The ABC of tensor

Andersen Ang

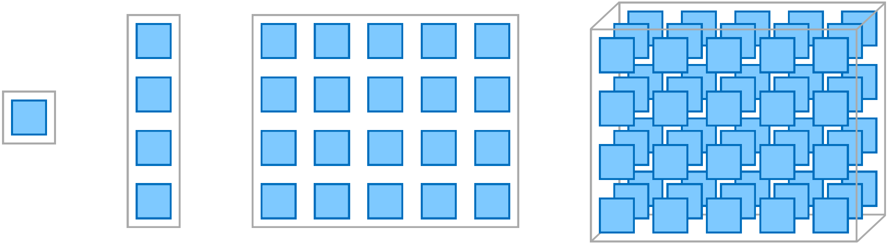
Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be Homepage: angms.science

First draft : October 21, 2020
Last update : October 21, 2020

What is tensor

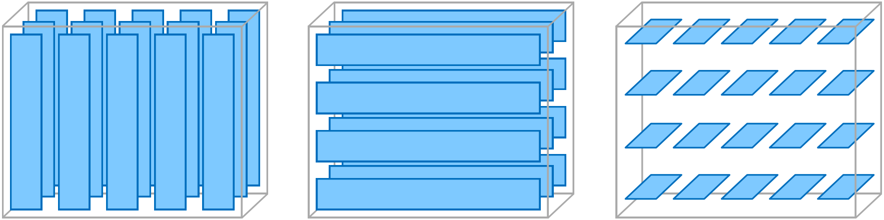
Tensor is the generalization of scalar, vector and matrix.



- The tensor here is an example of order-3 tensor.
- The column direction of the tensor is called mode-1
- The row direction of the tensor is called mode-2
- The tube direction of the tensor is called mode-3
- The (i, j, k)-th element of a tensor $\mathcal{T}$ is denoted as $\mathcal{T}_{i,j,k}$ or $\mathcal{T}(i, j, k)$.

Fiber

The vectors in a tensor are called fibers.



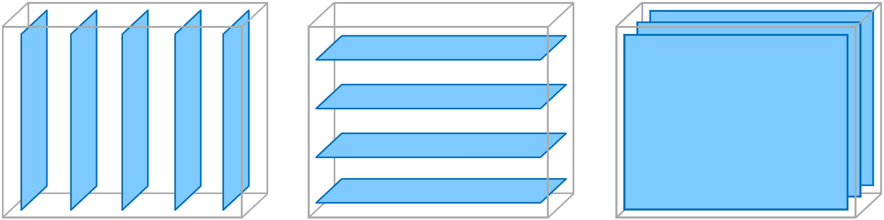
There are specific names for vectors in order-3 tensors:

- Mode-1 vector in tensor, denoted as $\mathcal{T}(:, j, k)$, is called column.
- Mode-2 vector in tensor, denoted as $\mathcal{T}(i, :, k)$, is called row.
- Mode-3 vector in tensor, denoted as $\mathcal{T}(i, j, :)$, is called tube.

There is no specific name for vectors in order-4,5,... tensors.

Slice

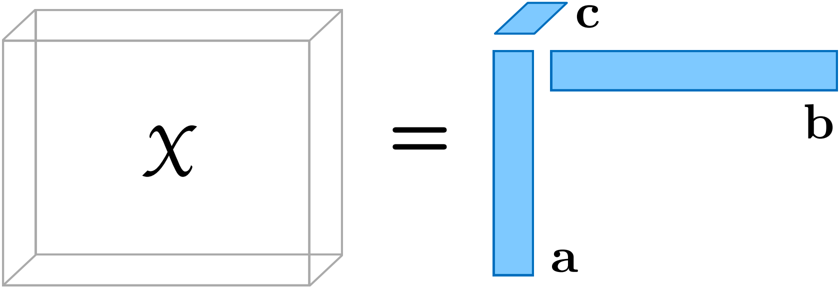
The matrices in a tensor are called slices.



There are specific names for vectors in order-3 tensors. For example, frontal slide refer to the matrices $\mathcal{T}(:,:,k)$.

Rank 1 tensor

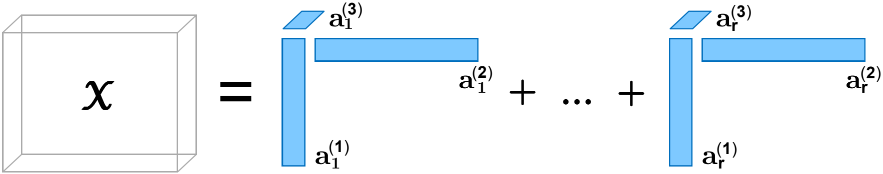
A matrix $\mathbf{X}$ is rank-1 if it can be expressed as the outer product of two vectors as $\mathbf{X} = \mathbf{u}\mathbf{v}^\top$. Similarly, an order-3 tensor is rank-1 if it can be expressed as the outer product of three vectors as $\mathbf{X} = \mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c}$.



Note: here $\otimes$ is the notation of outer product. There is no unified notation used in the literature.

CPD: CP decomposition

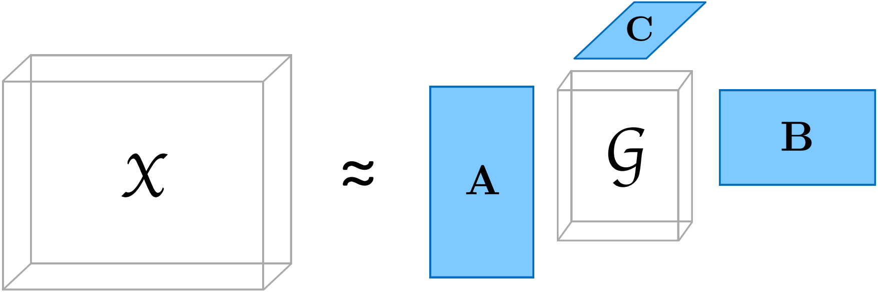
If a tensor $\mathcal{X}$ can be expressed as the sum of r rank-1 tensor, then the tensor is said to have rank r, and such decomposition is called CP decomposition.



- The ith rank-1 component and the jth rank-1 component are independent to each other in CPD.
- Historical note: CP stands for CANDECOMP/PARAFAC. And CPD stands for canonical polyadic decomposition.

Tucker decomposition

When the ith rank-1 component and the jth rank-1 component are not independent to each other, CPD becomes the Tucker decomposition.



We can also say that, when the core tensor/kernel $\mathcal{G}$ is a super-diagonal tensor, then Tucker decomposition reduces to CPD.

Solving CPD

(Approximate) CPD can be expressed as the following nonconvex optimization problem

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \|\mathcal{T} - \mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C}\|_F^2$$

Here we have a order-3 CPD problem: given a tensor $\mathcal{T}$, find the matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ such that their CPD product is as close to $\mathcal{T}$ as close as possible. Here $\mathbf{A}, \mathbf{B}, \mathbf{C}$ correspond to the mode-1, mode-2, mode-3 factor matrix of the CPD model.

BCD scheme to solve CPD

- (Approximate) CPD can be solve by using block coordinate descent: update each block at a time while fixing the other blocks.

- For CPD cost function $f(\mathbf{A},\mathbf{B},\mathbf{C}) = \|\mathcal{T} - \mathbf{A}\otimes\mathbf{B}\otimes\mathbf{C}\|_F^2$, when we update block $\mathbf{A}$, we only need the information of $\mathbf{B}$ and $\mathbf{C}$: by definition, the block $\mathbf{A}_k$ minimizes the function $f(\mathbf{A},\mathbf{B}_{k-1},\mathbf{C}_{k-1})$ as

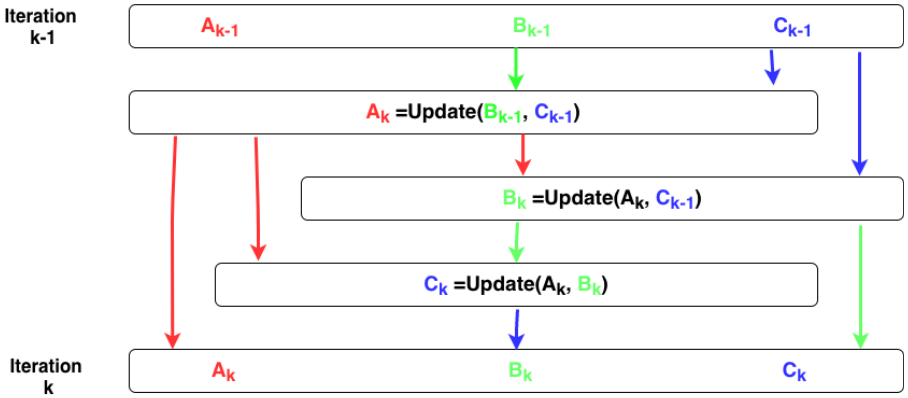
$$\mathbf{A}_k = \min_{\mathbf{A}} \|\mathcal{T} - \mathbf{A}\otimes\mathbf{B}_{k-1}\otimes\mathbf{C}_{k-1}\|_F^2$$

  Let $\mathbf{D}_{k-1} = \mathbf{B}_{k-1}\otimes\mathbf{C}_{k-1}$, then

$$\begin{aligned}\mathbf{A}_k &= \min_{\mathbf{A}} \|\mathcal{T} - \mathbf{A}\otimes\mathbf{D}_{k-1}\|_F^2 \\ &= \min_{\mathbf{A}} \|\mathcal{T}\|_F^2 - 2\langle\mathcal{T},\mathbf{A}\otimes\mathbf{D}_{k-1}\rangle + \|\mathbf{A}\otimes\mathbf{D}_{k-1}\|_F^2\end{aligned}$$

  Ignoring the constant term, $\mathbf{A}_k$ is thus the minimizer of the function $-2\langle\mathcal{T},\mathbf{A}\otimes\mathbf{D}_{k-1}\rangle + \|\mathbf{A}\otimes\mathbf{D}_{k-1}\|_F^2$. Take the derivative of this function w.r.t. to $\mathbf{A}$ and set the derivative to zero gives the update of $\mathbf{A}$ (we postpone the details of the algebra involved later).

BCD scheme to solve CPD



Last page - summary

Discussed

- What is tensor
- Fiber and slice
- Rank-1 tensor
- CPD and Tucker decomposition
- Solving CPD by BCD

Not discussed

- Existence of CPD/Tucker decomposition
- Uniqueness of the solution
- The issue of degeneracy
- Algebraic method to solve CPD

See here for details on CPD.

End of document