Matrix/Tensor Operations

Center for Artificial Intelligence Technologies (CAIT)
Skolkowo Institute of Science and Technology (SKOLTECH), Moscow, Russia

January 31, 2022

Part I

Basic Multilinear Algebra

- **Tensor**: a multi-way array of data $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$
  Matrices by bold capital letters, e.g.
  $\mathbf{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_R] \in \mathbb{R}^{I \times R}$,
  Vectors by bold italic letters, e.g. $\boldsymbol{a}_j$ or $\boldsymbol{I} = [I_1, I_2, \ldots, I_N]$.
- **Slice**: For an order-3 tensor $\mathcal{Y}$, $\mathbf{Y}_k = \mathbf{Y}_{::k}$ denote frontal slice, $\mathbf{Y}_{:j:}$, lateral slice, and $\mathbf{Y}_{i::}$ horizontal slice.
- **Tube**(fiber, vector) at a position (i, j) along the mode-3 is denoted by $\boldsymbol{y}_{ij:}$,
- **Tube** at a position $(i_1, \ldots, i_{n-1}, :, i_{n+1}, \ldots, i_N)$ along mode-n is an I_n vector

$$\mathcal{A}(i_1, \ldots, i_{n-1}, :, i_{n+1}, \ldots, i_N) = \begin{bmatrix} \mathcal{A}(i_1, \ldots, i_{n-1}, 1, i_{n+1}, \ldots, i_N) \\ \vdots \\ \mathcal{A}(i_1, \ldots, i_{n-1}, I_n, i_{n+1}, \ldots, i_N) \end{bmatrix}.$$

Definition (vectorization)

Vectorization of an order-N tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$: maps $\mathcal{A}$ to a column vector $\boldsymbol{a}$

$$\boldsymbol{a} = \text{vec}(\mathcal{A}) = \begin{bmatrix} \text{vec}\left(\mathcal{A}^{(1)}\right) \\ \vdots \\ \text{vec}\left(\mathcal{A}^{(I_N)}\right) \end{bmatrix} \tag{1}$$

where $\mathcal{A}^{(i_N)}$ is an $(N-1)$-order subtensor: $\mathcal{A}^{(i_N)}(i_1, i_2, \ldots, i_{N-1}) = \mathcal{A}(i_1, i_2, \ldots, i_{N-1}, i_N)$.

Linear index An entry $a_{\boldsymbol{i}} = \mathcal{A}(i_1, i_2, \ldots, i_N)$ will be an entry $\boldsymbol{a}(\text{ivec}(\boldsymbol{i}, \boldsymbol{I}))$

$$\text{ivec}(\boldsymbol{i}, \boldsymbol{I}) = i_1 + (i_2 - 1)I_1 + (i_3 - 1)I_1 I_2 + \cdots + (i_N - 1)I_1 \cdots I_{N-1}.$$

Definition (**Tensor transposition)**

If $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ and $\boldsymbol{p}$ is a permutation of $[1, 2, \ldots, N]$, then $\mathcal{A}^{<\boldsymbol{p}>} \in \mathbb{R}^{I_{p_1} \times \cdots \times I_{p_N}}$ denotes the $\boldsymbol{p}$-transpose of $\mathcal{A}$

$$\mathcal{A}^{<\boldsymbol{p}>}(i_{p_1}, \ldots, i_{p_N}) = \mathcal{A}(i_1, \ldots, i_N), \quad \mathbf{1} \leq \boldsymbol{i} \leq \boldsymbol{I} = [I_1, I_2, \ldots, I_N].$$

Definition (**Unfolding or matricization of tensor)**

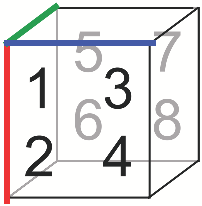
Mode-n unfolding of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is to horizontally concatenate all tubes of $\mathcal{A}$ along mode-n to yield an $I_n \times (\prod_{m \neq n} I_m)$ matrix

$$\mathbf{A}_{(n)} = \left[\ \mathrm{vec}\Big(\mathcal{A}^{(1,n)}\Big) \quad \cdots \quad \mathrm{vec}\Big(\mathcal{A}^{(I_n,n)}\Big) \ \right]^T$$

where $\mathcal{A}^{(i_n,n)}$ is an $(N-1)$-order subtensor of $\mathcal{A}$ whose the n-th index is fixed to i_n

$$\mathcal{A}^{(i_n,n)}(i_1, \ldots, i_{n-1}, i_{n+1}, \ldots, i_N) = \mathcal{A}(i_1, \ldots, i_n, \ldots, i_N).$$

Matricization



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

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

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

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

Corollary

$$\mathrm{vec}(\mathcal{A}) = \mathrm{vec}\big(\mathbf{A}_{(1)}\big) = \mathrm{vec}\big(\mathbf{A}_{(N)}^T\big)$$

Lemma

if $\mathcal{B} = \mathcal{A}^{<\boldsymbol{p}>}$*, and* $\boldsymbol{p} = [n, 1, \ldots, n-1, n+1, \ldots, N]$ *, then* $\mathbf{A}_{(n)} = \mathbf{B}_{(1)}$.

Question: How to efficiently perform mode-n unfolding?

Definition (**Reshaping)**

Reshaping a tensor $\mathcal{A}$ yields another tensor $\mathcal{B}$ with different shape but preserves its vectorization, i.e.,

$$\text{vec}(\mathcal{A}) = \text{vec}(\mathcal{B})$$

Example

- Reshape a vector of length $(I_1 I_2)$ to a matrix of size $I_1 \times I_2$
- Reshape a tensor of size $I_1 \times I_2 \times I_3 \times I_4 \times I_5$ to an order-3 tensor of size $(I_1 I_2) \times I_2 \times (I_3 I_4)$
- Reshaping: no need tensor permutation
- $\mathbf{Y}_{(1:n)} = \texttt{reshaping}(\mathcal{Y}, \texttt{I}_1\texttt{I}_2 \ldots \texttt{I}_\texttt{n}, \texttt{I}_{\texttt{n}+1} \ldots \texttt{I}_\texttt{N})$

Definition (**Outer product)**

The outer product of the tensors $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and $\mathcal{X} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_M}$ is given by

$$\mathcal{Z} = \mathcal{Y} \circ \mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N \times J_1 \times J_2 \times \cdots \times J_M}, \tag{2}$$

where

$$z_{i_1, i_2, \ldots, i_N, j_1, j_2, \ldots, j_M} = y_{i_1, i_2, \ldots, i_N} \; x_{j_1, j_2, \ldots, j_M}. \tag{3}$$

The tensor $\mathcal{Z}$ contains all the possible combinations of pair-wise products between the elements of $\mathcal{Y}$ and $\mathcal{X}$.
This operator is very closely related to the Kronecker product defined for matrices.

- Rank-one matrix: $\mathbf{A} = \boldsymbol{a} \circ \boldsymbol{b} = \boldsymbol{a}\boldsymbol{b}^T \in \mathbb{R}^{I \times J}$
- Rank-one tensor $\mathcal{Z} = \boldsymbol{a} \circ \boldsymbol{b} \circ \boldsymbol{c} \in \mathbb{R}^{I \times J \times Q}$, where $z_{ijq} = a_i \; b_j \; c_q$.

Definition (**Kronecker product**)

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\ \mathbf{B} & a_{12}\ \mathbf{B} & \cdots & a_{1J}\ \mathbf{B} \\ a_{21}\ \mathbf{B} & a_{22}\ \mathbf{B} & \cdots & a_{2J}\ \mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}\ \mathbf{B} & a_{I2}\ \mathbf{B} & \cdots & a_{IJ}\ \mathbf{B} \end{bmatrix} \tag{4}$$

Notation (**Kronecker product of matrices**)

Given set of N matrices $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R_n}$, Kronecker products among them are

$$\begin{aligned} \mathbf{A}^{\otimes} &= \bigotimes_{n=1}^{N} \mathbf{A}^{(n)} = \mathbf{A}^{(N)} \otimes \cdots \otimes \mathbf{A}^{(n)} \otimes \cdots \otimes \mathbf{A}^{(1)}, \\ \mathbf{A}^{\otimes_{-n}} &= \bigotimes_{k \neq n} \mathbf{A}^{(k)} = \mathbf{A}^{(N)} \otimes \cdots \otimes \mathbf{A}^{(n+1)} \otimes \mathbf{A}^{(n-1)} \otimes \cdots \otimes \mathbf{A}^{(1)}. \end{aligned}$$

Definition (**Khatri-Rao product)**

For $\mathbf{A} \in \mathbb{R}^{I \times J}$ and $\mathbf{B} \in \mathbb{R}^{T \times J}$, their Khatri-Rao product performs:

$$\begin{aligned} \mathbf{A} \odot \mathbf{B} &= [\boldsymbol{a}_1 \otimes \boldsymbol{b}_1 \;\; \boldsymbol{a}_2 \otimes \boldsymbol{b}_2 \;\; \cdots \;\; \boldsymbol{a}_J \otimes \boldsymbol{b}_J] \\ &= \left[\operatorname{vec}(\boldsymbol{b}_1 \boldsymbol{a}_1^T) \;\; \operatorname{vec}(\boldsymbol{b}_2 \boldsymbol{a}_2^T) \;\; \cdots \;\; \operatorname{vec}(\boldsymbol{b}_J \boldsymbol{a}_J^T)\right] \in \mathbb{R}^{IT \times J}. \end{aligned}$$

Notation (**Khatri-Rao product of matrices)**

Given set of N matrices $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$

$$\begin{aligned} \mathbf{A}^{\odot} &= \bigodot_{n=1}^{N} \mathbf{A}^{(n)} = \mathbf{A}^{(N)} \odot \cdots \odot \mathbf{A}^{(n)} \odot \cdots \odot \mathbf{A}^{(1)}, \\ \mathbf{A}^{\odot_{-n}} &= \bigodot_{k \neq n} \mathbf{A}^{(k)} = \mathbf{A}^{(N)} \odot \cdots \odot \mathbf{A}^{(n+1)} \odot \mathbf{A}^{(n-1)} \;\; \cdots \odot \mathbf{A}^{(1)}. \end{aligned}$$

Definition (**Hadamard product of two equal-size matrices)**

$$\mathbf{A} \circledast \mathbf{B} = \begin{bmatrix} a_{11}\, b_{11} & a_{12}\, b_{12} & \cdots & a_{1J}\, b_{1J} \\ a_{21}\, b_{21} & a_{22}\, b_{22} & \cdots & a_{2J}\, b_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}\, b_{I1} & a_{I2}\, b_{I2} & \cdots & a_{IJ}\, b_{IJ} \end{bmatrix}. \tag{5}$$

Notation (**Hadamard product of matrices)**

Given set of N matrices $\mathbf{A}^{(n)} \in \mathbb{R}^{I \times R}$, the following notation denotes Hadamard products among them

$$\begin{aligned} \mathbf{A}^{\circledast} &= \circledast_{n=1}^{N} \mathbf{A}^{(n)} = \mathbf{A}^{(N)} \circledast \cdots \circledast \mathbf{A}^{(n)} \circledast \cdots \circledast \mathbf{A}^{(1)}, \\ \mathbf{A}^{\circledast_{-n}} &= \circledast_{k \neq n} \mathbf{A}^{(k)} = \mathbf{A}^{(N)} \circledast \cdots \circledast \mathbf{A}^{(n+1)} \circledast \mathbf{A}^{(n-1)} \circledast \cdots \circledast \mathbf{A}^{(1)}. \end{aligned}$$

Properties: Vectorization I

- $\mathrm{vec}\left(\boldsymbol{a}\boldsymbol{b}^T\right) = \boldsymbol{b} \otimes \boldsymbol{a}$
- $\mathrm{vec}(\boldsymbol{a} \circ \boldsymbol{b} \circ \boldsymbol{c}) = \boldsymbol{c} \otimes \boldsymbol{b} \otimes \boldsymbol{a}$
- $\mathrm{vec}\left(\mathbf{A}\mathbf{B}^T\right) = (\mathbf{B} \odot \mathbf{A})\mathbf{1}_R$
- $\mathrm{vec}\left(\mathbf{A}\,\mathrm{diag}(\boldsymbol{s})\mathbf{B}^T\right) = (\mathbf{B} \odot \mathbf{A})\boldsymbol{s}$
- $\mathrm{vec}\left(\mathbf{A}\mathbf{G}\mathbf{B}^T\right) = (\mathbf{B} \otimes \mathbf{A})\,\mathrm{vec}(\mathbf{G})$
- $\mathrm{vec}(\mathbf{A} \circledast \mathbf{B}) = \mathrm{vec}(\mathbf{A}) \circledast \mathrm{vec}(\mathbf{B})$
- $\mathrm{vec}(\mathbf{A}\mathbf{B} - \mathbf{B}\mathbf{A}) = (\mathbf{I} \otimes \mathbf{A} - \mathbf{A}^T \otimes \mathbf{I})\,\mathrm{vec}(\mathbf{B})$

Properties I

Definition (**Commutation matrix)**

Given a matrix $\mathbf{A}$ of size $I \times J$, commutation matrix is a permutation matrix

$$\text{vec}(\mathbf{A}) = \mathbf{P}_{IJ}\,\text{vec}\left(\mathbf{A}^T\right)$$

$\mathbf{I}$ is identity matrix

$\mathbf{I} \otimes \mathbf{A} = \text{bdiag}(\mathbf{A}, \mathbf{A}, \ldots, \mathbf{A})$

$\mathbf{A} \otimes \mathbf{I} = ???$

$\mathbf{A} \otimes \mathbf{B}$ and $\mathbf{B} \otimes \mathbf{A}$???

$\mathbf{A} \otimes \mathbf{B} = (\mathbf{A} \otimes \mathbf{I})(\mathbf{I} \otimes \mathbf{B})$

Properties II

Given $\boldsymbol{a}$ of size $I \times 1$ and $\mathbf{B}$ of size $J \times S$

$$\begin{aligned}
\boldsymbol{a} \otimes \mathbf{B} &= [\boldsymbol{a} \otimes \boldsymbol{b}_1, \ldots, \boldsymbol{a} \otimes \boldsymbol{b}_S] \\
&= [\mathbf{P}_{I \times J}(\boldsymbol{b}_1 \otimes \boldsymbol{a}), \ldots, \mathbf{P}_{I \times J}(\boldsymbol{b}_S \otimes \boldsymbol{a})] \\
&= \mathbf{P}_{I \times J} [\boldsymbol{b}_1 \otimes \boldsymbol{a}, \ldots, \boldsymbol{b}_S \otimes \boldsymbol{a}] \\
&= \mathbf{P}_{I \times J} (\mathbf{B} \otimes \boldsymbol{a})
\end{aligned}$$

$$\mathbf{A} \otimes \mathbf{B} = \mathbf{P}_{I \times J}(\mathbf{B} \otimes \mathbf{A})\mathbf{P}_{R \times S}$$

Properties III

$(\mathbf{A} \otimes \mathbf{B})^T = \mathbf{A}^T \otimes \mathbf{B}^T$
Very important

$$
\begin{aligned}
(\mathbf{A} \otimes \mathbf{B})^T(\mathbf{C} \otimes \mathbf{D}) &= (\mathbf{A}^T\mathbf{C}) \otimes (\mathbf{B}^T\mathbf{D}) \\
(\mathbf{A} \otimes \mathbf{B})(\mathbf{E} \odot \mathbf{F}) &= (\mathbf{AE}) \odot (\mathbf{BF}) \\
(\mathbf{A} \odot \mathbf{B})^T (\mathbf{A} \odot \mathbf{B}) &= \mathbf{A}^T\mathbf{A} \circledast \mathbf{B}^T\mathbf{B}, \\
(\mathbf{A} \odot \mathbf{B})^\dagger &= [(\mathbf{A}^T\mathbf{A}) \circledast (\mathbf{B}^T\mathbf{B})]^{-1}(\mathbf{A} \odot \mathbf{B})^T
\end{aligned}
$$

Properties IV

- Orthogonal matrices **U** and **V**
  $(\mathbf{U} \otimes \mathbf{V})^T(\mathbf{U} \otimes \mathbf{V}) = (\mathbf{U}^T\mathbf{U}) \otimes (\mathbf{V}^T\mathbf{V}) = \mathbf{I}$
  $(\mathbf{U} \odot \mathbf{V})^T(\mathbf{U} \odot \mathbf{V}) = (\mathbf{U}^T\mathbf{U}) \circledast (\mathbf{V}^T\mathbf{V}) = \mathbf{I}$

If **A** and **B** are orthogonal matrices, then $\mathbf{F} = \mathbf{A} \otimes \mathbf{B}$ is also orthogonal

Orthogonal Procrustes problem I

Example (Orthogonal Procrustes problem)

Given two matrices $\mathbf{Y}$ and $\mathbf{A}$, find an orthogonal matrix $\mathbf{X}$ such that

$$\min_{\mathbf{X}} \quad f(\mathbf{X}) = \|\mathbf{Y} - \mathbf{X}\mathbf{A}\|_F^2 \qquad \text{s.t.} \quad \mathbf{X}^T\mathbf{X} = \mathbf{I}$$

Objective function: $f(\mathbf{X}) = \|\mathbf{Y}\|_F^2 + \|\mathbf{A}\|_F^2 - 2\,\mathrm{tr}(\mathbf{X}^T\mathbf{Y}\mathbf{A}^T)$
Let $\mathbf{Q} = \mathbf{Y}\mathbf{A}^T$, we solve a maximization problem

$$\max_{\mathbf{X}} \quad \mathrm{tr}(\mathbf{X}^T\mathbf{Q}) \qquad \text{s.t.} \quad \mathbf{X}^T\mathbf{X} = \mathbf{I}$$

Denote SVD of $\mathbf{Q} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ where $\mathbf{S} = diag(s_1, \ldots, s_R)$ is diagonal matrix of singular values of $\mathbf{Q}$.
The optimal $\mathbf{X}^\star = \mathbf{U}\mathbf{V}^T$.

Question:
Compute $\mathbf{X}$ when $\mathbf{Y}$ and $\mathbf{A}$ are big

Orthogonal Procrustes problem II

Let $\mathbf{X} = \mathbf{C} \otimes \mathbf{D}$, where $\mathbf{C}$ and $\mathbf{D}$ are two orthogonal matrices of smaller size than $\mathbf{X}$.
Instead of seeking $\mathbf{X}$, we

$$\min_{\mathbf{C},\mathbf{D}} \quad \|\mathbf{Y} - (\mathbf{C} \otimes \mathbf{D})\mathbf{A}\|_F^2 \qquad \text{s.t.} \quad \mathbf{C}^T\mathbf{C} = \mathbf{I}, \mathbf{D}^T\mathbf{D} = \mathbf{I}$$

Since $\mathbf{C}$ is orthogonal, $\mathbf{C} \otimes \mathbf{I}$ is orthogonal.
Keep $\mathbf{C}$ fixed, exploit

$$\mathbf{C} \otimes \mathbf{D} = (\mathbf{C} \otimes \mathbf{I})(\mathbf{I} \otimes \mathbf{D})$$

and reformulate the optimization problem for $\mathbf{D}$

$$\min_{\mathbf{D}} \quad \|(\mathbf{C}^T \otimes \mathbf{I})\mathbf{Y} - (\mathbf{I} \otimes \mathbf{D})\mathbf{A}\|_F^2 \qquad \text{s.t.} \quad \mathbf{D}^T\mathbf{D} = \mathbf{I}$$

Orthogonal Procrustes problem III

Taking into account that

$$(\mathbf{I}_R \otimes \mathbf{D})\mathbf{A} = \mathrm{bdiag}(\mathbf{D}, \mathbf{D}, \ldots, \mathbf{D}) \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_R \end{bmatrix} = \begin{bmatrix} \mathbf{D}\mathbf{A}_1 \\ \vdots \\ \mathbf{D}\mathbf{A}_R \end{bmatrix}$$

Denote $\tilde{\mathbf{Y}} = (\mathbf{C}^T \otimes \mathbf{I})\mathbf{Y}$

$$f(\mathbf{D}) \quad = \quad \|\tilde{\mathbf{Y}} - \begin{bmatrix} \mathbf{D}\mathbf{A}_1 \\ \vdots \\ \mathbf{D}\mathbf{A}_R \end{bmatrix}\|_F^2 = \|\begin{bmatrix} \tilde{\mathbf{Y}}_1 - \mathbf{D}\mathbf{A}_1 \\ \vdots \\ \tilde{\mathbf{Y}}_R - \mathbf{D}\mathbf{A}_R \end{bmatrix}\|_F^2$$

Denote $\hat{\mathbf{Y}} = \begin{bmatrix} \tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_R \end{bmatrix}$ and $\hat{\mathbf{A}} = \begin{bmatrix} \tilde{\mathbf{A}}_1, \ldots, \tilde{\mathbf{A}}_R \end{bmatrix}$
The problem becomes a standard Procrustes problem

$$\min_{\mathbf{D}} \quad f(\mathbf{D}) = \|\hat{\mathbf{Y}} - \mathbf{D}\hat{\mathbf{A}}\|_F^2 \quad \text{s.t.} \quad \mathbf{D}^T\mathbf{D} = \mathbf{I}.$$

Orthogonal Procrustes problem IV

$\mathbf{C}$ can be estimated similarly

$$\min_{\mathbf{C}} \quad \|\mathbf{Y} - (\mathbf{I} \otimes \mathbf{D})(\mathbf{C} \times \mathbf{I})\mathbf{A}\|_F^2$$

or

$$\min_{\mathbf{C}} \quad \|(\mathbf{I} \otimes \mathbf{D}^T)\mathbf{Y} - (\mathbf{C} \times \mathbf{I})\mathbf{A}\|_F^2$$

Properties I

- SVD $\mathbf{A} = \mathbf{U}_1 \Sigma_1 \mathbf{V}_1^T$ and $\mathbf{B} = \mathbf{U}_2 \Sigma_2 \mathbf{V}_2^T$ then

$$\mathbf{A} \otimes \mathbf{B} = (\mathbf{U}_1 \otimes \mathbf{U}_2)(\Sigma_1 \otimes \Sigma_2)(\mathbf{V}_1 \otimes \mathbf{V}_2)^T$$

  Singular values of $\mathbf{A} \otimes \mathbf{B}$ is Kronecker product of those of $\mathbf{A}$ and $\mathbf{B}$

$$\text{rank}(\mathbf{A} \otimes \mathbf{B}) = \text{rank}(\mathbf{A})\text{rank}(\mathbf{B})$$

Properties II

- For two square matrices $\mathbf{A}$, $\mathbf{B}$ of size $n \times n$ and $m \times m$, respectively

$$\operatorname{tr}(\mathbf{A} \otimes \mathbf{B}) = \operatorname{tr}(\mathbf{A}) \operatorname{tr}(\mathbf{B})$$

$$|\det(\mathbf{A} \otimes \mathbf{B})| = |\det(\mathbf{A})|^m |\det(\mathbf{B})|^n$$

Properties III

$$
\begin{aligned}
\mathbf{A}^{\odot T}\mathbf{A}^{\odot} &= \{\mathbf{A}^T\mathbf{A}\}^{\circledast} = \circledast_{n=1}^{N}\mathbf{A}^{(n)T}\mathbf{A}^{(n)} \\
\mathbf{A}^{\odot_{-n}T}\mathbf{A}^{\odot_{-n}} &= \{\mathbf{A}^T\mathbf{A}\}^{\circledast_{-n}} = \circledast_{k\neq n}\mathbf{A}^{(k)T}\mathbf{A}^{(k)}, \\
\mathbf{A}^{\otimes T}\mathbf{A}^{\otimes} &= \{\mathbf{A}^T\mathbf{A}\}^{\otimes} = \bigotimes_{n\neq 1}^{N}\mathbf{A}^{(n)T}\mathbf{A}^{(n)}, \\
\mathbf{A}^{\otimes_{-n}T}\mathbf{A}^{\otimes_{-n}} &= \{\mathbf{A}^T\mathbf{A}\}^{\otimes_{-n}} = \bigotimes_{k\neq n}\mathbf{A}^{(k)T}\mathbf{A}^{(k)}.
\end{aligned}
$$

Solve Large Linear System I

Example

Given square matrices $\mathbf{A}$ of size $I \times I$ and $\mathbf{B}$ of size $J \times J$, the system

$$(\mathbf{A} \otimes \mathbf{B})\boldsymbol{x} = \boldsymbol{y} \tag{6}$$

has optimal solution

$$\begin{aligned} \boldsymbol{x} &= (\mathbf{A} \otimes \mathbf{B})^{-1}\boldsymbol{y} \\ &= (\mathbf{A}^{-1} \otimes \mathbf{B}^{-1})\boldsymbol{y} \\ &= \mathrm{vec}\left(\mathbf{B}^{-1}\mathbf{Y}\mathbf{A}^{-1T}\right) \end{aligned} \tag{7}$$

where $\mathbf{Y} = \mathtt{reshape}(\boldsymbol{y}, \mathtt{I}, \mathtt{J})$

Solve Large Linear System II

Example

Given square matrices $\mathbf{A}$ of size $I \times R$ and $\mathbf{B}$ of size $J \times R$, the system

$$(\mathbf{A} \odot \mathbf{B})\boldsymbol{x} = \boldsymbol{y} \tag{8}$$

has optimal solution

$$\begin{aligned}
\boldsymbol{x} &= ((\mathbf{A} \odot \mathbf{B})^T(\mathbf{A} \odot \mathbf{B}))^{-1}(\mathbf{A} \odot \mathbf{B})^T\boldsymbol{y} \\
&= ((\mathbf{A}^T\mathbf{A}) \circledast (\mathbf{B}^T\mathbf{B}))^{-1}\begin{bmatrix} \boldsymbol{b}_1^T\mathbf{Y}\boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{b}_R^T\mathbf{Y}\boldsymbol{a}_R \end{bmatrix}
\end{aligned}$$

Sylvester matrix equations I

Example

Given $\mathbf{A}$ of size $I \times R$ and $\mathbf{B}$ of size $J \times S$, solve the following system

$$\mathbf{A}_1 \mathbf{X} \mathbf{B}_1^T + \mathbf{A}_2 \mathbf{X} \mathbf{B}_2^T = \mathbf{Y} \tag{9}$$

Let $\boldsymbol{y} = \text{vec}(\mathbf{Y})$ and $\boldsymbol{x} = \text{vec}(\mathbf{X})$, the optimal solution is given by

$$\boldsymbol{y} = (\mathbf{B}_1 \otimes \mathbf{A}_1 + \mathbf{B}_2 \otimes \mathbf{A}_2)\boldsymbol{x}$$

$$\boldsymbol{x}^\star = (\mathbf{B}_1 \otimes \mathbf{A}_1 + \mathbf{B}_2 \otimes \mathbf{A}_2)^\dagger \boldsymbol{y} \tag{10}$$

Question:

If $I, J > 100$ and $R, S > 100$, How to efficiently compute $(\mathbf{B}_1 \otimes \mathbf{A}_1 + \mathbf{B}_2 \otimes \mathbf{A}_2)^\dagger$

Sylvester matrix equations II

Solve a minimization problem with equality constraint

$$\begin{aligned} \min_{\mathbf{X}_1, \mathbf{X}_2} \quad & f(\mathbf{X}_1, \mathbf{X}_2) = \frac{1}{2}\|\mathbf{Y} - \mathbf{A}_1\mathbf{X}_1\mathbf{B}_1^T - \mathbf{A}_2\mathbf{X}_2\mathbf{B}_2^T\|_F^2 \\ \text{s.t.} \quad & \mathbf{X}_1 = \mathbf{X}_2 \end{aligned}$$

How to solve this constrained optimization?

Sylvester matrix equations III

Consider the augmented Lagrangian function

$$\mathcal{L} = f(\mathbf{X}_1, \mathbf{X}_2) - \frac{1}{\gamma}\operatorname{tr}(\mathbf{X}_1 - \mathbf{X}_2, \mathbf{T}) + \frac{1}{2\gamma}\|\mathbf{X}_1 - \mathbf{X}_2\|_F^2$$

or

$$\mathcal{L} = f(\mathbf{X}_1, \mathbf{X}_2) + \frac{1}{2\gamma}(\|\mathbf{X}_1 - \mathbf{X}_2 - \mathbf{T}\|_F^2 - \|\mathbf{T}\|_F^2)$$

Update $\mathbf{X}_1$

$$\mathbf{X}_1^\star = \arg\min_{\mathbf{X}_1} f(\mathbf{X}_1, \mathbf{X}_2) + \frac{1}{2\gamma}\|\mathbf{X}_1 - \mathbf{X}_2 - \mathbf{T}\|_F^2 \tag{11}$$

Sylvester matrix equations IV

Denote $\mathbf{Y}_1 = \mathbf{Y} - \mathbf{A}_2\mathbf{X}_2\mathbf{B}_2^T$ and $\boldsymbol{y}_1 = \mathrm{vec}(\mathbf{Y}_1)$, $\boldsymbol{x}_1 = \mathrm{vec}(\mathbf{X}_1)$, $\boldsymbol{x}_2 = \mathrm{vec}(\mathbf{X}_2)$, $\boldsymbol{t} = \mathrm{vec}(\mathbf{T})$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}_1} &= (\mathbf{B}_1^T \otimes \mathbf{A}_1^T)((\mathbf{B}_1 \otimes \mathbf{A}_1)\boldsymbol{x}_1 - \boldsymbol{y}_1) + \frac{1}{\gamma}(\boldsymbol{x}_1 - \boldsymbol{x}_2 - \boldsymbol{t}) \\ &= (\mathbf{B}_1^T\mathbf{B} \otimes \mathbf{A}_1^T\mathbf{A}_1 + \frac{1}{\gamma}\mathbf{I})\boldsymbol{x}_1 - (\mathbf{B}_1^T \otimes \mathbf{A}_1^T)\boldsymbol{y}_1 - \frac{1}{\gamma}(\boldsymbol{x}_2 + \boldsymbol{t})\end{aligned}$$

$$\boldsymbol{x}_1^\star = (\mathbf{B}_1^T\mathbf{B} \otimes \mathbf{A}_1^T\mathbf{A}_1 + \frac{1}{\gamma}\mathbf{I})^{-1}((\mathbf{B}_1^T \otimes \mathbf{A}_1^T)\boldsymbol{y}_1 + \frac{1}{\gamma}(\boldsymbol{x}_2 + \boldsymbol{t}))$$

Denote EVDs of $\mathbf{A}_1^T\mathbf{A}_1 = \mathbf{U}_1\Sigma_{a1}\mathbf{U}_1^T$ and $\mathbf{B}_1^T\mathbf{B}_1 = \mathbf{V}_1\Sigma_{b1}\mathbf{V}_1^T$
By expressing

$$\mathbf{Q} = \mathbf{B}_1^T\mathbf{B} \otimes \mathbf{A}_1^T\mathbf{A}_1 + \frac{1}{\gamma}\mathbf{I} = (\mathbf{V}_1 \otimes \mathbf{U}_1)\,\mathrm{diag}(\boldsymbol{\sigma}_{b1} \otimes \boldsymbol{\sigma}_{a1} + \frac{1}{\gamma})(\mathbf{V}_1^T \otimes \mathbf{U}_1^T)$$

Sylvester matrix equations V

we have

$$(\mathbf{B}_1^T\mathbf{B}\otimes\mathbf{A}_1^T\mathbf{A}_1+\frac{1}{\gamma}\mathbf{I})^{-1} = (\mathbf{V}_1\otimes\mathbf{U}_1)\,\mathrm{diag}(1./(\boldsymbol{\sigma}_{b1}\otimes\boldsymbol{\sigma}_{a1}+1/\gamma))(\mathbf{V}_1^T\otimes\mathbf{U}_1^T)$$

Let $\mathbf{K}_1 = \mathbf{A}_1^T\mathbf{Y}_1\mathbf{B}_1 + \frac{1}{\gamma}(\mathbf{X}_2 + \mathbf{T})$.

$$\mathbf{X}_1^\star = \mathbf{U}_1((\mathbf{U}_1^T\mathbf{K}_2\mathbf{V}_1)\oslash(\boldsymbol{\sigma}_{a1}\boldsymbol{\sigma}_{b1}^T + 1/\gamma))\mathbf{V}_1^T \tag{12}$$

Similarly, define $\mathbf{K}_2 = \mathbf{A}_2^T(\mathbf{Y} - \mathbf{A}_1\mathbf{X}_1\mathbf{B}_1^T)\mathbf{B}_2 + \frac{1}{\gamma}(\mathbf{X}_2 + \mathbf{T})$ we can update $\mathbf{X}_2$

$$\mathbf{X}_2^\star = \mathbf{U}_2((\mathbf{U}_2^T\mathbf{K}_2\mathbf{V}_2)\oslash(\boldsymbol{\sigma}_{a2}\boldsymbol{\sigma}_{b2}^T + 1/\gamma))\mathbf{V}_2^T \tag{13}$$

Tensor-matrix product I

Definition (mode-n tensor-matrix product)

$$\mathcal{Y} = \mathcal{G} \times_n \mathbf{A}\,, \quad \text{or} \quad \mathbf{Y}_{(n)} = \mathbf{A}\,\mathbf{G}_{(n)}.$$

For order-3 tensor $\mathcal{G}$ of size $I \times J \times K$, $\mathbf{A}$ ($L \times I$), $\mathbf{B}$ ($M \times J$), and $\mathbf{C}$ ($N \times K$)

$\mathcal{Y} = \mathcal{G} \times_1 \mathbf{A} \to \mathbf{Y}(:,:,k) = \mathbf{A}\mathbf{G}(:,:,k)$

$\mathcal{Y} = \mathcal{G} \times_2 \mathbf{B} \to \mathbf{Y}(:,:,k) = \mathbf{G}(:,:,k)\mathbf{B}^T$

$\mathcal{Y} = \mathcal{G} \times_3 \mathbf{C} \to \mathbf{Y}(:,j,:) = \mathbf{G}(:,j,:)\mathbf{C}^T$

Given $\mathcal{Y} = \mathcal{G} \times_1 \mathbf{A}$, then

$$\begin{aligned}
\mathbf{Y}_{(1)} &= \mathbf{A}\,\mathbf{G}_{(1)} \\
\mathbf{Y}_{(2)} &= \mathbf{G}_{(2)}(\mathbf{I}_{K\times K} \otimes \mathbf{A}^T) \\
\mathbf{Y}_{(3)} &= \mathbf{G}_{(3)}(\mathbf{I}_{J\times J} \otimes \mathbf{A}^T)
\end{aligned}$$

Tensor-matrix product II

Given $\mathcal{Y} = \mathcal{G} \times_2 \mathbf{B}$, then

$$\begin{aligned}
\mathbf{Y}_{(2)} &= \mathbf{B}\mathbf{G}_{(2)} \\
\mathbf{Y}_{(1)} &= \mathbf{G}_{(1)}(\mathbf{I}_{K\times K} \otimes \mathbf{B})^T \\
\mathbf{Y}_{(3)} &= \mathbf{G}_{(3)}(\mathbf{I}_{I\times I} \otimes \mathbf{B})^T
\end{aligned}$$

Tensor-matrix product III

Consider a rank-1 tensor $\mathcal{Y} = \boldsymbol{e}_1 \circ \boldsymbol{e}_2 \circ \boldsymbol{e}_3 \circ \boldsymbol{e}_4$.
Permutation of $\mathcal{Y}$ changes the order of components $\boldsymbol{e}_1$, $\boldsymbol{e}_2 \circ \boldsymbol{e}_3 \circ \boldsymbol{e}_4$

$$\mathcal{Y}^{(2,1,3,4)} = \boldsymbol{e}_2 \circ \boldsymbol{e}_1 \circ \boldsymbol{e}_3 \circ \boldsymbol{e}_4$$

Matricization of $\mathcal{Y}$

$$\begin{aligned}
\mathbf{Y}_{(1)} &= \boldsymbol{e}_1(\boldsymbol{e}_4 \otimes \boldsymbol{e}_3 \otimes \boldsymbol{e}_2)^T \\
\mathbf{Y}_{(2)} &= \boldsymbol{e}_2(\boldsymbol{e}_4 \otimes \boldsymbol{e}_3 \otimes \boldsymbol{e}_1)^T \\
\mathbf{Y}_{(3)} &= \boldsymbol{e}_3(\boldsymbol{e}_4 \otimes \boldsymbol{e}_2 \otimes \boldsymbol{e}_1)^T \\
\mathbf{Y}_{(4)} &= \boldsymbol{e}_4(\boldsymbol{e}_3 \otimes \boldsymbol{e}_2 \otimes \boldsymbol{e}_1)^T
\end{aligned}$$

Tensor-matrix product IV

$$\mathcal{X} = \mathcal{Y} \times_1 \mathbf{A} = (\mathbf{A}\boldsymbol{e}_1) \circ \boldsymbol{e}_2 \circ \boldsymbol{e}_3 \circ \boldsymbol{e}_4$$

Mode-2 matricization of $\mathcal{X}$

$$\begin{aligned}
\mathbf{X}_{(2)} &= \boldsymbol{e}_2(\boldsymbol{e}_4 \otimes \boldsymbol{e}_3 \otimes \mathbf{A}\boldsymbol{e}_1)^T \\
&= \boldsymbol{e}_2(\boldsymbol{e}_4 \otimes \boldsymbol{e}_3 \otimes \boldsymbol{e}_1)^T(\mathbf{I} \otimes \mathbf{A}^T) \\
&= \mathbf{Y}_{(2)}(\mathbf{I} \otimes \mathbf{A}^T).
\end{aligned}$$

For an arbitrary tensor $\mathcal{Y}$, we can always express

$$\mathcal{Y} = \sum_{i_1, i_2, i_3, i_4} y_{i_1 i_2 i_3 i_4} \boldsymbol{e}_1 \circ \boldsymbol{e}_2 \circ \boldsymbol{e}_3 \circ \boldsymbol{e}_4$$

and

$$\mathcal{X} = \mathcal{Y} \times_1 \mathbf{A} = \sum_{i_1, i_2, i_3, i_4} y_{i_1 i_2 i_3 i_4} (\mathbf{A}\boldsymbol{e}_1) \circ \boldsymbol{e}_2 \circ \boldsymbol{e}_3 \circ \boldsymbol{e}_4$$

Tensor-matrix product V

Hence unfolding of $\mathcal{X}$ can write as

$$\mathcal{X}_{(2)} = \mathbf{Y}_{(2)}(\mathbf{I} \otimes \mathbf{A}^T)$$

Tensor-matrix product VI

Given $\mathcal{Y} = \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B}$, then

$$\begin{aligned}
\mathbf{Y}_{(1)} &= \mathbf{A}\mathbf{G}_{(1)}(\mathbf{I}_{K\times K} \otimes \mathbf{B}^T) \\
\mathbf{Y}_{(2)} &= \mathbf{B}\mathbf{G}_{(2)}(\mathbf{I}_{K\times K} \otimes \mathbf{A}^T) \\
\mathbf{Y}_{(3)} &= \mathbf{G}_{(3)}(\mathbf{B} \otimes \mathbf{A})^T
\end{aligned}$$

$$\begin{aligned}
\mathcal{G} \times \{\mathbf{A}\} &= \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)}, \\
[\mathcal{G} \times \{\mathbf{A}\}]_{(n)} &= \mathbf{A}^{(n)}\mathbf{G}_{(n)}\left[\mathbf{A}^{(N)} \otimes \cdots \otimes \mathbf{A}^{(n+1)} \otimes \mathbf{A}^{(n-1)} \otimes \cdots \otimes \mathbf{A}^{(1)}\right]^T.
\end{aligned}$$

Tensor-vector product

Definition (mode-n tensor-vector product)

Mode-n multiplication of $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ by $\boldsymbol{a} \in \mathbb{R}^{I_n}$ is denoted by

$$\mathcal{Z} = \mathcal{Y} \,\bar{\times}_n\, \boldsymbol{a} \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times I_{n+1} \times \cdots \times I_N},$$

and product of $\mathcal{Y}$ with $\{\boldsymbol{a}\} = \left\{\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}, \ldots, \boldsymbol{a}^{(N)}\right\}$ is given by

$$\mathcal{Y} \,\bar{\times}\, \{\boldsymbol{a}\} = \mathcal{Y} \,\bar{\times}_1\, \boldsymbol{a}^{(1)} \,\bar{\times}_2\, \boldsymbol{a}^{(2)} \cdots \bar{\times}_N\, \boldsymbol{a}^{(N)} .$$

$$\mathcal{G} \times_{-n} \{\mathbf{A}\} = \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_{n-1} \mathbf{A}^{(n-1)} \times_{n+1} \mathbf{A}^{(n+1)} \cdots \times_N \mathbf{A}^{(N)} .$$

Tensor-matrix multiplication

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

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

$$\begin{aligned} y_{i,m,k} &= \sum_{j=1}^{J} x_{i,j,k}\, b_{m,j} \\ \mathbf{Y}_{:,:,k} &= \mathbf{X}_{:,:,k}\, \mathbf{B}^T \\ \mathbf{Y}_{i,:,:} &= \mathbf{B}\, \mathbf{X}_{i,:,:} \end{aligned}$$

$$\mathcal{Z} = \mathcal{X}\, \bar{\times}_3\, \boldsymbol{c} \quad \in \mathbb{R}^{I \times J}$$

$$\begin{aligned} z_{i,j} &= \sum_{k=1}^{K} x_{i,j,k}\, c_k \\ &= \mathbf{X}_{i,j,:}^T\, \boldsymbol{c} \end{aligned}$$

Tensor contraction I

Definition (Contraction between two tensors)

The contracted product of $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_M \times J_1 \times \cdots \times J_N}$ and $\mathcal{B} \in \mathbb{R}^{I_1 \times \cdots \times I_M \times K_1 \times \cdots \times K_P}$ along the first M modes yields a tensor of size $J_1 \times \cdots \times J_N \times K_1 \times \cdots \times K_P$, given by

$$\langle \mathcal{A}, \mathcal{B} \rangle_{1:M;1:M}(j_1, \ldots, j_N, k_1, \ldots, k_P) = \sum_{i_1=1}^{I_1} \cdots \sum_{i_M=1}^{I_M} a_{i_1,\ldots,i_M,j_1,\ldots,j_N}\, b_{i_1,\ldots,i_M,k_1,\ldots,k_P}.$$

Example

Contraction along mode-2 of a tensor $\mathcal{A} \in \mathbb{R}^{3\times4\times5}$, and mode-3 of a tensor $\mathcal{B} \in \mathbb{R}^{7\times8\times4}$ returns a tensor $\mathcal{C} = \langle \mathcal{A}, \mathcal{B} \rangle_{2;3} \in \mathbb{R}^{3\times5\times7\times8}$.

Contracted product along all modes but mode-n

$$\langle \mathcal{A}, \mathcal{B} \rangle_{-n} = \mathbf{A}_{(n)}\, \mathbf{B}_{(n)}^T \in \mathbb{R}^{I_n \times J_n}, \qquad (I_k = J_k, \ \ \forall k \neq n).$$

Tensor contraction II

The contracted product of two three-way tensors $\mathcal{A} \in \mathbb{R}^{I \times J \times K}$ and $\mathcal{B} \in \mathbb{R}^{P \times Q \times R}$ along the mode-1 returns a four-way tensor defined as

$$\mathcal{C} \quad = \quad \langle \mathcal{A}, \mathcal{B} \rangle_1 \quad \in \mathbb{R}^{J \times K \times Q \times R}, \quad c_{jkqr} = \sum_i a_{ijk} b_{iqr}, \quad (I = P),$$

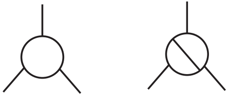
Contracted product along the two modes returns a matrix

$$\mathbf{F} \quad = \quad \langle \mathcal{A}, \mathcal{B} \rangle_{1,2} = \langle \mathcal{A}, \mathcal{B} \rangle_{-3} \quad \in \mathbb{R}^{K \times R}, \quad f_{kr} = \sum_{i,j} a_{ijk} b_{ijr},$$

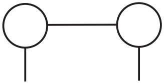
which can be expressed in a matrix multiplication form as $\mathbf{F} = \mathbf{A}_{(3)} \, \mathbf{B}_{(3)}^T$.
Inner product product of two tensors of the same dimension along all modes

$$\langle \mathcal{A}, \mathcal{B} \rangle_{1,\ldots,N} \quad = \quad \langle \mathcal{A}, \mathcal{B} \rangle.$$

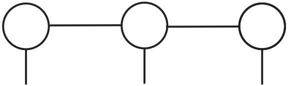
Graph representation of Tensor Network I



(a) Vector, matrix, order-3 tensor and diagonal tensor

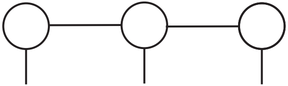


(b) Matrix multiplication
$y_{i,j} = \sum_r a_{i,r} b_{r,j}$



(c) TN of three tensors
$y_{i,j,k} = \sum_{r,s} a_{i,r} b_{r,j,s} c_{s,k}$

Graph representation of Tensor Network II



TN of three core tensors **A** of size $I \times R$, $\mathcal{B}$ of size $R \times J \times S$ and **C** of size $K \times S$

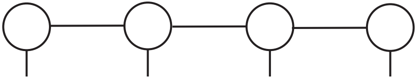
$$
\begin{aligned}
y_{i,j,k} &= \sum_{r,s} a_{i,r} b_{r,j,s} c_{k,s} \\
\mathcal{Y} &= \sum_{r,s} \mathbf{A}(:,r) \circ \mathcal{B}(r,:,s) \circ \mathbf{C}(:,s)
\end{aligned}
$$

Unfoldings

$$
\begin{aligned}
\mathbf{Y}_{(1)} &= \mathbf{A}[\mathcal{X}]_{(1)} = \mathbf{A}\mathbf{B}_{(1)}(\mathbf{C}^T \otimes \mathbf{I}) \\
\mathbf{Y}_{(2)} &= \mathbf{B}_{(2)}(\mathbf{C}^T \otimes \mathbf{A}^T) \\
\mathbf{Y}_{(3)} &= \mathbf{C}\mathbf{B}_{(3)}(\mathbf{I} \otimes \mathbf{A}^T)
\end{aligned}
$$

where $\mathcal{X}$ is a network of $\mathcal{B}$ and **C**.

Graph representation of Tensor Network III

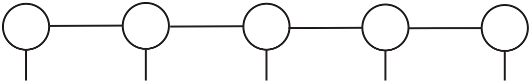


Tensor train

$$
\begin{aligned}
y_{i,j,k,l} &= \sum_{r,s,t} a_{i,r} b_{r,j,s} c_{s,k,t} d_{t,l} \\
\mathcal{Y} &= \sum_{r,s,t} \mathbf{A}_{:,r} \circ \mathcal{B}_{r,:,s} \circ \mathcal{C}_{s,:,t} \circ \mathbf{D}_{t,:}
\end{aligned}
$$

Matricization

$$
\begin{aligned}
\mathbf{Y}_{(1)} &= \mathbf{A}\mathbf{B}_{(1)}(\mathbf{C}_{(1)}(\mathbf{D} \otimes \mathbf{I}_K) \otimes \mathbf{I}_J) \\
&= \mathbf{A}\mathbf{B}_{(1)}(\mathbf{C}_{(1)} \otimes \mathbf{I}_J)(\mathbf{D} \otimes \mathbf{I}_{KJ}) \\
\mathbf{Y}_{(2)} &= \mathbf{B}_{(2)}(\mathbf{C}_{(1)}(\mathbf{D} \otimes \mathbf{I}_K) \otimes \mathbf{A}^T) \\
\mathbf{Y}_{(3)} &= \mathbf{C}_{(2)}(\mathbf{D} \otimes (\mathbf{B}_3(\mathbf{I}_K \otimes \mathbf{A}^T)) \\
\mathbf{Y}_{(4)} &= \mathbf{D}\mathbf{C}_{(3)}(\mathbf{I}_K \otimes \mathbf{B}_{(3)})(\mathbf{I}_{KJ} \otimes \mathbf{A}_{(3)})
\end{aligned}
$$



Tensor train

$$\mathcal{Y} = \sum_{r} \mathbf{A}_{:,r_1} \circ \mathcal{B}_{r_1,:,r_2} \circ \mathcal{C}_{r_2,:,r_3} \circ \mathcal{D}_{r_3,:,r_4} \circ \mathbf{E}_{r_4,:}$$

Mode-1 unfolding

$$\mathbf{Y}_{(1)} = \mathbf{A}\mathbf{B}_{(1)}(\mathbf{C}_{(1)} \otimes \mathbf{I}_{I_2})(\mathbf{D}_{(1)} \otimes \mathbf{I}_{I_3 I_2})(\mathbf{E}_{(1)} \otimes \mathbf{I}_{I_4 I_3 I_2})$$

Mode-2 unfolding

$$\begin{aligned} \mathbf{Y}_{(2)} &= \mathbf{B}_{(2)}(\mathbf{Y}_R \otimes \mathbf{Y}_L) \\ \mathbf{Y}_R &= \mathbf{C}_{(1)}(\mathbf{D} \otimes \mathbf{I}_{I_3})(\mathbf{E}_{(1)} \otimes \mathbf{I}_{I_4 I_3}) \\ \mathbf{Y}_L &= \mathbf{A}_{(3)} \end{aligned}$$

Graph representation of Tensor Network V

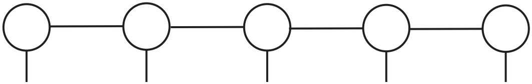
Mode-3 unfolding

$$
\begin{aligned}
\mathbf{Y}_{(3)} &= \mathbf{C}_{(2)}(\mathbf{Y}_R \otimes \mathbf{Y}_L) \\
\mathbf{Y}_R &= \mathbf{D}_{(1)}(\mathbf{E} \otimes \mathbf{I}_{I_4}) \\
\mathbf{Y}_L &= \mathbf{B}_{(3)}(\mathbf{I}_{I_2} \otimes \mathbf{A}_{(3)})
\end{aligned}
$$

Mode-4 unfolding

$$
\begin{aligned}
\mathbf{Y}_{(4)} &= \mathbf{D}_{(2)}(\mathbf{Y}_R \otimes \mathbf{Y}_L) \\
\mathbf{Y}_R &= \mathbf{E}_{(1)} \\
\mathbf{Y}_L &= \mathbf{C}_{(3)}(\mathbf{I}_{I_3} \otimes \mathbf{B}_{(3)})(\mathbf{I}_{I_3 I_2} \otimes \mathbf{A}_{(3)})
\end{aligned}
$$

Graph representation of Tensor Network VI



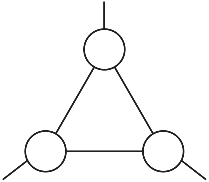
Mode-(1,2) unfolding

$$
\begin{aligned}
\mathbf{Y}_{(1,2)} &= \mathbf{Y}_L \mathbf{Y}_R \\
\mathbf{Y}_L &= (\mathbf{I}_{I_2} \otimes \mathbf{A}_{(3)}^T)\mathbf{B}_{(3)}^T \\
\mathbf{Y}_R &= \mathbf{C}_{(1)}(\mathbf{D}_{(1)} \otimes \mathbf{I}_{I_3})(\mathbf{E}_{(1)} \otimes \mathbf{I}_{I_3 I_4})
\end{aligned}
$$

Mode-(1-3) unfolding

$$
\begin{aligned}
\mathbf{Y}_{(1-3)} &= \mathbf{Y}_L \mathbf{Y}_R \\
\mathbf{Y}_L &= (\mathbf{I}_{I_2 I_3} \otimes \mathbf{A}_{(3)}^T)(\mathbf{I}_{I_3} \otimes \mathbf{B}_{(3)}^T)\mathbf{C}_{(3)}^T \\
\mathbf{Y}_R &= \mathbf{D}_{(1)}(\mathbf{E}_{(1)} \otimes \mathbf{I}_{I_4})
\end{aligned}
$$

Graph representation of Tensor Network VII



Tensor chain - Looped TN

$$
\begin{aligned}
y_{i,j,k} &= \sum_{r,s,t} a_{t,i,r} b_{r,j,s} c_{s,k,t} \\
\mathcal{Y} &= \sum_{r,s,t} \mathcal{A}(r,:,s) \circ \mathcal{B}(s,:,t) \circ \mathcal{C}(t,:,r)
\end{aligned}
$$

Graph representation of Tensor Network VIII

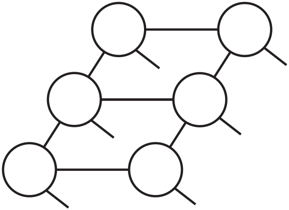
Unfolding

$$
\begin{aligned}
\mathbf{Y}_{(1)} &= \mathbf{A}_{(2),(3,1)} \left(\sum_r \mathcal{C}(r,:,:)^T \otimes \mathcal{B}(:,:,r) \right) \\
\mathbf{Y}_{(2)} &= \mathbf{B}_{(2),(1,3)} \left(\sum_r \mathcal{C}(:,:,r) \otimes \mathcal{A}(r,:,:)^T \right) \\
\mathbf{Y}_{(3)} &= \mathbf{C}_{(2),(3,1)} \left(\sum_r \mathcal{B}(r,:,:)^T \otimes \mathcal{A}(:,:,r) \right)
\end{aligned}
$$

Question:

Write unfoldings of a looped TN of four core tensors of order-3

Graph representation of Tensor Network IX



$$
\begin{aligned}
y_{i_1 i_2 \ldots i_6} &= \sum_{r_1 r_2 \ldots r_7} a_{i_1 r_7 r_1} b_{i_2 r_1, i_2, r_3} c_{i_3, r_3, r_4} d_{i_4, r_4, r_5} e_{i_5, r_5, r_2, r_4} f_{i_6, r_6, r_7} \\
\mathcal{Y} &= \sum_{r_1, r_2, \ldots, r_7} \mathcal{A}(r_1, :, r_2) \circ \mathcal{B}(r_2, :, r_3, r_4) \circ \mathcal{C}(r_4, :, r_5) \circ \\
&\quad \circ \mathcal{D}(r_5, :, r_6) \circ \mathcal{E}(r_6, :, r_3, r_7) \circ \mathcal{F}(r_7, :, r_1)
\end{aligned}
$$

First Example: Low-Rank Matrix Approximation I

- Consider an $I \times J$ matrix $\mathbf{X}$,
- Rank($\mathbf{X}$) = minimum R such that $\mathbf{X} = \sum_{r=1}^{R} \boldsymbol{a}_r \boldsymbol{b}_r^T$
- Low-rank approximation $\mathbf{X} \approx \mathbf{A}\mathbf{B}^T$

$$\min \quad \|\mathbf{X} - \mathbf{A}\mathbf{B}^T\|_F^2$$

  where $\mathbf{A} \in \mathbb{R}^{I \times R}$ and $\mathbf{B} \in \mathbb{R}^{J \times R}$, $R \leq I, J$
- Truncated SVD gives the best rank-R approximation

$$\mathbf{X} \approx \mathbf{U}\Sigma\mathbf{V}^T$$

  where $\mathbf{U}$ and $\mathbf{V}$ comprise leading singular vectors of $\mathbf{X}$, and $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_R)$.
- Low-rank decomposition of $\mathbf{X}$ is not unique
  Since for arbitrary non-singular $\mathbf{Q}$

$$\mathbf{A}\mathbf{B}^T = \mathbf{A}\mathbf{Q}\mathbf{Q}^{-1}\mathbf{B}^T = (\mathbf{A}\mathbf{Q})(\mathbf{B}\mathbf{Q}^{-1T})^T$$

First Example: Low-Rank Matrix Approximation II

- **Uniqueness**: if $\mathbf{A}$ is a mixing matrix, and $\mathbf{B}$ comprises source signals, separation of sources is not possible without incorporating additional information
  - Orthogonality is not sufficient
  - Nonnegativity is not sufficient.
    Nonnegative matrix factorization is not unique.
  - Statistical independence in Independent Component Analysis
  - Sparsity

Example I

Blind source separation for five linear mixtures of the sources

$$\begin{aligned} s_1(t) &= \sin(6\pi t) \\ s_2(t) &= \exp(10t)\sin(20\pi t), \end{aligned}$$

which were contaminated by white Gaussian noise, to give the mixtures

$$\mathbf{X} = \mathbf{AS} + \mathbf{E} \in \mathbb{R}^{5\times 60}$$

where $\mathbf{S}(t) = [s_1(t), s_2(t)]^T$ and $\mathbf{A} \in \mathbb{R}^{5\times 2}$ was a random matrix whose columns (mixing vectors) satisfy $\boldsymbol{a}_1^T \boldsymbol{a}_2 = 0.1$, $||\boldsymbol{a}_1|| = ||\boldsymbol{a}_2|| = 1$.

The 3Hz sine wave did not complete a full period over the 60 samples, so that the two sources had a correlation degree of $\frac{|\boldsymbol{s}_1^T \boldsymbol{s}_2|}{||\boldsymbol{s}_1||_2 ||\boldsymbol{s}_2||_2} = 0.35$.

Example II

- **PCA** failed since the mixing vectors were not orthogonal and the source signals were correlated, both violating the assumptions for PCA.
- **ICA** (using the JADE algorithm Cardoso and Souloumiac (1993)) failed because the signals were not statistically independent, as assumed in ICA.
- **Low rank tensor approximation:** a rank-2 CPD was used to estimate $\mathbf{A}$ as the third factor matrix, which was then inverted to yield the sources. The accuracy of CPD was compromised as the components of tensor $\mathcal{X}$ cannot be represented by rank-1 terms.

Example III

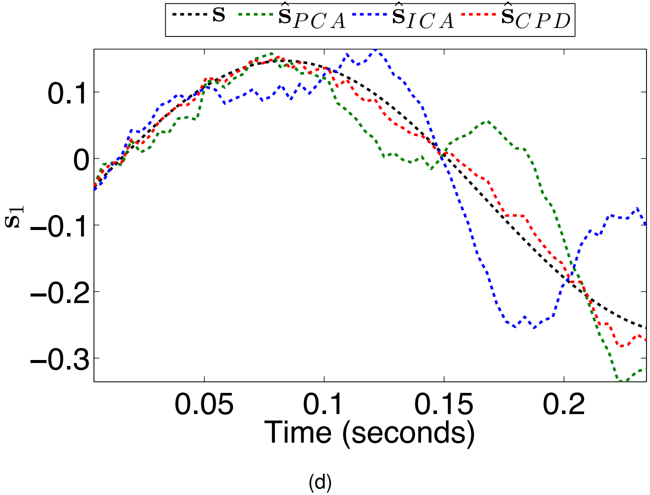
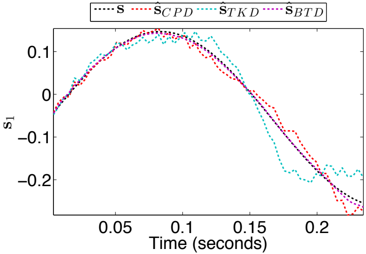


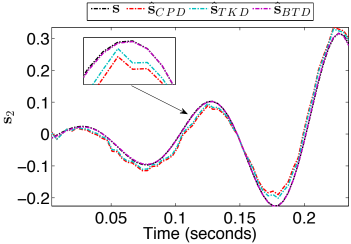
Figure: Blind separation of the mixture of a pure sine wave and an exponentially modulated sine wave using PCA, ICA, CPD. The sources

Example IV

s_1 and s_2 are correlated and of short duration; the symbols $\hat{s}_1$ and $\hat{s}_2$ denote the estimated sources.

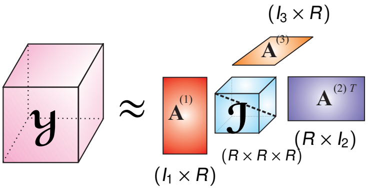


(a)



(b)

Canonical Polyadic Decomposition - PARAFAC



- $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is explained by N factor matrices $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$

$$
\begin{aligned}
\mathcal{Y} &\approx \sum_{r=1}^{R} \boldsymbol{a}_r^{(1)} \circ \boldsymbol{a}_r^{(2)} \circ \cdots \circ \boldsymbol{a}_r^{(N)} \qquad \|\boldsymbol{a}_r^{(n)}\|_2 = 1, \forall n \neq N \\
&= \mathcal{I} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)} = \hat{\mathcal{Y}}
\end{aligned}
$$

Rank and Border Rank I

- **Tensor Rank** the minimal rank-one tensor terms to fully represent the tensor
  The following tensor of size $2 \times 2 \times 2$

  $\mathcal{A} = \boldsymbol{a}_1 \circ \boldsymbol{a}_2 \circ \boldsymbol{a}_3 + \boldsymbol{a}_1 \circ \boldsymbol{b}_2 \circ \boldsymbol{b}_3 - \boldsymbol{b}_1 \circ \boldsymbol{a}_2 \circ \boldsymbol{b}_3 + \boldsymbol{b}_1 \circ \boldsymbol{b}_2 \circ \boldsymbol{a}_3,$

  has rank-3 over the real and rank-2 and a complex rank 3

  $$\mathcal{A} = \frac{1}{2}(\bar{\mathbf{z}}_1 \circ \mathbf{z}_2 \circ \bar{\mathbf{z}}_3 + \mathbf{z}_1 \circ \bar{\mathbf{z}}_2 \circ \mathbf{z}_3),$$

  $\boldsymbol{z}_k = \boldsymbol{a}_k + i\boldsymbol{b}_k$

Rank and Border Rank II

- **Border Rank**
  If there exists a sequence of tensors of rank at most $r < s$ whose limit is $\mathcal{A}$, the least value of s is the border rank of $\mathcal{A}$.
  The following tensor is of rank-3 but its border rank 2

$$\mathcal{A} = \boldsymbol{u} \circ \boldsymbol{u} \circ \boldsymbol{v} + \boldsymbol{u} \circ \boldsymbol{v} \circ \boldsymbol{u} + \boldsymbol{v} \circ \boldsymbol{u} \circ \boldsymbol{u}$$

with $||\boldsymbol{u}|| = ||\boldsymbol{v}|| = 1$ and $\boldsymbol{u}^T \boldsymbol{v} \neq 1$,
can be approximated with an arbitrary precision by the following sequence of rank-2 tensors $\mathcal{A}_n$ as $n \to \infty$

$$\begin{aligned}
\mathcal{A}_n &= n(\boldsymbol{u} + \frac{1}{n}\boldsymbol{v}) \circ (\boldsymbol{u} + \frac{1}{n}\boldsymbol{v}) \circ (\boldsymbol{u} + \frac{1}{n}\boldsymbol{v}) - n\boldsymbol{u} \circ \boldsymbol{u} \circ \boldsymbol{u} \\
&= \boldsymbol{u} \circ \boldsymbol{u} \circ \boldsymbol{v} + \boldsymbol{u} \circ \boldsymbol{v} \circ \boldsymbol{u} + \boldsymbol{v} \circ \boldsymbol{u} \circ \boldsymbol{u} \\
&\quad + \frac{1}{n}(\boldsymbol{u} \circ \boldsymbol{v} \circ \boldsymbol{v} + \boldsymbol{v} \circ \boldsymbol{u} \circ \boldsymbol{v} + \boldsymbol{v} \circ \boldsymbol{v} \circ \boldsymbol{u}) + \frac{1}{n^2}\boldsymbol{v} \circ \boldsymbol{v} \circ \boldsymbol{v}
\end{aligned}$$

Rank and Border Rank III

- **Diverging component - Degeneracy**
  Whenever $\mathcal{A}_n \to \mathcal{A}$ as $n \to \infty$, there should exist at least $1 \leq i \neq j \leq r$ such that

$$||\mathbf{a}_{i,n}^1 \circ \mathbf{a}_{i,n}^2 \circ \cdots \circ \mathbf{a}_{i,n}^d||_F \to \infty$$

  Degeneracy is often encountered when attempting to approximate a tensor using numerical optimization algorithms

Kruskal Rank and Uniqueness of CPD I

Definition (**Kruskal rank)**

A matrix **A** has k-rank k_A if and only if every subset of columns of **A** is full column rank, and this does not hold true for $k_A + 1$.

- **K-rank and the rank of a matrix**
  A matrix of rank-R, there is at least one subset of R linearly independent columns. In a matrix of k-rank k_A, every subset of k_A columns is of rank k_A.
- **Sufficient condition for the essential uniqueness of CPD**
  A d-th order tensor admits an essentially unique CPD if

$$\sum_{k=1}^{d} k - rank\{\mathbf{A}^{(d)}\} \geq 2R + d - 1$$

Kruskal Rank and Uniqueness of CPD II

- For generic matrices of size $I_n \times R$, the k-rank equals $\min(I_n, R)$. If matrices all have more columns than rows, the sufficient condition can be simplified to

$$\sum_{k=1}^{d} I_n \geq 2R + d - 1$$

Matrix Multiplication Algorithms I

Consider two matrices **E** and **F** of the sizes $P \times Q$ and $Q \times S$, respectively

Their matrix product $\mathbf{G} = \mathbf{EF}$ of the size $P \times S$

The matrix multiplication can be represented by a tensor $\mathcal{T}$ of the size $PQ \times QS \times PS$ such that

$$\mathrm{vec}(\mathcal{G}) = \mathcal{T} \times_1 \mathrm{vec}\left(\mathbf{E}^T\right)^T \times_2 \mathrm{vec}\left(\mathbf{F}^T\right)^T$$

$$\mathbf{T}_{(1)} = \left[\begin{array}{cccc|cccc|cccc|cccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

The tensor $\mathcal{T}$ has a CP representation of rank-7

$$\mathcal{T} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & -1 \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} 1 & 1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & -1 & 0 & 1 & 0 & 1 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 1 & -1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 1 & 0 \end{bmatrix},$$

Then

Matrix Multiplication Algorithms III

$$
\begin{aligned}
\mathrm{vec}(\mathcal{G}) &= \mathcal{T} \times_1 \mathrm{vec}\left(\mathbf{E}^T\right)^T \times_2 \mathrm{vec}\left(\mathbf{F}^T\right)^T \\
&= [\![\mathrm{vec}\left(\mathbf{E}^T\right)^T \mathbf{A}, \mathrm{vec}\left(\mathbf{F}^T\right)^T \mathbf{B}, \mathbf{C}]\!] \\
&= \mathbf{C}((\mathrm{vec}\left(\mathbf{E}^T\right)^T \mathbf{A}) \circledast (\mathrm{vec}\left(\mathbf{F}^T\right)^T \mathbf{B}))
\end{aligned}
$$

Kruskal Tensor: Properties I

Lemma

For a Kruskal tensor $\mathcal{Y}$ built up from N factors $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$

$$\mathcal{Y} = \mathcal{I} \times_1 \mathbf{A}^{(1)} \cdots \times_N \mathbf{A}^{(N)}. \tag{14}$$

1. $\sum_{i_1, i_2, \ldots, i_N} y_{i_1 i_2 \ldots i_N} = \left(\{\mathbf{1}^T \mathbf{A}\}^{\circledast} \right) \mathbf{1}.$
2. $\sum_{i_1, i_2, \ldots, i_N} y^2_{i_1 i_2 \ldots i_N} = \mathbf{1}^T \left(\{\mathbf{A}^T \mathbf{A}\}^{\circledast} \right) \mathbf{1}.$

Proof.

Kruskal Tensor: Properties II

Summation of all the entries of the tensor $\mathcal{Y}$ is given by

$$\begin{aligned}
\sum y_i &= \mathbf{1}^T \operatorname{vec}(\mathcal{Y}) \\
&= \left(\mathbf{1}_{I_N} \odot \mathbf{1}_{I_{N-1}} \odot \cdots \odot \mathbf{1}_{I_1}\right)^T \left(\mathbf{A}^{(N)} \odot \mathbf{A}^{(N-1)} \odot \cdots \odot \mathbf{A}^{(1)}\right) \mathbf{1} \\
&= \left(\left(\mathbf{1}^T \mathbf{A}^{(N)}\right) \circledast \left(\mathbf{1}^T \mathbf{A}^{(N-1)}\right) \circledast \cdots \circledast \left(\mathbf{1}^T \mathbf{A}^{(1)}\right)\right) \mathbf{1}\,. \qquad (15)
\end{aligned}$$

Frobenius norm of the tensor $\mathcal{Y}$

$$\begin{aligned}
\|\mathcal{Y}\|_F^2 &= \operatorname{vec}(\mathcal{Y})^T \operatorname{vec}(\mathcal{Y}) \\
&= \mathbf{1}^T \{\mathbf{A}\}^{\odot T} \{\mathbf{A}\}^{\odot} \mathbf{1} = \mathbf{1}^T \left\{\mathbf{A}^T \mathbf{A}\right\}^{\circledast} \mathbf{1}\,. \qquad (16)
\end{aligned}$$

□

Kruskal Tensor: Properties III

Unfolding. Mode-n matricization of a Krusal tensor $\mathcal{Y} = [\![\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)}]\!]$

$$\begin{aligned} \mathbf{Y}_{(n)} &= \mathbf{A}^{(n)} \left(\mathbf{A}^{(N)} \odot \cdots \odot \mathbf{A}^{(n+1)} \odot \mathbf{A}^{(n-1)} \cdots \odot \mathbf{A}^{(1)}\right)^T \\ &= \mathbf{A}^{(n)} \left\{\mathbf{A}\right\}^{\odot_{-n}T}, \qquad (n = 1, 2, \ldots, N), \end{aligned} \tag{17}$$

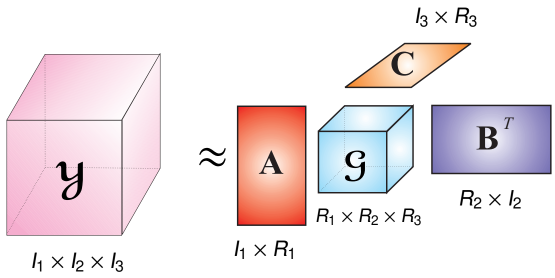
or as a summation of rank-one approximations

$$\begin{aligned} \mathbf{Y}_{(n)} &= \sum_{r=1}^{R} \boldsymbol{a}_r^{(n)} \left(\boldsymbol{a}_r^{(N)} \otimes \cdots \otimes \boldsymbol{a}_r^{(n+1)} \otimes \boldsymbol{a}_r^{(n-1)} \otimes \cdots \otimes \boldsymbol{a}_r^{(1)}\right)^T \\ &= \sum_{r=1}^{R} \boldsymbol{a}_r^{(n)} \left(\boldsymbol{a}_r^{\otimes_{-n}}\right)^T . \end{aligned} \tag{18}$$

$$\mathcal{Y}_{(n,m)} = [\![\mathbf{A}^{(n)}, \mathbf{A}^{(m)}, \mathbf{B}^{(nm)}]\!] \tag{19}$$

where $\mathbf{B}^{(nm)} = \bigodot_{k \neq n,m} \mathbf{A}^{(k)}$

TUCKER Decomposition



$$\begin{aligned}\mathcal{Y} &\approx \sum_{j_1=1}^{R_1}\sum_{j_2=1}^{R_2}\cdots\sum_{j_N=1}^{R_N} g_{j_1 j_2 \cdots j_N}\left(\boldsymbol{a}_{j_1}^{(1)} \circ \boldsymbol{a}_{j_2}^{(2)} \circ \cdots \circ \boldsymbol{a}_{j_N}^{(N)}\right)\\ &= \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)} = \mathcal{G} \times \{\mathbf{A}\}\end{aligned}$$

References I

Cardoso, J.-F. and Souloumiac, A. (1993). Blind beamforming for non-Gaussian signals. In *IEE Proceedings F (Radar and Signal Processing)*, volume 140, pages 362–370. IET.