Introduction to Tensors

8 May 2014

max planck institut
informatik

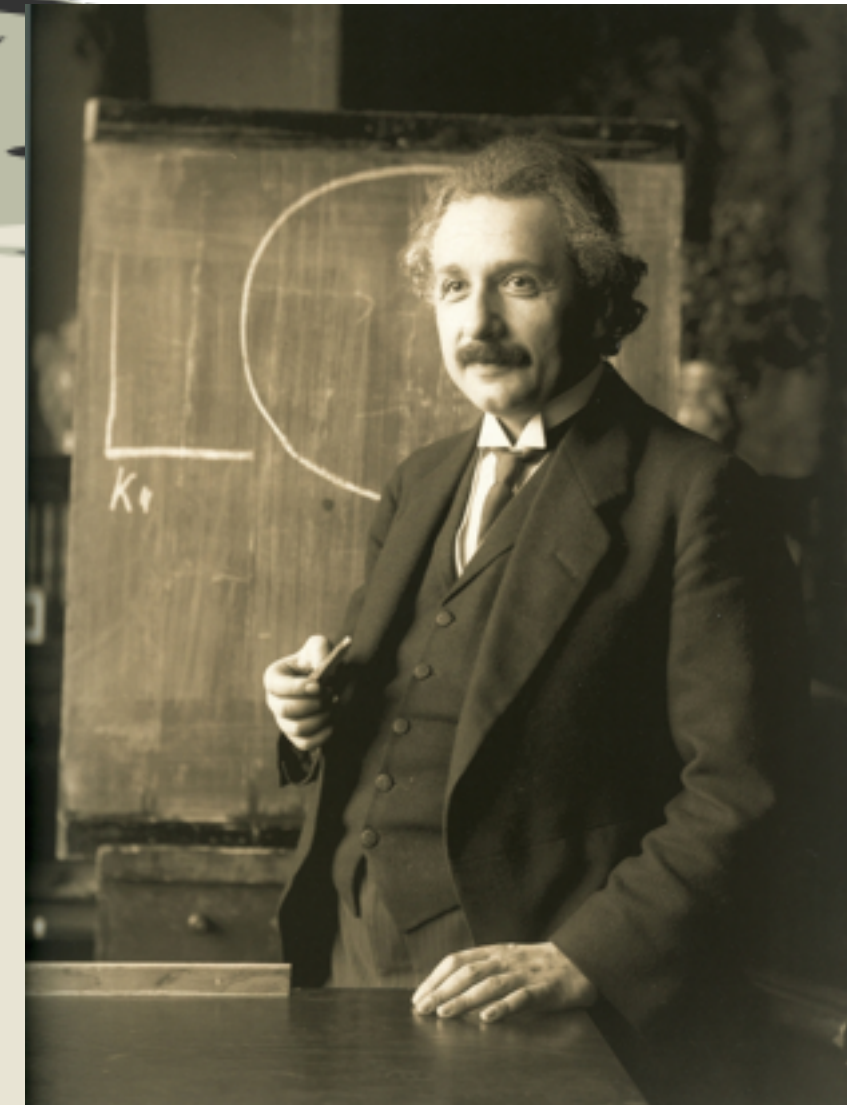
Introduction to Tensors

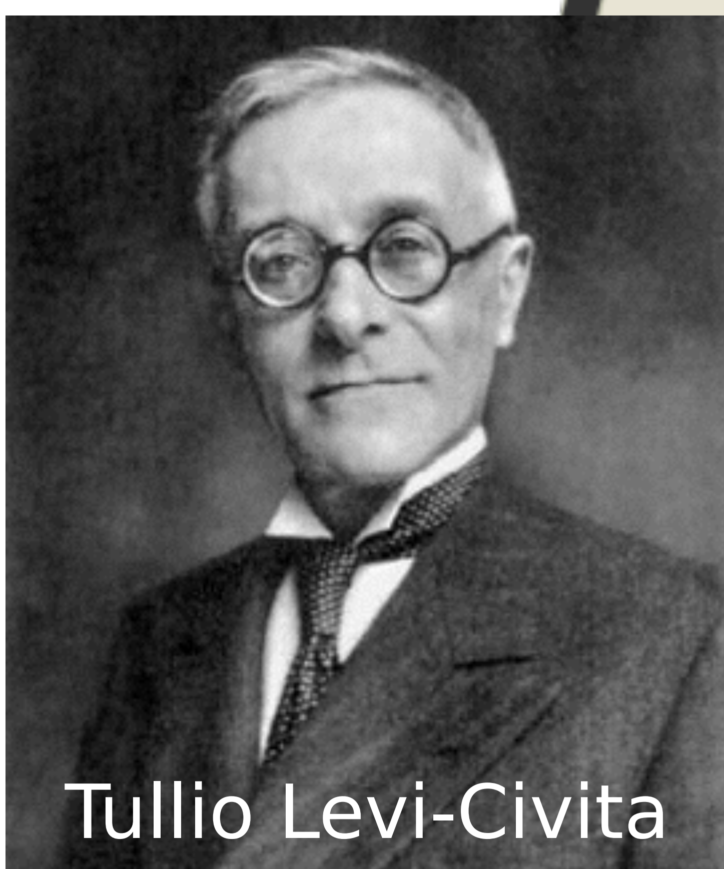
- What is a … tensor?
- Basic Operations
- CP Decompositions and Tensor Rank
- Matricization and Computing the CP

Dear Tullio,

I admire the elegance of your method of computation; it must be nice to ride through these fields upon the horse of true mathematics while the like of us have to make our way laboriously on foot.

Cheers, Albert

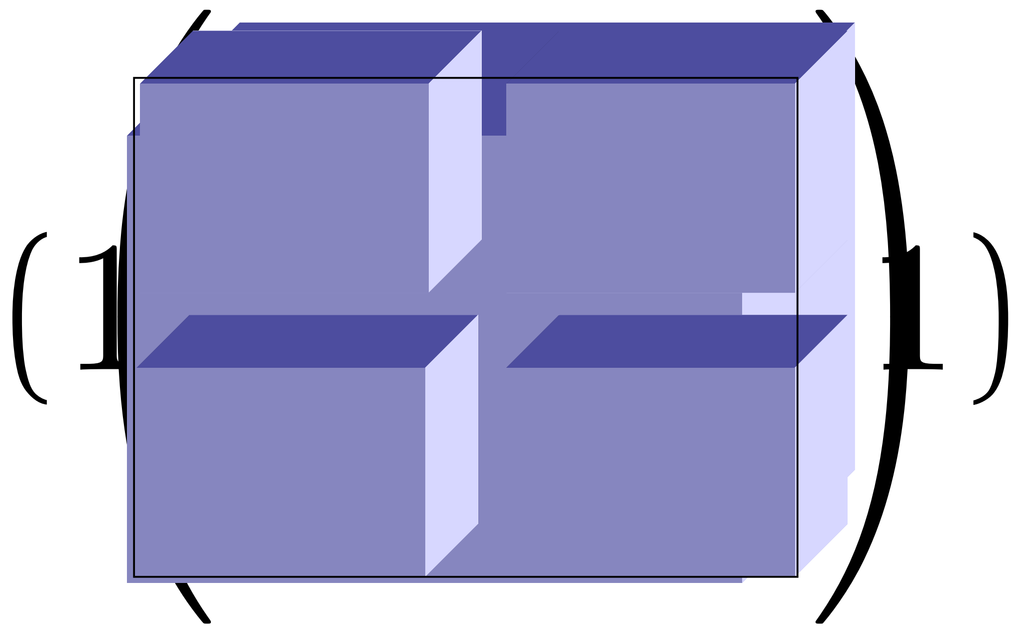




Tullio Levi-Civita

What is a … tensor?

- A tensor is a multi-way extension of a matrix
  - A multi-dimensional array
  - A multi-linear map
- In particular, the following are all tensors:
  - Scalars
  - Vectors
  - Matrices



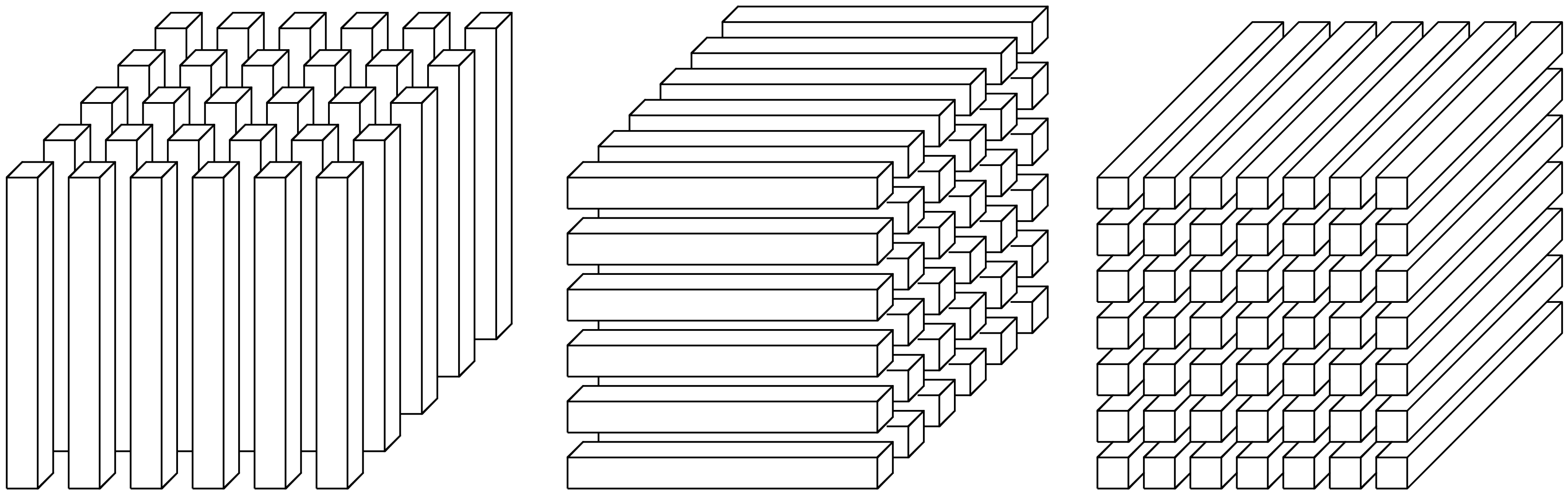
Why Tensors?

- Tensors can be used when matrices are not enough
- A matrix can represent a binary relation
  - A tensor can represent an n-ary relation
    - E.g. subject–predicate–object data
  - A tensor can represent a set of binary relations
    - Or other matrices
- A matrix can represent a matrix
  - A tensor can represent a series/set of matrices
  - But using tensors for time series should be approached with care

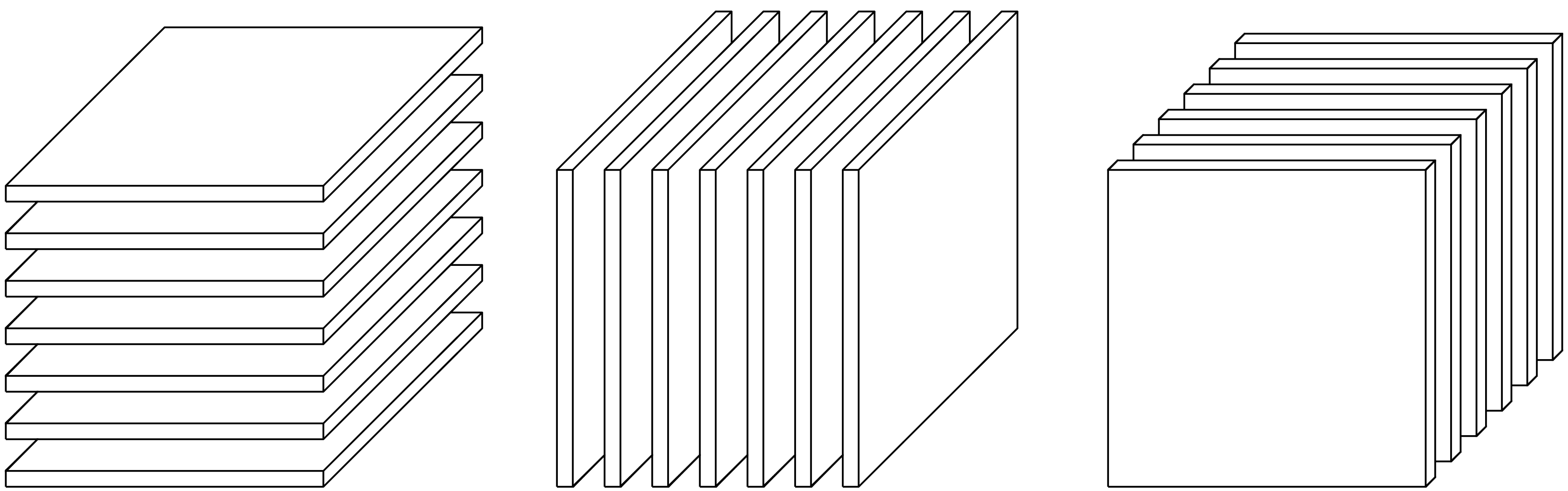
Terminology

- Tensor is N-way array
  - E.g. a matrix is a 2-way array
- Other sources use:
  - N-dimensional
    - But is a 3-dimensional vector a 1-dimensional tensor?
  - rank-N
    - But we have a different use for the word rank
- A 3-way tensor can be N-by-M-by-K dimensional
- A 3-way tensor has three modes
  - Columns, rows, and tubes

Fibres and Slices



(a) Mode-1 (column) fibers: $\mathbf{x}_{:jk}$ (b) Mode-2 (row) fibers: $\mathbf{x}_{i:k}$ (c) Mode-3 (tube) fibers: $\mathbf{x}_{ij:}$



(a) Horizontal slices: $\mathbf{X}_{i::}$ (b) Lateral slices: $\mathbf{X}_{:j:}$ (c) Frontal slices: $\mathbf{X}_{::k}$ (or $\mathbf{X}_k$)

Kolda & Bader 2009

Basic Operations

- Tensors require extensions to the standard linear algebra operations for matrices
- But before tensor operations, a recap on vectors and matrices

Basic Operations on Vectors

- A **transpose** $\boldsymbol{v}^T$ transposes a row vector into a column vector and vice versa
- If $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$, $\boldsymbol{v} + \boldsymbol{w}$ is a vector with $(\boldsymbol{v} + \boldsymbol{w})_i = v_i + w_i$
- For vector $\boldsymbol{v}$ and scalar α, $(\alpha\boldsymbol{v})_i = \alpha\boldsymbol{v}_i$
- A **dot product** of two vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ is $\boldsymbol{v}\cdot\boldsymbol{w} = \sum_i v_i w_i$
  - A.k.a. **scalar product** or **inner product**
  - Alternative notations: $\langle\boldsymbol{v}, \boldsymbol{w}\rangle$, $\boldsymbol{v}^T\boldsymbol{w}$ (for column vectors), $\boldsymbol{v}\boldsymbol{w}^T$ (for row vectors)

Basic Operations on Matrices

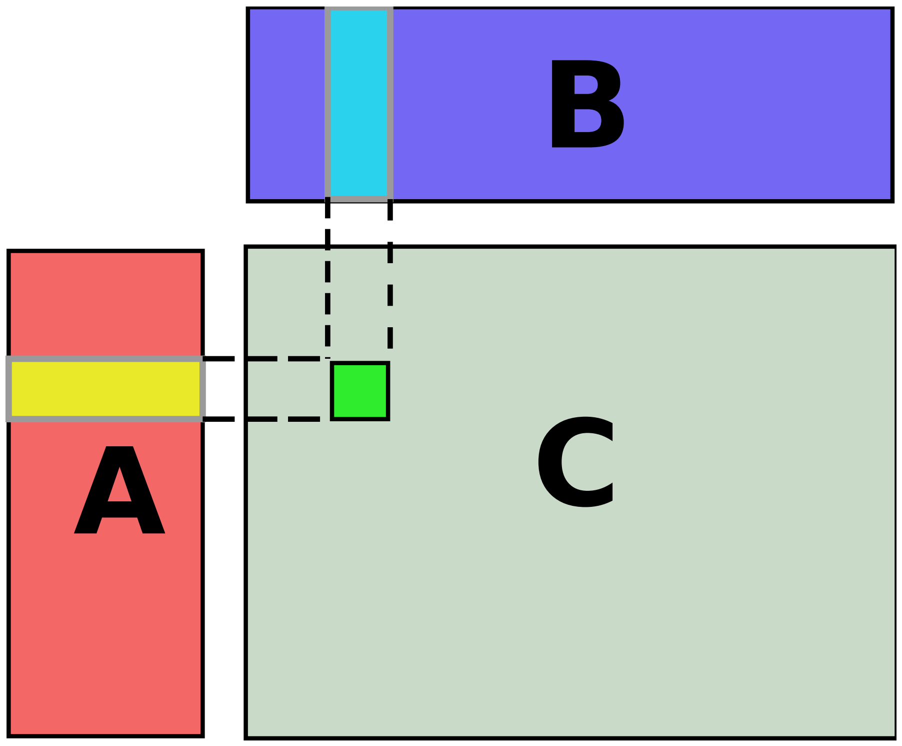
- Matrix **transpose** $\boldsymbol{A}^T$ has the rows of $\boldsymbol{A}$ as its columns
- If $\boldsymbol{A}$ and $\boldsymbol{B}$ are n-by-m matrices, then $\boldsymbol{A} + \boldsymbol{B}$ is an n-by-m matrix with $(\boldsymbol{A} + \boldsymbol{B})_{ij} = m_{ij} + n_{ij}$
- If $\boldsymbol{A}$ is n-by-k and $\boldsymbol{B}$ is k-by-m, then $\boldsymbol{AB}$ is an n-by-m matrix with

$$(\boldsymbol{AB})_{ij} = \sum_{\ell=1}^{k} a_{i\ell} b_{\ell j}$$

  - **Vector outer product** $\boldsymbol{vw}^T$ (for column vectors) is the matrix product of n-by-1 and 1-by-m matrices

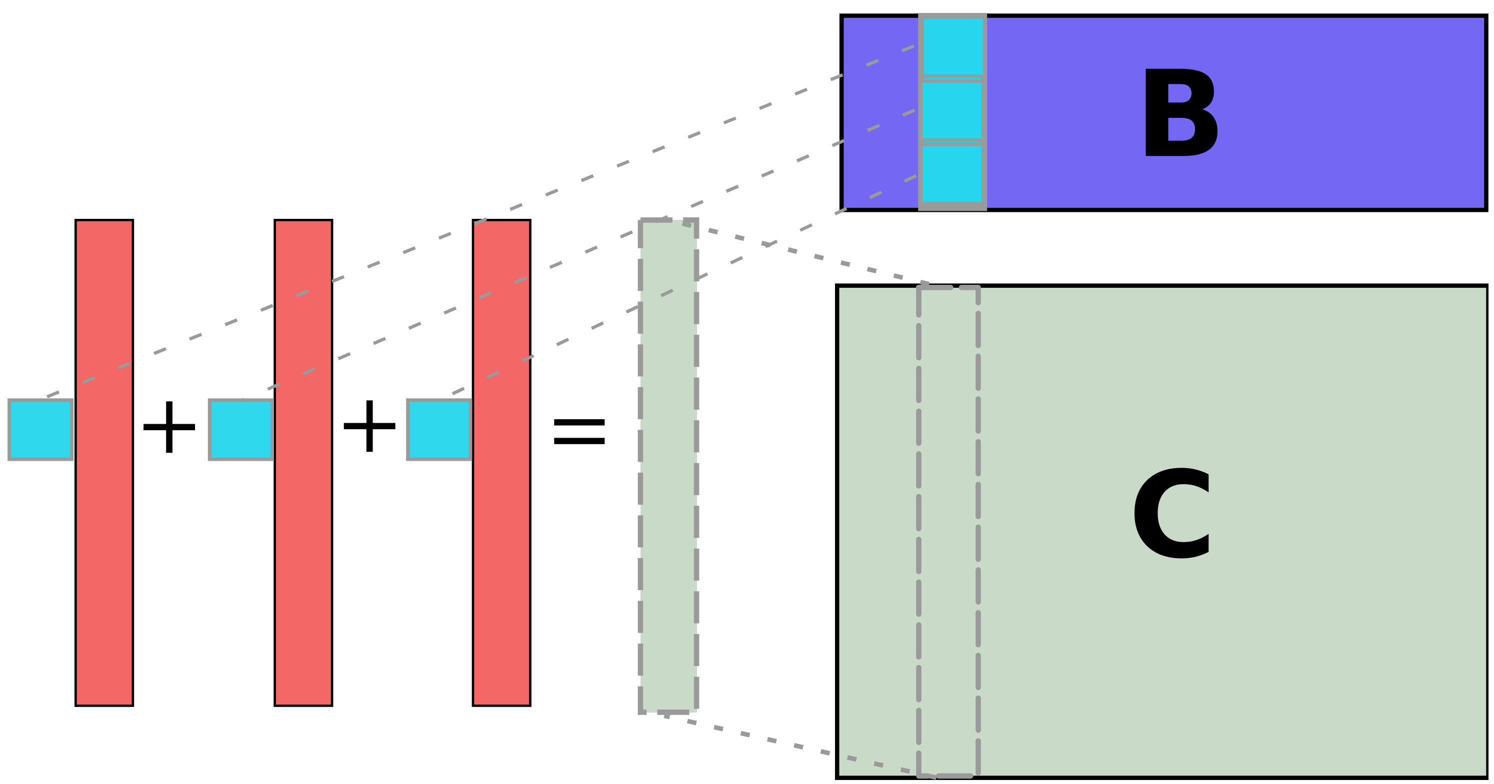
Intuition for Matrix Multiplication

- Element $(\boldsymbol{AB})_{ij}$ is the inner product of row i of $\boldsymbol{A}$ and column j of $\boldsymbol{B}$



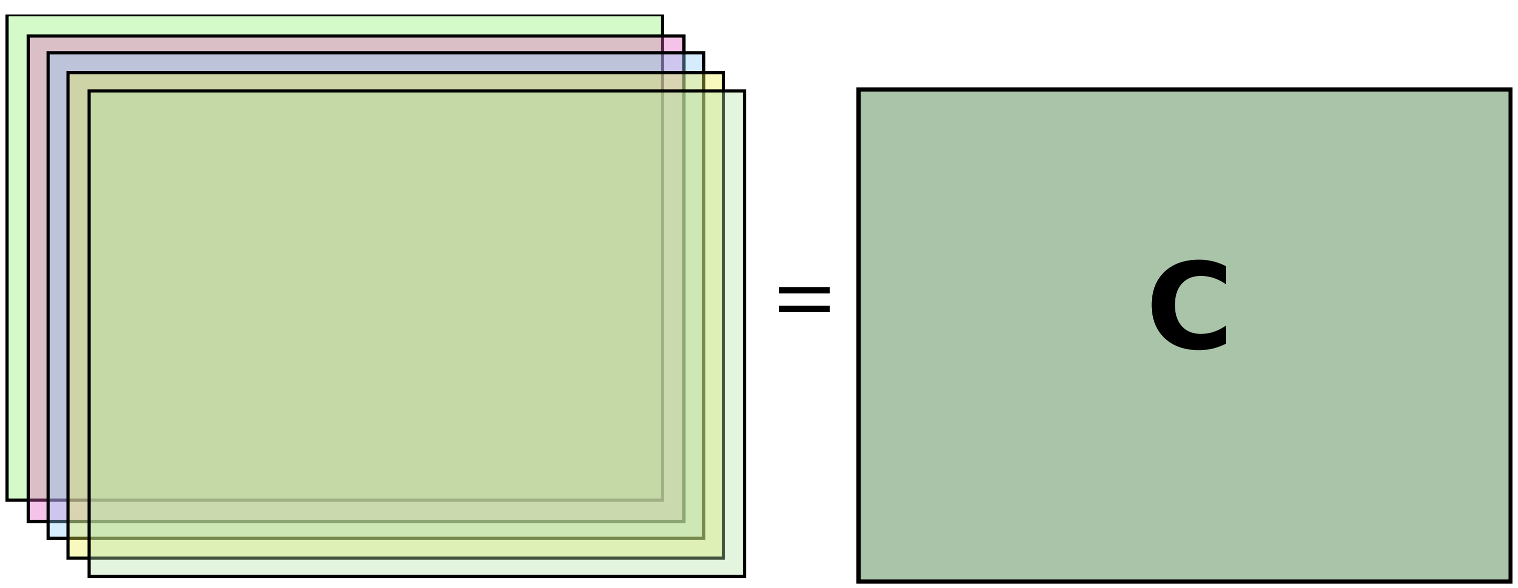
Intuition for Matrix Multiplication

- Column j of $\boldsymbol{AB}$ is the linear combination of columns of $\boldsymbol{A}$ with the coefficients coming from column j of $\boldsymbol{B}$



Intuition for Matrix Multiplication

- Matrix $\boldsymbol{AB}$ is a sum of k matrices $\boldsymbol{a}_l\boldsymbol{b}_l^T$ obtained by multiplying the l-th column of $\boldsymbol{A}$ with the l-th row of $\boldsymbol{B}$



Tensor Basic Operations

- A **multi-way vector outer product** is a tensor where each element is the product of corresponding elements in vectors:

$$(\boldsymbol{a} \circ \boldsymbol{b} \circ \boldsymbol{c})_{ijk} = a_i b_j c_k$$

- **Tensor sum** of two same-sized tensors is their element-wise sum $(\mathcal{X} + \mathcal{Y})_{ijk} = x_{ijk} + y_{ijk}$

- A **tensor inner product** of two same-sized tensors is the sum of the element-wise products of their values:

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i=1}^{I} \sum_{j=1}^{J} \cdots \sum_{z=1}^{Z} x_{ij\cdots z} y_{ij\cdots z}$$

Norms and Distances

- The **Frobenius norm** of a matrix $\boldsymbol{M}$ is $||\boldsymbol{M}||_F = (\Sigma_{i,j}\, m_{ij}{}^2)^{1/2}$
  - Can be used as a distance between two matrices: $d(\boldsymbol{M}, \boldsymbol{N}) = ||\boldsymbol{M} - \boldsymbol{N}||_F$
- Similar Frobenius distance on tensors is

$$d(\mathcal{X}, \mathcal{Y}) = \left(\sum_{i,j,k} (x_{ijk} - y_{ijk})^2\right)^{1/2}$$

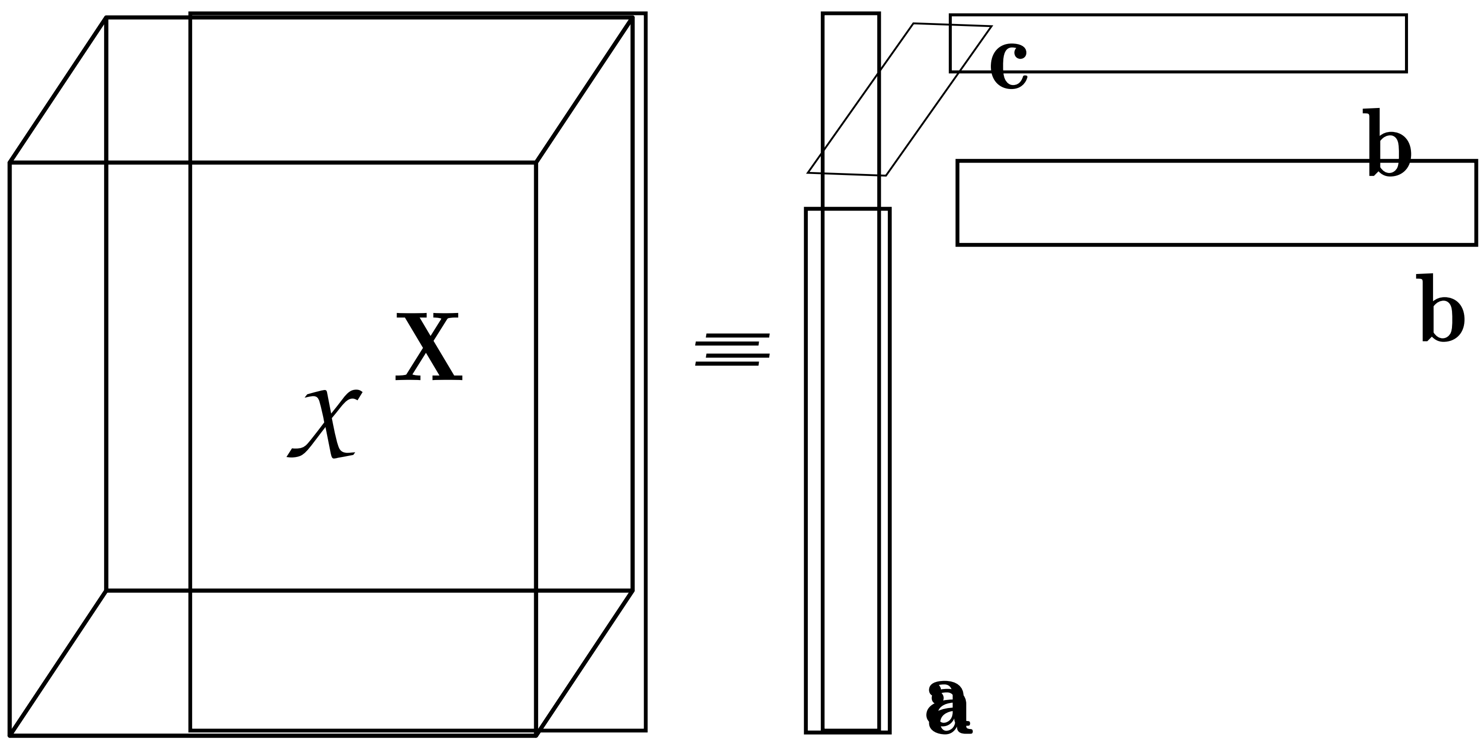
  - Equivalently $\sqrt{\langle \mathcal{X} - \mathcal{Y}, \mathcal{X} - \mathcal{Y} \rangle}$

CP Decomposition and Tensor Rank

- A matrix decomposition represents the given matrix as a product of two (or more) factor matrices
- The **rank** of a matrix ***M*** is the
  - Number of linearly independent rows (**row rank**)
  - Number of linearly independent columns (**column rank**)
  - Number of rank-1 matrices needed to be summed to get ***M*** (**Schein rank**)
    - A rank-1 matrix is an outer product of two vectors
  - They all are equivalent

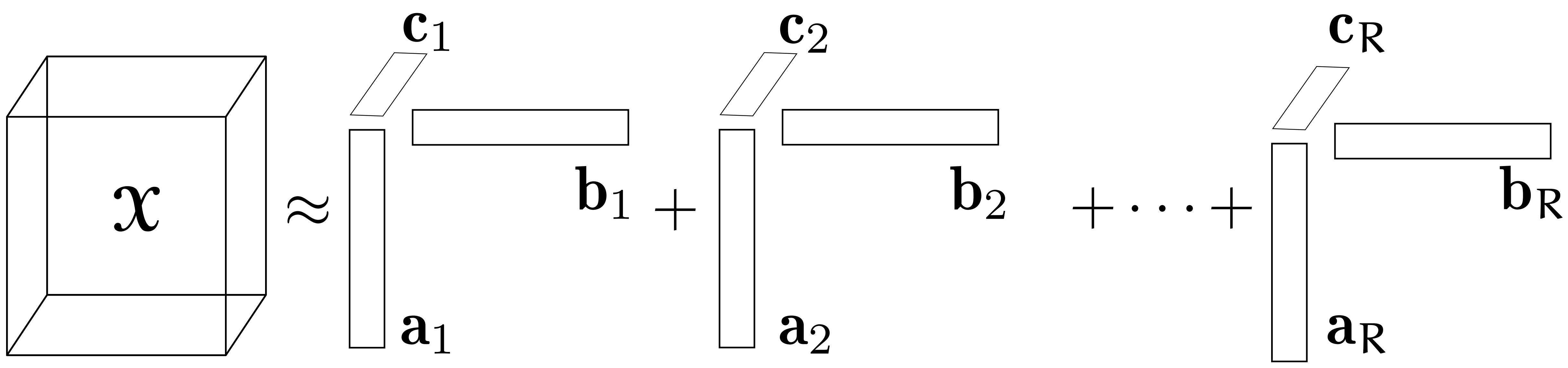
This we generalize

Rank-1 Tensors



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

The CP Tensor Decomposition



$$x_{ijk} \approx \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr}$$

More on CP

- The **size** of the CP decomposition is the number of rank-1 tensors involved
- The factorization can also be written using N factor matrices (for order-N tensor)
  - All column vectors are collected in one matrix, all row vectors in other, all tube vectors in third, etc.

CANDECOM, PARAFAC, ...

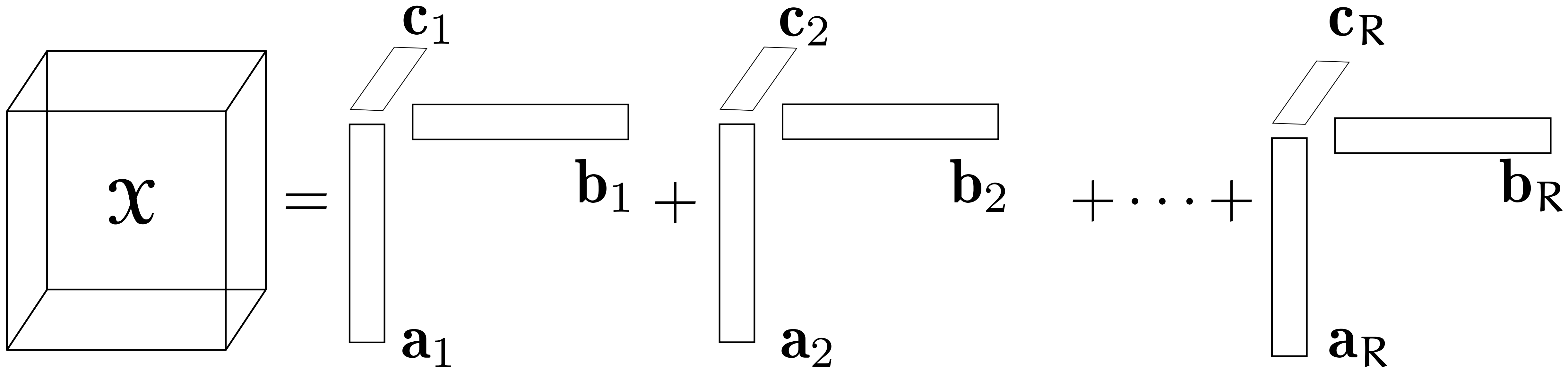
Name	Proposed by
Polyadic Form of a Tensor	Hitchcock, 1927 [105]
PARAFAC (Parallel Factors)	Harshman, 1970 [90]
CANDECOMP or CAND (Canonical decomposition)	Carroll and Chang, 1970 [38]
Topographic Components Model	Möcks, 1988 [166]
CP (CANDECOMP/PARAFAC)	Kiers, 2000 [122]

Table 3.1: Some of the many names for the CP decomposition.

Kolda & Bader 2009

Tensor Rank

- The **rank** of a tensor is the minimum number of rank-1 tensors needed to represent the tensor exactly
  - The CP decomposition of size *R*
  - Generalizes the matrix Schein rank



Tensor Rank Oddities #1

- The rank of a (real-valued) tensor is different over reals and over complex numbers.
  - With reals, the rank can be larger than the largest dimension
    - rank($\mathcal{X}$) ≤ min{IJ, IK, JK} for I-by-J-by-K tensor

$$\mathcal{X} = \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right)$$

$$\mathbf{A} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix},$$

$$\mathbf{B} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}$$

Tensor Rank Oddities #2

- There are tensors of rank R that can be approximated arbitrarily well with tensors of rank R' for some $R' < R$.
  - That is, there are no best low-rank approximation for such tensors.
  - **Eckart–Young-theorem** shows this is impossible with matrices.
  - The smallest such R' is called the **border rank** of the tensor.

Tensor Rank Oddities #3

- The rank-R CP decomposition of a rank-R tensor is essentially unique under mild conditions.
  - Essentially unique = only scaling and permuting are allowed.
  - Does not contradict #2, as this is the rank decomposition, not low-rank decomposition.
  - Again, not true for matrices (unless orthogonality etc. is required).

Tensor Matricization and New Matrix Products

- Tensor **matricization** unfolds an N-way tensor into a matrix
  - Mode-n matricization arranges the mode-n fibers as columns of a matrix, denoted $\boldsymbol{X}_{(n)}$
  - As many rows as is the dimensionality of the nth mode
  - As many columns as is the product of the dimensions of the other modes
- If $\mathcal{X}$ is an N-way tensor of size $I_1 \times I_2 \times \ldots \times I_N$, then $\boldsymbol{X}_{(n)}$ maps element $x_{i_1, i_2, \ldots, i_N}$ into (i_N, j) where

$$j = 1 + \sum_{k=1}^{N} (i_k - 1) J_k [k \neq n] \text{ with } J_k = \prod_{m=1}^{k-1} I_m [m \neq n]$$

Matricization Example

$$\mathcal{X} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

$$\mathbf{X}_{(1)} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \end{pmatrix}$$

$$\mathbf{X}_{(2)} = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

$$\mathbf{X}_{(3)} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 \end{pmatrix}$$

Another matricization example

$$\mathbf{X}_1 = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \qquad \mathbf{X}_2 = \begin{pmatrix} 5 & 7 \\ 6 & 8 \end{pmatrix}$$

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

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

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

Hadamard Matrix Product

- The element-wise matrix product
- Two matrices of size *n*-by-*m*, resulting matrix of size *n*-by-*m*

$$\boldsymbol{A} * \boldsymbol{B} = \begin{pmatrix} a_{1,1}b_{1,1} & a_{1,2}b_{1,2} & \cdots & a_{1,m}b_{1,m} \\ a_{2,1}b_{2,1} & a_{2,2}b_{2,2} & \cdots & a_{2,m}b_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1}b_{n,1} & a_{n,2}b_{n,2} & \cdots & a_{n,m}b_{n,m} \end{pmatrix}$$

Kronecker Matrix Product

- Element-per-matrix product
- n-by-m and j-by-k matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ give nj-by-mk matrix $\boldsymbol{A}\otimes\boldsymbol{B}$

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{pmatrix} a_{1,1}\boldsymbol{B} & a_{1,2}\boldsymbol{B} & \cdots & a_{1,m}\boldsymbol{B} \\ a_{2,1}\boldsymbol{B} & a_{2,2}\boldsymbol{B} & \cdots & a_{2,m}\boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1}\boldsymbol{B} & a_{n,2}\boldsymbol{B} & \cdots & a_{n,m}\boldsymbol{B} \end{pmatrix}$$

Khatri–Rao Matrix Product

- Element-per-column product
  - Number of columns must match
- n-by-m and k-by-m matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ give nk-by-m matrix $\boldsymbol{A} \odot \boldsymbol{B}$

$$\boldsymbol{A} \odot \boldsymbol{B} = \begin{pmatrix} a_{1,1}\boldsymbol{b}_1 & a_{1,2}\boldsymbol{b}_2 & \cdots & a_{1,m}\boldsymbol{b}_m \\ a_{2,1}\boldsymbol{b}_1 & a_{2,2}\boldsymbol{b}_2 & \cdots & a_{2,m}\boldsymbol{b}_m \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1}\boldsymbol{b}_1 & a_{n,2}\boldsymbol{b}_2 & \cdots & a_{n,m}\boldsymbol{b}_m \end{pmatrix}$$

Some identities

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{A}\mathbf{C} \otimes \mathbf{B}\mathbf{D}$$

$$(\mathbf{A} \otimes \mathbf{B})^{+} = \mathbf{A}^{+} \otimes \mathbf{B}^{+}$$

$$\mathbf{A} \odot \mathbf{B} \odot \mathbf{C} = (\mathbf{A} \odot \mathbf{B}) \odot \mathbf{C} = \mathbf{A} \odot (\mathbf{B} \odot \mathbf{C})$$

$$(\mathbf{A} \odot \mathbf{B})^{T}(\mathbf{A} \odot \mathbf{B}) = \mathbf{A}^{T}\mathbf{A} * \mathbf{B}^{T}\mathbf{B}$$

$$(\mathbf{A} \odot \mathbf{B})^{+} = \left((\mathbf{A}^{T}\mathbf{A}) * (\mathbf{B}^{T}\mathbf{B})\right)^{+}(\mathbf{A} \odot \mathbf{B})^{T}$$

$\mathbf{A}^{+}$ is the **Moore–Penrose pseudo-inverse**

Another View on the CP

- Using matricization and Khatri–Rao, we can re-write the CP decomposition
  - One equation per mode

$$X_{(1)} = A(C \odot B)^T$$
$$X_{(2)} = B(C \odot A)^T$$
$$X_{(3)} = C(B \odot A)^T$$

Solving CP: The ALS Approach

1. Fix $\boldsymbol{B}$ and $\boldsymbol{C}$ and solve $\boldsymbol{A}$
2. Solve $\boldsymbol{B}$ and $\boldsymbol{C}$ similarly
3. Repeat until convergence

$$\min_{\boldsymbol{A}} \left\| \boldsymbol{X}_{(1)} - \boldsymbol{A}(\boldsymbol{C} \odot \boldsymbol{B})^T \right\|_F$$

$$\boldsymbol{A} = \boldsymbol{X}_{(1)}\left((\boldsymbol{C} \odot \boldsymbol{B})^T\right)^+$$

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

R-by-R matrix

Wrap-up

- Tensors generalize matrices
- Many matrix concepts generalize as well
  - But some don’t
  - And some behave very differently
- We’ve only started with the basic of tensors…

Suggested Reading

- Skillicorn, D., 2007. *Understanding Complex Datasets: Data Mining with Matrix Decompositions*, Chapman & Hall/CRC, Boca Raton. Chapter 9
- Kolda, T.G. & Bader, B.W., 2009. Tensor decompositions and applications. *SIAM Review* 51(3), pp. 455–500.