(Interactive) Visualisation of Three-Way Data
Casper J. Albers¹, John C. Gower²

₁ Department of Psychometrics and Statistics, University of Groningen, The Netherlands, (e-mail: c.j.albers@rug.nl; url: casperalbers.nl)
₂ Department of Mathematics and Computing, The Open University, United Kingdom.

Keywords: Candecomp/Parafac, biplots, triplots, multiway data.

1 Introduction
We are concerned with three-way tables $X$ with elements $x_{ijk}$ ($i = 1, \ldots, I$; $j = 1, \ldots, J$; $k = 1, \ldots, K$). Thus, the factors used to classify the three ways have equal status (sometimes called modes) while the body of the table contains values of a quantitative variable that may be regarded as a dependent variable - as classically typified by a three-way table arising from agricultural experiments with fertilizer treatments as factors and crop yield as the response. Three-way tables are usually analysed by linear models containing additive terms representing main effects, two-factor interactions, and three-factor interactions. The number of factors can be readily extended to any number of “ways”. For reference, and to establish notation, we list the basic results for additive models. The model is

$$x_{ijk} = m + \{a_i + b_j + c_k\} + \{d_{jk} + e_{ik} + f_{ij}\} + g_{ijk} \quad (1)$$

where the terms with a single suffix represent main effects, those with double suffixes two factor interactions and $g_{ijk}$ represents contributions from three factor interactions. Some components of the interactions may be regarded as “error”. Estimation of main and biadditive effects is usually done through Analysis of Variance (ANOVA)-like methods; whereas specific three-way methods, such as Candecom/Parafac and Tucker3, are employed for the triadditive term.

When interactions have been estimated, there remains the problem of their interpretation. To help interpret overall representations of interaction, several simple approximations have been proposed. One possibility is to focus on the larger (positive or negative) terms. Another is to fit linear and quadratic polynomials to get, for example, linear $\times$ linear $\times$ quadratic estimates. Even the simpler of these can be difficult to interpret and, strictly speaking, such expressions are valid only when the classifying factors are numerical (like levels.
Another possibility, developed mainly for two-way tables, is to fit product terms like \( a_i b_j \). Products of two factors have bilinear regression interpretations and a nice geometrical representation that underpins useful visualisations of two-factor interaction.

When analysing three-way data, one often includes triple product terms. Such terms appear in three-mode principal component analysis (Tucker3), Individual Differences Scaling, and the Candecomp/Parafac model. The default Candecomp/Parafac rank-\( R \) approximation to \( X \) is given by

\[
\hat{x}_{ijk} = \sum_{r=1}^{R} u_i r v_j r w_k r,
\]

where the solution is found by minimising \( \sum_{i,j,k} (x_{ijk} - \hat{x}_{ijk})^2 \). Rather than approximating the array \( X \), one could also approximate the array \( G \) with the residuals after removing main and biadditive effects. It is clear that triple product terms may be potentially useful in many contexts and considered as a natural extension for representing triadditive interactions in a similar way that biadditive models may represent two-factor interactions.

The focus of the presentation lies on the creation of visual displays of these bi- and triadditive interactions, in such a way that these visualisations aid the researcher in grasping the key properties of the data. We focus on visualisation using the Candecomp/Parafac decomposition, but our theory is generalisable to other types of decomposition.

2 Three dimensional visualisation

For decompositions of rank up to \( R = 4 \), it is possible to visualise the triadditive terms in three dimensions. In the rank one case (\( R = 1 \)), the points for \( u_{i1} \) \((i = 1, \ldots, I)\); \( v_{j1} \) \((j = 1, \ldots, J)\); \( w_{k1} \) \((k = 1, \ldots, K)\) may be placed on separate orthogonal coordinate axes, which we shall label \( u \), \( v \) and \( w \). Then, \( u_{i1} v_{j1} w_{k1} \) is simply proportional to the volume of the tetrahedron with these three points and the origin as vertices (Figure 1, left).

When \( R = 2 \), the visualisation remains basically Euclidean in three dimensions and it may be interpreted in terms of tetrahedral volume where the vertices of the tetrahedra are confined to the origin and three orthogonal planes (Figure 1, right). The justification of this approach follows from the trilinear identity:

\[
\det \begin{pmatrix}
0 & u_{i1} & u_{i2} \\
v_{j2} & 0 & v_{j1} \\
w_{k1} & w_{k2} & 0
\end{pmatrix} = u_{i1} v_{j1} w_{k1} + u_{i2} v_{j2} w_{k2}.
\]
Figure 1. Rank $R = 1$ (left) and $R = 2$ (right) fits to triadditive terms for a given data set. For the $R = 1$ fit, all points lie on orthogonal axes, for the $R = 2$ fit, they all lie on orthogonal planes. The volumes of the tetrahedra corresponds to the interaction “low nitrogen $\times$ Edinburgh $\times$ Kinsman”.

The rows of the determinant on the left hand side may be interpreted as giving the coordinates of three points, one in each of three orthogonal dimensions, while the right hand side gives a term in the rank two triadditive model. (Visualisations of rank $R = 3$ and $R = 4$ are also possible, yet do not lend themselves for easy interpretation).

3 Two-dimensional visualisation

We shall show that, without loss of information, this rank $R = 2$ representation may be shown in two dimensions to give a visualisation which resembles a biplot, with one set of $K$ coplanar points and two sets of calibrated axes representing the remaining $IJ$ factors. Thus, it is a triplot rather than a biplot. The interpretation is similar to that of a biplot. That rank-two trilinear interactions may be shown in two dimensions, puts them on the same status as interactions for bilinear models and makes direct three-dimensional tetrahedral visualisations unnecessary. We believe that this is a major step forward. An example of such a triplot is given in Figure 2. One of the strengths of this method is that it allows for calibration of either the row-axes, the column-axes, or even both. This makes it easier to read off the interaction.

Rank two triplot displays in two dimensions, for a low number $I, J, K$, such as Figure 2, seem to be at the bounds of practical utility: larger arrays lead to visual overload. For larger data arrays, one needs either to be able to interact with the visualisation or make other, smart choices to reduce visual overload, perhaps at the cost of completeness. Figure 3 displays such a choice: here
Figure 2. Visualisation of the rank $R = 2$ approximation to the biadditive interaction between factors $B$ and $C$. Left, a regular biplot is given. The same information is shown on the right, now using calibrated axes for the locations.

Figure 3. Visualisation of Figure 2 through parallel axes.

all axes of Figure 2 have been displayed as consecutive parallel axes. Some information (the angles between axes) is lost, yet the interpretability is larger.

References


ALBERS, CASPER J., & GOWER, JOHN C. 2015. Visualising interaction in bi- and triadditive models for three-way tables. (Submitted for publication. Copies available upon request.).