Canonical Polyadic Decomposition

Basic notations

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

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

First draft : May 5, 2020
Last update : May 5, 2020

Notations

- $[N]$ denotes the interval $1 \leq x \leq N$. If it represents integer interval, then $[N] = [1, \ldots, N]$.
- A N-way array or N-th order tensor $\mathcal{T}$ is a multidimensional array in the product $\mathbb{R}^{I_1} \times \cdots \times \mathbb{R}^{N}$ of the vector spaces $\mathbb{R}^{I_i}$ for $i \in [N]$.
- A vector $\mathbf{x} \in \mathbb{R}_+^{I_1}$ is a first-order tensor, and a matrix $\mathbf{M} \in \mathbb{R}_+^{I_1 \times I_2}$ is a second-order tensor.
- The tensor product $\bigotimes$ over N real vector spaces $\mathbb{R}^{I_1}$, …, $\mathbb{R}^{I_N}$ is defined as

$$\left[\bigotimes_{i=1}^{N} \mathbf{a}^{(i)}\right]_{j_1,\ldots,j_N} := \prod_{i=1}^{N} \mathbf{a}^{(i)}(j_i),$$

  where $\mathbf{a}^{(i)} \in \mathbb{R}^{I_i}$ for all $i \in [N]$.

Tensor product, example of $N = 2$

- The tensor product $\bigotimes$ over N real vector spaces $\mathbb{R}^{I_1}$, ..., $\mathbb{R}^{I_N}$ is

$$\left[\bigotimes_{i=1}^{N} \mathbf{a}^{(i)}\right]_{j_1,\ldots,j_N} := \prod_{i=1}^{N} \mathbf{a}^{(i)}(j_i),$$

  where $\mathbf{a}^{(i)} \in \mathbb{R}^{I_i}$ for all $i \in [N]$.

- If $N = 2$:

$$[\mathbf{a}^{(1)} \otimes \mathbf{a}^{(2)}]_{j_1,j_2} = \mathbf{a}^{(1)}(j_1) \times \mathbf{a}^{(2)}(j_2).$$

  As 2nd order tensors are matrix, we can see that
  - $\mathbf{a}^{(1)} \otimes \mathbf{a}^{(2)}$ is a matrix, with size I_1-by-I_2.
  - $[\mathbf{a}^{(1)} \otimes \mathbf{a}^{(2)}]_{j_1,j_2}$ is the (j_1, j_2)-th element of the matrix $\mathbf{a}^{(1)} \otimes \mathbf{a}^{(2)}$, it is equal to $\mathbf{a}^{(1)}(j_1) \times \mathbf{a}^{(2)}(j_2)$, which is the j_1-th element of the vector $\mathbf{a}^{(1)}$ multiplies the j_2-th element of the vector $\mathbf{a}^{(2)}$.

Tensor product, example of $N = 3$

- The tensor product $\bigotimes$ over N real vector spaces $\mathbb{R}^{I_1}, \ldots, \mathbb{R}^{I_N}$ is

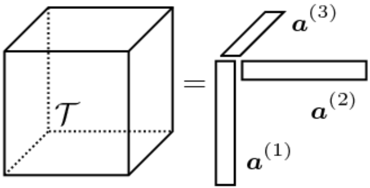
$$\left[\bigotimes_{i=1}^{N} \mathbf{a}^{(i)}\right]_{j_1,\ldots,j_N} := \prod_{i=1}^{N} \mathbf{a}^{(i)}(j_i),$$

  where $\mathbf{a}^{(i)} \in \mathbb{R}^{I_i}$ for all $i \in [N]$.

- If $N = 3$

$$[\mathbf{a}^{(1)} \otimes \mathbf{a}^{(2)} \otimes \mathbf{a}^{(3)}]_{j_1,j_2,j_3} = \mathbf{a}^{(1)}(j_1) \times \mathbf{a}^{(2)}(j_2) \times \mathbf{a}^{(3)}(j_3).$$

  Graphically it can be illustrated as

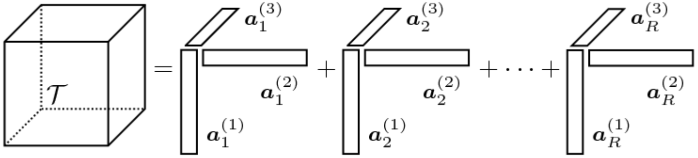


Canonical Polyadic Decomposition

- The Canonical Polyadic Decomposition (CPD) of a tensor $\mathcal{T}$ is

$$\mathcal{T} = \sum_{p=1}^{R} \bigotimes_{i=1}^{N} \mathbf{a}_p^{(i)}.$$

- If $N = 3$, graphically it can be illustrated as



- The smallest R such that the equation holds is called the (CP) rank of the tensor.

Equivalent notations to parameterize a low-rank tensor

Grouping components $\mathbf{a}_p^{(i)}$ as columns of factor matrices $\mathbf{A}^{(i)} = [\mathbf{a}_1^{(i)}, \ldots, \mathbf{a}_R^{(i)}]$, the following notations are equivalent:

$$
\begin{aligned}
\mathcal{T} &= \sum_{p=1}^{R} \bigotimes_{i=1}^{N} \mathbf{a}_p^{(i)} && \text{CPD} \\
&= [\![\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!] && \text{Kruskal notation} \\
&= \mathcal{I}_R \times_1 \mathbf{A}^{(1)} \times_2 \ldots \times_N \mathbf{A}^{(N)} && n\text{-mode product notation} \\
&:= \left(\bigotimes_{i=1}^{N}{}_a \mathbf{A}^{(i)} \right) \mathcal{I}_R,
\end{aligned}
$$

where $\mathcal{I}_R$ denotes the identity tensor, which the off-super-diagonal are 0 and the super-diagonal are 1, and $\bigotimes_a$ is a tensor product of linear maps induced by the tensor product $\otimes$ of vectors.

The last equation uses the fact that linear applications on tensor spaces of finite dimensions also form a tensor space with tensor product $(\mathbf{A} \otimes_a \mathbf{B})(\mathbf{x} \otimes \mathbf{y}) := \mathbf{A}\mathbf{x} \otimes \mathbf{B}\mathbf{y}$.

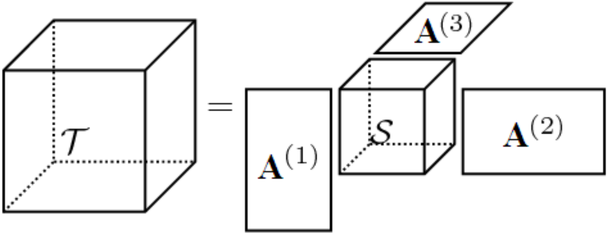
Tucker Decomposition

- In the notation

$$\mathcal{T} = \mathcal{I}_R \times_1 \mathbf{A}^{(1)} \times_2 \ldots \times_N \mathbf{A}^{(N)} := \left(\bigotimes_{i=1}^{N}{}_a \mathbf{A}^{(i)} \right) \mathcal{I}_R,$$

  if $\mathcal{I}_R$ is replaced by a general tensor $\mathcal{S}$, we have Tucker Decomposition.

- If $N = 3$, graphically it can be illustrated as



CPD factorization of a tensor

- To factorize a given tensor $\mathcal{T}$ into low rank structure, we find another tensor $\mathcal{X}$ that is low rank such that the distance between $\mathcal{T}$ and $\mathcal{X}$ is small. For example in Frobenius norm :

$$\min_{\mathcal{X}} \|\mathcal{T} - \mathcal{X}\|_F,$$

  where $\mathcal{X}$ has the form

$$\mathcal{X} = \sum_{p=1}^{R} \bigotimes_{i=1}^{N} \mathbf{a}_p^{(i)}.$$

- The factorization is called
  - exact if the distance $= 0$,
  - inexact (or approximate) if the distance > 0.
- Based on the equivalent notation in p.6, we have several formulations of this optimization problem.

NMF and NTF problems as constrained CPD

- In general, NMF and NTF are nonconvex optimization problems that falls into the following form

$$\min_{\substack{\mathbf{a}_p^{(i)} \\ i\in[N] \\ p\in[r]}} \left\{ \left\| \mathcal{T} - \sum_{p=1}^{r} \bigotimes_{i=1}^{N} \mathbf{a}_p^{(i)} \right\|_F^2 + \sum_{i,p} g_{i,p}\big(\mathbf{a}_p^{(i)}\big) \right\},$$

  where $g_{i,p}$ represent regularizers, or constraint (for example, the indicator function of the nonnegative orthant represents the nonnegativity constraint).

- This formulation stressed on the columns of the factors $\mathbf{A}^{(i)}$: as g are functions on $\mathbf{a}^{(i)}$, and the optimization variable here are the vectors $\mathbf{a}^{(i)}$.

NMF and NTF problems as constrained CPD

- In general, NMF and NTF are nonconvex optimization problems that falls into the following form

$$\min_{\substack{\mathbf{A}^{(i)} \\ i\in[N]}} \left\{ \left\| \mathcal{T} - \left(\bigotimes_{i=1}^{N}{}_{a} \mathbf{A}^{(i)} \right) \mathcal{I}_r, \right\|_F^2 + \sum_i g_i\big(\mathbf{A}^{(i)}\big) \right\},$$

  where g_i represent regularizers or constraint.

- This formulation stressed on the matrix form of the factors $\mathbf{A}^{(i)}$.

Matricization/unfolding of tensor and Kronecker product

- The unfolding/matricization of a tensor $\mathcal{X}$ is defined as

$$\mathbf{X}_{[i]} := \mathbf{a}^{(i)} \otimes \left(\underset{\substack{l \neq i \\ l=N}}{\overset{1}{\boxtimes}} \mathbf{a}^{(l)} \right) \in \mathbb{R}^{d},\ d = I_i \times \prod_{l \neq i} I_l,$$

where $\boxtimes$ is the Kronecker product of two matrices $\mathbf{A} \in \mathbb{R}^{I_1 \times J_1}$ and $\mathbf{B} \in \mathbb{R}^{I_2 \times J_2}$, it is defined as

$$\mathbf{A} \boxtimes \mathbf{B} = \left[\begin{array}{c|c|c} A_{1,1}\mathbf{B} & \dots & A(1, J_1)\mathbf{B} \\ \hline \vdots & \ddots & \vdots \\ \hline A_{I_1,1}\mathbf{B} & \dots & A(I_1, J_1)\mathbf{B} \end{array} \right] \in \mathbb{R}^{I_1 I_2 \times J_1 J_2},$$

where the Kronecker product of several matrices can be deduced by associativity.

- The mode-n unfolding can be seen as forming a matrix using the mode-n fiber of the tensor.

Khatri-Rao product

- Khatri-Rao product $\mathbf{A} \odot \mathbf{B}$ is the columns-wise Kronecker product. Setting $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_{J_1}]$ and $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_{J_1}]$,

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \boxtimes \mathbf{b}_1, \ldots, \mathbf{a}_{J_1} \boxtimes \mathbf{b}_{J_1}] \,.$$

- Eample $\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}, \mathbf{B} = \begin{bmatrix} g & h & i \\ j & k & l \end{bmatrix}$.

$$\begin{aligned} \mathbf{A} \odot \mathbf{B} &= \begin{bmatrix} \mathbf{a}_1 \boxtimes \mathbf{b}_1 & \mathbf{a}_2 \boxtimes \mathbf{b}_2 & \mathbf{a}_3 \boxtimes \mathbf{b}_3 \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} a \\ d \end{bmatrix} \boxtimes \begin{bmatrix} g \\ j \end{bmatrix} & \begin{bmatrix} b \\ e \end{bmatrix} \boxtimes \begin{bmatrix} h \\ k \end{bmatrix} & \begin{bmatrix} c \\ f \end{bmatrix} \boxtimes \begin{bmatrix} i \\ l \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} ag & bh & ci \\ aj & bk & cl \\ dg & eh & fi \\ dj & ek & fl \end{bmatrix} . \end{aligned}$$

Expressing CPD as a matrix factorization problem

- With unfolding and Khatri-Rao product, we have the following equivalence

$$\mathcal{X} = \left(\bigotimes_{i=1}^{N}{}_{a} \mathbf{A}^{(i)} \right) \mathcal{I}_r \quad \equiv \quad \forall i \in [N],\ \mathbf{X}_{[i]} = \mathbf{A}^{(i)} \left(\bigodot_{\substack{l \neq i \\ l=N}}^{1} \mathbf{A}^{(l)} \right)^{\top}.$$

- Define

$$\mathbf{B}^{(i)} = \left(\bigodot_{\substack{l \neq i \\ l=N}}^{1} \mathbf{A}^{(l)} \right) = \mathbf{A}^{(N)} \odot \cdots \odot \mathbf{A}^{(i+1)} \odot \mathbf{A}^{(i-1)} \cdots \odot \mathbf{A}^{(1)}.$$

- We have

$$\mathbf{X}_{[i]} = \mathbf{A}^{(i)} {\mathbf{B}^{(i)}}^{\top},$$

i.e. we have turned a tensor factorization problem written in tensor product form into a matrix factorization problem.

Solving the CPD using BCD

- The CPD problem

$$\min_{\substack{\mathbf{A}^{(i)} \\ i\in[N]}} \left\{ \left\| \mathcal{T} - \left(\bigotimes_{i=1}^{N}{}_{a} \mathbf{A}^{(i)} \right) \mathcal{I}_R, \right\|_F^2 + \sum_i g_i\big(\mathbf{A}^{(i)}\big) \right\},$$

  can be solved by solving the subproblem in the form

$$\min_{\mathbf{A}^{(i)}} \left\{ \left\| \mathbf{T}_{[i]} - \mathbf{A}^{(i)} \mathbf{B}^{(i)\top} \right\|_F^2 + g_i\big(\mathbf{A}^{(i)}\big) \right\}.$$

- By solving each of the subproblem, we arrive at the block coordinate descent (BCD) algorithm on solving CPD.

- Computationally speaking, the cost of the BCD includes the cost of performing the update and the cost of forming the matrix form expression. While forming $\mathbf{T}_{[i]}$ has low cost, forming $\mathbf{B}^{(i)}$ can be expensive as it contains several matrix multiplications. In fact, $\mathbf{B}^{(i)}$ should not be computed explicitly.

Last page - summary

- CPD parameterization of a low rank tensor

$$
\begin{aligned}
\mathcal{T} &= \sum_{p=1}^{R} \bigotimes_{i=1}^{N} \mathbf{a}_p^{(i)} && \text{CPD} \\
&= [\![\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}]\!] && \text{Kruskal notation} \\
&= \mathcal{I}_R \times_1 \mathbf{A}^{(1)} \times_2 \dots \times_N \mathbf{A}^{(N)} && n\text{-mode product notation} \\
&:= \left(\bigotimes_{i=1}^{N}{}_{a} \mathbf{A}^{(i)} \right) \mathcal{I}_R,
\end{aligned}
$$

- Tensor product, Kronecker product, Khatri-Rao product, unfolding
- CPD problem formulations.
- BCD approach to CPD.

End of document