Tamara G. Kolda
Computer Science and Mathematics Division
Oak Ridge National Laboratory
PO Box 2008, MS 6367
Oak Ridge, TN 37831

Time requested: 30
kolda@msr.epm.ornl.gov
423-574-0582
http://www.epm.ornl.gov/~kolda

Orthogonal Rank Decompositions for Tensors

Tamara G. Kolda

The theory of orthogonal rank decompositions for matrices is well understood, but the same is not true for tensors. For tensors, even the notions of orthogonality and rank can be interpreted several different ways. Tensor decompositions are useful in applications such as principal component analysis for multiway data. We present two types of orthogonal rank decompositions and describe methods to compute them. Furthermore, we conjecture an extension of the Eckart-Young theorem for one of these decompositions and provide a counterexample to show that it does not hold in the other case.

Let A be an $m_1 \times m_2 \times \cdots \times m_n$ *tensor* over $\Re$. The *order* of A is n. The *dimension* of A is $m \equiv \prod_{j=1}^{n} m_j$, and m_i is the ith *subdimension*. An element of A is specified as $A_{i_1 i_2 \cdots i_n}$ where $i_j \in \{1, 2, \ldots, m_j\}$ for $j = 1, \ldots, n$. If A and B are two tensors of the same size (that is, the order and all subdimensions are equal) then the *inner product* of A and B is defined as

$$A \cdot B \equiv \sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} \cdots \sum_{i_n=1}^{m_n} A_{i_1 i_2 \cdots i_n} B_{i_1 i_2 \cdots i_n}.$$

Correspondingly, the *norm* of A, $\|A\|$, is defined as $\|A\|^2 \equiv A \cdot A$. A tensor A is a *unit tensor* if $\|A\| = 1$. A *decomposed* tensor is a tensor that can be written as

$$x = x^{(1)} \otimes x^{(2)} \otimes \cdots \otimes x^{(n)},$$

where $x^{(j)} \in \Re^{m_j}$ for $j = 1, \ldots, n$. The vectors $x^{(j)}$ are called the *components* of x. Lower case letters denote decomposed tensors.

Lemma. *Let x and y be decomposed tensors. Then*

1. *The inner product $x \cdot y = \prod_{j=1}^{n} x^{(j)} \cdot y^{(j)}$.*
2. *The norm $\|x\| = \prod_{j=1}^{n} \|x^{(j)}\|_2$.*
3. *The sum of x and y is itself a decomposed tensor if and only if all but at most one of the components of x and y are equal (within a scalar multiple), i.e., $x^{(j)} = y^{(j)}$ for all j but at most one.*

Two unit decomposed tensors x and y are *orthogonal* if $x \cdot y = 0$. They are *strongly orthogonal* if, in addition, all components of x and y satisfy

$$x^{(j)} = y^{(j)} \text{ or } x^{(j)} \cdot y^{(j)} = 0.$$

From these definitions, we obtain two concepts of *orthogonal rank* as described in [1]. Both depend on decomposing a tensor A as,

$$A = \sum_{i=1}^{r} \sigma_i u_i, \tag{1}$$

with different restrictions on the u_i's. (Here, we abuse notation by letting the subscript i denote the tensor index rather than an index into the tensor.)

1. The minimal r for which A can be written as the weighted sum of unit tensors that are two-by-two orthogonal is the *orthogonal rank* of A, denoted $\text{rank}_\perp(A)$, and the decomposition is called the *orthogonal rank decomposition.*

2. The minimal r for which A can be written as the weighted sum of unit tensors that are two-by-two strongly orthogonal is the *strong orthogonal rank* of A, denoted $\text{rank}_{\perp\!\!\perp}(A)$, and the decomposition is called the *strongly orthogonal rank decomposition.*[1]

These decompositions can be computed using a type of "power method", the details of which are omitted here.

Example. Let a and b be two orthogonal vectors in $\Re^m$, and let $A \in \Re^m \times \Re^m \times \Re^m$ be defined by

$$\begin{aligned} A &= \sigma_1\, a \otimes b \otimes b + \sigma_2\, b \otimes b \otimes b + \sigma_3\, a \otimes a \otimes b \qquad (2) \\ &= \sqrt{\sigma_1^2 + \sigma_2^2}\; \frac{\sigma_1 a + \sigma_2 b}{\sqrt{\sigma_1^2 + \sigma_2^2}} \otimes b \otimes b + \sigma_3\, a \otimes a \otimes b, \end{aligned}$$

The first equation gives its strong orthogonal rank as three, and the second, its orthogonal rank as two.

Theorem. [1] *For a tensor A,*

$$\text{rank}_\perp(A) \leq \text{rank}_{\perp\!\!\perp}(A).$$

Furthermore, equality holds if the order of A is two. For orders greater than two, there exist tensors such that strict inequality holds.

[1] In [1], the term "free" is used rather than "strong".

Lemma. *Neither the orthogonal nor strongly orthogonal rank decompositions are unique.*

Example. The tensor in (2) from the previous example can also be expressed as

$$A = \hat{\sigma}_1 \hat{a} \otimes b \otimes b + \hat{\sigma}_2 \hat{b} \otimes a \otimes b + \hat{\sigma}_3 \hat{a} \otimes a \otimes b \ , \tag{3}$$

where

$$\hat{\sigma}_1 = \sqrt{\sigma_1^2 + \sigma_2^2} \ , \quad \hat{\sigma}_2 = \frac{\sigma_2 \ \sigma_3}{\hat{\sigma}_1} \ , \quad \hat{\sigma}_3 = \frac{\sigma_1 \ \sigma_3}{\hat{\sigma}_1} \ ,$$

$$\hat{a} = \frac{\sigma_1 \ a + \sigma_2 \ b}{\hat{\sigma}_1} \ , \quad \text{and} \quad \hat{b} = \frac{\sigma_2 \ a - \sigma_1 \ b}{\hat{\sigma}_1} \ .$$

Observe that $\hat{a} \perp \hat{b}$, so (3) is also a strongly orthogonal rank decomposition of A.

The uniqueness problem is not addressed in [1]. The partial 'fix' for lack of uniqueness is the following. Without loss of generality, assume that the σ_i's in (1) are always ordered so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. Then define the *unique* (strongly) orthogonal rank decomposition to be the (strongly) orthogonal rank decomposition that has the largest possible σ_1, and given that choice for σ_1, has the largest possible σ_2, and so forth. This decomposition is unique in the sense that the weights (σ_i's) are unique. The unit tensors are unique if and only if no two σ_i's are equal.

An Eckart-Young theorem for tensors is given in [1] for both the orthogonal rank decomposition and the strongly orthogonal rank decomposition; however, the proof in the first case is incorrect, and the second claim is false.

Conjecture. (Eckart-Young extended to tensors) *Let the unique orthogonal rank decomposition of a tensor A be given as in (1) and assume that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. Then the best orthogonal rank p ($p < r$) approximation to A satisfies*

$$\min_{\mathrm{rank}_\perp A_p = p} \|A - A_p\|^2 = \sum_{i=p+1}^{r} \sigma_i^2$$

and is given by

$$A_p \equiv \sum_{i=1}^{p} \sigma_i u_i.$$

We believe that an Eckart-Young type theorem holds in the orthogonal case, but we can produce a counterexample to an Eckart-Young theorem for strongly orthogonal rank decompositions.

Other research.

- *Asynchronous Parallel Direct Search Optimization.* A direct search method for optimization uses only function evaluations and never forms a gradient approximation. Recently such methods have gained respect within the mathematical community for their robustness and their ability to be easily parallelized. The parallel version, however, is constrained by the slowest processor because it has a synchronization step every iteration. We have developed a completely asynchronous parallel version of direct search that is guaranteed to converge (under the same assumptions as 'standard' direct search). This is joint research with Patricia Hough of Sandia National Labs and Virginia Torczon of the College of William & Mary.

- *Mathematical Methods for Karyotyping.* (See the abstract by John Conroy.) This is joint research with John Conroy of IDA Center for Computing Sciences and Dianne P. O'Leary of the University of Maryland.

- *Semidiscrete Decomposition* (SDD). The SDD is a 'storage efficient' matrix approximation defined as follows. An $m \times n$ matrix A is approximated as $A \approx \sum_{i=1}^{k} d_i x_i y_i^T$, where the d_i's are positive scalars, and the x_i's and y_i's are m- and n-vectors whose entries are constrained to be in the set $\{-1, 0, 1\}$. It has been used in image compression and information retrieval. We have obtained convergence results for the SDD. We have variations for the "weighted" problem as well as for tensors. This is joint research with O'Leary.

- *New Graph Models for Matrix Partitioning.* We propose a bipartite graph model for partitioning *rectangular* matrices as well as several partitioning algorithms for bipartite graphs. We are now proposing a hypergraph model for matrix partitioning in both the standard and rectangular case because it overcomes many of the limitations in the current models. This is joint research with Bruce Hendrickson of Sandia National Labs.

References

[1] D. Leibovici and R. Sabatier, *A singular value decomposition of a k-way array for principal component analysis of multiway data, PTA-k*, Linear Algebra Appl., 269 (1998), pp. 307–329.