Journal of Multivariate Analysis 132 (2014) 1–8

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A contribution to the visualisation of three-way arrays

Casper J. Albers [a,*], John C. Gower [b]

[a] Department of Psychometrics & Statistics, University of Groningen, Groningen, The Netherlands

[b] Department of Mathematics & Statistics, The Open University, Milton Keynes, United Kingdom

ARTICLE INFO

Article history:
Received 10 March 2014
Available online 7 August 2014

AMS subject classifications:
62H25
15A72
62-07
00A66

Keywords:
Biplots
Candecomp decomposition
PARAFAC
Multiway data

ABSTRACT

Visualisations of two-way arrays are well-understood. Here, a procedure, with geometric underpinning, is given for visualising rank-two three-way arrays in two-dimensions.

1. Introduction

In the analysis of multiway tables, model-fitting, interpretation and visualisation of interactions are important phases. Much of the standard exposition is based on linear models, generalised linear models [16], or generalised additive models [11]. In all such models, the estimation of additive parameters, which may include additive interaction terms, is well understood, but they are not conducive to visualisation. As early as [5], multiplicative/additive parametrisations were considered but because of computational limitations it was not until the computer revolution that they became popular, especially in genotype–environment studies. The estimates of biadditive model multiplicative parameters are routinely visualised in maps of two sets of points; one arising from the row-parameters and the other from the column-parameters. These biplot maps greatly help the interpretation of interactions between two factors. Usually, such maps are two-dimensional; though higher dimensional biadditive fits are easy to compute they are less easy to interpret by visual inspection. For a brief discussion of the term biadditive and its extension to triadditive as used below, see [4].

Here we discuss one possibility for mapping three-factor multiplicative interactions. As with visualisations of two-way arrays, our methods are rooted in Euclidean geometry, orthogonal projections and calibrated axes. Reference to these topics and their mathematical derivations can be found in, e.g., [15,7]. Consider a decomposition of a three-way array $\mathbf{X} = \{x_{ijk}\}$ ($i = 1, \ldots, I; j = 1, \ldots, J; k = 1, \ldots, K$):

$$x_{ijk} = \sum_{r=1}^{R} u_{ir} v_{jr} w_{kr} \tag{1}$$

which has the form of a Singular Value Decomposition of a matrix generalised to three-way arrays, although crucially without its nice orthogonal least squares properties.

* Corresponding author.
E-mail address: c.j.albers@rug.nl (C.J. Albers).

http://dx.doi.org/10.1016/j.jmva.2014.07.013
0047-259X/

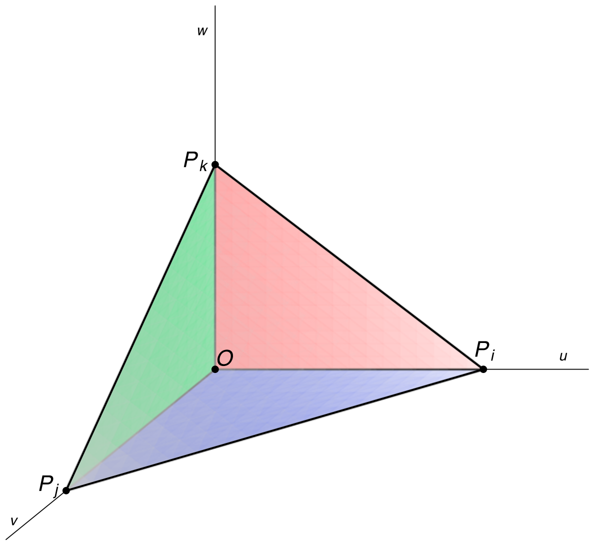

Fig. 1. Rank one visualisation in three dimensions. The value x_{ijk} is the volume of the tetrahedron with vertices O, P_i, P_j, P_k.

We adopt the terminology introduced by [14] of referring to a decomposition (1) as having rank R when it cannot be written in fewer terms. A growing literature on three-way rank has been summarised by [18], though it has little overlap with this paper. The basic algorithm for fitting (1) was independently due to [2,10], who termed it Candecomp and PARAFAC, respectively, although the basic idea goes back 43 years earlier [12]. Examples of the varieties of ways that three-way arrays may arise in applications can be found in [13,17]. Here, the neutral term "array" is used as an omnibus term to cover cross-classified tables, sets of data-matrices, configuration matrices and other three-way constructs. A three-way array $\mathbf{X}$ may represent raw data, or it may have been derived as an approximation to raw data, or it may be a term in some, possibly extensive, model. Thus, although one could analyse any rank-2 three-way array as raw data, often some preprocessing is required, as in our example below. All we seek here is a good visualisation of $\mathbf{X}$, however it may have been derived.

We define $\mathbf{U}$ with columns $\{\mathbf{u}_i\}$, $\mathbf{V}$ with columns $\{\mathbf{v}_j\}$, and $\mathbf{W}$ with columns $\{\mathbf{w}_k\}$, each with R columns. For reasons made clear below, we assume without loss of generality that $I \le J \le K$. Although matrices written $\mathbf{U}$ and $\mathbf{V}$ are often associated with expositions of the SVD, expressed in orthogonal forms, we emphasise here that no such restrictions apply in the following.

The paper is structured as follows. Section 2 gives (i) details of what is required for computations and (ii) the geometric and algebraic underpinning. Section 3 gives an example of our method and we finish with a discussion in Section 4.

2. Visualisation

It is clear from the previous section that visualisation is important in the interpretation of two-way arrays. Below, we explore to what extent three-way arrays may be visualised in two dimensions.

2.1. Basic method

In the rank one case ($R = 1$), the points for u_{i1} $(i = 1, \dots, I)$; v_{j1} $(j = 1, \dots, J)$; w_{k1} $(k = 1, \dots, K)$ may be placed on separate orthogonal coordinate axes, which we shall label u, v and w. Then, $x_{ijk} = u_{i1}v_{j1}w_{k1}$ is simply proportional to the volume of the tetrahedron with the three points $P_i = (u_{i1}, 0, 0)$, $P_j = (0, v_{j1}, 0)$, $P_k = (0, 0, w_{k1})$ and the origin as vertices cf. [7], as shown in Fig. 1.

When $R = 2$, we may inspect the two-dimensional sets of coordinates U, V and W but this ignores intrinsic three-way information. Then, one way of proceeding is to write

$$x_{ijk} = u_{i1}v_{j1}w_1 + u_{i2}v_{j2}w_2 \tag{2}$$

where (w_1, w_2) takes on the values (w_{k1}, w_{k2}) as k varies in the range $(1, \dots, K)$ while i and j remain fixed. The line represented by (2) may be envisaged in the two-dimensional plane of the K coordinates comprising the rows of $\mathbf{W}$. We term

$$(u_{i1}v_{j1})w_1 + (u_{i2}v_{j2})w_2 = 0$$

the *zero line*. The (Euclidean) distance of any point (w_{k1}, w_{k2}) from the zero line is

$$\kappa_{ij}(u_{i1}v_{j1}w_{k1} + u_{i2}v_{j2}w_{k2})$$

which is proportional to the required triadditive form with $R = 2$. The factor κ_{ij} ensures that the coefficients of (w_1, w_2) are normalised in the form of direction-cosines, as is required, and is given by $\kappa_{ij}^{-2} = (u_{i1}v_{j1})^2 + (u_{i2}v_{j2})^2$. When different values k' are used, the lines (2) remain parallel, i and j remaining fixed. The distance from the zero line increases with x_{ijk}. When