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 (PARAFAC) 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