

CHAU Dang Minh

TENSOR DECOMPOSITIONS
THEORY AND IMPLEMENTATION

1 CP Decomposition

By convention, let $\mathbf{A} \in \mathbb{R}^{I \times J}$ be a matrix, we shall denote by $\mathbf{a}_j \in \mathbb{R}^I$, where $1 \leq j \leq J$ the j -th column of \mathbf{A} . Let $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ be an order-3 tensor. The CP decomposition seeks for matrices $\mathbf{A} \in \mathbb{R}^{I \times R}$, $\mathbf{B} \in \mathbb{R}^{J \times R}$ and $\mathbf{C} \in \mathbb{R}^{K \times R}$ such that

$$\mathcal{X} = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \quad (1)$$

such that R is the smallest possible number. In such case, R is called the *rank* of \mathcal{X} . It is usually useful to standardize the column vectors of the matrices to unit norm by introducing scaling factors $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_R)^\top$, i.e.

$$\mathcal{X} = \sum_{r=1}^R \lambda_r \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r. \quad (2)$$

The decomposition can be described briefly as $\mathcal{X} = [[\boldsymbol{\lambda}, \mathbf{A}, \mathbf{B}, \mathbf{C}]]$. Let $\Lambda = \text{diag}(\boldsymbol{\lambda})$, we have the equivalent matricization expressions

$$\begin{cases} \mathbf{X}_{(1)} = \mathbf{A}\Lambda(\mathbf{C} \odot \mathbf{B})^\top \\ \mathbf{X}_{(2)} = \mathbf{B}\Lambda(\mathbf{C} \odot \mathbf{A})^\top \\ \mathbf{X}_{(3)} = \mathbf{C}\Lambda(\mathbf{B} \odot \mathbf{A})^\top. \end{cases} \quad (3)$$

There is no known finite algorithm for determining the rank of a tensor [2]. Therefore, for each $R = 1, 2, \dots$, we attempt to minimize the difference between \mathcal{X} and a rank- R tensor $\hat{\mathcal{X}} = [[\boldsymbol{\lambda}, \mathbf{A}, \mathbf{B}, \mathbf{C}]]$, i.e.

$$\min_{\hat{\mathcal{X}}} \|\mathcal{X} - \hat{\mathcal{X}}\|^2 \quad (4)$$

until the difference is less than a given tolerance. Using (3), we can write, for example, mode-1 matricization form of the problem as

$$\min_{\hat{\mathcal{X}}} \|\mathbf{X}_{(1)} - \mathbf{A}\Lambda(\mathbf{C} \odot \mathbf{B})\|. \quad (5)$$

The solution is

$$\mathbf{A}\Lambda = \mathbf{X}_{(1)}[(\mathbf{C} \odot \mathbf{B})^\top]^\dagger = \mathbf{X}_{(1)}(\mathbf{C} \odot \mathbf{B})(\mathbf{C}^\top \mathbf{C} \star \mathbf{B}^\top \mathbf{B})^\dagger \quad (6)$$

We use the last expression because it requires calculating the pseudoinverse of an $R \times R$ matrix rather than a $JK \times R$ matrix. Therefore, we can solve for each factor iteratively until a convergence criterion is met, particularly, the difference between the original tensor and the reconstructed tensor. The general case for N -way tensor is represented in the algorithm below.

Algorithm 1 CP Decomposition using Alternative Least Square

Input: $\mathcal{X} \in \mathbb{R}^{I_1, \dots, I_N}$, $\epsilon = 1e - 10$, $\text{maxIter} = 5000$

Initialize $\mathbf{A}_{(n)} \in \mathbb{R}^{I_n \times R}$ for $n = 1, \dots, N$ randomly

$\boldsymbol{\lambda} \leftarrow \mathbf{0}_R$

iter $\leftarrow 0$

while $\|\mathcal{X} - [[\boldsymbol{\lambda}, \mathbf{A}_{(1)}, \dots, \mathbf{A}_{(N)}]]\| > \epsilon$ and iter $< \text{maxIter}$ **do**

for $n = 1, \dots, N$ **do**

$\mathbf{V} \leftarrow \mathbf{A}_{(1)}^\top \mathbf{A}_{(1)} \star \dots \star \mathbf{A}_{(n-1)}^\top \mathbf{A}_{(n-1)} \star \mathbf{A}_{(n+1)}^\top \mathbf{A}_{(n+1)} \star \dots \star \mathbf{A}_{(N)}^\top \mathbf{A}_{(N)}$

$\mathbf{A}_{(n)} \leftarrow \mathbf{X}_{(n)}(\mathbf{A}_{(N)} \odot \dots \odot \mathbf{A}_{(n+1)} \mathbf{A}_{(n-1)} \odot \dots \odot \mathbf{A}_{(1)})\mathbf{V}^\dagger$

$\boldsymbol{\lambda} \leftarrow (\mathbf{a}_{(n)1}, \dots, \mathbf{a}_{(n)r})^\top$

 Normalize columns of $\mathbf{A}_{(N)}$

end for

end while

Output: $\boldsymbol{\lambda}, \mathbf{A}_{(1)}, \dots, \mathbf{A}_{(N)}$.

2 Tucker Decompositions: HOSVD and HOOI

Let $R_n = \text{rank}(\mathbf{X}_{(n)})$. Tucker decomposition of ranks S_1, \dots, S_N , where $S_n \leq R_n$ seeks a tensor $\mathcal{G} \in \mathbb{R}^{S_1 \times \dots \times S_N}$ and matrices $\mathbf{A}_{(n)} \in \mathbb{R}^{I_n \times S_n}$, where $n = 1, \dots, N$ to minimize

$$\|\mathcal{X} - \mathcal{G} \times_1 \mathbf{A}_{(1)} \dots \times_N \mathbf{A}_{(N)}\|. \quad (7)$$

The decomposition is written briefly as

$$\mathcal{X} \approx [[\mathcal{G}, \mathbf{A}_{(1)}, \dots, \mathbf{A}_{(N)}]]. \quad (8)$$

If $S_n = R_n$, for all $n = 1, \dots, N$, we have an exact such decomposition, called the higher-order SVD of \mathcal{X} (HOSVD).

Algorithm 2 Higher-order SVD

Input: $\mathcal{X} \in \mathbb{R}^{I_1, \dots, I_N}$
for $n = 1, \dots, N$ **do**
 $\mathbf{A}_{(n)} \leftarrow R_n$ leading left singular vectors of $\mathbf{X}_{(n)}$
end for
 $\mathcal{G} = \mathcal{X} \times_1 \mathbf{A}_{(1)}^\top \dots \times_N \mathbf{A}_{(N)}^\top$
Output: $\mathcal{G}, \mathbf{A}_{(1)}, \dots, \mathbf{A}_{(N)}$.

In the case there exists $S_n < R_n$ for some $n = 1, \dots, N$, taking S_n left singular vectors of $\mathbf{X}_{(n)}$ as in HOSVD does not lead to an optimal solution. We use the Higher-order Orthogonal Iteration (HOOI) approach with HOSVD as an initialization.

Algorithm 3 Higher-order Orthogonal Iteration

Input: $\mathcal{X} \in \mathbb{R}^{I_1, \dots, I_N}$, $\epsilon = 1e - 10$, $\text{maxIter} = 5000$.
 Initialize $\mathbf{A}_{(n)}, n = 1, \dots, N$ using HOSVD.
while $\|\mathcal{X} - [[\mathcal{G}, \mathbf{A}_{(1)}, \dots, \mathbf{A}_{(N)}]]\| > \epsilon$ and $\text{iter} < \text{maxIter}$ **do**
 for $n = 1, \dots, N$ **do**
 $\mathcal{Y} \leftarrow \mathcal{X} \times_1 \mathbf{A}_{(1)}^\top \dots \times_{n-1} \mathbf{A}_{(n-1)}^\top \times_{n+1} \mathbf{A}_{(n+1)}^\top \dots \times_N \mathbf{A}_{(N)}^\top$
 $\mathbf{A}_{(n)} \leftarrow S_n$ singular vectors of $\mathbf{Y}_{(n)}$
 end for
end while
 $\mathcal{G} = \mathcal{X} \times_1 \mathbf{A}_{(1)}^\top \dots \times_N \mathbf{A}_{(N)}^\top$
Output: $\mathcal{G}, \mathbf{A}_{(1)}, \dots, \mathbf{A}_{(N)}$.

3 Implementation Details

Implementation is organized into three main classes: Tensor, Tensor Builder and Matrix Product. Dependencies are shown in Figure 1.

Testcases can be found in `testcase_generator.m`, including analytically solvable tensors to check validity and convergence. Random tensors are also used to check convergence. Some $2 \times 2 \times 2$ tensors are shown in Table 1

The last case in Table 1 is a W-state [1], where the given rank-3 tensor can be approximated arbitrarily well by rank-2 tensors. Indeed, consider the following sequence of rank-2 tensors

$$\begin{aligned} \mathcal{A}_n &= n \left(\mathbf{u} + \frac{1}{n} \mathbf{v} \right) \otimes \left(\mathbf{u} + \frac{1}{n} \mathbf{v} \right) \left(\mathbf{u} + \frac{1}{n} \mathbf{v} \right) - n \mathbf{u} \otimes \mathbf{u} \otimes \mathbf{u} \\ &= \mathcal{A} + \frac{1}{n} (\mathbf{u} \otimes \mathbf{v} \otimes \mathbf{v} + \mathbf{v} \otimes \mathbf{u} \otimes \mathbf{v} + \mathbf{v} \otimes \mathbf{v} \otimes \mathbf{u}) \\ &\xrightarrow{n \rightarrow \infty} \mathcal{A}. \end{aligned}$$

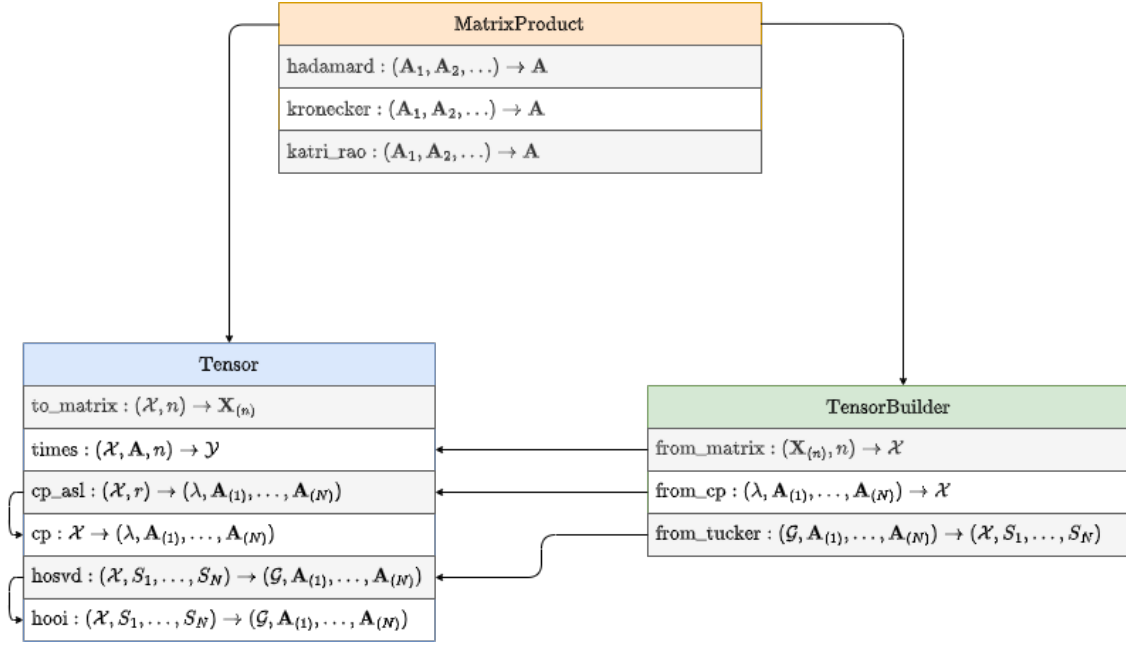


Figure 1: Dependencies between classes and methods

The result for such border rank is illustrated in Figure 2.

NO	Tensor	Analytic rank	Numerical rank
1	$\left(\begin{array}{cc cc} 1 & 2 & 3 & 6 \\ 4 & 5 & 6 & 9 \end{array} \right)$	1	1
2	$\left(\begin{array}{cc cc} 0 & 2 & 2 & 0 \\ 0 & 2 & 2 & 0 \end{array} \right)$	2	3
3	$\left(\begin{array}{cc cc} 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \end{array} \right)$	3	3
4	$\left(\begin{array}{cc cc} -1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array} \right)$	3	3
5	$\left(\begin{array}{cc cc} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{array} \right)$	unsolved	3
6	$\mathbf{u} \otimes \mathbf{u} \otimes \mathbf{v} + \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{u} + \mathbf{v} \otimes \mathbf{u} \otimes \mathbf{u}$, where $\ \mathbf{u}\ = \ \mathbf{v}\ = 1$ and $\langle \mathbf{u}, \mathbf{v} \rangle < 1$.	3	2, 3

Table 1: Some $2 \times 2 \times 2$ tensor decomposition results.

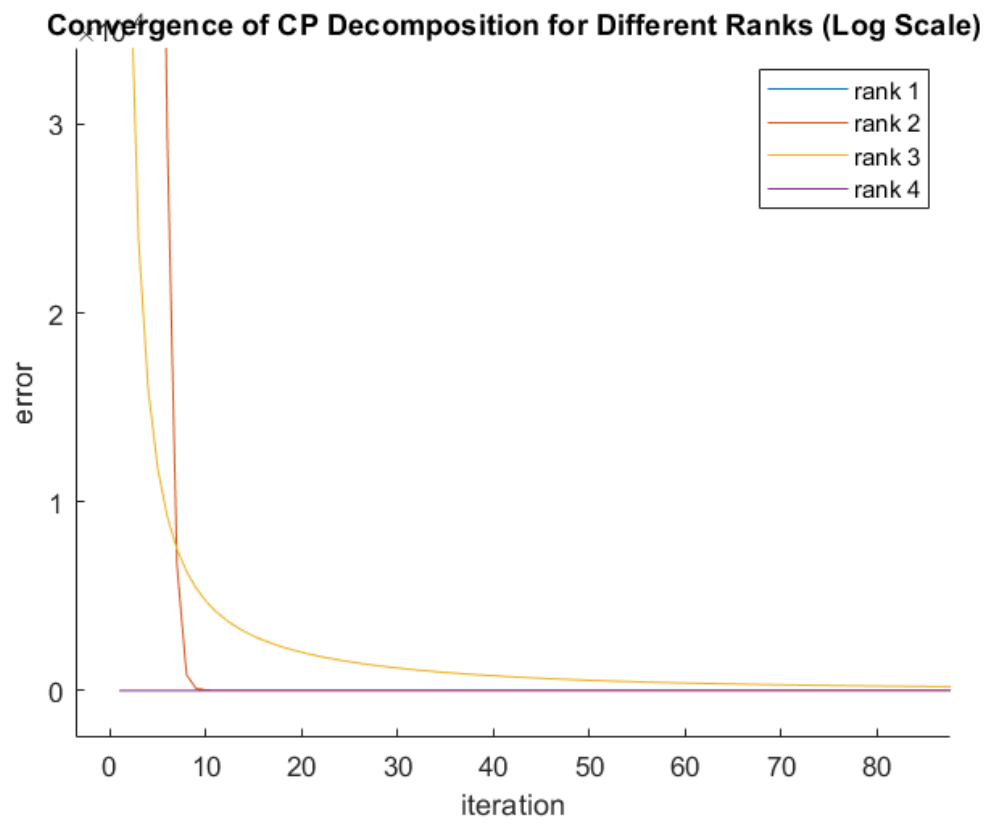


Figure 2: W-state approximation

```

>> testcase_generator
----- TESTCASE 1 -----
CP decomposition...
Tensor of dimensions:      2      2      2

Max possible rank: 4
Approximating by rank-1 tensor...
Reached maximum 5000 iterations.
Final error 8.000000e+00.
-----
Approximating by rank-2 tensor...
Converged after 1 iterations.
Final error 2.918785e-29.
-----
HOSVD...
Multilinear ranks not specified, computing max ones...
      1      2      2

HOOI with ranks...
      1      1      1

Maximum 5000 iterations reached without convergence.
Final error 8.000000e+00.
----- TESTCASE 2 -----
CP decomposition...

```

Figure 3: Test run result

References

- [1] Adán Cabello. “Bell’s theorem with and without inequalities for the three-qubit Greenberger-Horne-Zeilinger and W states”. In: *Physical Review A* 65.3 (2002), p. 032108.
- [2] Joseph B Kruskal. “Rank, decomposition, and uniqueness for 3-way and N-way arrays”. In: *Multiway data analysis*. 1989, pp. 7–18.