Introduction to Tensors

Suraj Kumar

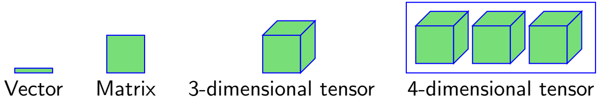
Inria & ENS Lyon

Email:*suraj.kumar@inria.fr*

CR12: September 2023

https://surakuma.github.io/courses/daamtc.html

Tensors (n-dimensional arrays)



- **Neuroscience**: measure of calcium fluorescence in a particular pixel during a time step of a single trial (Pixel $\times$ Time $\times$ Trial)
- **Combustion simulation**: value of a variable in a spatial grid during a time step (Grid length $\times$ Grid width $\times$ Grid height $\times$ Variable $\times$ Time)
- **Media**: rating of a movie by a user during a time slice (User $\times$ Movie $\times$ Time)
- **Molecular/quantum simulations**: interaction of electrons in d orbitals with a 4^d tensor

Notation convention: Matrix A, tensor $\mathcal{A}$

Table of Contents

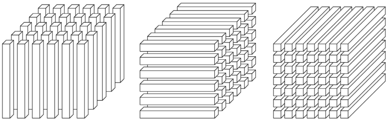
Tensor notations (following [Kolda and Bader, 2009])

Let $\mathcal{A}$ be a d-dimensional tensor of size $n_1 \times n_2 \times \cdots \times n_d$, $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_d}$.

- $d = 1$, first order tensors: vectors
- $d = 2$, second order tensors: matrices

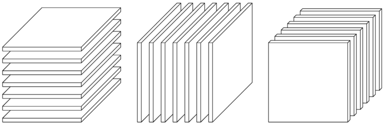
The element of $\mathcal{A}$ is denoted as $\mathcal{A}(i_1, i_2, \ldots, i_d)$.

- Fibers: defined by fixing all indices except one



Mode-1 (column) fibers: $\mathcal{A}(:, j, k)$, Mode-2 (row) fibers: $\mathcal{A}(i, :, k)$ and Mode-3 (tube) fibers: $\mathcal{A}(i, j, :)$ of a 3-dimensional tensor $\mathcal{A}$.

- Slices: defined by fixing all indices except two



Horizontal slices: $\mathcal{A}(i, :, :)$, Lateral slices: $\mathcal{A}(:, j, :)$ and Frontal slices: $\mathcal{A}(:, :, k)$ of a 3-dimensional tensor $\mathcal{A}$.

Figures from [Kolda and Bader, 2009].

Tensor preliminaries

- The norm of a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is analogous to the matrix Frobenius norm, and defined as

$$||\mathcal{A}|| = \sqrt{\sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_d=1}^{n_d} \mathcal{A}^2(i_1, i_2, \cdots, i_d)}$$

- The inner product of $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_d=1}^{n_d} \mathcal{A}(i_1, i_2, \cdots, i_d) \mathcal{B}(i_1, i_2, \cdots, i_d)$$

We can note that $\langle \mathcal{A}, \mathcal{A} \rangle = ||\mathcal{A}||^2$.

Specific tensors

- A rank one tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ can be written as the outer product of d vectors,

$$\mathcal{A} = u_1 \circ u_2 \circ \cdots \circ u_d$$

$$\mathcal{A}(i_1, i_2, \cdots, i_d) = u_1(i_1) u_2(i_2) \cdots u_d(i_d) \text{ for all } 1 \leq i_k \leq n_k$$

- A cubical tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ has same size in every mode,

$$n_1 = n_2 = \cdots = n_d$$

- A supersymmetric (or symmetric) tensor has the same element for any permutation of the indices
- A diagonal tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ has $\mathcal{A}(i_1, i_2, \cdots, i_d) \neq 0$ only if $i_1 = i_2 = \cdots = i_d$

Matricization or Unfolding of a tensor into a matrix

- The mode-j unfolding of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is represented by a matrix $A_{(j)} \in \mathbb{R}^{n_j \times n}$, where $n = n_1 n_2 \cdots n_{j-1} n_{j+1} \cdots n_d$
- Tensor element $\mathcal{A}(i_1, i_2, \cdots, i_d)$ maps to matrix element $A_{(j)}(i_j, k)$, where $k = 1 + \sum_{\ell=1, \ell \neq j}^{d} (i_\ell - 1) N_\ell$ with $N_\ell = \prod_{m=1, m \neq j}^{\ell-1} n_m$

Example with the frontal slices of $\mathcal{A} \in \mathbb{R}^{4 \times 2 \times 3}$:

$$\mathcal{A}(:,:,1) = \begin{pmatrix} 1 & 5 \\ 2 & 6 \\ 3 & 7 \\ 4 & 8 \end{pmatrix}, \quad \mathcal{A}(:,:,2) = \begin{pmatrix} 9 & 13 \\ 10 & 14 \\ 11 & 15 \\ 12 & 16 \end{pmatrix}, \quad \mathcal{A}(:,:,3) = \begin{pmatrix} 17 & 21 \\ 18 & 22 \\ 19 & 23 \\ 20 & 24 \end{pmatrix}$$

The three mode-j unfoldings are:

$$A_{(1)} = \begin{pmatrix} 1 & 5 & 9 & 13 & 17 & 21 \\ 2 & 6 & 10 & 14 & 18 & 22 \\ 3 & 7 & 11 & 15 & 19 & 23 \\ 4 & 8 & 12 & 16 & 20 & 24 \end{pmatrix}, \quad A_{(3)} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \\ 17 & 18 & 19 & 20 & 21 & 22 & 23 & 24 \end{pmatrix},$$

$$A_{(2)} = \begin{pmatrix} 1 & 2 & 3 & 4 & 9 & 10 & 11 & 12 & 17 & 18 & 19 & 20 \\ 5 & 6 & 7 & 8 & 13 & 14 & 15 & 16 & 21 & 22 & 23 & 24 \end{pmatrix}$$

Assignment 4 – deadline Oct 10

Question: Write a program in your preferred programming language to obtain mode-3 unfolding of $\mathcal{A} \in \mathbb{R}^{3\times 3\times 3}$. Elements of $\mathcal{A}$ are defined in the following way:

$$\mathcal{A}(i,j,k) = i + j^2 + k^3 \text{ for } 1 \leq i,j,k \leq 3.$$

If your preferred language support 0-based indexing then you can consider $0 \leq i,j,k \leq 2$.

Submission procedure : send your code to my ENS email address (suraj.kumar@ens-lyon.fr) by Oct 10.

Tensor multiplication along j-mode with a matrix

The j-mode product of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ with $U \in \mathbb{R}^{K \times n_j}$ is denoted by $\mathcal{A} \times_j U$ and is of size $n_1 \times \cdots n_{j-1} \times K \times n_{j+1} \times \cdots \times n_d$.

$$(\mathcal{A} \times_j U)(i_1, \cdots, i_{j-1}, k, i_{j+1}, \cdots i_d) = \sum_{i_j=1}^{n_j} \mathcal{A}(i_1, \cdots, i_d) U(k, i_j)$$

In terms of unfolded tensors:

$$\mathcal{B} = \mathcal{A} \times_j U \Leftrightarrow B_{(j)} = UA_{(j)}$$

Some properties of j-mode products:

- $\mathcal{A} \times_j U \times_k V = \mathcal{A} \times_k V \times_j U \quad (j \neq k)$
- $\mathcal{A} \times_j U \times_j V = \mathcal{A} \times_j VU$

Matrix products

- The Kronecker product of $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$ is $C \in \mathbb{R}^{mp\times nq}$,

$$C = A \otimes B = \begin{pmatrix} A(1,1)B & \cdots & A(1,n)B \\ \vdots & \ddots & \vdots \\ A(m,1)B & \cdots & A(m,n)B \end{pmatrix}$$

- The Khatri-Rao product of $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times n}$ is $C \in \mathbb{R}^{mp\times n}$,

$$C = A \odot B = \begin{pmatrix} A(:,1) \otimes B(:,1) & A(:,2) \otimes B(:,2) & \cdots & A(:,n) \otimes B(:,n) \end{pmatrix}$$

- The Hadamard product of $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times n}$ is $C \in \mathbb{R}^{m\times n}$,

$$C = A * B = \begin{pmatrix} A(1,1)B(1,1) & \cdots & A(1,n)B(1,n) \\ \vdots & \ddots & \vdots \\ A(m,1)B(m,1) & \cdots & A(m,n)B(m,n) \end{pmatrix}$$

Useful properties of matrix products

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

$$A \odot B \odot C = (A \odot B) \odot C = A \odot (B \odot C)$$

$$(A \odot B)^T (A \odot B) = A^T A * B^T B,$$

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

Here $A^{\dagger}$ denotes the Moore–Penrose pseudoinverse of A.

Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ and $U_j \in \mathbb{R}^{m_j \times n_j}$ for $1 \leq j \leq d$. Then,

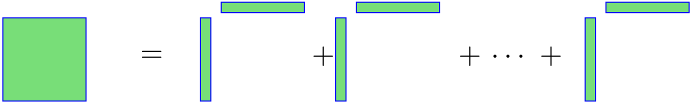
$$\mathcal{B} = \mathcal{A} \times_1 U_1 \times_2 U_2 \cdots \times_d U_d$$

$$\Leftrightarrow B_{(j)} = U_j A_{(j)} (U_d \otimes \cdots U_{j+1} \otimes U_{j-1} \otimes \cdots \otimes U_1)^T.$$

Table of Contents

Recap on Singular Value Decomposition (SVD)

- It decomposes a matrix $A \in \mathbb{R}^{m \times n}$ to the form $U \Sigma V^T$
  - U is an $m \times m$ orthogonal matrix
  - V is an $n \times n$ orthogonal matrix
  - Σ is an $m \times n$ rectangular diagonal matrix
- The diagonal entries $\sigma_i = \Sigma_{ii}$ of Σ are called singular values
  - $\sigma_i \geq 0$ and $\sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_{\min(m,n)}$
- The largest r such that $\sigma_r \neq 0$ is called the rank of the matrix
- SVD represents a matrix as the sum of r rank one matrices



Tensor decompositions

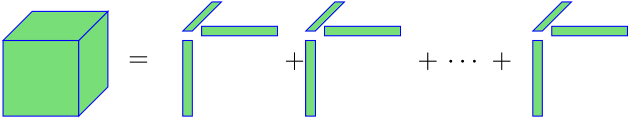
Popular higher-order extension of the matrix SVD:

- CANDECOMP/PARAFAC (CP) : proposed by Hitchcock in 1927
- Tucker decomposition: proposed by Tucker in 1963
- Tensor train decomposition: proposed by Oseledets in 2011, known in quantum chemistry community from a long time with the name of matrix product states

CP and Tucker decompositions are well suited to work with small and moderate dimension tensors ($d \leq 10$). Tensor train is preferred for high dimension tensors.

CP decomposition of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$

It factorizes a tensor into a sum of rank one tensors.



CP decomposition of a 3-dimensional tensor.

$$\mathcal{A} = \sum_{\alpha=1}^{r} U_1(:,\alpha) \circ U_2(:,\alpha) \circ \cdots \circ U_d(:,\alpha)$$

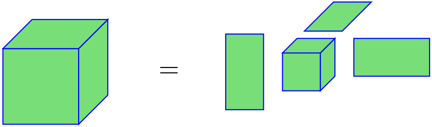
$$\mathcal{A}(i_1, \cdots, i_d) = \sum_{\alpha=1}^{r} U_1(i_1,\alpha) U_2(i_2,\alpha) \cdots U_d(i_d,\alpha)$$

The minimum r required to express $\mathcal{A}$ is called the rank of $\mathcal{A}$. The matrices $U_j \in \mathbb{R}^{n_j \times r}$ for $1 \leq j \leq d$ are called factor matrices.

- (+) The number of entries in a CP decomposition of $\mathcal{A} = \mathcal{O}((n_1 + \cdots + n_d)r)$
- (-) Determining the minimum value of r is an NP-complete problem
- (-) No robust algorithms to compute this representation

Tucker decomposition of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$

It represents a tensor with d matrices (usually orthogonal) and a small core tensor.



Tucker decomposition of a 3-dimensional tensor.

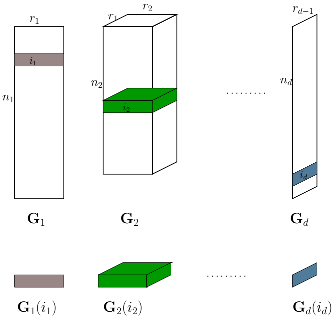
$$\mathcal{A} = \mathcal{G} \times_1 U_1 \cdots \times_d U_d$$

$$\mathcal{A}(i_1, \cdots, i_d) = \sum_{\alpha_1=1}^{r_1} \cdots \sum_{\alpha_d=1}^{r_d} \mathcal{G}(\alpha_1, \cdots, \alpha_d) U_1(i_1, \alpha_1) \cdots U_d(i_d, \alpha_d)$$

Here r_j for $1 \leq j \leq d$ denote a set of ranks. Matrices $U_j \in \mathbb{R}^{n_j \times r_j}$ for $1 \leq j \leq d$ are called factor matrices. The tensor $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times \cdots \times r_d}$ is called the core tensor.

- (+) SVD based stable algorithms to compute this decomposition
- (-) The number of entries $= \mathcal{O}(n_1 r_1 + \cdots + n_d r_d + \prod_{j=1}^{d} r_j)$

- A d-dimensional tensor is represented with 2 matrices and d-2 3-dimensional tensors.



$$\mathbf{A}(i_1, i_2, \cdots, i_d) = \mathbf{G}_1(i_1)\mathbf{G}_2(i_2)\cdots\mathbf{G}_d(i_d)$$

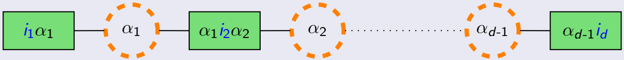
An entry of $\mathcal{A} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ is computed by multiplying corresponding matrix (or row/column) of each matrix/tensor.

Tensor Train decomposition

$\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is represented with cores $\mathcal{G}_k \in \mathbb{R}^{r_{k-1} \times n_k \times r_k}$, $k=1, 2, \cdots d$, $r_0 = r_d = 1$ and its elements satisfy the following expression:

$$\mathcal{A}(i_1, \cdots, i_d) = \sum_{\alpha_0=1}^{r_0} \cdots \sum_{\alpha_d=1}^{r_d} \mathcal{G}_1(\alpha_0, i_1, \alpha_1) \cdots \mathcal{G}_d(\alpha_{d-1}, i_d, \alpha_d)$$

$$= \sum_{\alpha_1=1}^{r_1} \cdots \sum_{\alpha_{d-1}=1}^{r_{d-1}} \mathcal{G}_1(1, i_1, \alpha_1) \cdots \mathcal{G}_d(\alpha_{d-1}, i_d, 1)$$



The ranks r_k are called TT-ranks.

- The number of entries in this decomposition = $\mathcal{O}(n_1 r_1 + n_2 r_1 r_2 + n_3 r_2 r_3 + \cdots + n_{d-1} r_{d-2} r_{d-1} + n_d r_{d-1})$