Multilinear Models: Applications in Spectroscopy

Sue Leurgans and Robert T. Ross

Abstract. Multilinear models are models in which the expectation of a multiway array is the sum of products of parameters, where each parameter is associated with only one of the ways. In spectroscopy, multilinear models permit mathematical decompositions of data sets when chemical decomposition of specimens is difficult or impossible. This paper presents a unified description of the models in an array notation. The spectroscopic context shows how to interpret one initialization of the nonlinear least-squares fits of these models. Several examples show that these models can be applied successfully.

Key words and phrases: Multi-mode factor analysis, nonlinear least-squares, PARAFAC, three-way arrays.

1. INTRODUCTION

We begin with an example. Figure 1 is a graph of light absorption by pea leaves as a function of the wavelength of the light. This light absorption is the sum of the absorption by several different assemblies of pigments, which are responsible for the capture of light and its conversion into chemical energy. Accurate information about light absorption by the individual assemblies is important in understanding photosynthesis, but chemical separation of the assemblies would alter the properties studied. This and related problems motivate the biophysicist's interest in mathematical modeling of light absorption data.

Light absorption and emission can often be measured under various conditions corresponding to different combinations of levels of several experimental variables such as wavelength and concentration. It is convenient to record the responses by using one subscript to denote the level of each independent variable and then to arrange the observed values in a multiway array, to be denoted Y below. If the experimental variables are wavelength and concentration, then μ, the array of expected values of the elements of Y, is a multilinear function of the wavelength and concentration parameters of the pigment assemblies. (See Sections 2 and 3 for definitions and explanations.) Although multilinear functions are nonlinear functions of the several sets of parameters (one set of parameters for each experimental variable) and although nonlinear regression models are generally intractable, some properties of multilin-

ear models are tractable if there are three or more experimental variables.

When two experimental variables are used, multilinear models are bilinear models, and the underlying spectra cannot be deduced from μ unless only one pigment assembly is absorbing light. However, if three or more experimental variables are used, then the underlying spectra for each experimental variable can generally be deduced from μ. (See Section 4.)

Statisticians have long employed bilinear models. The relevance of spectroscopy to bilinear models is implicitly recognized when the characteristic values of a matrix are called the "spectrum" of the matrix. Bellman (1970) attributes the term to Hilbert. Multilinear models are attracting more attention, and we hope that our description of multilinear models in spectroscopy will be another illustration of the advantages of multilinear models. This specific context, in which the models and their interpretations are well established, may assist readers, as it has us, when they think about arrays with more than two ways.

In Section 2, we introduce spectroscopy. In Section 3, we present multilinear models and show that several spectroscopic formulas imply multilinear models. The parameters of bilinear models are not identifiable, but the parameters of some of the trilinear models of Subsection 3.2 are identifiable. In Section 4, we review several methods of estimation. Section 4.1 is an intuitive derivation of an explicit decomposition for some trilinear models, a decomposition that represents the identifiability remarks in 3.2. Section 5 contains several examples.

Sue Leurgans is Associate Professor, Department of Statistics, Ohio State University, 1958 Neil Avenue, Columbus, Ohio 43210. Robert T. Ross is Professor, Department of Biochemistry, Ohio State University, 484 W. 12th Street, Columbus, Ohio 43210.
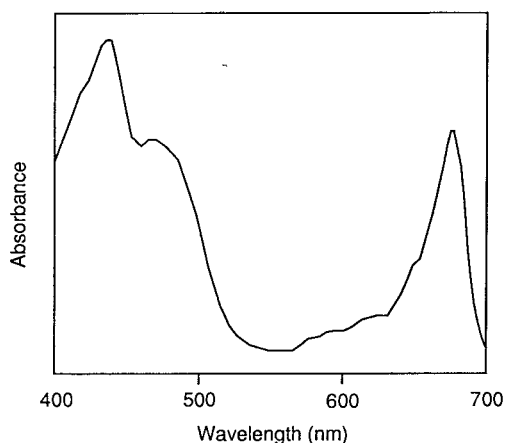
2. INTRODUCTION TO SPECTROSCOPY

In this section, we use the symbols and nomenclature used by chemists whenever possible. For further back-

FIG. 1. *The absorption of light by pea leaves as a function of wavelength.*

ground from the perspective of analytical chemistry, see Ingle and Crouch (1988). For a more mathematical presentation and applications to biophysics, see Cantor and Schimmel (1980). For a review of the use of multiple experimental variables, see Ndou and Warner (1991).

Spectroscopy is the measurement of the absorption of particles by a specimen, or the emission of particles from a specimen, as a function of the energy of the particles. The amount of absorption or emission as a function of particle energy is known as a spectrum. Information from the many kinds of spectroscopy used by physicists and chemists underlies most of our understanding of the structure of atoms and molecules.

Common spectroscopy uses electromagnetic radiation such as visible light. The particles are photons. The wavelength of the radiation is inversely proportional to photon energy, and wavelength is often used instead of energy as the independent variable. Different kinds of chemical spectroscopy use wavelengths ranging from 10^{-10} to 10^{+4} meters. The examples of spectroscopy presented in this article all involve electromagnetic radiation in the visible and near-visible range of wavelengths, from 10^{-7} to 10^{-6} m.

A typical specimen contains a very large number (10^{13} to 10^{16}) of each of several distinguishable light-absorbing entities, or *chromophores*. The light-absorption process may be almost entirely confined to a small part of a single molecule, and it is this part that is regarded as a chromophore. A large molecule may have several chromophores. For example, a small protein may be composed of 100 chemically linked amino acids; the protein is a single molecule, but for many purposes, parts of the individual amino acids will behave as independent chromophores.

When a beam of particles, such as photons, is directed at a specimen, the increment in intensity (dI) absorbed with an increment in distance (dx) is

(1) $dI = -\sigma n I_x dx$

where σs is the absorption cross-section (cm^2), n is the number density (per cm^3) of the chromophores and I_x is the intensity of the beam at distance x into the specimen.

In common applications of spectroscopy in chemistry, the specimen is a uniform solution of chromophores held in a tube of width L (usually 1 cm). Integration of dI/I in (1) with some unit changes gives what chemists call Beer's law:

(2) $A[\lambda] = -\log_{10}(I_L/I_0) = \varepsilon[\lambda]cL$

where A is known as the *absorbance*, I_L/I_0 is the fraction of the intensity of a beam of light of wavelength λ that passes through the specimen, and c is the concentration of chromophores. Figure 1 is a plot of the absorbance due to the pigments in pea leaves. The fundamental property of the chromophore is $\varepsilon[\lambda]$, the *extinction coefficient*, a quantity with units of (length $\times$ concentration)$^{-1}$ which is equivalent to absorption cross-section.

While the primary independent variable of spectroscopy is the energy or wavelength of the particles absorbed or emitted, an experiment may involve additional independent variables. With multiple chromophores f and wavelengths i and differing circumstances j in which concentrations of the chromophores vary, we have the bilinear equation

(3) $A[i,j] = \sum_f \varepsilon_f[i]c_f[j]L,$

where $\varepsilon_f[i]$ is the extinction coefficient of chromophore f at wavelength λ_i and $c_f[j]$ is the concentration of f in circumstance j. We will see in Subsection 3.1 that the parameters $\varepsilon_f[i]$ and $c_f[j]$ are not identifiable without reliable side conditions that restrict the parameter space. There are over a hundred publications on the bilinear analysis of absorbance as a function of wavelength and of some variable which affects the relative concentration of the chromophores in the specimen. Side conditions used include the positivity of concentrations and of extinction coefficients. The early work on this and other chemical applications of bilinear models (often called factor analysis by chemists) is reviewed in Malinowski and Howery (1980); the more recent work is surveyed in biannual reviews of the chemometric literature (Ramos et al., 1986; Brown et al., 1988; Brown, 1990).

The absorption of a photon puts a chromophore into a higher energy excited state. This excited state then loses its extra energy and decays to the ground state, usually in about 10^{-9} sec. Some decay can be accompanied by light emission, usually of a kind called fluorescence. Chromophores that emit a detectable amount of fluorescence are called *fluorophores*.

In fluorescence spectroscopy, the specimen is illumi-

nated with light of wavelength λ^{ex}, and the consequent emission of light is measured at wavelength λ^{em}. The sequence of events that occur during such an experiment is diagrammed in Figure 2. Incident light of wavelength λ^{ex} is absorbed by the chromophore, causing the chromophore to gain an amount of energy equal to the photon energy. Most of the additional energy is taken up by a single electron that moves from one orbital (probability pattern in space) to another that has a higher average distance from the positively-charged atomic nuclei and therefore has a higher energy. Because an electron has moved, the most stable geometry of the atomic nuclei in the excited state is somewhat different from that in the ground state. However, the absorption event ocurs so rapidly $(10^{-15}$ sec) that the nuclei, which are far more massive than electrons, do not have enough time to move. Following the electronic transition, the nuclei shift their positions to the more stable geometry in about 10^{-12} sec. As the nuclei shift, the properties of the excited state become independent of how it was created, so that the relative probability of any subsequent event is independent of how the excited state was created.

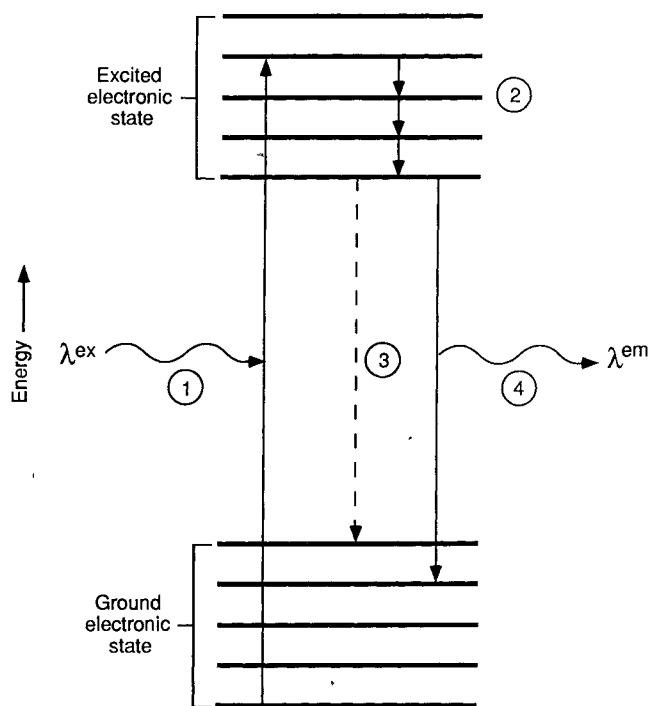The amount of light emission measured is separately linear in the number of photons absorbed and in the fraction of photons absorbed that lead to emission at wavelength λ^{em}. With multiple chromophores, we then have the trilinear equation

$$(4) \qquad \mu[i,j,k] = \sum_f \varepsilon_f[i]\pi_f[j]c_f[k]$$

where $c_f[k]$ is the concentration of chromophore f in circumstance k, $\varepsilon_f[i]$ is the relative absorption cross-section of chromophore f at wavelength λ_i^{ex} and $\pi_f[j]$ is the relative emission at the detection wavelength λ_j^{em}. We will see that equation (4) is the PARAFAC model (23).

Simple application of equation (4) requires that sample absorbance be small and that excitation not be transferred between chromophores (e.g., that absorption of a photon by chromophore 1 not lead to emission by chromophore 2); however, nonlinearity due to absorbance can be corrected, and energy transfer can be treated with a different multilinear model. [See (7) below.] The first application of a trilinear model to spectroscopy was by Appellof and Davidson (1981), who applied equation (4), with elution from a chromatography column serving to vary concentration.

There are two ways to expand equation (4) to obtain an additional independent variable. An expansion is especially important if one of the three variables in equation (4) is ineffective. For example, the relative concentrations of different fluorophores may be fixed, perhaps because they are all part of the same large molecule.

The first method of obtaining an additional independent variable is to treat the specimen in some manner that differentially affects the total amount of light emission from different fluorophores. Of all excitations, only some fraction lead to light emission. This fraction, φ, is the *fluorescence quantum yield*. The other excitations are lost via nonradiative decay. The experimenter can alter the fluorescence quantum yield by changing the rate of nonradiative decay. Most often, one decreases the fluorescence by providing an additional decay pathway. A chemical added to accomplish this is known as a *fluorescence quencher*. As long as the mean lifetime of the excited state remains much longer than the "forgetting" time of thermal equilibration, the amount of quenching will have no effect on the distribution of emission wavelengths. We then have the quadrilinear model

$$(5) \qquad \mu[i,j,k,l] = \sum_f c_f[i]\varepsilon_f[j]\varphi_f[k]\pi_f[l],$$

where $\varphi_f[k]$ is the fluorescence quantum yield of fluorophore f when subjected to treatment k and where c, ε and π are as defined previously. If relative concentration is fixed, (5) reduces to a trilinear equation. Some examples of our use of this equation are presented in Section 5.

All of the measurements described so far can be obtained using steady illumination of the specimen and

Fig. 2. *Events leading to fluorescence.* 1: *Energy from light of wavelength λ^{ex} moves an electron to an excited state.* 2: *Atomic nuclei move to a more stable geometry.* 3: *With probability $(1 - \varphi)$, the electron returns to the ground state without emitting light.* 4: *With probability φ, the electron returns to the ground state emitting light distributed over wavelengths λ^{em}.*

detection of light that ignores time as a variable. The second method of obtaining an additional variable requires measuring the rate of decay of the excited state. In one form of time-resolved spectroscopy, the specimen is excited at time zero by a brief intense flash of light, following which the light emitted by a particular category of fluorophore decays as some function $h(t)$. In simple cases, $h(t)$ is a negative exponential function. For a single chromophore, the intensity is the product of five terms

$$(6) \qquad \mu[i,j,k,l,m] = \varepsilon[i]\pi[j]c[k]\varphi[l]h[m],$$

where ε, π, c and φ are as in previous equations, and h is the relative population of the excited state at time m. If concentration and quantum yield are constant, (6) reduces to a trilinear equation, which was first applied by Russell, Gouterman and van Zee (1988) and Russell and Gouterman (1988a, b).

The other form of time-resolved emission spectroscopy uses the equivalence of information from the frequency domain. The exciting light is modulated at different frequencies. Consequently, the emitted light is modulated, but to a reduced degree and with a phase lag, due to the time delay between excitation and emission. Burdick et al. (1990) applied a trilinear model to phase-resolved fluorescence measurements.

Up to this point, we have not considered the possibility of any effect by one chromophore on another. However, in biological specimens it is quite common for the following sequence of events to occur: (a) light absorption by chromophore f_1 with a probability equal to A_{f_1}; (b) given light absorption by chromophore f_1, transfer of excitation and emission from chromophore f_2 with probability $Q_{f_1,f_2}[k]$ under condition k; (c) given emission from f_2, emission distribution $\pi_{f_2}[j]$. Thus

$$(7) \qquad \mu[i,j,k] = \sum_{f_1} \sum_{f_2} A_{f_1}[i]Q_{f_1,f_2}[k]\pi_{f_2}[j]$$

We will see in the next section that (7) is the T2 model, (26). Since the parameter matrices of a T2 model cannot be identified, implementing mathematical analysis of specimens with energy transfer requires that sensible side conditions be specified and imposed.

3. MULTILINEAR MODELS AND ARRAY FORMULAS

In this section, we present multilinear models and introduce convenient abstract notation that emphasizes the array structure. The abstract notation aids clearer understanding, because dummy summations are omitted. In Subsection 3.1, we discuss the familiar bilinear models and rewrite the model several times for comparison with the higher-way formulas developed in Subsections 3.2 and 3.3. In Subsection 3.2, we repeat the bilinear development in the 3-way setting, and we argue that some of the familiar matrix properties do not extend to 3-way arrays. The final subsection is

formulation of multilinear models for arrays with an arbitrary number of indices. Multilinear models from Section 2 are identified.

The following terminology and notation will be used in the rest of the paper. An *N-way array*, or an *array of order* N with dimensions $(I_1, \ldots, I_N)$ is a set of $\bar{I}_N = \Pi_{n=1}^N I_n$ numbers arranged in an $I_1 \times \cdots \times I_N$ table. If i_n is an integer between 1 and I_n (inclusive) for each $n = 1, \ldots, N$, the $(i_1, \ldots, i_N)$th element of the array is denoted by $A[i_1, \ldots, i_N]$. A single i_n will be called the *index* and refers to the i_nth level of the nth way of the array. The scalar upper limit (I_n) will not be confused with the $n \times n$ identity matrix $\boldsymbol{I}_n$, because the latter will be in bold face. A bold font $[A]$ will denote the entire array or a nonscalar subarray. A vector is a one-way array, and a matrix is a 2-way array. In formulas, vectors will be regarded as matrices with one column. A subset of indices will be used to denote a subarray: for example, $A[i_1,]$ is the i_1th row of a matrix (thought of as a 1-way array or a column vector). We shall use subscripts to refer to a finite sequence of arrays, such as A_1, A_2. We will exploit the natural identification with a finite set of N-way arrays with a single $(N + 1)$-way array, although we will preserve the notational distinction.

We caution those readers who are familiar with tensors that we are deliberately avoiding the word *tensor*, because tensors commonly have structure that we do not assume here or are identified with multilinear mappings of abstract linear spaces. The dimension vector of classical N-way tensors typically consists of N identical numbers. Much of the theory is devoted to determination of the consequences of various symmetry or invariance properties, such as those possessed by arrays of moments or of Nth-order mixed partial derivatives. The arrays we use do not exhibit any symmetry properties.

3.1 Bilinear Models for Matrices

The simplest truly multiway arrays are two-way arrays or matrices. The simplest multilinear model, a one-component bilinear model for a two-way array μ, requires that

$$(8) \qquad \mu[i,j] = a[i]\beta[j], \qquad i = 1, \ldots, I; j = 1, \ldots, J,$$

where a is an I-dimensional vector corresponding to the first way and β is a J-dimensional vector corresponding to the second way. The equations (8) show that the I by J matrix μ is a nonlinear function of the two vectors a and β, given by

$$(9) \qquad \mu = a\beta^T = a \times \beta,$$

where equation (9) defines the outer product $\times$. For fixed a, the elements of μ are linear functions of β. That is, μ is a conditionally linear function of β. Symmetrically, for fixed β, the elements of μ are linear in a. The

matrix μ is thus a bilinear function of α and β. If (8) holds, then the rank of μ is 1, and the ratios $\mu[i, j]/\mu[I, j]$ equal $a[i]/a[I]$ for all j's provided that $\mu[I, j] \neq 0$. If $\mu[I, j] = 0$, then either all J $\mu[I, j']$'s must be zero or all I $\mu[i, j]$'s must be zero. Therefore the vectors α and β can be determined up to constant multiples from μ so long as μ is not $0_{I, J}$.

Many two-way arrays μ have rank greater than 1. An F-component bilinear model requires that μ be the sum of F one-component models, or that

$$(10) \qquad \mu[i, j] = \sum_{f=1}^{F} a_f[i]\beta_f[j].$$

Using outer product notation, (10) becomes

$$(11) \qquad \sum_{f=1}^{F} a_f \times \beta_f.$$

Letting A be the $I \times F$ matrix whose columns are $a_1, \ldots, a_F$ and B be the $J \times F$ matrix whose columns are $\beta_1, \ldots, \beta_F$, then the model can be rewritten in matrix notation as (12):

$$(12) \qquad \mu = AB^T$$

The rank of μ will be F as long as the columns of A and of B are linearly independent. The model can be rewritten again to emphasize that every row $\mu[i, \cdot]$ is the same linear function of $A[i, \cdot]$:

$$\mu[i, \cdot] = A[i, \cdot]B^T, \qquad i = 1, \ldots, I.$$

Similarly, each column of μ is given by a linear function of the corresponding column of B^T:

$$\mu[, j] = AB^T[, j], \qquad j = 1, \ldots, J.$$

That is, μ is a conditionally linear function of $vec(A)$ and $vec(B)$ that respects the matrix structure of A and of B. Any formula for μ of the form (12) is a bilinear model.

The form (12) can be derived from conditional linearity if the matrix structure of μ reflects the structure of A and B. That is, if the $I \times J$ matrix μ is a function of the $I \times F_1$ matrix A and the $J \times F_2$ matrix B such that for fixed B, each row of μ is a linear function of the corresponding row of A and each column of μ is a linear function of the corresponding column of B, then there must exist a fixed $F_1 \times F_2$ matrix C such that

$$(13) \qquad \mu = ACB^T.$$

Note that if the matrix η satisfies $\eta = N_1 ACB^T N_2$, where N_1 and N_2 are fixed square matrices, then η is a conditionally linear function of A and of B. However, the rows (columns) of η will not be determined by the corresponding rows of A (B) unless N_1 and N_2 are both diagonal matrices. Therefore conditional linearity is a much weaker hypothesis than multilinearity.

The formula (13) shows that μ is linear in the matrix C. Alternatively, (13) could be thought of as expressing

the matrix μ as a linear function of the 4-way array $A \times B$ defined by

$$(A \times B)[i, f_1, j, f_2] = A[i, f_1]B[j, f_2].$$

We shall let $\{A \times B\}$ denote the 4-way array of dimension (I, J, F_1, F_2) obtained by exchanging the second and third ways of $A \times B$. (In formulas: $\{A \times B\}[, j, f_1,] = (A \times B)[, f_1, j,]$.) Next we define $A_1*_2A_2$ to be the product of the 4-way array A_1 with the 2-way array A_2 given by

$$(14) \qquad A_1*_2 A_2 = \sum_{\nu, \nu'} A_1[,, \nu, \nu']A_2[\nu, \nu'].$$

Of course, A_1 and A_2 are conformable for $*_2$ multiplication only if the last two dimensions of A_1 form the dimension vector of A_2. The operation in (14) is sometimes called reduction of the last two indices of A_1 by the first two indices of A_2. With these definitions, (13) is seen to be equivalent to

$$(15) \qquad \mu = \{A \times B\}*_2C,$$

so that the 2-way array μ is written as the product of a 4-way array and a 2-way array, where the 4-way array has a special form. These formulas will have convenient extensions when describing multilinearity for higher-way arrays. Equation (15) can be interpreted as expressing μ as a linear combination of outer products of the columns of A with the columns of B, with weights given by entries of C.

This notation expresses the well-known nonidentifiability of bilinear models compactly. The product of any nonsingular $F_1 \times F_1$ matrix R_1 (respectively $F_2 \times F_2$ matrix R_2) and its inverse can be inserted between A and C (respectively C and B^T) in (13) without changing μ. Using the associativity of matrix multiplication and the equivalence of (13) and (15), the following equation must also hold:

$$(16) \qquad \mu = \{(AR_1) \times (BR_2)\}*_2(\{R_1^{-1} \times R_2^{-1}\}*_2C).$$

Since (16) is (13) with A replaced by $A' = AR_1$, $B' = BR_2$ and C replaced by $C' = \{R_1^{-1} \times R_2^{-1}\}*_2C$ and since the matrices R_1 and R_2 are arbitrary nonsingular matrices, the matrices A, B and C cannot be identified from μ without side conditions. The most standard mathematical conditions are the requirements that $F_1 = F_2$, that both A and B have orthogonal columns, that the columns of A have unit length, that C be the identity matrix, and that the lengths of the columns of B be nondecreasing. Under these conditions, whenever μ has a unique singular value decomposition, μ has a factorization of the form (11) that is unique up to permutation and sign changes of columns. While mathematically elegant, orthogonality conditions may result in a factorization that is hard to interpret. For example, if A and B are interpretable only if all of their elements are nonnegative (as is usually true in

spectroscopy), the requirement that the columns are orthogonal is problematic.

The equation (3) has the form of the bilinear model (10), where $A[i, j]$ corresponds to $\mu[i, j]$, $\varepsilon_f[i]$ to $a_f[i]$, and $c_f[j]L$ to $\beta_f[j]$. Since parameters of bilinear models cannot be identified without further conditions, spectroscopic interpretation of bilinear models will require additional constraints.

3.2 Trilinear Models

The preceding subsection discusses many properties of bilinear models. We shall see that several natural extensions to 3-way arrays lead to different mathematical models when more than one component is present. In particular, the family of trilinear models obtained from multilinearity is different from the family obtained by adding together 1-component trilinear models.

A *triad* (Kruskal, 1977), or a *decomposable* array, is a 3-way array that is the outer product of three vectors, one for each way, so that

$$(17) \qquad \mu = \alpha \times \beta \times \gamma,$$

or

$$\mu[i, j, k] = \alpha[i] \times \beta[j] \times \gamma[k].$$

Notice that all nonzero subarrays obtained by fixing any index, that is, the arrays $\mu[i, ,]$, $\mu[, j,]$ and $\mu[, , k]$ have matrix rank 1. The vectors α, β and γ are identifiable from any two such matrices. It is very natural to define the array rank of μ satisfying (17) to be 1.

In (17) above, each element of μ is a conditionally linear function of α when β and γ are fixed, of β when α and γ are fixed, and of γ when α and β are fixed. Moreover, the vector obtained by fixing any two indices depends on the vectors for the corresponding way only through the elements corresponding to the fixed indices. For example, $\mu[i, j,]$ is determined by $\alpha[i]$, $\beta[j]$ and γ, and is linear in each when the other two are fixed. We now extend these hypotheses of separate linearity to arrays μ that are not decomposable. That is, the hypothesis of *trilinearity* is that μ is a conditionally linear function of A $(I \times F_1)$ given B $(J \times F_2)$ and Γ $(K \times F_3)$ such that $\mu[i, ,]$ depends only on $A[i,]$, B and Γ; that μ is a conditionally linear function of B given A and Γ such that $\mu[, j,]$ depends only on A, $B[j,]$ and Γ; and that μ is a conditionally linear function of Γ given A and B such that $\mu[, , k]$ depends only on A, B and $\Gamma[k,]$.

Trilinear models are easiest to formulate if suitable array operators extending $*_2$-multiplication are defined. A 6-way array A_1 is *conformable* for $*_3$-*multiplication* with a 3-way array A_2 if the last 3 dimensions of the 6-way array are the dimensions of the 3-way array. The $*_3$-product $A_1*_3A_2$ is the 3-way array

$$A_1*_3A_2 = \sum_{\nu} \sum_{\nu'} \sum_{\nu''} A_1[, , , \nu, \nu', \nu'']A_2[\nu, \nu', \nu''].$$

It will be convenient to let $\{A\}$ denote the array that results when the ways of A are exchanged as follows: $\{A\}[, i_3, i_5, i_2, i_4,] = A[, i_2, i_3, i_4, i_5,]$. If the 6-way array A is an outer product of three matrices M_1, M_2 and M_3, then the first three ways of $\{A\}$ correspond to the rows of M_1, M_2 and M_3 and the remaining 3 ways of $\{A\}$ correspond to the columns of the matrices. Proposition B.1 states that trilinearity implies that

$$(18) \qquad \mu = \{A \times B \times \Gamma\}*_3C,$$

where C is an $F_1 \times F_2 \times F_3$ array, to be referred to as a *core array*. The equation (18) is both an extension of (15) to 3-way arrays and an alternative notation for Kruskal's triple product of A, B and Γ with C. If either A or B or Γ fails to be of full column rank, then C will not be unique.

The trilinearity equation (18) can be written in terms of outer products of the columns of the A, B, and Γ as

$$(19) \qquad \mu = \sum_{f_1} \sum_{f_2} \sum_{f_3} \alpha_{f_1} \times \beta_{f_2} \times \gamma_{f_3} C[f_1, f_2, f_3].$$

In this form, the trilinearity equation can be recognized as a *Tucker3* or *T3 model*, named in reference to Tucker (1963). That is, (19) presents μ as a weighted sum of outer products of columns of A, B and Γ with weights given by the core array C. If A, B and Γ are unknown, they cannot be identified from μ, because it follows from the associativity property of $*_3$-products (Proposition C.1) that if M_1, M_2 and M_3 are nonsingular $F_1 \times F_1$, $F_2 \times F_2$ and $F_3 \times F_3$ matrices, then

$$\mu = \{A \times B \times \Gamma\}*_3C$$

$$(20) \qquad = \{(AM_1) \times (BM_2) \times (\Gamma M_3)\}*_3$$

$$\cdot (\{(M_1^{-1}) \times (M_2^{-1}) \times (M_3^{-1})\}*_3C).$$

That is, if the matrices A, B and Γ can be transformed by multiplication on the right, the array μ is preserved if C is countertransformed appropriately. This nonidentifiability is just the 3-way analog of the nonidentifiability of bilinear models (16), the nonidentifiability that is referred to as the "rotation problem" in factor analysis. To identify A, B, Γ and C, restrictions must be added. Two kinds of restrictions are common: restrictions on the matrices and restrictions on the core array. The countertransformation equation (20) implies that the columns of A, of B and of Γ can be selected to form orthonormal sets.

Since orthogonality conditions do not arise in spectroscopy, we now consider restrictions on C. Recall that if a matrix μ has an F-factor bilinear model, then the 2-way core matrix C can always be taken to be the identity matrix. A 3-way analog of an identity matrix is a "diagonal" 3-way "identity" array $I_{F, F, F}$ whose entries are 1 when all three indices coincide and are 0

otherwise. Therefore, a 3-way analog of the identifiable form of the bilinear model would be

$$(21) \qquad \mu = \{A \times B \times \Gamma\}*_3 I_{F,F,F},$$

where $F = min(F_1, F_2, F_3)$. This model was introduced by Carroll and Chang (1970) and by Harshman (1970) and was named Canonical Decomposition (CANDECOMP) in the first paper and Parallel Factors (PARAFAC) in the second. The matrices A, B and Γ cannot be identified solely from μ, because μ is unchanged if α_f, β_f and γ_f are multiplied by 3 scalars whose product is one or if the columns of A, B and Γ are subjected to the same permutation. Kruskal (1977) showed that frequently these (trivial) nonidentifiabilities are essentially the only nonidentifiabilities by proving (4 sets of) sufficient conditions for identifiability, up to the inherent trivial difficulties mentioned above. His proofs do not explicitly decompose μ. Subsection 4.1 below contains an explicit determination of the parameters under stronger conditions than Kruskal's, but without assuming that the columns of A, of B or of Γ are orthogonal. Indeed, if A, B and Γ are chosen without orthogonal columns, and if μ is defined via (21), the identifiability implies that orthogonality of the columns of the parameter matrices cannot be imposed without increasing F. Therefore, some T3 arrays with $F_1 = F_2 = F_3 = F$ do not have F-term PARAFAC expressions.

Two other representations for the PARAFAC model are convenient. The first form is an explicit outer product form:

$$(22) \qquad \mu = \sum_{f=1}^{F} \alpha_f \times \beta_f \times \gamma_f$$

or

$$(23) \qquad \mu[i, j, k] = \sum_{f=1}^{F} \alpha_f[i]\beta_f[j]\gamma_f[k],$$

where γ_f here corresponds to $\gamma_f C[f, f, f]$ of (19). The smallest value of F for which (22) holds is the *array rank* of μ. The second representation is a matrix representation that follows from (22) by fixing the third index of μ at k:

$$(24) \qquad \mu_k := \mu[, , k] = \sum_{f=1}^{F} \alpha_f \times \beta_f \gamma_f[k]$$
$$= A \Delta_k B^T, \qquad k = 1, \ldots, K,$$

where Δ_k is a F by F diagonal matrix whose elements are $\gamma_f[k]$, $f = 1, \ldots, F$. Equation (24) shows that an F-term PARAFAC model implies that the K matrices μ_k are weighted sums of the same F matrices $\alpha_f \times \beta_f$, with weights $\gamma_f[k]$.

Comparison of (22) with (19) shows that an F-term PARAFAC model is a T3 model with $F_1 = F_2 = F_3$ and no outer products in which f_1, f_2 and f_3 are not all

the same. In the spectroscopy setting, such a term corresponds to the properties of two or three different fluorophores. That is, the presence of several fluorophores results in an interaction among the fluorophores. A general T3 model permits all possible 3-way interactions to be present. Another subclass of models would permit all interactions among two ways and no interactions with the other way, which we take to be the third way. Such models are known as *T2 models*. Mathematically, T2 models have core arrays of order $F_1 \times F_2 \times (F_1 F_2)$ in which every $C[f_1, f_2,]$ has exactly one nonzero element. If the $F_3 = F_1 F_2$ vectors $\gamma_1, \ldots, \gamma_{F_3}$ are reindexed to have the two subscripts f_1 and f_2, then the core array can be absorbed into the γ's giving the mathematical form below for a T2 model:

$$(25) \qquad \mu = \sum_{f_1=1}^{F_1} \sum_{f_2=1}^{F_2} \alpha_{f_1} \times \beta_{f_2} \times \gamma_{f_1, f_2}$$

or

$$(26) \qquad \mu[i, j, k] = \sum_{f_1=1}^{F_1} \sum_{f_2=1}^{F_2} \alpha_{f_1}[i] \times \beta_{f_2}[j] \times \gamma_{f_1, f_2}[k].$$

The matrix form for a T2 model implies that

$$\mu_k = A \Phi_k B^T, \qquad k = 1, \ldots, K,$$

where the $F_1 \times F_2$ matrices Φ_k will not be diagonal unless a PARAFAC model holds. If μ satisfies a T3 model with $F_3 < K$, then μ_k can be represented as above, where the matrices Φ_k are not free to vary, but must be linear combinations of the F_3 matrices corresponding to fixing a level of the core array C:

$$\Phi_k = \sum_{f_3=1}^{F_3} \gamma_{f_3}[k]C[, , f_3].$$

This description of a T2 model uses an implicit core array. However, if a new $F_1 \times F_2 \times K$ core array Γ^* is constructed by setting $\Gamma^*[f_1, f_2, k]$ equal to $\gamma_{f_1, f_2}[k]$ of (25), then

$$\gamma_{f_1, f_2} = \sum_{k=1}^{K} I_K[, k]\Gamma^*[f_1, f_2, k]$$

and (25) is just

$$\mu = \{A \times B \times I_K\}*_3 \Gamma^*.$$

So μ satisfies a T2 model if a T3 model holds with $\Gamma = I_K$. It follows from (20) that the matrices A and B and the core array Γ^* cannot be identified for T2 models.

The element-by-element forms of the PARAFAC and T2 models have exactly the same form as the trilinear models introduced in Section 2. Equation (23) is just (4), where $\varepsilon_f[i]$ replaces $\alpha_f[i]$, $\pi_f[j]$ replaces $\beta_f[j]$, and $c_f[k]$ replaces $\gamma_f[k]$. Consequently, equation (4) is an F-term PARAFAC model with parameter matrices ε, π and c. Equation (7), describing fluorescence measure-

ments on a specimen with excitation transfer, is (26) with $A_{f_1}[i]$ replacing $a_{f_1}[i]$, $\pi_{f_2}[j]$ replacing $\beta_{f_2}[j]$ and $Q_{f_1,f_2}[k]$ replacing $\gamma_{f_1,f_2}[k]$. That is, (7) is a T2 model with parameter matrices A and π and core array Q.

3.3 General Multiway Arrays

All of the models for 3-way arrays can be extended to N-way arrays. Lastovicka (1981) discusses 4-way extensions of T3 models, and Kapteyn, Neudecker and Wansbeek (1986) discuss N-way extensions, using matrix notation. Carroll and Chang discussed CANDE-COMP models for r-way arrays, as well as for 3-way arrays. We outline N-way array notation, and indicate some spectroscopic applications. For more properties of the N-way product, see Leurgans (1991).

The multilinearity hypothesis is that an N-way array μ is a separately linear function of N parameter matrices $A_1, \ldots, A_N$ with the property that the separate linearity respects the array structure. The i_nth row of the nth matrix A_n is assumed to affect only those elements of μ with nth index equal to i_n. It can be shown that the multilinearity hypothesis implies that μ must be the weighted sum of N-way arrays that are the N-fold outer products of one column of each of the matrices A_n, $n = 1, \ldots, N$, to be abbreviated by

$$\mu = \{\times_{n=1}^{N} A_n\} *_N C,$$

where C is the fixed N-way core array of coefficients. If the core array is unrestricted, the matrices A_n will not be identifiable, because the $*_N$-product is associative, so the core array can be counter-rotated if the matrices A_n are multiplied on the right by nonsingular matrices.

For general $N \geq 2$, the array C cannot be forced to be diagonal, and assuming special hypotheses corresponds to considering special submodels. The only submodel that has received much attention is the N-way PARAFAC model, in which the elements C are assumed to be 0, unless all N indices are the same. These models can be written as

$$(27) \qquad \mu = \sum_{f=1}^{F} a_f^{(1)} \times \cdots \times a_f^{(N)} c_f,$$

where c_f denotes the element of C with all indices equal to f and $a_f^{(n)} = A_n[,f]$. Fixing all indices of μ except for the first two ways, (27) implies

$$(28) \qquad \mu[,,i_3, \ldots, i_N] = A_1 D_{i_3, \ldots, i_N} A_2^T,$$

where the $F \times F$ diagonal matrix $D_{i_3, \ldots, i_N}$ has fth diagonal element equal to $c_f \Pi_{n=3}^{N} A_n[i_n, f]$. But (28) just implies that a 3-way PARAFAC model holds for an $I_1 \times I_2 \times (\Pi_{n=3}^{N} I_n)$ 3-way array. Since the parameters of a 3-way PARAFAC model are identifiable (up to isolated singularities), the parameters of N-way PARAFAC models are identifiable, and many of the methods

for 3-way PARAFAC models can be extended to more ways.

Arrays with more than 3 ways occur in spectroscopy. Equation (5) is a 4-way PARAFAC model. Another 4-way PARAFAC model was considered by Durell et al. (1990). Equation (6) is a 1-term 5-way PARAFAC model.

4. ESTIMATION OF THE PARAMETERS OF PARAFAC MODELS

Least-squares estimation of the parameters of a PARAFAC model requires minimizing the quadratic function of the vectorized residuals

$$Q(A, B, \Gamma) = \| vec(Y) - vec(\mu) \|^2.$$

The criterion Q is a nonlinear function of (A, B, Γ), and is unchanged when columns of A, B and Γ are subjected to the same permutation. This permutation invariance of Q guarantees that Q has many local minima. There is no closed form solution to the least-squares optimization problem, so iterative algorithms are generally employed. The observation above that multiple minima should be anticipated suggests that good initializations are important. (Our numerical experience strongly supports the importance of initializations.) Gnanadesikan and Kettenring (1984) pointed out that one way to initialize algorithms is to find a method that recovers the parameters in the absence of noise. In our context, the array μ of (22) needs to be decomposed into its constituent parts. The first subsection discusses this problem and gives an intuitive explanation of the simultaneous decomposition presented in Sanchez and Kowalski (1988a, b, 1990) and in Burdick et al. (1990). The decomposition is also closely related to the algebraic decomposition of Sands and Young (1980). The explanation is similar in spirit to that of Sanchez and Kowalski (1988b).

A recursive approach is suggested by the observation that the model is the sum of F terms of the same form. Thus, it seems natural to fit the model by fitting a "leading" triad, subtracting it off, fitting another triad to the residuals, and so on. The second subsection demonstrates that even in the absence of noise, F-term PARAFAC models cannot be fitted by fitting a sequence of 1-term models. Our simple demonstration indicates that PARAFAC models should not be fitted recursively, but all parameters should be estimated simultaneously.

4.1 Simultaneous Estimation of F PARAFAC Terms

Consider an array μ satisfying (22) and ask whether it can be decomposed into a unique triple (A, B, Γ). From the discussion below (21), some Parameterization Conventions must be imposed. One natural convention used by Leurgans, Ross and Abel (1992) is that the

columns of A and of B have length 1. Kruskal (1977) gave several sets of conditions under which a decomposition is unique. Since the models of greatest practical importance are those with $F = 2, 3$ or 4, we concentrate on an explicit decomposition feasible under the strongest of Kruskal's uniqueness conditions. The decomposition has an intuitive explanation in the spectroscopic setting. We provide this explanation in four steps. First we argue that if the properties of the fluorophores are known, then their concentrations can be deduced from subarrays of μ. The second step is the fact that the model implies that the expected fluorescence intensities are transformed linearly when the chemical conditions change and excitation wavelength is fixed. The third step is the verification that the columns of B are eigenvectors of the matrix representing this linear transformation. The final step is the construction of the matrix directly from μ. We shall show that the final step is an easy consequence of properties of Moore-Penrose generalized inverses. The equivariance of the models under permutations of the ways of the array implies that A and Γ can be identified under similar conditions.

For this subsection, it is convenient to reparameterize so that all F vectors for each way have unit length. Therefore we now include a scalar parameter $\zeta[f]$ in the fth term of (22) and thus the model we consider for the expected fluorescence intensity at excitation level i, emission level j, and chemistry level k is

$$(29) \quad \mu[i, j, k] = \sum_{f=1}^{F} a_f[i]\beta_f[j]\gamma_f[k]\zeta[f],$$

$$i = 1, \ldots, I; \quad j = 1, \ldots, J; \quad k = 1, \ldots, K,$$

where $\|a_f\| = \|\beta_f\| = \|\gamma_f\| = 1, f = 1, \ldots, F$. The parameter $a_f[i]$ is proportional to the fraction of incoming light at wavelength i absorbed by fluorophore f, $\beta_f[j]$ is proportional to the fraction of emissions from fluorophore f emitted at wavelength j, $\gamma_f[k]$ is proportional to the fraction of photon absorptions by fluorophore f that yield fluorescence under chemistry conditions k, $\zeta[f]$ is the scaled concentration of fluorophore f in appropriate units, and F is the number of fluorophores. (The scaling of the concentration is multiplication of the concentration by physical constants such as the average extinction coefficient and the average quantum yield.) We assume that all fluorophores will fluoresce under some of the conditions, or that $a_f[i]\beta_f[j]\gamma_f[k] \neq 0$ for some i, j, k and every f. This property implies that the $\zeta[f]$'s are identifiable: The concentration of compounds that do not fluoresce under any of the conditions employed will not be relevant to the experiment.

Consider the IK J-vectors $\mu[i, , k]$ containing the expected fluorescence intensities at the various emission wavelengths when excitation (i) and chemistry (k) are fixed. We shall refer to this vector as the (i, k)th emis-

sion profile. We first argue that if the properties of the fluorophores (A, B and Γ) are known, then the scaled concentrations ζ can be deduced from any emission profile to which all fluorophores contribute.

We use the equation for ζ to show that the effect on the emission profiles of changing the chemical conditions is a linear transformation of the emission profiles. That is, if $k \neq k'$, then $\mu[i, , k']$ is a linear function of $\mu[i, , k]$. Moreover, we demonstrate that the linear function is the same for all excitation levels i, so the linear transformation corresponds to a matrix $\tau_{k, k'}$. The transformation matrix $\tau_{k, k'}$ depends on B and Γ, properties of the fluorophores. We will argue that the columns of B are eigenvectors of every transformation matrix, so that if a transformation matrix can be deduced from μ alone, then the columns of B can be determined from μ. The final step in our explanation is therefore the derivation of the transformation matrix from μ.

The model (29) implies that each element of μ is a linear function of ζ. Consequently, all subsets of $vec(\mu)$ are linear functions of ζ. For example, $\mu[i, , k]$ satisfies

$$(30) \quad \mu[i, , k] = \sum_{f=1}^{F} \beta_f a_f[i]\gamma_f[k]\zeta[f] = BD_{i,k}\zeta,$$

where $D_{i, k}$ is the F by F diagonal matrix whose fth diagonal element is $a_f[i]\gamma_f[k]$. The matrix $BD_{i,k}$ is the J by F matrix that maps ζ into $\mu[i, , k]$. Since μ is the expectation of an observable array, while ζ is not the expected value of a fluorescence intensity, the concentration vector ζ needs to be deduced by solving equation (30). The equation can be solved uniquely whenever the rank of $BD_{i,k}$ is $F \leq I$, or when the F columns of B are linearly independent and $D_{i,k}$ is nonsingular. The condition of linear independence can be interpreted as a requirement that no two different mixtures of fluorophores have the same fluorescence intensities at all emission wavelengths: no mixture can be substituted for any other mixture without some change in the expected fluorescence intensities. The nonsingularity condition is that $a_f[i]\gamma_f[k]$ be nonzero for every f, or that every fluorophore fluoresces at the fixed excitation i and chemistry k. If these conditions hold, then

$$(31) \quad \zeta = D_{i,k}^{-1}B^\dagger\mu[i, , k],$$

where the Moore-Penrose generalized inverse of B is $B^\dagger = (B^TB)^{-1}B^T$. Observe that if y is a hypothesized value for $\mu[i, , k]$, any component of y orthogonal to the column space of B will be mapped to $\zeta = 0_F$ by (31).

We emphasize that (31) shows that the true concentrations ζ are a known linear function of the emission profiles for fixed i and k. If equation (31) for i and k is substituted into (30) for i and k', it follows that

$$(32) \quad \mu[i, , k'] = BD_{i,k'}D_{i,k}^{-1}B^\dagger\mu[i, , k]$$

$$(33) \quad = \tau_{k, k'}\mu[i, , k],$$

where the $J \times J$ transformation matrix $\tau_{k,\,k'}$ maps the emission profiles under chemistry k to those for chemistry k' when excitation level i is fixed and implicit. However, the product of the diagonal matrices in (32) is a diagonal matrix with fth diagonal element $a_f[i]\gamma_f[k']/(a_f[i]\gamma_f[k]) = \gamma_f[k']/\gamma_f[k]$, a ratio independent of i. Defining D_k to be the diagonal matrix with fth diagonal element $\gamma_f[k]$, the transformation matrix is

$$(34) \qquad \tau_{k,k'} = BD_{k'}D_k^{-1}B^\dagger,$$

where the implicit excitation level i can be ignored. Therefore, the model (29) implies that changing chemical conditions induces a linear transformation of the emission profiles, and that the linear transformation is the same for every excitation wavelength. Note that (34) holds even when $\gamma_f[k'] = 0$ for some values of f.

We now explain how the spectra β can be obtained from the transformation matrices. Multiplying equation (34) on the right by B gives

$$(35) \qquad \tau_{k,k'}B = BD_{k'}D_k^{-1}.$$

Since the matrix $D_{k'}D_k^{-1}$ is diagonal, (35) establishes that the columns of B are eigenvectors of $\tau_{k,\,k'}$ with eigenvalues $\gamma_f[k']/\gamma_f[k]$. The spectroscopic interpretation of this property is that if only one fluorophore is present, then the vector of expected fluorescence intensities always has the shape of the spectrum of that fluorophore; only the overall intensity can change with chemistry. If any transformation matrix has F distinct nonzero eigenvalues, then every column of B can be deduced up to scalar multiples from that transformation matrix.

Therefore the emission spectra can be deduced from the transformation matrices, matrices that were defined as functions of the emission and chemistry spectra (34). We now show that the transformation matrix $\tau_{k,\,k'}$ is determined by the matrices $\mu[,\,,k']$ and $\mu[,\,,k]$. Intuitively, if the pairs of expected fluorescence vectors $\mu[i,\,,k']$ and $\mu[i,\,,k]$ are known for many $i = 1, \ldots, I$, then the transformation matrices $\tau_{k,\,k'}$ satisfying (33) can be determined directly from μ. It will follow that the emission spectra B can be deduced from μ even if A and Γ are unknown.

To establish this claim mathematically, observe that (34) implies trivially that

$$\tau_{k,k'} = BD_{k'}I_FD_k^{-1}B^\dagger.$$

If the excitation spectra are linearly independent, then $A^T(A^T)^\dagger = I_F$, so substitution for I_F in the equation above gives

$$\tau_{k,k'} = BD_{k'}A^T(A^T)^\dagger D_k^{-1}B^\dagger.$$

When the columns of A and of B are linearly independent and D_k is nonsingular, Corollary 5 of Theorem 2.16 of Pringle and Rayner (1971) states that the

Moore-Penrose inverse of (BD_kA^T) is the product of the Moore-Penrose inverses in reverse order, so that

$$\tau_{k,k'} = (BD_{k'}A^T)(BD_kA^T)^\dagger = \mu_{k'}^T(\mu_k^T)^\dagger,$$

and $\tau_{k,\,k'}$ can be determined directly from μ even when neither A and B are known.

We comment on two characteristics of this decomposition. The decomposition collapses if the third way has only one level because two or more chemistries must be observed if the effect of changing chemistry is to be deduced. Therefore, the decomposition does not contradict the known nonidentifiability of bilinear models. Secondly, we note that the array rank of μ is determined during the decomposition, because F will be the number of non-zero eigenvalues. All F columns of A, B and Γ are recovered simultaneously: the columns are not derived sequentially. Indeed, in Subsection 4.2 we show that sequential fitting is only possible if two of the parameter matrices have orthogonal columns.

To adapt this decomposition to initialize cycling-least-squares (or other iterative) algorithms, the array Y can be substituted for μ, giving sample transformation matrices

$$\hat{\tau}_{k,k'} = Y_{k'}^T(Y_k^T)^\dagger.$$

These matrices are square, but they are seldom symmetric, so their F leading eigenvalues may contain complex conjugate pairs when $F > 1$ if (29) does not hold exactly. Essentially, Leurgans, Ross and Abel (1992), Burdick et al. (1990) and Sanchez and Kowalski (1988a, b, 1990) replace $(Y_k^T)^\dagger$ with $(Y_+^T)^\dagger$, where $Y_+ = \Sigma_{k=1}^K Y_k$. Since sample discrepancies with (29) generally introduce additional nonzero eigenvalues, all three papers use variants in which the matrices Y_k are linearly transformed to matrices with F rows and columns, so that the resulting analogs of the transformation matrices are $F \times F$. If (29) holds exactly, then the parameters will be recovered. The requirement that each fluorophore have some emission for each chemistry can be weakened to the requirement that each fluorophore have some emission at some chemistry. See Leurgans, Ross and Abel (1992) for even weaker conditions, and for some discussion of our experience using these relationships to initialize least-squares calculations.

4.2 Recursive Estimation and 3-way PARAFAC Models

One useful property of real matrices is that the least-squares rank-P approximation to a matrix can be found by determining P rank-1 approximations. For example, the best rank-2 approximation is the sum of the best rank-1 approximation to the original matrix and the best rank-1 approximation to the matrix obtained by subtracting the first rank-1 approximation from the original matrix. The fact that rank-P approxi-

mations can be constructed recursively is a consequence of the singular value decomposition and is sometimes called the Eckart-Young Theorem (Eckart and Young, 1936).

The simplest 3-way generalization is rank-1 least-squares approximation of rank-2 arrays. In analogy with the matrix problem, we ask whether a rank-2 array is always the sum of the best rank-1 approximation to the rank-2 array and the best rank-1 approximation to the array difference of the original rank-2 array and the first rank-1 approximation. Equivalently, the question is whether the array difference has rank-1.

If Kruskal's conditions hold, then the rank-2 array has a unique representation as the sum of two rank-1 arrays. Therefore the array difference can only have rank 1 if the array difference is one of the rank-1 arrays in the original representation, which would force the best rank-1 approximation to be the other rank-1 array of the representation. The proposition below states that if the best rank-1 approximation coincides with one of the rank-1 arrays of the original representation, then at least two of the pairs of vectors in the original representation are orthogonal.

PROPOSITION 4.1. *Suppose* (α_1, α_2), (β_1, β_2) *and* (γ_1, γ_2) *are pairs of linearly independent nonzero I, J, K vectors, respectively. Define the 3-way array μ by*

$$\mu = \alpha_1 \times \beta_1 \times \gamma_1 + \alpha_2 \times \beta_2 \times \gamma_2.$$

If

$$\inf_{\alpha, \beta, \gamma} \| \mu - \alpha \times \beta \times \gamma \|$$

is achieved by $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (\alpha_f, \beta_f, \gamma_f)$, *for* $f = 1$, *then at least two of the three pairs of vectors* (α_1, α_2), (β_1, β_2) *and* (γ_1, γ_2) *must be orthogonal.*

PROOF. The conditional linearity of μ implies that the solution to the minimization must simultaneously solve the following three normal equations:

$$\sum_{j,k}\left(\mu[,j,k] - \hat{\alpha}\hat{\beta}[j]\hat{\gamma}[k]\right)\hat{\beta}[j]\hat{\gamma}[k] = 0_I,$$

$$\sum_{i,k}\left(\mu[i,,k] - \hat{\alpha}[i]\hat{\beta}\hat{\gamma}[k]\right)\hat{\alpha}[i]\hat{\gamma}[k] = 0_J,$$

$$\sum_{i,j}\left(\mu[i,j] - \hat{\alpha}[i]\hat{\beta}[j]\hat{\gamma}\right)\hat{\alpha}[i]\hat{\beta}[j] = 0_K.$$

If $\hat{\alpha} = \alpha_1$, $\hat{\beta} = \beta_1$ and $\hat{\gamma} = \gamma_1$, then substitution in the first equation results in the equation

$$\sum_{j,k}\left(\alpha_2\beta_2[j]\gamma_2[k]\right)\beta_1[j]\gamma_1[k] = 0_I.$$

Collecting terms in j and in k shows that

$$\alpha_2\left(\sum_j \beta_2[j]\beta_1[j]\right)\left(\sum_k \gamma_2[k]\gamma_1[k]\right)$$

$$= \alpha_2(B^TB)[2,1](\Gamma^T\Gamma)[2,1] = 0_I.$$

Since the vector α_2 is assumed to be nonzero, the inner product of β_1 with β_2 or of γ_1 with γ_2 must be 0. Similarly, the second and third equations imply that either α_1 must be orthogonal to α_2 or both (β_1, β_2) and (γ_1, γ_2) are orthogonal pairs of vectors. The conclusion follows. □

This proposition shows that unless orthogonality conditions can be imposed, fitting PARAFAC models with fewer terms than are actually present cannot recover the parameters of the larger model. Since the orthogonality conditions are tenable in few spectroscopic contexts, the initialization methods of Subsection 4.1 are to be preferred to increasing F and using the parameter estimates from the $(F - 1)$-term PARAFAC model to initialize the F-term model.

5. EXAMPLES

We present three examples: one with theoretical arrays, one with real data from a wet chemistry simulation, and one with data from biological specimens. In each of these applications, the data are fluorescence intensity as a function of three independent variables: excitation wavelength, emission wavelength and the concentration of a fluorescence quencher, so that the fluorescence from each component declines with increasing concentration. The mathematical example shows that when 3-way trilinear models hold, separate bilinear models, as implemented by separate singular value decompositions, may not reveal the underlying multilinear structure. The second is a wet chemistry simulation: the properties of 3 dyes were recovered when the dyes were mixed. The third example demonstrates the use of multilinear models to resolve the properties of components within a biological system: the components are individual amino acids within a protein.

5.1 Example 1: A Three-way Array without Noise

We started with 3 matrices A, B and Γ, chosen to be typical of excitation, emission and quenching spectra, of dimension 20 × 3, 25 × 3 and 5 × 3, respectively. The columns of each matrix are plotted against level number in Figure 3. The columns of A and B have been scaled to have Euclidean length 1. These vectors for the first two ways are unimodal and positive, with the peak for the second factor being between the peaks for the other two. The vectors for the third way are all decreasing and roughly linear. The inner products of the vectors for each way in Table 1 quantify these
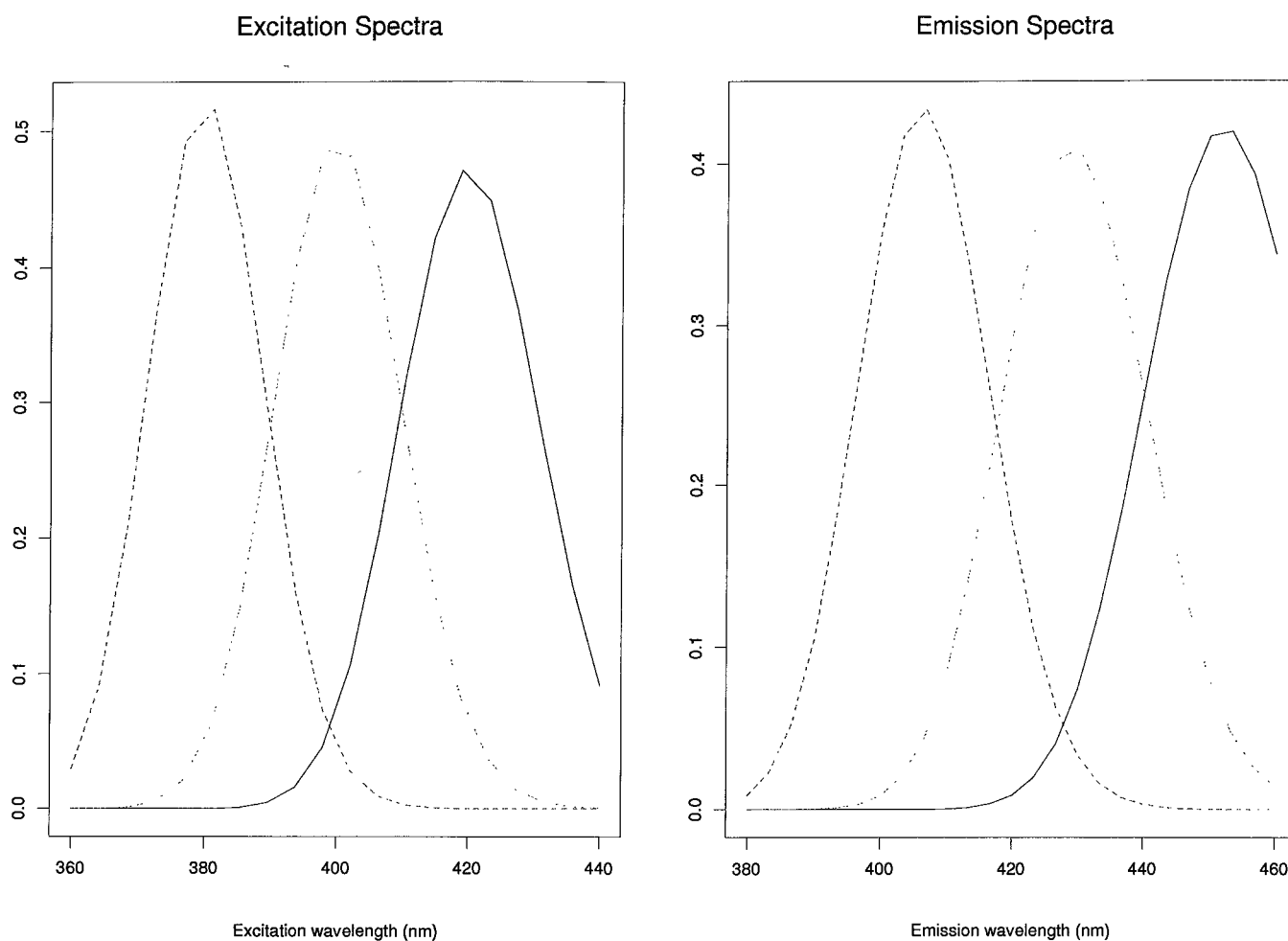
Excitation Spectra   Emission Spectra

Excitation wavelength (nm)   Emission wavelength (nm)

FIG. 3a.   *Excitation and emission spectra for three factors in mathematically synthesized system.*

observations: the inner products for the first two ways are 0.01 for factors 1 and 3, are near 0.4 for the factors 1 and 2, and near 0.3 for the factors 2 and 3. The columns of Γ are less distinct: all of the inner products for the third way exceed 0.9.

We then computed the $20 \times 25 \times 5$ array μ from (22). Figure 4 shows perspective plots of the elements of μ_1, μ_3 and μ_5 against the levels of the first and second ways. Each picture shows three clear bumps. Each bump is from one of the triads, or of the three values $f = 1, 2, 3$. The relative and absolute heights of the bumps vary with k, the level of the third way. The equation (24) guarantees that the 5 matrices μ_k can all be decomposed as the product of a 20×3 matrix, a

3×3 diagonal matrix and a 3×25 matrix using the same 20×3 and 3×25 matrices.

These product decompositions of the 5 matrices are not unique: the matrices μ_k all have rank 3, and must therefore have singular value decompositions $\bar{U}_k \bar{D}_k \bar{V}_k^T$, where the $\bar{U}_k$'s are 20×3 matrices with orthogonal columns, the $\bar{D}_k$'s are 3×3 diagonal matrices, and the $\bar{V}_k$'s are 25×3 matrices with orthogonal columns. The numerical accuracy of our calculations is supported by the singular values for the 5 matrices μ_k given in Table 2. These singular value decompositions are also product decompositions of the form (24), except that the 20×3 and 25×3 matrices are constructed independently for each k. Recall from Table 1 the columns

TABLE 1
Inner products of scaled vectors for Example 1, a three-way array without noise

Excitation (first way) $I = 20$ (α)			Emission (second way) $J = 25$ (β)			Quenching (third way) $K = 5$ (γ)		
1.000	–	–	1.000	–	–	1.000	–	–
0.380	1.00	–	0.409	1.000	–	0.972	1.000	–
0.014	0.306	1.000	0.015	0.306	1.000	0.932	0.990	1.000

Chemistry Parameters
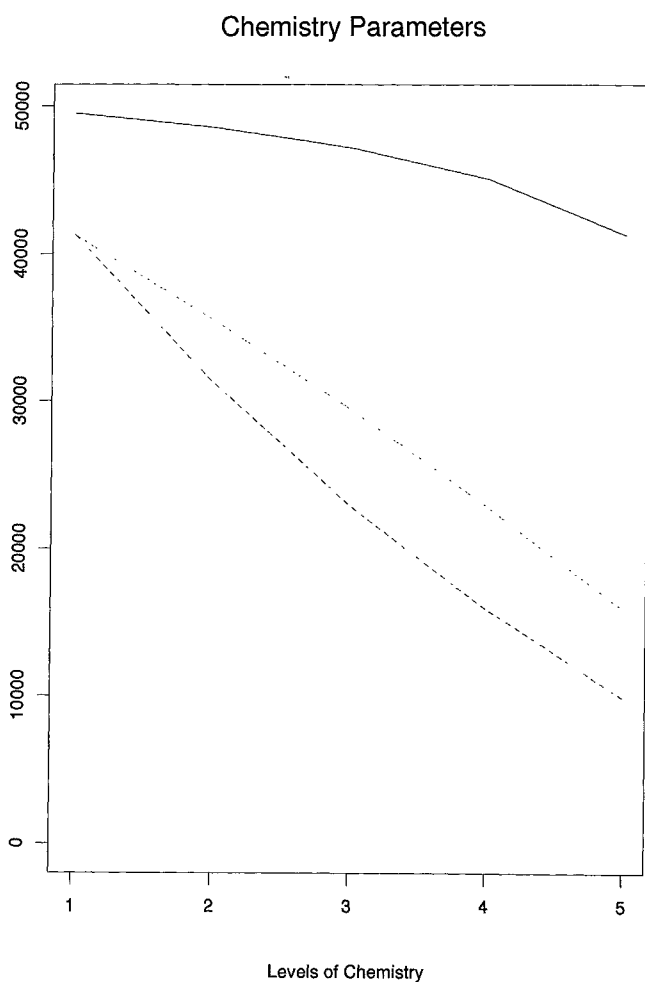
FIG. 3b. *Chemistry parameters for three factors in mathematically synthesized system.*

Quenching Level 1

A

Quenching Level 3

B

of A are not orthogonal, so the 20×3 matrices of left singular vectors will not coincide with A. Each panel of Figure 5 displays a column of A or of B with the appropriate columns of $\tilde{U}_k$ or $\tilde{V}_k$. Each row of the figure is for a different triad. The line in each panel connecting asterisks is the theoretical spectrum obtained from A or B. The other lines are columns of $\tilde{U}_k$ or $\tilde{V}_k$, for some k.

The columns from the singular value decompositions are different from the true spectra, so recovering multiplicative decompositions for each matrix μ_k does not reveal the underlying multilinear model, even in the absence of noise. The first left singular vectors do have positive elements, but they are shifted toward the second and third columns of A. The sets of second and third singular vectors have negative lobes forced by the requirement that singular vectors by mutually orthogonal. The first two singular vectors for the fifth level of the third way, which is the level with smallest elements of D_k, are nearer to the first two columns of A than the other singular vectors are.
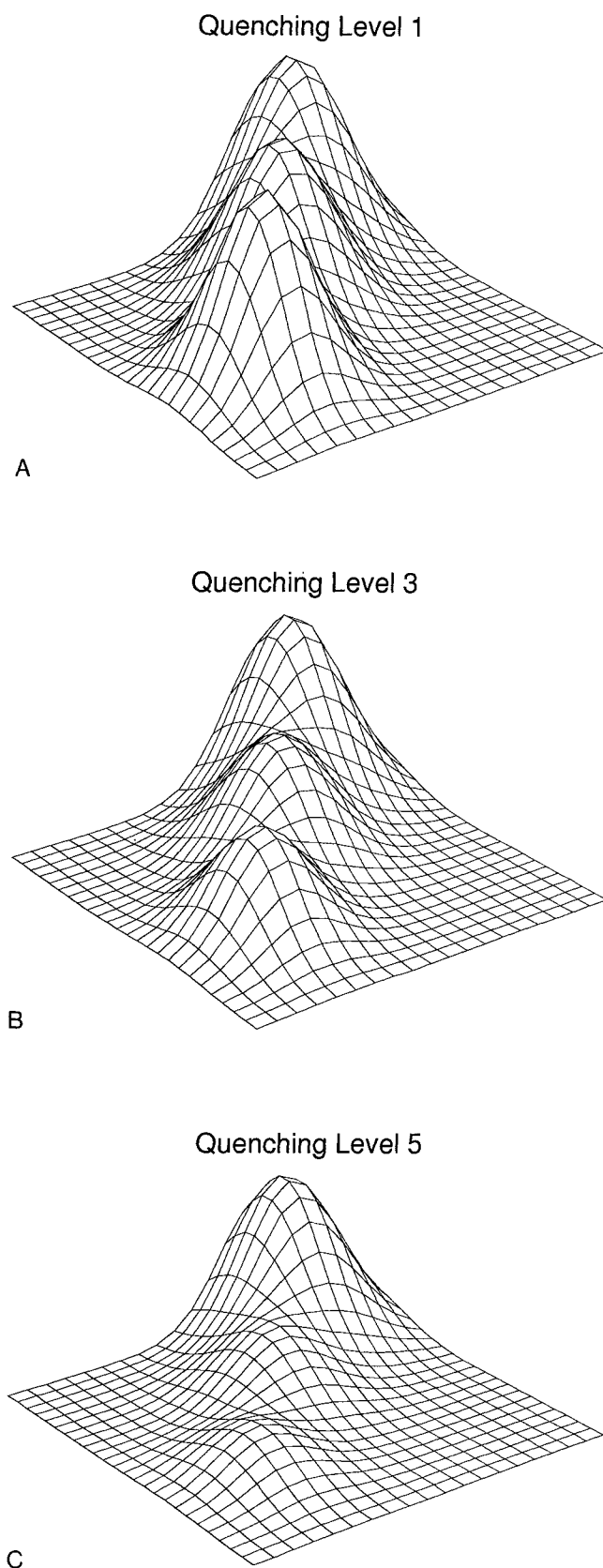
Quenching Level 5

C

FIG. 4. *Expected intensity as a function of emission and excitation wavelengths for three levels of chemistry for mathematically synthesized system. (A) Quenching level 1. (B) Quenching level 3. (C) Quenching level 5.*

Singular values of fluorescence matrices for 5 quenching levels corresponding to fixed levels of the third way

Singular value	Quenching level (k)				
	1	2	3	4	5
1	6.671×10^4	6.106×10^4	5.606×10^4	5.100×10^4	4.485×10^4
2	4.337×10^4	3.560×10^4	2.842×10^4	2.160×10^4	1.479×10^4
3	2.194×10^4	1.870×10^4	1.506×10^4	1.119×10^4	7.300×10^3
4	1.5×10^{-11}	7.9×10^{-12}	6.3×10^{-12}	5.4×10^{-12}	6.1×10^{-12}
	Ratio of first singular value to third singular value				
	3.04	3.27	3.72	4.56	6.14

Example 1, a three-way array without noise.

Since the singular vectors vary with k, although A and B do not, separate singular vector decompositions do not detect the shared structure of the 5 matrices μ_k. Moreover, we now show that the singular vectors from μ_k cannot be used to decompose μ_k' for $k' \neq k$. Observe that the coefficients $\gamma_f[k]$ could be recovered from μ_k, A and B by regressing the $25 * 20 = 500$ elements of μ_k on the three predictor variables determined by $a_f \times \beta_f$, $f = 1, 2, 3$. Formally, let Y_k be the vector obtained by stacking the columns of μ_k one above the other, so that $Y_k = vec(\mu_k)$. Take X to be the 500×3 matrix whose fth column is $vec(\beta_f \times a_f)$. By the construction of μ, the multiple correlation of Y_k with the columns of X is 1, for each k. To investigate how well the decomposition estimated from one level of the third way (k') can reconstruct the matrix for another level of the third way (k), we computed $R^2_{k',k}$, the multiple correlation coefficient of Y_k with the three columns of $X_{k'}$, defined by $X_{k'}[, f] = vec(\bar{V}_{k'}[, f] \times \bar{U}_{k'}[, f])$, $f = 1, 2, 3$.

These multiple correlation coefficients are provided in Table 3. The only multiple correlation coefficients that equal 1 are the diagonal elements of the table, elements that correspond to using the singular vectors from level k' to fit the matrix from the same level k'. The quality of the prediction of Y_k from the singular vectors of level k' decreases as k moves away from k'. The worst fit for each singular value decomposition k' is marked with an asterisk. The smallest multiple correlation coefficient is seen to increase with k'.

We see from this mathematical example that a trilinear model can recover structure that is not revealed when separate bilinear models are fit.

5.2 Example 2: A Mixture of Dyes

This example shows that the method works in a demonstration system. Three dyes were mixed. Fitting the PARAFAC model to data from the mixture accurately recovered the properties of the individual dyes. The original data from Lee (1988) (see also Lee, Kim and Ross, 1991) form a $25 \times 28 \times 6$ array, in which a

few entries were missing because the excitation and emission wavelengths are too close together to be able to measure fluorescence accurately. We have extracted the $16 \times 15 \times 6$ array that corresponds to excitation wavelengths from 440 nm to 560 in steps of 20; emission wavelengths from 570 to 610 in steps of 5, from 620 to 640 in steps of 10 and to 700 in steps of 20; and quenching levels 0, 15, 30, 50, 75 and 100 millimolar concentrations of potassium iodide.

Figure 6 shows perspective plots of the fluorescence data. There are several bumps visible, and their relative heights change with quenching level, but the exact number of bumps is less clear than in Figure 3. For fixed quenching level, the 6 induced matrices M_k are essentially rank 3 matrices, because the ordered singular values given in Table 4 settle down after the third singular value. The decrease is much less striking than with the mathematically synthesized data of Example 1. The singular vectors for the 6 arrays are displayed in Figure 7. While the 6 sets of vectors are remarkably similar, they are not identical. The multiple correlation coefficients of the responses observed at quenching level k with the predictors induced by the singular vectors obtained from quenching level k' are shown in Table 5.

We see that the multilinear model fits the laboratory array nearly as well as the multilinear model fits the mathematically synthesized array.

TABLE 3
Multiple correlations of fluorescence at level k with predictors from singular value decomposition at level k'

k'	Quenching level (k)				
	1	2	3	4	5
1	1.000	0.992	0.966	0.918	*0.843
2	0.994	1.000	0.993	0.967	*0.918
3	0.979	0.994	1.000	0.993	*0.967
4	*0.960	0.980	0.994	1.000	0.993

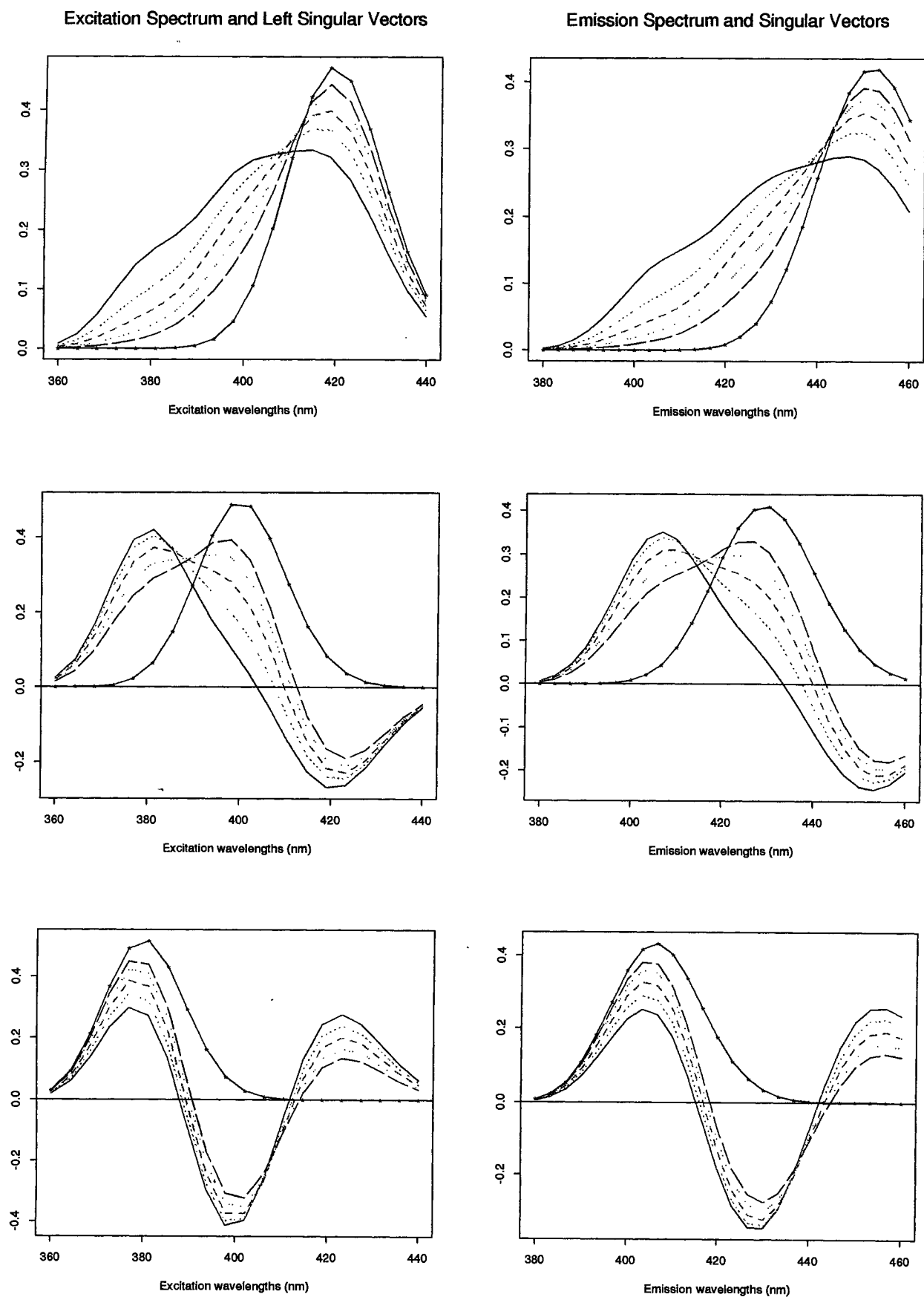Example 1, a three-way array without noise. The asterisk marks the lowest multiple correlation in each row.

Fig. 5. *Excitation and emission spectra for mathematically synthesized system with singular vectors from each level of chemistry superimposed.*
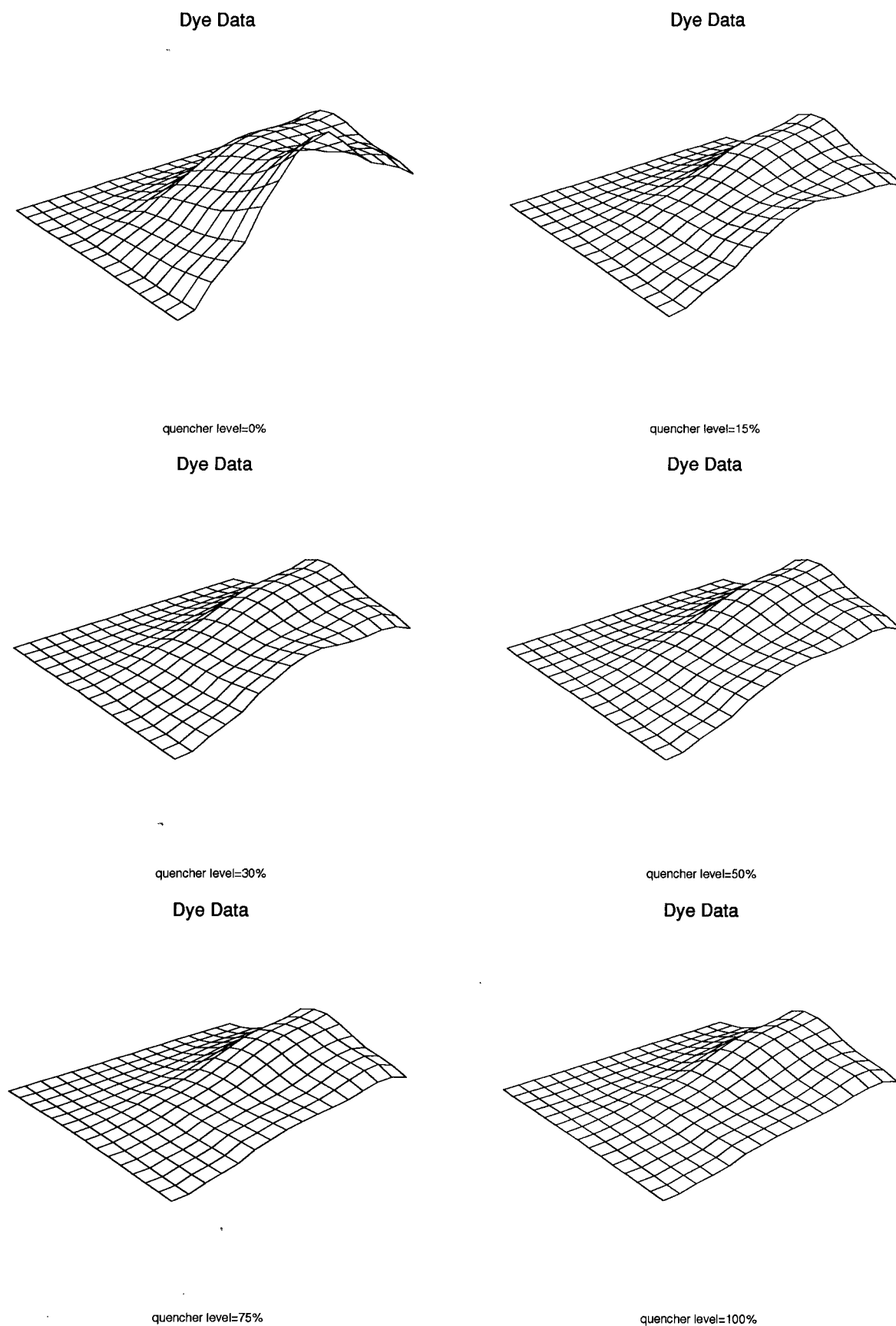
Dye Data   Dye Data

quencher level=0%   quencher level=15%

Dye Data   Dye Data

quencher level=30%   quencher level=50%

Dye Data   Dye Data

quencher level=75%   quencher level=100%

FIG. 6. *Observed intensity as a function of emission and excitation wavelengths for six concentrations of quencher for the mixture of dyes described in Section 5.2.*
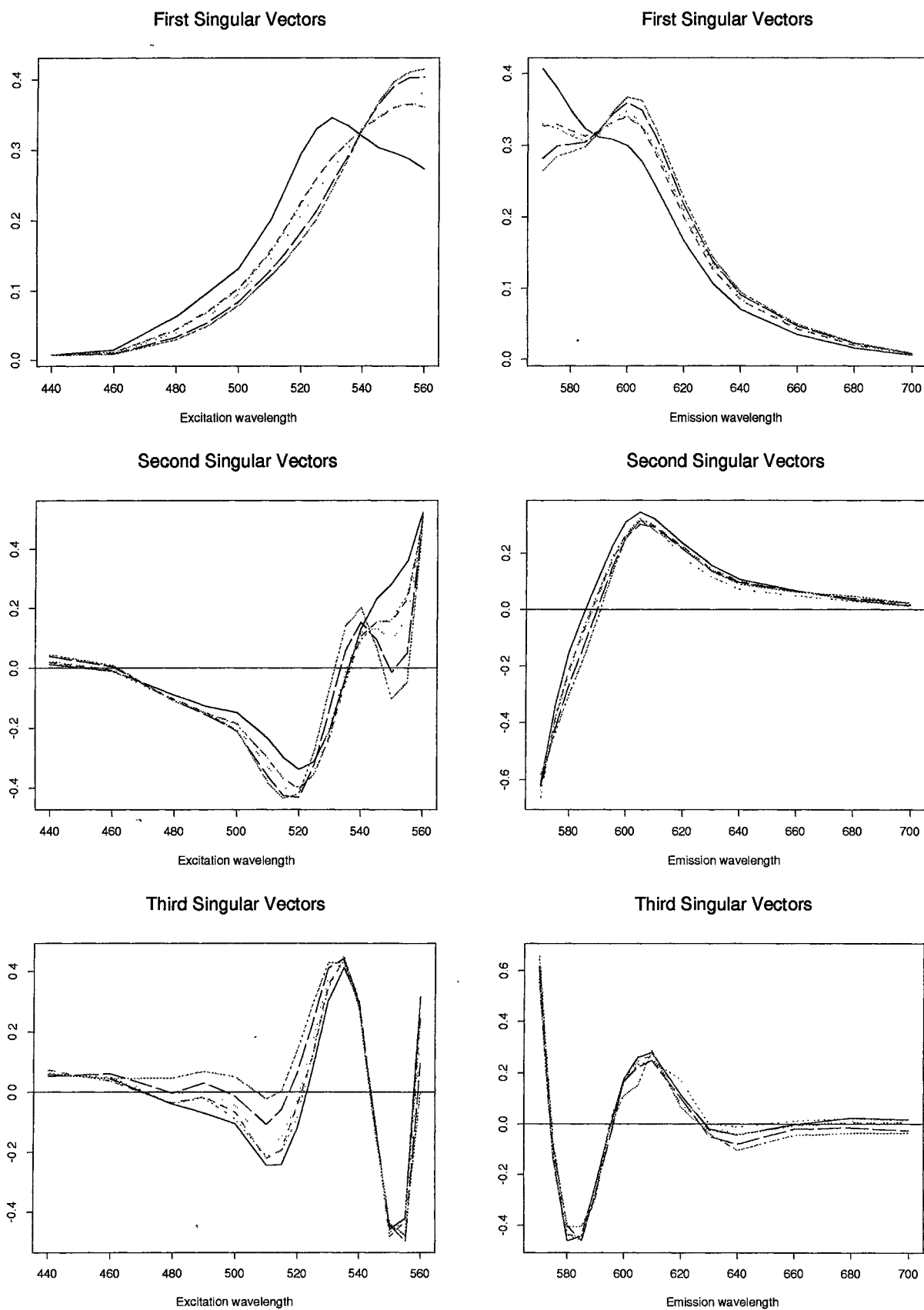
First Singular Vectors

Excitation wavelength

First Singular Vectors

Emission wavelength

Second Singular Vectors

Excitation wavelength

Second Singular Vectors

Emission wavelength

Third Singular Vectors

Excitation wavelength

Third Singular Vectors

Emission wavelength

FIG. 7. *Left and right singular vectors for all levels of quencher plotted against wavelength for mixture of dyes.*

TABLE 4

Singular values of fluorescence matrices for 6 quenching levels (k) corresponding to fixed levels of the third way

Singular value	Quenching level (k)					
	1	2	3	4	5	6
1	4.8387	3.0538	2.9512	2.8284	2.4325	2.1724
2	0.5661	0.3286	0.2776	0.2240	0.1482	0.1101
3	0.0474	0.0398	0.0366	0.0377	0.0337	0.0271
4	0.0039	0.0050	0.0041	0.0035	0.0043	0.0035
5	0.0034	0.0031	0.0026	0.0027	0.0030	0.0029
6	0.0026	0.0026	0.0023	0.0021	0.0028	0.0019

Example 2, mixture of dyes.

TABLE 5

Multiple correlations of fluorescence at level k with predictors from singular value decomposition at level k'

k'	Quenching level (k)					
	1	2	3	4	5	6
1	1.0000	0.9499	0.9527	0.9265	0.8810	*0.8515
2	*0.9508	1.0000	0.9994	0.9973	0.9851	0.9732
3	*0.9548	0.9994	1.0000	0.9971	0.9831	0.9700
4	*0.9312	0.9973	0.9971	1.0000	0.9941	0.9852
5	*0.8900	0.9855	0.9837	0.9942	1.0000	0.9980
6	*0.8612	0.9735	0.9708	0.9855	0.9980	1.0000

Example 2, mixture of dyes. The asterisk marks the lowest multiple correlation in each row.

5.3 Example 3: Components of a Biological System

Of the twenty amino acids most common in proteins, only tryptophan and tyrosine give off easily measurable amounts of fluorescence. In the experiments described here, excitation and emission wavelengths were chosen so that tyrosine fluorescence would not be observed, leaving tryptophan as the only expected source of fluorescence. A small protein may have two or three tryptophans, each of which has unique properties whose resolution provides information about its location in the protein. Alcohol dehydrogenase, the catalyst for the first reaction in the breakdown of alcohol, contains two tryptophans. Figure 8 shows the resolution of fluorescence from this protein into three components. Two correspond to the two tryptophans. One of them (amino acid number 15) is at the surface where it can be quenched by iodide; the other (number 314) is in the center of the protein, where it cannot be reached by the quencher. The third component was unexpected, and is due to impurities (Thampi, 1991).

This example demonstrates the resolution into three components of the fluorescence from amino acids in a protein. These spectra provide information about the structure of the intact protein, so that chemical separa-
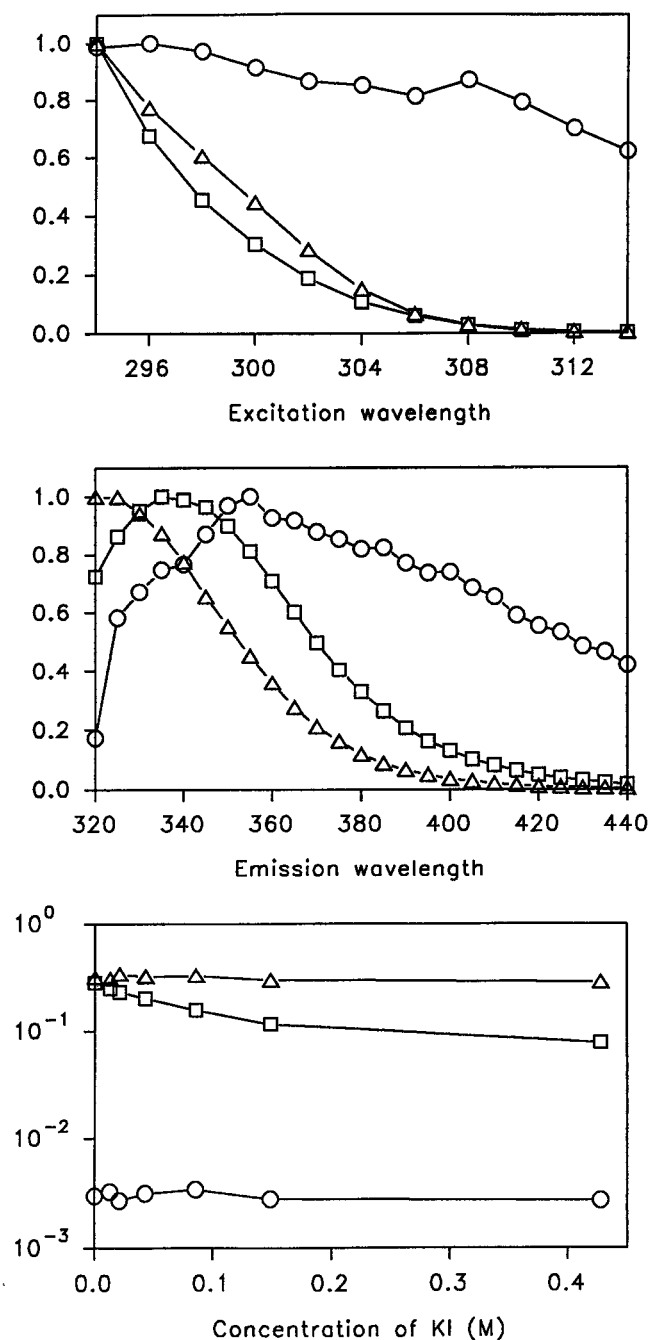

FIG. 8. *Resolution of fluorescence from the protein horse liver alcohol dehydrogenase into three components, identified as Tryptophan-314 (triangle), Tryptophan-15 (square), and impurity (circle).*

tion of the amino acids would have destroyed the system being studied.

6. CONCLUSION

Trilinear models can be used successfully to study complex specimens. Our interest in biological systems has caused us to emphasize mathematical resolution

of components that cannot be separated chemically. Analytical chemists are also interested in using multi-linear models to resolve spectra, in order to save the time and expense of chemical separation (Burdick et al., 1990). Kroonenberg (1983, 1991) cites other applications.

Figure 1 showed the absorption due to the photosynthetic pigments in a pea leaf. Here also, chemical separation would alter important properties of the system being studied. Our resolution of pea pigment fluorescence into two components was the first application of a trilinear model to the fluorescence spectra of a biological system (Ross et al., 1991). Our current work on photosynthetic pigments aims to resolve additional components, with excitation transfer between components represented by a T2 model.

Statistically, trilinear models behave differently from bilinear models. As seen in Section 4, although the parameters of some trilinear models are identifiable, the parameter estimates change when F increases. By contrast, the parameters of bilinear models are not estimable without mathematical constraints. And with these constraints, recursive estimation of the parameters as F increases is possible. Therefore the intuitions gained from work with bilinear models cannot be trusted as guides for multilinear models. The formal tractability of multilinear models should provide good test models for nonlinear least-squares theory. The attention given to multilinear models will increase; we hope this exposition will provide an especially concrete example of their value.

APPENDIX

This appendix contains three mathematical propositions referred to in the body of the paper. The statements and proofs are extracted from Leurgans (1991). The bilinear proof is in the first section. In the trilinear section, Lemma B.1 is used in the proof of Proposition B.1. The final section contains an explicit proof of a associativity property.

A. Bilinear Results

This section contains a self-contained argument that if every row of μ is a conditionally linear function of the corresponding row of A and every column of μ is a conditionally linear function of the corresponding row of B, then (36) holds. The notation here anticipates the notation for higher-way arrays.

PROPOSITION A.1. *If μ is $I \times J$, A is $I \times F_1$ and B is $J \times F_2$, and if*

$$\mu[i,] = M_1 A[i,], \qquad i = 1, \ldots, I,$$

$$\mu[,j] = M_2 B[j,], \qquad j = 1, \ldots, J,$$

where M_1 is a matrix function of B (independent of A)

and M_2 is a matrix function of A (independent of B), then there exists a conformable matrix C such that

$$(36) \qquad \mu = ACB^T.$$

PROOF. We first construct a matrix factorization for μ in terms of a matrix for each way and then show that we can deduce a factorization in terms of A and B.

Let R_1 (respectively R_2) be the dimension of $\mathcal{C}(\mu)$ ($\mathcal{C}(\mu^T)$), the column space of μ (of μ^T). By the definition of rank, there is an I by R_1 matrix $\overline{A}$ whose R_1 linearly independent columns span $\mathcal{C}(\mu)$. If $\overline{A}^\dagger$ denotes the Moore-Penrose generalized inverse of $\overline{A}$, then $\overline{A}\overline{A}^\dagger$ is the matrix that projects vectors in $\mathfrak{R}^I$ into $\mathcal{C}(\mu)$. Since the columns of μ are automatically in $\mathcal{C}(\mu)$, $\overline{A}\overline{A}^\dagger$ preserves the columns of μ, so that

$$\overline{A}\overline{A}^\dagger \mu = \mu.$$

Similarly, a $J \times R_2$ matrix $\overline{B}$ exists whose linearly independent columns are a basis for $\mathcal{C}(\mu^T)$, and $\overline{B}\overline{B}^\dagger \cdot \mu^T = \mu^T$. Combining these two equations gives $\overline{A}\overline{A}^\dagger \mu \overline{B}^{\dagger T} \cdot \overline{B}^T \mu = \mu$. Setting $\overline{C} = \overline{A}^\dagger \mu \overline{B}^{\dagger T}$, the displayed equation gives a matrix factorization for μ:

$$(37) \qquad \mu = \overline{A}\,\overline{C}\overline{B}^T.$$

The conclusion will now follow once $\overline{A}$ and $\overline{B}$ are written in terms of A and B.

To write $\overline{B}$ in terms of B, observe that the second hypothesis is that the columns of μ are linear combinations of the columns of B^T. Since the column space of a matrix is the set of linear combinations of its columns, the hypothesis is that the columns of μ belong to the column space of B^T, which implies that $\mathcal{C}(B^T)$ contains $\mathcal{C}(\mu)$. But the construction of $\overline{B}$ guarantees that $\mathcal{C}(\mu) = \mathcal{C}(\overline{B}^T)$, so each column of $\overline{B}^T$ must be a linear combination of columns of B^T, or that

$$\overline{B}^T = X_B B^T,$$

where the $R_2 \times F_2$ matrix X_B contains the coefficients for the linear combinations. A similar argument applied to $\mathcal{C}(\mu^T)$ shows that there exists an $R_1 \times F_1$ matrix of coefficients X_A such that

$$\overline{A} = AX_A.$$

Substituting for $\overline{A}$ and $\overline{B}$ in (37) and setting $C = X_A \overline{C} X_B^T$ completes the proof. $\square$

B. Trilinear Results

LEMMA B.1. *Let $\mathfrak{M}$ be a nonempty family of $I \times J \times K$ arrays. Let $\mathcal{C}_1$ denote the linear subspace of R^I generated by $\{\mu[, j, k], j = 1, \ldots, J; k = 1, \ldots, K; \mu \varepsilon \mathfrak{M}\}$, let $\mathcal{C}_2$ denote the linear subspace of $\mathfrak{R}^J$ generated by $\{\mu[i, , k], i = 1, \ldots, I; k = 1, \ldots, K; \mu\varepsilon\mathfrak{M}\}$, and let $\mathcal{C}_3$ denote the linear subspace of $\mathfrak{R}^K$ generated by $\{\mu[i, j,], i = 1, \ldots, I; j = 1, \ldots, J; \mu\varepsilon\mathfrak{M}\}$. Let $\overline{A}$ be any $I \times R_1$*

matrix whose R_1 linearly independent columns are a basis for $\mathfrak{C}_1$, let $\overline{B}$ be any $J \times R_2$ matrix whose R_2 linearly independent columns are a basis for $\mathfrak{C}_2$, and let $\overline{\Gamma}$ be any $K \times R_3$ matrix whose R_3 linearly independent columns are a basis for $\mathfrak{C}_3$. Define an $R_1 \times R_2 \times R_3$ array $\overline{C}$ in terms of the Moore-Penrose generalized inverses of $\overline{A}$, $\overline{B}$ and $\overline{\Gamma}$ by

$$\overline{C} = \{\overline{A}^\dagger, \overline{B}^\dagger, \overline{\Gamma}^\dagger\}*_3\mu.$$

Then every array $\mu \in \mathfrak{M}$ satisfies

$$\mu = \{\overline{A} \times \overline{B} \times \overline{\Gamma}\}*_3\, \overline{C}.$$

PROOF. Although this lemma is equivalent to the statement that the $R_1R_2R_3$ vectors $\overline{A}[, r_1] \times \overline{B}[, r_2] \times \overline{\Gamma}[, r_3]$ form a basis for the span of $\mathfrak{M}$, we provide an explicit proof based on an associativity property of $*_3$-multiplication, a property established for $*_N$-multiplication in Proposition C.1 below.

First observe that the $I \times I$ matrix $\overline{A}\overline{A}^\dagger$ projects $\mathfrak{R}^I$ into $\mathfrak{C}_1$. Since all vectors $\mu[, j, k]$ belong to $\mathfrak{C}_1$, they are invariant under projection:

$$\overline{A}\overline{A}^\dagger\mu[, j, k] = \mu[, j, k], \qquad j = 1, \ldots, J; k = 1, \ldots, K.$$

The JK equations above can be written as one equation using $*_3$:

(38) $$\{(\overline{A}\overline{A}^\dagger) \times I_J \times I_K\}*_3\, \mu = \mu.$$

By the same arguments, $\overline{B}\overline{B}^\dagger\, \mu[i, , k] = \mu[i, , k]$ and $\overline{\Gamma}\overline{\Gamma}^\dagger\, \mu[i, j,] = \mu[i, j,]$, or

(39) $$\{I_I \times (\overline{B}\overline{B}^\dagger) \times I_K\}*_3\, \mu = \mu$$

and

(40) $$\{I_I \times I_J \times (\overline{\Gamma}\overline{\Gamma}^\dagger)\}*_3\, \mu = \mu.$$

Substituting (39) and then (40) for μ in the left-hand side of (38) gives

$$\{(\overline{A}\overline{A}^\dagger) \times I_J \times I_K\}*_3(\{I_I \times (\overline{B}\overline{B}^\dagger) \times I_K\}*_3$$

$$(\{I_I \times I_J \times (\overline{\Gamma}\overline{\Gamma}^\dagger)\}*_3\, \mu)) = \mu.$$

But two applications of the associativity property reduce the three $*_3$-products to one:

$$\{(\overline{A}\overline{A}^\dagger) \times (\overline{B}\overline{B}^\dagger) \times (\overline{\Gamma}\overline{\Gamma}^\dagger)\}*_3\, \mu = \mu.$$

Applying associativity again to move the generalized inverses out of the 6-way array and into the 3-way array gives

$$\{\overline{A} \times \overline{B} \times \overline{\Gamma}\}*_3(\{\overline{A}^\dagger \times \overline{B}^\dagger \times \overline{\Gamma}^\dagger\}*_3\, \mu) = \mu.$$

The lemma now follows from substitution. □

PROPOSITION B.1 *If μ is a separately linear function of A, B and Γ such that $\mu[i, ,]$ depends on A only through $A[i,]$, $\mu[, j,]$ depends on B only through $B[, j,]$, and $\mu[, , k]$ depends on Γ only through $\Gamma[k,]$, then there exists a fixed $F_1 \times F_2 \times F_3$ array C such that*

$$\mu = \{A \times B \times \Gamma\}*_3\, C$$

for all $\mu \in \mathfrak{M}$.

PROOF. By Lemma B.1, $\mu = \{\overline{A} \times \overline{B} \times \overline{\Gamma}\}*_3\, \overline{C}$, where $\overline{\Gamma}$ is a basis for $\mathfrak{C}_3$, which contains all of the vectors $\mu[i, j,]$. But the parallel conditional linearity hypotheses imply that $\mu[i, j, k] = \Gamma[k,]^T v$, where v is an $F_3 \times 1$ vector function of $A[i,]$ and $B[, j,]$. Stacking the K equations gives $\mu[i, j,] = \Gamma v$, which shows that each vector $\mu[i, j,]$ is a linear combination of columns of Γ, so that $\mathfrak{C}_3$ must be a subset of $\mathfrak{C}(\Gamma)$. Since the columns of $\overline{\Gamma}$ are thus a basis for a subspace of the span of the columns of Γ, there must exist an $F_3 \times R_3$ matrix X_Γ such that $\overline{\Gamma} = \Gamma X_\Gamma$. Similarly, $\overline{B} = BX_B$ and $\overline{A} = AX_A$, so that substitution in the expression for μ gives

$$\mu = \{(AX_A) \times (BX_B) \times (\Gamma X_\Gamma)\}*_3\, \overline{C}.$$

By the associativity properties of $*_3$-multiplication, if the array C is set equal to $\{X_A \times X_B \times X_\Gamma\}*_3\ \overline{C}$, then

$$\mu = \{A \times B \times \Gamma\}*_3\overline{C}.$$

The array C cannot depend on A, B or Γ, because otherwise one of the conditional linearity hypotheses would fail. □

C. Associativity Property

The $*_N$-product of exchanged outer product arrays and N-way arrays has a useful associativity formula: If an N-way array A_1 is equal to the $*_N$-product of a $2N$-way exchanged outer product array and a second N-way array A, then the $*_N$-product of a second $2N$-way array with the first N-way array A_1 is equal to one $*_N$-product of a third exchanged outer product array with A. The matrices of the third outer product array are the ordinary matrix products of the corresponding matrices from the two exchanged outer product arrays. The extension of Proposition C.1 to arrays A with order greater than N is an immediate corollary. The lemma is the key to the proofs that multilinearity is equivalent to a $*_N$-product form.

PROPOSITION C.1 (Associativity of $*_N$-Products). *If each pair of matrices (M_n, N_n), $n = 1, \ldots, N$ is conformable for matrix multiplication and if $\{\times_{n=1}^N N_n\}$ is conformable for $*_N$-multiplication with the N-way array A, then*

(41)
$$\{\times_{n=1}^N M_n\}*_N(\{\times_{n=1}^N N_n\}*_N A)$$
$$= \{\times_{n=1}^N (M_n N_n)\}*_N A.$$

PROOF. The proof is an exercise in notation and in the distributive properties of addition and multiplication of real numbers. Let the dimension of N_n be (J_n, K_n). The lemma will be established by evaluating each side of (42) at a typical N-tuple of indices.

We begin by evaluating the formulas for the $2N$-way arrays of the left-hand side at a typical N-tuple of indices. A typical element of the exchanged product of the matrices M_n is

$$(42) \quad \{\times_{n=1}^N M_n\}[i_1, \dots, i_N, j_1, \dots, j_N] = \prod_{n=1}^N M_n[i_n, j_n],$$

and the exchanged product of the matrices N_n is

$$\{\times_{n=1}^N N_n\}[j_1, \dots, j_N, k_1, \dots, k_N] = \prod_{n=1}^N N_n[j_n, k_n].$$

So the second $*_N$-product on the left in (42) has the general form

$$\{\times_{n=1}^N N_n\}*_N A$$
$$(43) \qquad = \sum_{k_1, \dots, k_N} (\times_{n=1}^N N_n)[, \dots, k_1, \dots, k_N]$$
$$\cdot A[k_1, \dots, k_N]$$

and typical element

$$(\{\times_{n=1}^N N_n\}*_N A)[j_1, \dots, j_N]$$
$$= \sum_{k_1=1}^{K_1} \cdots \sum_{k_N=1}^{K_N} \left(\prod_{n=1}^N N_n[j_n, k_n] \right) A[k_1, \dots, k_N].$$

Writing out the definition of the left-most $*_N$-product in (41), the $(i_1, \dots, i_n)^{th}$ element of the left-hand side is

$$\sum_{j_1=1}^{J_1} \cdots \sum_{j_N=1}^{J_N} (\{\times_{n=1}^N M_n\}[i_1, \dots, i_N, j_1, \dots, j_N])$$
$$(\{\times_{n=1}^N N_n\}*_N A[j_1, \dots, j_N]).$$

Substitution of (42) and (43) in the equation above gives

$$\sum_{j_1=1}^{J_1} \cdots \sum_{j_N=1}^{J_N} \left(\prod_{n=1}^N M_n[i_n, j_n] \right)$$
$$\cdot \sum_{k_1=1}^{K_1} \cdots \sum_{k_N=1}^{K_N} \left(\prod_{n=1}^N N_n[j_n, k_n] \right) A[k_1, \dots, k_N].$$

Moving the product of elements of the M_n's inside the summation in the vector of indices k yields

$$\sum_j \sum_k \left(\left(\prod_{n=1}^N M_n[i_n, j_n] \right) \left(\prod_{n'=1}^N N_{n'}[j_{n'}, k_{n'}] \right) A[k_1, \dots, k_N] \right).$$

Now we combine the two products (Π) into one:

$$\sum_j \sum_k \left(\prod_{n=1}^N M_n[i_n, j_n] N_n[j_n, k_n] \right) A[k_1, \dots, k_N].$$

Interchanging the product in n and the sum in j gives

$$\sum_k \left(\prod_{n=1}^N \left(\sum_{j_n=1}^{J_N} M_n[i_n, j_n] N_n[j_n, k_n] \right) \right) A[k_1, \dots, k_N].$$

Since the summation in j_n is just an explicit formula for the (i_n, k_n) element of the matrix product $M_n N_n$, the expression above reduces to

$$\sum_k \left(\prod_{n=1}^N (M_n N_n)[i_n, k_n] \right) A[k_1 \dots, k_N].$$

The product in n is an element of the exchanged N-fold outer product of the matrices (M_n, N_n), giving

$$\sum_k (\times_{n=1}^N \{M_n N_n\})[i_1, \dots, i_N, k_1, \dots, k_N] A[k_1, \dots, k_N].$$

This last expression is the right side of (41), so the proof is now complete. $\square$

ACKNOWLEDGMENTS

REFERENCES

APPELLOF, C. J. and DAVIDSON, E. R. (1981). Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents. *Analytical Chemistry* 53 2053-2056.

BELLMAN, R. E. (1970). *Introduction to Matrix Analysis*, 2nd ed. McGraw-Hill, New York.

BROWN, S. D. (1990). Chemometrics. *Analytical Chemistry* 62 84R-101R.

BROWN, S. D., BARKER, T. Q., LARIVEE, R. J., MONFRE, S. L. and WILK, H. R. (1988). Chemometrics. *Analytical Chemistry* 60 252R-273R.

BURDICK, D. S., TU, X. M., McGOWN, L. B. and MILLICAN, D. W. (1990). Resolution of multicomponent fluorescent mixtures by analysis of the excitation-emission-frequency array. *Journal of Chemometrics* 4 15-28.

CANTOR, C. R. and SCHIMMEL, P. R. (1980). *Biophysical Chemistry*. W. H. Freeman, San Francisco.

CARROLL, J. D. and CHANG, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decompositions. *Psychometrika* 35 283-319.

DURELL, S. R., LEE, C.-H., ROSS, R. T. and GROSS, E. L. (1990). Factor analysis of the near-ultraviolet absorption spectrum of plastocyanin using bilinear, trilinear and quadrilinear models. *Archives of Biochemistry and Biophysics* 278 148-160.

ECKART, C. and YOUNG, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika* 1 211-218.

GNANADESIKAN, R. and KETTENRING, J. R. (1984). A pragmatic

review of multivariate methods in applications. In *Statistics: An Appraisal* (H. A. David and H. T. David, eds.) 309–337. Iowa State Univ. Press.

HARSHMAN, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "exploratory" multi-mode factor analysis. *UCLA Working Papers in Phonetics* **16** 1–84.

INGLE, JR., J. and CROUCH, S. R. (1988). *Spectrochemical Analysis.* Prentice Hall, Englewood Cliffs, NJ.

KAPTEYN, A., NEUDECKER, H. and WANSBEEK, T. (1986). An approach to n-mode components analysis. *Psychometrika* **51** 269–275.

KROONENBERG, P. M. (1983). Annotated bibliography of three-mode data by means of alternating least-squares algorithms. *British J. Math. Statist. Psych.* **36** 81–113.

KROONENBERG, P. M. (1991). The three-mode side of the three-way world. Technical report, Dept. Education, Leiden Univ.

KRUSKAL, J. B. (1977). Three-way arrays: Rank and uniqueness of trilinear decompositions with application to arithmetic complexity and statistics. *Linear Algebra Appl.* **18** 95–138.

LASTOVICKA, J. L. (1981). The extension of component analyses to 4-mode factor analysis. *Psychometrika* **46** 47–57.

LEE, C.-H. (1988). Multilinear analysis of fluorescence spectra of photosynthetic systems. Ph.D. dissertation, Ohio State Univ.

LEE, C.-H., KIM, K. and ROSS, R. T. (1991). Trilinear analysis for the resolution of overlapping fluorescence spectra. *Korean Biochemistry Journal* **24** 374–379.

LEURGANS, S. E. (1991). Multilinear models: Array formulas. Technical Report 473, Dept. Statistics, Ohio State Univ.

LEURGANS, S. E., ROSS, R. T. and ABEL, R. B. (1992). A decomposition for 3-way arrays. *SIAM J. Matrix. Anal. Appl.* To appear.

MALINOWSKI, E. R. and HOWERY, D. G. (1980). *Factor Analysis in Chemistry.* Wiley, New York.

NDOU, T. T. and WARNER, I. M. (1991). Applications of multidimensional absorption and luminescence spectroscopies in analytical chemistry. *Chemical Reviews* **91** 493–507.

PRINGLE, R. M. and RAYNER, A. A. (1971). *Generalized Inverse Matrices with Applications to Statistics.* Hafner, New York.

RAMOS, L. S., BEEBE, K. R., CAREY, W. P., SANCHEZ, E., ERICKSON, B. C., WILSON, B. E., WANGEN, L. E. and KOWALSKI, B. R. (1986). Chemometrics. *Analytical Chemistry* **58** 294R–315R.

ROSS, R. T., LEE, C.-H., DAVIS, C. M., FAYYAD, E. A. and LEURGANS, S. E. (1991). Resolution of fluorescence spectra of plant pigment-complexes using trilinear models. *Biochimica et Biophysica Acta* **1056** 317–320.

RUSSELL, M. D. and GOUTERMAN, M. (1988a). Excitation-emission-lifetime analysis of multicomponent systems—I. Principal component factor analysis. *Spectrochimica Acta* **44A** 857–861.

RUSSELL, M. D. and GOUTERMAN, M. (1988b). Excitation-emission-lifetime analysis of multicomponent systems—II. Synthetic model data. *Spectrochimica Acta* **44A** 863–872.

RUSSELL, M. D., GOUTERMAN, M. and VAN ZEE, J. A. (1988). Excitation-emission-lifetime analysis of multicomponent systems—III. Platinum, palladium and rhodium porphyrins. *Spectrochimica Acta* **44A** 873–882.

SANCHEZ, E. and KOWALSKI, B. R. (1988a). Tensorial calibration: I. First-order calculations. *Journal of Chemometrics* **2** 247–263.

SANCHEZ, E. and KOWALSKI, B. R. (1988b). Tensorial calibration: II. Second-order calibration. *Journal of Chemometrics* **2** 265–290.

SANCHEZ, E. J. and KOWALSKI, B. R. (1990). Tensorial resolution: A direct trilinear decomposition. *Journal of Chemometrics* **4** 29–45.

SANDS, R. and YOUNG, F. W. (1980). Component models for three-way data: An alternating least squares algorithm with optimal scaling features. *Psychometrika* **45** 39–67.

THAMPI, S. (1991). Resolution of intrinsic protein fluorescence using multilinear analysis. Master's thesis, Ohio State Univ.

TUCKER, L. R. (1963). Implications of factor analysis of three-way matrices for measurement of change. In *Problems in Measuring Change* (C. W. Harris, ed.) 122–137. Univ. Wisconsin Press.

Comment

Jan deLeeuw

INTRODUCTION

, Complicated multivariate models, and certainly the models used in multidimensional scaling, are most often used for exploratory purposes. The paper by Leurgans and Ross covers one of the fortunate, but rather exceptional, situations in which we can derive the form of the model from prior scientific knowledge. Another, similar, situation is the conformation of molecules using scaling techniques, and the seriation of artifacts in time or of genes along a chromosome. In this class of

Jan deLeeuw is Professor of Mathematics and Psychology, Director of the Social Statistics Program, Department of Mathematics, University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, California 90024-1555. e-mail: deleeuw@laplace.sscnet.ucla.edu.

applications, the physical information we have tells us that a multilinear model is appropriate—only the coefficients (mixtures) and dimensionality (number of components) are unknown and must be estimated.

OTHER AREAS

Leurgans and Ross discuss the multilinear models pretty much in the context in which they use them. Thus, it looks a bit as if these models were created for spectroscopy. This is perfectly appropriate in a paper such as this, which concentrates on a particular field of application. But to give a somewhat wider perspective, I'll list a number of other areas, both mathematical and nonmathematical, in which multilinear models have been studied or applied.

1. *Efficient computation of matrix products and*