Introduction to (third-order) tensors

Moulay Barkatou

A tensor is a mathematical object that generalizes scalars, vectors, and matrices to higher dimensions. In machine learning, tensors are used to represent multi-dimensional data. For example, a picture can be represented as a 3D tensor, with each pixel of the picture represented by a single value in the tensor.
In this lecture we introduce notation, definitions and operations on tensors: vectorization, matricization, inner product, norm, n-mode product, outer product, Kronecker, and Khatri-Rao products.

1 Definitions

1.1 Tensors

Definition: A Tensor is a *multi-dimensional* or *multi-way* array: $\mathcal{A} = (a_{i_1,i_2,\cdots,i_N}) \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. The *order* of a tensor is the number of its dimensions (ways). Each dimension (way) is called a *mode*.

Remark: Scalars are tensors of order 0, vectors are tensors of order 1 and matrices are tensors of order 2. A third-order tensor is a "cube".

Modes, fibers and slices of a tensor

- An order-N tensor has N modes. To refer to the nth mode, we use mode-n or n-mode.
- Vectors along any of the N modes are called *fibers*. Thus, a fiber is defined by fixing every index but one. The fibers of a tensor $\mathcal{A} = (a_{i_1,i_2,\cdots,i_N}) \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ are denoted by $\mathbf{a}_{i_1 i_2 \cdots i_{n-1} : i_{n+1} \cdots i_N}$.
- The *slices* of a tensor are defined by fixing all but two indices. The slices of a tensor $\mathcal{A} = (a_{i_1,i_2,\cdots,i_N})$ are denoted by $\mathcal{A}_{i_1 i_2 \cdots i_{n-1} : i_{n+1} \cdots i_{k-1} : i_{k+1} \cdots i_N}$.

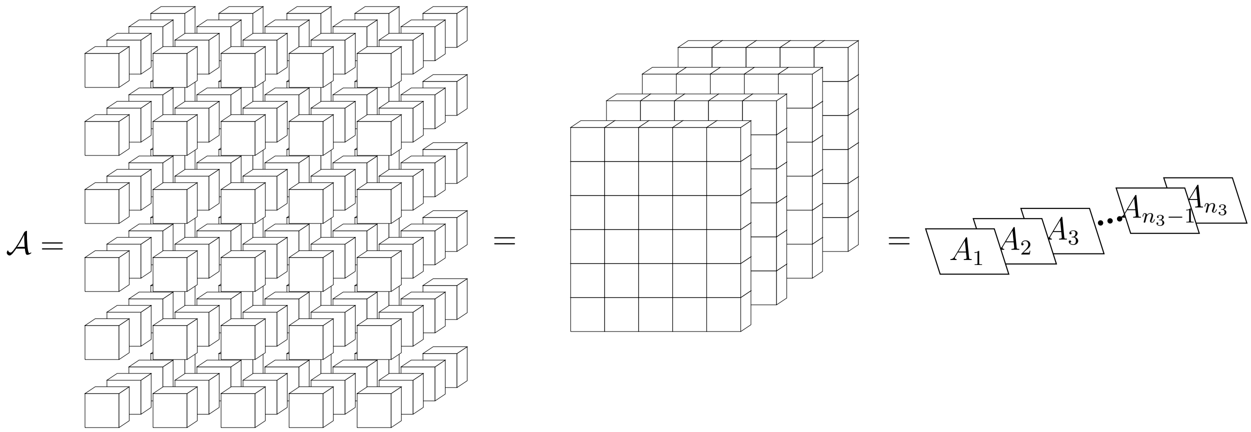
Remark:

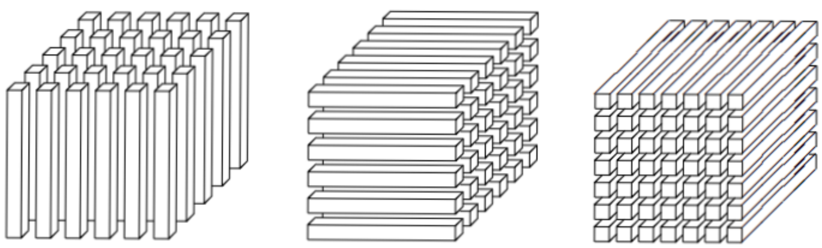
- A matrix column is a mode-1 fiber and a matrix row is a mode-2 fiber.

- A 3$-$way tensor has three modes: *columns, rows*, and *tubes fibers.*

- The frontal slices of a third-order tensor $\mathcal{A} \in \mathbb{R}^{I \times J \times K} = (a_{ijk})$ are simply denoted as $A_i := a_{(:,:,i)}$.

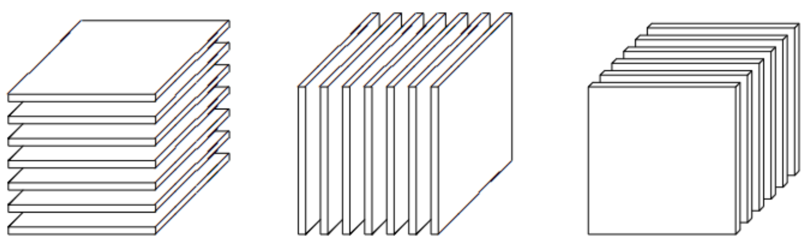
A tensor $\mathcal{A}$ can be visually depicted in the following ways:





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

Fibers of 3-order tensor



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

Slides of an order-3 tensor

1.2 Tensor Matricization

Definition: *Matricization* (also called *unfolding* or *flattening*) consists in rearranging the entries of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ so that it can be represented as a matrix.

Mode-n matricization arranges the mode-n fibers as columns of an $I_n \times \Pi_{\ell \neq n} I_\ell$ matrix, denoted $\mathbf{A}_{(n)}$.

For an order-3 tensor $\mathcal{A} \in \mathbb{R}^{I \times J \times K} = (a_{ijk})$, we have three modes of matricization:

$$\mathbf{A}_{(1)} \in \mathbb{R}^{I \times JK}, \mathbf{A}_{(2)} \in \mathbb{R}^{J \times IK}, \mathbf{A}_{(3)} \in \mathbb{R}^{K \times IJ}$$

Definition:

- *Vectorization* of an $m \times n$ matrix M consists in stacking its columns to obtain a vector of size mn, denoted by $\mathrm{vec}(M)$. Specifically, if we denote by $M_{.j}$ the jth column of M then

$$\mathrm{vec}(M) = \begin{bmatrix} M_{.1} \\ M_{.2} \\ \vdots \\ M_{.n} \end{bmatrix}$$

- *Vectorization* of a tensor is a process that transforms the tensor into a vector. One way to do it consists in first flattening the tensor along the first mode, and then vectorizing columns of the resulting matrix to obtain a vector. Specifically,

$$\mathrm{vec}(\mathcal{A}) = \mathrm{vec}(\mathbf{A}_{(1)}).$$

Example: Let $\mathcal{A} \in \mathbb{R}^{2 \times 2 \times 2}$ with $\mathcal{A}_{(:,:,1)} = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}$ and $\mathcal{A}_{(:,:,2)} = \begin{bmatrix} 0 & 1 \\ -1 & 4 \end{bmatrix}$ then,

$$\mathbf{A}_{(1)} = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 3 & 1 & -1 & 4 \end{bmatrix}, \ \mathbf{A}_{(2)} = \begin{bmatrix} 1 & 3 & 0 & -1 \\ 2 & 1 & 1 & 4 \end{bmatrix}, \ \mathbf{A}_{(3)} = \begin{bmatrix} 1 & 3 & 2 & 1 \\ 0 & -1 & 1 & 4 \end{bmatrix}.$$

$$\mathrm{vec}(\mathcal{A}) = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 1 \\ 0 \\ -1 \\ 1 \\ 4 \end{bmatrix}$$

1.3 Operations over tensors

Various operations can be performed over tensors.

1.3.1 Tensor sum

Definition:The sum of two tensors of the same size is defined component- wise:

If $\mathcal{A} = (a_{ijk}), \mathcal{B} = (b_{ijk}) \in \mathbb{R}^{I\times J\times K}$ then

$$(\mathcal{A}+\mathcal{B})_{ijk} = a_{ijk} + b_{ijk}.$$

1.3.2 Tensor Inner Product and Tensor Norm

Definition: The *inner product* of two tensors of $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I\times J\times K}$ is given by

$$\langle \mathcal{A}, \mathcal{B} \rangle := \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} a_{ijk} b_{ijk}.$$

The *Frobenius norm* is the norm associated to this inner product

$$\|\mathcal{A}\|_F = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle} = \sqrt{\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} |a_{i,j,k}|^2}.$$

One has :

$$||\, \mathcal{A}\, ||_F = \left\|\mathbf{A}_{(1)}\right\|_F = \left\|\mathbf{A}_{(2)}\right\|_F = \left\|\mathbf{A}_{(3)}\right\|_F$$

1.3.3 Vector outer Product

The *outer product* (or tensor product), is defined as follows.

- The outer product of two vectors $\mathbf{u} \in \mathbb{R}^m$ and $\mathbf{v} \in \mathbb{R}^n$ produces a matrix denoted by $\mathbf{u} \circ \mathbf{v} \in \mathbb{R}^{m\times n}$:

  $$(\mathbf{u} \circ \mathbf{v})_{ij} = \mathbf{u}_i \mathbf{v}_j$$

  In terms of the usual matrix product, one has

  $$\mathbf{u} \circ \mathbf{v} = \mathbf{u}\mathbf{v}^T.$$

- The outer product of three vectors produces a 3-dimensional tensor.

  Let $\mathbf{u} \in \mathbb{R}^I, \mathbf{v} \in \mathbb{R}^J$, and $\mathbf{w} \in \mathbb{R}^K$. Their outer product is given by:

  $$\mathbf{u} \circ \mathbf{v} \circ \mathbf{w} = \mathcal{A} = (a_{ijk}) \in \mathbb{R}^{I\times J\times K}, \text{ where } a_{i,j,k} = \mathbf{u}_i \mathbf{v}_j \mathbf{w}_k$$

- The outer product of N vectors $\mathbf{u}^{(n)} \in \mathbb{R}^{I_n}$ $(n = 1, \ldots, N)$, is given by

$$\mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \cdots \circ \mathbf{u}^{(N)} = \mathcal{A} = (a_{i_1 i_2 \cdots i_N}) \in \mathbb{R}^{I_1 \times I_2 \times \cdots I_N}, \text{ where } a_{i_1 i_2 \cdots i_N} = \mathbf{u}^{(1)}_{i_1} \mathbf{u}^{(2)}_{i_2} \cdots \mathbf{u}^{(N)}_{i_N}$$

Examples

1. The outer product of three vectors $a, b, c \in \mathbb{R}^2$

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \circ \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \circ \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 c_1 & a_1 b_2 c_1 \\ a_2 b_1 c_1 & a_2 b_2 c_1 \end{pmatrix} \begin{pmatrix} a_1 b_1 c_2 & a_1 b_2 c_2 \\ a_2 b_1 c_2 & a_2 b_2 c_2 \end{pmatrix}$$

2. For instance,

$$\begin{pmatrix} 1 \\ -1 \end{pmatrix} \circ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \circ \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$$

3. Let

$$\mathbf{u} = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} \in \mathbb{R}^3, \quad \mathbf{v} = \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix} \in \mathbb{R}^3, \quad \mathbf{w} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \in \mathbb{R}^2.$$

Then

$$\mathbf{u} \circ \mathbf{v} = \begin{pmatrix} 0 & 0 & 0 \\ 6 & 2 & 8 \\ 3 & 1 & 4 \end{pmatrix} \in \mathbb{R}^{3 \times 3}.$$

And $\mathcal{A} := u \circ v \circ w \in \mathbb{R}^{3 \times 3 \times 2}$, is given by its frontal slices:

$$\mathcal{A}(:, :, 1) = \begin{pmatrix} 0 & 0 & 0 \\ 12 & 4 & 16 \\ 6 & 2 & 8 \end{pmatrix}, \quad \mathcal{A}(:, :, 2) = \begin{pmatrix} 0 & 0 & 0 \\ 6 & 2 & 8 \\ 3 & 1 & 4 \end{pmatrix}.$$

Definition: An order-N tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ is called a *rank one* tensor if it can be written as the outer product of N vectors, i.e., $\mathcal{A} = \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \cdots \circ \mathbf{u}^{(N)}$.

1.3.4 Tensor times matrix

Let $\mathcal{A} = (a_{i_1, i_2, \cdots, i_N}) \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ and $U \in \mathbb{R}^{J \times I_n}$.

- The **mode-n matrix product** of $\mathcal{A}$ by U is the tensor $(\mathcal{A} \times_n U) \in \mathbb{R}^{I_1 \times \cdots I_{n-1} \times J \times I_{n+1} \cdots \times I_N}$ defined by

$$(\mathcal{A} \times_n U)_{i_1 \cdots i_{n-1} j i_{n+1} \cdots i_N} = \sum_{i_n = 1}^{I_n} a_{i_1 i_2 \cdots i_N} U_{j i_n}.$$

- Let $\mathcal{B} = \mathcal{A} \times_n U$ then each mode-n fibre of $\mathcal{B}$ is obtained by multiplying the corresponding mode-n fibre of $\mathcal{A}$ by the matrix U:

$$\mathbf{b}_{i_1 i_2 \cdots i_{n-1} : i_{n+1} \cdots i_N} = U\, \mathbf{a}_{i_1 i_2 \cdots i_{n-1} : i_{n+1} \cdots i_N}$$

- In terms of mode-n matricization, one has

$$\mathcal{B} = \mathcal{A} \times_n U \iff \mathbf{B}_{(n)} = U \mathbf{A}_{(n)}.$$

Example: Let $\mathcal{A} \in \mathbb{R}^{2\times 2\times 2}$ be the order-3 tensor given by

$$\mathcal{A}(:,:,1) = \begin{pmatrix} 2 & 4 \\ 3 & 5 \end{pmatrix}, \quad \mathcal{A}(:,:,2) = \begin{pmatrix} 1 & 5 \\ 0 & 3 \end{pmatrix}$$

and let

$$U = \begin{pmatrix} 6 & 3 \\ 1 & 1 \\ 0 & 4 \end{pmatrix} \in \mathbb{R}^{3\times 2}$$

We compute the mode-1 product $\mathcal{B} = \mathcal{A} \times_1 U \in \mathbb{R}^{3\times 2\times 2}$ by computing its mode-1 matricization:

$$B_{(1)} = U\, A_{(1)} = \begin{pmatrix} 6 & 3 \\ 1 & 1 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} 2 & 4 & 1 & 5 \\ 3 & 5 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 21 & 39 & 6 & 39 \\ 5 & 9 & 1 & 8 \\ 12 & 20 & 0 & 12 \end{pmatrix}.$$

So, the frontal slices of $\mathcal{B}$ are :

$$\mathcal{B}(:,:,1) = \begin{pmatrix} 21 & 39 \\ 5 & 9 \\ 12 & 20 \end{pmatrix}, \quad \mathcal{B}(:,:,2) = \begin{pmatrix} 6 & 39 \\ 1 & 8 \\ 0 & 12 \end{pmatrix}.$$

Remark: If $\mathcal{A}$ is a matrix (an order-2 tensor), U and V are matrices of appropriate size then

$$\mathcal{A} \times_1 U = UA$$
$$\mathcal{A} \times_2 V = AV^\top$$

Proposition: For matrices of appropriate size, the following properties hold

$$(\mathcal{A} \times_n U) \times_n V = \mathcal{A} \times_n (VU)$$

$$(\mathcal{A} \times_n U) \times_m V = \mathcal{A} \times_m V \times_n U \quad \textit{for } m \neq n$$

1.3.5 Tensor Times Vector

- The mode-1 multiplication of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ by a vector $\mathbf{v} \in \mathbb{R}^{I_1}$, is the tensor denoted $\mathcal{A} \,\bar{\times}_1\, \mathbf{v}$ and defined as

$$\mathcal{A} \,\bar{\times}_1\, \mathbf{v} = \mathcal{A} \times_1 \mathbf{v}^\top \in \mathbb{R}^{I_2 \times I_3}.$$

- For $1 \leq n \leq N$, the **mode-n vector product** of a tensor $\mathcal{A} = (a_{i_1,i_2,\cdots,i_N}) \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ by a vector $\mathbf{v} \in \mathbb{R}^{I_n}$ is the tensor $(\mathcal{A} \,\bar{\times}_n\, \mathbf{v}) \in \mathbb{R}^{I_1 \times \cdots I_{n-1} \times I_{n+1} \cdots \times I_N}$ defined by

$$(\mathcal{A} \,\bar{\times}_n\, \mathbf{v})_{i_1 \cdots i_{n-1} i_{n+1} \cdots i_N} = \sum_{i_n=1}^{I_n} a_{i_1 i_2 \cdots i_N} \mathbf{v}_{i_n}.$$

1.4 Hadamard, Kronecker and Khatri-Rao matrix products

Definition: [Hadamard Product] For two matrices A and B of the same size $m \times n$, the Hadamard product $A \star B$ is a matrix of the same size $m \times n$ with entries given by:

$$(A \star B)_{ij} = a_{ij} b_{ij}.$$

$$A \star B = \begin{bmatrix} a_{11}b_{11} & \cdots & a_{1n}b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1}b_{m1} & \cdots & a_{mn}b_{mn} \end{bmatrix}$$

Properties: The Hadamard product is commutative, associative and distributive with respect to addition. It has as identity the $m \times n$ matrix whose entries are equal to 1.

Definition: [Kronecker Product] For two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times q}$, the Kronecker product $A \otimes B \in \mathbb{R}^{mp \times nq}$ is defined by

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}$$

Remark:

- For two columns vectors $\mathbf{a} \in \mathbb{R}^{m \times 1}$ and $\mathbf{b} \in \mathbb{R}^{n \times 1}$ the Kronecker product $\mathbf{a} \otimes \mathbf{b} \in \mathbb{R}^{mn \times 1}$ and satisfies:

$$\mathbf{a} \otimes \mathbf{b} = \operatorname{vec}\left(\mathbf{b}\mathbf{a}^T\right) = \operatorname{vec}\left(\mathbf{b} \circ \mathbf{a}\right).$$

- If we denote the columns of A by $\mathbf{a}_1, \cdots, \mathbf{a}_n$ and the columns of B by $\mathbf{b}_1, \cdots, \mathbf{b}_q$, then the columns of $A \otimes B$ are given by $\mathbf{a}_j \otimes \mathbf{b}_\ell$ for $j = 1, \ldots, n$ and $\ell = 1, \ldots, q$:

$$A \otimes B = [\mathbf{a}_1 \otimes \mathbf{b}_1 \;\; \mathbf{a}_1 \otimes \mathbf{b}_2 \;\; \mathbf{a}_1 \otimes \mathbf{b}_3 \;\; \cdots \;\; \mathbf{a}_n \otimes \mathbf{b}_{\ell-1} \;\; \mathbf{a}_n \otimes \mathbf{b}_\ell] \, .$$

Properties of Kronecker Product:

1. The Kronecker product is bilinear and associative but it is not commutative.
2. Transposition (and Hermitian transposition) is distributive over the Kronecker product:

$$(A \otimes B)^\top = A^\top \otimes B^\top \quad (A \otimes B)^* = A^* \otimes B^*$$

3. Assuming compatible dimensions of the matrices, we have the following properties:

$$(A \otimes B)(C \otimes D) = AC \otimes BD$$

$$(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger$$

where $A^\dagger$ denotes the Moore-Penrose pseudoinverse of A.

In particular for invertible matrices A and B one has

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$$

4.

$$\mathrm{vec}\,(ACB) = (B^\top \otimes A)\mathrm{vec}\,(C)$$

5. Given an order-N tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ and N matrices $U^{(n)} \in \mathbb{R}^{J_n \times I_n}$, one has for for all $n \in \{1, \ldots, N\}$:

$$\mathcal{B} = \mathcal{A} \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_N U^{(N)} \iff B_{(n)} = U^{(n)} A_{(n)} \left(U^{(N)} \otimes \cdots \otimes U^{(n+1)} \otimes U^{(n-1)} \otimes \cdots \otimes U^{(1)} \right)^\top$$

6. For order-3 tensors:

$$\mathcal{B} = \mathcal{A} \times_1 U \times_2 V \times_3 W \iff \begin{cases} B_{(1)} & = U A_{(1)} (W \otimes V)^\top \\ B_{(2)} & = V A_{(2)} (W \otimes U)^\top \\ B_{(3)} & = W A_{(3)} (V \otimes U)^\top . \end{cases}$$

Exercise: Show the following properties

1. $tr(A \otimes B) = tr(A) tr(B)$
2. $\|A \otimes B\|_F = \|A\|_F \, \|B\|_F$
3. $rank(A \otimes B) = rank(A) rank(B)$

4. If A and B are positive-definite, then so is $A \otimes B$.

Definition: [Khatri-Rao Product] The Khatri-Rao product is defined for matrices having the same number of columns. If $A = \begin{bmatrix}\mathbf{a}_1 & \dots & \mathbf{a}_n\end{bmatrix} \in \mathbb{R}^{m\times n}, B = \begin{bmatrix}\mathbf{b}_1 & \dots & \mathbf{b}_n\end{bmatrix} \in \mathbb{R}^{p\times n}$, then the Khatri-Rao product denoted by $A \odot B$ is defined as:

$$A \odot B = [\mathbf{a}_1 \otimes \mathbf{b}_1 \;\; \mathbf{a}_2 \otimes \mathbf{b}_2 \;\; \cdots \;\; \mathbf{a}_n \otimes \mathbf{b}_n] \in \mathbb{R}^{mp\times n}$$

$$A \odot B = \begin{bmatrix} a_{11}\mathbf{b}_1 & a_{12}\mathbf{b}_2 & \cdots & a_{1n}\mathbf{b}_n \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\mathbf{b}_1 & a_{m2}\mathbf{b}_2 & \cdots & a_{mn}\mathbf{b}_n \end{bmatrix}$$

Properties of Khatri-Rao Product:

1. The Khatri-Rao product is bilinear and associative.
2. $$(A \odot B)^T (A \odot B) = (A^\top A) \star (B^\top B)$$
3. $$(A \odot B)^\dagger = \left((A^\top A) \star (B^\top B)\right)^\dagger (A \odot B)^T$$
4. $$\operatorname{vec}(ACB) = (B^\top \odot A)\operatorname{diag}(C)$$

   where diag($\mathbf{C}$) is a vector representing the main diagonal of C.

2 Tensor Decompositions

Recall that a Singular Value Decomposition of a matrix $A \in \mathbb{R}^{m\times n}$ is a decomposition of A of the form

$$A = U\Sigma V^\top$$

where U is an $m \times m$ orthogonal matrix, V an $n \times n$ orthogonal matrix and where Σ is a diagonal $m \times n$ matrix with non-negative real diagonal coefficient (the singular values of A.

The SVD of a matrix A can be formulated in the two following ways

- Using the outer product

  $$A = \sum_{i=1}^{r} \sigma_i \, (\mathbf{u}_i \circ \mathbf{v}_i)$$

  where $r = \operatorname{rank}(A)$, σ_i are the positive singular values of A, $\mathbf{u}_i$ and $\mathbf{v}_i$ are the left and right singular vectors (columns of U and V) respectively. and

- Using the n-mode products, then

$$A = \Sigma \times_1 U \times_2 V.$$

This can be extended to a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ as follows:

$$\mathcal{A} = \sum_{i=1}^{r_1} \sum_{j=1}^{r_2} \sum_{k=1}^{r_3} \sigma_{ijk} \ (\mathbf{u}_i \circ \mathbf{v}_j \circ \mathbf{w}_k) \quad \text{and} \quad \mathcal{A} = \mathcal{S} \times_1 U \times_2 V \times_3 W \tag{1}$$

where $\mathcal{S} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ is called the core tensor and $U \in \mathbb{R}^{n_1 \times r_1}, V \in \mathbb{R}^{n_2 \times r_2}, W \in \mathbb{R}^{n_2 \times r_3}$ are matrices.

However, it is not true that there always exists such a decomposition with the same requirements on the tensor $\mathcal{S}$ and the matrices U, V and W as in the usual SVD. We need to relax the constraints.

Here are three well-known extensions of SVD to tensors.

- The *CANDECOMP-PARAFAC decomposition (CP)* is given by (1) with the condition that $\mathcal{S}$ has to be a *hyper-diagonal* tensor (which means that $\sigma_{ijk} = 0$ if i, j and k are not equal) and hence the factor matrices U,V, and W are required to have the same number of columns. In that case, we can write $\mathcal{A}$ as a sum of rank-one tensors:

$$\mathcal{A} = \sum_{i=1}^{r} \sigma_i \ (u_i \circ v_i \circ w_i). \tag{2}$$

- The *TUCKER3 decomposition* is given by (1) without any more constraint: the factor matrices may have different number of columns (r_1, r_2, r_3 respectively) and the core tensor $\mathcal{S} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$.
- The *Higher Order Singular Value Decomposition (HOSVD)* is given by (1) where the factor matrices U,V, and W are required to be (column) orthogonal ($U^\top U = I_{n_1}$).

2.1 The CP Decomposition (Canonical Polyadic)

The CP decomposition, or CANDECOMP (canonical decomposition) / PARAFAC (parallel factors) aims to write a tensor as a sum of rank-one tensors (i.e. a sum outer products of vectors).

Definition:[CP Decomposition] Let $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then a CP decomposition represents $\mathcal{T}$ as a sum of rank-one tensors:

$$\mathcal{X} = \sum_{j=1}^{r} \mathbf{a}_j \circ \mathbf{b}_j \circ \mathbf{c}_j$$

where $\mathbf{a}_j \in \mathbb{R}^{n_1}$, $\mathbf{b}_j \in \mathbb{R}^{n_2}$ and $\mathbf{c}_j \in \mathbb{R}^{n_3}$ for $j = 1, \ldots, r$.

Notation: We will use the following shorthand notation for the CP decomposition:

$$\mathcal{X} = \sum_{j=1}^{r} \mathbf{a}_j \circ \mathbf{b}_j \circ \mathbf{c}_j = [[A, B, C]]$$

where the three *factor matrices* A, B and C are given by:

$$A = \begin{pmatrix} | & \cdots & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_r \\ | & \cdots & | \end{pmatrix} \in \mathbb{R}^{n_1 \times r},\ B = \begin{pmatrix} | & \cdots & | \\ \mathbf{b}_1 & \cdots & \mathbf{b}_r \\ | & \cdots & | \end{pmatrix} \in \mathbb{R}^{n_2 \times r} \text{ and } C = \begin{pmatrix} | & \cdots & | \\ \mathbf{c}_1 & \cdots & \mathbf{c}_r \\ | & \cdots & | \end{pmatrix} \in \mathbb{R}^{n_3 \times r}$$

Proposition: Let $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. If $\mathcal{X} = [[A, B, C]]$ then one has

$$\begin{aligned} \mathcal{X}_{(1)} &= A(C \odot B)^T \\ \mathcal{X}_{(2)} &= B(C \odot A)^T \\ \mathcal{X}_{(3)} &= C(B \odot A)^T. \end{aligned} \tag{3}$$

For the general case we have:

Proposition: Let $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$. If

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \ldots \mathbf{a}_r^{(n)} = [[A^{(1)}, A^{(2)}, \ldots, A^{(N)}]]$$

where $A^{(n)} \in \mathbb{R}^{I_n \times R}$, then

$$\mathcal{X}_{(n)} = A^{(n)}(A^{(N)} \odot \ldots \odot A^{(n+1)} \odot A^{(n-1)} \odot \ldots \odot A^{(1)})^T \tag{4}$$

Definition:[Tensor rank] The **tensor rank** of $\mathcal{X}$ is defined as the minimal integer r such that $\mathcal{X}$ can be written as in (2), i.e. the smallest integer r such that $\mathcal{X}$ can be written as a sum of r rank-1 tensors.

Definition:[$n-$rank] The n**-rank** of $\mathcal{X}$ is defined as the rank of $\mathcal{X}_{(n)}$ the mode-n matricization of $\mathcal{X}$. It is denoted by $\text{rank}_n(\mathcal{X})$.

$$\text{rank}_n(\mathcal{X}) = \text{rank}_n(\mathcal{X}_{(n)}).$$

The **multilinear rank** of $\mathcal{X}$ is the n-tuple $(\text{rank}_1(\mathcal{X}), \ldots, \text{rank}_N(\mathcal{X}))$.

Theorem: Let $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ of tensor rank R. Denote by (R_1, R_2, R_3) the multilinear rank of $\mathcal{X}$. Then the following inequalities hold:

$$\max(R_1, R_2, R_3) \leq R \leq \min(R_1 R_2, R_1 R_3, R_2 R_3).$$

Proof: Without loss of generality, it suffices to show that $R \leq R_2 R_3$.
Let

$$\mathcal{T} = \sum_{j=1}^{R} \mathbf{a}_j \circ \mathbf{b}_j \circ \mathbf{c}_j = [[A, B, C]]$$

a rank decomposition of $\mathcal{T}$.
Recall that by definition $R_n = \mathrm{rank}(\mathcal{T}_{(n)})$ for $n = 1, 2, 3$.

We first note that vectors $c_1 \otimes b_1, \ldots, c_R \otimes b_R$ are linearly independent. Otherwise, after reordering if necessary, there would exist scalars $\gamma_1, \ldots, \gamma_{R-1}$ such that

$$c_R \otimes b_R = \sum_{k=1}^{R-1} \gamma_k (c_k \otimes b_k) \,.$$

Using definition and linearity of Kronecker product, this would lead to

$$\begin{aligned} \mathrm{vec}(\mathcal{T}) &= \sum_{k=1}^{R-1} c_k \otimes b_k \otimes a_k + c_R \otimes b_R \otimes a_R \\ &= \sum_{k=1}^{R-1} c_k \otimes b_k \otimes (\gamma_k a_R + a_k) \,, \end{aligned}$$

which implies

$$\mathcal{T} = \sum_{k=1}^{R-1} (\gamma_k a_R + a_k) \circ b_k \circ c_k,$$

i.e. is $\mathcal{T}$ can be represented by a sum of $R-1$ rank one tensors, which is in contradiction with $\mathcal{T}$ having rank R.

Thus, the matrix

$$C \odot B = \begin{bmatrix} c_1 \otimes b_1 & \cdots & c_R \otimes b_R \end{bmatrix},$$

where $\odot$ denotes Khatri-Rao product, has rank R.
Now, we observe that the mode-1 matricization of $\mathcal{T}$ can be written as

$$\mathcal{T}_{(1)} = A (C \odot B)^\top \,.$$

Since $C \odot B$ is of full column rank, it follows that

$$R_1 = \mathrm{rank}(\mathcal{T}_{(1)}) = \mathrm{rank}(A) \leq R \,.$$

We can repeat all of the observations above for mode-2 and mode-3 matricization of $\mathcal{T}$ to conclude

$$R = \mathrm{rank}(C \odot B) = \mathrm{rank}(C \odot A) = \mathrm{rank}(B \odot A)$$

and

$$R_1 = \mathrm{rank}(A), \quad R_2 = \mathrm{rank}(B), \quad R_3 = \mathrm{rank}(C)\,.$$

Now, we have

$$R = \mathrm{rank}(C \odot B) \leq \mathrm{rank}(C \otimes B) = \mathrm{rank}(B)\,\mathrm{rank}(C) = R_2 R_3\,,$$

where inequality holds because each column of $C \odot B$ appears as a column of $C \otimes B$. Repeating this for matrices $C \odot A$ and $B \odot A$ proves the results.