linear constraints	No
DEDICOM	Asymmetric relationships between two modes that index the same object	Yes
…	…	…

Only tip of the iceberg — many others available!

What are the advantages?

- Exploit structure for improved interpretability
- Robust to noise and missing data
  (CP recovers components even with 99% missing in 3rd order tensor)
- Uniqueness properties (for some decompositions)

What are good resources?

- Tensor decompositions and applications
  (Kolda and Bader, 2009)

- Applications of tensor (multiway array) factorization and decompositions in data mining
  (Morup, 2011)

- Multi-way Analysis: Applications in the Chemical Sciences (Smilde, Bro, and Geladi, 2004)

- PARAFAC: Tutorial and applications (Bro, 1997)

What packages can I use?

- MATLAB
  - Tensor Toolbox
    http://www.sandia.gov/~tgkolda/TensorToolbox/index-2.5.html
  - N-way Toolbox
    http://www.models.life.ku.dk/nwaytoolbox
- Python
  - sckit-tensor
    https://github.com/mnick/scikit-tensor/
  - pytensor
    https://code.google.com/p/pytensor/