From *SIAM News*, Volume 37, Number 9, November 2004

# Decomposing a Tensor

*By Misha Elena Kilmer and Carla D. Moravitz Martin*

*Gene Golub conducts a tour of the Stanford campus for workshop participants.*

Are there analogues to the SVD, LU, QR, and other matrix decompositions for tensors (i.e., higher-order or multiway arrays)? What exactly do we mean by "analogue," anyway? If such decompositions exist, are there efficient ways to compute them?

These are some of the questions asked, and partially answered, at the Workshop on Tensor Decompositions, held July 19–23, 2004, at the American Institute of Mathematics in Palo Alto, California. Gene Golub, Tammy Kolda, James Nagy, and Charles Van Loan were the organizers. About 35 people—computer scientists, mathematicians, and a broad range of scholars who use tensor decompositions in their research—had come from eleven countries to participate in the week-long workshop. Large group discussions and smaller break-out sessions were interleaved with invited talks, making for lively exchanges and a creative learning environment.

Carla Martin, a graduate student at Cornell University, opened the week with an overview of multilinear algebra and tensor decompositions, beginning with the definition of a $p$th-order tensor $\mathcal{A}$ as a multiway array with $p$ indices. A third-order tensor, for example, is written

$$\mathcal{A} = (a_{ijk}) \in \mathbb{R}^{n_1 \times n_2 \times n_3}.$$

A second-order tensor is thus a matrix, a third-order tensor a "box," and so forth. If $x$ and $y$ are real-valued vectors, it is well known that $xy^T = x \circ y$ is a rank-one matrix ("$\circ$" denotes the outer product). Similarly, if $x^{(1)}, \ldots, x^{(p)}$ are real-valued vectors, then $\mathcal{A} = x^{(1)} \circ x^{(2)} \circ \cdots \circ x^{(p)}$ is a rank-one tensor with $\mathcal{A}(i_1, i_2, \ldots, i_p) = x^{(1)}_{i_1} x^{(2)}_{i_2} \cdots x^{(p)}_{i_p}$.

In one of the first technical talks, Pieter Kroonenberg (Leiden University, the Netherlands) described many situations in education and psychology in which it is necessary, for analytical reasons, to decompose a tensor into a sum of rank-one tensors. In one example, he discussed the use of multiway data analysis to identify correlations and relationships among different factors that influence infant behavior. Kroonenberg presented stochastic–nonstochastic models and parameter-estimation models currently used in applications. He also proposed technical areas for exploration during the workshop; among them were comparison of various techniques, properties of higher-order tensors, and rank of third-order tensors.

Readers not familiar with the area should find the following background information helpful in understanding current models used for tensor decompositions. The singular value decomposition of a matrix $A$ is a well-known, rank-revealing factorization. If the SVD of a matrix $A$ is given by $A = U \Sigma V^T$, then we write

$$A = \sum_{i=1}^{R} \sigma_i \left( u^{(i)} \circ v^{(i)} \right),$$

where $u^{(i)}$ and $v^{(i)}$ are the $i$th columns of $U$ and $V$, respectively, the numbers $\sigma_i$ on the diagonal of the diagonal matrix $\Sigma$ are the singular values of $A$, and $R$ is the rank of $A$. A matrix is a tensor of order 2. One question, discussed in depth at the workshop, is how the SVD concept can be extended to $p$th-order tensors when $p > 2$. To keep the notation simple, we focus here on third-order tensors.

## A Canonical Decomposition and Tensor Rank

To begin, we write a real, third-order, $n_1 \times n_2 \times n_3$ tensor $\mathcal{A}$ as a sum of rank-one tensors:

$$\mathcal{A} = \sum_{i=1}^{R} u^{(i)} \circ v^{(i)} \circ w^{(i)}, \tag{1}$$

where $u^{(i)} \in \mathbb{R}^{n_1}$, $v^{(i)} \in \mathbb{R}^{n_2}$, and $w^{(i)} \in \mathbb{R}^{n_3}$, for $i = 1, \ldots, R$. At this point, we do not make any other assumptions, such as orthogonality, on the vectors in the expansion. When $R$ is minimal, we say that the *tensor rank* is equal to $R$. In other words, the tensor rank is the smallest number of rank-one tensors that sum to $\mathcal{A}$ in linear combination.

Tensor rank is more complicated than matrix rank when $p > 2$. During the first group discussion, numerous participants cited small examples that illustrate the complexity of tensor rank. Unlike the case of the matrix SVD, the minimum representation is not always orthogonal (i.e., the vectors $u^{(i)}$, $v^{(i)}$, $w^{(i)}$ do not necessarily form orthonormal sets). For this reason, no orthogonality constraints on the vectors $u^{(i)}, v^{(i)}, w^{(i)}$ are imposed in the decomposition (1). A decomposition of the form (1), called a CANDECOMP–PARAFAC decomposition (CANonical DECOMPosition or PARAllel FACtors model), was independently proposed in [3] and [4]. As discussed below, this decomposition is very useful in applications even without orthogonality of the vectors. In some cases, in fact, orthogonality is not even desirable.

The maximum possible rank of a tensor is not given directly by the dimensions of the tensor, again in contrast to the matrix case. Indeed, the maximum possible rank of a matrix is the minimum of the dimensions, $R \leq \min(n_1,n_2)$, whereas $R \nleq \min(n_1,n_2,n_3)$ for a third-order tensor. In fact, the maximum possible tensor rank of an arbitrary tensor is unknown, although (usually loose) upper bounds, dependent on the dimensions of the tensor, do exist. Furthermore, in a result that does not mirror the Eckart–Young theorem for matrices, Lek-Heng Lim, a graduate student at Stanford University, showed that the best rank-$K$ ($K < R$) decomposition of a tensor might not not always exist. In his talk about degenerate solutions of (1), Richard Harshman (University of Western Ontario) used a different approach to identify the conditions for which a tensor has no best rank-$K$ approximation.

*Participants in the July Workshop on Tensor Decomposition, held at the American Institute of Mathematics.*

One nice aspect of the CANDECOMP–PARAFAC (CP) model is that it is unique under mild assumptions [5], although questions about uniqueness arose both during an "impromptu" talk by Pierre Comon (University of Nice and CNRS, France) and in a group discussion of tensor rank. Nevertheless, the CP decomposition is very useful in applications. Rasmus Bro (KVL, Denmark) presented results illustrating the effectiveness of a CP model with fluorescence spectral data in chemometrics. In particular, the uniqueness conditions on the model enable researchers to determine exactly the underlying factors (true spectra) of what is being measured. Bro argued that imposing orthogonality constraints in this case would ruin the physical interpretation of the vectors. He and his colleague Claus Andersson are the creators of the $N$-way MATLAB toolbox [1], which uses an efficient implementation based on the Khatri–Rao–Bro product to compute a CP decomposition of a tensor. The toolbox is available at http://www.models.kvl.dk/source/nwaytoolbox/. Tools are available for computing decompositions, but no such algorithm exists for computing the tensor rank a priori.

## Higher-Order SVDs

In contrast to the unconstrained model (1), orthogonality in a tensor decomposition is a desirable feature for many of the applications discussed at the workshop. Unfortunately, an orthogonal decomposition of the form (1) does not always exist! Therefore, a more general form is often used to guarantee existence of an orthogonal decomposition, as well as to better model certain data. If $\mathcal{A}$ is a third-order, $n_1 \times n_2 \times n_3$ tensor, its Tucker decomposition [7] has the form

$$A = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} \sigma_{ijk}\left(u^{(i)} \circ v^{(j)} \circ w^{(k)}\right), \tag{2}$$

where $u^{(i)} \in \mathbb{R}^{n_1}$, $v^{(j)} \in \mathbb{R}^{n_2}$, and $w^{(k)} \in \mathbb{R}^{n_3}$. Orthogonality constraints are not required in the general Tucker decomposition. However, if $u^{(i)},v^{(j)},w^{(k)}$ are columns of the orthogonal matrices $U,V,W$, (2) is referred to as the higher-order singular value decomposition, or HOSVD [6]. The $N$-way MATLAB toolbox [1] computes the generic Tucker decomposition as well as the HOSVD.

The tensor $\Sigma = (\sigma_{ijk})$ is called the *core tensor*. In the special case in which $\Sigma$ is *diagonal* (i.e., $\sigma_{ijk} = 0$ unless $i = j = k$) in the HOSVD, (2) reduces to (1) with orthogonality constraints.

## Applications of Tensor Decompositions

As many speakers made clear, the HOSVD is useful in a variety of applications. Participants enjoyed an animated (literally!) talk by M. Alex O. Vasilescu (Courant Institute of Mathematical Sciences, NYU) on computer facial recognition. Starting with a database of facial images—photographs of people with different expressions or head poses, taken under different lighting conditions or from different viewpoints, Vasilescu uses the HOSVD to get a representation called *TensorFaces*, which she uses to identify an unknown facial image. Berkant Savas (Linkoping University, Sweden) presented research in which he and his adviser, Lars Elden (Linkoping), use the HOSVD in handwriting analysis.

Lieven De Lathauwer (ETIS, France), who helped coin the term HOSVD with his explanation of the Tucker decomposition in SVD terminology, gave a talk on independent component analysis (ICA) in which he showed how tensor decompositions are used in higher-order statistics. He emphasized applications of ICA, including image processing, telecommunications, and biomedical applications, e.g., magnetic resonance imaging and electrocardiography.

One of the small break-out groups was asked to consider efficient algorithms for the HOSVD. Because of storage issues with higher-order tensors, memory access and communication can be a problem when computing the HOSVD. Michael Mahoney (Yale University) and Petros Drineas (Rensselaer Polytechnic Institute) are beginning to consider this issue in their research. In an impromptu talk, Mahoney discussed the extraction of structure from tensors via random sampling.

## SVD Generalizations

If the HOSVD is one type of generalization of the SVD, participants asked, are there others? A complete answer to the question did not emerge from the workshop, although several new insights did. Vince Fernando (Oxford University), for example, introduced an SVD generalization in which a special structure is enforced on the core tensor and proposed an algorithm to produce the decomposition.

Asked to list the desirable features of an SVD-like factorization, many break-out groups started by listing useful properties of the matrix SVD before going on to consider whether similar expectations are reasonable for a tensor SVD. The discussions led to new and varied questions: (a) Is it possible to find an LU or QR tensor equivalent? (b) Should there be a geometric interpretation, as there is for the matrix SVD? (c) Should the generalization of a $p$th-order tensor singular vector be a $(p-1)$th-order tensor? (d) What is the maximum number of zeros that can be introduced into the core tensor? These are only some of the open research questions proposed at the workshop.

*Workshop participants rated the workshop a success, admitting that they had raised at least as many questions as they answered.*

Groups discussing efficient algorithms posed questions about structured tensors. Participants considered how matrices with upper triangular, symmetric, Toeplitz, Vandermonde, or Hessenburg form generalize to tensors. The consensus was that considering structured tensors may simplify the algorithms. Prompted by Pierre Comon's research, one of the small groups discussed supersymmetric tensors (i.e., tensors whose entries are invariant under any permutation of the indices) and algorithms for computing their canonical decompositions.

Along with tensor decompositions, participants spent time considering matrix factorizations. Eugene Tyrtyshnikov (Institute of Numerical Mathematics, Russian Academy of Sciences) gave a talk on using Kronecker approximations and computing matrix inverses. Because of memory constraints, the quite large matrices that frequently arise in integral equations must be represented implicitly. Tyrtyshnikov showed how the truncated Kronecker product SVD (KPSVD) can be used as an approximation to a matrix. He also described a Newton-based method for computing matrix inverse approximations. Eduardo D'Azevedo (Oak Ridge National Lab) gave an impromptu talk on fast solvers for sums of Kronecker products of matrices, encouraging participants to suggest ideas for future research directions. Marko Huhtanen (Helsinki University of Technology, Finland) gave an impromptu talk on his research for extending SVD ideas to real linear operators. Orly Alter (University of Texas at Austin) presented work in which she used the generalized SVD (GSVD) to compare and contrast certain aspects of the yeast and human cell cycles.

## Workshop Wrap-up

The workshop focused as much on algorithmic issues of tensor decompositions as on theoretical multilinear algebra. Cleve Moler (The MathWorks, Inc.) was in attendance, gathering information on the best ways to represent tensors in MATLAB. MathWorks provided MATLAB licenses for participants during the workshop. Participants were encouraged to test a preliminary MATLAB object-oriented tensor class and related operations [2] developed by Brett Bader and Tammy Kolda (Sandia National Labs). The documentation is available on the workshop Web site.

Workshop participants rated the workshop a success, admitting that they had raised at least as many questions as they answered. By the end, though, all agreed that two things had become clear. First, there are important theoretical differences between the study of matrices and the study of tensors. Second, efficient algorithms are needed, for computing existing decompositions as well as those yet to be discovered.

*AIM and the National Science Foundation supported the workshop.*

## References

[1] C.A. Andersson and R. Bro, T*he N-way toolbox for MATLAB*, Chemo. Intell. Lab. Syst., 52 (2000), 1–4.

[2] B.W. Bader and T.G. Kolda, A Preliminary Report on the Development of MATLAB Tensor Classes for Fast Algorithm Prototyping, Technical Report SAND2004–3487, Sandia National Laboratories, Livermore, CA, July 2004.

[3] J.D. Carroll and J. Chang, *Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition*, Psychometrika, 35 (1970), 283–319.

[4] R.A. Harshman, *Foundations of the PARAFAC procedure: Model and conditions for an "explanatory" multi-mode factor analysis*, UCLA Working Papers in Phonetics, 16 (1970), 1–84.

[5] J.B. Kruskal, *Rank, decomposition, and uniqueness for 3-way and N-way arrays*, in *Multiway Data Analysis*, R. Coppi and S. Bolasco (eds.), Elsevier, Amsterdam, 1989, 7–18.

[6] L. De Lathauwer, B. De Moor, and J. Vandewalle, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), 1253–1278.

[7] L.R. Tucker, *Some mathematical notes of three-mode factor analysis*, Psychometrika, 31 (1966), 279–311.

*Interested readers will find a list of conference participants, talks, and related links at http://csmr.ca.sandia.gov/~tgkolda/tdw2004/. An extensive bibliography on multiway data analysis can be found on Pieter Kroonenberg's Web page at http://three-mode.leidenuniv.nl/.*

*Misha Elena Kilmer is an assistant professor in the Department of Mathematics at Tufts University. Carla Moravitz Martin is a PhD student at the Center for Applied Mathematics, Cornell University.*