Multidimensional Joint Graphical Display of Symmetric Analysis: Back to the Fundamentals

Shizuhiko Nishisato

Abstract The basic premise of dual scaling/correspondence analysis lies in the simultaneous or symmetric analysis of rows and columns of the data matrix, a task that resembles the analysis of principal component analysis of both the person-to-person correlation matrix and the item-by-item correlation matrix together. Our main quest: whether or not we can represent both analyses in the same Euclidean space. The traditional graphical methods are very problematic: symmetric display or French plot suffers from the discrepancy between the row space and the column space; non-symmetric display involves the projection of data onto standardized space, which does not contain coordinate information in the data; a variety of biplots, of which criticisms we rarely see, involve operations that do not typically maintain row and column measurements on the equal metrics, or if they do they are not the coordinates of the data. Thus, none of these provides a precise description of complex information in data, hence failing in the basic objective of symmetric data analysis. This paper will identify logical problems of the current practice and offers a justifiable alternative to joint graphical display. "Graphing is believing" may in reality remain to be a wishful thinking.

Keywords Duality • Joint space for rows and columns • Doubling multidimensional space

1 Introduction

This paper deals with graphical display of quantification theory, where the main interest lies in the joint analysis of rows and columns of the data matrix. This aspect is reflected by the word 'dual' of Canadian dual scaling (Nishisato 1980) used to treat rows and columns of a data matrix on the equal footing, that is, symmetric analysis of the data matrix. The technique is referred to by many other names such as British simultaneous linear regressions (Hirschfeld 1935), the

S. Nishisato (✉)
University of Toronto, Toronto, ON, Canada
e-mail: shizuhiko.nishisato@utoronto.ca

American method of reciprocal averages (Horst 1935), Hayashi's Japanese theory of quantification (1950), American principal component analysis of categorical data (Torgerson 1958), American optimal scaling (Bock 1960), French 'analyse des correspondances' (Escofier-Cordier 1969), and Dutch homogeneity analysis (De Leeuw 1973). See many other names in Nishisato (2007).

In the traditional multivariate analysis, we often use the least-squares procedure, which means projection of, for example, data onto the model space, meaning a one-directional analysis as opposed to the two-way symmetric analysis of equal norms. Graphical display of quantification results must be such that the norm of the row variables should be equal to the norm of the column variables. This is a difficult task for joint graphical display of quantification theory, and in the past a number of methods have been proposed, none of which, however, is satisfactory. The current paper starts with some basic premises of quantification, and then discusses how the perennial problem of joint graphical display should be dealt with. We start with some relevant basic points.

2 Fundamental One: Orthogonal Coordinates for *n* Variables

When we wish to show a graph of two sets of scores (e.g., Mathematics test and language test), it is a widely used practice to introduce the horizontal axis for the mathematics test and the vertical axis for the language test as if the two variates were orthogonal to each other. This is definitely wrong, but this practice has been used widely for many years. When we have a number of variables, say n, where $n > 1$, the first task for graphical display is to introduce an orthogonal coordinate system to accommodate these variables. There are an infinite number of such systems, and the most widely used choice, out of them, is to adopt principal coordinates, through principal component analysis: Given the subject-by-test data matrix, $\mathbf{F}$, we calculate the test-by-test correlation matrix R, which is then subjected to the eigenvalue decomposition, that is, $\mathbf{R} = \mathbf{X}' \mathbf{\Delta} \mathbf{X}$, where $\mathbf{X}$ is the subject-by-test matrix of coordinates and $\mathbf{\Delta}$ is the diagonal matrix of eigenvalues. The number of non-zero elements of $\mathbf{\Delta}$ is the required dimensionality of the space for multidimensional coordinates of n variables.

3 Fundamental Two: Coordinates of Framework and Variables

In this principal axis decomposition of data matrix $\mathbf{F}$, $\mathbf{X}$ is referred to as the matrix of *standard coordinates* and $\mathbf{\Delta}^{1/2}\mathbf{X}$ is called the matrix of *principal coordinates*. It is crucial for graphical display to distinguish between these two coordinates. Nishisato (1996) explained the important difference between them using a simple example

as follows: Consider principal component analysis of standardized variables, and suppose that the data are two-dimensional, then plotting principal coordinates of variables results in a perfect circle with the diameter 1, where all data points lie; suppose that the data are perfectly three-dimensional, then plotting the principal coordinates of the data reveals that all data points lie at a distance of 1 from the origin on the three-dimensional sphere, or on the perfect ball. If we plot standard coordinates, instead of principal coordinates, however, the two-dimensional data will show, not a perfect circle, but typically an elongated circle. If the first eigenvalue is comparatively larger than the second one, the graph will be elongated toward the second dimension. In other words, standard coordinates do not describe the structure of the data, but a function of the distribution of data under the condition that the sum of squares on each dimension is constant, thus the name standard (*i.e.,*the fewer the responses the larger the standard coordinates). The conclusion here is that the coordinates of variables in multidimensional space are given by principal coordinates.

4 Fundamental Three: Dual Relations

Quantification theory can be depicted as singular value decomposition of data matrix $\mathbf{F}$, that is, $\mathbf{Y}\mathbf{\Lambda}\mathbf{X}'$, where $\mathbf{Y}$ and $\mathbf{X}$ are standard coordinates of rows and columns, respectively, and $\mathbf{\Lambda}$ is the diagonal matrix of singular values. Because of the symmetry of this analysis, Nishisato (1980) called it dual scaling, based on the dual relations:

$$\rho_k y_{ik} = \frac{\sum_{j=1}^{m} f_{ij} x_{jk}}{f_i} \quad \text{and} \quad \rho_k x_{jk} = \frac{\sum_{i=1}^{n} f_{ij} y_{jk}}{f_j}$$

where ρ_k is the k-th singular value, f_{ij} is the element of the i-th row and the j-th column, $f_{i.}$ and $f_{.j}$ are respectively the sums of the i-th row and that of the j-th column of data matrix $\mathbf{F}$. In other words, for each component k, the mean of rows i of $\mathbf{F}$, weighted by column weights x_j is equal to the weight for row i times the singular value, and the mean of column j, weighted by row weights y_i is equal to the weight for column j times the singular value. This mutual reciprocal averaging relation holds for each component. Although ρ_k is the singular value of data matrix $\mathbf{F}$, it is also (1) Hirschfeld's simultaneous regression coefficient (1935), (2) Guttman's maximal row-column correlation (1941) and (3) Nishisato's (1980) projection operator from row space to column space or vice versa.

5 Fundamental Four: Discrepancy Between Row Space and Column Space

For a particular component, the dual relation shows that the mean of the row i, weighted by column weights x_j, is equal to the weight for row i times the singular value. In other words, the singular value is the projection operator of the row space onto the column space, or vice versa. Thus, it is possible to calculate the angle of discrepancy, θ_k, between the row space and the column space for component k by the following formula (Nishisato & Clavel 2008):

$$\theta_k = \cos^{-1}\rho_k$$

From this we know that only when the singular value is one the variables associated with rows and columns of the data matrix span the same space. In this regards, we should remember the famous warning by Lebart, Morineau and Tabard (1977) that one cannot calculate the exact distance between a row variable and a column variable from the symmetric scaling.

6 Lessons From Analysis of Contingency Table and Response-Pattern Table

Using an example from Nishisato (1980), some important aspects of joint graphical display can be illustrated to clarify the current controversies of joint graphical display.

Consider the following 2×3 contingency table, **C**, obtained by asking two multiple-choice questions:

Q1: Do you smoke? (yes, no)
Q2: Do you prefer coffee to tea? (yes, not always, no)

Suppose we obtained the following data indicated by **C**, which is the 'options of Q.1-by options of Q.2,' that is, a 2×3 table of joint frequencies. Nishisato (1980) has shown that the same data can be represented also as the traditional response-pattern table $\mathbf{F}_a$, which is the 'subjects-by-options of two items,' that is, 14×5 incidence table. He has also shown that this large table can be transformed into a condensed response-pattern table $\mathbf{F}_b$ by creating a table of distinct patterns with frequencies. In our example, the data in the three data formats are as follows:

$$\mathbf{C} = \begin{bmatrix} 3 & 2 & 1 \\ 1 & 2 & 4 \end{bmatrix}; \quad \mathbf{F}_a = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}; \quad \mathbf{F}_b = \begin{bmatrix} 3 & 0 & 3 & 0 & 0 \\ 2 & 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & 2 & 0 \\ 0 & 4 & 0 & 0 & 4 \end{bmatrix}$$

As Nishisato (1980) has shown, the two response-pattern formats yield identical quantification results. Therefore, for brevity we will use $\mathbf{F}$.

Suppose that two items have n and m options, respectively, and there are N respondents. Then, assuming that N is much larger than the sum of the response categories, the total number of components from the $n \times m$ contingency table, $K(\mathbf{C})$, is equal to the smaller of n and m minus 1, that is,

$$K(\mathbf{C}) = \min(n, m) - 1.$$

In the current example, $\min(2,3) - 1 = 2 - 1 = 1$. Assuming that N is much larger than $n + m$, the total number of components from the response-pattern table, $K(\mathbf{F})$, is equal to the total number of categories of two items minus 2, that is,

$$K(\mathbf{F}) = n + m - 2.$$

In the current example, $K(\mathbf{F}) = 2 + 3 - 2 = 3$.

According to the Young-Householder theorem (Young & Householder 1938), the variates within columns (or, rows) of the data matrix can be mapped in the same Euclidean space. Thus, the coordinates of those five columns of $\mathbf{F}$ can be mapped in the same Euclidean space. In contrast, we have already shown that the two rows and the three columns of $\mathbf{C}$ do not belong to the same space. From this comparison, we can draw the conclusion that the five options of the two items require three-dimensional space to be plotted together. Our numerical example (Table 1) yields the following coordinates on respective dimensions. Notice that the standard coordinates associated with $\mathbf{C}$ are exactly the same as the standard coordinates of the corresponding first component of $\mathbf{F}$:

Several years after Nishisato's book was published, Carroll, Green and Schaffer (1986) wrote a paper on the method called the CGS scaling, in which they maintained that the space discrepancy between row and column space of the

Table 1 Standard coordinates associated with the two formats of data

		The results of **C**	The results of $\mathbf{F}_b$		
Component		1	1	2	3
Smoking 'yes'		1.08	1.08	0.00	−1.08
Smoking 'no'		−0.93	−0.93	0.00	0.93
Coffee 'yes'		1.26	1.26	−1.14	1.26
	'not always'	0.17	0.17	2.11	0.16
	'no'	−1.14	−1.14	−0.77	1.14

contingency table could be solved by representing the rows and the columns of the contingency table into the same columns of the response-pattern table—this is exactly what was shown above. However, the CGS scaling was severely criticized by Greenacre (1989) as false, and his criticism resulted in the downfall of the CGS scaling. What these investigators completely missed was the point that the weights for the rows and those for the columns of the contingency table require more dimensions if they are represented in the same rows of the response-pattern table. In the above example, one needs three dimensions. In the above example, the singular value of the component associated with the contingency table is 0.4590, thus the discrepancy angle between the row axis and the column axis is 62.68°, leading to the conclusion that we need more than one dimension for the data. The idea of the CGS scaling should have been presented under the condition that the space dimensionality must be at least doubled from that of the contingency table.

7 Dimensionality of Total Space

In the above comparison of the contingency format and the response-pattern format, we concluded that those response options of the two items can be mapped in the same space, provided that the dimensionality of the space is expanded. There are two distinct views on how many dimension are needed. The first one is Nishisato's view of doubled multidimensional space (2012). His idea of 'doubling' comes from the consideration that for each component we must introduce two axes with the angle of $\cos^{-1}\rho_k$. His view looks reasonable, but we need another view on this: Based on the comparison between quantification of the contingency table and that of the corresponding response-pattern table, we need to double the dimensionality or more than double the dimensionality. This view stems from the following fact:

$K(\mathbf{F}) = 2 \times K(\mathbf{C})$, when $n = m$ and
$K(\mathbf{F}) > 2K(\mathbf{C})$, when $n \neq m$

In other words, only when the number of options of Item 1 is equal to that of Item 2, we need to double the dimensionality. Otherwise, as was the case of the above numerical example, we need more than double the dimensionality of the joint space.

8   From Joint Graphical Display to Cluster Analysis of Total Space

Nishisato (1997) wrote a paper on "Graphing is believing" in support of graphical display. With the current revelation, however, it seems generally impossible to summarize data in multidimensional space, for we are limited to grasp or understand only two- or three-dimensional graphs and the total space for the joint graphical display with principal coordinates is almost always greater than two or three. At this juncture, Nishisato and Clavel (2010) proposed total information analysis or comprehensive dual scaling: Extract all components from the data, calculate the within-row distance matrix, the between row-column distance matrix and the within-column distance matrix; subject this super-distance matrix to cluster analysis, to identify clusters in the total space as defined here. In this way, we do not have to concentrate only on major configurations, but can also look at other rare combinations of variables. (see Nishisato (2014) for a numerical example.) Total information analysis has not widely been applied to data analysis yet, but is definitely a logical and reasonable alternative to the traditional analysis via multidimensional joint graphical display.

9   Concluding Remarks

Historically, French correspondence analysis placed a major emphasis on joint graphical display. The current paper has identified a number of logical problems associated with joint graphical display, be it symmetric French plot, or non-symmetric plot, or biplot. A number of those logical problems prompted Nishisato and Clavel (2010) to propose total information analysis (TIA), which as explained in the current paper is free from any logical problems. It is hoped that through many applications of TIA to data we will learn further how practical and useful TIA is as an alternative to the traditional multidimensional joint graphical approach to data analysis.

References

Bock, R. D. (1960). Methods and applications of optimal scaling. *The University of North Carolina Psychometric Laboratory Research Memorandum*, No. 25.

Carroll, J. D., Green, P. E., & Schaffer, C. M. (1986). Interpoint distance comparison in correspondence analysis. *Journal of Marketing Research, 23*, 271–280.

De Leeuw, J. (1973). *Canonical analysis of categorical data.* Unpublished doctoral dissertation, Leiden University, Leiden, The Netherlands.

Escofier-Cordier, B. (1969). L'analyse factorielle des correspondances. *Bureau Universitaire de Recherche Operationnele. Cahiers, SérieRecherche* (Université de Paris), *13,* 25–29.

Greenacre, M. J. (1989). The Carroll-Green-Schaffer scaling in correspondence analysis: A theoretical and empirical appraisal. *Journal of Marketing Research, 26*, 358–365.

Guttman, L. (1941). The quantification of a class of attributes: A theory and method of scale construction. In P. Horst et al. (Eds.), *The Prediction of Personal Adjustment* (pp. 319–348). New York, NY: Social Research Council.

Hayashi, C. (1950). On the quantification of qualitative data from mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics, 2*, 35–47.

Hirschfeld, H.O. (1935). A connection between correlation and contingency. *Cambridge Philosophical Society Proceedings, 31,* 520–524.

Horst, P. (1935). Measuring complex attitudes. *Journal of Social Psychology, 6*, 369–374.

Lebart, L., Morineau, A., & Tabard, N. (1977). *Tecnniques de la description statistique: Méthodes et l'ogiciels pour l'analyse des grands tableaux*. Paris, France: Dunod.

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications*. Toronto, ON, Canada: University of Toronto Press.

Nishisato, S. (1996). Gleaning in the field of dual scaling. *Psychometrika, 1996*(61), 559–599.

Nishisato, S. (1997). Graphing is believing: Interpretable graphs for dual scaling. In J. Blasius & M. J. Greenacre (Eds.), *Visualization of categorical data* (pp. 185–196). London, England: Academic.

Nishisato, S. (2007). *Multidimensional nonlinear descriptive analysis*. Boca Raton, FL: Chapman & Hall/CRC.

Nishisato, S. (2012). Quantification theory: Reminiscence and a step forward. In W. Gaul, A. Geyer-Schults, I. Schmidt-Thieme, & J. Kunze (Eds.), *Challenges and the interface of data analysis, computer science and optimization* (pp. 109–119). New York, NY: Springer.

Nishisato, S. (2014). Structural representation of categorical data and cluster analysis through filters. In W. Gaul, A. Guyer-Schultz, A. Baba, & A. Okada (Eds.), *German-Japanese interchange of data analysis results* (pp. 81–90). New York, NY: Springer.

Nishisato, S., & Clavel, J. G. (2008). A note on between-set distances in dual scaling and correspondence analysis. *Behaviormetrika, 30*, 87–98.

Nishisato, S., & Clavel, J. G. (2010). Total information analysis: Comprehensive dual scaling. *Behaviormetrika, 37*, 15–32.

Torgerson, W. S. (1958). *Theory and Methods of Scaling*. New York, NY: Wiley.

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