Comment: PARAFAC in Three-Way Land

Pieter M. Kroonenberg

1. INTRODUCTION

The purpose of this comment is to provide a somewhat wider background to the PARAFAC model discussed in Leurgans and Ross' paper on three-way methods in spectroscopy.

First of all, I would like to express my appreciation for the paper by Leurgans and Ross. It seems to me that the opening up of a methodology to a new area of application is one of the more demanding, but also one of the more necessary tasks for statisticians and data analysts. When it goes hand in hand with further clarification and extension of the methodology, it is the more commendable.

In the present discussion, I will not so much attempt to provide a discussion of the content of the paper itself, but try to supply a wider perspective of the basic model presented in the paper. In particular, I will briefly sketch the literature on the PARAFAC model.

2. THREE-WAY ANALYSIS

Three-way multivariate analysis started with the seminal work of Tucker (especially 1966), and much work in this area including the development of the PARAFAC model is in one way or another derived from his basic ideas. As mentioned by Leurgans and Ross, the so-called PARAllel FACtor model (PARA-FAC) is primarily due to Harshman (1970, 1972; Harshman and Lundy, 1984a, b), but within the field of multidimensional scaling it was independently developed by Carroll and Chang (1970), who called it the CANonical DECOMPosition model (CANDECOMP). In a set of papers, Kiers (1988, 1991) showed how the PARAFAC model fits into a hierarchy of component models with increasing numbers of restrictions on the components. Geladi (1989) is a tutorial for three-way methods (including PARAFAC) with special reference to chemical applications such as spectrometry and chromatography. Other papers which discuss relationships between three-way component models are Carroll and Arabie (1980), Kroonenberg (1983, 1988), Snyder, Law and Hattie (1984), Kruskal (1984), Harshman and Lundy (1984a), Arabie, Carroll and DeSarbo (1987) and Smilde (1992).

Pieter M. Kroonenberg is Associate Professor of Statistics, Department of Education, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands.

3. PARAFAC MODEL

The PARAFAC model, which is the basic method discussed by Leurgans and Ross, is one of the generalisations of the singular value decomposition to three-way data. Several theoretical discussions about the nature of such generalisations can, for instance, be found in Kruskal (1984), Denis and Dhorne (1989), Franc (1989), Yoshizawa (1987) and Kroonenberg (1989). Other mathematical areas of interest are the *rank* of a three-way array (see Kruskal, 1976, 1977, 1989; Franc, 1989; Ten Berge, Kiers and deLeeuw, 1988; Ten Berge, 1991; and their references), and *optimality properties* of three-way methods (d'Aubigny and Polit, 1989).

As is evident from the paper by Leurgans and Ross, the model has generated much interest in chemistry (for related applications, see, in addition, Smilde et al., 1990; Ray and Cole, 1985). However, its roots lie with psychometrics, and it has now reached areas like agriculture and environmental studies. There is also a Russian connection (Lipovetskii, 1984), as well as a Japanese one (Hayashi and Hayashi, 1982; Hayashi, Yamaoka and Terao, 1982). Harshman and Lundy (1984a, b) is the most complete treatment of the model in psychometrics.

3.1 Uniqueness

As indicated by Leurgans and Ross, the PARAFAC model is identified under fairly general identification conditions, and Carroll and Chang (1970), Harshman (1972) and Kruskal (1984) have called attention to this fact (see also deLeeuw and Pruzansky, 1978).

Harshman uses the uniqueness property (or "intrinsic axis property" as he calls it) to search for "real" psychological factors. Whereas in psychology the existence of such proportional factors is a question of conjecture and empirical verification, in some sciences such as chemistry, explicit physical models of the PARAFAC form exist. As formulated by Sanchez and Kowalski (1990), there are "an abundance of instruments that can automatically collect precise third-order data arrays in a short time" (p. 33). In these areas, the PARAFAC model (sometimes referred to as an extension of "generalized rank annihilation"; see Appellof and Davidson, 1983) seems to be used more as a model for parameter estimation than for discovery.

3.2 Algorithms

The basic algorithm for the PARAFAC model is based on an alternating least squares approach, in

which all components are simultaneously searched for and in which, within each step of a major iteration in the simultaneous procedure, three regression problems are solved, each for one of the ways, with the other two ways fixed. Leurgans and Ross provide a proof that the components cannot be found one by one (or recursively) unless orthogonality conditions are imposed, because PARAFAC models with fewer terms than are actually present cannot recover the parameters of larger models. As reported by most authors, convergence is rather slow and alternative routes have been investigated. Appellof and Davidson (1981) reported testing a Fletcher-Powell algorithm which was anything but an improvement. They also investigated the performance of the so-called Aitken extrapolation method, which worked reasonably well in accelerating convergence. Harshman uses in his program a similar "overrelaxation method" (Harshman, 1970), and also Leurgans, Ross and Abel (1992) report using "ad hoc acceleration methods."

The present iterative algorithms seem to be somewhat sensitive to local minima, so that one needs several runs from different starting positions to evaluate the "best" minimum. Harshman and Lundy (1984a, b; Harshman and DeSarbo, 1984) discuss at length the technical and interpretational possibilities and problems of analysing three-way data with the PARAFAC model.

Kiers and Krijnen (1991) suggest, as a method for improving the basic algorithm for the PARAFAC model, estimating the parameters using variable-by-condition covariance matrices rather than the raw profile data. They report considerable decreases in computer time for random data. It was shown that at each iteration step the loss function is the same as that of the original algorithm. The modified PARAFAC algorithm only needs the (means and) covariances of two ways for the basic computations, so that the largest way can be eliminated. Therefore, given there is one stochastic way, the first and second order moments can be seen as sufficient statistics for the (nonincidental) parameters of the model, just as in the two-way case. The parameters of the eliminated way can be computed after the iterative solution for the other two ways has been found.

Several other authors have provided algorithms for estimating the PARAFAC model. Hayashi and Hayashi (1982; Hayashi, Yamaoka and Terao, 1982) provided an algorithm based on estimating successive differences in each iteration and solving sets of simultaneous equations to find the required parameters. Hayashi, Yamaoka and Terao (1982) reported different results than those of Harshman's PARAFAC program, but the paper in which the details are reported is unfortunately in Japanese. Lipovetskii (1984) discussed

the successive component solution that Leurgans and Ross have shown to be suboptimal.

3.3 Algebraic Solution

Sands and Young (1980), using a suggestion by Yoshio Takane, presented a decomposition which will provide a solution to the PARAFAC model in the case of a perfect fit. The basic idea of this decomposition goes back to Schönemann (1972; see also deLeeuw and Pruzansky, 1978), who proposed an algebraic solution for the scalar-product version of Carroll and Chang's (1970) INDSCAL model. Leurgans, Ross and Abel (1992), Sanchez and Kowalski (1990) and Burdick et al. (1990) independently developed a comparable procedure to Sands and Young with minor variations (for a comparison, see Leurgans, Ross and Abel, 1992). In the case of approximate solutions, the Sands and Young procedure and its variants can be used as an initialization for an alternating least squares algorithm.

3.4 Optimal Scaling

Sands and Young (1980) also developed an *optimal scaling* version of PARAFAC (ALSCOMP3) within the optimal scaling tradition of Takane, Young and deLeeuw (e.g., 1977). The inclusion of an optimal scaling phase allows for the treatment of data with different measurement levels, and for data with row, column or matrix conditionalities (see, e.g., Gifi, 1990, for a discussion of these terms). Harshman and Lundy (1984a) give their opinion on the relative merits of Sands and Young's ALSCOMP3 procedure and PARAFAC.

3.5 Constraints

As in most other models, one may impose constraints on the parameters of the PARAFAC model. Much work in this area has been done by Carroll and co-authors (Carroll, Pruzansky and Kruskal, 1980; Carroll and Pruzansky, 1984; Carroll, De Soete and Pruzansky, 1989). Their first two papers deal with finding solutions for the case that the constraints that various parameters are linearly related to prespecified variables or a design matrix have to be satisfied. The solutions are found by using the basic PARAFAC algorithm but now applied to a reduced problem which results from applying the design matrices to the components. This procedure was christened CANDELINC (CANonical DEcomposition with LINear Constraints). Carroll, De Soete and Pruzansky (1989) developed an algorithm using iterative reweighted least squares to produce best asymptotically normal estimates for the model parameters, which incorporated nonnegativity requirements at the same time. One may also put orthogonality constraints on PARAFAC components (see, e.g., Harshman and Lundy, 1984a), for instance to prevent certain kinds of degenerate solutions (Harshman,

Lundy and Kruskal, 1985; Kruskal, Harshman and Lundy, 1989; Lundy, Harshman and Kruskal, 1989).

Krijnen and Ten Berge (1991) developed variants of the basic PARAFAC algorithm to put nonnegativity constraints on the solution by using special least squares regression algorithms from Lawson and Hanson (1974). Durrell et al. (1990) refer to programs for three-way and four-way PARAFAC models (Lee, 1988) which also included nonnegativity constraints.

3.6 Additional Issues

In the above sections, the general focus has been on models and algorithms, but there are several issues in connection with these models which have not been mentioned so far. Very prominent, for instance, in Harshman's work, has been the question of preprocessing (i.e., centering and standardisation) of the data before the three-way analysis. Harshman and Lundy (1984b) discuss this issue in great detail touching on both algebraic and practical aspects (see also Kroonenberg, 1983). Ten Berge and Kiers (1989) and Ten Berge (1989) provide some theoretical results with respect to the iterative centering and standardisation proposed by Harshman and Lundy.

Another issue in this context is the postprocessing of output, that is, representation, graphing and transformations of the basic output of the programs to enhance interpretability (see especially Harshman and Lundy, 1984b; Kroonenberg, 1983).

Smilde (1992) raises the issue of variable selection for three-way data, as well as the problem of nonlinearities in the data and their effect on the solutions. These issues can also be seen as a serious concern in such areas like ecology where nonlinearities are the rule rather than the exception (see, e.g., Faith, Minchin and Belbin, 1987).

A final point is that within the framework of the analysis of covariance structures, McDonald (1984) has discussed the PARAFAC model, cited its limitations and proposed an altogether different (stochastic) approach to the kind of three-way data psychologists often encounter.

4. CONCLUSION

With the above comments, I have attempted to give a rough outline of research on the PARAFAC model. The model itself is only one of several conceivable models for three-way data, but a fully fledged exposé is not feasible here. What makes the PARAFAC model special is that it has a unique solution, a situation which is fairly unique in three-way land.

Comment

Donald S. Burdick

Multilinear models are fascinating because of the richness of their mathematical structure and the usefulness of their applications. The authors have done a fine job of presenting both of these features. I welcome their paper and hope that it has the effect of stimulating interest in this important topic.

Having said that, I must add my opinion that it is a mistake to shy away from tensors. The geometry of tensor products can be a source of valuable insight when struggling with the complicated details of multilinear algebra. The geometric perspective is especially useful when trying to make sense out of the nonuniqueness that occurs when model parameters are not identifiable.

For example, the concept of tensor products of vector

Donald S. Burdick is Associate Professor, Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina 27706.

spaces can shed light on the structure of the T3 model. Let Y denote an $I \times J \times K$ data array and write

$$Y = \mu + e$$

where μ is given by (19). The data array Y is unconstrained, which is tantamount to saying that Y is an arbitrary vector in $R^I \otimes R^J \otimes R^K$, the tensor product of real Euclidean spaces of dimensions I, J and K, respectively. The array μ, however, is constrained by expression (19). What is the nature of that constraint? Expression (19) stipulates that μ lie in $\mathcal{A} \otimes \mathcal{B} \otimes \mathcal{C}$, where $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are the respective subspaces of R^I, R^J and R^K spanned by the columns of A, B and Γ, respectively. The least squares fit of μ to Y is the projection of Y on $\mathcal{A} \otimes \mathcal{B} \otimes \mathcal{C}$. From a geometric perspective, the nonidentifiability is obvious, because the projection of a data vector on a subspace is unaffected by changes in the basis spanning the subspace. Replacing A by AM amounts to no more than a change of basis for $\mathcal{A}$.