CSE 392: Matrix and Tensor Algorithms for Data	Spring 2024
Lecture 15 — March 6	
Instructor: Shashanka Ubaru	*Scribe: Paulina Hoyos*

1 Introduction to Tensors

1.1 Motivation

We live in a multi-dimensional world. Much of real-world data is inherently multidimensional and is easily stored in and accessed through high-dimensional arrays. Examples of multi-dimensional data include medical imaging, videos and color images, and computer vision. Moreover, many operators can be inherently multi-dimensional. For example, when we model 3D objects (like planes) the operators used for this could be of high-order and thus require the construction of high-order grids. Even if we have 2D grids, taking snapshots of the model and putting them in a tensor might result in interesting observations and reveal interesting structures. One may also simulate multiple versions of the same model and store such data as a tensor.

The traditional approach for manipulating multidimensional data consists of flattening high-dimensional arrays, which are then stored as matrices or vectors and hence standard techniques can be applied. However, using tensors helps in memory allocation and fast access of data, so high-dimensional arrays become useful for data processing. Tensor decompositions give us a way to analyze, compress, and otherwise manipulate operators and data, which is far more useful and natural than flattening this multidimensional structure into a matrix and using matrix tools.

1.1.1 Applications

Some applications of tensors include:

- **Machine vision:** understanding the world in 3D, enable understanding phenomena such as perspective, occlusions, illumination.
- **Latent semantic tensor indexing:** common terms vs. entries vs. parts, co-occurrence of terms.
- **Medical imaging:** naturally involves 3D (spatio) and 4D (spatio-temporal) correlations.
- **Video surveillance and motion signature:** 2D images + 3rd dimension of time, 3D/4D motion trajectory.

1.1.2 The Power of Representation

Traditional matrix-based methods that use data vectorization are generally agnostic to possible high dimensional correlations. But representation matters! Some correlations can only be realized in an appropriate representation.

1.1.3 Data Organization Reveals Latent Structure

Suppose $y \in \mathbb{R}^{mn}$. Reshape y as an $m \times n$ matrix $Y = uv^T = u \circ v$. Then $y = v \otimes u = [v_1 u, \ldots, v_n u]^T$. This implies that storage is reduced from mn to $m+n$ numbers, and reveals only 1 important direction. We see how retaining the higher dimensional format reveals latent structure.

1.2 Basic Definitions and Notation

Definition. A dth order tensor $\mathcal{A}^{n_1 \times n_2 \times \cdots \times n_d}$ is an element of $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$.

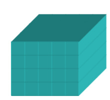
In other words, a tensor is a multidimensional array. The order of a tensor is its number of dimensions, also known as ways or modes.

Examples of tensors include:

1. 0th order tensor: scalar $a \in \mathbb{R}$

2. 1st order tensor: vector $\mathbf{a} \in \mathbb{R}^{n_1}$

3. 2nd order tensor: matrix $\mathbf{A} \in \mathbb{R}^{n_1 \times n_2}$
4. 3rd order tensor: "cube" $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$



Definition. The inner product of two order d tensors $\mathcal{A}$ and $\mathcal{B}$ is defined as

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_d}^{n_d} a_{i_1,\ldots,i_d} b_{i_1,\ldots,i_d}. \tag{1}$$

This is the sum of entry-wise elements.

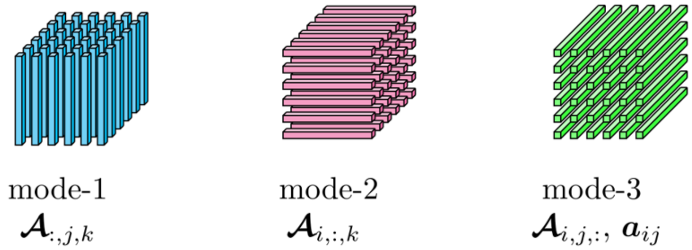
Definition. The Frobenious norm of an order d tensor is defined as

$$\|\mathcal{A}\|_F = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle} = \sqrt{\sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_d}^{n_d} |a_{i_1,\ldots,i_d}|^2}. \tag{2}$$

Unless otherwise specified, the norm of a tensor refers to the Frobenius norm.

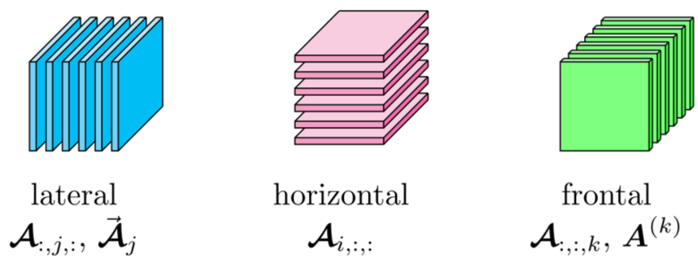
Subarrays of a tensor are formed when a subset of the indices is fixed. For matrices, these are rows and columns. We use Matlab notation, where a colon is used to indicate all elements of a mode. Different subarrays of a tensor fibers and slices.

Definition. A fiber is defined by fixing all but one index of a tensor.



Fibers are the higher-order analogue of matrix rows and columns.

Definition. A slice is a two-dimensional section of a tensor, defined by fixing all but two of its indices.



Example 1. Let $\mathcal{A}$ be the $4 \times 3 \times 2$ tensor with

$$\mathcal{A}_{:,:,1} = \begin{bmatrix} 5 & -1 & 0 \\ 0 & 2 & -1 \\ -3 & 8 & -2 \\ 1 & 4 & 0.25 \end{bmatrix}, \text{ and } \mathcal{A}_{:,:,2} = \begin{bmatrix} 0.5 & -1 & 0 \\ 6 & -5 & -1 \\ 7 & 0.5 & 3 \\ 1 & 0 & 1 \end{bmatrix}. \tag{3}$$

Then

1. $\mathcal{A}_{4,:,2} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \in \mathbb{R}^3$
2. $\mathcal{A}_{2,:,:} = \begin{bmatrix} 0 & 6 \\ 2 & -5 \\ -1 & -1 \end{bmatrix} \in \mathbb{R}^{3\times 2}$
3. $\mathcal{A}_{2,3,:} = \begin{bmatrix} -1 \\ -1 \end{bmatrix} \in \mathbb{R}^2$

Exercise 1. Compute $\|\mathcal{A}\|_F^2$.

2 Tensor-Matrix Products

Definition. The Kronecker product of an $m \times m$ matrix $\mathbf{G}$ with an $n \times n$ matrix $\mathbf{B}$ is the $mn \times mn$ matrix given by

$$\mathbf{A} = \mathbf{G} \otimes \mathbf{B} = \begin{bmatrix} g_{11}\mathbf{B} & g_{12}\mathbf{B} & \cdots & g_{1m}\mathbf{B} \\ g_{21}\mathbf{B} & g_{22}\mathbf{B} & \cdots & g_{2m}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ g_{m1}\mathbf{B} & g_{m2}\mathbf{B} & \cdots & g_{mm}\mathbf{B} \end{bmatrix}.$$

Kronecker Products are synonymous with the notion of separability, computational and storage efficiency. These come up a lot in tensor decompositions.

A tensor "matricization" refers to (specific) mappings of the tensor to a matrix.

Definition. Given an a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, its mth mode unfolding maps $\mathcal{A}$ to the matrix $\mathbf{A}_{(m)} \in \mathbb{R}^{n_m \times n_1 \cdots n_{m-1} n_{m+1} \cdots n_d}$ via $(i_1, \ldots, i_d) \mapsto (i_m, j)$ where

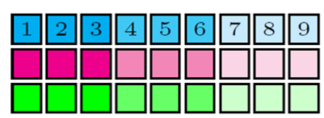
$$j = 1 + \sum_{k=1, k \neq m}^{d} (i_k - 1) \left(\prod_{l=1, l \neq m}^{k-1} n_l \right).$$

A graphical illustration is more illuminating:

(a) Original $\boldsymbol{\mathcal{A}}$.

(b) Mode-1 unfolding $\boldsymbol{\mathcal{A}}_{(1)}$.

(c) Mode-2 unfolding $\boldsymbol{\mathcal{A}}_{(2)}$.



(d) Mode-3 unfolding $\boldsymbol{\mathcal{A}}_{(3)}$.

Example 2. Let $\mathcal{A} \in \mathbb{R}^{4 \times 3 \times 2}$ be given by

$$\mathcal{A}_{:,:,1} = \begin{bmatrix} 5 & -1 & 0 \\ 0 & 2 & -1 \\ -3 & 8 & -2 \\ 1 & 4 & 0.25 \end{bmatrix} \text{ and } \mathcal{A}_{:,:,2} = \begin{bmatrix} 0.5 & -2 & 0 \\ 6 & -5 & -1 \\ 7 & 0.5 & 3 \\ 1 & 0 & 1 \end{bmatrix}.$$

The dimensions of its different unfoldings are

- Mode 1: $4 \times (3 \cdot 2)$

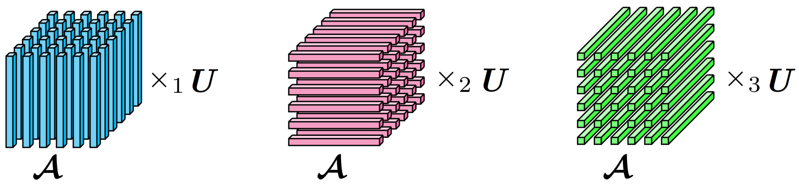
- Mode 2: $3 \times (2 \cdot 4)$
- Mode 3: $2 \times (4 \cdot 3)$

Definition. The k-mode multiplication of a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ with a matrix $\mathbf{U} \in \mathbb{R}^{j \times n_k}$, denoted by $\mathcal{A} \times_k \mathbf{U}$, is the tensor of size $n_1 \times \cdots \times n_{k-1} \times j \times n_{k+1} \times \cdots \times n_d$ whose entries are given by

$$(\mathcal{A} \times_k \mathbf{U})_{i_1 \cdots i_{k-1} j i_{k+1} \cdots i_d} = \sum_{i_k=1}^{n_k} a_{i_1 i_2 \cdots i_d} u_{j i_k}.$$

If we let $\mathbf{A}_{(k)}$ be the mth mode unfolding of the tensor $\mathcal{A}$, then

$$(\mathcal{A} \times_k \mathbf{U})_{(k)} = \mathbf{U}\mathbf{A}_{(k)}.$$



Example 3. Let $\mathcal{A} \in \mathbb{R}^{1 \times 1 \times n}$, that is, just a tube fiber, and let $\mathcal{M}$ be an $n \times n$ matrix. Then the equivalent matrix-arithmetic operation to $\mathcal{A} \times_3 \mathbf{M}$ is the matrix vector product $\mathbf{M}\mathcal{A}_{(3)}$.

Example 4. Let $\mathcal{A} \in \mathbb{R}^{2 \times 2 \times 2}$ be given by

$$\mathcal{A}_{:,:,1} = \begin{bmatrix} 1 & 2 \\ 3 & -2 \end{bmatrix} \text{ and } \mathcal{A}_{:,:,2} = \begin{bmatrix} -1 & 4 \\ -3 & 0 \end{bmatrix}$$

and consider the 2×2 matrix

$$\mathbf{X} = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}.$$

To compute the product $\mathcal{B} = \mathcal{A} \times_1 \mathbf{X}$, we first compute the matrix-matrix product

$$\mathcal{B}_{(1)} = \mathbf{X}\mathbf{A}_{(1)} = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & -11 & 4 \\ 3 & -2 & -3 & 0 \end{bmatrix} = \begin{bmatrix} -2 & 4 & 2 & 4 \\ 6 & -4 & -6 & 0 \end{bmatrix}.$$

Therefore, $\mathcal{B} \in \mathbb{R}^{2 \times 2 \times 2}$ is the tensor given by

$$\mathcal{B}_{:,:,1} = \begin{bmatrix} -2 & 4 \\ 6 & -4 \end{bmatrix}, \text{ and } \mathcal{B}_{:,:,2} = \begin{bmatrix} 2 & 4 \\ -6 & 0 \end{bmatrix}.$$

Example 5. Let $\mathcal{A} \in \mathbb{R}^{2 \times 2 \times 2}$ be given by

$$\mathcal{A}_{:,:,1} = \begin{bmatrix} 1 & 2 \\ 3 & -2 \end{bmatrix} \text{ and } \mathcal{A}_{:,:,2} = \begin{bmatrix} -1 & 4 \\ -3 & 0 \end{bmatrix}$$

and consider the 3×2 matrix

$$\mathbf{Y} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

To find $\mathcal{C} = \mathcal{A} \times_2 \mathbf{Y}$, we first compute the matrix-matrix product

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

So $\mathcal{C} \in \mathbb{R}^{2\times 3\times 2}$ is the tensor given by

$$\mathcal{C}_{:,:,1} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & -2 & 1 \end{bmatrix} \quad \text{and} \quad \mathcal{C}_{:,:,2} = \begin{bmatrix} -1 & 4 & 3 \\ -3 & 0 & -3 \end{bmatrix}.$$

Note that the second dimension (mode 2) expands from 2 to 3.

Example 6. Some tensor-matrix products result in contraction of a dimension. If $\mathcal{A}$ has size $5 \times 6 \times 7$ and $\mathbf{X}$ has size 2×7, then $\mathcal{A} \times_3 \mathbf{X}$ has size $5 \times 6 \times 2$.

There is a connection between mode-k products and the Kronecker product. Suppose $\mathcal{C}$ is a $n_1 \times n_2 \times \cdots \times n_d$ tensor, and for $j = 1, \ldots, d$, let $\mathbf{X}_j$ be a matrix with n_j columns. Let us define $\mathcal{A} = \mathcal{C} \times_{j=1}^{d} \mathbf{X}_j$. Then

$$\mathcal{A}_{(j)} = \mathbf{X}_j \mathcal{C}_{(j)} \left(\mathbf{X}_d^T \otimes \mathbf{X}_{d-1}^T \otimes \cdots \otimes \mathbf{X}_{j+1}^T \otimes \mathbf{X}_{j-1}^T \otimes \cdots \otimes \mathbf{X}_1^T \right).$$

This makes clear the separability of the mode-wise matrix products. Note that one or more of the X_j could be identity matrices.

Multiplication on each mode is independent, which gives commutativity:

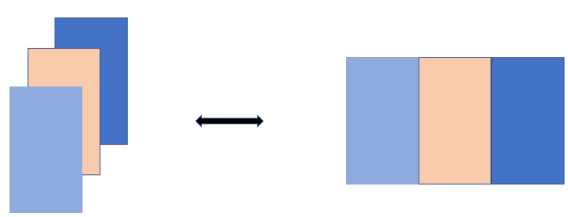
$$\mathcal{A} \times_m \mathbf{X} \times_n \mathbf{Y} = \mathcal{A} \times_n \mathbf{Y} \times_m \mathbf{X}.$$

Exercise 2. Prove the above identity.

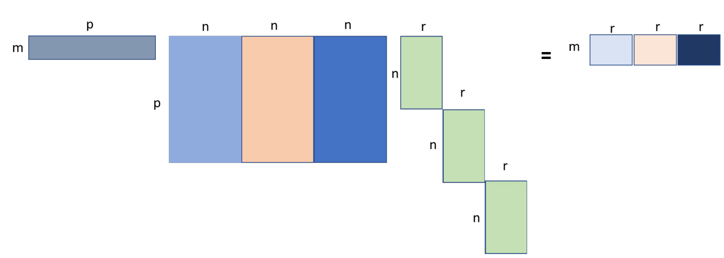
Example 7. Let $\mathcal{A}$ be a third order tensor. Then $\mathcal{B} = \mathcal{A} \times_1 \mathbf{X} \times_2 \mathbf{Y} = \mathcal{A} \times_1 \mathbf{X} \times_2 \mathbf{Y} \times_3 \mathbf{I}$ can be computed by $\mathcal{B}_{(1)} = \mathbf{X}\mathcal{A}_{(1)}(\mathbf{I} \otimes \mathbf{Y}^T)$ followed by folding these matrix along its first mode.

Separability implies that we have to do just one unfolding, as we can see in the following illustrations.

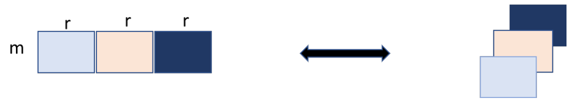
1. Unfold the tensor $\mathcal{A}$ to obtain $\mathcal{A}_{(1)}$



2. Compute the matrix products $\mathbf{X}\mathcal{A}_{(1)}(\mathbf{I} \otimes \mathbf{Y}^T)$



3. Fold the resulting matrix into a tensor:



Let $\mathcal{A} \in \mathbb{R}^{m\times p\times n}$, and consider orthogonal matrices $\mathbf{U} \in \mathbb{R}^{m\times m}$, $\mathbf{V} \in \mathbb{R}^{p\times p}$, $\mathbf{W} \in \mathbb{R}^{n\times n}$. Let $\widehat{\mathcal{A}} = \mathcal{A} \times_1 \mathbf{U}^T \times_2 \mathbf{V}^T \times_3 \mathbf{W}^T$. Then $\|\widehat{\mathcal{A}}\|_F = \|\mathcal{A}\|_F$. That is, the Frobenius norm is invariant under multiplication by orthogonal matrices. This is because the Kronecker product of orthogonal matrices is orthogonal.

Definition. A matrix $\mathrm{P} \in \mathbb{R}^{\mathrm{n}\times\mathrm{n}}$ is called an orthogonal production matrix if $\mathbf{P}^T = \mathbf{P}$ and $\mathbf{P}^2 = \mathbf{P}$.

Orthogonal projection matrices are not necessarily orthogonal - indeed, they are often not full rank: e.g. if $\mathbf{v}$ has unit length then $\mathbf{P} = \mathbf{v}\mathbf{v}^T$ is an orthogonal projector onto span$\{\mathbf{v}\}$.

In the tensor setting, projectors have the same behavior as in the matrix case with respect to the mode-k product: let $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$ and let $\mathbf{P}_k$ be a $n_k \times n_k$ orthogonal projection matrix, then

$$\|\mathcal{A} - \mathcal{A} \times_k \mathbf{P}_k\|_F = \|\mathcal{A} - \mathcal{A} \times_k (\mathbf{I} - \mathbf{P}_k)\|_F,$$

for $k = 1, 2, 3$.

Exercise 3. Show that $\mathbf{P}_k \otimes \mathbf{I}$ and $\mathbf{I} \otimes \mathbf{P}_k$ or orthogonal projection matrices.

Exercise 4. Now show that

$$\|\mathcal{A}-\mathcal{A}\times_1\mathbf{P}_1\times_2\mathbf{P}_2\times_3\mathbf{P}_3\|^2 = \|\mathcal{A}\times_1(\mathbf{I}-\mathbf{P}_1)\|^2+\|\mathcal{A}\times_1\mathbf{P}_1\times_2(\mathbf{I}-\mathbf{P}_2)\|^2+\|\mathcal{A}\times_1\mathbf{P}_1\times_2\mathbf{P}_2\times_3(\mathbf{I}-\mathbf{P}_3)\|^2.10$$