SIAM J. MATRIX ANAL. APPL.
Vol. 14, No. 4, pp. 1064–1083, October 1993

011

A DECOMPOSITION FOR THREE-WAY ARRAYS*

S. E. LEURGANS†, R. T. ROSS‡, AND R. B. ABEL§

Abstract. An I-by-J-by-K array has rank 1 if the array is the outer product of an I-, a J-, and a K-vector. The authors prove that a three-way array can be uniquely decomposed as the sum of F rank-1 arrays if the F vectors corresponding to two of the ways are linearly independent and the F vectors corresponding to the third way have the property that no two are collinear. Several algorithms that implement the decomposition are described. The algorithms are applied to obtain initial values for nonlinear least-squares calculations. The performances of the decompositions and of the nonlinear least-squares solutions on real and on simulated data are compared. An extension to higher-way arrays is introduced, and the method is compared with those of other authors.

Key words. alternating least-squares algorithm, array rank, multiway arrays

AMS subject classifications. 15A23, 15A69, 62H25, 62J99

1. Introduction.

This paper gives a method for identifying the parameters of a multilinear model proposed by Harshman [11] and by Carroll and Chang [4]. The models, described in §3 below, represent three-way arrays in terms of three sets of parameters, one set associated with each way. To work with these models, it is necessary to be able to identify the parameters for each way from the three-way array. Kruskal [14] gives conditions on the sets of parameters such that the parameters are identifiable, but his paper does not include a construction. Assuming conditions stronger than his weakest conditions, we show how to recover the parameters from the array. This method is related to the algebraic initialization method of Sands and Young [25]. Our method requires one eigenvector decomposition, one Moore–Penrose generalized inverse, and a finite number of arithmetic steps. These models have been applied recently in the chemical literature, as described in §7 below. In this paper we concentrate on a mathematical presentation to make these ideas more widely applicable. Using Moore–Penrose inverses gives a natural derivation, and we provide some evidence on the effectiveness of an appropriate decomposition for initializing alternating least-squares iterations.

Section 2 reviews some facts about matrices, or two-way arrays. The models for three-way arrays of interest to us are outlined in §3. We require a decomposition that returns the parameters ($\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{\Gamma}$ in the notation of §3) from a three-way array $\boldsymbol{\mu}$. Section 3 gives identifiability conditions under which a theorem is proven in §4. The theorem characterizes the parameters of the model. Section 5 shows how the theorem can be applied to give an algorithm. The section also reports some of our experience in applying this decomposition to initialize alternating least-squares calculations. We summarize the behavior of the decompositions and of the least-squares calculations for

* Received by the editors September 10, 1990; accepted for publication (in revised form) January 1, 1992. This research was supported by National Science Foundation grants DMS-88-05402 and DMS-89-02265 at Ohio State University and by the Ohio Supercomputer Center. Revision of the manuscript was supported by National Science Foundation grant HRD-9103314 while the first author was visiting the University of Chicago.

† Department of Statistics, Ohio State University, Columbus, Ohio 43210-1247 (leurgans@is.rpslmc.edu).

‡ Department of Biochemistry, Ohio State University, Columbus, Ohio 43210-1229 (rtr+@ohio-state.edu).

§ National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892-9905.

five sets of real data and a few sets of artificial data. In §6 we discuss decompositions for higher-way arrays. Section 7 compares our work with other papers.

The models of §3 can be interpreted as a representation of an array as the sum of rank-1 arrays. Since rank-1 arrays are often easy to interpret, these models can lead to simple interpretations of the array. For two-way arrays, or matrices, such representations are never unique without side conditions that often interfere with the simplicity of the interpretation. Frequently the condition that the vectors for each way be an orthogonal set is imposed. Since negative parameters have no physical meaning in many applications of these models, orthogonality cannot always provide representations that are interpretable unless the array is a rank-1 array. However, if the original array has more than two ways, the array often has a unique representation as the sum of rank-1 arrays. As has been emphasized in the literature on these models, side conditions, such as orthogonality, are not required. The sufficient conditions of §3 permit representations that do not sacrifice interpretability. Since representations for three-way arrays can lead to interpretable parameters, experiments that use three independent variables to generate three ways of an array can remove ambiguities unavoidable in experiments that use only two of the independent variables.

2. Decompositions of matrices. The decompositions for three-way arrays rely on results for two-way arrays, or matrices. The results we will use have been collected in this section. We state one proposition concerning the Moore–Penrose generalized inverse of products of matrices. This proposition implies that the Moore–Penrose generalized inverse can be obtained from a singular value decomposition.

A matrix has rank 1 if it is the outer product of one pair of vectors. An I-by-J matrix $\boldsymbol{\mu}$ has rank F if F is the smallest integer such that $\boldsymbol{\mu}$ is a sum of rank-1 matrices, or

$$\boldsymbol{\mu} = \sum_{f=1}^{F} \boldsymbol{\alpha}_f \times \boldsymbol{\beta}_f = \boldsymbol{A}\boldsymbol{B}^T, \tag{1}$$

where the columns of $\boldsymbol{A}$ are $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_F$ and those of $\boldsymbol{B}$ are $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_F$ and $\times$ denotes the outer product defined by $\boldsymbol{\alpha}_f \times \boldsymbol{\beta}_f$, the I-by-J matrix whose i, j entry is $\alpha_f[i]\beta_f[j]$. The model (1) is a bilinear model for $\boldsymbol{\mu}$. To avoid trivial reparameterizations we require the columns of $\boldsymbol{A}$ to be unit vectors whose first nonzero element is positive. When $F = 1$, the model is identifiable up to the sign convention: $\boldsymbol{\alpha}$ is any column of $\boldsymbol{\mu}$ divided by its length and $\boldsymbol{\beta}$ is the vector of the lengths of the columns, where the jth component is multiplied by the sign of the first nonzero element of the column. When $F > 1$, the parameters are not identifiable because $\boldsymbol{A}\boldsymbol{Q}(\boldsymbol{B}\boldsymbol{Q})^T = \boldsymbol{A}\boldsymbol{B}^T$ when $\boldsymbol{Q}$ is any orthogonal F-by-F matrix preserving the sign convention for $\boldsymbol{A}$. This nonidentifiability leads to the well-known rotation problem in factor analysis. If the columns of $\boldsymbol{A}$ and $\boldsymbol{B}$ are forced to be orthonormal, then the singular-value decomposition is the unique (up to permutation) decomposition of $\boldsymbol{\mu}$, provided that the singular values are unique [12, p. 414]. The singular-value decomposition of an I-by-J matrix $\boldsymbol{M}$ with rank F is an I-by-F matrix $\boldsymbol{U}$ with orthonormal columns, a J-by-F matrix $\boldsymbol{V}$ with orthonormal columns, and an F-by-F diagonal matrix $\boldsymbol{S}$ with nonincreasing positive diagonal elements satisfying $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$. The diagonal elements of $\boldsymbol{S}$ are the singular values of $\boldsymbol{M}$. We follow Greenacre [8] by taking singular values to be nonzero. The columns of $\boldsymbol{U}$ are the left singular vectors of $\boldsymbol{M}$, and the columns of $\boldsymbol{V}$ are the right singular vectors of $\boldsymbol{M}$. Mathematically, the requirement that the columns of $\boldsymbol{A}$ be mutually orthogonal is very natural, but practical interpretation is

frequently compromised.

The Moore–Penrose generalized inverse is a natural extension of matrix inversion to matrices that may be singular or even rectangular. (For defining properties see, for example, [26], [8], [20].) We use $\boldsymbol{M}^+$ to denote the Moore–Penrose generalized inverse of the matrix $\boldsymbol{M}$. We state a property of Moore–Penrose generalized inverses as a proposition.

PROPOSITION 2.1. *If $\boldsymbol{A}$ and $\boldsymbol{B}$ have equal full column rank and if the square matrix $\boldsymbol{D}$ is nonsingular, then*

$$(\boldsymbol{ADB}^T)^+ = (\boldsymbol{B}^+)^T\boldsymbol{D}^{-1}\boldsymbol{A}^+. \tag{2}$$

This proposition follows from Greville [9] and is given as Corollary 5 of Theorem 2.16 of Pringle and Rayner [20, p. 31]. Direct verification follows from routine applications of the defining properties of Moore–Penrose generalized inverses. If the hypotheses fail, Greville's results show that (2) can be false.

3. PARAFAC models for three-way arrays. In this section we describe the models of interest. Interpretation of these models for three-way arrays is much easier when the parameters can be identified from the array. We give three identifiability conditions that are sufficient for identifiability if three parameterization conventions are also imposed. The parameterization conventions can (essentially) always be imposed whenever the model holds. The conventions remove some trivial nonidentifiabilities. The results of Kruskal [14], discussed after the identifiability conditions are introduced, guarantee that IC1, IC2, and IC3 are sufficient for identifiability of all the parameters. Kruskal also provides weaker conditions, although his paper does not construct the parameters from the array. Since §5 provides such a construction, the identifiability of the parameters is verified. This section concludes with a presentation of some properties of the models.

The same models for three-way arrays were used in methods proposed independently by Harshman [10] and by Carroll and Chang [4]. The former paper calls the models *parallel factor models*, or PARAFAC, and the latter paper refers to *canonical decompositions*, or CANDECOMP. Both acronyms correspond to computer programs. For recent discussions of the methods in their original contexts, see [15], [2], [7]. Appellof and Davidson [1] first applied these models to spectroscopy.

The F-factor PARAFAC model for an I-by-J-by-K three-way array $\boldsymbol{\mu}$ is that $\boldsymbol{\mu}$ satisfies the *trilinear model*:

$$\boldsymbol{\mu} = \sum_{f=1}^{F} \boldsymbol{\alpha}_f \times \boldsymbol{\beta}_f \times \boldsymbol{\gamma}_f. \tag{3}$$

Since each array $\boldsymbol{\alpha}_f \times \boldsymbol{\beta}_f \times \boldsymbol{\gamma}_f$ is a rank-1 array, an F-term trilinear model (3) holds if $\boldsymbol{\mu}$ is the sum of F rank-1 arrays [14]. Since the trilinear model is preserved if $\boldsymbol{\alpha}_f$, $\boldsymbol{\beta}_f$, and $\boldsymbol{\gamma}_f$ are multiplied by constants whose product is 1 and since the values of f can be permuted without changing the sum, some parameterization conventions are necessary to provide identifiability. Before introducing either the identifiability conditions or a set of parameterization conventions, we rewrite the trilinear model in terms of matrices.

The trilinear model (3) can be written in terms of the K I-by-J matrices $\boldsymbol{\mu}_k$ as

follows:

$$\boldsymbol{\mu}_k = \boldsymbol{\mu}[,,k] = \sum_{f=1}^{F} \boldsymbol{\alpha}_f \times \boldsymbol{\beta}_f \gamma_f[k] = \sum_{f=1}^{F} \boldsymbol{\alpha}_f \gamma_f[k] \boldsymbol{\beta}_f^T = \boldsymbol{A}\boldsymbol{D}_k\boldsymbol{B}^T,$$

where $\boldsymbol{A}$ is a real matrix whose columns are the $\boldsymbol{\alpha}_f$'s, $\boldsymbol{B}$ is a real matrix whose columns are the $\boldsymbol{\beta}_f$'s, and $\boldsymbol{D}_k$ is the F-by-F diagonal matrix whose diagonal elements are $\gamma_f[k], f = 1, \ldots, F$. Kruskal [14] refers to the $\boldsymbol{\mu}_k$'s as 3-slabs of the array $\boldsymbol{\mu}$, and we refer to them as *slabs* below. Note that the diagonal of $\boldsymbol{D}_k$ is the kth row of $\boldsymbol{\Gamma}$. Therefore, the PARAFAC model for the array $\boldsymbol{\mu}$ is equivalent to requiring that the K matrices $\boldsymbol{\mu}_k$ can be factored into the product of an I-by-F, an F-by-F, and an F-by-J matrix in such a way that the K I-by-F matrices are identical, the K F-by-J matrices are identical, and the K F-by-F matrices are all diagonal. The diagonal matrices $\boldsymbol{D}_k$ vary with k and can be thought of as reflecting the relative importance of the $\boldsymbol{\alpha}_f$'s and the $\boldsymbol{\beta}_f$'s as k varies. The equation displayed above can be summarized as *trilinear matrix equations* (4):

$$\boldsymbol{\mu}_k = \boldsymbol{A}\boldsymbol{D}_k\boldsymbol{B}^T, \qquad k = 1, \ldots, K. \tag{4}$$

The trivial nonidentifiabilities alluded to in the preceding paragraph correspond to modification of $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{\Gamma}$ by multiplication on the right by diagonal matrices whose product is the identity matrix or by applying the same permutation to their columns.

The trivial sources of nonidentifiability are not the only ones: Conditions on $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{\Gamma}$ are required. One set of sufficient conditions is that all three matrices have the property that every pair of columns is linearly independent and that two of the matrices have all columns linearly independent. For specificity, we adopt the following *identifiability conditions*:

(IC1) The columns of $\boldsymbol{A}$ are linearly independent.
(IC2) The columns of $\boldsymbol{B}$ are linearly independent.
(IC3) Every pair of columns of $\boldsymbol{\Gamma}$ is linearly independent.

In §4, we show how to recover the parameters under these conditions.

Of course, when these three conditions hold, all F columns of $\boldsymbol{A}$ are linearly independent, all F columns of $\boldsymbol{B}$ are linearly independent, and every pair of columns of $\boldsymbol{\Gamma}$ are linearly independent. In Kruskal's notation, $I_o \geq F$, $J_o \geq F$, and $K_o \geq 2$, so that $I_o + J_o + K_o \geq 2F + 2$. This last inequality is sufficient for the parameters to be uniquely defined, by [14, Thm. 4a]. The same theorem implies that uniqueness can still hold when IC1 and IC2 fail if IC3 is strengthened. For example, if the last column of $\boldsymbol{A}$ is the sum of the preceding mutually linearly independent columns, if the last column of $\boldsymbol{B}$ is the sum of the preceding mutually linearly independent columns, and if every four columns of $\boldsymbol{\Gamma}$ are linearly independent, then Kruskal's results imply uniqueness, although our decomposition does not apply.

The identifiability conditions are also easy to interpret because each of the three conditions concerns the parameters associated with one way of the array. The first two identifiability conditions assert that there must truly be F factors present in two ways. The third condition requires that no two underlying factors respond to the other way in an exactly proportional manner. If $K > 1$, IC3 can hold when $F > K$. Observe that if $K = 1$, IC3 can hold only if $F = 1$. This observation corresponds to the nonidentifiability of bilinear models if $F > 1$. The conditions IC1 and IC2, however, can hold only if $F \leq I$ and $F \leq J$.

To see that condition IC3 cannot be omitted, suppose that $F = 2$ and that $\gamma_1 = c_1\gamma$ and $\gamma_2 = c_2\gamma$. Then the trilinear model implies that

$$\mu = \alpha_1 \times \beta_1 \times c_1\gamma + \alpha_2 \times \beta_2 \times c_2\gamma = \{c_1\alpha_1 \times \beta_1 + c_2\alpha_2 \times \beta_2\} \times \gamma.$$

The I-by-J matrix in braces is

$$(c_1\alpha_1 \vdots c_2\alpha_2)(\beta_1 \vdots \beta_2)^T = (c_1\alpha_1 \vdots c_2\alpha_2)MM^{-1}(\beta_1 \vdots \beta_2)^T,$$

where M is any nonsingular 2-by-2 matrix. The arbitrariness of M implies that A and B cannot be identified. These calculations can be extended to show that failure of IC3 always prevents identifiability of the columns of A and of B corresponding to the collinear columns of Γ.

The following *parameterization conventions* remove these trivial sources of nonidentifiability:

(PC1) The columns of A and of B are unit vectors.

(PC2) In each column of A, the element with largest magnitude is positive. If several elements in a column have magnitude equal to the largest magnitude, then the first such element is positive.

(PC3) The diagonal elements of

$$D_+ = \sum_{k=1}^{K} D_k$$

are nonincreasing and positive, that is,

$$1_K^T\gamma_1 = D_+[1,1] \geq \cdots \geq 1_K^T\gamma_f = D_+[f,f] \geq \cdots 1_K^T\gamma_F = D_+[F,F] > 0.$$

The parameterization conventions are easy to interpret. Convention PC1 asserts that the constant is carried in the third way. Convention PC2 establishes a sign convention for A by assuming that the first element of the column achieving the maximum absolute value for the column is positive. The convention that the columns of A are all unit vectors implies that PC2 is not vacuous in that every column of A must have strictly positive maximum absolute value. Convention PC3 determines the signs of the γ_f's, leaving the sign of the β_f's determined, and it also fixes the ordering of the factors when the inequalities are strict. If the inequalities are not strict, additional conditions will be required for identifiability. The requirement that all of the elements of D_+ be positive forces μ to satisfy a genuinely F-factor model. If μ satisfies a PARAFAC model with no more than $F-1$ factors, then $D_+[F,F] = 0$, the Fth column of Γ is entirely 0, and the Fth columns of A and B cannot possibly be well defined. It is possible for PC3 to hold with $D_+[F,F] = 0$ if two terms $\alpha_f \times \beta_f\gamma_f[k]$ and $\alpha_{f'} \times \beta_{f'}\gamma_{f'}[k]$ exactly cancel when summed over k. Since all elements of A, B, and Γ are generally nonnegative in the models we study, we will not explicitly present a decomposition as valid when cancellations occur. However, it will become apparent in §4 that D_+ can always be replaced in PC3 and in the decompositions by a matrix $D(w)$ as defined there.

Once some of the parameters of the trilinear model are known, it is easy to determine the others. We first suppose that A is known, and we then show that IC1 implies that B and Γ can be determined from μ and A. Identifiability condition IC1 implies that A^TA is nonsingular. It follows that $A^+ = (A^TA)^{-1}A^T$ and that A^+A is the F-dimensional identity matrix, or

$$A^+A = I_F. \tag{5}$$

Equation (5) and the trilinear matrix equations (4) imply that

$$\boldsymbol{A}^{+}\boldsymbol{\mu}_k = \boldsymbol{A}^{+}(\boldsymbol{A}\boldsymbol{D}_k\boldsymbol{B}^T) = \boldsymbol{D}_k\boldsymbol{B}^T, k = 1, \ldots, K. \tag{6}$$

Summing (6) over k, we obtain

$$\boldsymbol{A}^{+}\boldsymbol{\mu}_{+} = \boldsymbol{D}_{+}\boldsymbol{B}^T.$$

Since the rows of $\boldsymbol{B}^T$ have length 1 and $\boldsymbol{D}_{+}$ is diagonal, the fth diagonal element of $\boldsymbol{D}_{+}$ must be the length of the fth row of $\boldsymbol{A}^{+}\boldsymbol{\mu}_{+}$ and the fth column of $\boldsymbol{B}$ is just the unit vector formed by normalizing the fth row of $\boldsymbol{A}^{+}\boldsymbol{\mu}_{+}$. Algebraically, the diagonal of $\boldsymbol{D}_{+}$ is the square root of the diagonal of $\boldsymbol{A}^{+}\boldsymbol{\mu}_{+}(\boldsymbol{A}^{+}\boldsymbol{\mu}_{+})^T$ and

$$\boldsymbol{B} = \boldsymbol{\mu}_{+}^T(\boldsymbol{A}^{+})^T\boldsymbol{D}_{+}^{-1}. \tag{7}$$

The diagonal matrices $\boldsymbol{D}_k$, and hence $\boldsymbol{\Gamma}$, can now be calculated from (6). We note that this argument does not require $\boldsymbol{B}$ to have full column rank.

If both $\boldsymbol{A}$ and $\boldsymbol{B}$ are known and satisfy IC1 and IC2, a convenient expression is available for $\boldsymbol{D}_k$. Since $(\boldsymbol{B}^T)^{+} = (\boldsymbol{B}^{+})^T$, the derivation of (5) shows that IC2 implies that

$$\boldsymbol{B}^T(\boldsymbol{B}^T)^{+} = \boldsymbol{I}_F. \tag{8}$$

If both sides of (6) are multiplied on the right by $(\boldsymbol{B}^T)^{+}$, it follows that

$$\boldsymbol{D}_k = \boldsymbol{A}^{+}\boldsymbol{\mu}_k(\boldsymbol{B}^T)^{+}, \qquad k = 1, \ldots, K. \tag{9}$$

4. Mathematical decompositions. This section contains decompositions that can be used to obtain $\boldsymbol{A}$ from $\boldsymbol{\mu}$ when the identifiability conditions hold. Since (7) and (6) give $\boldsymbol{B}$ and $\boldsymbol{D}_k$ (and hence $\boldsymbol{\Gamma}$) from $\boldsymbol{\mu}$ and $\boldsymbol{A}$, all of the parameters can be identified from $\boldsymbol{\mu}$ if $\boldsymbol{A}$ can be determined from $\boldsymbol{\mu}$ when neither $\boldsymbol{B}$ nor $\boldsymbol{\Gamma}$ is known. The first part of Theorem 4.1 states that identifiability conditions on the first two ways imply that the columns of $\boldsymbol{A}$ are eigenvectors of certain matrices computed from the array $\boldsymbol{\mu}$. The second part of Theorem 4.1 states that if a weaker identifiability condition for the third way also holds, the only vectors that are eigenvectors of all of the matrices are scalar multiples of the columns of $\boldsymbol{A}$ and vectors in the intersection of all of the kernels of the matrices. The theorem therefore determines $\boldsymbol{A}$ up to permutation and sign changes of the columns when all three identifiability conditions hold. The analogues of the theorem obtained by permuting ways of the array are left to the reader. The two corollaries reexpress the decompositions in forms that are useful in applications. Corollary 4.3 restates Theorem 4.1 after reduction to an F-by-F-by-K array $\boldsymbol{\eta}$. Corollary 4.4 gives a variant appropriate when $I = J$ and all of the matrices $\boldsymbol{\mu}[, , k]$ are symmetric.

The key to the decomposition is to compare two linear combinations of the matrices $\boldsymbol{\mu}_k, k = 1, \ldots, K$. To that end, for every K-vector $\boldsymbol{w}$, define

$$\boldsymbol{\mu}(\boldsymbol{w}) = \sum_{k=1}^{K} w[k]\boldsymbol{\mu}[, , k]$$

and

$$\boldsymbol{D}(\boldsymbol{w}) = \sum_{k=1}^{K} w[k]\boldsymbol{D}_k,$$

so that $\boldsymbol{\mu}_+ = \boldsymbol{\mu}(\mathbf{1}_K)$ and $\boldsymbol{D}_+ = \boldsymbol{D}(\mathbf{1}_K)$. Theorem 4.1 below presents an eigenvalue problem whose solution is the columns of $\boldsymbol{A}$. See [17] for an interpretation of the matrices $\boldsymbol{\theta}_k$ and a heuristic rationale for the theorem.

THEOREM 4.1. *Assume that the trilinear matrix equations* (4) *hold with the parameterization conventions* PC1–PC3. *Using the Moore–Penrose inverse of* $\boldsymbol{\mu}_+$, *we define* $\boldsymbol{\theta}_k = \boldsymbol{\mu}_k \boldsymbol{\mu}_+^+$ *and* $\boldsymbol{\theta}(\boldsymbol{w}) = \boldsymbol{\mu}(\boldsymbol{w})\boldsymbol{\mu}_+^+$.

1. *If* IC1 *and* IC2 *hold, then for every nonzero* K*-vector* $\boldsymbol{w}$ *the columns of* $\boldsymbol{A}$ *are eigenvectors of* $\boldsymbol{\theta}(\boldsymbol{w})$ *with eigenvalues equal to the diagonal elements of* $\boldsymbol{D}(\boldsymbol{w})\boldsymbol{D}_+^{-1}, k = 1, \ldots, K$.

2. *If identifiability conditions* IC1–IC3 *hold, then there exists a* K*-vector* $\boldsymbol{w}$ *such that the only nontrivial eigenvectors of* $\boldsymbol{\theta}(\boldsymbol{w})$ *are scalar multiples of the columns of* $\boldsymbol{A}$.

Proof. It suffices to establish the first claim for vectors $\boldsymbol{w}$ that are columns of the K-dimensional identity matrix or to show that every column of $\boldsymbol{A}$ is an eigenvector of each matrix $\boldsymbol{\theta}_k$, $k = 1, \ldots, K$. First we argue that the columns of $\boldsymbol{A}$ are indeed eigenvectors of the $\boldsymbol{\theta}_k$'s.

By IC1, IC2, and PC3, Proposition 2.1 can be applied to give the Moore–Penrose generalized inverse of $\boldsymbol{\mu}_+$ in terms of the generalized inverses of $\boldsymbol{A}$, $\boldsymbol{D}_+$, and $\boldsymbol{B}$:

$$\boldsymbol{\mu}_+^+ = (\boldsymbol{B}^T)^+ \boldsymbol{D}_+^{-1} \boldsymbol{A}^+.$$

Substitution into the definition of $\boldsymbol{\theta}_k$ gives

$$\boldsymbol{\theta}_k = \boldsymbol{A}\boldsymbol{D}_k(\boldsymbol{B}^T)(\boldsymbol{B}^T)^+ \boldsymbol{D}_+^{-1} \boldsymbol{A}^+.$$

By (8), $\boldsymbol{\theta}_k = \boldsymbol{A}\boldsymbol{D}_k\boldsymbol{D}_+^{-1}\boldsymbol{A}^+$, so that

$$\boldsymbol{\theta}_k \boldsymbol{A} = \boldsymbol{A}\boldsymbol{D}_k\boldsymbol{D}_+^{-1}\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{A}\boldsymbol{D}_k\boldsymbol{D}_+^{-1}, \tag{10}$$

where the right-hand equality follows from (5). Since the matrix $\boldsymbol{D}_k\boldsymbol{D}_+{}^{-1}$ is diagonal, equation (10) demonstrates that the columns of $\boldsymbol{A}$ are right-eigenvectors of $\boldsymbol{\theta}_k$, with eigenvalues as claimed.

From part 1, each column of $\boldsymbol{A}$ is an eigenvector of $\boldsymbol{\theta}(\boldsymbol{w})$ for every $\boldsymbol{w}$. The second part of the theorem will follow if $\boldsymbol{\theta}(\boldsymbol{w})$ has F one-dimensional eigenspaces or if the matrix $\boldsymbol{\theta}(\boldsymbol{w})$ has F distinct real nonzero eigenvalues or if the diagonal elements of $\boldsymbol{D}(\boldsymbol{w})\boldsymbol{D}_+^{-1}$ are distinct and positive. Lemma 4.2 below shows that PC3 and IC3 imply that the diagonal elements of $\boldsymbol{D}(\boldsymbol{w})\boldsymbol{D}_+^{-1}$ are distinct and positive for almost all $\boldsymbol{w}$. The theorem follows. □

LEMMA 4.2. *If* PC3 *and* IC3 *hold, then the diagonal elements of* $\boldsymbol{D}(\boldsymbol{w})\boldsymbol{D}_+^{-1}$ *are distinct and positive unless* $\boldsymbol{w}$ *is in a set of Lebesgue measure* 0.

Proof. We first argue that the set of $\boldsymbol{w}$'s such that a diagonal element of $\boldsymbol{D}(\boldsymbol{w})\boldsymbol{D}_+^{-1}$ is zero has Lebesgue measure 0. By definition the fth diagonal element is 0 if and only if $\boldsymbol{w}^T\boldsymbol{\gamma}_f = 0$ or if $\boldsymbol{w}$ is in the orthogonal complement of $\boldsymbol{\gamma}_f$, a nonzero vector by PC3. The union of F orthogonal complements has Lebesgue measure 0, so that $\boldsymbol{D}(\boldsymbol{w})\boldsymbol{D}_+^{-1}$ has positive diagonal elements for almost all $\boldsymbol{w}$'s.

Next we show that if PC3 is assumed to hold, then if the fth and f'th diagonal elements of $\boldsymbol{D}(\boldsymbol{w})\boldsymbol{D}_+^{-1}$ are tied, then either IC3 fails or $\boldsymbol{w}$ is in a set of measure 0. The fth diagonal element of $\boldsymbol{D}(\boldsymbol{w})\boldsymbol{D}_+^{-1}$ is $\boldsymbol{w}^T\boldsymbol{\gamma}_f / \mathbf{1}_K^T\boldsymbol{\gamma}_f$. Therefore, the fth and f'th diagonal elements of $\boldsymbol{D}(\boldsymbol{w})\boldsymbol{D}_+^{-1}$ are equal if

$$\boldsymbol{w}^T\left(\frac{1}{\mathbf{1}_K^T\boldsymbol{\gamma}_f}\boldsymbol{\gamma}_f - \frac{1}{\mathbf{1}_K^T\boldsymbol{\gamma}_{f'}}\boldsymbol{\gamma}_{f'}\right) = 0, f \neq f'. \tag{11}$$

If the vector expression in parentheses above is zero,

$$\gamma_{f'} = \frac{\mathbf{1}_K^T \gamma_{f'}}{\mathbf{1}_K^T \gamma_f} \gamma_f.$$

Since PC3 implies that the coefficient of γ_f above is positive and finite, γ'_f is a scalar multiple of γ_f, or the columns f and f' of $\boldsymbol{\Gamma}$ are collinear, so that IC3 fails.

Thus if IC3 holds, none of the vectors in parentheses in (11) can be 0, or

$$\frac{1}{\mathbf{1}_K^T \gamma_f} \gamma_f - \frac{1}{\mathbf{1}_K^T \gamma_{f'}} \boldsymbol{\Gamma}_{f'} \neq 0, f' \neq f, f = 1, \ldots, F; f' = 1, \ldots, F.$$

That is, for a pair of diagonal elements of $\boldsymbol{D}(\boldsymbol{w})\boldsymbol{D}_+^{-1}$ to be tied, equation (11) implies that $\boldsymbol{w}$ must be in the orthogonal complement of one of the nonzero K-vectors above. Since each orthogonal complement is a hyperplane of dimension $K-1$, the diagonal elements will be distinct unless $\boldsymbol{w}$ lies in the union of $F(F-1)/2$ hyperplanes of dimension $K-1$. The lemma follows. □

When I (or J) is much larger than F, reparameterization can remove redundant parameters and reduce the order of the matrices whose eigenvalues are sought. Using PC3, IC1, and IC2, we see that the column space of $\boldsymbol{\mu}_+$ is the same as the column space of $\boldsymbol{A}$ and that the column space of each $\boldsymbol{\mu}_k$ is a subspace of the column space of $\boldsymbol{A}$. Therefore, if $\boldsymbol{U}$ is an I-by-F matrix whose orthonormal columns span the column space of $\boldsymbol{\mu}_+$ and if $\boldsymbol{A}' = \boldsymbol{U}^T\boldsymbol{A}$, then $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{A}'$. Therefore, $\boldsymbol{A}$ can be determined by determining the F-by-F matrix $\boldsymbol{A}'$. Similarly, if the orthonormal columns of $\boldsymbol{V}$ span the column space of $\boldsymbol{\mu}_+^T$, setting $\boldsymbol{B}' = \boldsymbol{V}^T\boldsymbol{B}$ implies that $\boldsymbol{B} = \boldsymbol{V}\boldsymbol{B}'$. Substituting for $\boldsymbol{A}$ and $\boldsymbol{B}$ in the trilinear matrix equations (4) gives $\boldsymbol{\mu}_k = \boldsymbol{U}\boldsymbol{A}'\boldsymbol{D}_k\boldsymbol{B}'^T\boldsymbol{V}^T$. Since both $\boldsymbol{U}$ and $\boldsymbol{V}$ have orthonormal columns,

$$\boldsymbol{\mu}'_k = \boldsymbol{U}^T\boldsymbol{\mu}_k\boldsymbol{V} = \boldsymbol{A}'\boldsymbol{D}_k\boldsymbol{B}'^T, \qquad k = 1, \ldots, K,$$

are K F-by-F matrices satisfying an F-factor PARAFAC model. Convenient matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are those obtained from the singular-value decomposition of $\boldsymbol{\mu}_+ = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$, so that $\boldsymbol{\mu}'_+ = \boldsymbol{S}$, a nonsingular diagonal matrix. Applying Theorem 4.1 to the matrices $\boldsymbol{\eta}_k = \boldsymbol{\mu}'_k\boldsymbol{S}^{-1}$ gives Corollary 4.3 below:

COROLLARY 4.3. *Let $\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$ be the singular value decomposition of $\boldsymbol{\mu}_+$. Define*

$$\underset{F \times F}{\boldsymbol{\eta}_k} = \underset{F \times I}{\boldsymbol{U}^T} \underset{I \times J}{\boldsymbol{\mu}_k} \underset{J \times F}{\boldsymbol{V}} \underset{F \times F}{\boldsymbol{S}^{-1}}, \qquad k = 1, \ldots, K.$$

Assume that the trilinear matrix equations (4) *and the parameterization conventions* PC1–PC3 *hold.*

1. *If identifiability conditions* IC1 *and* IC2 *are satisfied, then the columns of $\boldsymbol{Z} := \boldsymbol{U}^T\boldsymbol{A}$ are (right) eigenvectors of $\boldsymbol{\eta}_k$ with eigenvalues equal to the diagonal elements of $\boldsymbol{D}_k\boldsymbol{D}_+^{-1} = \boldsymbol{\Lambda}_k$ for every $k = 1, \ldots, K$.*

2. *If all three identifiability conditions* IC1–IC3 *hold, then the columns of $\boldsymbol{Z}$ are the only common eigenvectors of $\boldsymbol{\eta}_k$, $k = 1, \ldots, K$.*

Theorem 4.1 or Corollary 4.3 can be used to demonstrate that the trilinear matrix equations (4) do not hold. The matrices $\boldsymbol{\theta}_k, k = 1, \ldots, K$, need not be symmetric, so that they can fail to have real eigenvalues. Even if the K matrices have real eigenvalues, the eigenvectors can fail to be common.

In some settings the matrices $\boldsymbol{\mu}_k$ are square and symmetric. When all K matrices are symmetric, we shall use the suggestive notation $\boldsymbol{\Sigma}_k$ for these matrices. (The application suggested by this notation is beyond the scope of this paper.) If the trilinear matrix equations (4) hold, then $\boldsymbol{\Sigma}_k = \boldsymbol{A}\boldsymbol{D}_k\boldsymbol{A}^T$, so that $\boldsymbol{A} = \boldsymbol{B}$ and $\boldsymbol{\Sigma}_+ = \boldsymbol{A}\boldsymbol{D}_+\boldsymbol{A}^T$. If we let the singular-value decomposition of $\boldsymbol{\Sigma}_+$ be given by $\boldsymbol{U}\boldsymbol{S}\boldsymbol{U}^T$, the symmetric matrices

$$\boldsymbol{\Psi}_k = \boldsymbol{S}^{-1/2}\boldsymbol{U}^T\boldsymbol{\Sigma}_k\boldsymbol{U}\boldsymbol{S}^{-1/2}, \qquad k = 1, \ldots, K,$$

have the property that $\boldsymbol{\Psi}_+ = \boldsymbol{I}_F$. Corollary 4.4 below states that the decomposition can be obtained from the $\boldsymbol{\Psi}_k$'s, so that the eigenvalues needed are those of symmetric matrices.

COROLLARY 4.4. *If the K matrices $\boldsymbol{\Sigma}_k$ are square symmetric matrices and if* IC1 *and* IC3 *hold, then the only common eigenvectors of $\boldsymbol{\Psi}_k$ are proportional to the columns of $\boldsymbol{X}$, where*

$$\boldsymbol{X} = \boldsymbol{S}^{-1/2}\boldsymbol{U}^T\boldsymbol{A}.$$

The eigenvalues are the diagonal elements of the matrices $\boldsymbol{D}_k\boldsymbol{D}_+^{-1}$.

Proof. The columns of $\boldsymbol{X}$ will not generally be of unit length, so that normalizations will be required in order to meet the parameterization conventions PC1–PC3. The demonstration that the columns of $\boldsymbol{X}$ are eigenvectors uses the two representations for $\boldsymbol{\Sigma}_+^+$, one formula in terms of $\boldsymbol{A}^+$ and one formula in terms of $\boldsymbol{U}$. It is necessary to show that $\boldsymbol{\Psi}_k\boldsymbol{X} = \boldsymbol{X}\boldsymbol{\Lambda}_k$. The first step is to substitute for $\boldsymbol{\Sigma}_k$ in the left-hand side:

$$\begin{aligned}\boldsymbol{\Psi}_k\boldsymbol{X} &= (\boldsymbol{S}^{-1/2}\boldsymbol{U}^T(\boldsymbol{A}\boldsymbol{D}_k\boldsymbol{A}^T)\boldsymbol{U}\boldsymbol{S}^{-1/2})\boldsymbol{S}^{-1/2}\boldsymbol{U}^T\boldsymbol{A} \\ &= \boldsymbol{S}^{-1/2}\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{D}_k\boldsymbol{A}^T\boldsymbol{U}\boldsymbol{S}^{-1}\boldsymbol{U}^T\boldsymbol{A}.\end{aligned} \tag{12}$$

The product of the first three matrices is just $\boldsymbol{X}$. By Proposition 2.1, $\boldsymbol{U}\boldsymbol{S}^{-1}\boldsymbol{U}^T = (\boldsymbol{\Sigma}_+)^+$. Since $(\boldsymbol{\Sigma}_+)^+ = (\boldsymbol{A}\boldsymbol{D}_+\boldsymbol{A}^T)^+$, Proposition 2.1 also implies that $(\boldsymbol{\Sigma}_+)^+$ is $(\boldsymbol{A}^T)^+\boldsymbol{D}_+^{-1}\boldsymbol{A}^+$. Therefore, the matrix products to the right of $\boldsymbol{D}_k$ in (12) are $\boldsymbol{A}^T\boldsymbol{U}\boldsymbol{S}^{-1}\boldsymbol{U}^T\boldsymbol{A} = \boldsymbol{A}^T(\boldsymbol{A}^T)^+\boldsymbol{D}_+^{-1}\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{D}_+^{-1}$. Therefore, $\boldsymbol{\Psi}_k\boldsymbol{X} = \boldsymbol{X}\boldsymbol{D}_k\boldsymbol{D}_+^{-1}$, as claimed.

The remainder of the proof is similar to the proof of the second part of Theorem 4.1, and so the details will be omitted. □

5. A decomposition algorithm. In this section we present an algorithm that applies the decomposition. In the first subsection we list steps that will determine the parameters of an F-factor PARAFAC model in the absence of noise. We then indicate several modifications that will not change the results obtained when $\boldsymbol{\mu}$ satisfies the trilinear model. As described in §5.3 below, we have observed that a few of these modifications result in substantial improvements when applied to real or simulated arrays of fluorescence data. Subsection 5.2 discusses our implementation of this algorithm.

5.1. Steps in the algorithm.

1. Sum over the third index of the data array $\boldsymbol{Y}$ to obtain $\boldsymbol{Y}_+ = \sum_{k=1}^K \boldsymbol{Y}[,,k]$.
2. Fix $F > 1$.

When the array to be decomposed satisfies the trilinear model (3) and if PC3 holds, the rank of $\boldsymbol{\mu}_+$ will be F. If the context of the model ensures that all elements of $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{\Gamma}$ are nonnegative, then PC3 will hold for the smallest F such that the

trilinear model (3) holds. If $\boldsymbol{Y}$, the array to be decomposed, is a perturbation of an array $\boldsymbol{\mu}$ satisfying (3), the rank of $\boldsymbol{Y}_+$ need not be F. The value of F should be at least as large as the number of dominant singular values of $\boldsymbol{Y}_+$.

3. Obtain the F leading singular values and vectors of $\boldsymbol{Y}_+$. Let $\boldsymbol{S}$ be the F-by-F diagonal matrix whose diagonal elements are the singular values, let $\boldsymbol{U}$ be the I-by-F matrix whose orthonormal columns are the left singular vectors, and let $\boldsymbol{V}$ be the J-by-F matrix whose orthonormal columns are the right singular vectors.
4. Form $\hat{\boldsymbol{\eta}}_k = \boldsymbol{U}^T\boldsymbol{Y}[,,k]\boldsymbol{V}\boldsymbol{S}^{-1}$, $k = 1,\ldots,K$.
5. Find a K-vector of weights $\boldsymbol{w}$ such that $\hat{\boldsymbol{\eta}}(\boldsymbol{w}) = \sum_{k=1}^{K} w[k]\hat{\boldsymbol{\eta}}_k$ has F distinct real eigenvalues.
6. Let $\boldsymbol{Z}(\boldsymbol{w})$ be an F-by-F matrix whose columns are unit-length eigenvectors of $\hat{\boldsymbol{\eta}}(w)$.
7. Take $\hat{\boldsymbol{A}} = \boldsymbol{U}\boldsymbol{Z}(\boldsymbol{w})$.
8. Find

$$\boldsymbol{B}_o^T = \hat{\boldsymbol{A}}^+\boldsymbol{Y}_+ = (\boldsymbol{Z}(\boldsymbol{w}))^{-1}\boldsymbol{S}\boldsymbol{V}^T.$$

   The second equality follows from application of Proposition 2.1 to step 7.
9. Set $\boldsymbol{D}_+$ to be the diagonal matrix of lengths of columns of $\boldsymbol{B}_o$, and take $\hat{\boldsymbol{B}} = \boldsymbol{B}_o\boldsymbol{D}_+^{-1}$.
10. Find $\boldsymbol{D}_k = (\boldsymbol{Z}(\boldsymbol{w}))^{-1}\hat{\boldsymbol{\eta}}_k\boldsymbol{Z}(\boldsymbol{w})\boldsymbol{D}_+$ and set the kth row of $\hat{\boldsymbol{\Gamma}}$ equal to the diagonal of $\boldsymbol{D}_k$, $k = 1,\ldots,K$.
11. Permute the columns of $\hat{\boldsymbol{A}}$, $\hat{\boldsymbol{B}}$, and $\hat{\boldsymbol{\Gamma}}$, and change the signs of the columns to satisfy the parameterization conventions PC2 and PC3.

When $F = 1$, the matrices $\hat{\boldsymbol{\eta}}_k$, $k = 1,\ldots,K$, formed in step 4 are scalars, so that all linear combinations $\hat{\boldsymbol{\eta}}(\boldsymbol{w})$ automatically have a single ($F = 1$) nontrivial real eigenvalue. For the $F = 1$ PARAFAC model, steps 4–11 can be replaced by step 4_1:

4_1. Set $\hat{\boldsymbol{A}} = \boldsymbol{U}$, $\hat{\boldsymbol{B}} = \boldsymbol{V}$, and $\hat{\Gamma}[k,1] = \hat{\boldsymbol{A}}^T\boldsymbol{Y}[,,k]\hat{\boldsymbol{B}}$, $k = 1,\ldots,K$.

This modification is an application of (9) with $\boldsymbol{\mu}_k$ replaced by $\boldsymbol{Y}_k$ because the Moore–Penrose generalized inverse of an n-by-1 unit vector is the 1-by-n row vector that is its transpose.

The matrices $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{B}}$ are characterized differently in the steps above, even though the trilinear model (3) and the parameterization conventions PC1–PC3 treat the first two ways symmetrically. The proposition below establishes that this asymmetry is only in appearances: Exchanging the first two ways of an array will always give equivalent decompositions even if $\boldsymbol{Y}$ does not satisfy the trilinear model. Thus steps 8 and 9 are equivalent to 8′ and 9′ provided that the vector $\boldsymbol{w}$ of step 8′ is the vector used in step 5.

8′. Form $\hat{\boldsymbol{\eta}}^* = \boldsymbol{V}^T(\boldsymbol{Y}[,,k])^T\boldsymbol{U}\boldsymbol{S}^{-1}$, and set $\hat{\boldsymbol{\eta}}^*(\boldsymbol{w}) = \sum_{k=1}^{K} w[k]\hat{\boldsymbol{\eta}}_k^*$.
9′. Let $\boldsymbol{Z}^*$ be an F-by-F matrix whose columns are unit-length (right) eigenvectors of $\hat{\boldsymbol{\eta}}^*$. Take $\hat{\boldsymbol{B}} = \boldsymbol{V}\boldsymbol{Z}^*$.

PROPOSITION 5.1. *Let $\boldsymbol{Y}$ be an I-by-J-by-K data array. Define a J-by-I-by-K array $\boldsymbol{Y}^*$ by setting $\boldsymbol{Y}^*[,,k] := (\boldsymbol{Y}[,,k])^T$. Assume it is possible to take $\boldsymbol{w}$ such that step 5 is possible when applied to $\boldsymbol{Y}$. Let $\hat{\boldsymbol{B}}$ be the matrix obtained in step 9 when the algorithm is applied to $\boldsymbol{Y}$, and let $\hat{\boldsymbol{A}}^*$ be the matrix of parameters for the nominal first way when the algorithm is applied to $\boldsymbol{Y}^*$ (with the same vector $\boldsymbol{w}$). Then $\hat{\boldsymbol{B}} = \hat{\boldsymbol{A}}^*$.*

Proof. The superscript * will denote quantities derived from the permuted array $\boldsymbol{Y}^*$. To establish the proposition, the results of each $\boldsymbol{Y}^*$ step must be written in

terms of quantities computed from $\boldsymbol{Y}$. Since $\hat{\boldsymbol{A}}^* = \boldsymbol{U}^*\boldsymbol{Z}^*$ in step 7, the matrices $\boldsymbol{U}^*$ and $\boldsymbol{Z}^*$ need to be written in terms of unstarred quantities. The K-vector $\boldsymbol{w}$ will be omitted from the notation since the same $\boldsymbol{w}$ is used with both $\boldsymbol{Y}$ and $\boldsymbol{Y}^*$.

In step 1, $\boldsymbol{Y}_+^* = \boldsymbol{Y}_+^T$, so that the diagonal matrices of singular values in step 3 are equal and left and right singular vectors are exchanged: $\boldsymbol{S}^* = \boldsymbol{S}$, $\boldsymbol{U}^* = \boldsymbol{V}$, and $\boldsymbol{V}^* = \boldsymbol{U}$. In step 4, $\hat{\boldsymbol{\eta}}_k^* = \boldsymbol{V}^T(\boldsymbol{Y}_k)^T\boldsymbol{U}\boldsymbol{S}^{-1} = \boldsymbol{S}(\boldsymbol{S}^{-1}\boldsymbol{V}^T\boldsymbol{Y}_k^T\boldsymbol{U})\boldsymbol{S}^{-1} = \boldsymbol{S}\hat{\boldsymbol{\eta}}_k^T\boldsymbol{S}^{-1}$. In step 5, $\hat{\boldsymbol{\eta}}^* = \boldsymbol{S}(\hat{\boldsymbol{\eta}}^T)\boldsymbol{S}^{-1}$ or the matrix $\hat{\boldsymbol{\eta}}^*$ is similar to $\hat{\boldsymbol{\eta}}^T$ with diagonal similarity matrix $\boldsymbol{S}^{-1}$ [12, p. 44]. Therefore, $\hat{\boldsymbol{\eta}}^*$ and $\hat{\boldsymbol{\eta}}$ have the same eigenvalues, so that $\hat{\boldsymbol{\eta}}^*$ has F distinct positive real eigenvalues.

The matrix equation summarizing the fact that the columns of $\boldsymbol{Z}$ are (right) eigenvectors of $\hat{\boldsymbol{\eta}}$ with diagonal matrix of eigenvalues $\boldsymbol{\Lambda}$ is $\hat{\boldsymbol{\eta}}\boldsymbol{Z} = \boldsymbol{Z}\boldsymbol{\Lambda}$. When the eigenvalues are distinct, then $\boldsymbol{Z}$ is nonsingular, so that $\hat{\boldsymbol{\eta}} = \boldsymbol{Z}\boldsymbol{\Lambda}\boldsymbol{Z}^{-1}$ and $\hat{\boldsymbol{\eta}}^T = (\boldsymbol{Z}^T)^{-1}\boldsymbol{\Lambda}\boldsymbol{Z}^T$, implying that $\hat{\boldsymbol{\eta}}^T(\boldsymbol{Z}^T)^{-1} = (\boldsymbol{Z}^T)^{-1}\boldsymbol{\Lambda}$ or that the columns of $(\boldsymbol{Z}^T)^{-1}$ are eigenvectors of $\hat{\boldsymbol{\eta}}^T$. But the similarity of $\hat{\boldsymbol{\eta}}^*$ and $\hat{\boldsymbol{\eta}}^T$ implies that the columns of $\boldsymbol{S}(\boldsymbol{Z}^T)^{-1}$ are (unnormalized) eigenvectors of $\boldsymbol{\eta}^*$, so that the normalized eigenvectors $\boldsymbol{Z}^*$ are $\boldsymbol{S}(\boldsymbol{Z}^T)^{-1}\boldsymbol{D}^{-1}$, where $\boldsymbol{D}$ is the diagonal matrix of the lengths of the columns of $\boldsymbol{S}(\boldsymbol{Z}^T)^{-1}$. Since $\boldsymbol{V}$ is an orthogonal matrix, the lengths of the rows of $\boldsymbol{S}(\boldsymbol{Z}^T)^{-1}$ are the same as the lengths of the rows of $\boldsymbol{S}(\boldsymbol{Z}^T)^{-1}\boldsymbol{V}$, implying that $\boldsymbol{D} = \boldsymbol{D}_+$. Now substituting for $\boldsymbol{U}^*$ and $\boldsymbol{Z}^*$, we obtain $\boldsymbol{Z}^* = \boldsymbol{V}\boldsymbol{S}(\boldsymbol{Z}^T)^{-1}\boldsymbol{D}_+^{-1} = \boldsymbol{B}_o\boldsymbol{D}_+^{-1} = \hat{\boldsymbol{B}}$. The last equality is from step 9, and the preceding equality is from step 8. These equations complete the proof of the proposition. □

The computation of $\hat{\boldsymbol{\Gamma}}$ in step 10 relies on the fact that the columns of $\boldsymbol{Z}(\boldsymbol{w})$ will be eigenvectors of $\hat{\boldsymbol{\eta}}_k$ if the trilinear matrix equations (4) hold. But when $\boldsymbol{Y}$ does not satisfy the trilinear model (3), the matrices $\boldsymbol{D}_k$ may not be diagonal and the estimate of $\boldsymbol{\Gamma}$ may not be satisfactory. Since this decomposition is often intended to initialize a nonlinear least-squares calculation, step 10 can be replaced by step 10′ in the determination of the $\hat{\boldsymbol{\Gamma}}$ that gives the best least-squares fit to $\boldsymbol{Y}$. By the trilinear matrix equations (4), $\boldsymbol{\mu}[,,k]$ is a linear function of $\boldsymbol{\Gamma}[k,]$ when $\boldsymbol{A}$ and $\boldsymbol{B}$ are known, so that a conditional least-squares step will give $\boldsymbol{\Gamma}$. If the trilinear matrix equations (4) hold, steps 10 and 10′ are equivalent, but the numerical results below demonstrate that 10′ should be preferred to 10 for initializing nonlinear least-squares fit.

10′. Estimate $\boldsymbol{\Gamma}[k,]$ from $\boldsymbol{Y}[,,k]$ by using least-squares regression of the model (4) with $\boldsymbol{A} = \hat{\boldsymbol{A}}$ and $\boldsymbol{B} = \hat{\boldsymbol{B}}$.

This variant illustrates the nonuniqueness of decomposition algorithms: Either step 10 or step 10′ will recover $\boldsymbol{\Gamma}$ if (4) holds, but the algorithms differ away from the model.

5.2. Our implementation of the algorithm. We have implemented the decomposition in FORTRAN by using IMSL subroutines [13]. We use the decomposition to initialize an alternating least-squares algorithm equivalent to the iterative steps described by Carroll and Chang [4] and by Harshman [10]. Sands and Young [25] give equivalent algorithms involving slightly more compact expressions by exploiting the multilinear structure. Our implementation does not use the flexibility offered by our formulation of step 5 because we have generally found that one or more $\hat{\boldsymbol{\eta}}_k$ has F real eigenvalues. We restrict our attention to weight vectors $\boldsymbol{w}$ containing one 1 and $K-1$ 0's, so that $\hat{\boldsymbol{\eta}}(\boldsymbol{w})$ is always one of the matrices $\hat{\boldsymbol{\eta}}_k$. The simplest choice of weights is to select the k with largest minimum distance between F real eigenvalues. In §5.3.1 below the choice of k is delayed until after one cycle of an alternating least-squares algorithm has been completed.

The alternating least-squares algorithms exploit the conditional linearity of mul-

tilinear models. The alternating least-squares algorithm has been extremely slow for fluorescence data sets, and we now use ad hoc acceleration methods. At each step these methods extrapolate the recent parameter estimates to approximate the parameters attaining the minimum. Iteration stops when the ratio of the vector norm of the difference between the current estimate and the extrapolated estimate to the vector norm of the current parameter estimates is less than 10^{-4}.

When $F = 1$ the initialization is not used. We have found that the alternating least-squares algorithm converges rapidly (always in five or fewer cycles) and that the limit points and the speed of convergence are not especially sensitive to the initialization.

We have run this code on an IBM 3081D computer at the Instructional and Research Computing Center, Ohio State University; on Pyramid 90X, DEC3100, and DEC5400 computers in the Mathematics and Statistics Computer Laboratory at Ohio State University; and on CRAY computers (XMP; later YMP) at the Ohio Supercomputer Center. Like most programs in active use, this code evolves. The code implementing the algorithm and the tests described in §5.3 below is in [18, Appendix 1].

5.3. Our experience. We now review our experience with this decomposition to initialize iterative nonlinear least-squares calculations. Our experience with five real data sets is described in §5.3.1. Four sets of fluorescence data were collected from aqueous solutions of biochemicals. The rationale for applying the trilinear model to fluorescence data is explained in [17]. Our fifth data set is the tongue-position data reported by Harshman, Ladefoged, and Goldstein [11]. Fluorescence data were also simulated for a study reported in [16]. The simulation results are described in §5.3.2. Except for the tongue data set, the three ways correspond to excitation wavelength, emission wavelength, and concentration of a chemical that affects fluorescence intensity. We focus on the behavior of the decomposition and its variants as an initialization of nonlinear least-squares calculations.

5.3.1. Performance of decomposition on real data. Each of the real fluorescence data sets has some structurally missing entries because fluorescence intensity cannot be measured accurately when excitation and emission wavelengths are very close. Since complete arrays are required to attempt the decomposition and since the fraction of entries missing is small, we filled the missing entries with the expected fluorescence intensity determined from the parameter estimates from the model with $F-1$ when initializing the computations for the model with F terms. The constructed entries are used *only* for the decomposition; all iterative least-squares steps use only the observed data.

Table 1 is a summary of the behavior of the algorithm for five real data sets. The triple of integers below the code name of each data set is the dimension vector of the three-way data array. The negative number below the vector indicates the number of missing elements. For more description of the fluorescence data sets, see [18, Appendix 2].

We attempted K decompositions for each data set and for $F = 2, 3, 4$. For each decomposition the matrix $\hat{\boldsymbol{\eta}}(\boldsymbol{w})$ of step 5 was set equal to one of the matrices $\hat{\boldsymbol{\eta}}_k$ of step 4. All eigenvalues from step 5 are listed in the tables of [18, Appendix 2]. When any eigenvalues were complex, step 5 failed. The number of slabs for which step 5 failed is reported in the column of Table 1 headed "Failure." The number of failures of step 5 increases with F, the number of terms fit. None of the five decompositions

TABLE 1

Summary of real data sets. Four sets of fluorescence data from aqueous solutions of biochemicals were collected in the laboratory of the second author. The fifth data set is the tongue-position data of [11]. *The triple of integers below the code name of each data set is the dimension vector of the three-way data array. The negative number below the vector indicates the number of missing elements. For explanation of the four sums-of-squares* SS1–SS4, *see text.*

Data set	F	Failure	SS1	Slab	SS2	Slab	SS3	Slab	SS4
PEA	1								2.91
(17,15,9)	2	0	4.15	(3)	.96	(1)	.44	(1)	.42
$-(8 \times 9)$	3	3	4.83	(3)	.87	(6)	.31	(9)	.27
	4	5	58.86	(2)	.58	(9)	.25	(2)	.17
PCGK	1								3.237
(11,22,7)	2	1	.573	(6)	.568	(3)	.503	(2)	.498
$-(17 \times 7)$	3	4	.510	(3)	.442	(2)	.429	(5)	.422
	4	4	.669	(5)	.394	(5)	.354	(5)	.306
NAT	1								197.23
(8,22,7)	2	0	15.78	(3)	14.42	(3)	5.96	(3)	5.41
$-(8 \times 7)$	3	0	.81	(2)	.61	(2)	.50	(2)	.47
	4	2	1.12	(3)	.47	(2)	.33	(3)	.19
LADH	1								15.84
(15,27,7)	2	0	8.79	(1)	8.67	(1)	6.15	(1)	5.36
$-(16 \times 7)$	3	0	2.47	(5)	1.13	(3)	.65	(1)	.38
	4	2	1.69	(1)	.49	(1)	.40	(1)	.34
TONG	1								240.56
(10,13,5)	2	1	20.94	(2)	20.62	(4)	20.16	(4)	19.99
$-(0)$	3	2	24.43	(3)	18.88	(1)	17.82	(4)	18.43
	4	5							

attempted provided an initialization for the tongue data set (TONG) with $F = 4$. We note that in the original source the model with $F = 2$ was selected.

For every k for which step 5 was possible, the remaining steps were performed. The smallest residual sum of squares from the parameter estimates at the end of step 10 is designated SS1 in Table 1, with the number of the corresponding slab indicated in parentheses. The smallest residual sum of squares resulting when step 10′ replaces step 10 is designated SS2. Note that the slab corresponding to SS1 is not always the slab corresponding to the smallest SS2. The smallest residual sum of squares after one ALS cycle beyond the step giving SS2 is designated SS3. The alternating least-squares algorithm was iterated to convergence for the slab corresponding to SS3. The residual sum of squares at convergence is designated SS4. For the residual sums of squares for every slab, see [18, Appendix 2].

5.3.2. Performance of decomposition on simulated fluorescence data. We extracted some results from a small simulation study reported in [16], a preliminary study of several methods of determining F in the trilinear model (3). We first outline the simulation, and we then report the results and compare them with the real data.

For the simulation, five independent data arrays were generated for each of three deterministic 10-by-12-by-5 arrays $\boldsymbol{\mu}$ satisfying the trilinear model. Each deterministic array corresponds to plausible fluorescence parameters. The number of terms in the true array $\boldsymbol{\mu}$ was $\tilde{F} = 1, 2, 3$. The three theoretical arrays are referred to as NFAK1, NFAK2, and NFAK3, with the number denoting the number of terms in the true array $\boldsymbol{\mu}$. The elements of $\boldsymbol{\mu}$ were between 0 and 10^4. The data arrays were simulated by adding independent pseudorandom normal variables with mean 0 and standard deviation 100. For all 15 simulated data arrays the trilinear model was fit with F set

TABLE 2

*Summary of simulated data sets. Slab number is the value of k used in step 4. The entry in each cell of the table is the number (out of 5) of replicated data arrays for which step 5 was possible. The * denotes models and slabs for which the computer code gave complex eigenvalues even when no noise was present.*

		Slab number				
Model	F	1	2	3	4	5
NFAK1	2	4	4	5	5	5
	3	2*	1*	3*	4	2
	4	1*	1	1*	3*	1*
NFAK2	2	5	5	3	5	5
	3	5	4	5	5	5
	4	2*	2	3	3*	3
NFAK3	2	5	5	5	5	5
	3	5	5	3	5	5
	4	5	2	4	5	5

equal to 1, 2, 3, and 4.

Table 2 gives the number of slabs that generated successful step 5's. Figure 1 shows the actual eigenvalues: Each column of plots corresponds to a different theoretical array; each row of plots corresponds to a different number of terms fitted. The circles in each panel are sample eigenvalues (when real) plotted against slab number. The eigenvalues of the theoretical arrays are superimposed crosses. It is evident from the figure that the sample eigenvalues were close to the true eigenvalues when the correct number of terms were fit and when separation of the true eigenvalues was larger than the variation in the sample eigenvalues. When more eigenvalues were fit than were present, some of the sample eigenvalues were close to the population parameters but the other sample eigenvalues varied considerably. Examination of the three sums-of-squares fits showed that using conditional least-squares fits for $\hat{\boldsymbol{\Gamma}}$ is important.

6. Extensions to four or more ways. In this section we prove that arrays with four ways can be decomposed by using the decomposition for three-way arrays. We state an extension to more ways.

The four-way extension of the trilinear model (3) to an I-by-J-by-L-by-M four-way array $\boldsymbol{\mu}$ is called the *quadrilinear model*:

$$\boldsymbol{\mu} = \sum_{f=1}^{F} \boldsymbol{\alpha}_f \times \boldsymbol{\beta}_f \times \boldsymbol{\gamma}_f \times \boldsymbol{\delta}_f, \tag{13}$$

where the $\boldsymbol{\alpha}_f$'s, the $\boldsymbol{\beta}_f$'s, the $\boldsymbol{\gamma}_f$'s, and the $\boldsymbol{\delta}_f$'s are the F columns of the matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\Gamma}$, and $\boldsymbol{\Delta}$, respectively. The matrix $\boldsymbol{A}$ has I rows, $\boldsymbol{B}$ has J rows, $\boldsymbol{\Gamma}$ has L rows, and $\boldsymbol{\Delta}$ has M rows. The parameterization convention PC1 can be extended to include the third way by requiring that the columns of $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{\Gamma}$ all have unit length. The sign convention of PC2 can be extended to the matrix $\boldsymbol{B}$, so that $\boldsymbol{\Gamma}$ carries the sign.

The quadrilinear model (13) can be written in terms of three-way arrays. For example, if the levels of the last two ways of the four-way array define levels of a way with $K = LM$ levels, a three-way I-by-J-by-K array $\tilde{\boldsymbol{\mu}}$ is induced by

$$\tilde{\mu}[i, j, k] = \mu[i, j, l, m], \qquad k = l + L(m-1), \quad l = 1, \ldots, L; \quad m = 1, \ldots, M.$$

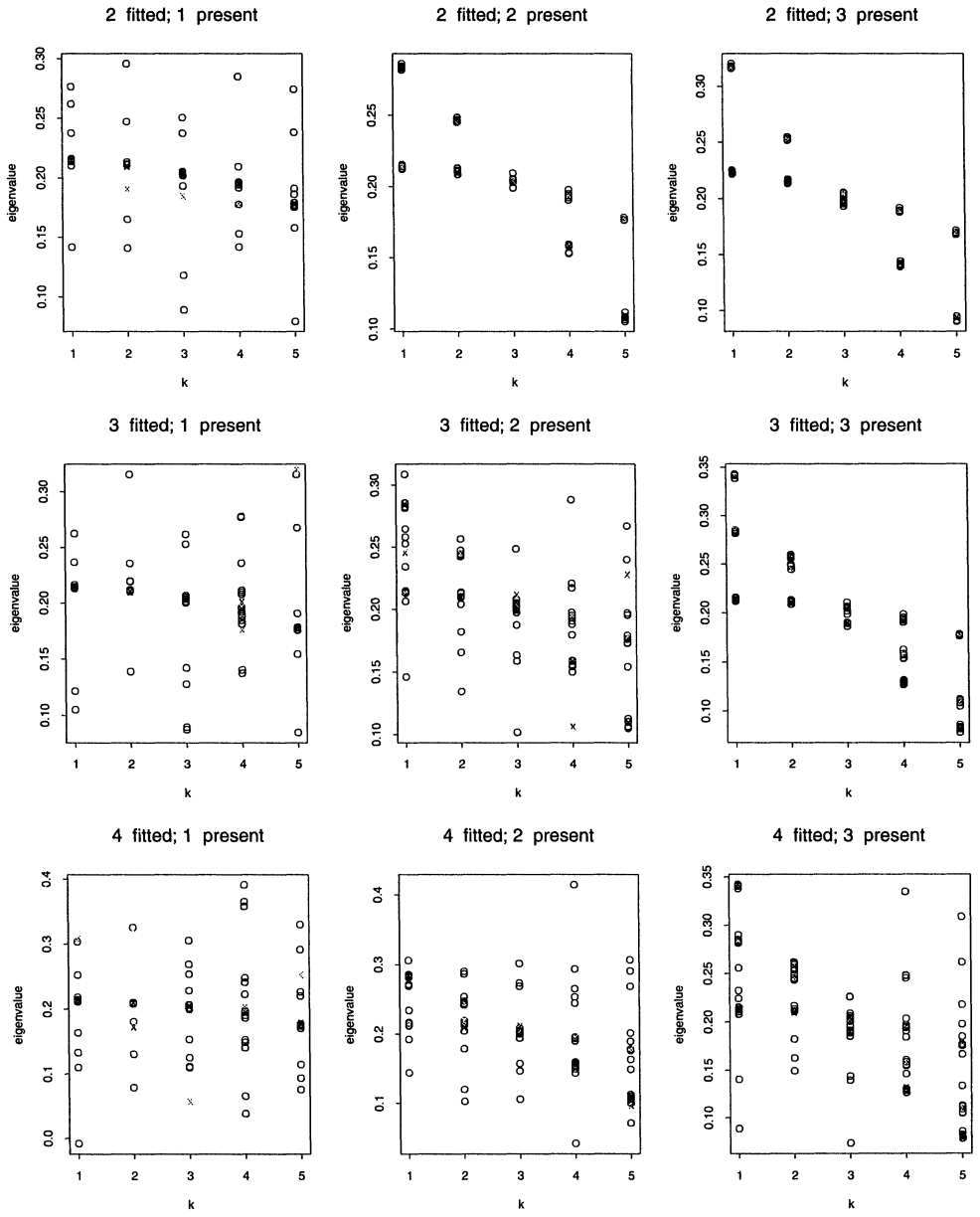

FIG. 1. *Eigenvalues from simulated arrays. Crosses are eigenvalues from theoretical arrays; open circles are eigenvalues from data arrays.*

The quadrilinear model (13) implies that the three-way array $\tilde{\mu}$ satisfies

$$\tilde{\mu} = \sum_{f=1}^{F} \alpha_f \times \beta_f \times \tilde{\gamma}_f, \tag{14}$$

where $\tilde{\gamma}_f$ is the K-vector equal to the Kronecker product of the L-vector γ_f with the M-vector δ_f, denoted $\gamma_f \otimes \delta_f$ and taken to have the first vector associated with the larger pattern and the second vector with the subpattern. Equation (14) has the form of the trilinear model (3), so that $\tilde{\mu}$ has an F-factor three-way PARAFAC decomposition.

Identifiability conditions IC1 and IC2 are unaffected by the reduction from four ways to three ways. Identifiability condition IC3 applies directly to the K-vectors $\tilde{\gamma}_1 \cdots \tilde{\gamma}_F$. Lemma 6.1 at the end of this section shows that IC3 for $\tilde{\Gamma}$ is equivalent to IC_43 and IC_44 below, so that the following *four-way identifiability conditions* are sufficient for identifiability when the parameterization conventions PC1–PC3 alluded to above are made:

(IC_41) The columns of A are linearly independent.
(IC_42) The columns of B are linearly independent.
(IC_43) No two columns of Γ are collinear.
(IC_44) No two columns of Δ are collinear.

To determine the L-by-F matrix Γ and the M-by-F matrix Δ from the K-by-F matrix $\tilde{\Gamma}$, the rows of $\tilde{\Gamma}$ are summed to add up the coefficients from Δ (or from Γ), leaving the coefficients from Γ (respectively, Δ). Explicitly, form the L-by-F matrix G by summing consecutive blocks of M rows of $\tilde{\Gamma}$:

$$G[l, f] = \sum_{m=1}^{M} \tilde{\Gamma}[(l-1)M + m, f], \qquad l = 1, \ldots, L, \quad f = 1, \ldots, F.$$

The quadrilinear model (13) then implies that

$$G[, f] = \Gamma[, f]\mathbf{1}_M^T \Delta[, f], \qquad f = 1, \ldots, F. \tag{15}$$

Since the columns of Γ are assumed to have length 1, Γ can now be determined from (15). Once the matrix Γ is known, the matrix Δ can be found by summing every Mth row of $\tilde{\Gamma}$ to get the M-by-B matrix $\tilde{G}$:

$$\tilde{G}[m, f] = \sum_{l=1}^{L} \tilde{\Gamma}[(l-1)M + m, f], \qquad m = 1, \ldots, M, \quad f = 1, \ldots, F.$$

It is easy to verify that $\Delta[, f] = \tilde{G}[, f]/\mathbf{1}_L^T \Gamma[, f], f = 1, \ldots, F$.

We now sketch the extension of the identifiability conditions to N-way arrays, where $N \geq 3$. Let μ be an N-way array with ν_n levels in the nth way such that a formula like (13) holds. Assume that the N matrices θ_n with ν_n rows and F columns containing the parameters for the nth way satisfy parameterization conventions like the parameterization conventions PC1–PC3, that is, the first $N-1$ matrices will be assumed to have length-1 columns and a sign convention will be assumed. Induction on N shows that if none of the matrices θ_n, $n = 1, \ldots, N$, have any collinear columns and if at least two of the matrices have linearly independent columns, then the N matrices θ_n can be identified from μ.

We conclude this section with a proof of the lemma containing the mathematical details.

LEMMA 6.1. *If $\tilde{\gamma}_f = \gamma_f \otimes \delta_f$, $f = 1, 2$, and if the lengths of $\tilde{\gamma}_1$ and of $\tilde{\gamma}_2$ are positive, then $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ are collinear K-vectors if and only if γ_1 and γ_2 are collinear L-vectors and δ_1 and δ_2 are collinear M-vectors.*

Proof. Suppose that $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ are collinear. Then

$$\tilde{\gamma}_1 = c\tilde{\gamma}_2$$

for some $c \neq 0$. Since $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ are Kronecker products, the L M-dimensional subvectors of γ_f are just

$$\gamma_1[l]\delta_1 = c\gamma_2[l]\delta_2, \qquad l = 1, \ldots, L.$$

Observe that $\gamma_1[l] = 0$ if and only if $\gamma_2[l] = 0$. Since the length of the Kronecker product of two vectors is the product of their lengths, there must be at least one l such that $\gamma_1[l] \neq 0$. For all such l

$$\delta_1 = c\frac{\gamma_2[l]}{\gamma_1[l]}\delta_2.$$

That is, the vector δ_1 is a multiple of δ_2, or δ_1 and δ_2 are collinear. The collinearity of γ_1 and γ_2 follows from a similar argument. □

7. Discussion. The essential features of the decomposition have been in both the psychometric and chemometric literatures concerning initialization of alternating least-squares algorithms. All of these papers discuss real data sets as well. We briefly compare their approaches with ours, and we then summarize our conclusions and recommendations. We conclude our discussion with a brief comparison of array and tensor decompositions.

The psychometric literature discusses several types of initialization, all of which are surveyed and compared for a setting with symmetric slabs by Carroll, DeSoete, and Pruzansky [6]. One simple method is to use one or more random starts. Harshman [10] recommends several random starts. Carroll, DeSoete, and Prozansky found this method to be dominated by other methods, and so did we. In our applications, the inner products of the columns of $\boldsymbol{A}$ (and of $\boldsymbol{B}$) often exceed 0.9; such configurations are not likely to arise from simple random schemes. Another simple method is to attempt to use bilinear models, such as the singular vectors of $\boldsymbol{Y}_+$. We found that imposing nonnegativity constraints did not provide useful initializations. A third approach is to postulate that the identifiability conditions almost fail or that the columns of $\boldsymbol{A}$, $\boldsymbol{B}$, or $\boldsymbol{\Gamma}$ are very close to a prespecified lower-dimensional subspace. In this approximate setting, Carroll, Pruzansky, and Kruskal's [5] linearly constrained PARAFAC model (CANDELINC) applies to reduce to another unconstrained PARAFAC model for an array of smaller dimension. We have not applied this approach directly because in our applications submodels can be specified only for $\boldsymbol{\Gamma}$, the parameter matrix with the smallest number of rows. However, reduction to η_k in Corollary 4.3 can be thought of as postulating data-dependent linear models. We do not, however, pursue alternating least-squares fits for these models. Other initialization methods are more similar to ours. The proposal of Sands and Young [25] differs in detail because they find the eigenvalues of

$$\frac{1}{K}\sum_{k=1}^{K}\hat{\eta}_k^r,$$

where r is an integer (often 2) rather than finding the eigenvalues of a matrix $\hat{\boldsymbol{\eta}}(\boldsymbol{w})$.

The chemometric literature has also used initializations based on eigenvalues. Sanchez and Kowalski [22] rediscovered the approach for I-by-J-by-2 arrays. Sanchez and Kowalski extend their earlier paper to I-by-J-by-K arrays in [21], [23], [24]. They also propose finding the eigenvectors of a matrix. To present their algorithm in an extension of our notation, define the IJ-by-K matrix $\tilde{\boldsymbol{\mu}}$ by stacking the matrices $\tilde{\boldsymbol{\mu}}_k, k = 1, \ldots, K$, one above the next. Let $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ be the first and second right singular vectors of $\tilde{\boldsymbol{\mu}}$. Sanchez and Kowalski use the eigenvectors of $\tilde{\boldsymbol{\mu}}(\boldsymbol{w}_2)^+\tilde{\boldsymbol{\mu}}(\boldsymbol{w}_1)$ to recover the decomposition. When noise is present Sanchez and Kowalski first project $\boldsymbol{Y}$ onto a larger class of multilinear arrays, a class that includes the PARAFAC models. Their preprocessing is different from ours, so that their algorithm will not behave exactly like ours. Our experiments with their approach suggest that their methods and ours behave similarly, with neither method dominating. Our approach has the advantage that we always try K decompositions, so that the impact of complex eigenvalues may be avoided.

Burdick, Tu, McGown, and Millican [3] also describe the same eigendecomposition; readers who compare our paper with theirs should notice that our f (respectively, F) is their k (respectively, K) and our k (respectively, K) is their f (respectively, F). They do not apply the decomposition directly to $\boldsymbol{Y}$ but first apply alternating eigenproblems to obtain better estimates of the column spaces of $\boldsymbol{A}$ and $\boldsymbol{B}$ and then apply the eigendecomposition to $\tilde{\boldsymbol{Y}}_k = \boldsymbol{U}\boldsymbol{U}^T\boldsymbol{Y}_k\boldsymbol{V}\boldsymbol{V}^T$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ have orthonormal columns and are such that their column spaces are estimates of the column spaces of $\boldsymbol{A}$ and $\boldsymbol{B}$. They are iteratively refining preprocessing as used by Sanchez and Kowalski [21]. Our preliminary experiments with our data indicate that the iterative eigendecomposition is more unstable than our direct initialization.

Sample variation from the trilinear model (3) can cause the decomposition to fail because complex eigenvalues can occur at step 5. Complex eigenvalues were observed for some k and some F for every data set. However, the imaginary part of the eigenvalues was not large: The absolute value of the ratio of the imaginary part to the real part of the complex eigenvalues never exceeded 0.189 for the fluorescence data sets. From simulations we see that complex eigenvalues can appear even when the correct model is fitted, so that complex eigenvalues should *not* be regarded as evidence that the trilinear model is inadequate. The set of arrays $\boldsymbol{Y}$ having real eigenvalues for every slab is a closed set, and those arrays having a slab generating at least one real eigenvalue with multiplicity at least 2 are on the boundary of this set, so that slight perturbations give arrays with complex roots. Indeed, slight perturbations of arrays having distinct real eigenvalues, some of which are not well separated, can also induce some complex roots. We have seen that nonunique eigenvalues occur when certain properties of the spectra nearly tie for two factors for some k. These ties can occur without the spectra being otherwise pathological, and so we recommend that several decompositions be attempted routinely. Table 1 shows that some decompositions are better than others when applied to real data, even though the decompositions give the same results when applied to arrays that exactly satisfy the trilinear model. We found the same comparison for simulated data. The original decomposition (with step 10) cannot be recommended because SS1 is smaller than SS4 from $F-1$ only for $F = 2$ (four data sets) and $F = 3$ (two data sets). The minimum SS1 fails to decrease as F increases for four of the five data sets. Sometimes SS1 is very large in real data sets: Two slabs of the PEA data exhibited this problem. Using step 10′ produces a substantial improvement: Only for $F = 3, 4$ of the PEA data does the smallest SS2

exceed SS4 from $F-1$. The conditional-least-squares property guarantees that SS3 will be less than SS2 (which is itself automatically less than SS1), and it is encouraging that the minimum SS3 was always less than SS4 from $F-1$, although the largest SS3 can exceed SS4 from $F-1$: This occurred in slab 5 of the PEA data when $F=4$.

A three-way array can be regarded as an array of components with respect to specified bases of a tensor in a tensor product of three finite-dimensional vector spaces [19]. If the bases in the vector spaces are unchanged, the array of components will change, although the underlying tensor does not. Decomposable tensors generate rank-1 arrays, and every tensor in a tensor product of finite-dimensional vector spaces can be written as a finite sum of decomposable tensors. If a tensor is written as a sum of decomposable tensors, a multilinear model is determined for each array of components of the tensor. However, determining the minimal number of terms in and the uniqueness of a decomposition is not routine in spaces that are tensor products of more than two vector spaces. For example, Kruskal, Harshman and Lundy introduced a 2-by-2-by-2 array (reproduced in [27]) that cannot be written as the sum of two rank-1 arrays. The array can be written as the sum of rank-1 arrays, but not uniquely, even under parameterization conventions PC1–PC3. This example contrasts with standard results for tensor products of two vector spaces (such as [19, p. 36]) and demonstrates again that decompositions for three-way arrays are not immediate from two-way results.

Acknowledgment. Particularly thorough comments from a referee have led us to improve the exposition and strengthen a lemma.

REFERENCES

[1] C. APPELLOF AND E. DAVIDSON, *Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents*, Anal. Chem., 53 (1981), pp. 2053–2056.

[2] P. ARABIE, J. D. CARROLL, AND W. S. DESARBO, *Three-Way Scaling and Clustering*, Quantitative Applications in the Social Sciences, Vol. 65, Sage, Newbury Park, CA, 1987.

[3] D. BURDICK, X. TU, L. MCGOWN, AND D. MILLICAN, *Resolution of multicomponent fluorescent mixtures by analysis of the excitation-emission-frequency array*, J. Chemometrics, 4 (1990), pp. 15–28.

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

[5] J. D. CARROLL, S. PRUZANSKY, AND J. B. KRUSKAL, CANDELINC: *A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters*, Psychometrika, 45 (1980), pp. 3–24.

[6] J. D. CARROLL, G. DESOETE, AND S. PRUZANSKY, *An evaluation of five algorithms for generating an initial configuration for* SINDSCAL, J. Classification, 6 (1989), pp. 105–119.

[7] R. COPPI AND S. BOLASCO, EDS., *Multiway Data Analysis*, North-Holland, New York, 1989.

[8] M. GREENACRE, *Theory and Applications of Correspondence Analysis*, Academic Press, Orlando, FL, 1984.

[9] T. GREVILLE, *Note on the generalized inverse of a matrix product*, SIAM Rev., 8 (1966), pp. 518–521.

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

[11] R. HARSHMAN, P. LADEFOGED, AND L. GOLDSTEIN, *Factor analysis of tongue shapes*, J. Acoust. Soc. Am., 62 (1977), p. 693.

[12] R. HORN AND C. JOHNSON, *Matrix Analysis*, Cambridge University Press, New York, 1985; corrected reprint, 1990.

[13] IMSL, INC., MATH/LIBRARY *User's Manual*, Houston, TX.

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

[15] H. LAW, C. SNYDER, JR., J. HATTIE, AND R. MCDONALD, EDS., *Research Methods for Multimode Data Analysis*, Praeger, New York, 1984.

[16] S. LEURGANS AND R. ROSS, *Cross-validation for multilinear models: Applications to biophysics*, 1989. Presented at Joint Statistical Meeting, Washington, DC.

[17] ———, *Multilinear models: Applications in spectroscopy*, Statist. Sci., 7 (1992), pp. 289–319.

[18] S. LEURGANS, R. ROSS, AND R. ABEL, *A Decomposition for 3-Way Arrays*, Tech. Report 448, Department of Statistics, Ohio State University, Columbus, OH, 1990.

[19] M. MARCUS, *Finite Dimensional Multilinear Algebra, Part* I, Marcel Dekker, New York, 1973.

[20] R. PRINGLE AND A. RAYNER, *Generalized Inverse Matrices with Applications to Statistics*, Hafner, New York, 1971.

[21] E. SANCHEZ AND B. KOWALSKI, *Tensorial resolution: A direct trilinear decomposition*, J. Chemometrics, 4 (1990), pp. 29–45.

[22] E. SANCHEZ AND B. R. KOWALSKI, *Generalized rank annihilation factor analysis*, Anal. Chem., 58 (1986), pp. 496–499.

[23] ———, *Tensorial calibration:* I. *First-order calculations*, J. Chemometrics, 2 (1988), pp. 247–263.

[24] ———, *Tensorial calibration:* II. *Second-order calculations*, J. Chemometrics, 2 (1988), pp. 265–290.

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

[26] G. STYAN, *Generalized inverses*, in Encyclopedia of Statistical Sciences, Vol. 3, S. Kotz, N. Johnson, and C. Read, eds., John Wiley & Sons, Inc., New York, 1983, pp. 334–337.

[27] J. M. TEN BERGE, H. A. KIERS, AND J. DE LEEUW, *Explicit* CANDECOMP/PARAFAC *solutions for a contrived* $2 \times 2 \times 2$ *array of rank three*, Psychometrika, 53 (1988), pp. 579–589.