Outcries of Dual Scaling: The Key Is Duality

Shizuhiko Nishisato

Abstract There are a number of points in the development of dual scaling which have escaped our attention. In my Beijing paper, problems with joint graphical display were discussed to fill the gap of understanding, and the current paper deals with some other points. These two papers can be regarded as a sequel to my paper, entitled "Gleaning in the field of dual scaling," written 20 years ago. Noting that the basic premise of dual scaling lies in duality of exhaustive analysis, we will look at a few more points in this paper. Outcry one is on linear and nonlinear analysis. As is well known, dual scaling is a method for simultaneous regressions of row variates and column variates on data, capturing all linear and nonlinear relations contained in the data. From this point of view, Likert scores, used as scores for data analysis, are far from satisfactory, for it is a strictly linear and data-independent procedure. Outcry two is on our definition of multidimensional quantification space, because the traditional framework needs to be modified so as to satisfy our objective, that is, describing both row and column structure of data in a symmetric comprehensive way. Outcry three is on a logical alternative to problem-plagued joint graphical display, and a recommended alternative is cluster analysis. Finally, outcry four is on the distinction between dual space and total space, leading to the suggestion that simple correspondence analysis fails to provide exhaustive analysis of information in data.

1 Introduction

In 1996, Nishisato presented his presidential address, entitled "Gleaning in the field of dual scaling," in which he identified a number of hidden or unsolved aspects

S. Nishisato (✉)

Professor Emeritus, University of Toronto, 30 Old Mill Road, Suite 308, Toronto, Ontario M8X 0A5, Canada

e-mail: shizuhiko.nishisato@utoronto.ca

of dual scaling (Nishisato 1996). It is 20 years since then, and one wonders if dual scaling is well understood and if some of the problems raised then have been solved to our satisfaction. Some major problems were discussed in the paper, entitled "Multidimensional joint graphical display of symmetric analysis: Back to the fundamentals" (Nishisato 2016a). The current paper supplements it with further discussion of the problems in quantification theory. Current concerns are with the nature of multidimensional space used in quantification, in particular about the point that we must at least double the dimensionality of the space to accommodate quantified variates, which makes us wonder if we should still pursue joint graphical display or consider an alternative to graphical display. Simple correspondence analysis is known as one of the main realms of quantification theory, and it is dual scaling of the contingency table. The current paper, however, will take us to the point at which we may have to say "farewell" to it. Let us discuss these problems as outcries.

1.1 Outcry 1: Linear and Nonlinear Analysis

This is a well-known aspect of quantification theory, but it seems that the point needs to be reemphasized. Suppose that we collect data on preference of tea under different water temperatures. Each subject is given ten cups of tea, ranging from freezing cold to boiling hot, and is asked to rate the preference of ten cups of tea on the 10-point scale, ranging from the worst to the best. If we use 10-point Likert scales for the temperature and for the preference ratings, the data can be presented as a 10-by-10 contingency table of choice frequencies. Typical analysis of the table using Likert scores without transformation, however, would not capture such a nonlinear relation as might be expected, namely, the preference being at the lowest (least liked) when the tea temperature is boiling hot, followed by freezing cold, then lukewarm, then ordinary cold ice tea, and finally optimally hot tea at the highest (most preferred). There are at least two distinct approaches to this kind of nonlinear relation. The first approach is to predict the preference Likert scores as a nonlinear function of the temperature of tea, indicated by Likert scores. Should we use a quadratic term, a cubic term, interaction terms, or higher order terms? The choice of these is not easy, but we must seek the best possible nonlinear function, and this is, however, not what most investigators would normally do—they do not consider any nonlinear function. Furthermore, what can we do to deal with multidimensional aspects of the data in this nonlinear regression approach? This is not a simple problem. The second approach is via correlation of the Likert scores of the two variables. In this approach, it is well known that Pearsonian correlation captures only linear relations; thus this is not an appropriate way to analyze nonlinear relations. One should realize then that Likert scores are predetermined quantities, independently of the data structure, and without additional operations of nonlinear transformations, one cannot generally expect exhaustive analysis of information in data through Likert scores. In contrast to these two approaches, dual scaling (correspondence analysis, homogeneity analysis, optimal scaling) is a method to find optimal scores for both the temperature

and the preference ratings as regressions on the data. In other words, dual scaling is used to transform Likert scores typically nonlinearly so as to make the regression of rows (ten cups of tea) on preference ratings and the regression of columns (ten preference values) on tea simultaneously linear (Hirschfeld 1935). Thus, this is a data-dependent method of scaling row values and column values in the optimal way, the reason why it is also called optimal scaling (Bock 1960), and multidimensional aspects of the data can be handled without problems. In this context, dual scaling is a method of projecting row values to the column values and column values to row values in the symmetric way. The common projection operators are known as singular values, which are also Hirschfeld's simultaneous regression coefficients and Guttman's maximal correlation coefficients between row quantification and column quantification. In terms of multidimensional decomposition, Nishisato (2006) has shown that dual scaling maximizes the Cramér's coefficient (Cramér 1946) and that this coefficient is the sum of the squared nonlinear correlation coefficients of principal components. This indicates how dual scaling deals with multidimensional nonlinear relations in the data.

In summary, Likert scores are predetermined scores, independently of the data, and should be used only for the purpose of data collection. Once data are collected, Likert scores should be subjected to transformation, typically nonlinear, so as to best describe the information in data.

There is a caution on the use of order constraints in analysis. Because the response categories are ordered (e.g., never < sometimes < often < always), one may wish to derive scores for these categories under the order constraint. This may sound reasonable, but one should not even be tempted to impose such an order constraint if the study aims to explore the information in the data, that is, if the research is exploratory. The reason is clear. The order constraint permanently wipes out the possibility of ever finding nonlinear relations in the data (e.g., one's ability to lift a heavy object increases as one gets older to a certain point and then decreases beyond a certain age). Thus, a general advice is not to use the order constraint in exploratory research. Note that there are many studies on ordered categories in quantification theory, but that the above advice should be kept in mind.

1.2   Outcry 2: Nature of Multidimensional Space for Symmetric Analysis

Dual scaling is based on the mathematical decomposition of data, called dual relations:

$$\frac{\sum_i^m f_{ij} y_{ik}}{f_{.j}} = \rho_k x_{kj}; \quad \frac{\sum_j^n f_{ij} y_{ki}}{f_{i.}} = \rho_k y_{ik}$$

where f_{ij} is the frequency of cell (i, j) of a contingency table, y_{ik} and x_{kj} are weights for row i and column j, called standard coordinates, of component k, $\rho_k x_{kj}$ and $\rho_k y_{ik}$ are the corresponding principal coordinates, and ρ_k is the singular value of component k. This is nothing but Hirschfeld's simultaneous linear regressions with the singular value as the regression coefficient, and the singular value is also Guttman's maximal correlation between the rows and the columns and also Nishisato's projection operator for rows onto columns and vice versa. From the last point, we can conclude that the row axis and the column axis for each component are separated by the angle $\theta_k = \cos^{-1} \rho_k$ (Nishisato and Clavel 2003, 2008). This space discrepancy indicates that if we analyze a two-by-two contingency table, we obtain a single component, but the fact of the matter is that dual scaling of this contingency table requires a two-dimensional graph, one for row variables and the other for column variables with the two axes separated by the angle θ. This means that one component of dual scaling outcome requires two dimensions and two components four dimensions. From this point of view, the currently most popular graphical methods used in quantification studies are all problematic. The first two (symmetric and nonsymmetric graphs) are traditional quantification approaches to graphical display (see, e.g., Benzécri et al. 1973; Nishisato 1980, 1994, 2007; Greenacre 1984; Lebart et al. 1984; Gifi 1990; Le Roux and Rouanet 2004; Beh and Lombardo 2014), and the third one (biplot) is a more general and mathematical invention with a variety of graphical choices (see, e.g., Gabriel 1971; Gower and Hand 1996).

1. Symmetric display or French plot: The two sets of principal coordinates, $\rho_k x_{kj}$ and $\rho_k y_{ik}$, are plotted in the same space (i.e., without taking the space discrepancy θ_k into consideration). In other words, a two-dimensional configuration of data points is plotted in a unidimensional graph; similarly, a four-dimensional configuration is plotted in a two-dimensional graph. Thus, unless the singular value ρ is *very close* to 1, the symmetric display does not offer a usable graph (see, for example, the warning by Lebart et al. 1977). Generally speaking, symmetric display is an illogical and obviously wrong graph for the data, but for its simplicity, it has unfortunately become a routine method for graphing quantification results. This practice should immediately be discarded.

2. Nonsymmetric display: This method plots the principal coordinates of one variable and the standard coordinate of the other variable, for example, $\rho_k x_{kj}$ and $y_{ik.}$ This is the projection of x onto the standard space of y. But, the standard coordinates are not the coordinates of the data, but artificially adjusted for the common variance, independently of the data at hand. Thus, projecting data onto these coordinates is not a logical way to describe data, thus making the joint graph not usable. See the demonstration (Nishisato 1996) that the standard coordinates associated with a small singular value are much further from the origin than those associated with a large singular value because the standard coordinates reciprocally compensate the frequencies of data points. One can consider the problem of principal component analysis, in which we start with a linear combination of variables, then find the principal axis, which is defined as the axis on which projections of data have the largest variance. Those projections of data on the principal axis are called principal coordinates. Therefore, principal

coordinates are the coordinates of data in the most informative way. Standard coordinates, on the other hand, do not represent projections of data points unless the singular value is 1.

3. Biplot: Consider the singular-value decomposition of a two-way data, $Y\Delta^{\alpha}\Delta^{1-\alpha}X$, where Y and X are, respectively, matrices of left and right singular vectors of the data matrix, Δ is the diagonal matrix of singular values, and α is bounded by 0 and 1. In biplot, graphical display of both variates are considered for various values of α. Notice, however, that only when α is either 0 or 1, it offers a plot comparable to the above two traditional plots, that is, nonsymmetric display of (2). In introducing coordinate systems for a set of variables, one of the most popular methods is through principal component analysis, where principal coordinates are the projections of data on principal axes. In this regard, principal coordinates represent data structure. It is true that the principal coordinate system is only one way of representing data, and there are an infinite number of coordinates systems, which, however, should be orthogonal transformations of the principal coordinates so long as we want to represent the data structure. Those variates used in biplots are not related to principal coordinates in any imaginable ways, except for one set of variates, Y or X, when α is 0 or 1. From the view of the graphical display in Euclidean space, therefore, it is the current author's personal view that a question mark has to be placed on the use of biplots for exploring data structure.

Considering that each of these popular methods for joint graphical display leaves a serious concern from the viewpoint that we wish to represent data in Euclidean multidimensional space, there seems to be an urgent problem of either finding a better method of graphical display or to give up a graphical display completely and look for a non-graphical way of summarizing the outcome of quantification.

1.3   Outcry 3: From "Graphing Is Believing" to Cluster Analysis

"Graphing is believing" (Nishisato 1997) was an attempt to legitimize joint graphical display of quantification results in Euclidean space. Since then, the author realized that a complete description of data requires a large number of dimensions, more precisely at least twice the dimensions that the traditional joint graphical display deals with. To clarify why we must at least double the dimensionality of space, Nishisato and Clavel (2010) proposed a framework for comprehensive dual scaling with doubled dimensions, and noting this aspect of expanded (doubled) dimensionality for graphical display, the authors proposed the use of cluster analysis as an alternative to the traditional graphical displays.

To illustrate their procedure, let us use an example from Stebbins (1950): 500 seeds of six varieties of barley were planted at six agricultural stations in the United States, and at the harvest time, 500 seeds at each station were randomly chosen

Table 1 Varieties of barley seeds after a number of years at different locations (from Stebbins 1950)

Locations						
	Arlington	Ithaca	St. Paul	Moccasin	Moro	Davis
Barley	(Virginia)	(New York)	(Minnesota)	(Montana)	(Oregon)	(California)
CT[a]	446	57	83	87	6	362
Ha	4	34	305	19	4	34
WS	4	0	4	241	480	65
Ma	1	343	2	21	0	0
Ga	13	9	15	58	0	1
Me	4	0	0	4	0	27

[a]Barley: *CT* Coast & Trebi, *Ha* Hanchen, *WS* White Smyrna, *Ma* Manchuria, *Ga* Gatemi, *Me* Meloy

and sorted into six varieties of barley, and those seeds were again planted in the following year, and at the harvest time, 500 randomly chosen seeds were again classified into six varieties, and so on. This experiment was repeated over a number of years to see if certain varieties of barley will become dominant at particular locations. The numbers of years of these experiments are not uniform but different from station to station. The final counts reported in Stebbins (1950) are summarized in the 6×6 contingency table (Table 1).

A complete dual scaling analysis of this data set is reported in Nishisato (1994), which shows the percentage contributions of the five components are, in the descending order, 38, 33, 25, 3, and 1%, showing the dominance of three components. Following Nishisato and Clavel (2003), the 12×12 super-distance matrix, consisting of the within-set distances of stations (between-station distances), the between-set distances (those between stations and barley varieties), and the within-set distances of barley varieties (between barley varieties), was calculated as given in Table 2.

This 12×12 matrix contains the distance information of all the variables in Euclidean space. Clavel and Nishisato (2008) and Nishisato and Clavel (2008) thoroughly analyzed this table by the hierarchical clustering method and the k-means clustering (see the results in their papers). Nishisato (2012) argued, however, that the investigators would typically be interested in the relations between row variables (stations) and column variables (varieties of barley), not relations within stations or within barley varieties, and therefore proposed that we should analyze only the between-set distance matrix, that is, the "barley varieties"-by-"locations" distance matrix. Although the current example of the between-set distance matrix is 6×6, the number of rows and the number of columns are not always equal; hence the between-set distance matrix is typically rectangular, as opposed to square. In order to deal with a rectangular distance matrix for clustering, Nishisato (2012) proposed a very simple and intuitive method of clustering, called clustering with the p-percentile filter. This method is very simple and does not require a complicated algorithm: calculate the p-percentile distance (the criterion distance) out of the

Table 2 Within-set and between-set distances in five-dimensional space (from Clavel and Nishisato 2008)

CT	0.0	(Symmetric upper portion is omitted)										
Ha	2.1	0.0										
WS	1.9	2.5	0.0									
Ma	2.6	2.9	2.8	0.0								
Ga	1.7	2.1	1.7	2.6	0.0							
Me	1.3	2.6	2.2	3.0	2.2	0.0						
Ar[a]	0.6	2.3	2.1	2.7	1.8	1.5	0.0					
It	2.3	2.7	2.6	0.9	2.4	2.8	2.5	0.0				
SP	2.0	0.9	2.5	2.9	2.1	2.5	2.3	2.7	0.0			
Mo	1.5	2.2	0.9	2.5	1.2	1.9	1.7	2.3	2.2	0.0		
Mo	2.1	2.6	0.7	2.9	1.9	2.4	2.4	2.8	2.7	1.1	0.0	
Da	0.6	2.1	1.8	2.6	1.7	1.3	0.6	2.4	2.1	1.4	2.0	0.0

[a]Locations: *Ar* Arlington, *It* Ithaca, *SP* St. Paul, *Mo* Mocasin, *Mo* Moro, *Da* Davis

Table 3 The 6 × 6 filtered between-set distance matrix, using 22-percentile criterion point

	CT[a]	Ha	WS	Ma	Ga	Me
Ar[b]	0.6	–	–	–	–	–
It	–	–	–	0.9	–	–
SP	–	0.9	–	–	–	–
Moc	–	–	0.9	–	1.2	–
Mor	–	–	0.7	–	–	–
Da	0.6	–	–	–	–	1.3

[a]*CT* Coast & Trebi, *Ha* Hanchen, *WS* White Smyrna, *Ma* Manchuria, *Ga* Gatemi, *Me* Meloy
[b]*Ar* Arlington, *It* Ithaca, *SP* St. Paul, *Mo* Mocasin, *Mo* Moro, *Da* Davis

elements of the between-set distance matrix, discard all distances which are larger than the criterion distance (i.e., variables which are widely separated do not belong to the same cluster), and see what clusters one can see among the remaining distances. The underlying idea is that we are interested only in those variables which are close to one another, thus we might as well discard all irrelevant distances from clustering. This method is simple and depending the value of p one chooses, the cluster can be tight or loose, and two clusters may or may not overlap. See its application in Nishisato (2014).

Let us apply the clustering with the *p*-percentile filter to the 6 × 6 matrix of the between-set distances, that is, the distance matrix between the six barley varieties and the six locations (see the 6 × 6 part of the left-bottom of the distance matrix). At the current stage of development, the choice of *p* is arbitrary, and for this example, $p = 22$ percentile was chosen, that is, all distances greater than this were discarded from the original 6 × 6 distance matrix, as shown in Table 3.

From this choice of the cutting point, we can identify the following clusters (Coast & Trebi at Arlington and Davis), (Hanchen at St. Paul), (White Smyrna at Mocasin and Moro), (Manchuria at Ithaca), (Gatemi at Mocasin), and (Meloy at Davis). As we can see, some clusters are overlapping. We can also guess that the overlapping can be eliminated by reducing the percentile point, although this may result in discarding some variables from analysis.

This filtering method is still at its infancy, and there are many studies needed before it can compete with other existing clustering methods, for example, how to determine the optimal p value for a given data set and how to calculate the distance between clusters. Its advantages over other methods are, among others, easiness or simplicity and capability to deal with rectangular matrices unlike some other existing methods. As for the traditional analysis through graphical display, see Nishisato (1994), noting that we must sacrifice much information in graphical display.

1.4   Outcry 4: Limitation of Simple Correspondence Analysis?

Traditionally, the French correspondence analysis identifies simple correspondence analysis and multiple correspondence analysis as two distinct forms of quantification. These "simple" and "multiple" methods correspond to dual scaling of the contingency table and that of multiple-choice data, respectively.

As was described in Nishisato (1980, 2016b), however, the two types of analysis are closely related to each other. Let us reproduce the example from Nishisato (2016b)— in response to the two multiple-choice questions:

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

The data can be represented in three forms as shown in Table 4.

Table 4 Three forms for representing the information in the contingency table

$$
C = \begin{bmatrix} 3 & 2 & 1 \\ 1 & 2 & 4 \end{bmatrix}
\quad
F_a = \begin{bmatrix}
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 \\
0 & 1 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 1
\end{bmatrix}
\quad
F_b = \begin{bmatrix}
3 & 0 & 3 & 0 & 0 \\
2 & 0 & 0 & 2 & 0 \\
1 & 0 & 0 & 0 & 1 \\
0 & 1 & 1 & 0 & 0 \\
0 & 2 & 0 & 2 & 0 \\
0 & 4 & 0 & 0 & 4
\end{bmatrix}
$$

If we are given one of the three forms, the other two forms can be generated from it. In this regard, the three formats are "equivalent" in some sense. Since the latter two forms yield identical singular values, let us eliminate the second format F_a from our discussion. The remaining formats are data forms respectively for simple correspondence analysis and multiple-correspondence analysis. But, as Nishisato (1980) has shown, the two formats yield different numbers of components. An $m \times n$ table of C yields the "smaller number of m and n minus 1" components, while the corresponding F_b provides "$m + n - 2$" components. Nishisato (2016b) calls twice of the space of C as "dual space" and the space for F_b as "total space." As is clear, when $m = n$, dual space and total space have the same number of dimensions, but when $m \neq n$, the dimensionality of total space is greater than that of dual space. What is the nature of these extra components in total space when $m \neq n$?

The implication of this discrepancy is that simple correspondence analysis, which deals with the format of C, fails to capture the total information in F_b when $m \neq n$, which is the format for multiple correspondence analysis. Thus, if we are to analyze the data information exhaustively, the conclusion is that we should always use multiple correspondence analysis, that is, dual scaling of multiple-choice data F_b, rather than simple correspondence analysis or dual scaling of contingency tables C. Does this mean "Limitation of simple correspondence analysis"?

We can stretch our imagination to the quantification of multimode contingency tables. For example, consider a three-mode data, which can be described as a trilinear decomposition of frequency f_{ijk}. The contingency table format will again restrict the total number of dimensions to the smallest number of categories of the variables minus 1. In this case, one can always represent the data in the response-pattern format with frequencies (e.g., F_b), which will most typically yield more components than the corresponding analysis of the three-way contingency table. Can we then abandon simple correspondence analysis completely and always use multiple correspondence analysis? The author's view is "yes, we can."

2   Concluding Remarks

Dual scaling quantifies categorical data in such a way that variates for the rows and those for the columns are determined as simultaneous regressions of them on the data in hand. As such, dual scaling provides the optimal way to explain the data. As is clear from such phrases as simultaneous linear regressions (Hirschfeld 1935), reciprocal averaging (Horst 1935), and dual scaling (Nishisato 1980), the basic premise of dual scaling lies in symmetric analysis of rows and columns of a data matrix. It was clarified in the current paper as well as my Beijing paper that we need to expand the multidimensional space to accommodate both variates. This awareness of expanded space has led to the criticism of the current methods of joint graphical display, leading to the suggestion for an alternative method of graphical display, that is, cluster analysis. In the same context, we were brought

back to Nishisato (1980) on the analytical comparisons between the contingency table format and its response-pattern format of the same data. When the number of rows is not equal to the number of columns of the data matrix, the response-pattern format of the same data yields more components than the contingency table format. If we are to pursue exhaustive analysis in the data, therefore, it is recommended that we should analyze the data represented in the response-pattern format rather than the contingency format. Data-dependent quantification, analysis in expanded multidimensional space, and exhaustive analysis using the response-pattern representation of the data are three major messages of the current paper.

References

E.J. Beh, R. Lombardo, *Correspondence Analysis: Theory, Practice and New Strategies* (Wiley, New York, NY, 2014)

J.P. Benzécri et al., *L'analyse Des Données: II. L'analyse Des Correspondances* (Dunod, Paris, 1973)

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

J.G. Clavel, S. Nishisato, in *New Trends in Psychometrics*, ed. by K. Shigemasu, A. Okada, T. Imaizumi, T. Hoshino. Joint analysis of within-set and between-set distances (Universal Academy Press, Tokyo, 2008), pp. 41–50

H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, NJ, 1946)

K.R. Gabriel, The biplot graphical display of matrices with applications to principal component analysis. *Biometrics* **58**, 453–476 (1971)

A. Gifi, *Nonlinear Multivariate Analysis* (Wiley, New York, NY, 1990)

J.C. Gower, D.J. Hand, *Biplots* (Chapman and Hall, London, 1996)

M.J. Greenacre, Theory and Applications of Correspondence Analysis (Academic, London, 1984)

H.O. Hirschfeld, A connection between correlation and contingency. *Camb. Philos. Soc. Proc.* **31**, 520–524 (1935)

P. Horst, Measuring complex attitudes. *J. Soc. Psychol.* **6**, 369–374 (1935)

B. Le Roux, H. Rouanet, *Geometric Data Analysis: From Correspondence Analysis to Structured Data* (Kluwer, Dordrecht, 2004)

L. Lebart, A. Morineau, N. Tabard, *Tecnniques de la Description Statistique: Méthodes et L'ogiciels Pour L'analyse des Grands Tableaux* (Dunod, Paris, 1977)

L. Lebart, A. Morineau, K.M. Warwick, *Multivariate Descriptive Statistical Analysis* (Wiley, New York, NY, 1984)

S. Nishisato, *Analysis of Categorical Data: Dual Scaling and its Applications* (University of Toronto Press, Toronto, ON, 1980)

S. Nishisato, *Elements of Dual Scaling: An Introduction to Practical Data Analysis* (Lawrence Erlbaum, Hillsdale, NJ, 1994)

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

S. Nishisato, in *Visualization of Categorical Data*, ed. by J. Blasius, M. J. Greenacre. Graphing is believing: Interpretable graphs for dual scaling (Academic Press, London, 1997), pp. 185–196

S. Nishisato, in *Multiple Correspondence Analysis and Related Methods*, ed. by M. J. Greenacre, J. Blasius. Correlational structure of multiple-choice data as viewed from dual scaling (Chapman and Hall/CRC, Boca Raton, FL, 2006), pp. 161–177

S. Nishisato, *Multidimensional Nonlinear Descriptive Analysis* (Chapman and Hall/CRC, Boca Raton, FL, 2007)

S. Nishisato, in *Challenges at the Interface of Data analysis, Computer Science and Optimization*, ed. by W. Gaul, A. Geyer-Schultz, L. Schmidt-Thiéme, I. Luntz. Quantification theory: reminiscence and a step forward (Springer, Heidelberg, 2012), pp. 109–119

S. Nishisato, Structural representation of categorical data and cluster analysis through filters, in German-Japanese Interchange of Data Analysis results, ed. by W. Gaul, A. Geyer-Schultz, Y. Baby, A. Okada (Springer, Berlin, 2014), pp. 81–91

S. Nishisato, in *Quantitative Psychology Research*, ed. by L. A. van der Ark, D. M. Bolt, W. C. Wang, J. A. Douglas, M. Wiberg. Multidimensional joint graphical display of symmetric analysis: back to the fundamentals (Springer, Berlin, 2016a), pp. 291–298. (Paper presented at the Annual Meeting of the Psychometric Society, Beijing)

S. Nishisato, Quantification theory: dual space and total space. Paper presented at the annual meeting of the Japanese behaviormetric Society, Sapporo, August (in Japanese), 2016b

S. Nishisato, J.G. Clavel, A note on between-set distances in dual scaling and Correspondence analysis. *Behaviormetrika* **30**, 87–98 (2003)

S. Nishisato, J.G. Clavel, in *New Trends in Psychometrics*, ed. by K. Shigemasu, A. Okada, T. Imaizumi, T. Hoshino. Interpreting data in reduced space: a case of what is not what in multidimensional data analysis (Universal Academy Press, Tokyo, 2008), pp. 347–356

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

C.L. Stebbins, *Variation and Evolution* (Columbia University Press, New York, NY, 1950)