Algorithm 862: MATLAB Tensor Classes for Fast Algorithm Prototyping

BRETT W. BADER and TAMARA G. KOLDA

Sandia National Laboratories

Tensors (also known as multidimensional arrays or *N*-way arrays) are used in a variety of applications ranging from chemometrics to psychometrics. We describe four MATLAB classes for tensor manipulations that can be used for fast algorithm prototyping. The tensor class extends the functionality of MATLAB's multidimensional arrays by supporting additional operations such as tensor multiplication. The tensor_as_matrix class supports the "matricization" of a tensor, that is, the conversion of a tensor to a matrix (and vice versa), a commonly used operation in many algorithms. Two additional classes represent tensors stored in decomposed formats: cp_tensor and tucker_tensor. We describe all of these classes and then demonstrate their use by showing how to implement several tensor algorithms that have appeared in the literature.

Categories and Subject Descriptors: G.4 [Mathematics of Computing]: Mathematical Software— Algorithm design and analysis; G.1.m [Mathematics of Computing]: Numerical Analysis—Miscellaneous

General Terms: Algorithms

Additional Key Words and Phrases: Higher-order tensors, multilinear algebra, N-way arrays, MATLAB $\,$

1. INTRODUCTION

A tensor is a multidimensional or N-way array of data; Figure 1 shows a threeway array of size $I_1 \times I_2 \times I_3$. Tensors arise in many applications, including chemometrics [Smilde et al. 2004], signal processing [Chen et al. 2002], and image processing [Vasilescu and Terzopoulos 2002]. In this article, we describe four MATLAB classes for manipulating tensors: tensor, tensor_as_matrix, cp_tensor, and tucker_tensor.

This research was supported by the Mathematical, Information, and Computational Sciences Division at the United States Department of Energy and by Sandia National Laboratory, a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.

Authors' addresses: B. W. Bader, T. G. Kolda, Sandia National Laboratories, Albuquerque, NM 87185 and Livermore, CA 94551; email: {bwbader,tgkolda}@sandia.gov.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org. © 2006 ACM 0098-3500/06/1200-0635 \$5.00

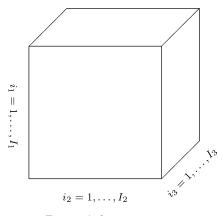


Fig. 1. A three-way array.

MATLAB is a high-level computing environment that allows users to develop mathematical algorithms using familiar mathematical notation. In terms of higher-order tensors, MATLAB supports multidimensional arrays (MDAs). Allowed operations on MDAs include elementwise operations, permutation of indices, and most vector operations (like sum and mean) [The MathWorks, Inc. 2004b]. More complex operations, such as the multiplication of two MDAs, are not supported by MATLAB. This article describes a tensor data type that extends MATLAB's MDA functionality to support tensor multiplication and more through the use of MATLAB's class functionality [The MathWorks, Inc. 2004a].

Basic mathematical notation and operations for tensors, as well as the related MATLAB commands, are described in Section 2. Tensor multiplication receives its own section, Section 3, in which we describe both notation and how to multiply a tensor times a vector, a tensor times a matrix, and a tensor times another tensor. Conversion of a tensor to a matrix (and vice versa) via the tensor_as_matrix class is described in Section 4.

A tensor may be stored in factored form as a sum of rank-one tensors. There are two commonly accepted factored forms. The first was developed independently under two names: the CANDECOMP model of Carroll and Chang [1970] and the PARAFAC model of Harshman [1970]. Following the notation in Kiers [2000], we refer to this decomposition as the CP model. The second decomposition is the Tucker [1966] model. Both models, as well as the corresponding MATLAB classes cp_tensor and tucker_tensor, are described in Section 5.

We note that these MATLAB classes serve a purely supporting role in the sense that they do not contain high-level algorithms—just the data types and their associated member functions. Thus, we view this work as complementary to those packages that provide algorithms for use with tensor data, for example, the N-way toolbox for MATLAB by Andersson and Bro [2000].

We use the following notational conventions. Indices are denoted by lowercase letters and span the range from 1 to the uppercase letter of the index, for example, n = 1, 2, ..., N. We denote vectors by lowercase boldface letters, for example, **x**; matrices by uppercase boldface, for example, **U**; and tensors by calligraphic letters, for example, A. Notation for tensor mathematics is still

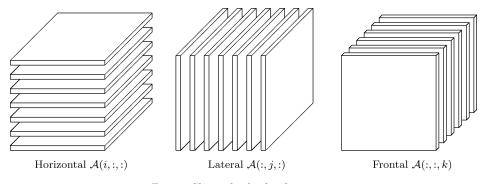


Fig. 2. Slices of a third-order tensor.

sometimes awkward. We have tried to be as standard as possible, relying on Harshman [2001] and Kiers [2000] for some guidance in this regard.

Throughout the article, we refer to numbered examples that the reader should run in MATLAB. These examples are located in the examples directory. Before running the examples, the user should change directories to the examples directory and run the setup script to set the path. The script runall runs all example scripts in sequence, and the file runall.out is a text file that shows sample output, which is useful for those who do not have MATLAB at hand or who wish to compare their output to ours. These methods have been tested on MATLAB version 7.0.1.24704 (R14) Service Pack 1.

2. BASIC NOTATION AND MATLAB COMMANDS FOR TENSORS

Let \mathcal{A} be a tensor of dimension $I_1 \times I_2 \times \cdots \times I_N$. The order of \mathcal{A} is N. The *n*th *dimension* (or *mode* or *way*) of \mathcal{A} is of size I_n .

A scalar is a zeroth-order tensor. An *n*-vector is a first-order tensor of size n. An $m \times n$ matrix is a second-order tensor of size $m \times n$. Of course, a single number could be a scalar, a 1-vector, a 1×1 matrix, etc. Similarly, an *n*-vector could be viewed as an $n \times 1$ matrix, or an $m \times n$ matrix could be viewed as a $m \times n \times 1$ tensor. It depends on the context, and our tensor class explicitly tracks the context, as described in Section 2.2.

We denote the index of a single element within a tensor by either subscripts or parentheses. Subscripts are generally used for indexing on matrices and vectors, but can be confusing for the complex indexing that is sometimes required for tensors. In general, we use $\mathcal{A}(i_1, i_2, \ldots, i_N)$ rather than $\mathcal{A}_{i_1 i_2 \cdots i_N}$.

We use colon notation to denote the full range of a given index. The *i*th row of a matrix **A** is given by $\mathbf{A}(i, :)$, and the *j*th column is $\mathbf{A}(:, j)$. For higher-order tensors, the notation is extended in an obvious way, but the terminology is more complicated. Consider a third-order tensor. In this case, specifying a single index yields a *slice* [Kiers 2000], which is a matrix in a specific orientation. So, $\mathcal{A}(i, :, :)$ yields the *i*th horizontal slice, $\mathcal{A}(:, j, :)$ the *j*th lateral slice, and $\mathcal{A}(:, :, k)$ the *k*th frontal slice; see Figure 2.

On the other hand, $\mathcal{A}(:, j, k)$ yields a column vector, $\mathcal{A}(i, :, k)$ yields a row vector, and $\mathcal{A}(i, j, :)$ yields a so-called *tube vector* [Kroonenberg 2004]; see Figure 3. Alternatively, these are called column fibers, row fibers, and depth fibers,

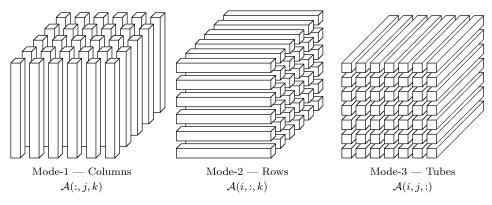


Fig. 3. Fibers of a third-order tensor.

respectively [Kiers 2000]. In general, a mode-n fiber is specified by fixing all dimensions except the nth.

2.1 Creating a tensor Object

In MATLAB, a higher-order tensor can be stored as an MDA. Our tensor class extends the capabilities of the MDA data type. An array or MDA can be converted to a tensor as follows.

T = tensor(A) or T = tensor(A,DIM) converts an array (scalar, vector, matrix, or MDA) to a tensor. Here, A is the object to be converted and DIM specifies the dimensions of the object.

A = double(T) converts a tensor to an array (scalar, vector, matrix, or MDA).

Run the script ex1 to see an example of converting an MDA to a tensor.

2.2 Tensors and Size

Out of necessity, the tensor class handles sizes differently than MATLAB arrays. Every MATLAB array has at least two dimensions; for example, a scalar is an object of size 1×1 and a column vector is an object of size $n \times 1$. On the other hand, MATLAB drops trailing singleton dimensions for any object of order greater than two. Thus, a $4 \times 3 \times 1$ object has a reported size of 4×3 . Our MATLAB tensor class explicitly stores trailing singleton dimensions; run the script ex2 to illustrate this. Furthermore, the tensor class allows for tensors of order zero (for a scalar) or one (for a vector); run the script ex3 to see this behavior. The function order returns the mathematical concept of order for a tensor, while the function ndims returns an algorithmic notion of the dimensions of a tensor, which is useful for determining the number of subscripts capable of accessing all of the elements in the data structure (i.e., one for the case of a scalar or vector). Note that the script ex3 also shows that the whos

• A + B or plus(A,B)	• A < B or lt(A,B)
• A - B or minus(A,B)	• A > B or gt(A,B)
• -A or uminus(A)	• A <= B or le(A,B)
• +A or uplus(A)	• $A \ge B \text{ or } ge(A,B)$
• A.*B or times(A,B)	• A ~= B or ne(A,B)
• A./B or rdivide(A,B)	• $A == B \text{ or } eq(A,B)$
• A.\B or ldivide(A,B)	• A & B or and(A,B)
• A. [^] B or power(A,B)	• A B or or(A,B)
	• ~A or not(A)

Fig. 4. Functions that behave identically for tensors and multidimensional arrays.

command does not report the correct sizes in the zero- or one-dimensional cases. The tensor constructor argument DIM must be specified whenever the order is intended to be zero or one or when there are trailing singleton dimensions.

2.3 Accessors

In MATLAB, indexing a tensor is the same as indexing a matrix:

```
A(i1,i2,...,iN) returns the (i_1, i_2, \ldots, i_N) element of \mathcal{A}.
```

Recall that $\mathcal{A}(:,:,k)$ denotes the *k*th frontal slice. The MATLAB notation is straightforward:

A(:,:,k) returns the *k*th submatrix (as a tensor) along the third dimension of the tensor A.

The script ex4 shows that the accessors for a tensor work generally the same as they would for an MDA.

2.4 General Functionality

In general, a tensor object will behave exactly as an MDA for all functions that are defined for an MDA; a list of these functions is provided in Figure 4.

3. TENSOR MULTIPLICATION

Notation for tensor multiplication is very complex. The issues have to do with specifying which dimensions are to be multiplied and how the dimensions of the result should be ordered. We approached this problem by developing notation that can be expressed easily by MATLAB. We describe three types of tensor multiplication: tensor times a matrix (Section 3.1), tensor times a vector (Section 3.2), and tensor times a tensor (Section 3.3). Both tensor times a matrix and tensor times a vector provide special instances of the functionality that is provided by the tensor times a tensor case.

3.1 Multiplying a Tensor Times a Matrix

The first question we consider is how to multiply a tensor times a matrix. With matrix multiplication, the specification of which dimensions should be multiplied is straightforward—it is always the inner product of the rows of the first matrix with the columns of the second matrix. A transpose on an argument swaps the rows and columns. Because tensors may have an arbitrary number of dimensions, the situation is more complicated. In this case, we need to specify which dimension of the tensor is multiplied by the columns (or rows) of the given matrix.

The adopted solution is the *n*-mode product [De Lathauwer et al. 2000a]. Let \mathcal{A} be an $I_1 \times I_2 \times \cdots \times I_N$ tensor, and let **U** be an $J_n \times I_n$ matrix. Then the *n*-mode product of \mathcal{A} and **U** is denoted by

$$\mathcal{A} \times_n \mathbf{U},$$

and defined (elementwise) as

$$(\mathcal{A} \times_n \mathbf{U})(i_1, \ldots, i_{n-1}, j_n, i_{n+1}, \ldots, i_N) = \sum_{i_n=1}^{I_n} \mathcal{A}(i_1, i_2, \ldots, i_N) \mathbf{B}(j_n, i_n).$$

The result is a tensor of size $I_1 \times \cdots \times I_{n-1} \times J_n \times I_{n+1} \times \cdots \times I_N$. Some authors call this operation the mode-*n* inner product and denote it as $\mathcal{A} \bullet_n \mathbf{U}$ (see, e.g., Comon [2001]).

To understand *n*-mode multiplication in terms of matrices (i.e., order-two tensors), suppose **A** is $m \times n$, **U** is $m \times k$, and **V** is $n \times k$. It follows that

$$\mathbf{A} \times_1 \mathbf{U}^T = \mathbf{U}^T \mathbf{A}$$
 and $\mathbf{A} \times_2 \mathbf{V}^T = \mathbf{A} \mathbf{V}$.

Further, the matrix SVD can be written as

$$\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^T = \Sigma \times_1 \mathbf{U} \times_2 \mathbf{V}.$$

The following MATLAB commands can be used to calculate *n*-mode products.

B = ttm(A,U,n) calculates "tensor times matrix" in mode-*n*, that is, $B = A \times_n U$.

 $B = ttm(A, \{U, V\}, [m, n])$ calculates two sequential *n*-mode products in the specified modes, that is, $B = A \times_m \mathbf{U} \times_n \mathbf{V}$.

The *n*-mode product satisfies the following property [De Lathauwer et al. 2000a]: Let \mathcal{A} be a tensor of size $I_1 \times I_2 \times \cdots \times I_N$. If $\mathbf{U} \in \mathbb{R}^{J_m \times I_m}$ and $\mathbf{V} \in \mathbb{R}^{J_n \times I_n}$, then

$$\mathcal{A} \times_m \mathbf{U} \times_n \mathbf{V} = \mathcal{A} \times_n \mathbf{V} \times_m \mathbf{U}.$$
 (1)

Run the script ex5 to see *n*-mode products that demonstrate this property, and run ex6 to revisit the same example using cell arrays to calculate the *n*-mode products.

It is often desirable to calculate the product of a tensor and a sequence of matrices. Let \mathcal{A} be an $I_1 \times I_2 \times \cdots \times I_N$ tensor, and let $\mathbf{U}^{(n)}$ denote a $J_n \times I_n$ matrix for $n = 1, \ldots, N$. Then the sequence of products

$$\mathcal{B} = \mathcal{A} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_N \mathbf{U}^{(N)}$$
(2)

is of size $J_1 \times J_2 \times \cdots \times J_N$. We propose new, alternative notation for this operation that is consistent with the MATLAB notation for cell arrays:

 $\mathcal{B} = \mathcal{A} \times \{ \mathbf{U} \}.$

This mathematical notation will prove useful in presenting some algorithms, as shown in Section 6.

The following equivalent MATLAB commands can be used to calculate *n*-mode products with a sequence of matrices.

B = ttm(A, {U1, U2, ..., UN}, [1:N]) calculates $\mathcal{B} = \mathcal{A} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_n \mathbf{U}^{(N)}$. Here, Un is a MATLAB matrix representing $\mathbf{U}^{(n)}$. B = ttm(A,U) calculates $\mathcal{B} = \mathcal{A} \times \{\mathbf{U}\}$. Here, U = {U1, U2, ..., UN} is a MATLAB cell array and Un is as described earlier.

Another frequently used operation is multiplying by *all but one* of a sequence of matrices:

$$\mathcal{B} = \mathcal{A} \times_1 \mathbf{U}^{(1)} \cdots \times_{n-1} \mathbf{U}^{(n-1)} \times_{n+1} \mathbf{U}^{(n+1)} \cdots \times_N \mathbf{U}^{(N)}.$$

We propose new, alternative notation for this operation:

 $\mathcal{B} = \mathcal{A} \times_{-n} \{\mathbf{U}\}.$

This notation will prove useful in presenting some algorithms in Section 6.

The following MATLAB commands can be used to calculate *n*-mode products with all but one of a sequence of matrices.

B = ttm(A,U,-n) calculates $\mathcal{B} = \mathcal{A} \times_{-n} \{U\}$. Here, $U = \{U1,U2,\ldots,UN\}$ is a MATLAB cell array; the *n*th cell is simply ignored in the computation.

Note that $B = ttm(A, \{U1, \ldots, U4, U6, \ldots, U9\}, [1:4, 6:9])$ is equivalent to B = ttm(A, U, -5), where $U = \{U1, \ldots, U9\}$; both calculate $\mathcal{B} = \mathcal{A} \times_{-5} \{U\}$.

3.2 Multiplying a Tensor Times a Vector

In our opinion, one source of confusion in n-mode multiplication is what to do to the singleton dimension in mode n that is introduced when multiplying a tensor times a vector. If the singleton dimension is dropped (as is sometimes desired), then the commutativity of the multiplies in Eq. (1) outlined in the

previous section no longer holds because the order of the intermediate result changes and \times_n or \times_m applies to the wrong mode.

Although we can usually determine the correct order of the result via the context of the equation, it is impossible to do this automatically in MATLAB in any robust way. Thus, we propose an alternate name and notation in the case when the newly introduced singleton dimension indeed should be dropped.

Let \mathcal{A} be an $I_1 \times I_2 \times \cdots \times I_N$ tensor, and let **b** be an I_n -vector. We propose that the *contracted* n-mode product, which drops the nth singleton dimension, be denoted by

$$\mathcal{A}\,\bar{\times}_n\,\mathbf{b}.$$

The result is of size $I_1 \times \cdots \times I_{n-1} \times I_{n+1} \times \cdots \times I_N$. Note that the order of the result is N - 1, one less than the original tensor. The entries are computed as:

$$(\mathcal{A} \times_{n} \mathbf{u})(i_{1}, \ldots, i_{n-1}, i_{n+1}, \ldots, i_{N}) = \sum_{i_{n}=1}^{I_{n}} \mathcal{A}(i_{1}, i_{2}, \ldots, i_{N}) \mathbf{u}(i_{n}).$$

The following MATLAB command computes the contracted *n*-mode product.

 $\mathbf{B}=\mathtt{ttv}(\mathtt{A},\mathtt{u},\mathtt{n})$ calculates "tensor times vector" in mode-n, that is, $B=\mathcal{A}\,\bar{\times}_n\,\mathbf{u}.$

Observe that $\mathcal{A} \times_n \mathbf{u}$ and $\mathcal{A} \times_n \mathbf{u}^T$ produce identical results, except for the order and shape of the results; that is, $\mathcal{A} \times_n \mathbf{u}$ is of size $I_1 \times \cdots \times I_{n-1} \times I_{n+1} \times \cdots \times I_N$, whereas $\mathcal{A} \times_n \mathbf{u}^T$ is of size $I_1 \times \cdots \times I_{n-1} \times 1 \times I_{n+1} \times \cdots \times I_N$. Run extension of two and ttm.

Because the contracted *n*-mode product drops the *n*th singleton dimension, it is no longer true that multiplication is commutative; that is,

$$(\mathcal{A} \,\bar{\times}_m \,\mathbf{u}) \,\bar{\times}_n \,\mathbf{v} \neq (\mathcal{A} \,\bar{\times}_n \,\mathbf{v}) \,\bar{\times}_m \,\mathbf{u}.$$

However, a different statement about commutativity may be made. If we assume m < n, then

$$(\mathcal{A}\,\bar{\times}_m\,\mathbf{u})\,\bar{\times}_{n-1}\,\mathbf{v}=(\mathcal{A}\,\bar{\times}_n\,\mathbf{v})\,\bar{\times}_m\,\mathbf{u}.$$

For the sake of clarity in a sequence of contracted *n*-mode products, we assume that the order reduction happens after all products have been computed. This assumption obviates the need to explicitly place parentheses in an expression and appropriately decrement any *n*-mode product indices. In other words,

$$\mathcal{A}\,\bar{\mathbf{x}}_{m}\,\mathbf{u}\,\bar{\mathbf{x}}_{n}\,\mathbf{v} \equiv \begin{cases} (\mathcal{A}\,\bar{\mathbf{x}}_{m}\,\mathbf{u})\,\bar{\mathbf{x}}_{n}\,\mathbf{v} & : & m > n, \\ (\mathcal{A}\,\bar{\mathbf{x}}_{m}\,\mathbf{u})\,\bar{\mathbf{x}}_{n-1}\,\mathbf{v} & : & m < n. \end{cases}$$
(3)

As before with matrices, it is often useful to calculate the product of a tensor and a sequence of vectors; for example,

$$\mathcal{B} = \mathcal{A} \, \bar{\times}_1 \, \mathbf{u}^{(1)} \, \bar{\times}_2 \, \mathbf{u}^{(2)} \cdots \, \bar{\times}_N \, \mathbf{u}^{(N)}$$

Algorithm 862: Tensor Classes • 643

or

$$\mathcal{B} = \mathcal{A} \,\bar{\mathbf{x}}_1 \, \mathbf{u}^{(1)} \cdots \,\bar{\mathbf{x}}_{n-1} \, \mathbf{u}^{(n-1)} \,\bar{\mathbf{x}}_{n+1} \, \mathbf{u}^{(n+1)} \cdots \,\bar{\mathbf{x}}_N \, \mathbf{u}^{(N)}$$

We propose the following alternative notation for these two cases:

 $\mathcal{B} = \mathcal{A} \,\bar{\times} \, \{ \mathbf{u} \}$

and

 $\mathcal{B} = \mathcal{A} \,\bar{\times}_{-n} \, \{\mathbf{u}\},\,$

respectively.

In practice, we must be careful, when calculating a sequence of contracted *n*-mode products, to perform the multiplications starting with the highest mode and proceeding sequentially to the lowest. The following MATLAB commands automatically sort the modes in correct order.

B = ttv(A, u1, u3, [1,3]) computes $B = A \bar{\times}_1 \mathbf{u}^{(1)} \bar{\times}_3 \mathbf{u}^{(3)}$, where u1 and u3 correspond to vectors $\mathbf{u}^{(1)}$ and $\mathbf{u}^{(3)}$, respectively.

B = ttv(A,u) computes $B = A \times \{u\}$, where u is a cell array whose *n*th entry is the vector $\mathbf{u}^{(n)}$.

 $B = ttv(A,u,-n) \text{ computes } B = \mathcal{A} \,\bar{\times}_{-n} \, \{\mathbf{u}\}.$

Note that the result of the second expression is a scalar, and the result of the third is a vector of size I_n ; run ex8 for examples of a tensor times a sequence of vectors.

3.3 Multiplying a Tensor Times Another Tensor

The last category of tensor multiplication to consider is the product of two tensors. We consider three general scenarios for tensor-tensor multiplication: outer product, contracted product, and inner product.

The outer product of two tensors is defined as follows. Let \mathcal{A} be of size $I_1 \times \cdots \times I_M$, and let \mathcal{B} be of size $J_1 \times \cdots \times J_N$. The outer product $\mathcal{A} \circ \mathcal{B}$ is of size $I_1 \times \cdots \times I_M \times J_1 \times \cdots \times J_N$ and is given by

 $(\mathcal{A} \circ \mathcal{B})(i_1, \ldots, i_M, j_1, \ldots, j_N) = \mathcal{A}(i_1, \ldots, i_M)\mathcal{B}(j_1, \ldots, j_N).$

In MATLAB, the command is as follows.

C=ttt(A,B) computes "tensor times tensor," that is, the outer product $\mathcal{C}=\mathcal{A}\circ\mathcal{B}.$

Run ex9 to see an example the outer product.

The contracted product of two tensors is a generalization of the tensor-vector and tensor-matrix multiplications discussed in the previous two subsections. The key distinction is that the modes to be multiplied and the ordering of the resulting modes is handled specially in the matrix and vector cases. In this

general case, let \mathcal{A} be of size $I_1 \times \cdots \times I_M \times J_1 \times \cdots \times J_N$ and \mathcal{B} be of size $I_1 \times \cdots \times I_M \times K_1 \times \cdots \times K_P$. We can multiply both tensors along the first M modes, and the result is a tensor of size $J_1 \times \cdots \times J_N \times K_1 \times \cdots \times K_P$, given by

$$\langle \mathcal{A}, \mathcal{B} \rangle_{\{1,\ldots,M;1,\ldots,M\}} (j_1,\ldots,j_N,k_1,\ldots,k_P)$$

= $\sum_{i_1=1}^{I_1} \cdots \sum_{i_M=1}^{I_M} \mathcal{A}(i_1,\ldots,i_M,j_1,\ldots,j_N) \mathcal{B}(i_1,\ldots,i_M,k_1,\ldots,k_P).$

With this notation, the modes to be multiplied are specified in the subscripts that follow the angle brackets. The remaining modes are ordered such that those from \mathcal{A} come before \mathcal{B} , which is different from the tensor-matrix product case considered previously, where the leftover matrix dimension of **B** replaces I_n , rather than being moved to the end. In MATLAB, the command for the tensor contracted product is as follows.

$$C = ttt(A,B,[1:M],[1:M]) \text{ computes } C = \langle A, B \rangle_{\{1,\dots,M;1,\dots,M\}}$$

The arguments specifying the modes of \mathcal{A} and those of \mathcal{B} for contraction need not be consecutive, as shown in the previous example. However, the sizes of the corresponding dimensions must be equal. That is, if we call ttt(A,B,ADIMS,BDIMS), then size(A,ADIMS) and size(B,BDIMS) must be identical. Run ex10 to see examples of contracted products.

The inner product of two tensors requires that both have equal dimensions. Assuming both are of size $I_1 \times I_2 \times \cdots \times I_N$, their inner product is given by

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} \mathcal{A}(i_1, i_2, \dots, i_N) \mathcal{B}(i_1, i_2, \dots, i_N).$$

In MATLAB this is accomplished via the following command.

ttt(A,B, [1:N]) calculates $\langle \mathcal{A}, \mathcal{B} \rangle$; the result is a scalar.

See ex11 for examples of using each notation.

Using this definition of inner product, the Frobenius norm of a tensor is then given by

$$\|\mathcal{A}\|_{F}^{2} = \langle \mathcal{A}, \mathcal{A} \rangle = \sum_{i_{1}=1}^{I_{1}} \sum_{i_{2}=1}^{I_{2}} \cdots \sum_{i_{N}=1}^{I_{N}} \mathcal{A}(i_{1}, i_{2}, \dots, i_{N})^{2}.$$

In MATLAB the norm can be calculated as follows.

norm(A) calculates $\|\mathcal{A}\|_F$, the Frobenius norm of a tensor.

4. MATRICIZE: TRANSFORMING A TENSOR INTO A MATRIX

It is often useful to rearrange the elements of a tensor so that they form a matrix. Although many names for this process exist, we call it "matricizing," following Kiers [2000], because matricizing a tensor is analogous to vectorizing a matrix. De Lathauwer et al. [2000a] call this process "unfolding." It is sometimes also called "flattening" (see, e.g., Vasilescu and Terzopoulos [2002]).

4.1 General Matricize

Let \mathcal{A} be an $I_1 \times I_2 \times \cdots \times I_N$ tensor, and suppose we wish to rearrange this to be a matrix of size $J_1 \times J_2$ (or a vector if $J_2 = 1$). Clearly, the number of entries in the matrix must equal the number of entries in the tensor; in other words, $\prod_{n=1}^{N} I_n = J_1 J_2$. Given J_1 and J_2 satisfying the aforementioned property, the mapping can be done any number of ways, so long as we have a one-to-one mapping π such that

$$\pi: \{1,\ldots,I_1\}\times\{1,\ldots,I_2\}\times\cdots\times\{1,\cdots,I_N\}\to\{1,\ldots,J_1\}\times\{1,\ldots,J_2\}.$$

The tensor_as_matrix class supports the conversion of a tensor to a matrix as follows. Let the set of indices be partitioned into two disjoint subsets: $\{1, \ldots, N\} = \{r_1, \ldots, r_K\} \cup \{c_1, \ldots, c_L\}$. The set $\{r_1, \ldots, r_K\}$ defines those indices that will be mapped to the row indices of the resulting matrix and the set $\{c_1, \ldots, c_L\}$ defines those that will likewise be mapped to the column indices. In this case,

$$J_1 = \prod_{k=1}^K I_{r_k} \quad ext{and} \quad J_2 = \prod_{\ell=1}^L I_{c_\ell}$$

Then we define $\pi(i_1, i_2, ..., i_N) = (j_1, j_2)$, where

$$j_1 = 1 + \sum_{k=1}^{K} \left[(i_{r_k} - 1) \prod_{\hat{k}=1}^{k-1} I_{r_{\hat{k}}} \right] \quad ext{and} \quad j_2 = 1 + \sum_{\ell=1}^{L} \left[(i_{r_\ell} - 1) \prod_{\hat{\ell}=1}^{\ell-1} I_{r_{\hat{\ell}}} \right].$$

Note that the sets $\{r_1, \ldots, r_K\}$ and $\{c_1, \ldots, c_L\}$ can be in any order and are not necessarily ascending. The following MATLAB commands convert a tensor to a matrix, and ex12 shows some examples.

 $A = tensor_as_matrix(T, RDIMS)$ matricizes T such that the dimensions (or modes) specified in RDIMS map to the rows of the matrix (in the order given), and the remaining dimensions (in ascending order) map to the columns.

 $A = \texttt{tensor_as_matrix}(\texttt{T},\texttt{RDIMS},\texttt{CDIMS})$ matricizes $\mathcal T$ such that the dimensions specified in <code>RDIMS</code> map to the rows of the matrix, and the dimensions specified in <code>CDIMS</code> map to the columns, both in the order given.

A tensor_as_matrix object can be converted to a matrix as follows.

B = double(A) converts the tensor_as_matrix object to a matrix.

Also, the size of the corresponding tensor, the tensor indices corresponding to the matrix rows, and tensor indices corresponding to the matrix columns can be extracted as follows.

sz = A.tsize gives the size of the corresponding tensor.

ridx = A.rindices gives the indices that have been mapped to the rows, that is, $\{r_1, \ldots, r_K\}$.

cidx = A.cindices gives the indices that have been mapped to the columns, that is, $\{c_1, \ldots, c_L\}$.

With overloaded functions in MATLAB, the tensor_as_matrix class allows multiplication between tensors and/or matrices. More precisely, mtimes(A,B) is called for the syntax A * B when A or B is a tensor_as_matrix object. The result is another tensor_as_matrix object that can be converted back into a tensor object, as described next. The multiplication is analogous to the functionality provided by ttt for multiplying two tensor objects. Run ex13 to see tensor-tensor multiplication using tensor_as_matrix objects.

Given a tensor_as_matrix object, we can automatically rearrange its entries back into a tensor by passing the tensor_as_matrix object into the constructor for the tensor class. The tensor_as_matrix class contains the mode of matricization and original tensor dimensions, making the conversion transparent to the user. The following MATLAB command can convert a matrix to a tensor, and ex14 shows such a conversion.

tensor(A) creates a tensor from A, which is a tensor_as_matrix object.

4.2 Mode-*n* Matricize

Typically, a tensor is matricized such that all of the fibers associated with a particular *single* dimension are aligned as columns of the resulting matrix. In other words, we align the fibers of dimension n of a tensor \mathcal{A} to be the columns of the matrix. This is a special case of the general matricize, where only one dimension is mapped to the rows so that K = 1 and $\{r_1\} = \{n\}$. The resulting matrix is typically denoted by $\mathbf{A}_{(n)}$.

The columns can be ordered in any way. Two standard, but different, orderings are used by De Lathauwer et al. [2000a] and Kiers [2000]. Both are cyclic, but the order is reversed. For De Lathauwer et al. [2000a], the ordering is given

Algorithm 862: Tensor Classes • 647

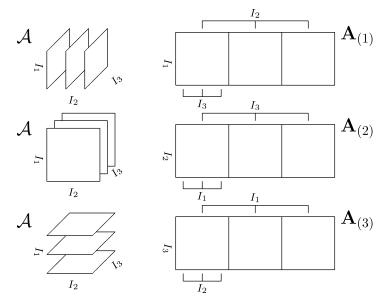


Fig. 5. Backward cyclic matricizing a three-way tensor.

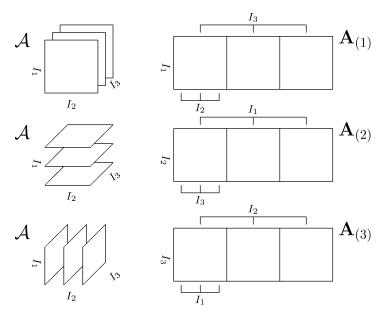


Fig. 6. Forward cyclic matricizing a three-way tensor.

by $\{c_1, \ldots, c_L\} = \{n - 1, n - 2, \ldots, 1, N, N - 1, \ldots, n + 1\}$, and we refer to this ordering as *backward cyclic*, or "bc" for short. For Kiers [2000], the ordering is given by $\{c_1, \ldots, c_L\} = \{n + 1, n + 2, \ldots, N, 1, 2, \ldots, n - 1\}$, and we refer to this ordering as *forward cyclic* or "fc" for short. Figure 5 shows backward cyclic, ordering and Figure 6 shows forward cyclic ordering.

The following MATLAB commands can convert a tensor to a matrix according to the two preceding definitions, and ex15 shows two examples of matricizing a tensor.

tensor_as_matrix(T,n, 'bc') computes $\mathbf{T}_{(n)}$, that is, matricizing \mathcal{T} using a backward cyclic ordering. The equivalent general command is tensor_as_matrix(T,n, [n-1:-1:1 ndims(T):-1:n+1]).

tensor_as_matrix(T,n, 'fc') computes $\mathbf{T}_{(n)}$, that is, matricizing \mathcal{T} using a forward cyclic ordering. The equivalent general command is tensor_as_matrix(T,n, [n+1:ndims(T) 1:n-1]).

One benefit of matricizing is that tensors stored in matricized form may be manipulated as matrices, reducing *n*-mode multiplication, for example, to a matrix-matrix operation. If $\mathcal{B} = \mathcal{A} \times_n \mathbf{M}$, then

$$\mathbf{B}_{(n)} = \mathbf{M}\mathbf{A}_{(n)}.$$

Moreover, the series of *n*-mode products in Eq. (2), when written as a matrix formulation, can be expressed as a series of Kronecker products involving the **U** matrices. Consider the ordering of the tensor dimensions that map to the column space of the matrix (e.g., for forward cyclic ordering about mode-3 on a fourth-order tensor; then $\{r_1\} = \{3\}$ and $\{c_1, c_2, c_3\} = \{4, 1, 2\}$). The series of *n*-mode products in Eq. (2) is given by

$$\mathbf{B}_{(n)} = \mathbf{U}^{(n)} \mathbf{A}_{(n)} \big(\mathbf{U}^{(c_{n-1})} \otimes \mathbf{U}^{(c_{n-2})} \otimes \cdots \otimes \mathbf{U}^{(c_1)} \big)^T.$$

The script ex16 shows an example of computing the series of *n*-mode products using tensor_as_matrix and Kronecker products. The tensor_as_matrix object is converted into a standard MATLAB matrix for matrix-matrix multiplication, and the result must be converted back to a tensor with further matrix manipulation. Because this approach requires that the user code some lowerlevel details, this example highlights the simplicity of the tensor class, which accomplishes the same computation in one function call to ttm.

5. DECOMPOSED TENSORS

As mentioned previously, we have also created two additional classes to support the representation of tensors in decomposed form, that is, as the sum of rankone tensors. A rank-one tensor is a tensor that can be written as the outer product of vectors, that is,

$$\mathcal{A} = \lambda \ \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \cdots \circ \mathbf{u}^{(N)},$$

where λ is a scalar and each $\mathbf{u}^{(n)}$ is an I_n -vector, for n = 1, ..., N. The \circ symbol denotes the outer product; so, in this case the $(i_1, i_2, ..., i_N)$ entry of \mathcal{A} is given by

$$\mathcal{A}(i_1, i_2, \ldots, i_N) = \lambda \mathbf{u}_{i_1}^{(1)} \mathbf{u}_{i_2}^{(2)} \cdots \mathbf{u}_{i_N}^{(N)},$$

where \mathbf{u}_i denotes the *i*th entry of vector \mathbf{u} . We focus on two different tensor decompositions: CP and Tucker.

5.1 CP Tensors

Recall that "CP" is shorthand for CANDECOMP [Carroll and Chang 1970] and PARAFAC [Harshman 1970], which are identical decompositions that were developed independently for different applications. The CP decomposition is a weighted sum of rank-one tensors, given by

$$\mathcal{A} = \sum_{k=1}^{K} \lambda_k \, \mathbf{U}_{:k}^{(1)} \circ \mathbf{U}_{:k}^{(2)} \circ \cdots \circ \mathbf{U}_{:k}^{(N)}.$$

$$\tag{4}$$

Here, λ is a vector of size K and each $\mathbf{U}^{(n)}$ is a matrix of size $I_n \times K$, for $n = 1, \ldots, N$. Recall that the notation $\mathbf{U}_{k}^{(n)}$ denotes the kth column of the matrix $\mathbf{U}^{(n)}$.

The following MATLAB command creates a CP tensor.

 $T = cp_tensor(lambda,U)$ creates a cp_tensor object. Here, lambda is a K-vector and U is a cell array whose nth entry is the matrix $U^{(n)}$ with K columns.

 $T = cp_tensor(U)$ creates a cp_tensor object where all the λ_k 's are assumed to be one.

A CP tensor can be converted to a dense tensor as follows, and ex17 provides an example.

B = full(A) converts a cp_tensor object to a tensor object.

Addition and subtraction of CP tensors is handled in a special manner. The λ 's and $\mathbf{U}^{(n)}$'s are concatenated. To add or subtract two CP tensors (of the same order and size), use the + and - signs. An example is shown in ex18

- A + B computes the sum of two CP tensors.
- A B computes the difference of two CP tensors.

To determine the value of K for a CP tensor, execute the following MATLAB command.

r = length(T.lambda) returns the number of components of tensor T.

5.2 Tucker Tensors

The Tucker decomposition [Tucker 1966], also called a Rank- (K_1, K_2, \ldots, K_N) decomposition [De Lathauwer et al. 2000b], is another way of summing decomposed tensors and is given by

$$\mathcal{A} = \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} \cdots \sum_{k_N=1}^{K_N} \lambda(k_1, k_2, \dots, k_N) \mathbf{U}_{:k_1}^{(1)} \circ \mathbf{U}_{:k_2}^{(2)} \circ \cdots \circ \mathbf{U}_{:k_N}^{(N)}.$$
 (5)

Here, λ is itself a tensor of size $K_1 \times K_2 \times \cdots \times K_N$, and each $\mathbf{U}^{(n)}$ is a matrix of size $I_n \times K_n$, for $n = 1, \ldots, N$. As before, the notation $\mathbf{U}_{:k}^{(n)}$ denotes the kth column of the matrix $\mathbf{U}^{(n)}$. The tensor λ is often called the "core array" or "core tensor."

A Tucker tensor can be created in MATLAB as follows, and ex19 shows an example.

 $T = tucker_tensor(lambda,U)$, where lambda is a $K_1 \times K_2 \times \cdots \times K_N$ tensor and U is a cell array whose *n*th entry is a matrix with K_n columns.

A Tucker tensor can be converted to a dense tensor as follows.

B = full(A) converts a tucker tensor object to a tensor object.

6. EXAMPLES

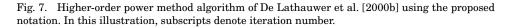
We demonstrate the use of the tensor, cp_tensor, and tucker_tensor classes for algorithm development by implementing the higher-order generalizations of the power method and orthogonal iteration presented by De Lathauwer et al. [2000b].

The first example is the higher-order power method, Algorithm 3.2 of De Lathauwer et al. [2000b], which is a multilinear generalization of the best rank-one approximation problem for matrices. This is also the same as the Alternating Least Squares algorithm for fitting a rank-one CP model. The best rank-one approximation problem is that given a tensor \mathcal{A} , we want to find a \mathcal{B} of the form

$$\mathcal{B} = \lambda \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \cdots \circ \mathbf{u}^{(N)}$$

such that $||\mathcal{A} - \mathcal{B}||$ is as small as possible. The higher-order power method computes a \mathcal{B} that approximately solves this problem. Essentially, this method works as follows. It fixes all **u**-vectors except $\mathbf{u}^{(1)}$, and then solves for the optimal $\mathbf{u}^{(1)}$, likewise for $\mathbf{u}^{(2)}$, $\mathbf{u}^{(3)}$, and so on, cycling through the indices until the specified number of iterations is exhausted. In Figure 7, we show the algorithm using our new notation, and the file examples/hopm.m shows the MATLAB code

HIGHER-ORDER POWER METHOD In: \mathcal{A} of size $I_1 \times I_2 \times \cdots \times I_N$. Out: \mathcal{B} of size $I_1 \times I_2 \times \cdots \times I_N$, an estimate of the best rank-one approximation of \mathcal{A} . (1) Compute initial values: Let $\mathbf{u}_0^{(n)}$ be the dominant left singular vector of $\mathbf{A}_{(n)}$ for $n = 2, \dots, N$. (2) For $k = 0, 1, 2, \dots$ (until converged), do: For $n = 1, \dots, N$, do: $\tilde{\mathbf{u}}_{k+1}^{(n)} = \mathcal{A} \times -_n \{\mathbf{u}_k\}$. $\lambda_{k+1}^{(n)} = \|\tilde{\mathbf{u}}_{k+1}^{(n)}\|$ $\mathbf{u}_{k+1}^{(n)} = \tilde{\mathbf{u}}_{k+1}^{(n)}/\lambda_{k+1}^{(n)}$ (3) Let $\lambda = \lambda_K$ and $\{\mathbf{u}\} = \{\mathbf{u}_K\}$, where K is the index of the final result of step(2). (4) Set $\mathcal{B} = \lambda \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \cdots \circ \mathbf{u}^{(n)}$.



that implements the algorithm. $\operatorname{Run}[ex20]$ to find a rank-one CP approximation to a tensor using the higher-order power method.

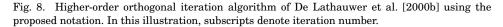
The second example is the higher-order orthogonal iteration, which is the multilinear generalization of the best rank-R approximation problem for matrices. Algorithm 4.2 in De Lathauwer et al. [2000b] is the higher-order orthogonal iteration and finds the best rank- (R_1, R_2, \ldots, R_N) approximation of a higher-order tensor. We have reproduced this algorithm in Figure 8 using our new notation. Our corresponding MATLAB implementation is listed in examples/hooi.m. Run ex21 to generate a rank-(1,1,1) Tucker approximation to the same tensor used in ex20, and note that the answers are identical. Run ex22 to generate a rank-(2,2,1) Tucker approximation and note that the quality of the approximation has improved. Run ex23 to generate a model with a core array of the same size as the original tensor and observe that the approximation is equal to the original tensor.

7. CONCLUSIONS

We have described four new MATLAB classes for manipulating dense and factored tensors. These classes extend MATLAB's built-in capabilities for multidimensional arrays in order to facilitate rapid algorithm development when dealing with tensor data types.

The tensor class simplifies the algorithmic details for implementing numerical methods for higher-order tensors by hiding the underlying matrix operations. It was previously the case that users had to know how to appropriately reshape the tensor into a matrix, execute the desired operation using matrix commands, and then appropriately reshape the result into a tensor. This can be nonintuitive and cumbersome, and we believe using the tensor class will be much simpler.

HIGHER-ORDER ORTHOGONAL ITERATION In: \mathcal{A} of size $I_1 \times I_2 \times \cdots \times I_N$ and desired rank of output. <u>Out</u>: \mathcal{B} of size $I_1 \times I_2 \times \cdots \times I_N$, an estimate of the best rank- (R_1, R_2, \ldots, R_N) approximation of \mathcal{A} . (1) Compute initial values: Let $\mathbf{U}_0^{(n)} \in \mathbb{R}^{I_n \times R_n}$ be an orthonormal basis for the dominant R_n -dimensional left singular subspace of $\mathbf{A}_{(n)}$ for n = 2, ..., n. (2) For k = 0, 1, 2, ... (until converged), do: For $n = 1, ..., N_{-}$, do: $\tilde{\mathcal{U}} = \mathcal{A} \times_{-n} \{ \mathbf{U}_k^T \}$ Let \mathbf{W} of size $I_n \times R_n$ solve: max $\left\| \tilde{\mathcal{U}} \times_n \mathbf{W}^T \right\|$ subject to $\mathbf{W}^T \mathbf{W} = \mathbf{I}$. $\mathbf{U}_{k+1}^{(n)''} = \mathbf{W}.$ (3) Let $\{\mathbf{U}\} = \{\mathbf{U}_K\}$, where K is the index of the final result of step 2. (4) Set $\lambda = \mathcal{A} \times \{\mathbf{U}^T\}$. (5) Set $\mathcal{B} = \lambda \times \{\mathbf{U}\}.$



The tensor_as_matrix class offers a way to convert a higher-order tensor into a matrix. Many existing algorithms in the literature that deal with tensors rely on matrix-matrix operations. The tensor_as_matrix functionality offers a means to implement these algorithms more easily, without the difficulty of reshaping and permuting tensor objects to the desired shapes.

The tucker_tensor and cp_tensor classes give users an easy way to store and manipulate factored tensors, as well as the ability to convert such tensors into nonfactored (or dense) format.

At this stage, our MATLAB implementations are not optimized for performance or memory usage; however, we have striven for consistency and ease-ofuse. In the future, we plan to further enhance these classes and add additional functionality.

Over the course of this code development effort, we have relied on published notation, especially from Kiers [2000] and De Lathauwer et al. [2000b]. To address ambiguities that we discovered in the class development process, we have proposed extensions to the existing mathematical notation, particularly in the area of tensor multiplication, that we believe more clearly denote mathematical concepts that were difficult to write succinctly with the existing notation.

We have demonstrated our new notation and MATLAB classes by revisiting the higher-order power method and the higher-order orthogonal iteration method from De Lathauwer et al. [2000b]. In our opinion, the resulting algorithm and code are more easily understood using our consolidated notation and MATLAB classes.

ACKNOWLEDGMENTS

For their many helpful suggestions, we gratefully acknowledge the participants at the Tensor Decompositions Workshop held at the American Institute

of Mathematics (AIM) in Palo Alto, California on July 19–23, 2004. We also thank AIM and the National Science Foundation for sponsoring the workshop. We also thank the referees for carefully reading the manuscript and code and for suggesting many useful improvements and corrections.

REFERENCES

- ANDERSSON, C. A. AND BRO, R. 2000. The N-way toolbox for MATLAB. Chemometrics Intel. Lab. Syst. 52, 1, 1-4. http://www.models.kvl.dk/source/nwaytoolbox/.
- CARROLL, J. D. AND CHANG, J. J. 1970. Analysis of individual differences in multidimensional scaling via an n-way generalization of 'Eckart-Young' decomposition. *Psychometrika 35*, 283–319.
- CHEN, B., PETROPOLU, A., AND DE LATHAUWER, L. 2002. Blind identification of convolutive mim systems with 3 sources and 2 sensors. *Appl. Signal Processing* 5, 487–496.
- COMON, P. 2001. Tensor decompositions. In *Mathematics in Signal Processing V*, J. G. McWhirter and I. K. Proudler, Eds. Oxford University Press, Oxford, UK, 1–24.
- DE LATHAUWER, L., DE MOOR, B., AND VANDEWALLE, J. 2000a. A multilinear singular value decomposition. SIAM J. Matrix Anal. Appl. 21, 4, 1253–1278.
- DE LATHAUWER, L., DE MOOR, B., AND VANDEWALLE, J. 2000b. On the best rank-1 and rank- (R_1, R_2, \ldots, R_N) approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.* 21, 4, 1324–1342.
- HARSHMAN, R. A. 1970. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. UCLA Working Papers in Phonetics 16, 1–84.
- HARSHMAN, R. A. 2001. An index formulism that generalizes the capabilities of matrix notation and algebra to *n*-way arrays. J. Chemometrics 15, 689–714.
- KIERS, H. A. L. 2000. Towards a standardized notation and terminology in multiway analysis. J. Chemometrics 14, 105–122.
- KROONENBERG, P. 2004. Applications of three-mode techniques: Overview, problems, and prospects. Presentation at the AIM Tensor Decompositions Workshop (Palo Alto, CA, July 19-23, 2004. \tthttp://csmr.ca.sandia.gov/~tgkolda/tdw2004/Kroonenberg%20-%20Talk.pdf.
- SMILDE, A., BRO, R., AND GELADI, P. 2004. Multi-Way Analysis: Applications in the Chemical Sciences. Wiley.
- THE MATHWORKS, INC. 2004a. Documentation: MATLAB: Programming: Classes and objects. http://www.mathworks.com/access/helpdesk_r13/help/techdoc/matlab_prog/ch14_oop.html.
- THE MATHWORKS, INC. 2004b. Documentation: MATLAB: Programming: Multidimensional arrays.http://www.mathworks.com/access/helpdesk_r13/help/techdoc/matlab_prog/ch12_ndf. html.
- TUCKER, L. R. 1966. Some mathematical notes on three-mode factor analysis. *Psychometrika 31*, 279–311.
- VASILESCU, M. A. O. AND TERZOPOULOS, D. 2002. Multilinear analysis of image ensembles: Tensorfaces. In Proceedings of the 7th European Conference on Computer Vision (ECCV'02). Copenhagen, Denmark, May. Lecture Notes in Computer Science vol. 2350. Springer Verlag, 447–460.

Received October 2004; revised July 2005; accepted February 2006