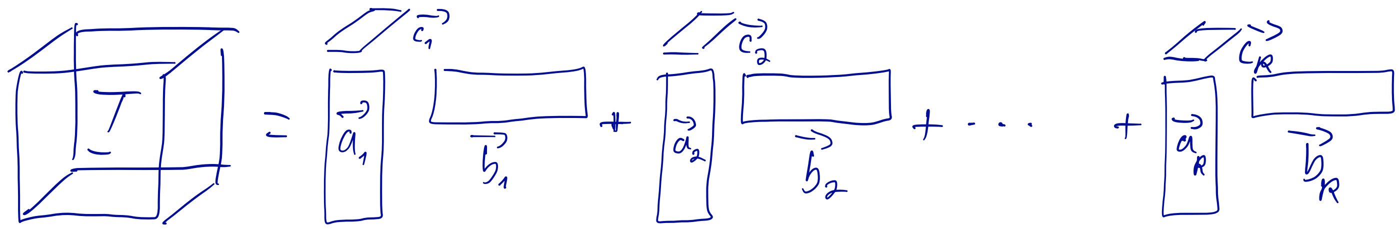
Tensor rank

An N-way tensor $\underline{T}$ is <u>rank-1</u> if it is an outer product of N vectors

$$\underline{T} = \vec{a}^{(1)} \circ \vec{a}^{(2)} \circ \cdots \circ \vec{a}^{(N)}.$$

Tensor $\underline{T}$ has <u>rank R</u> if it is a sum of R rank-1 tensors (and no less), e.g.

$$\underline{T} = \vec{a}_1 \circ \vec{b}_1 \circ \vec{c}_1 + \vec{a}_2 \circ \vec{b}_2 \circ \vec{c}_2 + \cdots + \vec{a}_R \circ \vec{b}_R \circ \vec{c}_R$$



Equivalently, the rank of a tensor $\underline{T}$ is the least R such that $\underline{T}$ has exact CP decomposition to R components. If $\underline{T}$ is all-zero tensor, its rank is agreed to be 0.

Compare this definition with that of matrix rank. The vectors don't have to be linearly independent, but the formulation is analoguous to the so-called Schein rank of a matrix: the rank of a matrix M is the least R s.t. M has decomposition $M = AB$ with A having R columns.

Tensor_rank_oddities

While seemingly similar to matrix rank, tensor rank behaves very differently in many cases.

Computational_complexity

Matrix rank is easy to compute, save precision issues, using, for instance, SVD. Tensor rank is NP-hard to compute. This also means that deciding whether a tensor has an exact rank-R decomposition is NP-hard.

$\mathbb{R}$ vs. $\mathbb{C}$

Rank of a matrix $M \in \mathbb{R}^{I \times J}$ is the same irrespective of whether we take the factorization over $\mathbb{R}$ or $\mathbb{C}$. With tensors, this is not the case. Consider

$$T_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad T_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

This tensor has rank 3 over $\mathbb{R}$, but rank 2 over $\mathbb{C}$. For example

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \text{ and } C = \begin{pmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \end{pmatrix}$$

over $\mathbb{R}$, but over $\mathbb{C}$ we have

$$A = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}, \quad B = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}, \text{ and } C = \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}.$$

Maximal rank

If $M \in \mathbb{R}^{I \times J}$, we know that $\operatorname{rank}(M) \leq \min\{I, J\}$. With tensors, this is not necessarily the case, as we see above. For $\underline{T} \in \mathbb{R}^{I \times J \times K}$, we only know the weak upper bound

$$\operatorname{rank}(\underset{\sim}{T}) \le \min\{IJ, IK, JK\}.$$

Typical rank

Typical rank is any rank that occurs with probability greater than zero if we sample over $\mathbb{R}^{I_1 \times I_2 \times \cdots I_N}$. (Notice that this is not the same as sampling over the tensors represented by the floating point numbers – the set of all of those tensors has measure zero). With matrices $M \in \mathbb{R}^{I \times J}$, typical rank is again $\min\{I, J\}$, that is, all random matrices have a full rank. With tensors, this is not the case. For instance, tensors in $\mathbb{R}^{2\times 2\times 2}$ have typical ranks 2 and 3 over $\mathbb{R}$ (experiments suggest that about 79% of 2-by-2-by-2 tensors have rank 2 and 21% have rank 3; rank-1 tensors occur with zero probability).

Uniqueness of the rank decomposition

Matrix factorizations are generally <u>not</u> unique: if $M = XY$, we can always have

$$M = X'Y', \text{ where } X' = XZ \text{ and } Y' = Z^{-1}Y$$

for some invertible Z. SVD is unique only because of the orthogonality constraints and the scaling matrix Σ.

CP decomposition, on the other hand, is often unique (up to self-cancelling scaling and permutation of the factors). It is always possible to scale

$$\underline{T} = \sum_{r=1}^{R} (\alpha_r \vec{a}_r) \circ (\beta_r \vec{b}_r) \circ (\gamma_r \vec{c}_r),$$

provided that $\alpha_r \beta_r \gamma_r = 1$ for all $r \in [R]$.
We can also permute the components

$$\underline{T} = \sum_{r=1}^{R} \vec{a}_{\sigma(r)} \circ \vec{b}_{\sigma(r)} \circ \vec{c}_{\sigma(r)}$$

for any permutation $\sigma : [R] \to [R]$.

A sufficient condition for the uniqueness of the exact CP decomposition can be expressed using the concept of a <u>k-rank</u>: the k-rank of a matrix A, denoted k_A, is the largest k such that <u>any</u> k columns of A are linearly independent (cf. normal rank, that requires that <u>some</u> k columns are lin. independent). The condition for 3-way CP decomposition $\underline{T} = [\![A, B, C]\!]$ is

$$k_A + k_B + k_C \geq 2R + 2.$$

As $\max\{k_A, k_B, k_C\} \leq R$, it's enough that, e.g., A and B have full rank and C has $k_C = 2$.

For N-way tensors, the sufficient condition is

$$\sum_{n=1}^{N} k_{A^{(n)}} \geq 2R + (N-1).$$

A necessary condition in 3-way case is

$$\min\{\mathrm{rank}(A \odot B), \mathrm{rank}(A \odot C), \mathrm{rank}(B \odot C)\} = R.$$

Border rank

In approximate decompositions, the situation is reversed. The Eckart-Young theorem states that the best rank-R approximation of a matrix is its rank-R truncated SVD. This provides a hierarchy: best rank-$(R-1)$ factorization is a part of the best rank-R approximation. CP decomposition doesn't have such hierarchy: the best rank-1 approximation might not be part of any higher-rank optimal approximations, for example.

The Eckart-Young theorem also shows that there's a clear difference between the best rank-$(R-1)$ and rank-R decomposition:

$$\| U_{R-1} \Sigma_{R-1} V_{R-1}^T - U_R \Sigma_R V_R^T \|_\xi = \sigma_R \quad \text{for } \xi \in \{2, F\}.$$

With tensors, it is possible to get arbitrarily close to the one higher rank decomposition. For an example, consider $\underline{T} \in \mathbb{R}^{I \times J \times K}$ with

$$\underline{T} = \vec{a}_1 \circ \vec{b}_1 \circ \vec{c}_2 + \vec{a}_1 \circ \vec{b}_2 \circ \vec{c}_1 + \vec{a}_2 \circ \vec{b}_1 \circ \vec{c}_1,$$

where the columns of $A, B,$ and C are linearly independent. Hence $\mathrm{rank}(\underline{T}) = 3$.
Let

$$\underline{S} = \alpha\left(\vec{a}_1 + \frac{1}{\alpha}\vec{a}_2\right) \circ \left(\vec{b}_1 + \frac{1}{\alpha}\vec{b}_2\right) \circ \left(\vec{c}_1 + \frac{1}{\alpha}\vec{c}_2\right) - \alpha \vec{a}_1 \circ \vec{b}_1 \circ \vec{c}_1$$

Now $\mathrm{rank}(\underline{S}) = 2$ and

$$\|\underline{T} - \underline{S}\| = \frac{1}{\alpha}\left\|\vec{a}_2 \circ \vec{b}_2 \circ \vec{c}_1 + \vec{a}_2 \circ \vec{b}_1 \circ \vec{c}_2 + \vec{a}_1 \circ \vec{b}_2 \circ \vec{c}_2 + \frac{1}{\alpha}\vec{a}_2 \circ \vec{b}_2 \circ \vec{c}_2\right\| \xrightarrow{\alpha \to \infty} 0.$$

Hence, we can make $\underline{S}$ arbitrarily close to $\underline{T}$. Such $\underline{T}$ are called <u>degenerate</u>. The degenerate matrices have positive Lebesgue measure (positive probability) for at least some ranks. Thus, the problem is not "rare".

The <u>border rank</u> is defined as the minimum number of rank-1 tensors needed to obtain arbitrarily good approximation of the tensor. Tensor $\underline{I}$ in the previous example has border rank 2.

Notice that this does not contradict the uniqueness of the rank decomposition: the exact decomposition is unique, even if the approximate ones are not.