

A BRIEF SURVEY OF ASYMMETRIC MDS AND SOME OPEN PROBLEMS

Naohito Chino*

A brief review is made of a body of extant asymmetric MDS models and methods, given a one-mode, two-way asymmetric square relational data matrix whose elements are similarity or dissimilarity measures between objects, or a special two-mode, three-way asymmetric relational data matrix which is composed of one-mode, two-way asymmetric square relational data matrices, and several open problems are discussed.

1. Introduction

Asymmetric relationships among objects are frequently observed in daily life as well as in designed experiments. They are diverse in character. One-sided love and hate among members of any informal groups are typical examples. Table 1 shows a sociomatrix among 10 male students in a senior-high school (Chino, 1978). A group of hens and cocks shows a special asymmetric relationship called pecking order among the group (e.g., Rushen, 1982). In recognition experiments the proportion of times response R_j was made when stimulus S_i was presented is not necessarily equal to the proportion of response R_i when stimulus S_j was presented (e.g., Nakatani, 1972). Number of references from one journal to another also gives rise to an asymmetric data matrix (e.g., Coombs, 1964). Amount of trade each nation in the world has with other nations sometimes shows huge imbalance. Table 2 is a set of trade data among 10 countries (including two regions) (Chino, 1978), which was obtained from *Statistical Yearbook* (1974) by United Nations. Amount of migration from one region to another is a typical example in geography (Tobler, 1976–1977). The number of browsing times as well as the spread of computer viruses (e.g., Balthrop et al., 2004) from one site to another on the Internet also yields asymmetric relations. The *asymmetric coupling* matrix defined in a dynamical network consisting of identical nodes is a typical example in neural networks (e.g., Amari, 1971; Fukai & Shiino, 1990; Kanter, 1988; Kree & Zippelius, 1995; Lie, 2008; Parisi, 1986).

In general, these asymmetric relationships observed at a specific point in time or during a certain period of time are summarized in a square matrix whose elements are similarity or dissimilarity measures between objects. Such a matrix is characterized as a *one-mode, two-way square asymmetric matrix*. In contrast, if such a matrix is obtained at several points in time or during several periods of time, the data matrix is represented by a special two-mode, three-way matrix which is composed of a set of

Key Words and Phrases: asymmetric MDS, Bayesian inference, Chino-Shiraiwa theorem, circular hierarchy, Hilbert space, multiple-judgment sampling, tensor, tests for symmetry, maximum likelihood, Young-Householder theorem

* Dept. of Psychology, Aichi Gakuin University, Nissin-city, Aichi, Japan
Mail Address: chino@dpc.agu.ac.jp

Table 1: Sociometric data of 10 students in a senior-high school.

Rater\Ratee	1	2	3	4	5	6	7	8	9	10
1	4	3	4	3	5	5	6	4	7	7
2	2	3	4	6	7	6	5	5	4	5
3	4	4	3	5	5	4	4	3	4	5
4	4	7	6	3	7	7	4	6	4	5
5	1	7	6	7	4	7	6	6	6	5
6	4	5	4	6	5	7	4	4	4	4
7	4	5	4	4	3	3	6	6	4	6
8	2	4	4	4	5	4	5	2	4	4
9	6	5	5	5	5	6	5	5	4	6
10	4	4	4	3	4	4	4	4	4	4

(Adapted from Table 1 of Chino (1978).)

Table 2: Trade data between 10 countries including two regions. The abbreviated country names, EC, EA, US, JN, AN, AA, ME, OA, UR, and EP, respectively denote EEC, EFTA, USA, Japan, Australia, New Zealand, Africa, Middle East, Other Asia, USSR, and Centrally planned economics.

from\to	EC	EA	US	JN	AN	AA	ME	OA	UR	EP
EC		25.67	15.67	2.84	2.42	9.62	5.59	5.07	2.66	6.49
EA	17.99		2.43	0.70	0.35	1.33	0.76	0.82	0.78	1.51
US	16.38	2.23		8.18	1.68	1.51	2.21	6.85	1.19	0.60
JN	4.40	1.32	9.55		1.46	2.48	1.61	8.93	0.49	0.32
AN	1.70	0.12	1.14	3.08		0.08	0.16	1.01	0.24	0.13
AA	11.60	0.88	1.92	0.98	0.04		0.20	0.35	0.63	0.67
ME	10.57	1.21	1.27	5.06	0.33	0.57		2.22	0.35	0.23
OA	5.33	0.46	6.94	6.96	0.79	0.63	0.59		0.62	0.36
UR	2.88	1.04	0.19	0.84	0.01	0.66	0.67	0.45		10.03
EP	5.38	1.49	0.33	0.14	0.55	0.63	0.64	0.30	10.56	

(Adapted from Table 3.2 of Chino (1997).)

one-mode, two-way square asymmetric matrices. Sometimes such a matrix is observed per individual, and yields the same two-mode, three-way matrix.

Asymmetric MDS (hereafter abbreviated as AMDS) extends symmetric MDS (hereafter abbreviated simply as MDS) to handle such asymmetric relationships among objects. As for symmetric MDS, Cox and Cox (2001) give two definitions, i.e., a *narrow definition* and a *wider definition*. According to their narrow definition, MDS is a search for a low dimensional space, usually Euclidean, in which points in the space represent objects, one for each object, and such that the distances between the points in the space match as well as possible the original dissimilarities. In order to clarify the definition of MDS and also the difference between MDS and AMDS, we shall redefine their narrow definition of MDS, adding a bit stricter constraints, as follows:

The following five conditions are assumed in our narrow definition of MDS. That is,

1. Data matrix is a *one-mode, two-way square symmetric matrix* whose elements are similarity or dissimilarity measures or a *special two-mode, three-way matrix* which is composed of a set of one-mode, two-way square symmetric matrices.

2. Dissimilarities must be measured at the ordinal level or higher, or measured as the count data.
3. MDS searches for a low dimensional *real metric space*, *Minkowski's r-metric* (a special case is Euclidean) in which points representing objects are embedded.
4. Points in the space represent the objects, each representing one object.
5. The distances between the points in the space match as well as possible the original dissimilarities.

Although Cox and Cox refer to neither the data type nor the scale level for MDS, we have added them as conditions 1 and 2. Especially, the reader should pay special attention to the term “count data” in condition 2. Although counting is sometimes viewed as an ordinal scale or higher, we shall clearly distinguish between counting and ordinal scale or higher. In this paper, following Suppes and Zinnes (1963, p.9), we shall use counting as an example of an *absolute scale*, in which there is no arbitrary choice of unit or zero available. In contrast, in the case of, say, a ratio scale, the choice of a unit is an arbitrary decision made by an individual or group of individuals.

Furthermore, if we consider an inferential procedure, the most rational analysis of count data is to treat them, not as, say, measured at an ordinal scale level, but as measured at a ratio scale level and take their natural logarithm further. The reason for this is that (1) multinomial distribution as well as Poisson distribution as standard models for count data belongs to a usual *exponential family*, and can be written in *canonical form*, (2) canonical parameters of these two distributions are not their population parameters themselves but the natural logarithms of their function or the natural logarithm of the population parameter, and (3) statistics corresponding to these canonical parameters have sufficient information of data (e.g., Andersen, 1980).

According to their wider definition, MDS can subsume several techniques of multivariate data analysis. It covers any technique which produces a graphical representation of objects from multivariate data. It is apparent that such techniques do not necessarily satisfy all of the above conditions. Therefore, we shall adopt the narrow definition of MDS in this paper.

Two basic theorems underpinned MDS in the 1930's. These are the Eckart and Young (1936) theorem and the Young and Householder (1938) theorem. The former is concerned with the lower rank fit to the data matrix \mathbf{S} , and the latter with the necessary and sufficient condition that the coordinates of objects in a multidimensional space are real points in the Euclidean space. Drawing upon these theorems, Richardson (1938) proposed a method of MDS and Torgerson (1954, 1958) developed it further (e.g., Tucker & Messick, 1963). Their method is now called *classical MDS*. There exists a large body of literature which extends classical MDS. We can divide it first into two, i.e., *descriptive MDS* and *inferential MDS*. We may further divide the latter into three. These are the so-called *special probabilistic MDS*, *maximum likelihood MDS*, and *Bayesian MDS*.

The descriptive MDS is a class of MDS which does not accompany any statistical inferences on population parameters. We shall only refer to representative ones below.

The classical MDS was first extended to the case when the dissimilarities are measured at an ordinal scale level by Kruskal (1964a, b), which is called nonmetric MDS. Guttman and his colleagues (e.g., Guttman, 1968; Lingoes, 1973) followed Kruskal, and proposed another method of nonmetric MDS called the smallest space analysis (abbreviated as SSA). Carroll and Chang (1970) also extended classical MDS in such a way to handle individual differences in dissimilarity judgments, and therefore their method is called individual differences MDS. Takane, Young, and De Leeuw (1977) proposed another method of individual differences MDS. In contrast with Carroll and Chang' method, their method is applicable to dissimilarity data measured on a wider class of scale levels.

As the special probabilistic MDS, we include a class of MDS methods which assumes a normal distribution on coordinates of objects to be estimated, and as a result assumes a noncentral χ^2 distribution with some degrees of freedom and a specified noncentrality parameter on squared distances. Such an MDS technique goes back to Hefner (1958), and several papers exist in the literature (e.g., Ramsay, 1969; Suppes & Zinnes, 1963; Zinnes & Mackay, 1983).

The maximum likelihood MDS is another class of inferential MDS techniques, which assumes a normal distribution and/or log-normal distribution on observed dissimilarities (e.g., Ramsay, 1977, 1978, 1982; Takane, 1978a,b, 1981; Takane & Carroll, 1981). In this class of MDS methods, it is assumed that dissimilarities are obtained by the pair comparison method or by a certain rating method.

The Bayesian MDS is another inferential MDS, which is based on the Bayesian inference, and has recently been proposed by several researchers (e.g., Fong et al., 2010; Je et al., 2008; Lee (2008); Oh & Raftery, 2001, 2007; Okada & Mayekawa, 2011; Okada & Shigemasu, 2010; Park et al., 2008).

As is apparent from the brief review of MDS made above, symmetric MDS has a long history and has been almost fully developed in that both descriptive and inferential methods are now available. Furthermore, major books on MDS which thoroughly review MDS have been published (e.g., Borg & Groenen, 2005; Cox & Cox, 2001; Saito, 1980; Takane, 1980). In contrast, such books on AMDS are rare (e.g., Chino, 1997), although the books on MDS mentioned above partly introduce AMDS (e.g., Borg & Groenen, 2005; Cox & Cox, 2001). Recently, Saito and Yadohisa (2005) review AMDS fairly extensively. One possible reason for the relative paucity of literature on AMDS may be that it is still an active area of research. As a result, the notion of AMDS is still vague, and it seems that the precise definition of AMDS has not been established. Therefore, it seems appropriate and necessary to give some definition of AMDS, before we make a critical review of AMDS in this paper. In fact, depending on this definition, the history of AMDS will have to focus on different aspects. It might be possible to make a narrow definition of AMDS as well as a wider definition in a manner similar to Cox & Cox (2001) for MDS. In this paper, we shall further divide the narrow definition of AMDS into three, i.e., a *narrow definition*, a *narrower definition*, and finally the *narrowest definition*.

The following six conditions are assumed for the narrowest definition of AMDS:

1. Data matrix is a *one-mode, two-way square asymmetric matrix* whose elements are similarity or dissimilarity measures or a *special two-mode, three-way matrix* which is composed of one-mode, two-way square asymmetric matrices.
2. Dissimilarities are measured at the ordinal level or higher.
3. It involves a search for a low dimensional *real metric space, complex metric space, or real asymmetric metric space*.
4. Points in the space represent the objects, each representing one object. This condition is the same as in MDS.
5. Some additional parameters other than distance are assumed in the case of the real metric space.
6. The distances between the points in the space match as well as possible the original dissimilarities. This condition is the same as condition 5 in MDS.

Condition 2 excludes a body of asymmetric methods for count data (i.e., Chino, 1997). As the metric spaces in condition 3, we include the Minkowski r -metric space (the Euclidean space as a special case) for the *real* metric space, the Hilbert space for the *complex* metric space (i.e., Chino & Shiraiwa, 1993; Saito & Yadohisa, 2005), and the asymmetric Minkowski space for a *real asymmetric* metric space (i.e., Sato, 1988). This condition excludes a familiar method for the skew-symmetric data, i.e., the *Gower diagram* (or sometimes called the canonical analysis of skew-symmetry (abbreviated as *CASK* by Chino, 1997), because the *area property* in *CASK* has the *symplectic* but not Euclidean structure (Chino & Shiraiwa, 1993).

As will be discussed in detail later, *CASK* decomposes the skew-symmetric part of a squared asymmetric matrix into the weighted sum of a special quantity (see, Eq. (20) in Section 2). It is well known that this quantity is an oriented *area* of the parallelogram spanned by the two location vectors corresponding to two objects.

It is apparent from the above definition of AMDS that we have inherited the concept of *metric space* from the narrow definition of MDS discussed previously. As is well known, a metric space is a nonempty set M equipped with a positive real-valued function $d : M \times M \rightarrow R$, called the distance function, that satisfies the following axioms:

1. $d(X, Y) \geq 0$, and $d(X, Y) = 0 \Leftrightarrow X = Y$ (positive-definite),
2. $d(X, Y) = d(Y, X)$ (symmetric),
3. $d(X, Y) + d(Y, Z) \geq d(X, Z)$ (triangle inequality),

for all $X, Y, Z \in M$.

Minkowski's r -metric as well as Hilbert space satisfies all of these axioms, while a more general Minkowski space, which we call here the asymmetric Minkowski space, does not satisfy the second axiom. As defined elsewhere (e.g., Matsumoto, 1986; Sato, 1988), a Minkowski space M is a finite dimensional real vector space such that the length of a vector $\mathbf{x} \in M$ is given by the value $L(\mathbf{x})$ of a function L on M , where L is assumed to satisfy the following conditions,

1. $L(\mathbf{x}) \geq 0$, for any $\mathbf{x} \in M$, $L(\mathbf{x}) = 0$ if and only if $\mathbf{x} = \mathbf{0}$, (nonnegativity),

2. $L(\alpha \mathbf{x}) = \alpha L(\mathbf{x})$, for any $\alpha > 0$, $\mathbf{x} \in M$, (positively homogeneous of degree one),
3. $L(\mathbf{x} + \mathbf{y}) \leq L(\mathbf{x}) + L(\mathbf{y})$, for any $\mathbf{x}, \mathbf{y} \in M$, (convexity),
4. $L(\mathbf{x})$ is differentiable at any non-zero \mathbf{x} .

The function L is called *Minkowski metric function*. If $L(\mathbf{x}) = L(-\mathbf{x})$ for any \mathbf{x} , then the metric is considered to be symmetric.

The narrower definition of AMDS, on the other hand, adds the count data in condition 2. This gives a criterion for checking and classifying the seemingly different approaches to an asymmetric dissimilarity data matrix using the notion of *quasisymmetry* in the log-linear model, as discussed in a later section.

The narrow definition of AMDS adds the *indefinite metric space* in condition 3. The narrow definition also adds a *one-mode, three-way square matrix* in condition 1. We shall discuss open problems on these topics in the discussion section.

According to the wider definition of AMDS, it can subsume several techniques of multivariate data analysis, and covers any techniques which produce a graphical representation of objects from multivariate data, as in MDS. It is apparent that such techniques do not necessarily satisfy all of the above conditions for AMDS. Therefore, we shall adopt the narrow definition of AMDS in this paper.

The organization of this paper is as follows. In the next section we shall briefly review a body of AMDS methods in the narrowest sense. Then, we shall briefly review a body of AMDS methods in the narrower sense, which assumes count data, and embeds objects in some metric space, and discuss the implications of checking and classifying various features of these two bodies of AMDS in the third section. In the discussion section, we shall list up the open problems on AMDS in the near future.

2. A Brief Review of AMDS in the Narrowest Sense

In the narrowest sense, AMDS may go back to the work of Young (1975). In his ASYMSCAL model, the squared distance from object i to object j is defined by the equation

$$d_{ij}^2 = \sum_{a=1}^r w_{ia}(x_{ia} - x_{ja})^2, \quad w_{ia} \geq 0, \quad (1)$$

where w_{ia} is the weight of object i on dimension a , x_{ia} is the coordinate of object i on dimension a , and where there are n objects and r dimensions. The distance used in this model is clearly an extension of the familiar Euclidean distance. As a result we need not only the coordinates of objects but also the object weights (which he calls the stimulus weights) in order to explain the asymmetric dissimilarity data.

A number of models have been proposed since Young proposed his ASYMSCAL. As in MDS, we can divide them into two, i.e., *the descriptive AMDS* and *the inferential AMDS*. According to the narrowest definition, almost all the extant AMDS models remain descriptive. Representative methods in this category are Borg and Groenen (2005), Chino (1978, 1990), Chino and Shiraiwa (1993), Constantine and Gower (1978), Escoufier and Grorud (1980), Gower (1977), Harshman (1978), Harsh-

man et al. (1982), Kiers and Takane (1994), Krumhansl (1978), Loisel and Takane (2011), Okada and Imaizumi (1984, 1987, 1997), Rocci and Bove (2002), Saito (1991), Saito and Takeda (1990), Sato (1988), ten Berge (1997), Tobler (1976–77), Trendafilov (2002), Weeks and Bentler (1982), Yadohisa and Niki (1999), Young (1975), and Zielman and Heiser (1996).

Chino and Okada (1996) and Chino (1997) further divide the above methods (except for the models proposed after 1996) into three groups, i.e., *the augmented distance model*, *the non-distance model*, and *the extended distance model*. The augmented distance model is a family of AMDS in which some parameters are added to the metric distance between objects in order to handle asymmetry, and includes Borg and Groenen (2005), some of Gower's (1977) models, Krumhansl (1978), Okada and Imaizumi (1984, 1987, 1997), Saito (1991), Saito and Takeda (1990), Tobler (1976–1977), Weeks and Bentler (1982), Yadohisa and Niki (1999), Young (1975), and Zielman and Heiser (1993). Young's ASYMSCAL can be classified into this family of models.

Tobler (1976–1977) proposed a unique geographical AMDS model. His *wind model* assumes a kind of wind for the observed asymmetry, and estimates it using a *vector field model*. In order to estimate the vector field from the original similarity measures, the following model is assumed

$$t_{ij} = \frac{d_{ij}}{r + c_{ij}}, \quad c_{ij} = r \frac{t_{ji} - t_{ij}}{t_{ij} + t_{ji}}, \quad (2)$$

where t_{ij} is a travel effort (time, cost, etc.), and it is assumed that t_{ij} is aided by a flow c_{ij} (Tobler denotes it as \vec{c}_{ij}) in the direction of movement from place i to place j . Moreover, r is a rate of travel, independent of position and of direction, and is in the same units as c_{ij} . Once the vector field is estimated, it is decomposed into divergence and curl-free parts, and the scalar and vector potentials are calculated. Finally, the gradient vectors of the scalar potentials are drawn on the estimated configuration, which explain the travel flows. If the distances between two places are known, Tobler's method cannot be classified into AMDS in the narrowest sense, but if it is estimated from the symmetric part of t_{ij} using some appropriate MDS, his method can be viewed as an AMDS.

Yadohisa and Niki (1999) proposed a vector field model similar to the Toblers' model. They assume that the locations of objects have already been determined from the symmetric part of the data via some suitable MDS method. Given the configuration of objects, they estimate vectors at those locations as well as the estimated scalar potentials from the skew-symmetric part of the data.

Two of the models which Gower (1977) proposed, i.e., the *jet-stream model* and the *cyclone model*, are very similar to Tobler's wind model. The jet-stream model was conceived by imaging a plane flying at a constant velocity V between two towns P_i and P_j which are d_{ij} distance apart. If there is a jet-stream, velocity v , making an angle θ_{ij} with line $P_i P_j$ then the flight times t_{ij} and t_{ji} are

$$t_{ij} = \frac{d_{ij}}{V + v \cos \theta_{ij}}, \quad t_{ji} = \frac{d_{ij}}{V - v \cos \theta_{ij}}. \quad (3)$$

As a result, if the ratio v/V is sufficiently small to ignore v^2/V^2 , the symmetric part and the skew-symmetric part of t_{ij} can be written as

$$t_{ij}(s) \doteq \frac{d_{ij}}{V}, \quad t_{ji}(sk) \doteq v d_{ij} \cos\theta_{ij}/V^2, \quad (4)$$

respectively. In any case, this model is formally the same as that of Tobler, if we reparameterize $c_{ij} = v \cos\theta_{ij}$ and $c_{ji} = -v \cos\theta_{ij}$. Although Gower was interested in analyzing symmetric parts and skew-symmetric parts separately, the jet-stream model may be thought of as an AMDS model if we analyze them simultaneously. Borg and Groenen (2005) proposed a similar model to the jet-stream model in that both symmetric parts and skew-symmetric parts are analyzed simultaneously. It is called the *hill-climbing model*, and is written as

$$d_{ij}^* = d_{ij} + \frac{(\mathbf{x}_i - \mathbf{x}_j)^t \mathbf{z}}{d_{ij}}. \quad (5)$$

The cyclone model also proposed by Gower (1977) is similar to the jet-stream model. This model is written as

$$t_{ij} = \frac{d_{ij}}{V + \omega h_{ij}}, \quad t_{ji} = \frac{d_{ij}}{V - \omega h_{ij}}. \quad (6)$$

As Gower points out, this model is similar to the jet-stream model except that the jet-stream is replaced by a cyclonic wind rotating about its center C at a constant angular velocity. Moreover, it is the same as Tobler's wind model, if we reparameterize $c_{ij} = \omega h_{ij}$ and $c_{ji} = -\omega h_{ij}$. As with the jet-stream model, the cyclone model may also be thought of as an AMDS model if we analyze them simultaneously. Later, we shall discuss an AMDS method proposed by Sato (1988, 1989) which generalizes the jet stream model from a mathematically more sophisticated view point.

Krumhansl's *distance-density model* augments the distance d_{ij} as follows:

$$d_{ij}^* = d_{ij} + \alpha\delta(\mathbf{x}_i) + \beta\delta(\mathbf{x}_j), \quad (7)$$

where d_{ij}^* is an augmented distance, and $\delta(\mathbf{x}_i)$ and $\delta(\mathbf{x}_j)$ are measures of spatial density in the neighborhoods of objects i and j , respectively, while α and β are the corresponding weights applied to the densities.

By contrast, Weeks and Bentler (1982) proposed an augmented distance model (the W-B model) which simplifies the distance-density model. According to the *W-B model*,

$$d_{ij}^* = b d_{ij} + c_i - c_j + a, \quad (8)$$

where a is an additive constant, and $b = -1$ if the data is composed of similarity measures instead of dissimilarity measures. The Euclidean distance, d_{ij} , may be replaced by d_{ij}^2 in Eq. (8).

Okada and Imaizumi (1984) proposed a more general model than the distance-density model as well as the W-B model. This model is written as

$$d_{ij}^* = d_{ij} + \alpha c(i, j, t) + \beta c(j, i, t), \quad (9)$$

where α and β are constant weight parameters, and $c(i, j, t)$ and $c(j, i, t)$ are the terms which represent the skew-symmetric component and are assumed to be positive. They consider several sub-models, one of which is the *O-I model* (Okada & Imaizumi, 1987)

$$d_{ij}^* = d_{ij} - r_i + r_j. \quad (10)$$

Okada and Imaizumi (1997) extended it to the two-mode, three-way case.

Saito and Takeda (1990) proposed models similar in form to the O-I model. Model 2, which is the most general one, is written as,

$$d_{ij}^* = d_{ij} + a\theta_i + b\theta_j + r, \quad (11)$$

where d_{ij} is the Minkowski's r -metric, and r is an additive constant. This model is apparently a special case of Eq. (9) but it may be more appropriately regarded as an extension of Eqs. (7), (8), and (10). For example, although $\delta(\mathbf{x}_i)$ and $\delta(\mathbf{x}_j)$ in Eq. (7) as well as r_i and r_j in Eq. (10) are all positive by definition, there exist no such restrictions on a , b , θ_i , and θ_j in Eq. (11). It is evident that the second and third right-hand terms in (11) are extensions of those in (8), although the first right-hand term of (11) is a special case of that in (8). If θ_i 's are positive, they are interpreted as stimulus specific effects, analogous to spatial densities in the distance-density model.

Later, Saito (1991) proposed the following model,

$$d_{ij}^* = d_{ij} + \theta_i + \phi_j + \gamma, \quad (12)$$

which is partly a generalization of Eq. (8), because parameters in the second and third right-hand terms are distinct in Eq. (12) but they belong to the same set of parameters in Eq. (8). However, as for the first right-hand term, this model assumes a special case of that of the W-B model.

Most of the above models are considered as a special case of the *Holman model* (Holman, 1979) stated as

$$s_{ij} = F[m_{ij} + r_i + c_j], \quad (13)$$

where s_{ij} is a similarity between objects i and j which Holman calls the proximity data and F is some strictly increasing function, and m_{ij} is a *symmetric function*. If m_{ij} is parametrized further by the coordinates of objects in a metric space, this model can be said to be an AMDS model in the narrowest sense. In any case, the only exception for the Holman model is a generalized version of the O-I model described by Eq. (10). Nosofsky (1991) calls the Holman model the *additive similarity and bias model*.

Zielman and Heiser (1993) proposed an algorithm to fit the *slide-vector model*, which had been suggested by Kruskal in 1973. This model is written as:

$$d_{ij} = \left\{ \sum_{t=1}^r (x_{it} - x_{jt} + z_t)^2 \right\}^{1/2}, \quad (14)$$

where x_{it} and x_{jt} , respectively, are the coordinates of objects i and j on dimension t , and (z_1, z_2, \dots, z_r) constitutes the slide-vector \mathbf{z} . This model has a distinguishing feature that the diagonal elements of the model are non-zero, unlike a regular distance model. They showed how the coordinates and the slide-vector can be obtained by using an unfolding algorithm by Heiser (1987). This means that the slide-vector is a special version of the *unfolding model* originated by Coombs (1964). They also proposed a three-way generalization of the slide-vector model,

$$d_{ijk} = \left\{ \sum_{t=1}^r u_{kt}(x_{it} - x_{jt} + z_t)^2 \right\}^{1/2}. \quad (15)$$

However, we consider the unfolding model, especially the *multidimensional unfolding models* (e.g., Bennet & Hays, 1960; Hays & Bennett, 1961; Schönemann, 1970) as they are neither a family of the augmented distance models nor a family of AMDS's in the narrowest sense, because in general the multidimensional unfolding models do not necessarily satisfy the conditions 1, 4, and 6 in our definition of AMDS in the narrowest sense. Incidentally, Zielman and Heiser (1993) also proposed the *multiple slide-vector model* as well as the *row-weighted slide-vector model* as candidates for other possible generalizations.

The non-distance model is based on some quantity other than the metric distance, e.g., the inner product to the similarity measures (e.g., Chino, 1977, 1978, 1990; Constantine & Gower, 1978; DeSarbo et al., 1992; Escoufier & Grorud, 1980; Gower, 1977; Harshman, 1978; Harshman et al., 1982; Kiers & Takane, 1994; Loisel & Takane, 2011; Trendafilov, 2002).

Chino (1977, 1978) proposed an ASYMSCAL model different from Young (1975). In contrast with Young's augmented distance model, he proposed a special inner product model for AMDS, which is written as follows,

$$s_{ij} = a(x_{i1}x_{j1} + x_{i2}x_{j2}) + b(x_{i1}x_{j2} - x_{i2}x_{j1}) + c, \quad (16)$$

where s_{ij} is a similarity between objects i and j , and c is an additive constant, while x_{il} is the coordinate of object i on dimension l . It is apparent that the quantity in the first parentheses on the right-hand side of Eq. (16) and that in the second parentheses are, respectively, the inner product and the cross-product (outer product) of position vectors corresponding to two objects i and j on a two-dimensional plane. The cross-product is equivalent to the area of the parallelogram formed by the two position vectors.

Figure 1 shows the two-dimensional configuration of 10 nations including two regions for the trade data in Table 2. The positive direction in this figure is crucial in interpreting the direction of skewness in the amounts of trade between two nations. The sign of b in Eq. (16) determines this direction. It is apparent from the positive direction as well as the magnitudes of the parallelogram spanned by two position vectors corresponding to two nations that Japan is in a state called "unilateral love". In other words, Japan's trade surplus of exports over imports was prominent among the

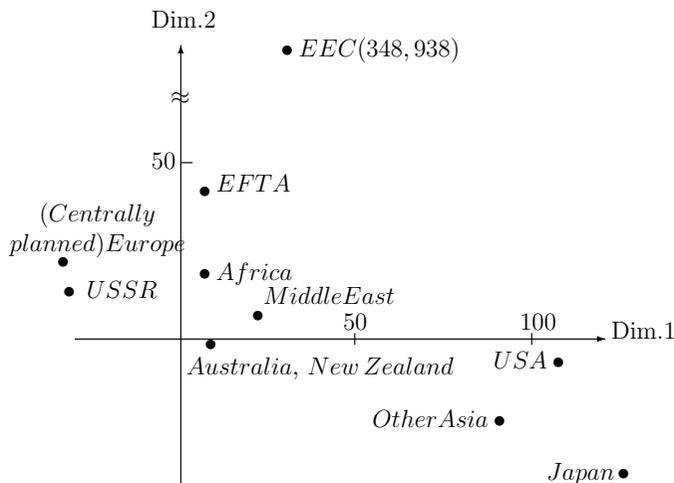


Figure 1: Configuration of 10 nations including two regions obtained via Chino's ASYMSCAL (Adapted from Figure 3.7 of Chino (1997)). In this figure, the positive direction is counterclockwise.

10 nations.

Although Chino's ASYMSCAL was confined within the three-dimensional space, he later extended it to a generalized inner product model called GIPSCAL (Chino, 1990). GIPSCAL is written in matrix form as,

$$\mathbf{S} = a \mathbf{X} \mathbf{X}^t + b \mathbf{X} \mathbf{L}_q \mathbf{X}^t + c \mathbf{1}_N \mathbf{1}_N^t, \quad (17)$$

where $\mathbf{S} = \{s_{ij}\}$, a and b are constants, and c is an additive constant, while \mathbf{X} is an $N \times q$ coordinate matrix, and \mathbf{L}_q is a special skew-symmetric matrix (also, Chino, 1980; Gower, 1984).

Later, Kiers and Takane (1994) pointed out that the alternating least squares (ALS) algorithm for fitting the GIPSCAL off-diagonal elements did not constitute a true ALS, and hence need not decrease the objective function value monotonically. Furthermore, they simplified the \mathbf{L}_q in Eq. (17) in order to facilitate the interpretation of the skew-symmetric part of the data as follows,

$$\mathbf{S} = a \tilde{\mathbf{X}} \tilde{\mathbf{X}}^t + b \tilde{\mathbf{X}} \mathbf{\Delta} \tilde{\mathbf{X}}^t + c \mathbf{1}_N \mathbf{1}_N^t, \quad (18)$$

where $\mathbf{\Delta}$ is a fixed matrix with singular values of the matrix \mathbf{L}_q in skew-symmetric 2×2 blocks along the diagonal. To be precise, $\mathbf{L}_q = \mathbf{U} \mathbf{\Delta} \mathbf{U}'$ and $\mathbf{X} \mathbf{U} = \tilde{\mathbf{X}}$. Rocci and Bove (2002) proposed a special case of (18).

Trendafilov (2002) proposed another unique method for solving GIPSCAL efficiently. In his approach, the GIPSCAL problem is reformulated into an *initial value problem for matrix ordinary differential equations* on manifolds defined by the constraints of the original least-squares problems. This algorithm has been found to produce solutions which give better fits to the data than the algorithms of Chino

(1978) as well as Kiers and Takane (1994). Trendafilov also proposed a three-way GIPSCAL and its algorithm.

Recently, Loisel and Takane (2011) proposed a fast convergent algorithm for GIPSCAL with acceleration by the minimal polynomial extrapolation. They adapted their basic algorithm to various extensions of GIPSCAL, including off-diagonal DEDECOM/GIPSCAL, and three-way GIPSCAL.

Gower (1977) proposed several methods for the analysis of asymmetry, one of which includes in part the same quantity as Chino's ASYMSCAL in the two-dimensional case. The Gower diagram or CASK, referred to in the introduction section, decomposes the skew-symmetric part of the square asymmetric data matrix \mathbf{S} via the singular value decomposition,

$$\mathbf{S}_{sk} = \mathbf{X} \mathbf{\Lambda} \mathbf{K} \mathbf{X}^t, \quad (19)$$

where \mathbf{S}_{sk} denotes the skew-symmetric part of \mathbf{S} , while \mathbf{X} is an $N \times N$ orthogonal matrix, $\mathbf{\Lambda}$ is a special diagonal matrix of singular values such that $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_1, \lambda_2, \lambda_2, \dots, (0))$, and \mathbf{K} is a skew identity matrix consisting of

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

as diagonal blocks. In scalar form, Eq. (19) is written as

$$s_{ij}(sk) = \sum_t^p \lambda_t (x_{i,2t-1} x_{j,2t} - x_{i,2t} x_{j,2t-1}), \quad (20)$$

where p is the largest integer not exceeding $N/2$. Eq.(20) is nothing but the second term on the right-hand side without constant b in Chino's ASYMSCAL given in (16). Chino (1977) and Gower (1977) introduced this quantity independently.

However, Gower took a philosophically different approach from Chino's. Holman points out that there exist two approaches to asymmetric proximity data. One considers the symmetric part and skew-symmetric part of the data to be inseparable parts of the same fundamental process, and the other considers the two parts to reflect different processes that can be distinguished by appropriate analysis. Chino (1977) took the former approach, while Gower (1977) took the latter, although at least some other methods proposed by Gower, i.e., the *jet-stream model* and the *cyclone model* include the distance between two geographical points, as discussed earlier. In any case, we shall not regard at least CASK as an AMDS model because it possesses a symplectic structure which is different from that of the Euclidean metric structure (e.g., Arnold, 1978; Chino & Shiraiwa, 1993).

Harshman and his colleagues (Harshman, 1978; Harshman et al., 1982) proposed a simple non-distance model called DEDICOM, which stands for the DEcomposition into DIrectional COMponents,

$$\mathbf{S} = \mathbf{Y} \mathbf{A} \mathbf{Y}^t, \quad (21)$$

where \mathbf{Y} is the $N \times p$ loading matrix of N objects on a few basic types of objects, while \mathbf{A} is a small asymmetric matrix of order p giving the directional relationships among the basic p types or dimensions. He called DEDICOM a non-spatial approach, and thus did not discuss its metric structure. However, Chino and Shiraiwa (1993) have proven that DEDICOM has an explicit metric structure under a mild condition. Formally, Chino's ASYMSCAL is a special case of DEDICOM. Harshman et al. (1982) discuss a two-mode, three-way version of DEDICOM. Takane and Kiers (1997) proposed the *latent class DEDICOM* for square contingency tables, which is a hybrid model of a latent class model and a special constrained DEDICOM.

Escoufier and Grorud (1980) proposed a similar model in form to Chino's ASYMSCAL from a linear algebraic point of view. Although they did not give any formal name to their model, Chino (1991) called it the *Hermitian canonical model* abbreviated as *HCM*. HCM first decomposes the original similarity matrix \mathbf{S} into the symmetric part \mathbf{S}_s and the skew-symmetric part \mathbf{S}_{sk} , and compute an Hermitian matrix \mathbf{H} such that $\mathbf{H} = \mathbf{S}_s + i\mathbf{S}_{sk}$, where i is a pure imaginary number. Then, they solved the eigenvalue problem of \mathbf{H} using a well-known method which constructs a double sized, real symmetric matrix \mathbf{T} such that

$$\mathbf{T} = \begin{pmatrix} \mathbf{S}_s & -\mathbf{S}_{sk} \\ \mathbf{S}_{sk} & \mathbf{S}_s \end{pmatrix}. \tag{22}$$

It is well known that if the eigenvector corresponding to the l -th eigenvalue of \mathbf{H} is denoted as $\mathbf{w}_l = \mathbf{u}_l + i\mathbf{v}_l$, then the eigenvalues of \mathbf{T} come in pairs $\lambda_1, \lambda_1, \lambda_2, \lambda_2, \dots$, and the eigenvectors corresponding to an identical eigenvalue λ_l are $\mathbf{a}_l = (\mathbf{u}_l^t, \mathbf{v}_l^t)^t$ and $\mathbf{b}_l = (-\mathbf{v}_l^t, \mathbf{u}_l^t)^t$ (e.g., Wilkinson, 1965). Using these results, we have

$$\mathbf{S}_s = \sum_{l=1}^N \lambda_l (\mathbf{u}_l \mathbf{u}_l^t + \mathbf{v}_l \mathbf{v}_l^t), \tag{23}$$

and

$$\mathbf{S}_{sk} = \sum_{l=1}^N \lambda_l (\mathbf{v}_l \mathbf{u}_l^t - \mathbf{u}_l \mathbf{v}_l^t). \tag{24}$$

In scalar form, these are approximated as follows with only the largest eigenvalue λ_1 ,

$$s_{ij}(s) \doteq \lambda_1 (u_{i1} u_{j1} + v_{i1} v_{j1}), \quad s_{ij}(sk) \doteq \lambda_1 (v_{i1} u_{j1} - u_{i1} v_{j1}). \tag{25}$$

This is the result obtained by Escoufier and Grorud.

DeSarbo et al. (1992) proposed a spatial MDS procedure called *TSCALE* based on *Tversky's contrast model*. They assume a two-mode, three-way proximity data. Letting d_{ijr}^* be the dissimilarity value on the r -th replication between the two objects i and j , for example, one of the *TSCALE* models is written as

$$d_{ijr}^* = \sum_{t=1}^T \alpha_r (x_{it} - x_{jt})_+ + \sum_{t=1}^T \beta_r (x_{jt} - x_{it})_+ + \sum_{t=1}^T \theta_r \min(x_{it}, x_{jt}), \tag{26}$$

for Tversky's original linear contrast analog. Here, the function $(u - v)_+ = \max(u - v, 0)$, α_r is the impact or salience that is distinctive to the first object, i , in pair ij , presented on the r th replication (e.g., subject r), β_r is the impact or salience that is distinctive to the second object, j , in pair ij , presented on the r th replication (e.g., subject r), and θ_r is the impact or salience that is common to the object pair ij presented on the r th replication. They also proposed a ratio, distinctive feature model. For details, see DeSarbo et al. (1992).

The extended distance model is a family of AMDS in which some distance structure is assumed other than the traditional *real metric structure*, i.e., the Minkowski's r -metric. One is the the (real) asymmetric Minkowski's metric structure, and the other the (complex) Hilbert space structure. Sato (1988, 1989) proposed to embed objects in a certain real asymmetric Minkowski space, given a set of asymmetric dissimilarity measures. Here, it should be noticed that the Minkowski space he uses is not the familiar Minkowski's r -metric, but has an asymmetric metric function. The two-dimensional (asymmetric) Minkowski metric model is written as follows,

$$d_{ij}^* = \frac{d_{ij}}{r_\beta(\psi | \gamma)}, \quad (27)$$

where d_{ij} is a usual Euclidean distance, and $r_\beta(\psi | \gamma)$ is a transformed indicatrix such that, using an original indicatrix $r(\theta | \gamma)$,

$$x_1 - \beta = r(\theta | \gamma) \cos(\theta), \quad x_2 - \beta = r(\theta | \gamma) \sin(\theta), \quad (28)$$

where (x_1, x_2) is the current coordinate of point in the plane, and the original indicatrix is written as

$$r(\theta | \gamma) = \sum_{k=1}^m \gamma_k r_k(\theta). \quad (29)$$

Here, the component indicatrix in Eq. (29) is

$$r_k(\theta) = \mu_k + \nu_k \cos(k(\theta - \pi/4)), \quad (30)$$

where $\mu_k = (k^2 + 2)/(k^2 + 3)$ and $\nu_k = 1/(k^2 + 3)$. As Sato (1988) point out, if we choose the indicatrix with $k = 1$, the Minkowski metric defined by Eq. (27) is equivalent to the jet stream model by Gower (1977). Sato (1989) extended his model to the three-dimensional case in which another asymmetric metric function, *Rander's metric function* (Randers, 1941), was used.

The other is to utilize the (complex) Hilbert space structure. As discussed earlier, Escoufier and Grorud (1980) did not point out the metric structure in their HCM model. In contrast, Chino and Shiraiwa (1993) reformulated HCM and discovered the Hilbert space structure. In contrast with Escoufier and Grorud's approach, they directly treated the eigenvalue problem of the Hermitian matrix obtained from the data matrix \mathbf{S} , and deduced the following equation:

$$\mathbf{H} = \mathbf{X}\mathbf{\Omega}_s\mathbf{X}^t + i\mathbf{X}\mathbf{\Omega}_{sk}\mathbf{X}^t, \quad (31)$$

where

$$\Omega_s = \begin{pmatrix} \Lambda, & \mathbf{O} \\ \mathbf{O}, & \Lambda \end{pmatrix}, \quad \Omega_{sk} = \begin{pmatrix} \mathbf{O}, & -\Lambda \\ \Lambda, & \mathbf{O} \end{pmatrix}, \quad (32)$$

and Λ is a diagonal matrix of order n which is the number of non-zero eigenvalues of \mathbf{H} , i.e., $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$. The matrix \mathbf{X} is the special *real* $N \times 2n$ coordinate matrix of objects, i.e., $\mathbf{X} = (\mathbf{U}_r, \mathbf{U}_c)$, where $\mathbf{U}_1 = \mathbf{U}_r + i\mathbf{U}_c$ and \mathbf{U}_1 are composed of the *complex* eigenvectors of \mathbf{H} corresponding to its non-zero eigenvalues. It should be noticed that all the eigenvalues of \mathbf{H} is *real*.

It is easy to show that Eq. (31) can be rewritten as

$$\mathbf{S} = \mathbf{X}\Omega_s\mathbf{X}^t + \mathbf{X}\Omega_{sk}\mathbf{X}^t. \quad (33)$$

Chino and Shiraiwa (1993) called Eq. (31) the *Hermitian form model* abbreviated as *HFM*. They showed that DEDICOM, GIPSCAL, and HCM are special cases of the Hermitian form model if their complex counterparts are considered. Furthermore, Chino and Shiraiwa (1993) proved that a necessary and sufficient condition for these models to be expressible in terms of (complex) Hilbert space is the *positive semi-definiteness* of \mathbf{H} . This is an extension of the Young-Householder theorem on MDS to the case of the complex space.

Saito and Yadohisa (2005) contended that there existed some unclear points of this extension, and took another look at the extension. That is, they divided the Chino-Shiraiwa theorem into two, i.e., theorems on sufficiency and necessity, which they called theorem 4.2 and theorem 4.3, respectively. Moreover, they emphasized the role of the distance between objects j and k , d_{jk} ($= \|\mathbf{w}_j - \mathbf{w}_k\|$, where \mathbf{w}_j and \mathbf{w}_k are coordinate vectors corresponding to objects j and k in the Hilbert space) in discussing the extension of the Young-Householder theorem. However, in theorem 4.2 they did not discuss the other important distance \bar{d}_{jk} ($= \|\mathbf{w}_j - i\mathbf{w}_k\|$, where $i^2 = -1$) in the Hilbert space, which is indispensable to the Chino-Shiraiwa theorem. For further details, see Chino and Shiraiwa (1993) as well as Saito and Yadohisa (2005).

All the AMDS models discussed up to now are *descriptive models* in which no statistical inference is involved. Traditionally, the AMDS in the narrowest sense might have been applied to asymmetric relational data matrices under an implicit assumption that the data are sufficiently asymmetric enough to warrant AMDS. However, the question arises: Given an \mathbf{S} , how should we ascertain whether it is sufficiently symmetric or not? Kruskal (1964b), for example, discusses nonsymmetry of dissimilarities in his seminal paper on nonmetric MDS. He recommends to average the (i, j) and (j, i) elements of \mathbf{S} if they are measurements on the same underlying quantity and differ only because of random fluctuation. However, he did not provide any method for checking whether they differ only randomly. In order to answer this question empirically, it seems appropriate to develop some inferential method for it.

Recently, Saburi and Chino (2008) proposed a maximum likelihood asymmetric MDS called ASYMMAXSCAL, which is an extension of MAXSCAL proposed by Takane (1981). MAXSCAL is a maximum likelihood (ML) MDS method specifically

designed to analyze the symmetric similarity data measured on rating scales with a relatively small number of observation categories. ASYMMAXSCAL inherited almost all the characteristics which MAXSCAL possesses. As a result, the scale level assumed in ASYMMAXSCAL is usually higher than or equal to the ordinal level. However, as Takane (1981) points out, as a special case, the same-different judgments like the confusion data can also be analyzed as if they were two-category judgments. As a result, whatever the scale level may be, the original data set for applying ASYMMAXSCAL is composed of one-dimensional frequency distributions of proximity judgments on the pairs of objects, which is called the *Type A design data*. By examining the estimated values of the category boundaries of the rating scale under study, one can diagnose the scale level of measurement as either the ordinal level or the interval level.

ASYMMAXSCAL provides two methods for checking whether the data are sufficiently asymmetric or not. One is a special symmetry test prior to fitting any asymmetric MDS model, and the other is some symmetry tests in a subsequent scaling step. To do the first, we rearrange the Type A data into a special $n \times n \times M$ three-way contingency table whose stratified variable is composed of the rating categories. It is called the *Type B design data*.

In ASYMMAXSCAL, a special conditional symmetry hypothesis, i.e., $H_0^{(cs)}$: $p_{ijm} = p_{jim}$ ($1 \leq i < j \leq N$; $1 \leq m \leq M - 1$) in Type A design is tested using the LR test statistic. The other method for checking the asymmetry of data in ASYMMAXSCAL is composed of two kinds of symmetry tests in the scaling step. One is the test for the symmetry hypothesis based on the *saturated representation model* (the SR model). Although any extant AMDS model may be chosen as the representation model in ASYMMAXSCAL, the SR model is the representation model with no structure. The other symmetry test in the scaling step is the test for the symmetry hypothesis based on a specified representation model under study. For example, Saburi and Chino (2008) choose the O-I model as the specified model. As a result, the hypothesis is represented by $H_0^{(s/oi)}$: $r_i = r_{i+1}$ ($1 \leq i \leq N - 1$). In ASYMMAXSCAL, several models including the above mentioned symmetry hypotheses can be compared using AIC, instead of using the statistical tests. Such a strategy has often been taken in the literature of asymmetric MDS as well as analysis of contingency tables (Takane, 1981, 1987; Tomizawa, 1992).

Figure 2 shows the three-dimensional configuration and spheres for the O-I model under the ordinal scale assumption via the application of ASYMMAXSCAL to the friendship data among nations in East Asia and the USA, which was adapted from Saburi and Chino (2008). In this survey, four hundred Japanese university students participated, and the single-judgment sampling was chosen. Each subject rated the extent to which he or she felt the government of a nation as friendly or hostile to that of another nation on a 5-point rating scale. Since the sphere associated with Japan is larger than those of the other nations, it is interpreted from the O-I model that Japan is perceived as more friendly to other nations than the other way around.

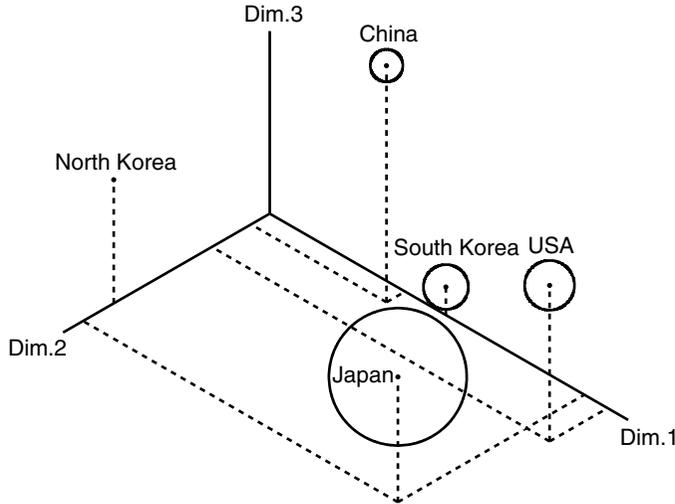


Figure 2: Configuration of 5 nations and spheres surrounding each of them for the O-I model obtained via ASYMMAXSCAL (Adapted from Figure 4 of Saburi & Chino (2008)).

3. A Brief Review of AMDS in the Narrower Sense and a Criterion for Checking Features of AMDS Models

In this section, we shall first review briefly major AMDS methods for count data. As was done in the review of AMDS in the narrowest sense, we shall divide a body of AMDS in the narrower sense into the descriptive AMDS and the inferential AMDS. As was the case for MDS, the latter AMDS grew out of the descriptive methods.

In stimulus recognition experiments in psychology, typical data can be summarized in a *confusion matrix* \mathbf{C} , each element of which is the proportion of times response R_j was made when stimulus S_i was presented, and it is an estimate of the conditional confusion probability $P(R_j | S_i)$. According to Ashby and Perrin (1988), the *biased-choice model* (sometimes called the *similarity choice model*, e.g., Nosofsky, 1986) has been most successful in predicting a wide variety of confusion matrices over the last 20 years (e.g., Luce, 1963; Shepard, 1957). This model is written as

$$P(R_j | S_i) = \frac{\beta_j \eta_{ij}}{\sum_m \beta_m \eta_{im}}, \tag{34}$$

where η_{ij} is the similarity of stimulus S_i to stimulus S_j , and β_j is the *bias* toward response R_j . Here, it is assumed that $\eta_{ij} = \eta_{ji}$ and all self-similarities are equal. Eq. (34) is sometimes called the *unrestricted similarity-choice model* in contrast with the restricted version discussed below (Takane & Shibayama, 1992).

The biased-choice model itself cannot be viewed as an AMDS model in the narrower sense. However, if we assume that η_{ij} is a function of the distance between the two stimuli, and the distance is parameterized by coordinates of stimuli, such a model can be considered as an AMDS model. In fact, Shepard (1957) suggested replacing

η_{ij} in Eq. (34) by the exponential decay function, $\exp(-d_{ij})$, i.e.,

$$P(R_j | S_i) = \frac{\beta_j \exp(-d_{ij})}{\sum_m \beta_m \exp(-d_{im})}. \quad (35)$$

This model is called the *MDS-choice model* (e.g., Nosofsky, 1985a, b, 1986) or Shepard (1958a) also considered the Gaussian function, i.e., $\eta_{ij} = \exp(-d_{ij}^2)$, and sometimes called, respectively, the exponential MDS-choice model and the Gaussian MDS-choice model (e.g., Takane & Shibayama).

Although Ashby and Perrin (1988) refer to several models as versions of the MDS-choice model, some of them cannot be viewed as AMDS models in the narrower sense, while some other can be. Moreover, some of them are descriptive, but others are inferential. We shall briefly discuss the descriptive AMDS first. Shepard (1957, 1958b) discussed not only Eq. (35) but also proposed a procedure for obtaining coordinates of stimuli using a version of classical MDS.

Getty et al. (1979) conducted an experiment to demonstrate that the perceptual space derived from similarity judgments in a pairwise similarity-judgment task can then be used to predict behavior in an identification task. Judged (symmetric) similarity data among 8 visual stimuli were analyzed by INDSCAL (Carroll & Chang, 1970) first, and the set of spatial coordinates for each of the stimuli were obtained. Then, the weighted Euclidean distances were computed and a set of confusion weights between stimuli were obtained using an exponential decay function. Finally, the conditional probabilities of giving the response R_j assigned to stimulus S_j when stimulus S_i was presented using a version of Luce's choice model, which assumed no differential response bias. Results indicated that this model predicts well the confusion matrices of the identification task.

Nosofsky (1984) related Medin and Schaffer's (1978) *context theory* to a more general theoretical framework for the modeling of choice and similarity. Their context theory assumes two rules. One is the *response-ratio rule* which states the probability of classifying any test stimuli as a member of some category, and the other the *multiplicative rule* for computing stimulus similarity. As for the response rule, Nosofsky considers it as a bias-free extension of Luce's choice theory. As regards the multiplicative rule, he points out that it arises as a special case of psychological distance between stimuli conforming to the city-block metric, and of stimulus similarity being an exponential decay function of psychological distance.

Keren and Baggen (1981) proposed and tested a feature analytic model for recognition of alphanumeric characters based on Tversky's contrast model. They applied their model, to Gilmore et al.'s (1979) data. However, since distances between stimuli are not resolved into coordinates of stimuli, which is consistent with the philosophy of their model, their model cannot be viewed as asymmetric MDS in the narrower sense.

Appelman and Mayzner (1982) applied Krumhansl's distance-density model to the confusion matrices in three published studies which conducted a typical letter recognition experiment using a dot matrix for displaying letters. Although they examine, for example, the functional relationship between mean proportion of confusions per

letter pair and the distance between letters, they do not utilize any MDS procedure to estimate coordinates of letters. For this reason, their study cannot be considered as that of asymmetric MDS in the narrower sense.

In contrast, Nakatani (1972) proposed a slightly different model from the MDS-choice models discussed above. He proposed the *confusion-choice model* which attempts to model the recognition process with a fairly complicated perceptual process incorporating concepts from signal detection theory. His model is written as

$$P(R_j | S_i) = \sum_{k=1}^m e_{ik} \frac{b_j(r_{jk})}{\sum_{l=1}^n b_l(r_{lk})}, \tag{36}$$

where b_j is the bias probability, $m = 2^n$, r_{jk} is the value of r_j for confusion state s_k , and b_j is a function of r_{jk} such that

$$b_j(r_{jk}) = \begin{cases} b_j, & \text{when } r_{jk} = 1, \\ 0, & \text{when } r_{jk} = 0. \end{cases}$$

Moreover, r_j and s_k are defined as follows:

First, a binary-valued mediating variable, r_j , which carries the information as to whether R_j is or is not acceptable, is defined as,

$$r_j = \begin{cases} 1, & \text{for } R_j \text{ is acceptable} \\ 0, & \text{for } R_j \text{ is not acceptable.} \end{cases}$$

Second, a confusion state is defined as the n -tuple,

$$s_k = \langle r_{1k}, \dots, r_{jk}, \dots, r_{nk} \rangle .$$

Moreover, e_{ik} is called the equivocation probability

$$e_{ik} = P(s_k = \langle r_{1k} \dots r_{jk} \dots r_{nk} \rangle | S_i) = \prod_{j=1}^n a_{ij}(r_{jk}), \tag{37}$$

where

$$a_{ij}(r_{jk}) = \begin{cases} a_{ij}, & \text{when } r_{jk} = 1, \\ 1 - a_{ij}, & \text{when } r_{jk} = 0. \end{cases}$$

Here, a_{ij} is called the *acceptance probability*,

$$a_{ij} = P(R_j \text{ acceptable} | S_i) = P(q_{ij} \leq t'_j) = \int_{-\infty}^{t'_j - u_{ij}} \varphi(z) dz, \tag{38}$$

where $\varphi(\bullet)$ is the normal density function, $q_{ij} = u_{ij} + \epsilon_j$, u_{ij} is the L-dimensional Euclidean distance between S_i and R_j , i.e., $u_{ij} = \left\{ \sum_{l=1}^L (x_{il} - x_{jl})^2 \right\}^{1/2}$, and ϵ_j is the channel noise, $N(0, 1)$.

To sum up, it is evident that Nakatani's confusion-choice model is an AMDS model

in the narrower sense, because the distance between each stimulus S_i and its response R_i is represented by a common point in a metric space. Shepard's MDS-choice model and Getty et al.'s weighted Euclidean distance model are also considered as the AMDS models in the narrower sense among several MDS-choice models.

In contrast with the descriptive AMDS models in the narrower sense, some researchers proposed inferential AMDS models in the 1980's, as will be discussed below: Nosofsky (1986) proposed a unified quantitative approach to modeling subject's identification and categorization of multidimensional perceptual stimuli. He examined a traditional similarity choice model with the Minkowski's r -metric for the former and a new categorization model which generalized the context theory of classification developed by Medin and Schaffer (1978) for the latter.

It is interesting to notice here that, on the one hand, the count data obtained by an identification experiment are summarized in an $N \times N$ confusion matrix \mathbf{C} , where c_{ij} in each cell is the frequency with which stimulus i was identified as stimulus j . On the other hand, in a categorization experiment the N stimuli are classified into $m < N$ groups, each group assigned a distinct response.

Moreover, in this case the count data are summarized in an $N \times m$ confusion matrix, where c_{ij} is the frequency with which stimulus i was classified in category j . Although the new categorization model is interesting in that it takes the INDSCAL approach (Carroll & Chang, 1970) in which selective attention is modeled by differential weighting of the component dimensions in the psychological space (Nosofsky, 1986, p.41), we will not consider it as an AMDS model in the narrower sense because the data matrix obtained by the categorization experiment does not satisfy the conditions, 1 and 4, of AMDS. The reason is that in Nosofsky's unified approach, he assumes the two kinds of confusion matrices.

As the function η_{ij} in Eq. (34) of the similarity choice model, he considers both the exponential decay function and the Gaussian function. He uses an ML criterion to estimate parameters of both models. The MDS-choice model yielded its best fits to the identification data by assuming a Gaussian function and a Euclidean metric. However, his method seems to be insufficient as an inferential method, because he neither utilize information criteria nor perform statistical tests in order to compare several models in terms of their goodness of fit (GIF).

In contrast, Takane and Shibayama (1986, 1992) compared various models intensively through the AIC statistic (Akaike, 1974) by developing a standard ML estimation procedure. The likelihood of the total set of observations is stated in the same way for all the models, i.e., $L = \prod_{i,j} p_{ij}^{f_{ij}}$, where p_{ij} is the conditional probability specified by each of the models, assuming the multinomial design.

The data used in Takane and Shibayama (1986) are Keren and Baggen's (1981) recognition experiment data. Models compared in their analysis are the null model (the saturated model in the context of the log-linear model) (e.g., Bishop et al., 1975; Birch, 1963), the unrestricted similarity-choice model (i.e., the similarity choice model), the Euclidean distance-choice model (i.e., a version of the MDS choice model), and the unique feature-choice model (a general version of Keren-Baggen's (1981)

model). Here, the Keren-Baggen's model is a feature analytic model for recognition of alphanumeric characters based on Tversky's (1977) contrast model (Keren & Baggen, 1981). In any case, it is apparent that the Euclidean distance-choice model is the only inferential AMDS model in the narrower sense among the several models discussed in Takane and Shibayama (1986). Here, it should be noticed that only the Euclidean distance model among them parameterizes the distance assumed in this model by coordinates of stimuli. Takane and Shibayama (1992) conducted a more extensive examination of these models and the related models using four data sets including Keren and Baggen's (1981) data.

Nosofsky (1991) proposed a *bias-supplemented MDS model*

$$p_{ij} = F[ur_i + vc_j - wd_{ij}], \quad (39)$$

where u , v , and w are positive constants. He assume that the distance in the multi-dimensional space is computed using a city-block metric.

De Rooij and Heiser (2003, 2005) proposed models called *distance-association models* for the analysis of square contingency tables, each of which might be said to be a hybrid of the log-linear model and a distance model which is closely related to MDS. In the *one-mode distance association model*,

$$\ln[E(f_{ij})] = \lambda + \lambda_i^R + \lambda_j^C - \sum_m (x_{im} - x_{jm})^2, \quad (40)$$

where f_{ij} is the observed count in the (i, j) cell of the a square contingency table, and $E(f_{ij})$ is the expected cell value corresponding to f_{ij} . De Rooij and Heiser (2003) denote $E(f_{ij})$ by F_{ij} . They point out that it is a reduced rank version of the quasi-symmetry model. Here, at a glance, Eq. (40) is similar to Shepard's (Shepard, 1958a) Gaussian model (sometimes called *MDS-choice model* (e.g., Nosofsky, 1985a), or the *Euclidean distance-choice model* (Takane & Shibayama, 1986), but the second right-hand term is slightly different in that λ_i^R in Eq. (40) includes no distances.

In contrast, the two-mode distance-association model (De Rooij & Heiser, 2005) is no longer classified as the AMDS in the narrower sense, because it assumes two sets of coordinate matrix. As an inferential method, they apply the Pearson χ^2 -test as well as the likelihood ratio (abbreviated hereafter as LR) χ^2 -test (e.g., Bishop et al., 1975) in the former paper. In the latter, they utilize the %AAF (Goodman, 1971) instead of the G^2 , considering the fact that the traditional chi-squared distributed statistics tend to dismiss all models except the saturated model as the sample size increases indefinitely.

The AMDS models in the narrower sense also include those in the narrowest sense which were reviewed in the preceding section. Although these models are diverse in character, there seems to exist at least one criterion for checking various features of AMDS models. This is the property of *quasi-symmetry*, which was proposed by Caussinus (1965) in the context of the square contingency table. As for the models for the count data, especially confusion data matrices as a typical example of such data discussed in this section, it has been pointed out (e.g., Smith, 1982; Takane &

Shibayama, 1986; Townsend & Landon, 1982) that the similarity choice model has this property. Of course, the distance-association models proposed by De Rooij and Heiser (2003, 2005) also have this property, because they associate some symmetric metric function with the interaction term of the log-linear model.

As for the models assuming the ordinal level or higher in the preceding section, we can apply this criterion in checking these models if the data are obtained by a rating scale judgments as in ASYMMAXSCAL. For, in such a case, no matter what the scale level may be, we can rearrange the original Type A design data into the Type B design data, i.e., a special $N \times N \times M$ contingency table, where M is the number of rating categories in the rating scale, and check whether the quasi-symmetry hypothesis holds or not. Moreover, many of the AMDS models in the narrowest sense have this property.

For example, Chino and Saburi (2009) consider several extant AMDS models as the *quasi-symmetry-like AMDS models*. These are Saito's (1991) model, the distance-association model, the distance-density model, the Saito and Takeda's (1990) model, the slide-vector model, the W-B model, and the O-I model. The reason for labeling 'quasi-symmetry-like' models is that many of these models are not necessarily designed specifically to analyze the count data. A crucial problem may be whether the quasi-symmetry property is a universal property which holds for any asymmetric phenomena. We shall discuss this problem in more detail in the discussion section.

4. Discussion

In this section, we shall briefly introduce problems that are presently unsolved in AMDS to facilitate its further developments of AMDS in the near future.

Problem 1 (Tests of Symmetry, Quasi-Symmetry, and Marginal Homogeneity)

As pointed out in the last paragraph of the previous section, it may be necessary and appropriate to check at least whether the data to be analyzed by AMDS has the quasi-symmetry property possibly *before* some AMDS model is applied to the data. On the one hand, for the count data discussed in the previous section, we can directly test the quasi-symmetry hypothesis by Caussinus (1965), for example, as De Rooij and Heiser (2003) did. On the other hand, we can apply the special conditional quasi-symmetry hypothesis (e.g., Bishop et al., 1975) to the Type B data discussed in the preceding section, if the data are obtained on a rating scale (Saburi & Chino, 2008). The reason for checking this property before any AMDS is applied is that it is logically invalid to apply any AMDS with the quasi-symmetry property to the data which have no such property.

In connection with the above method for checking the quasi-symmetry hypothesis, Saburi and Chino (2008) proposed two possibilities, i.e., a symmetry test prior to fitting the AMDS model, and some symmetry tests in the scaling step. Alternatively, Chino and Saburi (2006) have suggested that the conditional quasi-symmetry hypothesis plays a fundamental role in AMDS, and have proposed some sequential tests for the symmetry and related hypotheses.

Recently, Chino and Saburi have attempted to prove the following conjecture on the independence of the three hypotheses, i.e., the quasi-symmetry hypothesis for testing $H_0^{QS} : \boldsymbol{\theta} \in \omega_{QS}$ against $H_1^{QS} : \boldsymbol{\theta} \in \Omega - \omega_{QS}$ (Caussinus, 1965), a special symmetry hypothesis for testing $H_0^S : \boldsymbol{\theta} \in \omega_S$ against $H_1^{S*} : \boldsymbol{\theta} \in \omega_{QS} - \omega_S$, and a marginal homogeneity hypothesis for testing $H_0^{MH_0} : \boldsymbol{\theta} \in \omega_{MH_0}$ against $H_1^{MH_0^*} : \boldsymbol{\theta} \in \omega_S - \omega_{MH_0}$. Here, the parameter spaces, Ω , ω_{QS} , ω_S , and ω_{MH_0} are defined below in Eqs. (41) through (44). It should be noticed that there exists the relation of inclusion, i.e., $\Omega \supset \omega_{QS} \supset \omega_S \supset \omega_{MH_0}$, in these spaces:

Conjecture 1 *The three LR statistics, G_{QS}^2 to test H_0^{QS} against H_1^{QS} , $G_{S^*}^2$ to test H_0^S against H_1^{S*} , and $G_{MH_0^*}^2$ to test $H_0^{MH_0}$ against $H_1^{MH_0^*}$, are mutually independent stochastically, if the three hypotheses are tested seriatim and if the H_0^{QS} is accepted first and the H_0^S is accepted second.*

Here, parameter spaces pertaining to the above three hypotheses, using the log-linear model (Birch, 1963), are as follows:

First, the total parameter space for $\boldsymbol{\theta} = (\theta_{ij}^{(12)}, \theta_i^{(1)}, \theta_j^{(2)}, \theta^{(0)})$ corresponding to an $N \times N$ cross classification table is

$$\Omega = \left\{ (\theta_{ij}^{(12)}, \theta_i^{(1)}, \theta_j^{(2)}, \theta^{(0)}), \right. \\ \left. -\infty < \theta_{ij}^{(12)}, \theta_i^{(1)}, \theta_j^{(2)}, \theta^{(0)} < \infty \quad i, j = 1, 2, \dots, N \right\}. \quad (41)$$

Second, the parameter space related to the quasi-symmetry hypothesis is

$$\omega_{QS} = \left\{ (\theta_{ij}^{(12)}, \theta_i^{(1)}, \theta_j^{(2)}, \theta^{(0)}), \quad -\infty < \theta_{ij}^{(12)} = \theta_{ji}^{(12)} < \infty, \right. \\ \left. -\infty < \theta_i^{(1)}, \theta_j^{(2)}, \theta^{(0)} < \infty, \quad i, j = 1, 2, \dots, N \right\}. \quad (42)$$

Third, the parameter space pertaining to the symmetry hypothesis is

$$\omega_S = \left\{ (\theta_{ij}^{(12)}, \theta_i^{(1)}, \theta_j^{(2)}, \theta^{(0)}), \quad -\infty < \theta_{ij}^{(12)} = \theta_{ji}^{(12)} < \infty, \right. \\ \left. -\infty < \theta_i^{(1)} = \theta_i^{(2)} < \infty, \quad -\infty < \theta^{(0)} < \infty, \quad i, j = 1, 2, \dots, N \right\}. \quad (43)$$

Fourth, the parameter space pertaining to the marginal homogeneity hypothesis in terms of the log-linear model (Andersen, 1980, pp.208–209) is

$$\omega_{MH_0} = \left\{ (\theta_{ij}^{(12)}, \theta_i^{(1)}, \theta_j^{(2)}, \theta^{(0)}), \quad \theta_{ij}^{(12)} = 0, \quad -\infty < \theta_i^{(1)} = \theta_i^{(2)} < \infty, \right. \\ \left. -\infty < \theta^{(0)} < \infty, \quad i, j = 1, 2, \dots, N \right\}. \quad (44)$$

However, this conjecture has not been proven yet (Chino & Saburi, 2009; 2010). The most difficult problem is whether the LR-statistic for testing the quasi-symmetry hypothesis, G_{QS}^2 , is an *ancillary statistic* (e.g., Lehmann, 1983) regarding its nuisance parameters. If G_{QS}^2 is an *ancillary statistic*, we can prove the above conjecture using the theorems by Basu (1955), Lehmann (1983), and Hogg and Craig (1956). Then,

we can perform a sequential test for testing these three hypothesis, and can control the type 1 error completely.

Problem 2 (Multiple-Judgment Sampling)

Computation of the likelihood of the data is essential in constructing any inferential statistical method. For example, ASYMMAXSCAL assumes that each subject judges only one pair of objects, which is called the *single-judgment condition* or the *single-judgment sampling*. Computation of the likelihood as a product of likelihoods of observations is justified in such a case.

By contrast, if each subject judges all pairs of objects, it is called the *multiple-judgment condition* or the *multiple-judgment sampling* (e.g., Bock & Jones, 1968). In this case, statistical dependencies among these multiple judgments are inevitable. In practice, however, it is sometimes convenient to use multiple-judgment sampling. For example, it is usual in the sociometric test to ask members of a classroom to rate all the members with a rating scale. Table 1 in the introductory section shows one such set of data.

Although Bock and Jones (1968), for example, discuss the cases in which this requirement is relaxed somewhat in the treatment of paired comparisons, the simplest way to cope with the violation of independence of the data in the multiple-judgment condition may be to assume some appropriate statistical distribution for such data. Promising candidates for it may be the *multivariate (Bernoulli) multinomial distribution* (Wishart, 1949) and some version of *multivariate Poisson distribution* (e.g., Johnson et al., 1997, Krumpalauer, 1998).

Problem 3 (Bayesian Inferences)

As mentioned in the preceding section, various Bayesian inferential MDS methods have already been developed since Oh and Raftery (2001). However, as far as we know, there had been no such method for AMDS until recently. Okada (2011) has proposed such a method. Bayesian AMDS methods might be expected to overcome some shortfalls of ML methods (e.g., Oh & Raftery, 2001; Press, 1989). A shortfall of ML MDS methods, which was pointed out by Cox (1982), is that the asymptotic theory on which ML relies may not apply in high dimensions because the number of parameters to be optimized typically grows. A well-known shortfall of ML estimation in a general setting is the so-called Neyman-Scott problem (i.e., Neyman & Scott, 1948), in which the small-sample bias of some ML estimator persists as sample size increases in the case of a non-identical parent distribution.

However, the Bayesian inference cannot seem to be omnipotent. For example, as a reason for the schism within the field of statistics over the Bayesian position, Price states, “Some statisticians feel that the evidence against the frequentist approach to statistical inference and decision making is not yet sufficiently cogent to warrant discarding a large collection of procedures that by and large have worked pretty well (Price, 1989, p.48)”. Bernardo states as follows in answering a question about the axioms of probability by Irony and Singpurwalla (Irony & Singpurwalla, 1996, p.161): “These are proven *existence results*; they imply that the common sense ‘a prior does not exist’ is a mathematical fallacy: for mathematical consistency one must be a

Bayesian. However, these are only existence results; they leave open the question of specifying a particular prior in each problem”.

Problem 4 (Multiple Comparisons in Quasi-Symmetry)

Although the importance of the test for quasi-symmetry in the context of AMDS has already been pointed out by some researchers as discussed in the preceding section, the *multiple comparisons for the hypotheses of quasi-symmetry* have rarely been done in the context of AMDS, although several statistical theories on this problem have already been developed (e.g., Hirotsu, 1983; Kastenbaum, 1960). For example, the row-wise (or column-wise) multiple comparison procedure proposed by Hirotsu (1983) enables us to examine in which pairs of rows (or columns) the quasi-symmetry is detected. As is well known, the quasi-symmetry hypothesis is usually defined as $H_0^{QS} : \theta_{ij}^{(12)} = \theta_{ji}^{(12)}$ for all i and j , where $\theta_{ij}^{(12)}$ is the interaction term of the log-linear model. In contrast, Hirotsu utilizes an alternative definition of quasi-symmetry, i.e., $H_0^{QS} : p_{ij}p_{jk}p_{ki} = p_{ji}p_{ik}p_{kj}$ for the triplet i, j, k being all different as well as any pair of the triplet being equal, where p_{ij} is the population probability corresponding to the observed frequency f_{ij} . If H_0^{QS} is rejected, one may consider, say, the hypothesis on its row-wise partition, $h_0^{K_{ij}} : \mathbf{D}_{ij}^t \mathbf{q} = \mathbf{0}$, where \mathbf{D}_{ij}^t is an $(n-2) \times N^2$ matrix with \mathbf{d}_{ijh}^t , $h = 1, \dots, N; h \neq i, j$ as its $N-2$ rows, \mathbf{d}_{ijh}^t being an N^2 dimensional vector defined for a triplet such that the $N(i-1)+j$ th, the $N(j-1)+h$ th, the $N(h-1)+i$ th elements are 1, the $N(i-1)+h$ th, the $N(j-1)+i$ th, and the $N(h-1)+j$ th elements are -1, and all the other elements 0 if $1 \leq i < j < h \leq N$ and similar for other permutation. Moreover, \mathbf{q} is an N^2 -dimensional vector with $\ln p_{ij}$ as the $N(i-1)+j$ th element. Such comparisons might uncover interesting features contained in the contingency tables peculiar to AMDS.

Problem 5 (Positive Semidefinite Programming)

As Kumagai (2010) points out, recently there has been increasing attention to the problems and applications of the *semidefinite programming* (e.g., Alfakih et al., 1999; Laurent, 2001; Weinberger & Saul, 2004; So & Ye, 2007). Kumagai (2010) proposed to solve the classical MDS problem by Torgerson with asymmetric relational data in a different fashion, utilizing the semidefinite programming. His method cannot be classified into AMDS in the narrowest sense, because it does not satisfy condition 4. However, his idea might be extended further to solve various AMDS problems from the view point of semidefinite programming, because (positive) semidefiniteness plays an essential role not only in MDS but also in AMDS, as the Chino and Shiraiwa’s theorem points out. Solving HFM assuming a positive semidefinite Hermitian matrix may be an easy problem to be solved in the near future.

Problem 6 (Longitudinal Asymmetric Relational Data)

In the traditional psychometric approach to a special two-mode, three-way matrix which is composed of one-mode, two-way square asymmetric matrices, the second mode usually represents individual. However, we frequently encounter a three-way matrix whose second mode denotes time. Longitudinal sociometric matrices and longitudinal trade data are typical examples. An ambitious approach to such three-way

data may be to construct some *vector field model* based on *dynamical system theory* in mathematics.

On the one hand, Tobler (1976–1977) and Yadohisa and Niki (1999) estimate such a vector field, given a single asymmetric data matrix among objects. On the other hand, a DYNAmical System SCALing method (abbreviated as *DYNASCAL*) proposed by Chino and Nakagawa (1990), estimates the vector field at each point in time (which is not restricted to the *gradient field of the scalar potentials*), given a set of longitudinal asymmetric relational data matrices, such as longitudinal attraction data which were gathered by Newcomb (1961). In this method, the configuration of members of a dormitory at each time is estimated from each of the longitudinal asymmetric data matrices via *MULTISCALE* proposed by Ramsay (1982).

DYNASCAL estimates further some important features of the vector field at each time, utilizing the *qualitative theories of singularities* as well as those of *bifurcations* in interpreting the estimated vector fields. Figure 3 shows one of the estimated vector fields, some estimated trajectories peculiar to the vector field, and some estimated attractors (in particular, singularities and a limit cycle) of the field for the Newcomb data via DYNASCAL. These features of the vector field provide us not only with the configuration of members in a group at a specific point in time but also with changes in the local as well as the global group structure of members over time.

A limitation of these methods seems to be their assumption about the *state space*. For, all of these assume that it is Euclidean. By contrast, the Chino and Shiraiwa theorem suggests that the state space is a *finite-dimensional (complex) Hilbert space*. In fact, Chino (2005) proposed a *complex difference equation model* of interpersonal interaction. But, it is a very restrictive model, as Chino pointed out. For example, it assumes that members will not move in his or her psychological space if there exists no skewness of sentiment between any two members. A more promising model should be constructed based on a deeper insight into the principle of interpersonal interactions and on ample empirical facts regarding interpersonal attraction.

Problem 7 (Indefinite Metric Space)

Indefinite metric space models for asymmetric MDS can generally be defined as models whose Hermitian matrix constructed directly or estimated from a one-mode, two-way square asymmetric matrix is *indefinite* (e.g., Chino & Shiraiwa, 1993). Spaces with such a metric is not so familiar among social and behavioral scientists, but they are well known in physics. For example, Minkowski space in a *special relativity theory* in physics is known to possess a typical example of indefinite metric structure (e.g., Arfken & Weber, 1995). In this space, the space-time interval, ds^2 , is defined as

$$ds^2 = dx_0^2 - dx_1^2 - dx_2^2 - dx_3^2,$$

where $x_0 = ct$ and c is the velocity of light.

Berlin and Kac (1952) proposed a mathematical model of a ferromagnet, in which it is assumed that there is a spin at each site of a regular lattice of N sites. Immediately before discussing some models for the interaction energy between neighboring spins, they examined the eigenvalue-eigenvector problem of the real *circulant matrix*

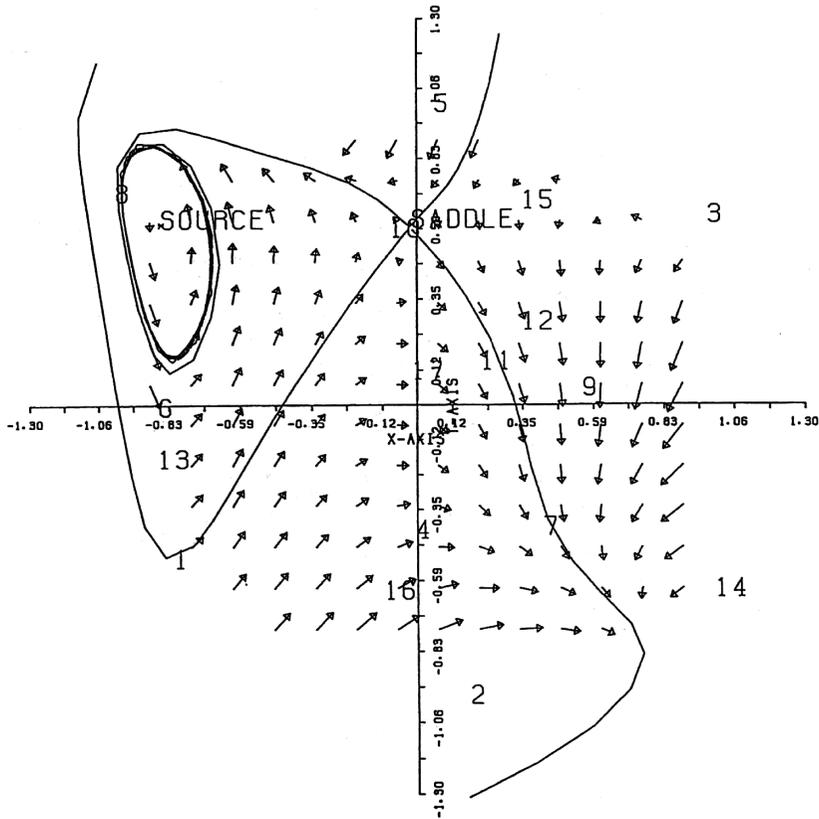


Figure 3: Estimated vector field at Week 0 of Newcomb's longitudinal data (Reproduced from Figure 5 of Chino and Nakagawa (1990)). In this figure, numbers indicate the locations of members in the dormitory.

(or cyclic matrix),

$$M_c = \begin{pmatrix} c_1 & c_2 & c_3 & \cdots & c_{N-1} & c_N \\ c_N & c_1 & c_2 & \cdots & c_{N-2} & c_{N-1} \\ c_{N-1} & c_N & c_1 & \cdots & c_{N-3} & c_{N-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ c_2 & c_3 & c_4 & \cdots & c_N & c_1 \end{pmatrix}. \tag{45}$$

This matrix is asymmetric in general, but can be symmetric by imposing necessary conditions on the coefficients. They wrote down all the eigenvalues of M_c and the corresponding eigenvectors in the appendix.

Chino (2001) examined the eigenvalue-eigenvector problem of the Hermitian matrix constructed from a special 4×4 circulant matrix whose elements are composed of $c_1 = 1$, $c_2 = 2$, $c_3 = 3$, and $c_4 = 4$ using HFM, and showed that it has the indefinite metric structure. Kosugi (2004) considered the possible 16 patterns of the triadic asymmetric relationships among members of a group (Harary, 1968) from the viewpoint of the above theories by examining the eigenvalue-eigenvector structures of

It will be necessary and desirable to gather empirical evidence further on circular hierarchy as well as indefinite metric structure, to examine the genetic background for such structures, and to establish substantive theories on these structures. For example, experiments conducted by Chadwick-Furman and Rinkevich (1994) with a complex allorecognition system in a reef-building coral, in which circular hierarchies of overgrowth interactions were exhibited in some of the colonies of a coral reef, are interesting in that they aimed at formulating a genetic basis for their allorecognition system.

Problem 8 (One-Mode, Three-Way Asymmetric Relations)

Recently, several MDS as well as AMDS methods which assume some triadic distances have been developed, given a one-mode, three-way square data matrix. Methods for analyzing this type of triadic distance data go back to Hayashi (1972). Representative methods in this category are Cox et al. (1991), Daws (1996), De Rooij and Gower (2003), Gower and De Rooij (2003), Hayashi (1972, 1989), Heiser and Benani (1997), Joly and Le Calvé (1995), Nakayama (2005), and Nakayama and Okada (2011). These are *descriptive MDS* methods which assume certain *symmetric, triadic distance models*. In contrast, De Rooij (2002) and De Rooij and Heiser (2000) proposed an *inferential AMDS* method which assumes some *asymmetric, triadic distance models*.

These MDS as well as AMDS methods assume some functions of *dyadic distances* except for Hayashi's (Hayashi, 1972) area model. Stated another way, neither of these models except Hayashi's deals with the triadic models which include *triadic interaction* (e.g., De Rooij and Gower, 2003). However, there exist some triadic relational data in which it is inappropriate to reduce holistic triadic relationships into diadic relationships. Typical examples can be found in some *balance theories* (e.g., Heider, 1946; Newcomb, 1953) and also in *the cognitive dissonance theory* (Festinger, 1957) in social psychology.

As already discussed in Problem 7, Kosugi (2004) examined the possible 16 patterns of triadic asymmetric relationships from the viewpoint of the above theories in social psychology by checking the eigenvalue-eigenvector structures of the Hermitian matrices. That is, he considers these structures from the balance theoretic viewpoint.

A possible method for handling the triadic interaction discussed above is to utilize a special skew-symmetric tensor called the *exterior form* which is usually denoted as *k-form* (or an exterior form of degree k). Here, a k -form is a function of k vectors which is *k-linear and antisymmetric* (e.g., Arnold, 1978). Let us suppose that we embed objects in an oriented p -dimensional Euclidean space R^p . Then, the three-dimensional oriented volume of the projection of the parallelepiped with edges $\xi_1, \xi_2, \xi_3 \in R^p$ onto R^3 is a *3-form* on R^p . We may fit the square of this volume as a measure of the triadic interaction to a one-mode, three-way square *symmetric* relational data. On the other hand, for a one-mode, three-way square *asymmetric* relational data, one way to handle the skew-symmetric components of the data may be to use the 3-form itself. In this case, it might be better to make some origin shift in defining the square of this volume to, say, the centroid of the coordinates of objects.

Of course, the idea of utilizing exterior forms is not new in the literature of MDS or, AMDS. For example, Chino's ASYMSCAL, GIPSCAL, and the Gower diagram fit area quantities to the skew-symmetric part of any one-mode, two-way square asymmetric matrix, as has already been discussed in section 2. These area quantities are all 2-forms. In contrast, Hayashi's area model for triadic data utilizes a special 3-form, which is associated with the area of simplex composed of the three vertices corresponding to three objects for handling triadic data. The idea of utilizing exterior forms can naturally be extended to *MDS* as well as *AMDS for n -tuples* (e.g., Cox et al., 1991).

However, caution should be exercised when utilizing tensors in general for fitting any statistical model to data. As has already been mentioned elsewhere (e.g., Silva & Lim, 2008; Stegeman & Comon, 2010), the following best rank- r approximation problem, $Approx(\mathbf{A}, r)$, has no solution in general for $r = 2, \dots, \min(d_1, \dots, d_k)$ and $k \geq 3$, where an order- k tensor \mathbf{A} is an element of a *tensor product* of k real vector space $V_1 \otimes V_2 \otimes \dots \otimes V_k$, as defined in any standard algebra textbook:

$Approx(\mathbf{A}, r)$: Given an order- k tensor $\mathbf{A} \in R^{d_1 \times \dots \times d_k}$, determine vectors $\mathbf{x}_i \in R^{d_1}, \mathbf{y}_i \in R^{d_2}, \dots, \mathbf{z}_i \in R^{d_k}, i = 1, \dots, r$, which minimize

$$\|\mathbf{A} - \mathbf{x}_1 \otimes \mathbf{y}_1 \otimes \dots \otimes \mathbf{z}_1 - \dots - \mathbf{x}_r \otimes \mathbf{y}_r \otimes \dots \otimes \mathbf{z}_r\|, \quad (47)$$

where $\|\bullet\|$ denotes some choice of norm on $R^{d_1 \times \dots \times d_k}$.

Stated another way, the best low-rank approximation problem for tensor is ill-posed for all orders (higher than 2), all norms, and many ranks, and the set of tensors that fails to have a best low-rank approximation has positive volume (De Silva & Lim, 2008). These results may be contrasted with the case when $k = 2$, i.e., the problem with the Eckart-Young theorem, which was discussed in the introductory section.

As reviewed extensively in Stegeman and Comon (2010), there has already been an extensive literature on this problem and related topics, and several psychometricians have partly contributed to them (e.g., Krijnen, Dijkstra, & Stegeman, 2008; Kruskal, 1989; Stegeman, 2006, 2007, 2008; Stegeman & Lathauwer, 2009; Ten Berge & Kiers, 1999; Ten Berge, Kiers, & De Leeuw, 1988; Ten Berge, Sidiropoulos, & Rocci, 2004).

It might be possible to overcome possible difficulties in introducing some appropriate tensors in the analysis of the triadic interaction problem discussed above by utilizing future methods for overcoming the ill-posedness of the lower-rank approximation problem. This is because the approach to this problem is still an active area of research, which has recently received increasing attention in statistics as well as in mathematics.

Acknowledgements

The author is indebted to reviewers for their valuable suggestions and comments on earlier version of this paper. The author is also indebted to Gregory L. Rohe for proofreading of an earlier version of this paper.

REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.
- Alfakih, A.Y., Khandani, A., & Wolkowicz, H. (1999). Solving Euclidean distance matrix completion problems via semidefinite programming. *Computational Optimization and Applications*, **12**, 13–30.
- Amari, S. (1971). Characteristics of randomly connected threshold-element networks and network systems. *Proceedings of the IEEE*, **59**, 35–47.
- Andersen, E.B. (1980). *Discrete Statistical Models with Social Science Applications*. Amsterdam: North Holland.
- Appelman, I.B., & Mayzner, M.S. (1982). Application of geometric models to letter recognition: Distance and density. *Journal of Experimental Psychology, General*, **111**, 60–100.
- Appleby, M.C. (1983). The probability of linearity in hierarchies. *Animal Behavior*, **31**, 600–6008.
- Arfken, G.B., & Weber, H.J.J. (1995). *Mathematical methods for physicists*. 4th ed. New York: Academic Press.
- Arnold, V.I. (1978). *Mathematical Methods of Classical Mechanics*. Berlin: Springer Verlag.
- Ashby, F.G., & Perrin, N.A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, **95**, 124–150.
- Bass, F.M., Pessemer, E.A., & Lehmann, D.R. (1972). An experimental study of relationships between attitudes, brand preference, and choice. *Behavioral Science*, **17**, 532–541.
- Basu, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhyā*, **15**, 377–380.
- Bennett, J.F., & Hays, W.L. (1960). Multidimensional unfolding: Determining the dimensionality of ranked preference data. *Psychometrika*, **25**, 27–43.
- Birch, M.W. (1963). Maximum likelihood in three-way contingency tables. *Journal of the Royal Statistical Society, Series B*, **25**, 220–233.
- Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W. (1975). *Discrete Multivariate Analysis-Theory and Practice*. Massachusetts: The MIT press.
- Bock, R.D., & Jones, L.V. (1968). *The Measurement and Prediction of Judgment and Choice*. Holden-Day, San Francisco.
- Bodenreider, O. (2001). Circular hierarchical relationships in the UMLS: Etiology, diagnosis, treatment, complications and prevention. *Proceedings of the AMIA symposium*, pp.57–61.
- Bond, C.F. Jr., Horn, E.M., & Kenny, D.A. (1997). A model for triadic relations. *Psychological Methods*, **2**, 79–94.
- Bond, C.F. Jr., Kenny, D.A., Broome, E.H., Stokes-Zoota, J.J., & Richard, F. D. (2000). Multivariate analysis of triadic relations. *Multivariate Behavioral Research*, **35**, 397–426.
- Borg, I., & Groenen, P.J.F. (2005). *Modern Multidimensional Scaling - Theory and Applications*. 2nd Ed. New York: Springer.
- Carroll, D.J., and Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an N -Way generalization of "Eckart-Young" decomposition. *Psychometrika*, **35**, 283–319.
- Caussinus, H. (1965). Contribution à l'analyse statistique des tableaux de corrélation. *Annales de la Faculté des Sciences de l'Université de Toulouse*, **29**, 77–182.
- Chadwick-Furman, N., & Rinkevich, B. (1994). A complex allorecognition system in a reef-building coral: delayed responses, reversals and nontransitive hierarchies. *Coral Reefs*, **13**, 57–63.
- Chino, N. (1977). N -ko no taisyo kan no hitaisyo na kankei wo zusikika suru tameno ichi-giho. [A graphical technique for representing the asymmetric relationships between N objects]. *Proceedings of the 5th annual meeting of the Behaviormetric Society of Japan*, 146–149 (in Japanese).

- Chino, N. (1978). A graphical technique for representing the asymmetric relationships between N objects. *Behaviormetrika*, **5**, 23–40.
- Chino, N. (1980). A unified geometrical interpretation of the MDS techniques for the analysis of asymmetry and related techniques. *Paper presented at the symposium on "Asymmetric multidimensional scaling" at the Spring Meeting of the Psychometric Society*, Iowa, U.S.A.
- Chino, N. (1990). A generalized inner product model for the analysis of asymmetry. *Behaviormetrika*, **27**, 25–46.
- Chino, N. (1991). A critical review for the analysis of asymmetric relational data. *Bulletin of The Faculty of Letters of Aichi Gakuin University*, **21**, 31–52.
- Chino, N. (1997). *Hitaisho Tajigen Syakudo Koseiho* [Asymmetric Multidimensional Scaling]. Kyoto: Gendai Sugaku (in Japanese).
- Chino, N. (1998). Hilbert space theory in psychology -(1) Basic concepts and possible applications. *Bulletin of the Faculty of Letters of Aichi Gakuin University*, No.**28**, 45–66.
- Chino, N. (2005). Abnormal behaviors of members predicted by a complex difference system model. *Bulletin of the Faculty of Psychological & Physical Science*, **1**, 69–73.
- Chino, N., & Nakagawa, M. (1990). A bifurcation model of change in group structure. *The Japanese Journal of Experimental Social Psychology*, **29**, 25–38.
- Chino, N., & Okada, A. (1996). Hitaisho tajigen syakudo Koseiho to sono syuhen [Asymmetric multidimensional scaling and related topics], *Kodokeiryogaku*, **23**, 130–152 (in Japanese).
- Chino, N., & Saburi, S. (2006). Tests of symmetry in asymmetric MDS. *Paper presented at the 2nd German Japanese Symposium on Classification - Advances in data analysis and related new techniques & application*. Berlin, Germany, March 7–8.
- Chino, N., & Saburi, S. (2009). Features of quasi-symmetry-like asymmetric MDS models and independence of some tests for symmetry. *Proceedings of the 37th annual meeting of the Behaviormetric Society of Japan*, pp.24–25, Oita University.
- Chino, N., & Saburi, S. (2010). Controlling the two kinds of error rate in selecting an appropriate asymmetric MDS model. *Journal of the Institute for Psychological and Physical Science*, **2**, 37–42.
- Chino, N., & Shiraiwa, K. (1993). Geometrical structures of some non-distance models for asymmetric MDS. *Behaviormetrika*, **20**, 35–47.
- Constantine, A.G. & Gower, J.C. (1978). Graphical representation of asymmetric matrices. *Applied Statistics*, **27**, 297–304.
- Coombs, C.H. (1964). *A Theory of Data*. New York: Wiley.
- Cox, T.F., & Cox, M.A.A. (1991). Multidimensional scaling for n -tuples. *British Journal of Mathematical and Statistical Psychology*, **44**, 195–206.
- Cox, T.F., & Cox, M.A.A. (2001). *Multidimensional Scaling*, 2nd ed., London: Chapman & Hall/CRC.
- Cox, T.F., Cox, M.A.A., & Branco, J.A. (1991). Multidimensional scaling for n -tuples. *British Journal of Mathematical and Statistical Psychology*, **44**, 195–206.
- Daws, J.T. (1996). The analysis of free-sorting data: Beyond pairwise co-occurrences. *Journal of Classification*, **13**, 57–80.
- De Rooij, M. (2002). Studying triadic distance models under a likelihood function. In S. Nishisato, Y. Baba, H. Bozdogan, & K. Kanefuji (Eds.), *Measurement and Multivariate Analysis* (pp.69–76). Tokyo: Springer.
- De Rooij, M., & Gower, J.C. (2003). The geometry of triadic distances. *Journal of Classification*, **20**, 181–220.
- De Rooij, M., & Heiser, W.J. (2000). Triadic distance models for the analysis of asymmetric three-way proximity data. *British Journal of Mathematical and Statistical Psychology*, **53**, 99–119.
- De Rooij, M., & Heiser, W.J. (2003). A distance representation of the quasi-symmetry model

- and related distance models. In H. Yanai et al. (Eds.) *New Developments in Psychometrics* (pp.487–494). Tokyo: Springer.
- De Rooij, M., & Heiser, W.J. (2005). Graphical representations and odds ratios in a distance-association model for the analysis of cross-classified data. *Psychometrika*, **70**, 99–122.
- DeSarbo, W.S., Johnson, M., Manrai, A., & Edwards, E. (1992). TSCALE: A new multidimensional scaling procedure based on Tversky's contrast model. *Psychometrika*, **57**, 43–70.
- DeSarbo, W.S., & De Soete, G. (1984). On the use of hierarchical clustering for the analysis of nonsymmetric proximities. *Journal of Consumer Research*, **11**, 601–610.
- De Silva, V., & Lim, L-H. (2008). Tensor rank and the ill-posedness of the best low-rank approximation problem. *The SIAM Journal of Matrix Analysis and Applications*, **30**, 1084–1127.
- De Vries, H. (1995). An improved test of linearity in dominance hierarchies containing unknown or tied relationships. *Animal Behavior*, **50**, 1375–1389.
- Digby, R.G.N., & Kempton, R.A. (1987). *Multivariate Analysis of Ecological Communities*. London: Chapman and Hall.
- Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, **1**, 211–218.
- Escoufier, Y., & Grorud, A. (1980). Analyse factorielle des matrices carrees non symetriques [Factor analysis of square asymmetric matrices]. In E. Diday et al. (Eds.), *Data Analysis and Informatics* (pp.263–276). Amsterdam: North Holland.
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.
- Fong, D.K.H., DeSarbo, W.S., Park, J., & Scott, C.J. (2010). A Bayesian vector multidimensional scaling procedure for the analysis of ordered preference data. *Journal of the American Statistical Association*, **105**, 482–492.
- Fukai, T., & Shiino, M. (1990). Asymmetric neural networks incorporating the Dale hypothesis and noise-driven chaos. *Physical Review Letters*, **64**, 1465–1468.
- Getty, D.J., Swets, J.A., Swets, J.B., & Green, D.M. (1979). On the prediction of confusion matrices from similarity judgments. *Perception & Psychophysics*, **26**, 1–19.
- Goodman, L.A. (1971). The analysis of multidimensional contingency tables: stepwise procedures and direct estimation methods for building models for multiple classifications. *Technometrics*, **13**, 33–61.
- Gower, J.C. (1977). The analysis of asymmetry and orthogonality. In J.R. Barra, F. Brodeau, G. Romer, & B. van Cutsem (Eds.), *Recent Developments in Statistics* (pp.109–123). Amsterdam: North Holland.
- Gower, J.C. (1984). Multivariate analysis: Ordination, multidimensional scaling and allied topics. In E.H. Llokyd (Ed.), *Handbook of Applicable Mathematics* (pp.727–781). Vol. VI, Chichester: Wiley.
- Gower, J.C., & De Rooij, M. (2003). A comparison of the multidimensional scaling of triadic and dyadic distances. *Journal of Classification*, **20**, 115–136.
- Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, **33**, 469–506.
- Harary, F. J(1968). *Graph Theory*. Massachusetts: Addison-Wesley.
- Harshman, R.A. (1978). Models for analysis of asymmetrical relationships among N objects or stimuli. *Paper presented at the First Joint Meeting of the Psychometric Society and the Society for Mathematical Psychology*, Hamilton, Canada.
- Harshman, R.A. (1981). *DEDICOM multidimensional analysis of skew-symmetric data. Part I: Theory*. Unpublished manuscript.
- Harshman, R.A., Green, P.E., Wind, Y., & Lundy, M.E. (1982). A model for the analysis of asymmetric data in marketing research. *Marketing Science*, **1**, 205–242.
- Hayashi, C. (1972). Two dimensional quantifications based on a measure of dissimilarity among three elements. *Annals of the Institute of Statistical Mathematics*, **25**, 251–257.

- Hayashi, C. (1989). Multiway data matrices and method of quantification of qualitative data as a strategy of data analysis. In R. Coppi and B. Bolasco (Eds.), *Multiway data analysis* (pp.131–142). Amsterdam: North-Holland.
- Hays, W.L., & Bennett, J.F. (1961). Multidimensional unfolding: Determining configuration from complete rank order preference data. *Psychometrika*, **26**, 221–238.
- Hefner, R.A. (1958). Extensions of the law of comparative judgment to discriminable and multi-dimensional stimuli. *Doctoral dissertation*, University of Michigan.
- Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology*, **21**, 107–112.
- Heiser, W.J. (1987). Joint ordination of species and sites: the unfolding technique. In P. Legendre & L. Legendre (Eds.), *Developments in numerical ecology* (pp.189–221). Berlin: Springer Verlag.
- Heiser, W.J., & Bannani, M. (1997). Triadic distance models: Axiomatization and least squares representation. *Journal of Mathematical Psychology*, **41**, 189–206.
- Hirotsu, C. (1983). Defining the pattern of association in two-way contingency tables. *Biometrika*, **70**, 579–589.
- Hogg, R.V., & Craig, A.T. (1956). Sufficient statistics in elementary distribution theory. *Sankhyā*, **17**, 209–216.
- Holman, E.W. (1979). Monotonic models for asymmetric proximities. *Journal of Mathematical Psychology*, **20**, 1–15.
- Horn, R.A., & Johnson, C.R. (1985). *Matrix analysis*. Cambridge: Cambridge University Press.
- Irony, T.Z., & Singpurwalla, N.D. (1997). Non-informative priors do not exist: A dialogue with José M. Bernardo. *Journal of Statistical Planning and Inference*, **65**, 159–189.
- Je, H., Kim, D., & Pohang, P. (2008). Bayesian multidimensional scaling for multi-robot localization. *Proceedings of IEEE International Conference on Networking, Sensing and Control*, 926–931.
- Johnson, N., Kotz, S., & Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. New York: Wiley.
- Joly, S., & Le Calvé, G. (1995). Three-way distances. *Journal of Classification*, **12**, 191–205.
- Kanter, I. J.(1988). Inhomogeneous magnetization in dilute asymmetric and symmetric systems. *Physical Review Letters*, **60**, 1891–1894.
- Kastenbaum, M.A. (1960). A note on the additive partitioning of chi-square in contingency tables. *Biometrics*, **16**, 416–422.
- Kendall, M.G. (1962). *Rank Correlation Methods* (3rd Ed.). London: Charles Griffin.
- Kendall, M.G., & Smith, B.B. (1940). On the method of paired comparisons. *Biometrika*, **31**, 324–345.
- Keren, G., & Baggen, S. (1981). Recognition models of alphanumeric characters. *Perception & Psychophysics*, **29**, 234–246.
- Kiers, H.A.L., & Takane, Y. (1994). A generalization of GIPSCAL for the analysis of asymmetric data. *Journal of Classification*, **11**, 79–99.
- Kosugi, K. (2004). *Shinri-Ronri to Taido-ron he no Suri-apurouchi [Mathematical Approaches to Interpersonal Psycho-Logic and Attitude Theory]*. Unpublished doctoral dissertation, Kwansai-Gakuin University, Nishinomiya.
- Kree, R., & Zippelius, A. (1995). Asymmetrically diluted neural networks. In E. Domany, J.L. van Hemmen, and K. Schulten (Eds.), *Models of Neural Networks I* (2nd ed.) (pp.201–220).
- Krijnen, W.P., Dijkstra, T.K., & Stegeman, A. (2008). On the non-existence of optimal solutions and the occurrence of "degeneracy" in the CANDECOP/ PARA- FAC model. *Psychometrika*, **73**, 431–439.
- Krumhansl, C.L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, **85**, 445–463.
- Krummenauer, F. (1998). Limit theorems for multivariate discrete distributions. *Metrika*, **47**,

47–69.

- Kruskal, J.B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, **29**, 1–27.
- Kruskal, J.B. (1964b). Multidimensional scaling: A numerical method. *Psychometrika*, **29**, 115–129.
- Kruskal, J.B. (1989). Rank, decomposition, and uniqueness for 3-way and N -way arrays. In R. Coppi and S. Bolasco (Eds.), *Multway Data Analysis* (pp.7–18). Amsterdam: North Holland.
- Kumagai, A. (2010). Kotenteki Tajigen Schakudo Koseiho ni motozuku Hitaisyo Kanrensei deta no Bunseki [Analysis of asymmetric relational data based on classical multidimensional scaling]. *Nihon Oyo Sugakukai Ronbunshi*, **20**, 57–66 (in Japanese).
- Kuriki, S. (1991). A construction of simultaneous confidence regions and multiple comparison procedures concerning the paired comparison model. *Ouyo Tokeigaku*[Applied Statistics], **20**, 127–137 (in Japanese).
- Laurent, M. (2001). Matrix completion problems. *Encyclopedia of Optimization*, **3**, 221–229.
- Lee, M.D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, **15**, 1–15.
- Lehmann, E.L. (1983). *Theory of Point Estimation*. New York: Wiley.
- Li, Z. (2008). Exponential stability of synchronization in asymmetrically coupled dynamical networks. *Chaos*, **18**, 1–11.
- Lingoes, J.C. (1973). *The Guttman-Lingoes nonmetric program series*. Ann Arbor: Mathesis Press.
- Luce, R.D. (1963). Detection and recognition. In R.D. Luce, R.R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp.103–189). New York: Wiley.
- Loisel, S., & Takane, Y. (2011). Generalized GIPSCAL re-revisited: a fast convergent algorithm with acceleration by the minimal polynomial extrapolation. *Advances in Data Analysis and Classification*, **5**, 57–75.
- Masure, R.H., & Allee, W.C. (1934). The social order in flocks of the common chicken and the pigeon. *The Auk*, **51**, 306–327.
- Matsumoto, M. (1986). *Foundation of Finsler Geometry and Special Finsler Space*. Ohtsu: Kaishisha.
- Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification Learning. *Psychological Review*, **85**, 207–238.
- Nakatani, L.H. (1972). Confusion-choice model for multidimensional psychophysics. *Journal of Mathematical Psychology*, **9**, 104–127.
- Nakayama, A. (2005). A multidimensional scaling model for three-way data analysis. *Behaviormetrika*, **32**, 95–110.
- Nakayama, A., & Okada, A. (2011). Reconstructing one-mode three-way asymmetric data for multidimensional scaling. *Proceedings of the 2010 GFKL meeting*. (in print).
- Newcomb, T.M. J(1953). An approach to the study of communicative acts. *Psychological Review*, **60**, 393–404.
- Newcomb, T.M. (1961). *The Acquaintance Process*. New York: Holt, Rinehart and Winston.
- Neyman, J., & Scott, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrika*, **16**, 1–32.
- Nosofsky, R.M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology*, **10**, 104–114.
- Nosofsky, R.M. (1985a). Luce's choice model and Thurstone's categorical judgment model compared: Kornbrot's data revisited. *Perception & Psychophysics*, **37**, 89–91.
- Nosofsky, R.M. (1985b). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Perception & Psychophysics*, **38**, 415–432.

- Nosofsky, R.M. (1986). Attention, similarity, and the identification-categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **13**, 87–108.
- Nosofsky, R.M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, **23**, 94–140.
- Oh, M-S, and Raftery, A.E. (2001). Bayesian multidimensional scaling and choice of dimension. *Journal of the American Statistical Association*, **96**, 1031–1044.
- Oh, M-S, and Raftery, A.E. (2007). Model-based clustering with dissimilarities: A Bayesian approach. *Journal of Computational and Graphical Statistics*, **16**, 559–585.
- Okada, A., & Imaizumi, T. (1984). Geometric models for asymmetric similarity. *Research Reports of School of Social Relations, Rikkyo (St. Paul's) University*.
- Okada, A., & Imaizumi, T. (1987). Nonmetric multidimensional scaling of asymmetric proximities. *Behaviormetrika*, **21**, 81–96.
- Okada, A., & Imaizumi, T. (1997). Asymmetric multidimensional scaling of two-mode, three-way proximities. *Journal of Classification*, **14**, 195–224.
- Okada, K. (2011). Beizu suitei ni yoru hitaisyo MDS. [Asymmetric MDS by Bayesian estimation]. *Proceedings of the 39th annual meeting of the Behaviormetric Society of Japan*. Okayama Science University (in print).
- Okada, K., & Mayekawa, S. (2011). Bayesian nonmetric successive categories multidimensional scaling. *Behaviormetrika*, **38**, 17–32.
- Okada, K., & Shigemasu, K. (2010). Bayesian multidimensional scaling for the estimation of a Minkowski exponent. *Behavior Research Methods*, **42**, 899–905.
- Parisi, G. (1986). Asymmetric neural networks and the process of learning. *Journal of Physics A: Mathematical and General*, **19**, L675–L680.
- Park, J., DeSarbo, W.S., & Liechty, J. (2008). A hierarchical Bayesian multidimensional scaling methodology for accommodating both structural and preference heterogeneity. *Psychometrika*, **73**, 451–472.
- Press, S.J. (1989). *Bayesian Statistics - Principles, models, and applications*. New York: Wiley.
- Ramsay, J.O. (1969). Some statistical considerations in multidimensional scaling. *Psychometrika*, **34**, 167–182.
- Ramsay, J.O. (1977). Maximum likelihood estimation in multidimensional scaling. *Psychometrika*, **42**, 241–266.
- Ramsay, J.O. (1978). Confidence regions for multidimensional scaling analysis. *Psychometrika*, **43**, 145–160.
- Ramsay, J.O. (1982). Some statistical approaches to multidimensional scaling data. *The Journal of the Royal Statistical Society, Series A (General)*, **145**, 285–312.
- Randers, G. (1941). On an asymmetrical metric in the four-space of general relativity. *Physical Review*, **59**, 195–199.
- Richardson, M.W. (1938). Multidimensional Psychophysics. *Psychological Bulletin*, **35**, 659.
- Rocci, R., & Bove, G. (2002). Rotation techniques in asymmetric multidimensional scaling. *Journal of Computational and Graphical Statistics*, **11**, 405–419.
- Rushen, J. (1982). The peck orders of chickens: How do they develop and why are they linear? *Animal Behavior*, **30**, 1129–1137.
- Saburi, S. & Chino, N. (2008). A maximum likelihood method for an asymmetric MDS model. *Computational Statistics and Data Analysis*, **52**, 4673–4684.
- Saito, T. (1980). *Tajigen Shakudo Kouseiho [Multidimensional Scaling]*. Tokyo: Asakura Book Company (in Japanese).
- Saito, T. (2002). Circle structure derived from decomposition of asymmetric data matrix. *Journal of the Japanese Society of Computational Statistics*, **15**, 1–18.
- Saito, T. (1991). Analysis of asymmetric proximity matrix by a model of distance and additive terms. *Behaviormetrika*, **29**, 45–60.

- Saito, T., & Takeda, S. (1990). Multidimensional scaling of asymmetric proximity: model and method. *Behaviormetrika*, **28**, 49–80.
- Saito, T., & Yadohisa, H. (2005). *Data Analysis of Asymmetric Structures - Advanced Approaches in Computational Statistics*. New York: Marcel Dekker.
- Sato, Y. (1988). An analysis of sociometric data by MDS in Minkowski space. In K. Matsusita (Ed.), *Statistical Theory and Data Analysis II*, (pp.385–396). Amsterdam: North Holland.
- Sato, Y. (1989). Minkowski keiryō o mochiita hitaisyo ruijisei no kyori hyogen ni tsuite. [On the distance representation of asymmetric dissimilarities by Minkowskimetric function]. *Computational Statistics*, **2**, 35–45. (in Japanese).
- Schönemann, P.H. (1970). On metric multidimensional unfolding. *Psychometrika*, **35**, 349–366.
- Shepard, R.N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, **22**, 325–345.
- Shepard, R.N. (1958a). Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review*, **65**, 242–256.
- Shepard, R.N. (1958b). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, **55**, 509–523.
- Shepard, R.N. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America*, **36**, 2346–2353.
- Smith, J.E.K. (1982). Recognition models evaluated: A commentary on Keren and Baggen. *Perception & Psychophysics*, **31**, 183–189.
- So, A. M-C., & Ye, Y. (2007). Theory of semidefinite programming for sensor network localization. *Mathematical Programming*, **109**, 367–384.
- Stegeman, A. (2006). Degeneracy in CANDECOMP/PARAFAC explained for $p \times p \times 2$ arrays of rank $p + 1$ or higher. *Psychometrika*, **71**, 483–501.
- Stegeman, A. (2007). Degeneracy in CANDECOMP/PARAFAC explained for several three-sliced arrays with a two-valued typical rank. *Psychometrika*, **72**, 601–619.
- Stegeman, A. (2008). Low-rank approximation of generic $p \times q \times 2$ arrays and diverging components in the CANDECOMP/PARAFAC model. *SIAM Journal on Matrix Analysis and Applications*, **30**, 988–1007.
- Stegeman, A., & Comon, P. (2010). Subtracting a best rank-1 approximation may increase tensor rank. *Linear Algebra and its Applications*, **433**, 1276–1300.
- Stegeman, A., & De Lathauwer, L. (2009). A method to avoid diverging components in the CANDECOMP/PARAFAC model for generic $I \times J \times 2$ arrays. *SIAM Journal on Matrix Analysis and Applications*, **30**, 1614–1638.
- Suppes, P., & Zinnes, J.L. (1963). Basic measurement theory. In R.D. Luce et al. (Eds.), *Handbook of mathematical psychology*, Vol.1, New York: Wiley.
- Takane, Y. (1978a). A maximum likelihood method for nonmetric multidimensional scaling: 1. The case in which all empirical pairwise orderings are independent- Theory. *Japanese Psychological Research*, **20**, 7–17.
- Takane, Y. (1978b). A maximum likelihood method for nonmetric multidimensional scaling: 2. The case in which all empirical pairwise orderings are independent- Evaluation. *Japanese Psychological Research*, **20**, 105–114.
- Takane, Y. (1880). Tajigen Shakudoho [Multidimensional Scaling]. Tokyo: Tokyo University Press (in Japanese).
- Takane, Y. (1981). Multidimensional successive categories scaling: A maximum likelihood method. *Psychometrika*, **46**, 9–28.
- Takane, Y. (1987). Analysis of contingency tables by ideal point discriminant analysis. *Psychometrika*, **52**, 493–513.
- Takane, Y., and Carroll, J.D. (1981). Nonmetric maximum likelihood multidimensional scaling from directional rankings of similarities. *Psychometrika*, **46**, 389–405.

- Takane, Y., & Kiers, H.A.L. (1997). Latent class DEDICOM. *Journal of Classification*, *14*, 225–247.
- Takane, Y., & Shibayama, T. (1986). Comparison of models for stimulus recognition data. In J. de Leeuw, et al. (Eds.), *Multidimensional Data Analysis* (pp.119–148). Leiden: DSWO Press.
- Takane, Y., & Shibayama, T. (1992). Structures in stimulus identification data. In F.G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp.335–362). New Jersey: Lawrence Erlbaum Associates.
- Takane, Y., Young, F.W., and de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternative least squares method with optimal scaling features. *Psychometrika*, *42*, 1–68.
- Ten Berge, J.M.F. (1997). Reduction of asymmetry by rank-one matrices. *Computational Statistics & Data Analysis*, *24*, 357–366.
- Ten Berge, J.M.F., & Kiers, H.A.L. (1999). Simplicity of core arrays in three-way principal component analysis and the typical rank of $p \times q \times 2$ arrays. *Linear Algebra and its Applications*, *294*, 169–179.
- Ten Berge, J.M.F., & Kiers, H.A.L., & De Leeuw (1988). Explicit CANDECOMP/PARAFAC solutions for a contrived $2 \times 2 \times 2$ array of rank three. *Psychometrika*, *53*, 579–584.
- Ten Berge, J.M.F., Sidiropoulos, N.D., & Rocci, R. (2004). Typical rank and INDSCAL dimensionality for symmetric three-way arrays of order $I \times 2$ or $I \times 3 \times 3$. *Linear Algebra and its Applications*, *388*, 363–377.
- Tobler, W. (1976–77). Spatial interaction patterns. *Journal of Environmental Systems*, *6*, 271–301.
- Tomizawa, S. (1992). Multiplicative models with further restrictions on the usual symmetry model. *Communications in Statistics - Theory and Methods*, *21*, 693–710.
- Torgerson, W.S. (1954). A law of categorical judgment. *The American Psychologist*, *9*, 483.
- Torgerson, W.S. (1958). *Theory and Methods of Scaling*. Wiley, New York.
- Townsend, J. t., & Landon, D.E. (1982). An experimental and theoretical investigation o the constant-ratio rule and other models of visual letter confusion. *Journal of Mathematical Psychology*, *25*, 119–162.
- Trendafilov, N.T. (2002). GIPSCAL revisited. A projected gradient approach. *Statistics and Computing*, *12*, 135–145.
- Tucker, L.R. and Messick, S. (1963). An individual differences model for multidimensional scaling. *Psychometrika*, *28*, 333–367.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.
- Yadohisa, H., Niki, N. (1999). Vector field representation of asymmetric proximity data. *Communications in Statistics, Theory and method*, *28*, 35–48.
- Young, F.W. (1975). An asymmetric Euclidean model for multi-process asymmetric data. *Paper presented at U.S.-Japan Seminar on MDS*, San Diego, U.S.A.
- Young, G., & Householder, A.S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika*, *3*, 19–22.
- Weeks, D.G., & Bentler, P.M. (1982). Restricted multidimensional scaling models for asymmetric proximities. *Psychometrika*, *47*, 201–208.
- Weinberger, K.Q., & Saul, L.K. (2004). Unsupervised learning of image manifolds by semidefinite programming. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. *70*, 77–90.
- Wilkinson, J.H. (1965). *The Algebraic Eigenvalue Problem*. Oxford: Clarendon Press.
- Wishart, J. (1949). Cumulants of multivariate multinomial distributions. *Biometrika*, *36*, 47–58.
- Zielman, B., & Heiser, W.J. (1993). Analysis of asymmetry by a slide vector. *Psychometrika*, *58*, 101–114.

- Zielman, B., & Heiser, W.J. (1996). Models for asymmetric proximities. *British Journal of Mathematical and Statistical Psychology*, **49**, 127–146.
- Zinnes, J.L., & MacKay, D.B. (1983). Probabilistic multidimensional scaling: complete and incomplete data. *Psychometrika*, **48**, 27–48.

(Received June 1 2011, Revised November 28 2011)