SANDIA REPORT

SAND2006-2081
Unclassified Unlimited Release
Printed April 2006

Multilinear operators for higher-order decompositions

Tamara G. Kolda

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94-AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone:	(865) 576-8401
Facsimile:	(865) 576-5728
E-Mail:	reports@adonis.osti.gov
Online ordering:	http://www.doe.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone:	(800) 553-6847
Facsimile:	(703) 605-6900
E-Mail:	orders@ntis.fedworld.gov
Online ordering:	http://www.ntis.gov/ordering.htm

SAND2006-2081
Unclassified Unlimited Release
Printed April 2006

Multilinear operators for higher-order decompositions

Tamara G. Kolda
Computational Science and Mathematics Science Research Department
Sandia National Laboratories
Livermore, CA 94551-9159
tgkolda@sandia.gov

Abstract

We propose two new multilinear operators for expressing the matrix compositions that are needed in the Tucker and PARAFAC (CANDECOMP) decompositions. The first operator, which we call the Tucker operator, is shorthand for performing an n-mode matrix multiplication for every mode of a given tensor and can be employed to consisely express the Tucker decomposition. The second operator, which we call the Kruskal operator, is shorthand for the sum of the outer-products of the columns of N matrices and allows a divorce from a matricized representation and a very consise expression of the PARAFAC decomposition. We explore the properties of the Tucker and Kruskal operators independently of the related decompositions. Additionally, we provide a review of the matrix and tensor operations that are frequently used in the context of tensor decompositions.

Contents

1 Introduction . . . 5
2 Notation . . . 5
3 Review of standard operations . . . 7
3.1 Matrix operations . . . 7
3.2 The outer product of vectors . . . 8
3.3 Tensor multiplication: the n-mode product . . . 8
3.4 Matricization of a tensor . . . 10
3.5 Norm and inner product of a tensor . . . 12
4 The Tucker operator . . . 13
4.1 Definition of the Tucker operator . . . 13
4.2 Tucker operator properties . . . 13
4.3 The Tucker decomposition . . . 15
4.4 Finding an optimal rank-$(J_1, J_2, \ldots, J_N)$ approximation . . . 16
4.5 Derivatives . . . 17
5 The Kruskal operator . . . 18
5.1 Definition of the Kruskal operator . . . 19
5.2 Kruskal operator properties . . . 19
5.3 The PARAFAC decomposition . . . 20
5.4 Computing the PARAFAC decomposition . . . 21
5.5 Derivatives of the Kruskal operator . . . 22
6 Conclusions . . . 23
References . . . 24

Appendix

A LaTeX formatting . . . 27

Figures

1 Three-way tensors. . . . 6
2 Fibers of a 3rd-order tensor. . . . 6
3 Slices of a 3rd-order tensor. . . . 7
4 Illustration of mode-1 matricization—the column fibers are aligned to form a matrix. 11
5 Illustration of the Tucker decomposition: $\mathcal{X} = [\![\mathcal{G}\,; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$ . . . 15
6 Illustration of the PARAFAC decomposition: $\mathcal{X} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$ . . . 21

Algorithms

1 Tucker: Higher Order Orthogonal Iteration . . . 18
2 PARAFAC: Alternating Least Squares (ALS) . . . 22

Multilinear operators for higher-order decompositions

1 Introduction

Higher-order tensor decompositions are in frequent use today in a variety of fields including psychometrics [43, 12, 19], chemometrics [6, 37], image analysis [45, 36, 46], graph analysis [27, 26], signal processing [13, 33], and much more [1, 38, 32]. The two most commonly used decompositions are Tucker [43] and PARAFAC (also known as CANDECOMP) [12, 19], which can be thought of as higher-order generalizations of the matrix singular value decomposition.

Unfortunately, the notation for these decompositions is not standardized because there are no operators to denote the multilinear compositions of matrices that are needed. Kiers [24], Harshman [20], and Bader and Kolda [8] have provided guidance on general notation for higher-order operations but did not focus on notation for higher-order decompositions. Typically, these decompositions are written in terms of elementary tensor operations (like n-mode multiplication) or by using a matricized representation. The difficulty is that this notation is non-intuitive and obscures the multilinear properties of the underlying operations.

To remedy this problem, we propose two new operators: a Tucker operator, which denotes a series of n-mode multiplication operations, and a Kruskal operator, which is a special case of the Tucker operator and useful for the PARAFAC decomposition. In this paper, we define these operators, examine their properties, and demonstrate how their use enables a better understanding of the Tucker and PARAFAC decompositions.

For example, the Tucker operator simplifies the notation for the Tucker decomposition. Let $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ be a third-order tensor. It will turn out that

$$\mathcal{X} = [\![\mathcal{G} \,; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \qquad \text{replaces} \qquad x_{ijk} = \sum_{r=1}^{R} \sum_{s=1}^{S} \sum_{t=1}^{T} g_{rst} \, a_{ir} b_{js} c_{kt}.$$

The third-order tensor $\mathcal{G} \in \mathbb{R}^{R \times S \times T}$ is called the core array, and the matrices $\mathbf{A} \in \mathbb{R}^{I \times R}$, $\mathbf{B} \in \mathbb{R}^{J \times S}$, and $\mathbf{C} \in \mathbb{R}^{K \times T}$ are called the factors. Likewise, the Kruskal operator simplifies expression of the PARAFAC decomposition. Here it turns out that

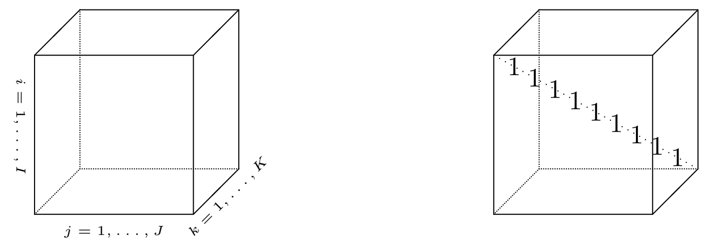
$$\mathcal{X} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \qquad \text{replaces} \qquad x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr}.$$

In the PARAFAC case, there is no core array, only the factor matrices $\mathbf{A} \in \mathbb{R}^{I \times R}$, $\mathbf{B} \in \mathbb{R}^{J \times R}$, and $\mathbf{C} \in \mathbb{R}^{K \times R}$, which are now constrained to have equal numbers of columns. Kruskal proposed identical PARAFAC notation with the exception of the type of bracket; he used $[\mathbf{A}, \mathbf{B}, \mathbf{C}]$ [29].

The paper is organized as follows. Notation for n-way arrays can be complex, so we explain ours in §2. We review standard matrix and tensor operations and their properties in §3. The Tucker operator is covered in §4 and the Kruskal operator in §5. Lesser-known LaTeX formatting commands that are necessary to reproduce the symbols in this paper are provided in the appendix.

2 Notation

Multiway arrays (a.k.a. tensors) are denoted by boldface Euler script letters, e.g., $\mathcal{X}$. The *order* of a tensor is the number of dimensions, also known as ways or modes. Figure 1(a) illustrates a three-way tensor. We use $\mathcal{I}$ to denote the *identity tensor* with ones on the superdiagonal and zeros elsewhere; see Figure 1(b).

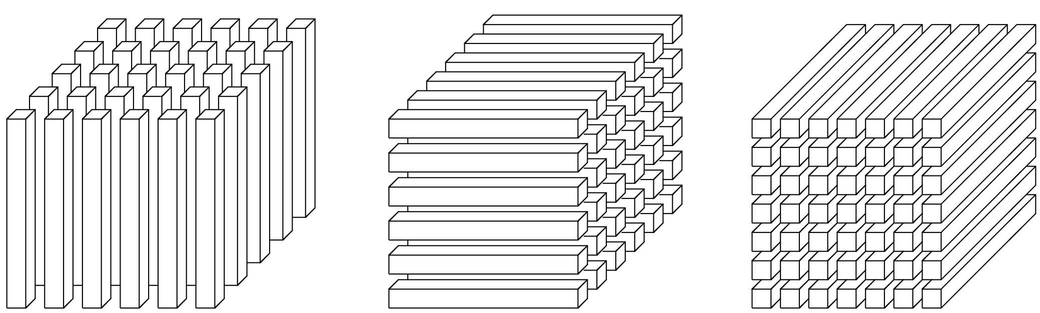


(a) $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ (b) $\mathcal{I} \in \mathbb{R}^{I \times I \times I}$

Figure 1. Three-way tensors.

Matrices are denoted by boldface capital letters, e.g., $\mathbf{A}$. We use $\mathbf{I}$ to denote the identity matrix. Vectors are denoted by boldface lowercase letters, e.g., $\mathbf{a}$. Scalars are denoted by lowercase letters, e.g., a. We have attempted to keep this notation consistent. Thus, the ith entry of a vector $\mathbf{a}$ is denoted by a_i, the jth column of $\mathbf{A}$ is denoted by $\mathbf{a}_{:j}$, the ith row by $\mathbf{a}_{i:}$, element (i, j) by a_{ij}, and element (i, j, k) element of a 3-way tensor $\mathcal{X}$ is denoted by x_{ijk}.

The higher order analogue of matrix rows and columns are called *fibers*. A matrix column is a mode-1 fiber and a matrix row is a mode-2 fiber. For a third order tensor, we have column, row, and tube fibers, which are denoted by $\mathbf{x}_{:jk}$, $\mathbf{x}_{i:k}$, and $\mathbf{x}_{ij:}$, respectively; see Figure 2. For orders higher than three, the fibers no longer have special names. We always assume that fibers are column vectors.

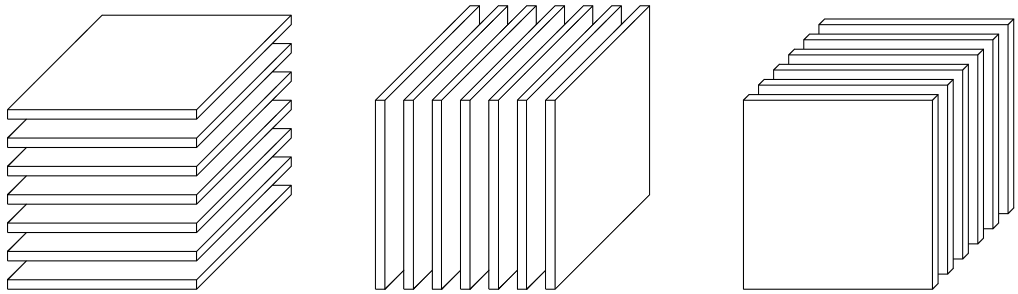


(a) Columns: $\mathbf{x}_{:jk}$ (b) Rows: $\mathbf{x}_{i:k}$ (c) Tubes: $\mathbf{x}_{ij:}$

Figure 2. Fibers of a 3rd-order tensor.

For third-order tensors, it is often useful to consider the two-dimensional slices. Figure 3 shows the horizontal, lateral, and frontal slides of a third-order tensor $\mathcal{X}$, denoted by $\mathbf{X}_{i::}$, $\mathbf{X}_{:j:}$, and $\mathbf{X}_{::k}$, respectively.

Finally, indices typically range from 1 to their capital version, e.g., $i = 1, \ldots, I$. Multiple indices have subscripts, e.g., $i_n = 1, \ldots, I_n$. Sets are denoted in calligraphic font, e.g., $\mathcal{R} = \{r_1, r_2, \ldots, r_P\}$. We denote a set of indexed indices by $I_{\mathcal{R}} = \{I_{r_1}, I_{r_2}, \ldots, I_{r_P}\}$.



(a) Horizontal: $\mathbf{X}_{i::}$ (b) Lateral: $\mathbf{X}_{:j:}$ (c) Frontal: $\mathbf{X}_{::k}$

Figure 3. Slices of a 3rd-order tensor.

3 Review of standard operations

We present a comprehensive survey of standard operations and concepts that are used in multiway analysis.

3.1 Matrix operations

The Kronecker product (also sometimes known as the tensor product), Khatri-Rao product, and Hadamard product are matrix operations that we use in this paper.

The *Kronecker product* of matrices $\mathbf{A} \in \mathbb{R}^{I\times J}$ and $\mathbf{B} \in \mathbb{R}^{K\times L}$ is denoted by $\mathbf{A} \otimes \mathbf{B}$ and the $(IK) \times (JL)$ result is defined by

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1J}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2J}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}\mathbf{B} & a_{I2}\mathbf{B} & \cdots & a_{IJ}\mathbf{B} \end{bmatrix}.$$

Certain of its properties (listed in Proposition 3.1) will prove useful in our discussions. Van Loan [44] provides a more general overview of the Kronecker product and its uses.

Proposition 3.1 (Kronecker product [44]) *Let* $\mathbf{A} \in \mathbb{R}^{I\times J}, \mathbf{B} \in \mathbb{R}^{K\times L}$. *Then*

(a) $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$, *and*
(b) $(\mathbf{A} \otimes \mathbf{B})^\dagger = \mathbf{A}^\dagger \otimes \mathbf{B}^\dagger$.

The Khatri-Rao product [31, 35, 10, 37] is the columnwise Kronecker product. The *Khatri-Rao product* of matrices $\mathbf{A} \in \mathbb{R}^{I\times K}$ and $\mathbf{B} \in \mathbb{R}^{J\times K}$ is denoted by $\mathbf{A} \odot \mathbf{B}$ and its $(IJ) \times K$ result is defined by

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} \mathbf{a}_{:1} \otimes \mathbf{b}_{:1} & \mathbf{a}_{:2} \otimes \mathbf{b}_{:2} & \cdots & \mathbf{a}_{:K} \otimes \mathbf{b}_{:K} \end{bmatrix}.$$

We will see later that the Khatri-Rao product is very important for expressing the PARAFAC decomposition. Observe that the matrices in a Khatri-Rao product all have the same number of columns. Furthermore, if $\mathbf{a}$ and $\mathbf{b}$ are vectors, then the Khatri-Rao and Kronecker products are identical, i.e., $\mathbf{a} \otimes \mathbf{b} = \mathbf{a} \odot \mathbf{b}$. The Khatri-Rao product has properties that involve the Hadamard product, which is the elementwise matrix product; i.e., the *Hadamard product* of matrices $\mathbf{A}$ and $\mathbf{B}$, both of size $I \times J$, is denoted by $\mathbf{A} * \mathbf{B}$ and its $I \times J$ result is defined by

$$\mathbf{A} * \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1J}b_{1J} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2J}b_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}b_{I1} & a_{I2}b_{I2} & \cdots & a_{IJ}b_{IJ} \end{bmatrix}.$$

In particular, the pseudo-inverse of a Khatri-Rao product has a special form that involves the Hadamard product.

Proposition 3.2 (Khatri-Rao product [37]) *Let* $\mathbf{A} \in \mathbb{R}^{I \times L}$, $\mathbf{B} \in \mathbb{R}^{J \times L}$, $\mathbf{C} \in \mathbb{R}^{K \times L}$. *Then*

(a) $\mathbf{A} \odot \mathbf{B} \odot \mathbf{C} = (\mathbf{A} \odot \mathbf{B}) \odot \mathbf{C} = \mathbf{A} \odot (\mathbf{B} \odot \mathbf{C})$,
(b) $(\mathbf{A} \odot \mathbf{B})^{\mathsf{T}}(\mathbf{A} \odot \mathbf{B}) = \mathbf{A}^{\mathsf{T}}\mathbf{A} * \mathbf{B}^{\mathsf{T}}\mathbf{B}$, *and*
(c) $(\mathbf{A} \odot \mathbf{B})^{\dagger} = \left((\mathbf{A}^{\mathsf{T}}\mathbf{A}) * (\mathbf{B}^{\mathsf{T}}\mathbf{B})\right)^{\dagger} (\mathbf{A} \odot \mathbf{B})^{\mathsf{T}}$.

3.2 The outer product of vectors

The outer product of two vectors yields a matrix and is typically written as $\mathbf{X} = \mathbf{a}\mathbf{b}^{\mathsf{T}}$. To extend the outer product to higher dimensions, we cannot reply solely on the transpose operator; instead, we use the symbol $\circ$ to denote the outer product, so we write $\mathbf{X} = \mathbf{a} \circ \mathbf{b}$ for the outer product of two vectors. Let $\mathcal{N} = \{1, \ldots, N\}$ and $\mathbf{a}^{(n)} \in \mathbb{R}^{I_n}$ for all $n \in \mathcal{N}$. Then the outer product of these N vectors is an Nth-order tensor and defined elementwise as

$$\left(\mathbf{a}^{(1)} \circ \mathbf{a}^{(2)} \circ \cdots \circ \mathbf{a}^{(N)}\right)_{i_1 i_2 \cdots i_N} = a^{(1)}{i_1} a^{(2)}{i_2} \cdots a^{(N)}{i_N} \text{ for } 1 \leq i_n \leq I_n, n \in \mathcal{N}.$$

Sometimes the notation $\otimes$ is used (see, e.g., [25]), but we reserve that notation in this paper for the matrix Kronecker product.

3.3 Tensor multiplication: the n-mode product

The n-mode product [14] defines multiplication of a tensor by a matrix in mode n. Though other types of tensor multiplication exist, see, e.g., [8], we only need to consider the n-mode product in this paper.

The *n-mode (matrix) product* of a tensor $\mathcal{Y} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$ with a matrix $\mathbf{A} \in \mathbb{R}^{I \times J_n}$ is denoted by $\mathcal{Y} \times_n \mathbf{A}$. The result is of size $J_1 \times \cdots \times J_{n-1} \times I \times J_{n+1} \times \cdots \times J_N$ and is defined elementwise as

$$(\mathcal{Y} \times_n \mathbf{A})_{j_1 \cdots j_{n-1} i \, j_{n+1} \cdots j_N} = \sum_{j_n=1}^{J_n} y_{j_1 j_2 \cdots j_N} \, a_{i j_n}.$$

There are many ways of considering n-mode multiplication. For example, let $\mathcal{Y} \in \mathbb{R}^{I \times J \times K}$, $\mathbf{B} \in \mathbb{R}^{L \times J}$, and $\mathcal{X} = \mathcal{Y} \times_2 \mathbf{B}$. One interpretation is that each mode-2 fiber of $\mathcal{X}$ is the result of multiplying the corresponding mode-2 fiber of $\mathcal{Y}$ by $\mathbf{B}$:

$$\mathbf{x}_{i:k} = \mathbf{B}\mathbf{y}_{i:k} \text{ for each } i = 1, \ldots, I, k = 1, \ldots, K.$$

Example 3.3 (n-mode matrix product) *Let* $\mathcal{Y}$ *be the following* $3 \times 4 \times 2$ *tensor:*

$$\mathbf{Y}_{::1} = \begin{bmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{bmatrix}, \quad \mathbf{Y}_{::2} = \begin{bmatrix} 13 & 16 & 19 & 22 \\ 14 & 17 & 20 & 23 \\ 15 & 18 & 21 & 24 \end{bmatrix}. \tag{1}$$

Let $\mathbf{A}$ *be the following* 2×3 *matrix:*

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}. \tag{2}$$

Note that the number of columns in $\mathbf{A}$ *is equal to the size of mode 1 of* $\mathcal{Y}$*. Thus we can compute* $\mathcal{Y} \times_1 \mathbf{A}$*, which is of size* $2 \times 4 \times 2$ *and*

$$(\mathcal{Y} \times_1 \mathbf{A})_{::1} = \begin{bmatrix} 22 & 49 & 76 & 103 \\ 28 & 64 & 100 & 136 \end{bmatrix}, \quad (\mathcal{Y} \times_1 \mathbf{A})_{::2} = \begin{bmatrix} 130 & 157 & 184 & 211 \\ 172 & 208 & 244 & 280 \end{bmatrix}.$$

Proposition 3.4 (n-mode matrix product [14]) *Let* $\mathcal{Y} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$ *be an* N*-way tensor.*

(a) *Given matrices* $\mathbf{A} \in \mathbb{R}^{I_m \times J_m}$, $\mathbf{B} \in \mathbb{R}^{I_n \times J_n}$,

$$\mathcal{Y} \times_m \mathbf{A} \times_n \mathbf{B} = (\mathcal{Y} \times_m \mathbf{A}) \times_n \mathbf{B} = (\mathcal{Y} \times_n \mathbf{B}) \times_m \mathbf{A} \quad (m \neq n).$$

(b) *Given matrices* $\mathbf{A} \in \mathbb{R}^{I \times J_n}$, $\mathbf{B} \in \mathbb{R}^{K \times I}$,

$$\mathcal{Y} \times_n \mathbf{A} \times_n \mathbf{B} = \mathcal{Y} \times_n (\mathbf{BA}).$$

(c) *Moreover, if* $\mathbf{A} \in \mathbb{R}^{I \times J_n}$ *with full column rank, then*

$$\mathcal{X} = \mathcal{Y} \times_n \mathbf{A} \quad \Rightarrow \quad \mathcal{Y} = \mathcal{X} \times_n \mathbf{A}^{\dagger}.$$

(d) *Consequently, if* $\mathbf{A} \in \mathbb{R}^{I \times J_n}$ *is orthonormal, then*

$$\mathcal{X} = \mathcal{Y} \times_n \mathbf{A} \quad \Rightarrow \quad \mathcal{Y} = \mathcal{X} \times_n \mathbf{A}^{\mathsf{T}}.$$

Example 3.5 *Here we illustrate Proposition 3.4(d). Let* $\mathcal{Y}$ *be given by (1). Define the orthonormal matrix*

$$\mathbf{C} = \begin{bmatrix} 0.58 & 0.00 \\ 0.58 & -0.71 \\ 0.58 & 0.71 \end{bmatrix}.$$

Then $\mathcal{X} = \mathcal{Y} \times_3 \mathbf{C}$ *is*

$$\mathbf{X}_{::1} = \begin{bmatrix} 0.58 & 2.31 & 4.04 & 5.77 \\ 1.15 & 2.89 & 4.62 & 6.35 \\ 1.73 & 3.46 & 5.20 & 6.93 \end{bmatrix},$$

$$\mathbf{X}_{::2} = \begin{bmatrix} -8.62 & -9.00 & -9.39 & -9.78 \\ -8.74 & -9.13 & -9.52 & -9.91 \\ -8.87 & -9.26 & -9.65 & -10.04 \end{bmatrix},$$

$$\mathbf{X}_{::3} = \begin{bmatrix} 9.77 & 13.62 & 17.48 & 21.33 \\ 11.05 & 14.91 & 18.76 & 22.61 \\ 12.34 & 16.19 & 20.05 & 23.90 \end{bmatrix}.$$

We have then that $\mathcal{Z} = \mathcal{X} \times_3 \mathbf{C}^{\mathsf{T}}$ *is*

$$\mathbf{Z}_{::1} = \begin{bmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{bmatrix}, \quad \mathbf{Z}_{::2} = \begin{bmatrix} 13 & 16 & 19 & 22 \\ 14 & 17 & 20 & 23 \\ 15 & 18 & 21 & 24 \end{bmatrix}.$$

In other words, we have recovered the original $\mathcal{Y}$*.*

3.4 Matricization of a tensor

Especially in computations, it is important to be able to transform the indicies of a tensor so that it can be represented as a matrix, and vice versa [8]. In order to fully capture all the salient information, we need to explicitly track three pieces of information in addition to the data itself: the size of the tensor, the modes that are mapped to the rows of the matrix, and the modes that are mapped to the columns of the matrix.

The *matricization* of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is defined as follows. Let the ordered sets $\mathcal{R} = \{r_1, \ldots, r_L\}$ and $\mathcal{C} = \{c_1, \ldots, c_M\}$ be a partitioning of the modes $\mathcal{N} = \{1, \ldots, N\}$. Recall that $I_\mathcal{N}$ denotes the size of the tensor: $\{I_1, \ldots, I_N\}$. The matricized tensor can then be specified by

$$\mathbf{X}_{(\mathcal{R} \times \mathcal{C} \,:\, I_\mathcal{N})} \in \mathbb{R}^{J \times K} \quad \text{with} \quad J = \prod_{n \in \mathcal{R}} I_n \quad \text{and} \quad K = \prod_{n \in \mathcal{C}} I_n.$$

The indices in $\mathcal{R}$ are mapped to the rows and the indices in $\mathcal{C}$ mapped to the columns. Specifically,

$$\left(\mathbf{X}_{(\mathcal{R} \times \mathcal{C} \,:\, I_\mathcal{N})}\right)_{jk} = x_{i_1 i_2 \cdots i_N}$$

with

$$j = 1 + \sum_{\ell=1}^{L} \left[(i_{r_\ell} - 1) \prod_{\ell'=1}^{\ell-1} I_{r_{\ell'}} \right] \quad \text{and} \quad k = 1 + \sum_{m=1}^{M} \left[(i_{r_m} - 1) \prod_{m'=1}^{m-1} I_{r_{m'}} \right].$$

It may be easier to understand matriticization in MATLAB notation. Suppose `X` is a multidimensional array, and let the sets `R` and `C` be defined. Then the following code converts to a matrix and back again to a tensor.

```
X = rand(5,6,4,2); R = [2 3]; C = [4 1];
I = size(X); J = prod(I(R)); K = prod(I(C));
Y = reshape(permute(X,[R C]),J,K);  % convert X to matrix Y
Z = ipermute(reshape(Y,[I(R) I(C)]),[R C]); % convert back to tensor
```

Note that we must explicitly recall the sizes of the original tensor dimensions in order to convert the matrix *back* to a tensor. This is generally not called out explicitly in notation. For example, if $\mathcal{R} = \{1, 2\}$ and $\mathcal{C} = \{3, \ldots, N\}$, then $\mathbf{X}_{(\mathcal{R} \times \mathcal{C} \,:\, I_\mathcal{N})}$ is more typically written as

$$\mathbf{X}^{I_1 I_2 \times I_3 I_4 \cdots I_N} \quad \text{or} \quad \mathbf{X}_{(I_1 I_2 \times I_3 I_4 \cdots I_N)}.$$

In other words, the size of the original tensor is generally treated implicitly. However, we will see that explicitly listing the sizes (i.e., the argument following the colon) proves useful in certain cases such as Proposition 3.7(b) and Proposition 3.7(d).

An important special case is whenever $\mathcal{R}$ is a singleton. This means that the fibers of mode n are aligned as the columns of the resulting matrix. The *n-mode matricization* of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is a special case of matricization given by

$$\mathbf{X}_{(n)} \equiv \mathbf{X}_{(\mathcal{R} \times \mathcal{C} \,:\, I_\mathcal{N})} \text{ with } \mathcal{R} = \{n\} \text{ and } \mathcal{C} = \{1, \ldots, n-1, n+1, \ldots, N\}.$$

Here we adhere to the standard notation. See also the illustration in Figure 4.

In general, the order of the modes within $\mathcal{C}$ is irrelevant so long as all operations with the transformed modes are consistent. Different authors use different orderings for the columns of the resulting matrix; see, e.g., [14] versus [24].

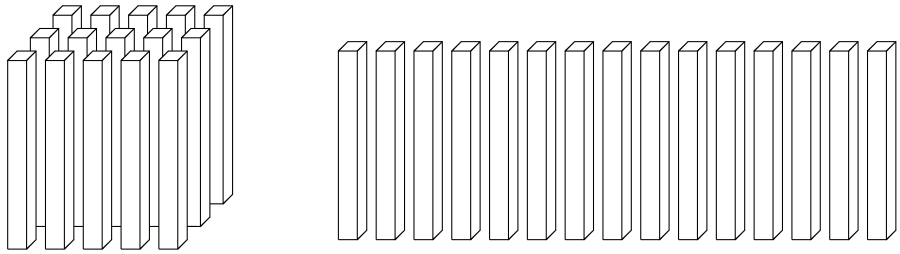


Figure 4. Illustration of mode-1 matricization—the column fibers are aligned to form a matrix.

In addition to converting a tensor to a matrix, it can also be converted to a vector, which is just a special case of matricization where all the modes become row modes; i.e., the *vectorized* tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is given by

$$\text{vec}(\mathcal{X}) \equiv \mathbf{X}_{(\mathcal{N} \times \emptyset \,:\, I_\mathcal{N})}.$$

Example 3.6 (Matricization) *Let $\mathcal{Y}$ be given by (1). Then*

$$\mathbf{Y}_{(\{3,1\} \times \{2\} \,:\, \{3,4,2\})} = \begin{bmatrix} 1 & 4 & 7 & 10 \\ 13 & 16 & 19 & 22 \\ 2 & 5 & 8 & 11 \\ 14 & 17 & 20 & 23 \\ 3 & 6 & 9 & 12 \\ 15 & 18 & 21 & 24 \end{bmatrix},$$

$$\mathbf{Y}_{(1)} = \mathbf{Y}_{(\{1\} \times \{2,3\} \,:\, \{3,4,2\})} = \begin{bmatrix} 1 & 4 & 7 & 10 & 13 & 16 & 19 & 22 \\ 2 & 5 & 8 & 11 & 14 & 17 & 20 & 23 \\ 3 & 6 & 9 & 12 & 15 & 18 & 21 & 24 \end{bmatrix},$$

and

$$\text{vec}(\mathcal{Y}) = \mathbf{Y}_{(\mathcal{N} \times \emptyset \,:\, I_\mathcal{N})} = \begin{bmatrix} 1 & 2 & \cdots & 24 \end{bmatrix}^\mathsf{T}.$$

Converting a tensor to a matrix in useful both computationally and theoretically because there are useful connections between the n-mode matrix product, matricization, and Kronecker products.

Proposition 3.7 *Let $\mathcal{Y} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$ and $\mathcal{N} = 1, \ldots, N$.*

(a) *If $\mathbf{A} \in \mathbb{R}^{I \times J_n}$. Then*

$$\mathcal{X} = \mathcal{Y} \times_n \mathbf{A} \quad \Leftrightarrow \quad \mathbf{X}_{(n)} = \mathbf{A}\mathbf{Y}_{(n)}.$$

(b) *Let $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$ for all $n \in \mathcal{N}$. If $\mathcal{R} = \{r_1, \ldots, r_L\}$ and $\mathcal{C} = \{c_1, \ldots, c_M\}$ partition $\mathcal{N}$, then*

$$\mathcal{X} = \mathcal{Y} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)} \quad \Leftrightarrow$$
$$\mathbf{X}_{(\mathcal{R} \times \mathcal{C} \,:\, J_\mathcal{N})} = \left(\mathbf{A}^{(r_L)} \otimes \cdots \otimes \mathbf{A}^{(r_1)} \right) \mathbf{Y}_{(\mathcal{R} \times \mathcal{C} \,:\, I_\mathcal{N})} \left(\mathbf{A}^{(c_M)} \otimes \cdots \otimes \mathbf{A}^{(c_1)} \right)^\mathsf{T}$$

(c) *Consequently, if $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$ for all $n \in \mathcal{N}$, for any specific $n \in \mathcal{N}$ we have*

$$\mathcal{X} = \mathcal{Y} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)} \quad \Leftrightarrow$$
$$\mathbf{X}_{(n)} = \mathbf{A}^{(n)} \mathbf{Y}_{(n)} \left(\mathbf{A}^{(N)} \otimes \cdots \otimes \mathbf{A}^{(n+1)} \otimes \mathbf{A}^{(n-1)} \otimes \cdots \otimes \mathbf{A}^{(1)} \right)^\mathsf{T}.$$

(d) *Moreover, if* $\mathcal{C} = \{c_1, \ldots, c_M\} \subseteq \mathcal{N}$ *and* $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$ *for* $n \in \mathcal{C}$*, defining* $\mathcal{R} = \mathcal{N} \setminus \mathcal{C}$ *yields*

$$\mathbf{\mathcal{X}} = \mathbf{\mathcal{Y}} \times_{c_1} \mathbf{A}^{(c_1)} \times_{c_2} \mathbf{A}^{(c_2)} \cdots \times_{c_M} \mathbf{A}^{(c_M)} \quad \Leftrightarrow$$

$$\mathbf{X}_{(\mathcal{R} \times \mathcal{C} \,:\, K_{\mathcal{N}})} = \mathbf{Y}_{(\mathcal{R} \times \mathcal{C} \,:\, I_{\mathcal{N}})} \left(\mathbf{A}^{(c_M)} \otimes \cdots \otimes \mathbf{A}^{(c_1)} \right)^{\mathsf{T}} \text{ with } K_n \equiv \begin{cases} I_n & \text{if } n \in \mathcal{C} \\ J_n & \text{if } n \in \mathcal{R} \end{cases}.$$

3.5 Norm and inner product of a tensor

The norm and inner product are most easily thought of in terms of the vectorized tensor. The *inner product* of two tensors $\mathbf{\mathcal{X}}, \mathbf{\mathcal{Y}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is given by

$$\langle \mathbf{\mathcal{X}}, \mathbf{\mathcal{Y}} \rangle = \operatorname{vec}(\mathbf{\mathcal{X}})^{\mathsf{T}} \operatorname{vec}(\mathbf{\mathcal{Y}}) = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \cdots i_N} y_{i_1 i_2 \cdots i_N}.$$

The *norm* of a tensor $\mathbf{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is given by

$$\| \mathbf{\mathcal{X}} \|^2 = \langle \mathbf{\mathcal{X}}, \mathbf{\mathcal{X}} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x^2_{i_1 i_2 \cdots i_N}.$$

The norm of a tensor can be transformed to a matrix or vector norm by using the matricized or vectorized version of the tensor (Proposition 3.8). Moreover, the norm of the difference of two tensors can be rewritten to instead involve the inner product of the two tensors (Proposition 3.9). The inner product of two rank-1 tensors can be simplified to be the product of the individual dot products of the components (Proposition 3.10). Finally, Mode-n multiplication commutes with respect to the inner product (Proposition 3.11).

Proposition 3.8 *Let* $\mathbf{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ *and* $\mathcal{N} = \{1, \ldots, N\}$*.*

(a) *Let sets* $\mathcal{R}$ *and* $\mathcal{C}$ *be a partitioning of* $\mathcal{N}$*. Then* $\| \mathbf{\mathcal{X}} \| = \left\| \mathbf{X}_{(\mathcal{R} \times \mathcal{C} \,:\, I_{\mathcal{N}})} \right\|_F$.
(b) *Let* $n \in \mathcal{N}$*. Then* $\| \mathbf{\mathcal{X}} \| = \left\| \mathbf{X}_{(n)} \right\|_F$.
(c) $\| \mathbf{\mathcal{X}} \| = \| \operatorname{vec}(\mathbf{\mathcal{X}}) \|_2$.

Proposition 3.9 *Let* $\mathbf{\mathcal{X}}, \mathbf{\mathcal{Y}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$*. Then*

$$\| \mathbf{\mathcal{X}} - \mathbf{\mathcal{Y}} \|^2 = \| \mathbf{\mathcal{X}} \|^2 - 2 \langle \mathbf{\mathcal{X}}, \mathbf{\mathcal{Y}} \rangle - \| \mathbf{\mathcal{Y}} \|^2.$$

Proposition 3.10 *Let* $\mathbf{\mathcal{X}}, \mathbf{\mathcal{Y}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ *with* $\mathbf{\mathcal{X}} = \mathbf{a}^{(1)} \circ \mathbf{a}^{(2)} \circ \cdots \circ \mathbf{a}^{(N)}$ *and* $\mathbf{\mathcal{Y}} = \mathbf{b}^{(1)} \circ \mathbf{b}^{(2)} \circ \cdots \circ \mathbf{b}^{(N)}$*. Then*

$$\langle \mathbf{\mathcal{X}}, \mathbf{\mathcal{Y}} \rangle = \prod_{n=1}^{N} \langle \mathbf{a}^{(n)}, \mathbf{b}^{(n)} \rangle.$$

Proposition 3.11 *Let* $\mathbf{\mathcal{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N}$*,* $\mathbf{\mathcal{Y}} \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times K \times I_{n+1} \times \cdots \times I_N}$*, and* $\mathbf{A} \in \mathbb{R}^{J \times K}$*. Then*

$$\langle \mathbf{\mathcal{X}}, \mathbf{\mathcal{Y}} \times_n \mathbf{A} \rangle = \langle \mathbf{\mathcal{X}} \times_n \mathbf{A}^{\mathsf{T}}, \mathbf{\mathcal{Y}} \rangle.$$

An interesting corollary of the previous result is that mode-n multiplication of a tensor with an orthogonal matrix does not change its norm.

Proposition 3.12 *Let* $\mathbf{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ *and let* $\mathbf{Q}$ *be a* $J \times I_n$ *orthonormal matrix. Then*

$$\| \mathbf{\mathcal{X}} \| = \| \mathbf{\mathcal{X}} \times_n \mathbf{Q} \|.$$

4 The Tucker operator

Now that we have reviewed essential matrix and tensor operations, we can proceed to defining our multilinear operators. In this section, we consider the Tucker operator and its application to the Tucker decomposition.

4.1 Definition of the Tucker operator

The *Tucker operator* is an efficient representation for multi-mode multiplication, which we formally define as follows.

Definition 4.1 *Let* $\mathcal{Y} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$ *and* $\mathcal{N} = \{1, \ldots, N\}$*. Suppose we have matrices* $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$ *for* $n \in \mathcal{N}$*. Then the* Tucker operator *is defined as:*

$$[\![\mathcal{Y} \,; \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)}]\!] \equiv \mathcal{Y} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)}. \tag{3}$$

The result is of size $I_1 \times I_2 \times \cdots \times I_N$*.*

The Tucker operator can be defined on a subset of modes $\{k_1, \ldots, k_P\} \subset \mathcal{N}$ *via a subscript on the operator as follows:*

$$[\![\mathcal{Y} \,; \mathbf{A}^{(k_1)}, \mathbf{A}^{(k_2)}, \ldots, \mathbf{A}^{(k_P)}]\!]_{\{k_1,\ldots,k_P\}} \equiv \mathcal{Y} \times_{k_1} \mathbf{A}^{(k_1)} \times_{k_2} \mathbf{A}^{(k_2)} \cdots \times_{k_P} \mathbf{A}^{(k_P)}.$$

Grigorascu and Regalia [18] have proposed notation for the same concept as the Tucker operator,

$$\mathbf{A}^{(1)} \overset{\mathcal{Y}}{\star} \mathbf{A}^{(2)} \overset{\mathcal{Y}}{\star} \cdots \overset{\mathcal{Y}}{\star} \mathbf{A}^{(N)},$$

which they refer to as the *weighted Tucker product* (the unweighted version has $\mathcal{Y} = \mathcal{I}$, the identity tensor). Note that the case of using only a subset of modes is equivalent to replacing the missing modes with $J_n \times J_n$ identity matrices.

4.2 Tucker operator properties

The properties of the Tucker operator follow directly from the properties of n-mode multiplication (see §3.3).

Proposition 4.2 *Let* $\mathcal{Y} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$ *and let* $\mathcal{N} = \{1, \ldots, N\}$*.*

(a) *Given matrices* $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$, $\mathbf{B}^{(n)} \in \mathbb{R}^{K_n \times I_n}$ *for all* $n \in \mathcal{N}$*, we have*

$$[\![\, [\![\mathcal{Y} \,; \mathbf{A}^{(1)}, \cdots, \mathbf{A}^{(N)}]\!] \,; \mathbf{B}^{(1)}, \cdots, \mathbf{B}^{(N)}]\!] = [\![\mathcal{Y} \,; \mathbf{B}^{(1)}\mathbf{A}^{(1)}, \cdots, \mathbf{B}^{(N)}\mathbf{A}^{(N)}]\!]$$

(b) *Given matrices* $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$ *for all* $n \in \mathcal{N}$ *with full column rank, we have*

$$\mathcal{X} = [\![\mathcal{Y} \,; \mathbf{A}^{(1)}, \cdots, \mathbf{A}^{(N)}]\!] \Rightarrow \mathcal{Y} = [\![\mathcal{X} \,; \mathbf{A}^{(1)^\dagger}, \cdots, \mathbf{A}^{(N)^\dagger}]\!]$$

(c) *Given orthonormal matrices* $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$ *for all* $n \in \mathcal{N}$*, we have*

$$\mathcal{X} = [\![\mathcal{Y} \,; \mathbf{A}^{(1)}, \cdots, \mathbf{A}^{(N)}]\!] \Rightarrow \mathcal{Y} = [\![\mathcal{X} \,; \mathbf{A}^{(1)\mathsf{T}}, \cdots, \mathbf{A}^{(N)\mathsf{T}}]\!]$$

Proof. Part (a) follows from the definition of the Tucker operator and the properties of mode-n multiplication (Proposition 3.4(b)). Likewise, Parts (b) and (c) follow from other properties of mode-n multiplication (Proposition 3.4(c) and Proposition 3.4(d), respectively). □

The Tucker operator also has various expressions in terms of matricized tensors and the Kronecker product.

Proposition 4.3 *Let $\mathcal{Y} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$ and $\mathcal{N} = 1, \ldots, N$.*

(a) *Let $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$ for all $n \in \mathcal{N}$. If $\mathcal{R} = \{r_1, \ldots, r_L\}$ and $\mathcal{C} = \{c_1, \ldots, c_M\}$ partition $\mathcal{N}$, then*

$$\mathcal{X} = [\![\mathcal{Y} \,; \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \cdots, \mathbf{A}^{(N)}]\!] \quad \Leftrightarrow$$
$$\mathbf{X}_{(\mathcal{R} \times \mathcal{C} \,:\, J_{\mathcal{N}})} = \left(\mathbf{A}^{(r_L)} \otimes \cdots \otimes \mathbf{A}^{(r_1)} \right) \mathbf{Y}_{(\mathcal{R} \times \mathcal{C} \,:\, I_{\mathcal{N}})} \left(\mathbf{A}^{(c_M)} \otimes \cdots \otimes \mathbf{A}^{(c_1)} \right)^{\mathsf{T}}$$

(b) *Consequently, if $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$ for all $n \in \mathcal{N}$, for any specific $n \in \mathcal{N}$ we have*

$$\mathcal{X} = [\![\mathcal{Y} \,; \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \cdots, \mathbf{A}^{(N)}]\!] \quad \Leftrightarrow$$
$$\mathbf{X}_{(n)} = \mathbf{A}^{(n)} \mathbf{Y}_{(n)} \left(\mathbf{A}^{(N)} \otimes \cdots \otimes \mathbf{A}^{(n+1)} \otimes \mathbf{A}^{(n-1)} \otimes \cdots \otimes \mathbf{A}^{(1)} \right)^{\mathsf{T}}.$$

(c) *Moreover, if $\mathcal{C} = \{c_1, \ldots, c_M\} \subseteq \mathcal{N}$ and $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$ for $n \in \mathcal{C}$, defining $\mathcal{R} = \mathcal{N} \setminus \mathcal{C}$ yields*

$$\mathcal{X} = [\![\mathcal{Y} \,; \mathbf{A}^{(c_1)}, \mathbf{A}^{(c_2)}, \cdots, \mathbf{A}^{(c_M)}]\!]_{\mathcal{C}} \quad \Leftrightarrow$$
$$\mathbf{X}_{(\mathcal{R} \times \mathcal{C} \,:\, K_{\mathcal{N}})} = \mathbf{Y}_{(\mathcal{R} \times \mathcal{C} \,:\, I_{\mathcal{N}})} \left(\mathbf{A}^{(c_N)} \otimes \cdots \otimes \mathbf{A}^{(c_1)} \right)^{\mathsf{T}} \text{ with } K_n \equiv \begin{cases} J_n & \text{if } n \in \mathcal{C} \\ I_n & \text{if } n \in \mathcal{R} \end{cases}.$$

Proof. The proof follows from the connections between the Tucker operator and n-mode multiplication (Proposition 4.2) and the connections between n-mode multiplication and matricization (Proposition 3.7). □

Results such as these help to yield insight into the properties of the Tucker operator. Consider the following proposition that says that the norm of a large tensor can be calculated by considering a much smaller tensor.

Proposition 4.4 ([7]) *Let $\mathcal{Y} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$ and let $\mathcal{N} = \{1, \ldots, N\}$. Suppose we have matrices $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$ for all $n \in \mathcal{N}$. Let the QR decomposition of each matrix be denoted by*

$$\mathbf{A}^{(n)} = \mathbf{Q}^{(n)} \mathbf{R}^{(n)} \text{ for } n \in \mathcal{N},$$

where $\mathbf{Q}^{(n)} \in \mathbb{R}^{I_n \times J_n}$ is orthonormal and $\mathbf{R}^{(n)} \in \mathbb{R}^{J_n \times J_n}$ is upper triangular. Then

$$\left\| [\![\mathcal{Y} \,; \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!] \right\| = \left\| [\![\mathcal{Y} \,; \mathbf{R}^{(1)}, \ldots, \mathbf{R}^{(N)}]\!] \right\|.$$

Proof. From the properties of the Tucker operator (Proposition 4.2(a)), the definition of the Tucker operator (Definition 4.1), and the property that orthonormal matrices in n-mode multiplication do not change the norm (Proposition 3.12), respectively, we have

$$\begin{aligned} \left\| [\![\mathcal{X} \,; \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!] \right\| &= \left\| [\![\, [\![\mathcal{X} \,; \mathbf{R}^{(1)}, \ldots, \mathbf{R}^{(N)}]\!] \,; \mathbf{Q}^{(1)}, \ldots, \mathbf{Q}^{(N)}]\!] \right\| \\ &= \left\| \left([\![\mathcal{X} \,; \mathbf{R}^{(1)}, \ldots, \mathbf{R}^{(N)}]\!] \right) \times_1 \mathbf{Q}^{(1)} \cdots \times_N \mathbf{Q}^{(N)} \right\| \\ &= \left\| [\![\mathcal{X} \,; \mathbf{R}^{(1)}, \ldots, \mathbf{R}^{(N)}]\!] \right\|. \end{aligned}$$

□

Consequently, suppose that we have a tensor $\mathcal{X} \in I_1 \times I_2 \times \cdots \times I_N$ such that

$$\mathcal{X} = [\![\mathcal{Y} \,; \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!].$$

If $J_n \ll I_n$, the norm of $\mathcal{Y}$ is the same as the much smaller tensor $\mathcal{Z} \in J_1 \times J_2 \times \cdots \times J_N$ where

$$\mathcal{Z} = [\![\mathcal{Y} \,; \mathbf{R}^{(1)}, \ldots, \mathbf{R}^{(N)}]\!].$$

4.3 The Tucker decomposition

The Tucker decomposition [43] of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is given by

$$\mathcal{X} = [\![\mathcal{G} \,; \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)}]\!]. \tag{4}$$

Here $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$ and $\mathcal{G} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$. If $\mathcal{G}$ is the same size as $\mathcal{X}$, the Tucker decomposition is simply a change of basis. More often, we are interested in using a change of basis to compress $\mathcal{X}$, thereby resulting in a tensor $\mathcal{G}$ that is smaller than $\mathcal{X}$; see Figure 5. The *n-rank* of a tensor $\mathcal{X}$ is defined as the rank of $\mathbf{X}_{(n)}$ [14]. If we let J_n be the n-rank of $\mathcal{X}$ for each n, the we can always reproduce $\mathcal{X}$ exactly. Otherwise, the "decomposition" may not be exact but instead produce an approximation to the tensor.

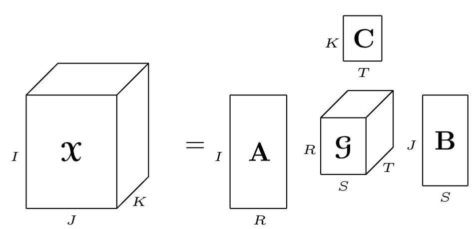


Figure 5. Illustration of the Tucker decomposition: $\mathcal{X} = [\![\mathcal{G} \,; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$

Tucker [43] dealt only with 3-way arrays but the basic principals have since been extended to N-way arrays. For the three-way case, the terms Tucker3, Tucker2, and Tucker1 have been coined [28]: Tucker3 is the decomposition presented here with $N = 3$ (three modes are free), Tucker2 constrains one mode to be the identity matrix (so that 2 modes are free), and Tucker1 constrains two modes to be identity matrices (1 mode is free).

In general, the Tucker decomposition is not unique. For example, let $\mathbf{B}$ be an orthogonal matrix of size $J_1 \times J_1$. Then, recalling Proposition 4.2(a),

$$\mathcal{X} = [\![\mathcal{G} \,; \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!] = [\![\mathcal{G} \times_1 \mathbf{B} \,; \mathbf{A}^{(1)}\mathbf{B}, \ldots, \mathbf{A}^{(N)}]\!].$$

Many researchers have considered the problems of rotating the core $\mathcal{G}$ to something that is more interpretable; see, e.g., [21, 22, 23, 2].

The new Tucker operator replaces the following options for expressing the Tucker decomposition:

- Mode-n multiplication: $\mathcal{X} = \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)}$,
- Matricized form: $\mathbf{X}_{(n)} = \mathbf{A}^{(n)} \mathbf{G}_{(n)} \left(\mathbf{A}^{(n)} \otimes \cdots \otimes \mathbf{A}^{(n+1)} \otimes \mathbf{A}^{(n-1)} \otimes \cdots \otimes \mathbf{A}^{(1)} \right)$,
- Outer products: $\mathcal{X} = \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \cdots \sum_{j_N=1}^{J_N} g_{j_1 j_2 \cdots j_N} \, \mathbf{a}^{(1)}_{:j_1} \circ \mathbf{a}^{(2)}_{:j_2} \circ \cdots \circ \mathbf{a}^{(N)}_{:j_N}$, or
- Elementwise: $x_{i_1 i_2 \cdots i_N} = \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \cdots \sum_{j_N=1}^{J_N} g_{j_1 j_2 \cdots j_N} \, a^{(1)}_{i_1 j_1} a^{(2)}_{i_2 j_2} \cdots a^{(n)}_{i_N j_N}$.

4.4 Finding an optimal rank-$(J_1, J_2, \ldots, J_N)$ approximation

Given a tensor $\mathcal{X}$ and a desired rank of the core tensor $\mathcal{G}$, we can consider the problem of computing a Tucker decomposition with the least amount of error. The goal is to find the best possible Tucker decomposition (4) given a tensor $\mathcal{X}$ or, in other words, to solve

$$\begin{aligned} \min_{\mathcal{G},\mathbf{A}^{(1)},\ldots,\mathbf{A}^{(N)}} \quad & \left\| \mathcal{X} - [\![\mathcal{G} \,; \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!] \right\| \\ \text{subject to} \quad & \mathcal{G} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N} \\ & \mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n} \text{orthonormal}, n = 1, \ldots, N. \end{aligned} \tag{5}$$

We assume J_n is strictly less than the n-rank of $\mathcal{X}$ in at least one mode — otherwise the solution is trivial and exact. We reformulate the problem so that $\mathcal{G}$ is eliminated by considering the problem of finding the optimal $\mathcal{G}$ given that all the matrices $\mathbf{A}^{(n)}$ are fixed. We present an alternative proof that explicitly uses the properties of the Tucker operator.

Theorem 4.5 (Theorems 4.1 and 4.2 in [14]) *Let $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. Assuming the matrices $\mathbf{A}^{(n)}$ are fixed, the optimal $\mathcal{G}$ for (5) is*

$$\mathcal{G} = [\![\mathcal{X} \,; \mathbf{A}^{(1)\mathsf{T}}, \ldots, \mathbf{A}^{(N)\mathsf{T}}]\!] \tag{6}$$

Consequently, the optimal matrices $\mathbf{A}^{(n)}$ for (5) are given by the solution to

$$\max_{\mathbf{A}^{(1)},\ldots,\mathbf{A}^{(N)}} \left\| [\![\mathcal{X} \,; \mathbf{A}^{(1)\mathsf{T}}, \ldots, \mathbf{A}^{(N)\mathsf{T}}]\!] \right\|. \tag{7}$$

Proof. From Proposition 4.3(a) with $\mathcal{R} = \{1, \ldots, N\}$, we can rewrite the norm in matrix form:

$$\left\| \mathcal{X} - [\![\mathcal{G} \,; \mathbf{A}^{(1)\mathsf{T}}, \ldots, \mathbf{A}^{(N)\mathsf{T}}]\!] \right\| = \left\| \mathrm{vec}(\mathcal{X}) - \left(\mathbf{A}^{(N)} \otimes \cdots \otimes \mathbf{A}^{(1)} \right) \mathrm{vec}(\mathcal{G}) \right\|$$

This is a classic linear least squares problem, and the solution is given by

$$\mathrm{vec}(\mathcal{G}) = \left(\mathbf{A}^{(N)} \otimes \cdots \otimes \mathbf{A}^{(1)} \right)^{\dagger} \mathrm{vec}(\mathcal{X}).$$

By Proposition 3.1(b) and the fact that the matrices are orthonormal, we can conclude

$$\mathrm{vec}(\mathcal{G}) = \left(\mathbf{A}^{(N)\mathsf{T}} \otimes \cdots \otimes \mathbf{A}^{(1)\mathsf{T}} \right) \mathrm{vec}(\mathcal{X})$$

Equation (6) follows from Proposition 4.3(a), so we assume (6) holds for the remainder of the proof.

Next, from Proposition 3.9, we have

$$\begin{aligned} &\left\| \mathcal{X} - [\![\mathcal{G} \,; \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!] \right\| \\ &\qquad = \| \mathcal{X} \|^2 - 2 \langle \mathcal{X}, [\![\mathcal{G} \,; \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!] \rangle + \left\| [\![\mathcal{G} \,; \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!] \right\|^2. \end{aligned}$$

From Proposition 3.11 and (6),

$$\langle \mathcal{X}, [\![\mathcal{G} \,; \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!] \rangle = \langle [\![\mathcal{X} \,; \mathbf{A}^{(1)\mathsf{T}}, \ldots, \mathbf{A}^{(N)\mathsf{T}}]\!], \mathcal{G} \rangle = \| \mathcal{G} \|^2$$

From Proposition 3.12,

$$\left\| [\![\mathcal{G} \,; \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!] \right\| = \| \mathcal{G} \|$$

Hence,

$$\left\| \mathcal{X} - [\![\mathcal{G} \,; \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!] \right\| = \| \mathcal{X} \|^2 - \| \mathcal{G} \|^2 .$$

It follows that minimizing (6) is equivalent to maximizing $\| \mathcal{G} \|$; hence, the claim. □

Consequently, from Theorem 4.5, the minimization problem (5) can be reformulated as:

$$\begin{aligned} \max_{\mathbf{A}^{(1)},\ldots,\mathbf{A}^{(N)}} \quad & \left\| [\![\mathcal{X} \,; \mathbf{A}^{(1)\mathsf{T}}, \ldots, \mathbf{A}^{(N)\mathsf{T}}]\!] \right\| \\ \text{subject to} \quad & \mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n} \text{ orthonormal }, n = 1, \ldots, N. \end{aligned} \tag{8}$$

Next, we can consider the question of how to find each $\mathbf{A}^{(n)}$ without making any assumptions about the other factors, i.e., we solve the following problem.

$$\begin{aligned} \max_{\mathbf{A}^{(n)}} \quad & \left\| [\![\mathcal{X} \,; \mathbf{I}, \ldots, \mathbf{I}, \mathbf{A}^{(n)\mathsf{T}}, \mathbf{I}, \ldots, \mathbf{I}]\!] \right\| \\ \text{subject to} \quad & \mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n} \text{ orthonormal.} \end{aligned} \tag{9}$$

The objective function is equivalent to $\left\| \mathcal{X} \times_n \mathbf{A}^{(n)\mathsf{T}} \right\| = \left\| \mathbf{A}^{(n)\mathsf{T}} \mathbf{X}_{(n)} \right\|$. In matrix format, we can see that the J_n leading left singular vectors of $\mathbf{X}_{(n)}$ yield the optimal solution. If we solve for each $\mathbf{A}^{(n)}$ for $n = 1, \ldots, N$ in this manner, than we have what is has been popularized as the Higher-Order Singular Value Decomposition (HO-SVD) [14]. Unlike its matrix counterpart, the HO-SVD does not yield an optimal rank-$J_1, J_2, \ldots, J_N$ approximation to $\mathcal{X}$. However, it is a good starting point for an alternating algorithm.

Consider next the problem of how to find the optimal $\mathbf{A}^{(n)}$ given that all the other factors are known and fixed, which yields the following subproblem for matrix n:

$$\begin{aligned} \max_{\mathbf{A}^{(n)}} \quad & \left\| [\![\mathcal{X} \,; \mathbf{A}^{(1)\mathsf{T}}, \ldots, \mathbf{A}^{(N)\mathsf{T}}]\!] \right\| \\ \text{subject to} \quad & \mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n} \text{ orthonormal.} \end{aligned} \tag{10}$$

More simply, defining $\mathcal{Z} \equiv [\![\mathcal{X} \,; \mathbf{A}^{(1)\mathsf{T}}, \ldots, \mathbf{A}^{(n-1)\mathsf{T}}, \mathbf{I}, \mathbf{A}^{(n+1)\mathsf{T}}, \ldots, \mathbf{A}^{(N)\mathsf{T}}]\!]$ and $\mathbf{B} \equiv \mathbf{A}^{(n)}$, the problem becomes

$$\begin{aligned} \max_{\mathbf{B}} \quad & \left\| \mathcal{Z} \times_n \mathbf{B}^{\mathsf{T}} \right\| \equiv \left\| \mathbf{B}^{\mathsf{T}} \mathbf{Z}_{(n)} \right\|_F \\ \text{subject to} \quad & \mathbf{B} \in \mathbb{R}^{J_n \times I_n} \text{ orthonormal.} \end{aligned} \tag{11}$$

The solution to this subproblem is easily realized via the matrix SVD of $\mathbf{Z}_{(n)}$, i.e., setting the columns of $\mathbf{B}$ to be the J_n leading left singular vectors of $\mathbf{Z}_{(n)}$ yields the optimal solution.

This leads naturally an alternating algorithm [15, 28, 39] to compute an approximate Tucker decomposition, shown in Algorithm 1.

4.5 Derivatives

Before we continue, we consider the derivatives of the Tucker operator. Let $\mathcal{N} = \{1, \ldots, N\}$, $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$, and $\mathcal{G} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$. Define a function $\mathcal{F}$ as the Tucker operator, i.e.,

$$\mathcal{F}(\mathcal{G}, \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}) = [\![\mathcal{G} \,; \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!].$$

Algorithm 1 Tucker: Higher Order Orthogonal Iteration

in: Tensor $\mathfrak{X}$ of size $I_1 \times I_2 \times \cdots \times I_N$.
in: Desired rank of core: $J_1 \times J_2 \times \cdots \times J_N$.
for n=1,...,N **do** {initialization via HO-SVD}
  $\mathbf{A}^{(n)} \leftarrow J_n$ leading eigenvalues of $\mathbf{X}_{(n)}\mathbf{X}_{(n)}^{\mathsf{T}}$
end for
while not converged **do** {main loop}
  for n=1,...,N **do**
    $\mathfrak{Z} \leftarrow [\![\mathfrak{X}\,; \mathbf{A}^{(1)\mathsf{T}}, \ldots, \mathbf{A}^{(n-1)\mathsf{T}}, \mathbf{I}, \mathbf{A}^{(n+1)\mathsf{T}}, \ldots, \mathbf{A}^{(N)\mathsf{T}}]\!]$
    $\mathbf{A}^{(n)\mathsf{T}} \leftarrow J_n$ leading eigenvalues of $\mathbf{Z}_{(n)}\mathbf{Z}_{(n)}^{\mathsf{T}}$
  end for
end while
$\mathfrak{G} \leftarrow [\![\mathfrak{X}\,; \mathbf{A}^{(1)\mathsf{T}}, \ldots, \mathbf{A}^{(N)\mathsf{T}}]\!]$
out: $\mathfrak{G}$ of size $J_1 \times J_2 \times \cdots \times J_N$ and orthonormal matrices $\mathbf{A}^{(n)}$ of size $I_n \times J_n$ such that $\mathfrak{X} \approx [\![\mathfrak{G}\,; \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!]$.

Consider the partial derivative of $\mathfrak{F}$ with respect to $\mathfrak{G}$. The result is a $2N$-way array such that

$$\left(\frac{\partial \mathfrak{F}}{\partial \mathfrak{G}}\right)_{i_1 i_2 \cdots i_N j_1 j_2 \cdots j_N} = a^{(1)}_{i_1 j_1} a^{(2)}_{i_2 j_2} \cdots a^{(N)}_{i_N j_N}.$$

In matricized form, this is

$$\left(\frac{\partial \mathfrak{F}}{\partial \mathfrak{G}}\right)_{(\mathfrak{I}\times\mathfrak{J})} = \mathbf{A}^{(N)} \otimes \mathbf{A}^{(N-1)} \otimes \cdots \otimes \mathbf{A}^{(1)}.$$

Consider the derivative of $\mathfrak{F}$ with respect to $\mathbf{A}^{(n)}$. The result is an $(N+2)$-way array, such that

$$\left(\frac{\partial \mathfrak{F}}{\partial \mathbf{A}^{(n)}}\right)_{i_1 i_2 \cdots i_N i_n j_n} = \sum_{j_1=1}^{J_1} \cdots \sum_{j_{n-1}=1}^{J_{n-1}} \sum_{j_{n+1}=1}^{J_{n+1}} \cdots \sum_{j_N=1}^{J_N} g_{j_1 j_2 \cdots j_N}\, a^{(1)}_{i_1 j_1} \cdots a^{(n-1)}_{i_{n-1} j_{n-1}} a^{(n+1)}_{i_{n+1} j_{n+1}} \cdots a^{(N)}_{i_N j_N}.$$

Another way to see this is as follows. In matricized form, we have

$$\mathbf{F}_{(n)} = \mathbf{A}^{(n)} \mathbf{G}_{(n)} \left(\mathbf{A}^{(N)} \otimes \cdots \otimes \mathbf{A}^{(n+1)} \otimes \mathbf{A}^{(n-1)} \otimes \mathbf{A}^{(1)}\right)^{\mathsf{T}}.$$

Thus, from matrix calculus (see, e.g., [17]), we have

$$\frac{\partial \mathbf{F}_{(n)}}{\partial \mathbf{A}^{(n)}} = \left[\left(\mathbf{A}^{(N)} \otimes \cdots \otimes \mathbf{A}^{(n+1)} \otimes \mathbf{A}^{(n-1)} \otimes \mathbf{A}^{(1)}\right) \mathbf{G}_{(n)}^{\mathsf{T}}\right] \otimes \mathbf{I}$$

where $\mathbf{I}$ is the $I_n \times I_n$ identity matrix.

5 The Kruskal operator

The Kruskal operator provides shorthand notation for the sum of the outer products of the columns of a set of matrices. This turns out to be a special case of the Tucker operator where the core tensor is the identity tensor. Unlike the Tucker operator, which can be written using n-mode multiplication, there is no concise multidimensional representation for this special case. The result is that this operation is usually expressed in matricized form, which tends to obscure its multidimensional properties.

5.1 Definition of the Kruskal operator

The Kruskal operator is a special case of the Tucker operator (Definition 4.1) where the core tensor $\mathcal{G}$ is the $R \times R \times \cdots \times R$ identity tensor and all the matrices $\mathbf{A}^{(n)}$ have R columns.

Definition 5.1 *Let $\mathcal{N} = \{1, \ldots, N\}$. Suppose we have matrices $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ for $n \in \mathcal{N}$. Then the* Kruskal operator *is defined as:*

$$[\![\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)}]\!] \equiv [\![\mathcal{I} \,; \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)}]\!], \tag{12}$$

where $\mathcal{I}$ is the identity tensor, i.e., it has ones along the superdiagonal and zeros elsewhere.

See Figure 1(b) for an illustration of the identity tensor. We call this operator the Kruskal operator since such an operator was proposed by Kruskal [29].

5.2 Kruskal operator properties

The properties of the Kruskal operator are much more interesting than those of the Tucker operator because they do not always directly from the n-mode multiplication results.

The following proposition shows what happens when a Kruskal operator is the core tensor of a Tucker operator, which can happen when compression is used as the first step in the calculation of a PARAFAC decomposition [4].

Proposition 5.2 *Let $\mathcal{N} = \{1, \ldots, N\}$. Suppose we have matrices $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$, $\mathbf{B}^{(n)} \in \mathbb{R}^{K_n \times I_n}$ for all $n \in \mathcal{N}$. Then*

$$[\![\, [\![\mathbf{A}^{(1)}, \cdots, \mathbf{A}^{(N)}]\!] \,; \mathbf{B}^{(1)}, \cdots, \mathbf{B}^{(N)}]\!] = [\![\mathbf{B}^{(1)}\mathbf{A}^{(1)}, \cdots, \mathbf{B}^{(N)}\mathbf{A}^{(N)}]\!].$$

We can also consider the relationship between the Kruskal operator, matricization, and the Khatri-Rao product. This is the analogue of Proposition 4.3 for the Tucker operator.

Proposition 5.3 *Let $\mathcal{N} = 1, \ldots, N$. Let $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ for all $n \in \mathcal{N}$.*

(a) *If $\mathcal{R} = \{r_1, \ldots, r_L\}$ and $\mathcal{C} = \{c_1, \ldots, c_M\}$ partition $\mathcal{N}$, then*

$$\mathcal{X} = [\![\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \cdots, \mathbf{A}^{(N)}]\!] \quad \Leftrightarrow \quad \mathbf{X}_{(\mathcal{R} \times \mathcal{C} \,:\, I_{\mathcal{N}})} = \left(\mathbf{A}^{(r_L)} \odot \cdots \odot \mathbf{A}^{(r_1)}\right) \left(\mathbf{A}^{(c_M)} \odot \cdots \odot \mathbf{A}^{(c_1)}\right)^{\mathsf{T}}$$

If $\mathcal{R} = \emptyset$, then the first multiplicand is replaced by a length-R row vector of all ones; conversely, if $\mathcal{C} = \emptyset$, then the second multiplicand is replaced by a length-R column vector of all ones. In other words,

$$\mathbf{X}_{(\emptyset \times \mathcal{N} \,:\, I_{\mathcal{N}})} \equiv \mathbf{1}^{\mathsf{T}} \left(\mathbf{A}^{(N)} \odot \cdots \odot \mathbf{A}^{(1)}\right)^{\mathsf{T}} \text{ and } \mathbf{X}_{(\mathcal{N} \times \emptyset \,:\, I_{\mathcal{N}})} \equiv \left(\mathbf{A}^{(N)} \odot \cdots \odot \mathbf{A}^{(1)}\right) \mathbf{1}.$$

(b) *Consequently, for any specific $n \in \mathcal{N}$ we have*

$$\mathcal{X} = [\![\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \cdots, \mathbf{A}^{(N)}]\!] \quad \Leftrightarrow \quad \mathbf{X}_{(n)} = \mathbf{A}^{(n)} \left(\mathbf{A}^{(N)} \odot \cdots \odot \mathbf{A}^{(n+1)} \odot \mathbf{A}^{(n-1)} \odot \cdots \odot \mathbf{A}^{(1)}\right)^{\mathsf{T}}.$$

The norm of the Kruskal operator has a very special form because it can be reduced to summing the entries of the Hadamard product of N matrices of size $R \times R$.

Proposition 5.4 *Let $\mathcal{N} = 1, \ldots, N$ and $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ for all $n \in \mathcal{N}$. Then*

$$\left\| [\![\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \cdots, \mathbf{A}^{(N)}]\!] \right\|^2 = \sum_{j=1}^{R} \sum_{k=1}^{R} \left((\mathbf{A}^{(1)\mathsf{T}} \mathbf{A}^{(1)}) * (\mathbf{A}^{(2)\mathsf{T}} \mathbf{A}^{(2)}) * \cdots * (\mathbf{A}^{(N)\mathsf{T}} \mathbf{A}^{(N)}) \right)_{jk}$$

Proof. The proof is a matter of using the definition of the Kruskal operator (Definition 5.1) and rearranging the terms appropriately.

$$\begin{aligned}
\left\| [\![\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \cdots, \mathbf{A}^{(N)}]\!] \right\|^2 &= \langle [\![\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \cdots, \mathbf{A}^{(N)}]\!], [\![\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \cdots, \mathbf{A}^{(N)}]\!] \rangle \\
&= \langle \sum_{j=1}^{R} \mathbf{a}_{:j}^{(1)} \circ \cdots \circ \mathbf{a}_{:j}^{(N)}, \sum_{k=1}^{R} \mathbf{a}_{:k}^{(1)} \circ \cdots \circ \mathbf{a}_{:k}^{(N)} \rangle \\
&= \sum_{j=1}^{R} \sum_{k=1}^{R} \left(\mathbf{a}_{:j}^{(1)\mathsf{T}} \mathbf{a}_{:k}^{(1)} \right) \cdots \left(\mathbf{a}_{:j}^{(N)\mathsf{T}} \mathbf{a}_{:k}^{(N)} \right) \\
&= \sum_{j=1}^{R} \sum_{k=1}^{R} \left(\mathbf{A}^{(1)\mathsf{T}} \mathbf{A}^{(1)} \right)_{jk} \cdots \left(\mathbf{A}^{(N)\mathsf{T}} \mathbf{A}^{(N)} \right)_{jk}
\end{aligned}$$

The third step used Proposition 3.10. □

Proposition 5.5 *Let $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and $\mathcal{N} = 1, \ldots, N$ and $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ for all $n \in \mathcal{N}$.*

(a) *The inner product of $\mathcal{X}$ and the Kruskal product yields:*

$$\begin{aligned}
&\langle \mathcal{X}, [\![\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \cdots, \mathbf{A}^{(N)}]\!] \rangle \\
&\quad = \sum_{r=1}^{R} \mathcal{X} \bar{\times}_1 \mathbf{a}_{:r}^{(1)} \bar{\times}_2 \mathbf{a}_{:r}^{(2)} \cdots \bar{\times}_N \mathbf{a}_{:r}^{(N)} \\
&\quad = \sum_{j=1}^{J} \sum_{k=1}^{K} \left(\mathbf{X}_{(\mathcal{R} \times \mathcal{C} \,:\, I_N)} * \left[\left(\mathbf{A}^{(r_L)} \odot \cdots \odot \mathbf{A}^{(r_1)} \right) \left(\mathbf{A}^{(c_M)} \odot \cdots \odot \mathbf{A}^{(c_1)} \right)^{\mathsf{T}} \right] \right)_{jk} \\
&\quad = \langle \operatorname{vec}(\mathcal{X}), \left(\mathbf{A}^{(N)} \odot \cdots \odot \mathbf{A}^{(1)} \right) \mathbf{1} \rangle
\end{aligned}$$

(b) *The norm of the difference of $\mathcal{X}$ and the Kruskal product is:*

$$\left\| \mathcal{X} - [\![\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!] \right\|^2 = \| \mathcal{X} \|^2 - 2 \langle \mathcal{X}, [\![\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!] \rangle + \left\| [\![\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!] \right\|^2$$

5.3 The PARAFAC decomposition

The PARAFAC decomposition of $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is given by

$$\mathcal{X} = [\![\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)}]\!].$$

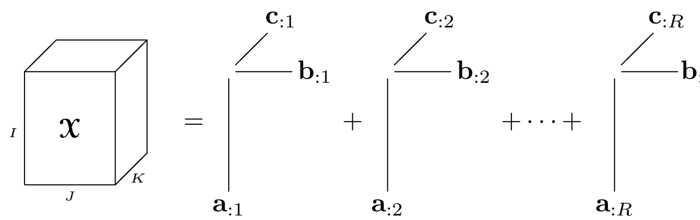


Figure 6. Illustration of the PARAFAC decomposition: $\mathcal{X} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$

Here $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$, for $n = 1, \ldots, N$. The PARAFAC decomposition of a three-way tensor is illustrated in Figure 6.

In the case that R is minimal, then R is the *rank* of $\mathcal{X}$ [29]. It can be the case that $R > \min\{I_n \mid n \in \mathcal{N}\}$. For example, the maximal rank of a $2 \times 2 \times 2$ tensor is 3 [40, 30]. Moreover, the typical rank of a $2 \times 2 \times 2$ tensor is 2 (79% of the time) and 3 (21% of the time) [30]. See [9, 10] for an overview of PARAFAC and related decompositions.

The new Kruskal operator replaces the following options for expressing the PARAFAC decomposition:

- Elementwise: $x_{i_1 i_2 \cdots i_N} = \sum_{r=1}^{R} a^{(1)}_{i_1 r} \cdot a^{(2)}_{i_2 r} \cdot \cdots \cdot a^{(N)}_{i_N r}$,
- Sum of outer products: $\mathcal{X} = \sum_{r=1}^{R} \mathbf{a}^{(1)}_{:r} \circ \mathbf{a}^{(2)}_{:r} \circ \cdots \circ \mathbf{a}^{(N)}_{:r}$,
- Matricized: $\mathbf{X}_{(n)} = \mathbf{A}^{(n)} \left(\mathbf{A}^{(N)} \odot \cdots \odot \mathbf{A}^{(n+1)} \odot \mathbf{A}^{(n-1)} \odot \cdots \odot \mathbf{A}^{(1)} \right)^{\mathsf{T}}$, and
- Vectorized: $\mathrm{vec}(\mathcal{X}) = \left(\mathbf{A}^{(N)} \odot \cdots \odot \mathbf{A}^{(1)} \right) \mathbf{1}$ where $\mathbf{1}$ is a ones vector or length R.
- Slice notation (three-way only): If $\mathcal{X} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \in \mathbb{R}^{I \times J \times K}$, then we can for example write each frontal slice (see Figure 3) as

$$\mathbf{X}_{::k} = \mathbf{A}\mathbf{D}^{(k)}\mathbf{B}^{\mathsf{T}} \qquad \text{for } k = 1, \ldots, K,$$

  where the $R \times R$ diagonal matrix $\mathbf{D}^{(k)}$ is defined by $\mathbf{D}^{(k)} = \mathrm{diag}(\mathbf{c}_{k:})$. Slice notation can be used in the other directions as well:

$$\begin{aligned} \mathbf{X}_{i::} &= \mathbf{B}\mathrm{diag}(\mathbf{a}_{i:})\mathbf{C}^{\mathsf{T}} && \text{for } i = 1, \ldots, I, \text{ and} \\ \mathbf{X}_{:r:} &= \mathbf{A}\mathrm{diag}(\mathbf{b}_{j:})\mathbf{C}^{\mathsf{T}} && \text{for } j = 1, \ldots, J. \end{aligned}$$

5.4 Computing the PARAFAC decomposition

Faber et al. [16] present an overview of different methods for fitting a PARAFAC decomposition, and alternating least squares continues to be the workhorse algorithm (i.e., slow but steady) and thus is our focus here.

Given a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and a desired rank R, the alternating least squares (ALS) algorithm is used to compute a PARAFAC factorization. In general, it is not known how to choose R in advance; see, e.g., [11] for more discussion of this issue. Here, we assume that R is known.

As is the case with computing the Tucker approximation, the idea behind ALS is that we solve for each factor in turn, leaving all the other factors fixed. Thus, the subproblem at each iteration is as follows: Suppose that all factors $\mathbf{A}^{(m)}$, $m \neq n$, are fixed and solve for $\mathbf{B} \equiv \mathbf{A}^{(n)}$. This can be cast as the following optimization problem:

$$\min_{\mathbf{B}\in\mathbb{R}^{I_n\times R}} \left\| \mathcal{X} - [\![\mathbf{A}^{(1)},\ldots,\mathbf{A}^{(n-1)},\mathbf{B},\mathbf{A}^{(n+1)},\ldots,\mathbf{A}^{(N)}]\!] \right\|.$$

From Proposition 5.3(b), this can be expressed in matrix form as

$$\min_{\mathbf{B}\in\mathbb{R}^{I_n\times R}} \left\| \mathbf{X}_{(n)} - \mathbf{B}\left(\mathbf{A}^{(N)}\odot\ldots\odot\mathbf{A}^{(n+1)}\odot\mathbf{A}^{(n-1)}\odot\ldots\odot\mathbf{A}^{(1)}\right)^{\mathsf{T}} \right\|,$$

which is a classic least squares problem. Using Proposition 3.2(c), the optimal solution is easily computed as

$$\begin{aligned}\mathbf{B}^{\mathsf{T}} &= \left(\mathbf{A}^{(N)}\odot\ldots\odot\mathbf{A}^{(n+1)}\odot\mathbf{A}^{(n-1)}\odot\ldots\odot\mathbf{A}^{(1)}\right)^{\dagger}\mathbf{X}_{(n)}^{\mathsf{T}}\\ &= \mathbf{V}^{\dagger}\left(\mathbf{A}^{(N)}\odot\ldots\odot\mathbf{A}^{(n+1)}\odot\mathbf{A}^{(n-1)}\odot\ldots\odot\mathbf{A}^{(1)}\right)^{\mathsf{T}}\mathbf{X}_{(n)}^{\mathsf{T}}\end{aligned}$$

where

$$\mathbf{V} = (\mathbf{A}^{(N)\mathsf{T}}\mathbf{A}^{(N)}) * \cdots * (\mathbf{A}^{(n+1)\mathsf{T}}\mathbf{A}^{(n+1)}) * (\mathbf{A}^{(n-1)\mathsf{T}}\mathbf{A}^{(n-1)}) * \cdots * (\mathbf{A}^{(1)\mathsf{T}}\mathbf{A}^{(1)}).$$

Note that $\mathbf{V}$ is of size $R\times R$ and symmetric. An interesting observation is worth making here, which is that the pseudoinverse can be recast using the Kruskal operator. Define

$$\mathcal{Z} = [\![\mathbf{A}^{(1)},\ldots,\mathbf{A}^{(n-1)},\mathbf{V}^{\dagger},\mathbf{A}^{(n+1)},\ldots,\mathbf{A}^{(N)}]\!],$$

which is of size $I_1\times\cdots I_{n-1}\times R\times I_{n+1}\times\cdots\times I_N$. Then we have

$$\mathbf{B} = \mathbf{X}_{(n)}\mathbf{Z}_{(n)}^{\mathsf{T}} \in \mathbb{R}^{I_n\times R}.$$

A basic ALS algorithm is shown in Algorithm 2.

Algorithm 2 PARAFAC: Alternating Least Squares (ALS)

in: Tensor $\mathcal{X}$ of size $I_1\times I_2\times\cdots\times I_N$.
in: Desired rank of result: $R>0$.
for $n=1,\ldots,N$ **do** {initialization}
  Initialize $\mathbf{A}^{(n)}$ in some way (e.g., random or HO-SVD).
  Normalize columns of $\mathbf{A}^{(n)}$.
  $\mathbf{B}^{(n)} \leftarrow \mathbf{A}^{(n)\mathsf{T}}\mathbf{A}^{(n)}$.
end for
while not converged **do** {main loop}
  for $n=1,\ldots,N$ **do**
    $\mathbf{V} \leftarrow \mathbf{B}^{(N)} * \cdots * \mathbf{B}^{(n+1)} * \mathbf{B}^{(n-1)} * \cdots * \mathbf{B}^{(1)}$.
    $\mathbf{A}^{(n)} \leftarrow \mathbf{X}_{(n)}\left(\mathbf{A}^{(N)}\odot\ldots\odot\mathbf{A}^{(n+1)}\odot\mathbf{A}^{(n-1)}\odot\ldots\odot\mathbf{A}^{(1)}\right)^{\mathsf{T}}\mathbf{V}^{\dagger}$
    if $n\neq N$ **then**
      Normalize columns of $\mathbf{A}^{(n)}$.
    end if
    Set $\mathbf{B}^{(n)} \leftarrow \mathbf{A}^{(n)\mathsf{T}}\mathbf{A}^{(n)}$.
  end for
end while
out: $\mathbf{A}^{(n)}\in\mathbb{R}^{I_n\times R}$ for $n=1,\ldots,N$ such that $\mathcal{X}\approx[\![\mathbf{A}^{(1)},\ldots,\mathbf{A}^{(N)}]\!]$.

5.5 Derivatives of the Kruskal operator

Finally, we consider derivatives of the Kruskal operator. Let $\mathcal{N}=\{1,2\ldots,N\}$ and $\mathbf{A}^{(n)}\in\mathbb{R}^{I_n\times R}$ for all $n\in\mathcal{N}$. Define a function $\mathcal{F}$ as the Kruskal operator, i.e.,

$$\mathcal{F}(\mathbf{A}^{(1)},\ldots,\mathbf{A}^{(N)}) = [\![\mathbf{A}^{(1)},\ldots,\mathbf{A}^{(N)}]\!].$$

Consider the derivative of $\mathcal{F}$ with respect to $\mathbf{A}^{(n)}$. The result is an $(N+2)$-way array, such that

$$\left(\frac{\partial \mathcal{F}}{\partial \mathbf{A}^{(n)}}\right)_{i_1 i_2 \cdots i_N i_n r} = a^{(1)}_{i_1 r} \cdots a^{(n-1)}_{i_{n-1} r} a^{(n+1)}_{i_{n+1} r} \cdots a^{(N)}_{i_N r}.$$

Another way to see this is as follows. In matricized form, we have

$$\mathbf{F}_{(n)} = \mathbf{A}^{(n)} \left(\mathbf{A}^{(N)} \odot \cdots \odot \mathbf{A}^{(n+1)} \odot \mathbf{A}^{(n-1)} \odot \mathbf{A}^{(1)} \right)^{\mathsf{T}}.$$

Using the fact that $\operatorname{vec}(\mathbf{XYZ}) = (\mathbf{Z}^{\mathsf{T}} \otimes \mathbf{X}) \operatorname{vec}(\mathbf{Y})$ (see, e.g., [17]), we can rewrite the previous expression as

$$\operatorname{vec}\left(\mathbf{F}_{(n)}\right) = \left[\left(\mathbf{A}^{(N)} \odot \cdots \odot \mathbf{A}^{(n+1)} \odot \mathbf{A}^{(n-1)} \odot \mathbf{A}^{(1)} \right) \otimes \mathbf{I} \right] \operatorname{vec}\left(\mathbf{A}_{(n)}\right)$$

where $\mathbf{I}$ is the $I_n \times I_n$ identity matrix. Thus, from matrix calculus, we can define

$$\mathbf{J}^{(n)} \equiv \frac{\partial \operatorname{vec}\left(\mathbf{F}_{(n)}\right)}{\partial \operatorname{vec}\left(\mathbf{A}^{(n)}\right)} = \left(\mathbf{A}^{(N)} \odot \cdots \odot \mathbf{A}^{(n+1)} \odot \mathbf{A}^{(n-1)} \odot \mathbf{A}^{(1)} \right) \otimes \mathbf{I}$$

Note that the size of $\mathbf{J}^{(n)}$ is $\left(\prod_{n=1}^N I_n\right) \times (I_n R)$. Thus, each partial derivative has the same number of rows but a different number of columns. A full Jacobian for vec($\mathcal{F}$) can be constructed using the partials, but the rows have to be reordered for consistency [41]. Define $\mathbf{P}^{(n)}$ to a permutation matrix of size $\prod_{n=1}^N I_n$ that reorders $\mathbf{X}^{(n)}$ to be $\mathbf{X}^{(1)}$, i.e.,

$$\mathbf{X}_{(1)} = \mathbf{P}^{(n)} \mathbf{X}_{(n)}.$$

Then the full Jacobian of vec($\mathcal{F}$) is of size $\left(\prod_{n=1}^N I_n\right) \times \left(\sum_{n=1}^N I_n R\right)$ and defined by

$$\frac{d \operatorname{vec}(\mathcal{F})}{d\left(\left[\operatorname{vec}\left(\mathbf{A}^{(1)}\right)^{\mathsf{T}} \quad \cdots \quad \operatorname{vec}\left(\mathbf{A}^{(N)}\right)^{\mathsf{T}} \right]^{\mathsf{T}} \right)} = \left[\mathbf{J}^{(1)} \quad \mathbf{P}^{(2)} \mathbf{J}^{(2)} \quad \cdots \quad \mathbf{P}^{(N)} \mathbf{J}^{(N)} \right].$$

6 Conclusions

We consider two new operators that are useful for expressing and understanding higher-order tensor decompositions: The Tucker operator is shorthand for all-mode matrix multiplication, and the Kruskal operator is shorthand for the sum of the rank-1 tensors that are formed as outer products of the columns of the component matrices. By using these new operators, we can more easily express and understand the multilinear nature of the Tucker and PARAFAC decompositions because matricized representations can be avoided or at least easy to switch between. We have gathered together many commonly known and used properties but expressed them here in their native multilinear contexts, avoiding the potential confusion that comes about due to the numerous options for matricization and vectorization.

We have reviewed the ALS methods for both Tucker and PARAFAC using the new operators, though there are many approaches including those that handle constraints (see, e.g., [3, 16, 42, 47]). Moreover, many other approaches rely on gradient information (see, e.g., [34, 41]), so we have included derivatives of our operators. We also note that MATLAB software exists for working with tensors [8] and for efficiently computing the various decompositions [5].

References

[1] E. Acar, S. A. Çamtepe, M. S. Krishnamoorthy, and B. Yener, *Modeling and multiway analysis of chatroom tensors*, in ISI 2005: IEEE International Conference on Intelligence and Security Informatics, vol. 3495 of Lecture Notes in Computer Science, Springer Verlag, 2005, pp. 256–268.

[2] C. Andersson and R. Henrion, *A general algorithm for obtaining simple structure of core arrays in N-way PCA with application to fluorimetric data*, Comput. Stat. Data. An., 31 (1999), pp. 255–278.

[3] C. A. Andersson and R. Bro, *Improving the speed of multi-way algorithms: Part I: Tucker3*, Chemometr. Intell. Lab., 42 (1998), pp. 93–103.

[4] ——, *Improving the speed of multi-way algorithms: Part II: Compression*, Chemometr. Intell. Lab., 42 (1998), pp. 105–113.

[5] C. A. Andersson and R. Bro, *The N-way toolbox for MATLAB*, Chemometr. Intell. Lab., 52 (2000), pp. 1–4. See also `http://www.models.kvl.dk/source/nwaytoolbox/`.

[6] C. J. Appellof and E. R. Davidson, *Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents*, Anal. Chem., 53 (1981), pp. 2053–2056.

[7] B. W. Bader. private communication, 2005.

[8] B. W. Bader and T. G. Kolda, *Algorithm xxx: MATLAB tensor classes for fast algorithm prototyping*, ACM Transactions on Mathematical Software. To appear. See also `http://csmr.ca.sandia.gov/~tgkolda/TensorToolbox`.

[9] R. Bro, *PARAFAC. tutorial and applications*, Chemometr. Intell. Lab., 38 (1997), pp. 149–171.

[10] ——, *Multi-way analysis in the food industry. models, algorithms, and applications*, PhD thesis, University of Amsterdam, 1998. Available at `http://www.models.kvl.dk/research/theses/`.

[11] R. Bro and H. A. L. Kiers, *A new efficient method for determining the number of components in parafac models*, J. Chemometr., 17 (2003), pp. 274–286.

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

[13] B. Chen, A. Petropolu, and L. De Lathauwer, *Blind identification of convolutive MIM systems with 3 sources and 2 sensors*, Applied Signal Processing, (2002), pp. 487–496. (Special Issue on Space-Time Coding and Its Applications, Part II).

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

[15] ——, *On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensors*, SIAM J. Matrix Anal. A., 21 (2000), pp. 1324–1342.

[16] N. K. M. Faber, R. Bro, and P. K. Hopke, *Recent developments in CANDECOMP/PARAFAC algorithms: a critical review*, Chemometr. Intell. Lab., 65 (2003), pp. 119–137.

[17] P. L. Facker, *Notes on matrix calculus*. Available from `http://www4.ncsu.edu/~pfackler/MatCalc.pdf`, Sept. 2005.

[18] V. S. Grigorascu and P. A. Regalia, *Fast Reliable Algorithms for Matrices with Structure*, SIAM, 1999, ch. Tensor displacement structures and polyspectral matching, pp. 245–276.

[19] R. A. Harshman, *Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multi-modal factor analysis*, UCLA working papers in phonetics, 16 (1970), pp. 1–84.

[20] ——, *An index formulism that generalizes the capabilities of matrix notation and algebra to n-way arrays*, J. Chemometr., 15 (2001), pp. 689–714.

[21] R. Henrion, *Body diagonalization of core matrices in three-way principal components analysis: Theoretical bounds and simulation*, J. Chemometr., 7 (1993), pp. 477–494.

[22] ——, *N-way principal component analysis theory, algorithms and applications*, Chemometr. Intell. Lab., 25 (1994), pp. 1–23.

[23] H. A. Kiers, *Joint orthomax rotation of the core and component matrices resulting from three-mode principal components analysis*, J. Classif., 15 (1998), pp. 245 – 263.

[24] H. A. L. Kiers, *Towards a standardized notation and terminology in multiway analysis*, J. Chemometr., 14 (2000), pp. 105–122.

[25] T. G. Kolda, *Orthogonal tensor decompositions*, SIAM J. Matrix Anal. A., 23 (2001), pp. 243–255.

[26] T. G. Kolda and B. W. Bader, *The TOPHITS model for higher-order web link analysis*, in Workshop on Link Analysis, Counterterrorism and Security, 2006.

[27] T. G. Kolda, B. W. Bader, and J. P. Kenny, *Higher-order web link analysis using multilinear algebra*, in ICDM 2005: Proceedings of the 5th IEEE International Conference on Data Mining, IEEE Computer Society, 2005, pp. 242–249.

[28] P. M. Kroonenberg and J. De Leeuw, *Principal component analysis of three-mode data by means of alternating least squares algorithms*, Psychometrika, 45 (1980), pp. 69–97.

[29] J. B. Kruskal, *Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra Appl., 18 (1977), pp. 95–138.

[30] 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 Science Publishers B.V., 1989.

[31] R. P. McDonald, *A simple comprehensive model for the analysis of covariance structures*, Brit. J. Math. Stat. Psy., 33 (1980), p. 161. Cited in [10].

[32] M. Młrup, L. K. Hansen, C. S. Herrmann, J. Parnas, and S. M. Arnfred, *Parallel factor analysis as an exploratory tool for wavelet transformed event-related eeg*, NeuroImage, (2005). In Press, Corrected Proof, Available online 26 September 2005.

[33] D. Muti and S. Bourennane, *Multidimensional filtering based on a tensor approach*, Signal Process., 85 (2005), pp. 2338–2353.

[34] P. Paatero, *The multilinear engine - a table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model*, J. Comput. Graph. Stat., 8 (1999), pp. 854–888.

[35] C. R. Rao and S. Mitra, *Generalized inverse of matrices and its applications*, Wiley, New York, 1971. Cited in [10].

[36] B. Savas, *Analyses and tests of handwritten digit recognition algorithms*, master's thesis, Linkoping University, Sweden, 2003.

[37] A. Smilde, R. Bro, and P. Geladi, *Multi-way analysis: applications in the chemical sciences*, Wiley, 2004.

[38] J.-T. Sun, H.-J. Zeng, H. Liu, Y. Lu, and Z. Chen, *CubeSVD: a novel approach to personalized Web search*, in WWW 2005: Proceedings of the 14th international conference on World Wide Web, 2005, pp. 382–390.

[39] J. Ten Berge, J. De Leeuw, and P. M. Kroonenberg, *Some additional results on principal components analysis of three-mode data by means of alternating least squares algorithms*, Psychometrika, 52 (1987), pp. 183–191.

[40] J. M. F. Ten Berge, H. A. L. Kiers, and J. de Leeuw, *Explicit CANDECOMP/PARAFAC solutions for a contrived 222 array of rank three*, Psychometrika, 53 (1988), pp. 579–583.

[41] G. Tomasi, *Use of the properties of the Khatri-Rao product for the computation of Jacobian, Hessian, and gradient of the PARAFAC model under MATLAB*. 2005.

[42] G. Tomasi and R. Bro, *A comparison of algorithms for fitting the PARAFAC model*, Comput. Stat. Data. An., (2005).

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

[44] C. F. Van Loan, *The ubiquitous Kronecker product*, J. Comput. Appl. Math., 123 (2000), pp. 85–100.

[45] M. A. O. Vasilescu and D. Terzopoulos, *Multilinear analysis of image ensembles: TensorFaces*, in ECCV 2002: 7th European Conference on Computer Vision, vol. 2350 of Lecture Notes in Computer Science, Springer-Verlag, 2002, pp. 447–460.

[46] H. Wang and N. Ahuja, *Facial expression decomposition*, in ICCV 2003: 9th IEEE International Conference on Computer Vision, vol. 2, 2003, pp. 958–965.

[47] T. Zhang and G. H. Golub, *Rank-one approximation to high order tensors*, SIAM J. Matrix Anal. A., 23 (2001), pp. 534–550.

A LaTeX formatting

The double brackets used to denote the Tucker and Kruskal operators are produced as follows:

```
\usepackage{stmaryrd}          % provides \llbracket and \rrbracket
$\llbracket ... \rrbracket$    % here are the brackets
```

The boldface Euler script letters that are used to denote tensors are produced as follows:

```
\usepackage{amsmath}           % provides \boldsymbol
\usepackage[mathscr]{eucal}    % provides \mathscr (Euler script)
$\boldsymbol{\mathscr{X}}$     % here's a tensor X
```

DISTRIBUTION:

2 MS 9018
Central Technical Files, 8945-1

2 MS 0899
Technical Library, 4536

1 MS 0188
D. Chavez, LDRD Office, 1011