ELSEVIER

Chemometrics and Intelligent Laboratory Systems 28 (1995) 229–237

Chemometrics and intelligent laboratory systems

Tutorial

An introduction to tensor products with applications to multiway data analysis

Donald S. Burdick

Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708, USA

Received 27 January 1994; accepted 5 May 1994

Abstract

The concepts of tensor algebra and vector space geometry provide a unifying framework for multilinear data analysis which simplifies notation and leads to economy of thought. Avoiding too much abstraction too soon in defining tensor products makes these concepts accessible. Examples are given of the use of tensor algebra in the analysis of bilinear and trilinear models arising in fluorescence spectroscopy.

Contents

1. Introduction

The terminology and notation of tensor products and vector space geometry offer a framework which greatly simplifies the notational complexity often associated with multiway data analysis. This simplification leads to an economy of thought which makes the salient features of multilinear models and algorithms easier to understand.

Unfortunately, although the concepts of tensor algebra have been around for a long time, most of the extant literature on the subject is unnecessarily esoteric and emphasizes material with little relevance to multiway data analysis. It can be discouraging after struggling to understand the concepts of covariance and contravariance to discover that they have little to do with the analysis of multiway data.

This Tutorial has a twofold objective. The first is to present basic tensor product concepts in a form that is readily accessible to a reader with a knowledge of introductory linear algebra. The

SSDI 0169-7439(94)0048-N

second is to give some examples that illustrate the usefulness of these concepts in the analysis of two-way and three-way data arrays.

2. Definitions and notation

We begin by reviewing some basic terminology from linear algebra.

Definition 1 (vectors and vector spaces)

The symbol R^N denotes the Euclidean vector space consisting of all vectors $\mathbf{x}$ *with real coordinates* $\{x_1, \ldots, x_N\}$. *If* $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_R\}$ *is a set of vectors in* R^N, *then the set* $\mathscr{U}_X$ *of all linear combinations of the vectors in X is the linear span of X, written* $Sp[X]$. *The linear span* $\mathscr{U}_X$ *is a linear subspace of* R^N. *If the vectors in X are linearly independent, then X is a basis for* $\mathscr{U}_X$. *If* $\mathscr{U}$ *is a linear subspace of* R^N, *the dimension of* $\mathscr{U}$ *is the number of vectors in any basis for* $\mathscr{U}$.

Definition 1 is given in formal language for the sake of rigor and precision. Informally, we can think of vectors as entities that can be added together or multiplied by scalars. Any mathematical construct, e.g. function, which has the fundamental properties of vector addition and scalar multiplication can be regarded as a vector. For our purposes, though, a vector will be an array of real numbers and a scalar will also be a real number. Even with this limitation we have flexibility in the arrangement of the numbers in the array. Horizontal rows of numbers, vertical columns of numbers, matrices, and three-way arrays can all be regarded as vectors because they can be added and multiplied by scalars.

Definition 2 (matrices)

The vector space of all $I \times J$ *matrices will be denoted by* $\mathscr{M}[I, J]$. *The subspace of* R^I *spanned by the columns of an* $I \times J$ *matrix M is called the column space of M. Similarly, the row space of M is the subspace of* R^J *spanned by the rows of M. The rank of M is defined to be the dimension of its column space.*

There are other ways to define the notion of matrix rank. The dimension of the row space is an equivalent alternative, because the row and column spaces of a matrix can readily be shown to have the same dimension.

If $\mathbf{x}$ and $\mathbf{y}$ are nonzero column vectors in R^I and R^J, respectively, the $I \times J$ matrix $\mathbf{x}\mathbf{y}^\top$ (where $\mathbf{y}^\top$ denotes the transpose of $\mathbf{y}$) is a matrix of rank 1. Conversely, every rank 1 matrix can be expressed as the matrix product of a nonzero column vector by a nonzero row vector. More generally, an $I \times J$ matrix M has rank R if and only if there exist linearly independent vectors $\mathbf{x}_1, \ldots, \mathbf{x}_R$ in R^I and linearly independent vectors $\mathbf{y}_1, \ldots, \mathbf{y}_R$ in R^J such that

$$M = \sum_{r=1}^{R} \mathbf{x}_r \mathbf{y}_r^T$$

We are now ready to define the tensor product of vectors, which serves as the fundamental basis for subsequent definitions of tensor product of vector spaces and tensor product of linear transformations.

Definition 3 (tensor products of vectors)

Let $\mathbf{x}$ *be a vector with I coordinates given by* $\{x_i\}$ *and* $\mathbf{y}$ *be a vector with J coordinates given by* $\{y_j\}$. *A / The tensor product of* $\mathbf{x}$ *and* $\mathbf{y}$, *denoted* $\mathbf{x} \otimes \mathbf{y}$, *is a vector with IJ coordinates given by* $\{x_i y_j\}$.

The vagueness in Definition 3 concerning the arrangement of the vector coordinates is deliberate. Either of the vectors $\mathbf{x}$ and $\mathbf{y}$ or their tensor product could be represented by a multiway array or a column vector, and the representation, especially for the tensor product, could change from one context to the next.

Examples are:

1. If $\mathbf{x}$ is an $I \times 1$ column vector and $\mathbf{y}$ is a $J \times 1$ column vector, the tensor product defined by

$$\mathbf{x} \otimes \mathbf{y} = \mathbf{x}\mathbf{y}^T$$

is an $I \times J$ matrix. In particular, if $I = 3 = J$ and

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

the tensor product $\mathbf{x}\mathbf{y}^{\mathrm{T}}$ is

$$\begin{pmatrix} x_1y_1 & x_1y_2 & x_1y_3 \\ x_2y_1 & x_2y_2 & x_2y_3 \\ x_3y_1 & x_3y_2 & x_3y_3 \end{pmatrix}$$

2. if $\mathbf{x} \in R^I$ and $\mathbf{y} \in R^J$ as above, the tensor product defined by

$$\mathbf{x} \otimes \mathbf{x} = \left(x_1\mathbf{y}^{\mathrm{T}}, x_2\mathbf{y}^{\mathrm{T}}, \ldots, x_I\mathbf{y}^{\mathrm{T}}\right)^{\mathrm{T}}$$

is an $IJ \times 1$ column vector. Here for the particular case $I = 3 = J$ we have

$$\mathbf{x} \otimes \mathbf{y} = \begin{pmatrix} x_1y_1 \\ x_1y_2 \\ x_1y_3 \\ x_2y_1 \\ x_2y_2 \\ x_2y_3 \\ x_3y_1 \\ x_3y_2 \\ x_3y_3 \end{pmatrix}$$

3. The Kronecker product of matrices is a tensor product. If $\mathbf{x} = (x_{kl})$ is a $K \times L$ matrix and $\mathbf{y} = (y_{mn})$ is an $M \times N$ matrix, the Kronecker product of $\mathbf{x}$ and $\mathbf{y}$ is the $KM \times LN$ matrix defined by

$$\mathbf{x} \otimes \mathbf{y} = \begin{pmatrix} x_{11}\mathbf{y} & \cdots & x_{1L}\mathbf{y} \\ \vdots & \ddots & \vdots \\ x_{K1}\mathbf{y} & \cdots & x_{KL}\mathbf{y} \end{pmatrix} \in \mathcal{M}[KM, LN]$$

In particular if

$$\mathbf{x} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}$$

and

$$\mathbf{y} = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix}$$

then

$$\mathbf{x} \otimes \mathbf{y} = \begin{pmatrix} x_{11}y_{11} & x_{11}y_{12} & x_{12}y_{11} & x_{12}y_{12} \\ x_{11}y_{21} & x_{11}y_{22} & x_{12}y_{21} & x_{12}y_{22} \\ x_{21}y_{11} & x_{21}y_{12} & x_{22}y_{11} & x_{22}y_{12} \\ x_{21}y_{21} & x_{21}y_{22} & x_{22}y_{21} & x_{22}y_{22} \end{pmatrix}$$

Examples 1 and 2 illustrate the ambiguity in the arrangement of the coordinates of a tensor product. Both $\mathbf{x}$ and $\mathbf{y}$ are column vectors in these examples, but their tensor product is an $I \times J$ matrix in Example 1 and an $IJ \times 1$ column vector in Example 2. The difference between these arrangements is, in a sense, trivial because we can readily obtain the column vector from the matrix and vice versa. If we want to distinguish carefully between the arrangements, we can call Example 1 a tensor product of $\mathbf{x}$ and $\mathbf{y}$ and Example 2 another.

Alternatively, we could decide that the particular arrangement is a minor detail that should not be an encumbrane to the concept of tensor product. In that case we may define *the* tensor product as a collection of coordinates that we are free to rearrange at our convenience.

The latter alternative is an example of the abstraction process in which only the essence of a concept that applies to various particular situations is incorporated into the definition of that concept. Details that serve only to distinguish the particular applications and are irrelevant to the abstract concept are stripped from its definition by the abstraction process. Abstractions of this sort are the heart and soul of mathematics and are extremely valuable aids to thought. It is very useful to have the notion of the number 'two' without having to ask whether it refers to two apples, two dollars, or two inches.

Although abstraction is important and potentially very useful, too much too soon can be a pedagogical mistake. To appreciate fully the value of an abstract concept, the student should be allowed to participate in the abstraction process. If an uninitiated student is presented at the outset with a rigorous, highly abstracted definition, he or she may find the concept impossible to appreciate. That, in this author's opinion, is a major problem with much of the extant literature on tensors and helps to explain why this potentially useful subject is not more widely used in practice. Avoidance of the pedagogical trap of too much abstraction too soon is an important objective of this Tutorial.

The following properties of tensor products of vectors are worth emphasizing:

Extendability. The definition is easily extended to tensor products of three or more vectors. For $\mathbf{x} \in R^I$, $\mathbf{y} \in R^J$, $\mathbf{z} \in R^K$, the tensor product $\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}$ is a vector with coordinates $x_i y_j z_k$.

Independence. Tensor products preserve linear independence. In other words, if the R vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_R\}$ are linearly independent, and the S vectors $\{\mathbf{y}_1, \ldots, \mathbf{y}_S\}$ are linearly independent, then the RS vectors $\{\mathbf{x}_r \otimes \mathbf{y}_s\}$ will also be linearly independent. The proof of this result is straightforward but somewhat tedious, and will be omitted for the sake of brevity.

Nonclosure. In general linear combinations of tensor products are not themselves tensor products. In particular if $\mathbf{x}_1$ is linearly independent of $\mathbf{x}_2$ and $\mathbf{y}_1$ is linearly independent of $\mathbf{y}_2$, it is impossible to find vectors $\mathbf{u}$ and $\mathbf{v}$ for which

$$\mathbf{x}_1 \otimes \mathbf{y}_1 + \mathbf{x}_2 \otimes \mathbf{y}_2 = \mathbf{u} \otimes \mathbf{v} \tag{1}$$

The impossibility of satisfying (1) is easy to see from Example 1 where tensor products are rank 1 matrices. Under the stipulated conditions of linear independence the sum of two rank 1 matrices will have rank equal to 2 and cannot therefore be a tensor product of two vectors. In formal jargon the tensor product of vectors is a bilinear map whose range is not closed under vector addition.

By the nonclosure property not every $I \times J$ matrix is a tensor product, i.e. a rank 1 matrix. However, every matrix in $\mathcal{M}[I, J]$ can be expressed as a sum of rank 1 matrices. This feature forms the basis for a definition of a/the tensor product of vector spaces.

Definition 4 (tensor products of vector spaces)
Suppose a tensor product $\mathbf{x} \otimes \mathbf{y}$ *has been defined for* $\mathbf{x} \in \mathcal{U}$ *and* $\mathbf{y} \in \mathcal{V}$, *where* $\mathcal{U} \subset R^I$ *and* $\mathcal{V} \subset R^J$ *vector spaces. The tensor product of* $\mathcal{U}$ *and* $\mathcal{V}$, *denoted by* $\mathcal{U} \otimes \mathcal{V}$, *is defined to be the vector space consisting of all linear combinations of vectors of the form* $\mathbf{x} \otimes \mathbf{y}$, *where* $\mathbf{x} \in \mathcal{U}$ *and* $\mathbf{y} \in \mathcal{V}$. *Continuing our examples yields*:

1. $R^I \otimes R^J = \mathcal{M}[I, J]$. *If* A *is an* $I \times J$ *matrix with* $\mathcal{U}$ *and* $\mathcal{V}$ *as its column and row spaces, respectively, then* $A \in \mathcal{U} \otimes \mathcal{V}$.
2. $R^I \otimes R^J = R^{IJ}$
3. $\mathcal{M}[K, L] \otimes \mathcal{M}[M, N] = \mathcal{M}[KM, LN]$

Some important properties of tensor products of vector spaces are:

Extendability. The extension to tensor products of more than two vector spaces is straightforward.

Dimensionality. If $\{\mathbf{x}_1, \ldots, \mathbf{x}_R\}$ is a basis for $\mathcal{U}$ and $\{\mathbf{y}_1, \ldots, \mathbf{y}_S\}$ is a basis for $\mathcal{V}$, then preservation of linear independence implies that the RS vectors $\{\mathbf{x}_r \otimes \mathbf{y}_s\}$ are a basis for $\mathcal{U} \otimes \mathcal{V}$. Thus, if $\dim(\mathcal{U}) = R$ and $\dim(\mathcal{V}) = S$, it follows that $\dim(\mathcal{U} \otimes \mathcal{V}) = RS$.

Factorability. The tensor product $\mathcal{U} \otimes \mathcal{V}$ is a linear subspace of $R^I \otimes R^J$, but not every linear subspace of $R^I \otimes R^J$ can be factored as $\mathcal{U} \otimes \mathcal{V}$ with $\mathcal{U} \subset R^I$ and $\mathcal{V} \subset R^J$. This property is analogous to the nonclosure property, since not every matrix can be expressed as the tensor product of vectors.

A tensor product of vector spaces is a vector space with extra structure. This structure permits it to be factored into a product of lower dimensional vector spaces. A *tensor* is just a vector in a tensor product vector space. A natural consequence of the tensor product structure is the use of multiple subscripts for the coordinates of a tensor. When we write $\mathbf{x} \in R^4 \otimes R^3$ instead of $\mathbf{y} \in R^{12}$, it is natural to use (y_{ij}) for $i = 1, 2, 3, 4$ and $j = 1, 2, 3$ as the coordinates rather than (y_i) for $i = 1, \ldots, 12$.

Finally, we define the tensor product of linear transformations.

Definition 5 (tensor products of linear transformations)
Let P *be a linear transformation from* R^I *to* R^K *and let* Q *be a linear transformation from* R^J *to* R^L. *The tensor product of* P *and* Q *is the linear transformation from* $R^I \otimes R^J$ *to* $R^K \otimes R^L$ *defined by*

$$(P \otimes Q)[\mathbf{x} \otimes \mathbf{y}] = P\mathbf{x} \otimes Q\mathbf{y} \tag{2}$$

where $\mathbf{x}$ *and* $\mathbf{y}$ *are arbitrary vectors in* R^I *and* R^J, *respectively. Notice that Expression (2) defines* $(P \otimes Q)[\mathbf{w}]$ *explicitly only for those vectors* $\mathbf{w} \in R^I \otimes R^J$ *which happen to be tensor products. If instead* $\mathbf{w} = \sum_{r=1}^{R} (\mathbf{x}_r \otimes \mathbf{y}_r)$, *the linearity of* $P \otimes Q$ *implies that*

$$(P \otimes Q)[\mathbf{w}] = \sum_{r=1}^{R} (P\mathbf{x}_r \otimes Q\mathbf{y}_r)$$

Turning to our examples, we have:

1. If $\mathbf{x} \otimes \mathbf{y}$ is the $I \times J$ matrix $\mathbf{x}\mathbf{y}^{\mathrm{T}}$ and $P\mathbf{x} \otimes Q\mathbf{y}$ is the $K \otimes L$ matrix $(P\mathbf{x})(Q\mathbf{y})^{\mathrm{T}}$, then for any matrix A in $\mathcal{M}[I, J]$

$$(P \otimes Q)[A] = PAQ^T$$

2. If the tensor product of $\mathbf{x}$ and $\mathbf{y}$ and the tensor product of $P\mathbf{x}$ and $Q\mathbf{y}$ are both Kronecker products as in Example 2, then $P \otimes Q$ is a linear transformation from R^{IJ} to R^{KL}. As such, it has a representation as a $KL \times IJ$ matrix, which in fact is the Kronecker product of the $K \times I$ and $L \times J$ matrices associated with P and Q, respectively. Symbolically, we have

$$P\mathbf{x} \otimes Q\mathbf{y} = (P \otimes Q)[\mathbf{x} \otimes \mathbf{y}]$$

where all tensor products are interpreted as Kronecker products. The demonstration of this result for Kronecker products is straightforward but tedious and will be omitted for the sake of brevity.

Some important properties of tensor products of linear transformations are:

Extendability. The definition can be easily extended to tensor products of three or more linear transformations.

Composition. Using juxtaposition to denote composition of linear transformations or, when appropriate, matrix multiplication, we have

$$(P_1 \otimes Q_1)(P_2 \otimes Q_2) = P_1P_2 \otimes Q_1Q_2$$

Factorability. The tensor product $P \otimes Q$ is a linear transformation from $R^I \otimes R^J$ to $R^K \otimes R^L$, but not every linear transformation from $R^I \otimes R^J$ to $R^K \otimes R^L$ can be factored as the tensor product of a linear transformation from R^K to R^I and a linear transformation from R^L to R^J.

The tensor products of vectors, vector spaces, and linear transformations are concepts that can be usefully applied in many areas of multivariate statistics as well as multiway analysis of variance. It is beyond the scope of this Tutorial to cover all the potential uses of tensor algebra in statistical data analysis. Instead we focus multilinear models to illustrate applications of the tensor product concepts we have introduced.

3. Multilinear models

In this section we describe the specification and fitting of multilinear models for two-way and three-way data arrays using the ideas of tensor algebra. Specifying a model for an array involves identifying special characteristics that the array is to have. Fitting a model to a particular data array involves finding another array which has the specified characteristics and best represents the data.

For example, in fluorescence spectroscopy it may be desired to find a rank R matrix which best approximates an $I \times J$ excitation–emission matrix. Specifying the rank to be R is an example of a bilinear model specification. Other natural ways that come to mind of specifying and fitting bilinear models turn out to be equivalent. Fitting R principal components to an observations by variables data matrix, for example, is equivalent to finding the best rank R approximation to the matrix after its columns have been centered.

The situation for trilinear models is more complicated. Different approaches that would in the end coincide for two-way data lead to the quite distinct PARAFAC and Tucker models when applied to three-way data. Tensor algebra can help to explain the greater complexity and the associated greater opportunities for obtaining useful results afforded by trilinear models.

3.1. Bilinear models

In this section we present two equivalent definitions of the main bilinear model. This approach is taken to facilitate comparison with trilinear models. Each bilinear model definition has a natural extension to three-way data, but the two definitions of trilinear models are not equivalent.

Each model under discussion represents the data array as the sum of a signal array and a noise array, i.e.

$$A = S + N$$

where the signal array S is assumed to have some special structure. The nature of the special structure is what characterizes the model. Different structures lead to different models. In the bilinear case there are two alternative model defini-

tions which are equivalent because the structures they describe are identical. The first represents the signal matrix as a sum of rank one terms [SOROT].

Definition 6 (SOROT)
The R-component SOROT model for an $I \times J$ data matrix A assumes that the signal matrix S can be written as the sum of R rank one matrices. In other words, there exist vectors $\{\mathbf{x}_r\}$ in R^I and $\{\mathbf{y}_r\}$ in R^J such that

$$S = \sum_{r=1}^{R} \mathbf{x}_r \otimes \mathbf{y}_r$$

The rank of S will be less than or equal to R with equality holding if both the $\{\mathbf{x}_r\}$ and $\{\mathbf{y}_r\}$ are linearly independent. Conversely, if S has rank R or less, then S satisfies the R-component SOROT model.

The special structure required of S by the SOROT model is that it have rank R (or less). Since rank is by definition the dimension of the column (or row) space of a matrix, we have the following alternative model definition:

Definition 7 (TPS)
For an $I \times J$ data array the tensor product subspace (TPS) model with dimension parameters R and Q assumes that there exist linear subspaces $\mathscr{U} \subset R^I$ and $\mathscr{V} \subset R^J$ with $\dim(\mathscr{U}) = R$ and $\dim(\mathscr{V}) = Q$ such that

$$A = S + N$$

and

$$S \in \mathscr{U} \otimes \mathscr{V}$$

The use of two dimension parameters in the definition of the two-way TPS model anticipates the three-way TPS model to be defined in the next section. One dimension parameter would be enough for two-way data. If the rank of S is R then $S \in \mathscr{U} \otimes \mathscr{V}$ where $\mathscr{U}$ is the R-dimensional column space of S and $\mathscr{V}$ is the R-dimensional row space of S. If $\mathscr{W} \supset \mathscr{V}$ and $\dim(\mathscr{W}) = Q > R$, then also

$$S \in \mathscr{U} \otimes \mathscr{W}$$

but the increased dimensionality for $\mathscr{W}$ does not increase the set of possibilities for the signal matrix S. For three-way data, however, the story is different.

To avoid the too much abstraction too soon syndrome, we turn to fluorescence spectroscopy for an example of the SOROT model. If fluorescence emission intensity from a sample is measured at each of J wavelengths when the sample is stimulated at each of I excitation wavelengths, the resulting $I \times J$ data matrix A is called an excitation–emission matrix (EEM). Each fluorescent component in the sample has an excitation spectrum $\mathbf{x}$ whose coordinates x_i reflect its propensity to absorb energy at the ith excitation wavelength and an emission spectrum $\mathbf{y}$ whose coordinates y_j reflect its propensity to emit fluorescent light at the jth emission wavelength. If there are R components in the sample and their concentrations c_r are not too large, they contribute additively to the signal EEM, yielding

$$S = \sum_{r=1}^{R} c_r \mathbf{x}_r \mathbf{y}_r^{\mathrm{T}} \tag{3}$$

Factoring the c_r into either $\mathbf{x}$ or $\mathbf{y}$ converts (3) into the defining expression for the SOROT model.

The SOROT model can be used to estimate the number of fluorescent components in the sample, at least when the signal-to-noise is sufficiently high. Linear independence of the constituent spectra in an R component sample implies that the EEM would have rank R in the absence of noise. If the noise is not too large, we would expect a good fit of the data to the R-component model, but not to the $(R-1)$-component model.

If we could also extract the individual rank 1 terms in (3), we could estimate the excitation and emission spectra of the individual components, which might enable us to identify them. Unfortunately, this is not generally possible because the representation of a rank R matrix as the sum of rank 1 matrices is highly nonunique whenever $R \geqslant 2$. In statistical jargon the coordinates of the excitation and emission spectra are nonidentifiable parameters of the model.

It is instructive to explore further the nature of the nonidentifiability of the spectral resolution

parameters. The rank R of an $I \times J$ matrix S is also the dimension of its column space. Let $\{\mathbf{x}_r\}$ be an arbitrary basis for the column space of S. Then each column of S can be expressed as a linear combination of the $\mathbf{x}_r$. Let $\{y_{jr}\}$ denote the coefficients of these linear combinations so that the jth column of S is given by

$$\sum_{r=1}^{R} y_{jr}\mathbf{x}_r \tag{4}$$

If $\mathbf{y}_r$ is the vector in R^J with coordinates $\{y_{jr}\}$, it follows from (4) that

$$S = \sum_{r=1}^{R} \mathbf{x}_r \otimes \mathbf{y}_r \tag{5}$$

In other words there is a distinct SOROT representation for S corresponding to each choice of a basis for its column space. A similar relationship exists between SOROT representations and bases for the row space of S.

The nonuniqueness associated with an arbitrary choice of basis occurs fairly often. The rotation problem is the common name for this type of nonuniqueness.

3.2. Trilinear models

The SOROT and TPS model definitions can be readily extended to the threeway case.

Definition 8 (PARAFAC)
The R-component SOROT model for an $I \times I \times K$ data array A assumes that there exist vectors $\{\mathbf{x}_r\} \subset R^I$, $\{\mathbf{y}_r\} \subset R^J$, and $\{\mathbf{z}_r\} \subset R^K$ such that

$$S = \sum_{r=1}^{R} \mathbf{x}_r \otimes \mathbf{y}_r \otimes \mathbf{z}_r$$

where S is the signal array for the model. The three-way SOROT model is commonly called the PARAFAC model [1] *and that term will be used in lieu of three-way SOROT for the rest of this Tutorial.*

Definition 9 (Tucker3)
For an $I \times J \times K$ array A the tensor product subspace (TPS) model with dimension parameters R, S, T assumes there exist linear subspaces $\mathscr{U} \subset R^I$, $\mathscr{V} \subset R^J$, and $\mathscr{W} \subset R^K$ with $dim(\mathscr{U}) = R$, $dim(\mathscr{V}) = S$, and $dim(\mathscr{W}) = T$ such that the tensor product subspace $\mathscr{U} \otimes \mathscr{V} \otimes \mathscr{W}$ contains the signal array for A. The three-way TPS model is also called the Tucker3 or T3 model.

The extensions of the SOROT and TPS definitions to the three-way case are straightforward and natural, but there are striking differences between the trilinear and bilinear versions. One is that PARAFAC and Tucker3 are distinct models, while the SOROT and TPS bilinear models are equivalent. Another important difference is the fact that under quite general conditions the PARAFAC resolution is unique, apart from trivial reindexing and rescaling [2]. This stands in stark contrast to the bilinear SOROT model, where the expression of a rank R matrix as the sum of R rank 1 matrices is highly nonunique.

The uniqueness of the PARAFAC resolution has important applications in analytical chemistry. Returning to fluorescence spectroscopy, suppose there are two samples, the first containing R components at known concentrations $\{c_{1r}\}$ and the second containing those same components at unknown concentrations $\{c_{2r}\}$. Taking EEMs for both samples yields an $I \times J \times 2$ array and the PARAFAC model with the mode Z vectors given by $\mathbf{z}_r = (ac_{1r}, ac_{2r})^T$. The factor a is included because of the scale indeterminacy in the resolution, but it does not affect the ratio $z_{2r}/z_{1r} = c_{2r}/c_{1r}$. Thus, the ratios of the unknown concentrations to the known concentrations can be extracted by fitting an R-component PARAFAC model to the $I \times J \times 2$ data array. The generalized rank annihilation method [3] is a data analytic method for fitting the PARAFAC model in order to obtain this type of calibration.

Using new samples at different concentrations is not the only way to build excitation–emission data into a trilinear three-way array. Unique resolutions have been successfully obtained for a single sample by fitting the PARAFAC model to three-way data by the technique of phase-resolved fluorescence spectroscopy [4,5]. Examples of resolutions obtained by fitting the PARAFAC model to other types of three-way data in chemistry can also be found in the literature [6,7].

The PARAFAC model forms the basis for the definition of rank for three-way arrays.

Definition 10 (rank)
The rank of a three-way array A is defined to be the smallest value such that the R-component PARAFAC model fits A exactly.

This definition extends the concept of matrix rank to three-way arrays. A similar definition, based on the SOROT model, of matrix rank would agree with the definition of rank as the dimension of the column (or row) space.

Rows and columns are subarrays of a matrix which are quite useful for describing its properties. Extending these concepts to three-way arrays is also quite useful but somewhat more complicated than one might at first expect.

Definition 11 (fibers and slabs)
Let A be an $I \times J \times K$ array with coordinates a_{ijk}. The Mode X fibers of A are vectors $\mathbf{a}_{jk}$ in R^I obtained from A by fixing the index j and k and letting the index i vary. The Mode X slabs of A are the $J \times K$ matrices obtained as subarrays of A by fixing the i index and letting the j and k indices vary. The Mode Y (or Mode Z) fibers and slabs are similarly defined as subarrays obtained by letting j (or k) be the only varying or the only fixed index. The fiber spaces of A are the subspaces spanned by its fibers e.g. the Mode X fiber space of A is the subspace of R^I spanned by the Mode X fibers of A.

If $A \in \mathscr{U} \otimes \mathscr{V} \otimes \mathscr{W}$, then the Mode X, Mode Y, and Mode Z fiber spaces of A must be subspaces of $\mathscr{U}$, $\mathscr{V}$, and $\mathscr{W}$, respectively. An $I \times J \times K$ array can be arranged as an $I \times JK$ matrix, in which case its columns are the Mode X fibers. Similar arrangements are the $J \times IK$ matrix with the Mode Y fibers as its columns and the $K \times IJ$ matrix with the Mode Z fibers as its columns. This process of representing a three-way array as a matrix is sometimes called unfolding.

We conclude by demonstrating how tensor concepts applied to the unfolded three-way array can help to show why trilinear models differ so markedly from their bilinear analogs. Let $\mathscr{X}$ be the Mode X fiber space for an $I \times J \times K$ array and suppose $\dim(\mathscr{X}) = R$. Let $\{\mathbf{x}_r\}$ be a basis for $\mathscr{X}$ and let A be represented in its unfolded $I \times JK$ form. As in (5) we can find vectors $\{\mathbf{w}_r\}$ in R^{JK} such that

$$A = \sum_{r=1}^{R} \mathbf{x}_r \otimes \mathbf{w}_r \tag{6}$$

Expression (6) gives an R-component bilinear SOROT representation of A as an $I \times JK$ matrix, but it is not a trilinear PARAFAC representation. Even though each $\mathbf{w}_r$ is in $R^{JK} = R^J \otimes R^K$, the vectors $\mathbf{w}_r$ are not necessarily themselves tensor products. If each of them were, (6) would indeed be an R-component PARAFAC representation of A.

This analysis demonstrates that the rank of a three-way array must be at least as great as the dimension of its Mode X fiber space. In fact, writing Y and $\mathscr{Z}$ for the Mode Y and Mode Z fiber spaces, respectively, we have that

$$\operatorname{rank}(A) \geq \max\{\dim(\mathscr{X}), \dim(\mathscr{Y}), \dim(\mathscr{Z})\}$$

This brief introduction to bilinear and trilinear models provides only a sample of the uses of tensor algebra and notation. For more about tensor algebra in the context of multilinear models, see Refs. [8,9]. A discussion of multilinear models that does not explicitly involve tensor terminology is in Ref. [10]. For an extensive survey of the literature on trilinear models up to 1983, including many psychometric applications, see Ref. [11].

4. Conclusion

The value of tensor algebra to the practicing data analyst can be compared to the value of matrix algebra. Either subject requires some effort to learn and neither is necessary if one is willing to use formulas with lots of subscripts and summation signs. However, the rewards that accrue from learning these subjects are substantial. A valuable economy of thought results from concepts and notation that disencumber the tedious details from key ideas. The effort spent to learn matrix algebra is rarely regretted.

Experience with tensor algebra has been less positive. It is not uncommon for someone to begin a study of tensor algebra and then quit in

discouragement when he or she encounters the obstacles of too much abstraction too soon or applications to physics that seem irrelevant to data analysis of multilinear models. Under these circumstances the effort to achieve the rewards that result from learning tensor algebra has been stymied by the obstacles along the way. These obstacles need not loom so large. Tensor algebra does not have to be a difficult esoteric subject for data analysts.

This Tutorial introduces basic concepts of tensor algebra in a way that should be accessible to most readers. It also gives a brief introduction to bilinear and trilinear models as an illustration of the utility of these concepts. There are many more things that could be said about multilinear models. The point is that the concepts and notation of tensor algebra provide a convenient unifying framework in which to say them.

References

[1] R.A. Harshman and M.E. Lundy, The PARAFAC model for three-way factor analysis and multidimensional scaling, in H.G. Law, C.W. Snyder, Jr., J.P. Hattie and R.P. McDonald (Editors), *Research Methods for Multilinear Data Analysis*, Praeger, 1984, pp. 122–215.

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

[3] E. Sanchez and B.R. Kowalski, Tensorial calibration: II. Second-order calibration, *Journal of Chemometrics*, 2 (1988) 265–280.

[4] D.S. Burdick, X.M. Tu, L.B. McGown and D.W. Millican, Resolution of multicomponent fluorescent mixtures by analysis of the excitation–emissionfrequency array, *Journal of Chemometrics*, 4 (1990) 15–28.

[5] D.W. Millican and L.B. McGown, Fluorescence lifetime resolution of spectra in the frequency domain using multiway analysis, *Analytical Chemistry*, 62 (1990) 2242–2247.

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

[7] R.T. Ross, C.-H. Lee, C.M. Davis, E.A. Fayyad and S.E. Leurgans, Resolution of fluorescence spectra of plant pigment-complexes using trilinear models, *Biochimica et Biophysica Acta*, 1056 (1991) 317–320.

[8] A. Franc, Multiway matrices: some algebraic results, in R. Coppi and S. Bolasco (Editors), *Multiway Data Analysis*, Elsevier, Amsterdam, 1989, pp. 19–30.

[9] A. Franc, Étude algebrique des multitableaux: Apports de l'algèbre tensorielle (An algebraic study of multi-way tables: Contributions of tensor algebra), Ph.D. thesis, Université de Montpellier II, France, 1992.

[10] S.E. Leurgans and R.T. Ross, Multilinear models: applications in spectroscopy, *Statistical Science*, 3 (1992) 289–319.

[11] P.M. Kroonenberg, Annotated bibliography of three-mode factor analysis, *British Journal of Mathematical and Statistical Psychology*, 36 (1983) 81–113.