Carroll, J.D. & Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an *N*-way generalization of "Eckart-Young" decomposition. *Psychometrika, 35,* 283–319 (3320 citations in Google Scholar as of 4/1/2016).

No doubt this paper was Doug Carroll's most influential contribution. It introduced a method for estimating all parameters of the weighted Euclidean model for multidimensional scaling, in which the weights provide a simple way to describe individual differences in similarity data. That model had been originally proposed by Horan (1969), but after publication of the Carroll and Chang paper it became widely known as the INDSCAL model[1]. Their method for finding parameter estimates for the weighted Euclidean model relied on the new CANDECOMP model and algorithm, which stands for CANonical DECOMPosition of *N*-way tables. Jih-Jie Chang had done all the programming.

It is important to appreciate the distinction between model and algorithm, both carrying the same name here, because they brought two distinct innovations. The CANDECOMP model is a generalization of the Eckart-Young decomposition mentioned in the title which was published in the first volume of *Psychometrika* (Eckart and Young, 1936), while the CANDECOMP algorithm is an extension of the iterative estimation scheme for principal components analysis, proposed by the econometrician Herman Wold (1966) under the name Nonlinear Iterative Least Squares (NILES). These novel ideas were presented in full generality, and they energized the emerging field of three-way analysis. That is why the Carroll and Chang paper is not only cited by quantitative psychologists, but also by a much larger three-way community.

In multidimensional scaling (MDS), the interest is to analyze the mutual similarity among stimuli in terms of a few basic dimensions; if we have m individuals, each judging the similarity of pairs formed out of n stimuli, the data can be collected in an $n \times n \times m$ three-way array. In earlier approaches such three-way similarity data were usually averaged across individuals, on the assump-

[1] The Horan paper was submitted to *Psychometrika* in September 1964, but because the author was killed in a road accident before the completion of his PhD work, his supervisor John Ross took care of finalizing its publication.

tion that the same basic dimensions would be operative in each of them, and that individual differences would be merely random variation, which could be ignored. One notable exception had been the approach proposed by Tucker and Messick (1963), which first factor-analyzed the m similarity matrices (strung out into vectors) to form p factors of individuals called "Points of View" (PoV), and then performed p separate multidimensional scaling (MDS) analyses on the weighted average similarities (with the factor loadings as weights and the mean similarities reassembled into a matrix for each PoV).

Carroll and Chang started by clearly voicing earlier criticisms of PoV analysis: "Perhaps the most cogent criticism is that the method is little more powerful than doing separate scalings on the individual subjects—and it makes no explicit assumptions about possible or probable communality of the dimensional structures for different real or idealized individuals. It would be very surprising if the various configurations had no structure in common" (C&C, p. 284). Then they assumed again that one common set of r dimensions would be *potentially* operative in each individual, but now each dimension may have a different relative *salience*, or *importance*, for different people. Suppose x_{jt} denotes the coordinate value of the jth stimulus on the tth dimension of the common space ($j = 1, \ldots, n$ and $t = 1, \ldots, r$). Then the "modified" Euclidean distance for the ith individual was defined as

$$d_{ijk} = \sqrt{\sum_{t=1}^{r} w_{it}\left(x_{jt} - x_{kt}\right)^2}. \tag{1}$$

Here, the weights w_{it} represent the salience of dimension t for individual i (with $i = 1, \ldots, m$). It is easily seen that if we define

$$y_{ijt} = w_{it}^{1/2} x_{jt}, \tag{2}$$

then the ordinary Euclidean distances based on the y-values defined in (2) produce the individual distances defined in (1). So each individual is assumed to perceive distances in a space where the common dimensions have been (differentially) expanded or contracted, and ignored if $w_{it} = 0$.

The next step was to apply the classic Young-Householder-Torgerson transformation (Young and Householder, 1938; Torgerson, 1952) to the distances defined in (1); that is, for each individual separately, the row- and column means of the matrix with elements $-\frac{1}{2}d_{ijk}^2$ are subtracted from its entries (called *double centering*). This transformation gives m matrices of scalar products with elements b_{ijk} related to the parameters of the model as

$$b_{ijk} = \sum_{t=1}^{r} y_{ijt} y_{ikt} = \sum_{t=1}^{r} w_{it} x_{jt} x_{kt} , \tag{3}$$

by using (1) and substituting (2). It is worth noting that Horan (1969) had already obtained result (3), and that he had used it for arguing that the average of the scalar products across individuals would result in an $n \times n$ matrix **B**, of which the eigenvector-eigenvalue decomposition could be used to estimate the stimulus coordinates x_{jt}.

It was here that Carroll and Chang proceeded differently, using a clever maneuver. Instead of (3), they considered the more general model

$$z_{ijk} = \sum_{t=1}^{r} a_{it} b_{jt} c_{kt} , \tag{4}$$

which is the CANDECOMP model already mentioned. Any method to estimate the parameters of the model (4) could be used to get estimates of the parameters of the INDSCAL model, by just identifying $a_{it} = w_{it}$, $b_{jt} = x_{jt}$, and $c_{kt} = x_{kt}$. Presumably, the symmetry of the data would carry over to the fitted expected values having the property $\hat{z}_{ijk} = \hat{z}_{ikj}$ for all i, j and k, and that in turn would guarantee that the estimates of the matrices { b_{jt} } and { c_{kt} } coincide to yield one estimate of { x_{jt} }.

To switch from (3) to (4) was a stroke of genius, and the most original contribution of the paper. Note that in (3) there are terms that are quadratic in the x's for $j = k$, while all terms in (4) are trilinear, implying that equation (4) is a linear model in the a's when the b's and c's are fixed, linear in the b's when the other two sets of parameters are fixed, and likewise for the c's. Given some initial estimates of the three sets of parameters, the CANDECOMP algorithm iteratively solves for the parameters of one set given temporarily fixed values of the other two sets by linear regression, and continues with partial solutions until the process stabilizes. Then the INDSCAL algorithm is just the CANDECOMP algorithm with the identification stated before, with certain normalization constraints.

About the origin of the idea of the CANDECOMP model and algorithm, Carroll and Chang gave the following credits: “This method of “canonical decomposition” was suggested to us by a paper by Herman Wold (1966) […], in which a related method of decomposition of two-way tables was discussed […], involving what Wold calls a NILES (for “Nonlinear Iterative Least Squares”) procedure. In the same paper […] Wold suggested the more general three-way model discussed here, but did not describe a computational scheme for this model (except for the special case of one dimension)” (C&C, p. 312).

It may have occurred to the reader that in the one-dimensional case, least squares estimation under model (4) of the *a*’s when the *b*’s and *c*’s are fixed amounts to a series of *simple* linear regressions, while in the more-dimensional case one needs a series of *multiple* linear regressions — not quite a big conceptual leap, I would say. In addition, Carroll and Chang appeared to be unaware of other work that Wold and his co-workers had done, reported in Wold and Lyttkens (1969). They considered a whole range of estimation procedures under the new name NIPALS (Nonlinear Iterative Partial Least Squares), which was later abbreviated to PLS. Among these procedures, the full-blown CANDECOMP algorithm was described by Alf Israelson in a section called “Three-way (or second order) component analysis” (o.c., pp. 32-33)[2]. It also became soon apparent that the CANDECOMP model had been conceived independently by Richard Harshman (1970) under the name PARAFAC (for *parallel factors*), for which Bob Jennrich had developed an algorithm that was identical to CANDECOMP (o.c., p. 32).

The larger part of the Carroll and Chang paper consists of:

- Extensive descriptions of two illustrative datasets collected by their colleagues Sandy Pruzansky and Mike Wish from Bell Labs;
- Persistent discussions of the strong points of the INDSCAL model (unique and meaningful orientation of the axes, thus obviating the rotational problem; easier interpretation of individual differences);
- Detailed comparisons with other work in three-way MDS and three-way component analysis, and several possible modifications of INDSCAL and CANDECOMP to accommodate these.

[2] I am indebted to Pieter Kroonenberg for pointing out this reference to me.

This emphasis on strong properties and flexibility of the methods, as well as on new perspectives that three-way data may bring over and above two-way data makes it a prototypical *programmatic* paper. It generated a lot of follow-up work about theoretical and technical aspects of three-way models and their extensions. But the high citation count is especially due to the fact that INDSCAL and CANDECOMP became the methods of choice in a wide variety of applications, not only in the behavioral sciences and psychometrics, but notably also in chemometrics and signal processing (*cf*. Kroonenberg, 2014).

Willem Heiser, July 2016

References

Eckart, C. & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika, 1*, 211-218.

Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis. *UCLA Working Papers in Phonetics, 16,* 1-84.

Horan, C.B. (1969). Multidimensional scaling: Combining observations when individuals have different perceptual structures. *Psychometrika, 34*, 139–165.

Kroonenberg, P.M. (2014). History of multiway component analysis and three-way correspondence analysis. In Blasius, J. & Greenacre, M.J. (Eds.). *Visualization and verbalization of data* (pp. 77–94). Boca Raton: Chapman & Hall/CRC.

Torgerson, W.S. (1952). Multidimensional scaling: I Theory and Method. *Psychometrika, 17*, 401-419.

Tucker, L.R. & Messick, S. (1963). An individual difference model for multidimensional scaling. *Psychometrika, 28*, 333-367.

Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In: P.R. Krishnaiah (Ed.), *Multivariate Analysis*, pp. 391-420. New York: Academic Press.

Wold, H. & Lyttkens, E. (Eds.) (1969). Nonlinear iterative partial least squares (NIPALS) estimation procedures. *Bulletin of the International Statistical Institute, 43*, 29-51.

Young, G. & Householder, A.S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika 3*, 19-22.