What's possible and what's not possible in tensor decompositions — a freshman's views

Lek-Heng Lim
Scientific Computing/Computational Mathematics
Stanford University

Workshop on Tensor Decompositions
American Institute of Mathematics

July 19–23, 2004

Tensor products

Vector Spaces $\supseteq$ Norm Spaces $\supseteq$ Inner Product Spaces

Tensor products of vector spaces: rank, decomposability, covariance/contravariance, symmetry/antisymmetry

Tensor products of norm spaces: Eckart-Young theorem, low-rank approximations

Tensor products of inner product spaces: orthogonal decompositions, singular value decompositions

Tensor products of other objects: modules, algebras, Banach and Hilbert spaces, C^*-algebras, vector bundles

Tensor fields, ie. tensor-valued functions: higher order derivatives, stress tensor, Riemann/Ricci curvature tensor

Stuff marked in blue are pretty much irrelevant to this workshop.

Important to distinguish between properties marked in red, for instance:

- there are many norms on tensors that do not arise from (and are incompatible with) inner products;

- rank, as defined in Slide 4, is essentially an algebraic concept while approximation is an analytic one — the fact that they are closely related for matrices doesn't necessarily carry over to higher order tensors;

- in similar vein, there's no reason to expect that a best low *orthogonal* rank approximation (w.r.t. some inner product) would turn out to be also the best low rank approximation (w.r.t. some compatible norm) cf. [Kolda 2001/2003].

Tensor products of vector spaces

$V_1, \ldots, V_k$ all real (or all complex) vector spaces, $\dim(V_i) = d_i$.

Tensor product space $V_1 \otimes \cdots \otimes V_k$ is a vector space of dimension $d_1 d_2 \ldots d_k$; element $\mathbf{t} \in V_1 \otimes \cdots \otimes V_k$ is called a tensor of order k.

A tensor of the form $v^1 \otimes \cdots \otimes v^k$ with $v^i \in V_i$ is called a decomposable tensor.

Fix a choice of basis $e^i_1, \ldots, e^i_{d_i}$ for each V_i, $i = 1, \ldots, k$, then $\mathbf{t}$ has coordinate representation

$$\mathbf{t} = \sum_{j_1=1}^{d_1} \cdots \sum_{j_k=1}^{d_k} t_{j_1,\ldots,j_k} e^1_{j_1} \otimes \cdots \otimes e^k_{j_k}.$$

The coefficients form a k-way array, $[\![t_{j_1,\ldots,j_k}]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_k}$.

A k-way array $[\![t_{j_1,\dots,j_k}]\!] \in \mathbb{R}^{d_1\times\cdots\times d_k}$ is decomposable if there exists $(a_1^1,\dots,a_{d_1}^1) \in \mathbb{R}^{d_1},\dots,(a_1^k,\dots,a_{d_k}^k) \in \mathbb{R}^{d_k}$ such that

$$t_{j_1,\dots,j_k} = a_{j_1}^1 a_{j_2}^2 \dots a_{j_k}^k.$$

Tensor product space $V_1 \otimes \cdots \otimes V_k$ is *really* a vector space of dimension $d_1 d_2 \cdots d_k$ together with a mapping of the form

$$V_1 \times \cdots \times V_k \to V_1 \otimes \cdots \otimes V_k,$$
$$(v_1,\dots,v_k) \mapsto v_1 \otimes \cdots \otimes v_k.$$

This structure is lost when one 'unfolds' or 'vectorizes'

$$\mathbb{R}^m \otimes \mathbb{R}^n \otimes \mathbb{R}^l \xrightarrow{\text{unfold}} \mathbb{R}^{mn} \otimes \mathbb{R}^l.$$

Good read: Timothy Gowers, "How to lose your fear of tensor products," `http://www.dpmms.cam.ac.uk/~wtg10/tensors3.html`

Tensorial Rank

F.L. Hitchcock, "The expression of a tensor or a polyadic as a sum of products," *J. Math. and Phys.*, **6** (1), 1927, pp. 164–189.

Definition. If $\mathbf{t} \neq \mathbf{0}$, the **rank** of $\mathbf{t}$, denoted $\mathrm{rank}(\mathbf{t})$, is defined as the minimum $r \in \mathbb{N}$ such that $\mathbf{t}$ may be expressed as a sum of r decomposable tensors:

$$\mathbf{t} = \sum_{i=1}^{r} v_i^1 \otimes \cdots \otimes v_i^k$$

with $v_i^j \in V_j$, $j = 1, \ldots, k$. We set $\mathrm{rank}(\mathbf{0}) = 0$.

Well-defined, ie. there exists a unique $r = \mathrm{rank}(\mathbf{t})$ for every $\mathbf{t} \in V_1 \otimes \cdots \otimes V_k$, and it agrees with the usual definition of matrix rank when $k = 2$.

Computing the rank of an order 3 tensor over a finite field is NP-complete while computing it over $\mathbb{Q}$ is NP-hard [Håstad 1990].

Aside: Alfeld-Trefethen Bet

Knowing the rank can be a useful thing (the concept cannot be completely replaced by other notions such as strong/free orthogonal rank)

25 June 1985 L.N. Trefethen hereby bets Peter Alfeld that by December 31, 1994, a method will have been found to solve $Ax = b$ in $O(n^{2+\varepsilon})$ operations for any $\varepsilon > 0$. Numerical stability is not required. The winner gets $\$100$ from the loser. Signed: *Peter Alfeld*, *Lloyd N. Trefethen* Witnesses: *Per Erik Koch*, *S.P. Norsett*.

Trefethen paid up in 1996. Bet renewed for another 10 years (1 January 2006).

Observe for $A = (a_{ij}), B = (b_{jk}) \in \mathbb{R}^{n\times n}$,

$$AB = \sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{i=1}^{n} a_{ik}b_{kj}e_{ij} = \sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{i=1}^{n} \varepsilon_{ik}(A)\varepsilon_{kj}(B)e_{ij}$$

where $e_{ij} = (\delta_{ip}\delta_{jq}) \in \mathbb{R}^{n\times n}$.

$$\varepsilon_k^i \otimes \varepsilon_j^k \otimes e_i^j = \sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{i=1}^{n} \varepsilon_{ik} \otimes \varepsilon_{kj} \otimes e_{ij}$$

where the first term is just the second written with Einstein summation convention (rule: sum over an index when it appears twice — once in superscript and once in subscript).

$O(n^{2+\varepsilon})$ algorithm for multiplying two $n\times n$ matrices gives $O(n^{2+\varepsilon})$ algorithm for solving system of n linear equations [Stassen 1969].

Corollary. Trefethen wins if and only if $\log_2(\mathrm{rank}(\varepsilon_k^i \otimes \varepsilon_j^k \otimes e_i^j)) \leq 2 + \varepsilon$ for ε arbitrarily small.

Best known result: $O(n^{2.376})$ [Coppersmith-Winograd 1987]

Tensor products of norm spaces

To discuss approximations, need norm on $V_1 \otimes \cdots \otimes V_k$. Assume that vector spaces $V_1, \ldots, V_k$ are equipped with norms $\|\cdot\|_1, \ldots, \|\cdot\|_k$.

Canonical norm defined first on the decomposable tensors by

$$\|v^1 \otimes \cdots \otimes v^k\| := \|v^1\|_1 \cdot \cdots \cdot \|v^k\|_k$$

and then extended to all $\mathbf{t} \in V_1 \otimes \cdots \otimes V_k$ by taking infimum over all possible representations of $\mathbf{t}$ as a sum of decomposable tensors:

$$\|\mathbf{t}\| = \inf\left\{ \sum_{i=1}^{n} \|v_i^1\|_1 \cdot \cdots \cdot \|v_i^k\|_k \;\middle|\; \mathbf{t} = \sum_{i=1}^{n} v_i^1 \otimes \cdots \otimes v_i^k \right\}.$$

Let $e_1^i, \ldots, e_{d_i}^i$ be a basis of unit vectors for each V_i, $i = 1, \ldots, k$ (ie. $\|e_j^i\|_i = 1$) and let coordinate representation of $\mathbf{t}$ be

$$\mathbf{t} = \sum_{j_1=1}^{d_1} \cdots \sum_{j_k=1}^{d_k} t_{j_1,\ldots,j_k} e_{j_1}^1 \otimes \cdots \otimes e_{j_k}^k.$$

Frobenius norm of $[\![t_{j_1,\dots,j_k}]\!] \in \mathbb{R}^{d_1\times\cdots\times d_k}$ is defined by

$$\|[\![t_{j_1,\dots,j_k}]\!]\|_F^2 := \sum_{j_1=1}^{d_1} \cdots \sum_{j_k=1}^{d_k} t_{j_1,\dots,j_k}^2.$$

Easy to see: $\|\mathbf{t}\| = \|[\![t_{j_1,\dots,j_k}]\!]\|_F$

Definition. A **best rank-r approximation** to a tensor $\mathbf{t} \in V_1 \otimes \cdots \otimes V_k$ is a tensor $\mathbf{s}_{\text{min}}$ with

$$\|\mathbf{s}_{\text{min}} - \mathbf{t}\| = \inf_{\text{rank}(\mathbf{s})\leq r} \|\mathbf{s} - \mathbf{t}\|.$$

Eckart-Young problem: find a best rank-r approximation for tensors of order k.

A fact that's often overlooked: in a norm space, the minimum distance of a point $\mathbf{t}$ to a non-closed set $\mathcal{S}$ *may not be attained* by any point in $\mathcal{S}$

Non-existence of low rank approximations

x, y two linearly independent vectors in V, $\dim(V) = 2$. Consider tensor $\mathbf{t}$ in $V \otimes V \otimes V$,

$$\mathbf{t} := x \otimes x \otimes x + x \otimes y \otimes y + y \otimes x \otimes y.$$

If unaccustomed to abstract vector spaces, may take $V = \mathbb{R}^2$, $x = (1, 0)^t$, $y = (0, 1)^t$ and

$$\mathbf{t} = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right] \in \mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \mathbb{R}^2.$$

We will show that $\operatorname{rank}(\mathbf{t}) = 3$ and that $\mathbf{t}$ has no best rank-2 approximation.

$\mathbf{t}$ is a rank-3 tensor: easy to verify.

$\mathbf{t}$ has no best rank-2 approximation: consider sequence $\{\mathbf{s}_n\}_{n=1}^{\infty}$ in $V \otimes V \otimes V$,

$$\mathbf{s}_n := x \otimes x \otimes (x - ny) + \left(x + \frac{1}{n}y\right) \otimes \left(x + \frac{1}{n}y\right) \otimes ny$$

Clear that $\operatorname{rank}(\mathbf{s}_n) \leq 2$ for all n. By multilinearity of $\otimes$,

$$
\begin{aligned}
\mathbf{s}_n = x \otimes x \otimes x - nx \otimes x \otimes y + nx \otimes x \otimes y \\
+ x \otimes y \otimes y + y \otimes x \otimes y + \frac{1}{n} y \otimes y \otimes y \\
= \mathbf{t} + \frac{1}{n} y \otimes y \otimes y.
\end{aligned}
$$

For any choice of norm on $V \otimes V \otimes V$,

$$\|\mathbf{s}_n - \mathbf{t}\| = \frac{1}{n} \|y \otimes y \otimes y\| \to 0 \qquad \text{as } n \to \infty.$$

Another example

Previous example not pathological. Examples of tensors with no best low-rank approximation easy to construct. Let $V = \operatorname{span}\{x, y, z, w\}$, $\dim(V) = 4$. Define

$$\mathbf{v} := x\otimes x\otimes x + x\otimes y\otimes z + y\otimes z\otimes x + y\otimes w\otimes z + z\otimes x\otimes y + y\otimes z\otimes w$$

and sequence

$$\begin{aligned}\mathbf{u}_n := \left(y + \frac{1}{n}x\right)\otimes\left(y + \frac{1}{n}w\right)\otimes nz + \left(y + \frac{1}{n}x\right)\otimes nx\otimes\left(x + \frac{1}{n}y\right)\\ - ny\otimes y\otimes\left(x + z + \frac{1}{n}w\right) - nz\otimes\left(x + y + \frac{1}{n}z\right)\otimes x\\ + n(y+z)\otimes\left(y + \frac{1}{n}z\right)\otimes\left(x + \frac{1}{n}w\right)\end{aligned}$$

May check that: $\operatorname{rank}(\mathbf{u}_n) \le 5$, $\operatorname{rank}(\mathbf{v}) = 6$ and $\|\mathbf{u}_n - \mathbf{v}\| \to 0$.

$\mathbf{v}$ is a rank-6 tensor that has no best rank 5 approximations.

A third example

Here's an example that can 'jump rank' by more than 1.

x, y, a, b four linearly independent vectors in V, $\dim(V) = 4$. Consider tensor $\mathbf{t}$ in $V \otimes V \otimes V$,

$$\mathbf{t} := x \otimes x \otimes x + x \otimes y \otimes y + y \otimes x \otimes y \\ + a \otimes a \otimes a + a \otimes b \otimes b + b \otimes a \otimes b.$$

$\mathbf{t}$ is a rank-6 tensor: tedious but straightforward.

$\mathbf{t}$ has no best rank-4 approximation: $\{\mathbf{s}_{m,n}\}_{m,n=1}^{\infty}$ in $V \otimes V \otimes V$,

$$\mathbf{s}_{m,n} := x \otimes x \otimes (x - my) + \left(x + \frac{1}{m}y\right) \otimes \left(x + \frac{1}{m}y\right) \otimes my \\ + a \otimes a \otimes (a - nb) + \left(a + \frac{1}{n}b\right) \otimes \left(a + \frac{1}{n}b\right) \otimes nb$$

Clearly $\operatorname{rank}(\mathbf{s}_{m,n}) \leq 4$ and $\lim_{m,n\to\infty} \mathbf{s}_{m,n} = \mathbf{t}$.

Norm independence

The choice of norm in the previous slides is inconsequential because of the following result.

Fact. All norms on finite-dimensional spaces are equivalent and thus induce the same topology (the Euclidean topology).

Since questions of convergence and whether a set is closed depend only on the topology of the space, the results here would all be independent of the choice of norm on $V_1 \otimes \cdots \otimes V_k$, which is finite-dimensional.

Exceptional cases: order-2 tensors and rank-1 tensors

Set of tensors of rank not more than r,

$$\mathcal{S}(k,r) := \{\mathbf{t} \in V_1 \otimes \cdots \otimes V_k \mid \operatorname{rank}(\mathbf{t}) \leq r\}.$$

When $k = 2$ (matrices) and when $r = 1$ (decomposable tensors), $\mathcal{S}(k,r)$ is closed — Eckart-Young problem solvable in these cases.

Proposition. For any $r \in \mathbb{N}$, the set $\mathcal{S}(2,r) = \{A \in \mathbb{R}^{m\times n} \mid \operatorname{rank}(\mathbf{s}) \leq r\}$ is closed in $\mathbb{R}^{m\times n}$ under any norm-induced topology.

Corollary. Let U and V be vector spaces. The set $\mathcal{S}(2,r) = \{\mathbf{s} \in U \otimes V \mid \operatorname{rank}(\mathbf{s}) \leq r\}$ is closed in $U \otimes V$.

Proposition. The set of decomposable tensors, $\mathcal{S}(k,1) = \{\mathbf{s} \in V_1 \otimes \cdots \otimes V_k \mid \operatorname{rank}(\mathbf{s}) \leq 1\}$, is closed in $V_1 \otimes \cdots \otimes V_k$ under any norm-induced topology.

[Thanks to Pierre Comon for help with the last proposition]

Topological properties of matrix rank

Set of tensors of rank exactly r,

$$\mathcal{R}(k,r) := \{\mathbf{t} \in V_1 \otimes \cdots \otimes V_k \mid \mathsf{rank}(\mathbf{t}) = r\}.$$

$\mathcal{R}(k,r)$ not closed even in the case $k = 2$ — higher-rank matrices converging to lower-rank ones easily constructed:

$$\begin{bmatrix} 1 & 1+\frac{1}{n} \\ 1 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \qquad \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{n} \end{bmatrix} \to \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

$\mathcal{R}(2,r)$ not closed often a source of numerical instability: the problem of defining matrix rank in a finite-precision context [Golub-Van Loan 1996], the inherent difficulty of computing a Jordan canonical form [Golub-Wilkinson 1976], may all be viewed as consequences of the fact that $\mathcal{R}(2,r)$ is not closed.

However, the closure of $\mathcal{R}(2, r)$ can be easily described. The same is not true for higher-order tensors.

Proposition. With $\mathcal{R}(2, r) = \{A \in \mathbb{R}^{m \times n} \mid \text{rank}(A) = r\}$ and $\mathcal{S}(2, r) = \{A \in \mathbb{R}^{m \times n} \mid \text{rank}(A) \leq r\}$, we have

$$\overline{\mathcal{R}(2, r)} = \mathcal{S}(2, r).$$

Here $\overline{\mathcal{R}}$ denotes the topological closure of a non-empty set $\mathcal{R}$.

That is, matrices of rank r are dense in matrices of rank $\leq r$.

Aside: ambiguity in Eckart-Young theorem

Eckart-Young problem for matrices often stated in the form

$$B_{\text{min}} = \operatorname*{argmin}_{B \in \mathcal{R}(2,r)} \|B - A\| = \operatorname*{argmin}_{\text{rank}(B)=r} \|B - A\|,$$

rather than

$$B_{\text{min}} = \operatorname*{argmin}_{B \in \mathcal{S}(2,r)} \|B - A\| = \operatorname*{argmin}_{\text{rank}(B)\leq r} \|B - A\|.$$

Latter form not uncommon either [Golub-Hoffman-Stewart 1987].

The two forms are really equivalent in practice (when rank$(A) > r$) — consequence of the fact that $\overline{\mathcal{R}(2,r)} = \mathcal{S}(2,r)$ and

$$\inf_{B \in \mathcal{R}} \|B - A\| = \inf_{B \in \overline{\mathcal{R}}} \|B - A\|.$$

Better to use the latter form — since $\mathcal{R}(2,r)$ is not closed, one runs into difficulties when rank$(A) < r$.

Topological properties of tensor rank

We have shown the following [L., 2004].

Proposition. Let $k \geq 3$ and $V_1, \dots, V_k$ be vector spaces with $\dim(V_i) \geq 2$. Then the Eckart-Young problem in $V_1 \otimes \cdots \otimes V_k$ has no solution in general (in any norm).

The result may be further refined [L., 2004].

Theorem. Let $k \geq 3$ and $2 \leq r \leq \operatorname{rank}_{\mathsf{max}}(V_1 \otimes \cdots \otimes V_k) - 1$. The set $\mathcal{S}(k, r) := \{\mathbf{s} \in V_1 \otimes \cdots \otimes V_k \mid \operatorname{rank}(\mathbf{s}) \leq r\}$ is not closed in $V_1 \otimes \cdots \otimes V_k$ in any norm-induced topology.

When $r \geq \operatorname{rank}_{\mathsf{max}}(V_1 \otimes \cdots \otimes V_k)$, $\mathcal{S}(k, r) = V_1 \otimes \cdots \otimes V_k$ and so this trivial case has to be excluded in the theorem.

Message. Eckart-Young problem has no solution in general for $k > 2$ and $r > 1$.

How about imposing orthogonality?

Assume that $V_1 \otimes \cdots \otimes V_k$ has an inner product (not always possible) and require tensor decompositions to have some form of orthogonality.

We have shown the following [L., 2004].

Result. There can be no globally convergent algorithm for determining rank, orthogonal rank or singular values of a tensor or for determining the best rank r approximation, with or without orthogonality.

Rough idea: Algorithms for finding rank or singular values constrain one to move on iso-rank or iso-singular-values surfaces (proper nomenclature: orbit under some group action).

For matrices, there is only one iso-rank surface for each value of rank, one iso-singular-values surface for each tuple of singular-values (arranged in non-increasing order). Cleverly designed algorithms will move on such surfaces and, after a finite or infinite number of steps, reach a point (e.g. rank revealing, diagonal matrix of singular values) where such information is easy to deduce.

For higher order tensors, there may be several or even infinitely many such iso-rank or iso-singular-values surfaces, all disconnected from each other. Any algorithm will be constrained to move on just one and may never reach the required solution lying on another surface.

Possible way around this problem: Sampling-based randomized algorithms [Drineas-Kannan-Mahoney, 2004]

What's possible

Order 2 tensors — best rank r approximation always exists

Order k tensors — best rank 1 approximation always exists

Efficient algorithms to find these (for data of moderate size)

What's not possible

Order k tensors — best rank r approximations may not exist, $k \geq 3$, $r \geq 2$

Globally convergent algorithms for determining rank, orthogonal rank, singular values, best low orthogonal rank approximations, $k \geq 3$, $r \geq 2$