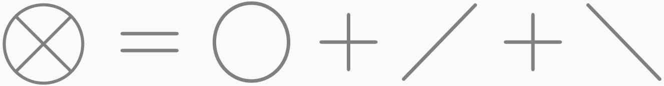
What is Tensor Decomposition?

"What is..." mathematics talks at CSB Dresden

MAX PLANCK INSTITUTE
FOR MATHEMATICS
IN THE SCIENCES

Leonie Kayser
leokayser.github.io

March 8, 2024



Warm-up: Matrix decomposition

- **Given:** matrix $M \in \mathbb{C}^{m \times n}$ (representing m student's n test results)
- **Goal:** Uncover (low-rank) decomposition/explanation "students ⇝ skill ⇝ test"
- Matrix decomposition $M = UV^\intercal$ with $U \in \mathbb{C}^{m \times r}$, $V \in \mathbb{C}^{n \times r}$
- Possible if (and only if) $\operatorname{rk} M \leq r$

$$M = S \cdot \begin{bmatrix} 1 & 0 & & & 0 \\ 0 & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ 0 & & & & 0 \end{bmatrix} \cdot T, \qquad S \in \mathrm{GL}(m, \mathbb{C}), T \in \mathrm{GL}(n, \mathbb{C})$$

- Similar to SVD: $M = U\Sigma V^\dagger$ (U, V unitary, Σ diagonal with singular values)

What is a tensor?

A tensor...

- ...is an object that transforms like a tensor
- ...is an element of a tensor product of vector spaces $\mathbb{U} \otimes \mathbb{V} \otimes \mathbb{W}$
- ...is a bilinear map $\mathbb{U} \times \mathbb{V} \to \mathbb{W}^*$ / a trilinear map $\mathbb{U} \times \mathbb{V} \times \mathbb{W} \to \mathbb{C}$ / ...
- ...is a multidimensional array of numbers $A = (A_{ijk})_{i,j,k}$
  - 1-tensor = vector
  - 2-tensor = matrix
  - 3-tensor = cube of numbers

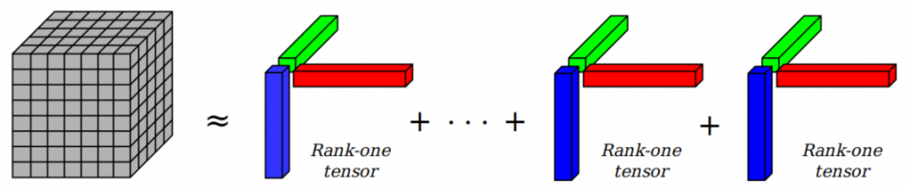
This is Tensor Decomposition!

- A tensor of the form $u \otimes v \otimes w \hat{=} (u_i v_j w_k)_{i,j,k}$ is simple
- **Given** a tensor, **find** a short(est) sum of simple tensors

$$T = \sum_{i=1}^{r} \lambda_i u_i \otimes v_i \otimes w_i \qquad (\heartsuit)$$

- The smallest such r is the tensor rank of T, $(\heartsuit)$ is a tensor decomposition
- Many names: Canonical Polyadic Decomposition, Parallel Factorization, . . .
- Generalizes matrix rank: $A = S \cdot \operatorname{diag}(\underbrace{1, \ldots, 1}_{\operatorname{rank} A}, 0, \ldots, 0) \cdot T = \sum_{i=1}^{r} u_i v_i^\intercal$
- If the simple tensors are unique up to scaling, then A is called identifiable

A small example



$$\begin{bmatrix}2&3\\4&5\end{bmatrix}\begin{bmatrix}0&1\\0&3\end{bmatrix} = \begin{bmatrix}0&-1\\0&-3\end{bmatrix}\begin{bmatrix}0&1\\0&3\end{bmatrix} + \begin{bmatrix}2&4\\4&8\end{bmatrix}\begin{bmatrix}0&0\\0&0\end{bmatrix}$$

$$= \begin{bmatrix}1\\3\end{bmatrix} \otimes \begin{bmatrix}0&1\end{bmatrix} \otimes \begin{bmatrix}-1\end{bmatrix}\begin{bmatrix}1\end{bmatrix} + \begin{bmatrix}2\\4\end{bmatrix} \otimes \begin{bmatrix}1&2\end{bmatrix} \otimes \begin{bmatrix}0\end{bmatrix}\begin{bmatrix}1\end{bmatrix}$$

The simultaneous decomposition algorithm

- **Notation:** Arrange the r vectors each into a matrix:

$$T = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i =: [\![U, V, W]\!], \qquad U \in \mathbb{C}^{l \times r},\ V \in \mathbb{C}^{m \times r},\ W \in \mathbb{C}^{n \times r}$$

- Assume U, V, W have maximal rank $= r$, then unique up to scaling & permuting
- Algorithm:
  1. Pick random $c, c' \in \mathbb{C}^n$
  2. $T(\cdot,\cdot,c) := \left(\sum_{k=1}^{n} T_{ijk} c_k\right)_{i,j} = U D_c V^\intercal, \quad T(\cdot,\cdot,c') := \left(\sum_{k=1}^{n} T_{ijk} c'_k\right)_{i,j} = U D_{c'} V^\intercal$
  3. $A := T(\cdot,\cdot,c) \cdot T(\cdot,\cdot,c')^{-1}, \quad B := T(\cdot,\cdot,c)^\intercal \cdot T(\cdot,\cdot,c')^{-\intercal}$ (really: Moore–Penrose inverse)

$$A = U D_c V^\intercal (V^\intercal)^{-1} D_{c'}^{-1} U^{-1} = U D_A U^{-1}$$

  4. Diagonalize $A = U D_A U^{-1}$ and $B = V D_B V^{-1}$

     Eigenbasis $\tilde{U}$ of A is U up to scaling & permutation, eigenbasis $\tilde{V}$ of B is V
  5. $D_A \hat{=} D_B \rightsquigarrow$ Order/pair eigenvectors to get matrices $\tilde{U}, \tilde{V}$
  6. Find $\tilde{W}$ using linear equation $T = \sum_{i=1}^{r} \tilde{u}_i \otimes \tilde{v}_i \otimes \tilde{w}_i$

The example, demystified

- $T = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 3 \end{bmatrix}$
- $c = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \rightsquigarrow T(\cdot, \cdot, c) = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}, \quad c' = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \rightsquigarrow T(\cdot, \cdot, c') = \begin{bmatrix} 2 & 1 \\ 4 & -1 \end{bmatrix}$
- $A = T(\cdot, \cdot, c) \cdot T(\cdot, \cdot, c')^{-1} = \frac{1}{3} \begin{bmatrix} 7 & -2 \\ 12 & -3 \end{bmatrix}$ $\rightsquigarrow$ eigenpairs $(1, \begin{bmatrix} 1 \\ 2 \end{bmatrix}), (\frac{1}{3}, \begin{bmatrix} 1 \\ 3 \end{bmatrix})$
- $B = T(\cdot, \cdot, c)^{\intercal} \cdot T(\cdot, \cdot, c')^{-\intercal} = \frac{1}{3} \begin{bmatrix} 3 & 0 \\ 4 & 1 \end{bmatrix}$ $\rightsquigarrow$ eigenpairs $(\frac{1}{3}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}), (1, \begin{bmatrix} 1 \\ 2 \end{bmatrix})$
- Actual decomposition: $U = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, V = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}, W = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$

Hidden algebraic geometry: joins and secants

- Let $X, Y \subseteq \mathbb{C}^N$ be affine cones (= zero sets of homogeneous polynomials)
- $J^\circ(X,Y) \coloneqq \{\, x + y \mid x \in X, y \in Y \,\} = \bigcup$ lines through X, Y
- How large is $J^\circ(X,Y)$? In how many ways can one write $p \in J^\circ$ as $x + y$?
- If $\dim X = n$ and $\dim Y = m$, then $\dim J^\circ(X,Y) \leq \min\{m+n, N\}$

Example

- $X = Y =$ simple tensors in $\mathbb{C}^{2\times2\times2}$, $\dim X = 4$
- Expected dimension of $\sigma_2^\circ X \coloneqq J^\circ(X,X)$ is $4 + 4 = 8$
- But actually contained in hypersurface of dim 7
- A random tensor lies outside this hypersurface and has rank 3

Applications and outlook

- Plenty of applications: Machine learning, (algebraic) statistics, signal processing, phylogenetic trees, ...
- Special kinds of tensors: Symmetric tensor decomposition = Waring decomposition

$$F \in \mathbb{C}[x_1, \ldots, x_n]_d, \qquad F = \sum_{i=1}^{r} L_i^d, \qquad L_i \in \mathbb{C}[x_1, \ldots, x_n]_1$$

- Dimension and identifiability questions still open for many cases $\mathbb{C}^{l \times m \times n}$
- Often interested in best-approximations with low rank $\rightsquigarrow$ generalized SVD
- Limits of rank r tensors may have rank $< r$

$$T = w \otimes v \otimes v + v \otimes w \otimes v + v \otimes v \otimes w = \lim_{n \to \infty} n(v + \tfrac{1}{n} w)^{\otimes 3} - nv \otimes v \otimes v$$

- Many other notions of rank: Border rank, subrank, cactus rank, slice rank, ...

Thank you! Questions?

Image credit

- Slide 4: `https://redirect.cs.umbc.edu/2019/06/talk-tensor-decomposition-of-nd-data-arrays-umbc/`