JOURNAL OF MATHEMATICAL PSYCHOLOGY **41**, 189–206 (1997)
ARTICLE NO. MP971166

# Triadic Distance Models: Axiomatization and Least Squares Representation

Willem J. Heiser

*Leiden University*

and

Mohammed Bennani

*Université de Haute Bretagne Rennes II*

Distance models for three-way proximity data, which consist of numerical values assigned to triples of objects that indicate their joint (lack of) homogeneity or resemblance, require a generalization of the usual distance concept defined on pairs of objects. An axiomatic framework is given for characterizing triadic dissimilarity, triadic similarity, and triadic distance, where the term triadic implies that each element of the triple is treated on an equal footing. Two kinds of distance models are studied in detail: the Minkowski-$p$ or $M_p$ model, which is based upon dyadic components and includes the perimeter model as an important special case, and several models based on presence–absence variables. They are shown to satisfy the tetrahedral inequality, a condition that is characteristic for the present axiomatization. Two monotonically convergent algorithms are described that find weighted least squares representations of three-way proximity data under the Euclidean $M_1$ model and the Euclidean $M_2$ model. To enable a scalefree evaluation of the quality of the fit, an additive decomposition of the sum of squares of the dissimilarities is derived. As illustrated in one of the examples, distance analysis of three-way, three-mode tables is possible by a suitable manipulation of the least squares weights. © 1997 Academic Press

## INTRODUCTION

Three-way proximity data consist of numerical values assigned to triples of objects that indicate their joint (lack of) homogeneity or resemblance. There has been a growing interest in the analysis of three-way data, but the mathematical structure of the models used is predominantly multilinear (Law, Snyder, Hattie, and McDonald, 1984; Coppi and Bolasco, 1989). Three-way generalizations of distance models, which form the characteristic structure of multidimensional scaling and hierarchical clustering representations, have been studied by only a few authors.

In Tucker's (1964) terminology, this paper is primarily concerned with proximality data collected in a *three-way*, *one-mode* table, in which each dimension has indices that refer to the same set of objects. To clearly distinguish this case from others, we will refer to the elements in these tables as *triadic proximity data*, while proximities defined on pairs will be called *dyadic*. Triadic proximities are not often directly collected. Daws (1996) used the free sorting method, in which subjects are instructed to produce a partition of the set of stimuli into any number of classes, according to any self-selected criterion. One of the usual procedures in this method is to construct dyadic proximity data from the partitions by counting the number of times $\rho_{ij}$ that each pair of stimuli $(i, j)$ has been placed in the same class. The following example, due to Daws, shows how the reduction from a distribution over all subset patterns to a dyadic similarity implies loss of information about the way in which the individuals have classified the objects. Suppose that free sorting by two groups of subjects yields the two distributions given in Table 1. Here (123)–(4) indicates that objects 1, 2, and 3 have been put together, but 4 has been left apart, and so on. It is easily verified that for both groups we obtain the dyadic similarities $\rho_{12}=6$, $\rho_{13}=6$, $\rho_{23}=7$, $\rho_{14}=0$, $\rho_{24}=4$, $\rho_{34}=3$. So, if one uses only dyadic information, the two groups cannot be recognized, because they do not distinguish pairs differently. Now, let $\rho_{ijk}$ denote the number of individuals that have classified objects $i$, $j$, and $k$ into the same subset of their partition. In the example, we obtain

$$\text{for Group 1: } \rho_{123}=5, \quad \rho_{124}=0, \quad \rho_{134}=0, \quad \rho_{234}=1,$$
$$\text{for Group 2: } \rho_{123}=1, \quad \rho_{124}=0, \quad \rho_{134}=0, \quad \rho_{234}=2.$$

Use of the triadic similarities clearly leads to the conclusion that two groups have classified the objects in a different

A large part of this work is based on the unpublished dissertation of the second author, entitled *Analyses Métriques á Trois Voies*, defended at the Université de Haute Bretagne Rennes II, France, on October 22, 1993.

Correspondence and requests for reprints should be addressed to Willem J. Heiser, Department of Data Theory, University of Leiden, P.O. Box 9555, 2300 RB Leiden, The Netherlands. E-mail: Heiser@Rulfsw.LeidenUniv.nl.



0022-2496/97 $25.00

**TABLE 1**

**Frequency of Subset Choice in Two Groups of Subjects (Daws' Example)**

| Subset pattern | Group 1 | Group 2 |
|---|---|---|
| (1234) | 0 | 0 |
| (123)–(4) | 5 | 1 |
| (124)–(3) | 0 | 0 |
| (134)–(2) | 0 | 0 |
| (1)–(234) | 1 | 2 |
| (12)–(34) | 0 | 1 |
| (13)–(24) | 1 | 2 |
| (14)–(23) | 0 | 0 |
| (12)–(3)–(4) | 1 | 4 |
| (13)–(2)–(4) | 0 | 3 |
| (1)–(23)–(4) | 1 | 4 |
| (14)–(2)–(3) | 0 | 0 |
| (1)–(24)–(3) | 2 | 0 |
| (1)–(2)–(34) | 2 | 0 |
| (1)–(2)–(3)–(4) | 5 | 1 |
| Total | 18 | 18 |

fashion, and we may ask ourselves how to construct geometric models accounting for these differences.

Hayashi (1972) has presumably been the first to develop a triadic model for the dissimilarity among three elements from the same set; he proposed the *surface area model*, using a special quantification method in two dimensions for modeling the data by the squared area of the triangle formed by three points (also see Hayashi, 1989). A similar method has been developed by Pan and Harris (1991), who used the area model to define triadic similarities in high-dimensional space and then modeled these in low-dimensional space with the square of what we will call the *generalized Euclidean model*, or $M_2$ model. Independently, Cox, Cox, and Branco (1991) developed a least squares method for the $M_2$ model and extended it to the nonmetric case and to $K$-adic relations. Joly and Le Calvé (1995) initiated the axiomatic study of distance models for triads, introduced the *three-way ultrametric* and the *semi-perimeter model*, and stated the necessary and sufficient conditions for Euclidean embeddability of the $M_2$ model.

Various concepts of triadic distance emphasize diverse aspects of the differences among three objects. Consider, for instance, the six points in Fig. 2, an example that will be discussed more extensively later, and regard in particular the triads $(1, 3, 5)$ and $(3, 5, 6)$. The area of the triangle $(1, 3, 5)$ is 5.085, while the area of triangle $(3, 5, 6)$ is 2.970; by contrast, the Euclidean semi-perimeter (half the sum of the length of the sides of the triangle) of $(1, 3, 5)$ is 6.256, while the comparable semi-perimeter of $(3, 5, 6)$ is 11.728. Thus, while under one model the lack of homogeneity for the elongated triangle $(3, 5, 6)$ is almost two times $(11.728/6.256 = 1.875)$ larger than for the more compact triangle $(1, 3, 5)$, under the other model this ratio is the reverse $(2.970/5.085 = 0.584)$. These differences suggest that we may except quite different representations of the same data under different kinds of models. It turns out that the area of a triangle does not satisfy the definition of triadic distance to be proposed shortly, while the semi-perimeter does.

We will start our discussion of triadic distance models by setting up an axiomatic framework, in which triadic similarity and dissimilarity are defined as mappings from the three-way Cartesian product of the set of objects into the nonnegative reals that satisfy very general conditions, the most important of which is *three-way symmetry*. Then the ordinary concept of distance between pairs is extended to the case of triads by posing an inequality to be called the *tetrahedral inequality*, which bounds the triadic distance associated with one face of a tetrahedron in terms of the sum of the triadic distances associated with the other three faces, and we establish some properties of the resultant class of functions. Two kinds of models are discussed in detail: the *Minkowski-p* or $M_p$ model, of which the perimeter model and the $M_2$ model are special cases, and several models based on presence–absence variables, including Daws' (1996) model, which generalize some of the common dyadic (dis)similarity indices. For all models it is shown that they satisfy the stated conditions for being a triadic distance. We will then describe two convergent algorithms—one for the Euclidean perimeter model, the other for the $M_2$ model—to find least squares representations of triadic dissimilarities and demonstrate how the total dispersion can be broken down into a residual component and a percentage dispersion accounted for. Since they incorporate weights and are based on a general algorithmic strategy, the proposed procedures are very versatile; as shown in one of the examples, they can be used for three-way, three-mode data as well.

## TRIADIC DISSIMILARITY, TRIADIC SIMILARITY, AND TRIADIC DISTANCE

Let $O$ be a finite set of $n$ elements, denoted by $O = \{1, 2, ..., i, j, k, ..., n\}$, which are the labels of the modeling units, or *objects*. We start by defining the concept of triadic dissimilarity, then provide a dual treatment of similarity, and finally move to triadic distance by adding two more properties to the dissimilarity definition, one shared by all semi-metrics, and another comparable to the triangle inequality for ordinary metrics.

A triadic dissimilarity on $O$ measures the lack of resemblance between objects in $O$ taken three at a time, which is zero if there is no lack of resemblance, and positive otherwise. Mathematically, it is represented as a mapping $T$ of $O \times O \times O$ into $\mathbb{R}^+$, the nonnegative reals. Thus the first property that a triadic dissimilarity $\tau_{ijk} = T(i, j, k)$ has to satisfy is *nonnegativity*,

$$\tau_{ijk} \geqslant 0, \tag{1a}$$

for all $i$, $j$, and $k$ in $O$. It is conceptually undesirable if the (lack of) resemblance between three objects would depend on the order in which they are listed. Therefore, the second natural requirement is *three-way symmetry*: that is, for all permutations $\pi$ of $\{i, j, k\}$ we must have

$$\pi_{ijk} = \tau_{\pi(i)\,\pi(j)\,\pi(k)}, \tag{1b}$$

for all $i$, $j$, and $k$. Lack of resemblance of an object with itself should not be different from zero, so we require *minimality*:

$$\tau_{ijk} = 0 \quad \text{if } i = j = k. \tag{1c}$$

When one of the objects is identical to one of the others, the lack of resemblance between the two nonidentical objects should remain invariant regardless of which two are the same, so that a last natural requirement is

$$\tau_{iji} = \tau_{ijj}, \tag{1d}$$

for all $i$ and $j$. Requirement (1d) is called *diagonal-plane equality*, because it requires equality of the three matrices $\{\tau_{iij}\}$, $\{\tau_{iji}\}$, and $\{\tau_{ijj}\}$, which are formed by cutting the three-way block diagonally, starting at one of the three edges joining at the corner $\tau_{111}$. By symmetry, we must also have $\tau_{iij} = \tau_{jji}$, $\tau_{iji} = \tau_{jij}$, $\tau_{ijj} = \tau_{jii}$, and so on. The quantities in the diagonal planes of the three-way table simply measure the *dyadic dissimilarity*, defined by $\delta_{ij} = \tau_{iij}$. Note that where the three planes intersect we have the elements of the (three-way) diagonal $\delta_{ii} = \tau_{iii} = 0$. Summarizing, a *triadic dissimilarity* on $O$ is a mapping of the threefold Cartesian product of $O$ into the real numbers that satisfies (1a)–(1d).

Triadic dissimilarity is defined here on triples of elements selected from a single set $O$. The case where the three elements are selected from three distinct sets (mentioned in the Introduction) is a special case, obtained by partitioning $O$ into $\{O_1, O_2, O_3\}$ and considering triples from the subset $O_1 \times O_2 \times O_3$ only. Thus three-way, three-mode data are regarded as a special case of three-way, one-mode data, just as two-way, two-mode data (e.g., individual preferences) can be regarded as a special case of two-way, one-mode data (in the example: similarities between two distinct groups of objects). It is also important to note that the notion of triadic dissimilarity is entirely different from the notion of three-way, two-mode dissimilarity (used in Carroll and Chang's (1970) INDSCAL model), because the latter generally lacks three-ways symmetry and diagonal plane equality.

It is possible to define in a dual way the notion of *triadic similarity* as a measure of resemblance on triples of objects. Formally, a triadic similarity function is a mapping $R$ of $O \times O \times O$ into $\mathbb{R}^+$ such that, $\forall i, j, k \in O$, and $\forall \pi$ on $\{i, j, k\}$, $\rho_{ijk} = R(i, j, k)$ satisfies

$$\rho_{ijk} \geqslant 0, \tag{2a}$$

$$\rho_{ijk} = \rho_{\pi(i)\,\pi(j)\,\pi(k)}, \tag{2b}$$

$$\rho_{iii} = \rho_{jjj} = \rho_{kkk} \geqslant \rho_{ijk}, \tag{2c}$$

$$\rho_{iji} = \rho_{ijj}. \tag{2d}$$

The notions of dissimilarity and similarity play opposite roles. Like in the two-way case, we may associate each triadic similarity $R$ with one triadic dissimilarity $T$ by specifying a decreasing function. As an extension of classic transformations (Gower, 1986) we could specify, for example, $\tau_{ijk} = \rho_{iii} - \rho_{ijk}$, $\tau_{ijk} = (\rho_{iii} - \rho_{ijk})^{1/2}$, or $\tau_{ijk} = (\rho_{iii}^2 - \rho_{ijk}^2)^{1/2}$.

Triadic dissimilarity functions can be embedded in a vector space of dimension $n(n^2-1)/6$, and the set of all such functions is a closed convex cone, in fact the nonnegative orthant of this vector space. Not all these functions allow a simple representation, and therefore it is often desirable to require a stronger, metric structure. In the notation we will distinguish triadic distances from triadic dissimilarities by writing $t_{ijk}$ for the former and $\tau_{ijk}$ for the latter. Analogous to the dyadic case, where a basic requirement already for semi-metric structure is *definiteness*, a first addition is

$$t_{ijk} = 0 \quad \textit{only} \text{ if } i = j = k. \tag{3a}$$

Thus, while zero triadic dissimilarity is not excluded in triads of different objects, the notion of a semi-metric space implies that coinciding points are regarded as identical. In the role of the *triangle inequality*, we propose the *tetrahedral inequality* as a second additional constraint, defined as, $\forall i, j, k, l \in O$:

$$2t_{ijk} \leqslant t_{ikl} + t_{jkl} + t_{ijl}. \tag{3b}$$

This compares, in the tetrahedron formed by four corners $i$, $j$, $k$, and $l$, the size of the face made up by $i$, $j$, and $k$ with the sum of the sizes of the faces that join at corner $l$. The corresponding empirical principle is as follows: three objects that all resemble a fourth object cannot be very different, while three objects being very different implies that at least two of them are very different from the fourth. If neither of them resembles the fourth object they still could be arbitrarily similar among themselves.

Summarizing, a *triadic distance function* is defined as a mapping $T$ of $O^3$ into $\mathbb{R}^+$ satisfying (1a)–(1d) and (3a)–(3b). Certain axioms in the definition are redundant; for example, three-way symmetry (1b) and the tetrahedral inequality (3b) imply positivity, as can been seen by adding the three inequalities $2t_{ikl} \leqslant t_{ijk} + t_{ijl} + t_{jkl}$, $2t_{jkl} \leqslant t_{ijk} + t_{ikl} + t_{ijl}$, and $2t_{ijl} \leqslant t_{ijk} + t_{ikl} + t_{jkl}$ member by member, which yields $t_{ijk} \geqslant 0$. We also have the following lemma.

LEMMA 1. *The tetrahedral inequality together with three-way symmetry implies* $t_{ijk} \leqslant t_{ikl} + t_{jkl}$ *for all* $i, j, k, l \in O$.

*Proof.* Interchanging the roles of $k$ and $l$ in (3b) and dividing by 2, we have

$$t_{ijl} \leqslant \tfrac{1}{2}(t_{ijk} + t_{ikl} + t_{jkl}).$$

Adding this inequality to (3b) and multiplying by $\frac{2}{3}$ gives the desired inequality. ▮

The present triadic distance is different from the concept of *three-way distance* used in Joly and Le Calvé (1995). The Joly–Le Calvé three-way distance is based on the inequalities

$$t_{iij} \leqslant t_{ijk} \quad \text{and} \quad \max(t_{ijk}, t_{ijl}) \leqslant t_{ikl} + t_{jkl}. \tag{4}$$

As can be readily verified by setting $k = l$ and writing $d_{ij} = \frac{1}{2}t_{iij}$, these inequalities imply the triangle inequality $d_{ij} \leqslant d_{ik} + d_{jk}$. The tetrahedral inequality does *not* imply the triangle inequality, but we do have the following result.[1]

LEMMA 2. *Given a triadic distance function* $T$ *(that is,* $T$ *satisfies* (1a)–(1d) *and* (3a)–(3b)), *define* $d_{ij} = \frac{1}{2}t_{iij}$. *Then the inequalities*

$$\tfrac{1}{3}(d_{ij} + d_{ik} + d_{jk}) \leqslant t_{ijk} \leqslant \tfrac{4}{3}(d_{ij} + d_{ik} + d_{jk}), \tag{5}$$

$$d_{ij} \leqslant \tfrac{5}{4}(d_{ik} + d_{jk}), \tag{6}$$

*hold for all* $i, j, k \in O$.

*Proof.* Setting $k$ equal to $j$ in (3b) and then substituting $l$ by $k$, we obtain

$$2d_{ij} \leqslant t_{ijk} + d_{jk}. \tag{I}$$

Setting $l$ equal to $k$ in (3b) gives

$$t_{ijk} \leqslant 2(d_{ik} + d_{jk}). \tag{II}$$

Interchanging $k$ and $i$ in (I), we obtain $d_{jk} \leqslant \frac{1}{2}t_{ijk} + \frac{1}{2}d_{ij}$, and adding this inequality to (I) yields

$$d_{ij} \leqslant t_{ijk}. \tag{III}$$

Interchanging $k$ with $i$ and $j$, respectively, in (II), and then adding up all three variants of (II) yields the upper bound for $t_{ijk}$ asserted in (5). The corresponding lower bound is obtained in a similar way from (III) and its two variants. Finally, addition of (I) and (II) gives

$$2d_{ij} \leqslant 2d_{ik} + 3d_{jk}.$$

Adding this inequality to the corresponding one where $i$ and $j$ are interchanged yields the asserted inequality (6). ▮

Note that (6) is an instance of a parametrized triangle inequality with parameter $\frac{5}{4}$ in the sense of Andreae and Bandelt (1995). The examples in Fig. 1 show that the inequalities (5) and (6) are the best possible. In each triangle the value $t_{ijk}$ is inscribed, whereas the derived values $d_{ij}$, $d_{ik}$, and $d_{jk}$ are labeling the corresponding sides. For example (a) the first inequality of (5) is tight, and, for (b) the second one is tight. For (c), inequality (6) becomes an equality. Example (a) also shows that $t_{iij} \leqslant t_{ijk}$ is not necessarily true, whence a triadic distance need not be a Joly–Le Calvé distance. Conversely, a Joly–Le Calvé distance satisfies the above inequalities (II) and (III), but rather $2d_{ij} \leqslant t_{ijk} + 2d_{jk}$ than (I); example (d) shows that (I) can in fact be violated. Still, a Joly–Le Calvé distance satisfies the following parametrized tetrahedral inequality with parameter $\frac{3}{4}$:

$$2t_{ijk} \leqslant \tfrac{4}{3}(t_{ikl} + t_{jkl} + t_{ijl}).$$

For a triadic distance, validity of $t_{iij} \leqslant t_{ijk}$ allows one to sharpen the inequalities (5) and (6).

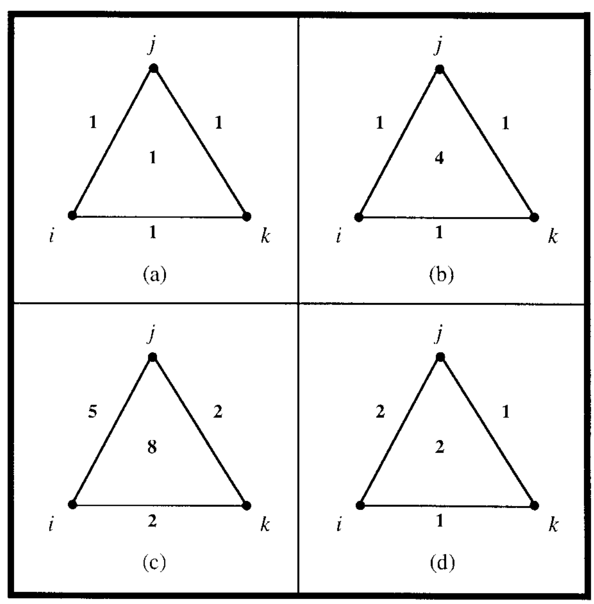


**FIG. 1.** Schematic examples of the relationship between a triadic distance (inscribed in the triangle) and the dyadic distances between each pair of points. A triadic distance must be at least as large as the average dyadic distances (a) and at most four times larger than the average (b); under the triadic model, a direct path from $i$ to $j$ may be slightly longer than the detour via $k$ (c); a Joly–Le Calvé distance which is not a triadic distance (d).

[1] The authors are obliged to H.-J. Bandelt for pointing out an error in a previous version of this lemma, and suggesting its present content and way of proof.

Lemma 2′. *A triadic distance function obeying* $t_{iij} \leqslant t_{ijk}$ (*that is*, $2d_{ij} \leqslant t_{ijk}$) *satisfies*

$$\tfrac{2}{3}(d_{ij}+d_{ik}+d_{jk}) \leqslant t_{ijk} \leqslant \tfrac{4}{3}(d_{ij}+d_{ik}+d_{jk}), \tag{5′}$$

$$d_{ij} \leqslant d_{ik}+d_{jk}, \tag{6′}$$

*for all* $i, j, k \in O$.

The proof is immediate, and a modification of the examples confirms that (5′) and (6′) are sharp. Note that the inequality $t_{iij} \leqslant t_{ijk}$ implies monotonicity of the dissimilarity function when viewed as a function on the ordered system of subsets with 1 to 3 elements. The next proposition gives a way to construct a new triadic distance from a given one.

Proposition 1. *Let c be a strictly positive real number and T a triadic distance function. Then* $T^*$ *defined as* $T^*(i, j, k) = T(i, j, k)/(c+T(i, j, k))$ *is also a triadic distance function.*

*Proof.* Obviously, $T^*$ is a triadic dissimilarity and is definite when $T$ is a triadic distance. Let the quantity $a$ be defined as

$$a=(c+t_{ijk})(c+t_{ikl})(c+t_{jkl})(c+t_{ijl})(t^*_{ikl}+t^*_{jkl}+t_{ijl}-2t^*_{ijk});$$

then $T^*$ satisfies the tetrahedral inequality provided that $a$ is nonnegative. Expanding the above equation into polynomial form, we obtain

$$\begin{aligned} a &= c^3(t_{ikl}+t_{jkl}+t_{ijl}-2t_{ijk}) \\ &\quad + c^2(2t_{ikl}t_{ijl}+2t_{ikl}t_{jkl}+2t_{ijl}t_{jkl}-t_{ijk}t_{ijl}-t_{ijk}t_{ikl}-t_{ijk}t_{jkl}) \\ &\quad + 3c\, t_{ikl}t_{jkl}t_{ijl}+t_{ikl}t_{jkl}t_{ijl}t_{ijk}. \end{aligned}$$

Because the cubic coefficient and the linear terms are nonnegative by the definition of $T$, we only have to verify that the coefficient of $c^2$, denoted as $b$, is nonnegative. We have

$$\begin{aligned} 2b &= t_{ijl}(t_{ikl}+t_{jkl}+t_{ijl}-2t_{ijk})+t_{ikl}(t_{ikl}+t_{jkl}+t_{ijl}-2t_{ijk}) \\ &\quad + t_{jkl}(t_{ikl}+t_{jkl}+t_{ijl}-2t_{ijk}) \\ &\quad + 2t_{ikl}t_{ijl}+2t_{ikl}t_{jkl}+2t_{ijl}t_{jkl}-t^2_{ijl}-t^2_{ikl}-t^2_{jkl}. \end{aligned}$$

Since the first three terms of this expression are nonnegative because $T$ is a triadic distance, it is sufficient to check the sign of the remaining terms. By repeated use of Lemma 1, with the role of $t_{ijk}$ taken by $t_{ijl}$ and $t_{ikl}$, respectively, and rearranging the inequalities, we may write $t^2_{ikl} \geqslant (t_{ijl}-t_{jkl})^2$, $t^2_{jkl} \geqslant (t_{ijl}-t_{ikl})^2$, and $t^2_{ijl} \geqslant (t_{ikl}-t_{jkl})^2$. Adding these, we obtain

$$t^2_{ijl}+t^2_{jkl}+t^2_{ikl} \leqslant 2t_{ijl}t_{jkl}+2t_{ijl}t_{ikl}+t_{ikl}t_{jkl},$$

which proves that $b \geqslant 0$, and hence $a \geqslant 0$. ∎

This result closes our discussion of the general concept of triadic distance function. Let us now turn to two groups of special cases. The first group of models is constructed on the basis of dyadic distances, while the second group is built on the basis of presence–absence variables.

## TRIADIC DISTANCES DEFINED ON DYADIC DISTANCES

In this section we will give several specific examples of classes of functions with the property that all their members satisfy the tetrahedral inequality. To start relatively simple, we first introduce the perimeter model, which will then be generalized to the Minkowski-$p$ or $M_p$ model.

### *Perimeter Model*

This model is a variant of the semi-perimeter model used as an illustration in the Introduction, where we considered dyadic distances among pairs of points in Euclidean space. More generally, given dyadic dissimilarities $\delta_{ij} = \Delta(i, j) = \Delta(j, i)$, the *perimeter dissimilarity* is defined by

$$\tau_{ijk} = \delta_{ij}+\delta_{ik}+\delta_{jk}.$$

The following result establishes the relationship between the triadic metricity of the perimeter model and the metricity of its dyadic constituents.

Proposition 2. *Let D be a mapping of* $O \times O$ *into* $\mathbb{R}^+$, *with* $d_{ij} = D(i, j) = D(j, i)$ *and* $d_{ij} = 0$ *if and only if* $i = j$. *The mapping T defined by* $t_{ijk} = d_{ij}+d_{ik}+d_{jk}$ *is a triadic distance function if and only if D satisfies the triangle inequality.*

*Proof.* It is easily checked that $T$ satisfies (1a)–(1d) and (3a). Inserting the definition of $T$ into the tetrahedral inequality (3b), we must have

$$2d_{il}+2d_{jl}+2d_{kl} \geqslant d_{ij}+d_{ik}+d_{jk}. \tag{7}$$

If we put $k = l$, it follows that $d_{ij} \leqslant d_{ik}+d_{jk}$. Conversely, the inequality (7) follows from adding the three triangle inequalities that are related to each side of the triangle $(i, j, k)$ and the extra point $l$: $d_{il}+d_{jl} \geqslant d_{ij}$, $d_{il}+d_{kl} \geqslant d_{ik}$, and $d_{jl}+d_{kl} \geqslant d_{jk}$. ∎

Note that inequality (7) is an equality if and only if $D$ is a *star* or (*center*) distance (Le Calvé, 1985), with $l$ as the center point. Hence, the tetrahedral inequality (3b) becomes an equality under the perimeter model with star distances.

### *Minkowski-p Model*

Let $\Delta$ be again a dyadic dissimilarity on $O$ and let $T$ be a mapping defined by, for all $1 \leqslant p \leqslant \infty$,

$$\tau_{ijl} = [\delta^p_{ij}+\delta^p_{ik}+\delta^p_{jk}]^{1/p}. \tag{8}$$

This class of triadic dissimilarity functions generalizes the perimeter model, which has $p=1$, and is called the Minkowski-$p$ dissimilarity, or $M_p$ dissimilarity for short.

THEOREM 1. *If $\Delta$ is a dyadic distance D, then the $M_p$ dissimilarity is a triadic distance function, called the $M_p$ distance.*

*Proof.* First suppose $p<\infty$. From the definition of the $M_p$ distance it is evident that it satisfies three-way symmetry (1b) and diagonal-plane equality (1d) and is positive and definite (1a), (1c), and (3a). To show that the tetrahedral inequality (3b) holds, we first note that, adding the base distance to both sides of the triangle equality, we have $2d_{ij} \leqslant d_{ij}+d_{il}+d_{jl}$, $2d_{ik} \leqslant d_{ik}+d_{il}+d_{kl}$, and $2d_{jk} \leqslant d_{jk}+d_{jl}+d_{kl}$. Raising both sides of each of these inequalities to the $p$th power and adding, we obtain

$$2^p(d_{ij}^p+d_{ik}^p+d_{jk}^p) \leqslant (d_{ij}+d_{il}+d_{jl})^p+(d_{ik}+d_{il}+d_{kl})^p + (d_{jk}+d_{jl}+d_{kl})^p,$$

which implies

$$2(d_{ij}^p+d_{ik}^p+d_{jk}^p)^{1/p} \leqslant [(d_{ij}+d_{il}+d_{jl})^p+(d_{ik}+d_{il}+d_{kl})^p + (d_{jk}+d_{jl}+d_{kl})^p]^{1/p}. \tag{9}$$

The left-hand side of inequality (9) equals two times the $M_p$ distance between $i$, $j$, and $k$, and the right-hand side is always smaller than the sum of the other three triadic distances that form the right-hand side of the tetrahedral inequality. That the latter statement is correct follows from Minkowski's inequality, here extended to three terms, which asserts that

$$\left[\sum_k (a_k+b_k+c_k)^p\right]^{1/p} \leqslant \left[\sum_k a_k^p\right]^{1/p}+\left[\sum_k b_k^p\right]^{1/p}+\left[\sum_k c_k^p\right]^{1/p}.$$

With the substitution $a_1=d_{ij}$, $a_2=d_{il}$, $a_3=d_{jl}$, $b_1=d_{il}$, $b_2=d_{ik}$, $b_3=d_{kl}$, $c_1=d_{jl}$, $c_2=d_{kl}$, $c_3=d_{jk}$ inequality (9) and Minkowski's inequality can be combined, showing that the $M_p$ distance satisfies the tetrahedral inequality. For $p=\infty$, the result still holds by the fact that the tetrahedral inequality is preserved by limit considerations. ∎

For $p=\infty$, the $M_p$ model becomes equivalent to $\tau_{ijk}=\max(d_{ij}, d_{ik}, d_{jk})$, called the *maximum model*. Note that the triadic dissimilarity $\tau_{ijk}=\min(d_{ij}, d_{ik}, d_{jk})$ is *not* an $M_p$ distance, nor is it even a triadic distance.

### *Generalized Euclidean Model*

Apart from the perimeter model for $p=1$ and the maximum model for $p=\infty$, another special case of the $M_p$ distance of particular interest is the *generalized Euclidean model* for $p=2$ and with $D$ Euclidean. As an aside, if $D$ is Euclidean, all $M_p$ distance models are invariant under rotations and translations, because in that case they are built up from invariant dyadic parts; so this invariance property is not restricted to the generalized Euclidean model, as suggested by Cox *et al.* (1991). Hayashi (1989) considered what he calls a distance measure $H$ defined by

$$H(i,j,k)=d_{ij}^2+d_{ik}^2+d_{jk}^2,$$

with $D$ Euclidean. Thus $H$ is the *square* of a generalized Euclidean triadic distance. Hayashi dismissed $H$ with the argument that it does not model the three-way information in the data beyond the two-way marginals. It is not hard to show by counterexample that $H$ also does not satisfy the present definition of triadic distance. In fact, it is proportional to the *inertia* of triads of points (Joly and Le Calvé, 1995), defined as the sum of squared distances toward their center of gravity. Because it is a special of the $M_p$ model, the generalized Euclidean distance satisfies the tetrahedral inequality. However, it is not a Joly–Le Calvé distance, since it does not necessarily satisfy the first inequality in (4) (under this model, $t_{iij} \leqslant t_{ijk}$ implies $d_{ij}^2 \leqslant d_{ik}^2+d_{jk}^2$, which can be false, as for example in a triangle with sides of length 3, 3, and 5).

## TRIADIC DISTANCES DEFINED ON PRESENCE–ABSENCE VARIABLES

In this section, we theoretically discuss a class of applications in which a three-way, one-mode table is constructed from a two-mode table with presence–absence data. Our primary interest will be to characterize the entries of the newly constructed table as genuine triadic distances.

### *Daws' Model: Degree of Triadic Distinguishability in Partitions*

As mentioned in the Introduction, Daws (1996) studied dyadic and triadic similarities in the context of the free sorting method of data collection, where $m$ individuals classify $n$ objects into any number of mutually exclusive classes; i.e., the data consists of $m$ partitions. We can say that two objects in different classes are distinguishable, while two objects in the same class are indistinguishable. Let $\rho_{ij}$ be the number of individuals that have put object $i$ and object $j$ into the same class, and $\rho_{ijk}$ be the number of individuals that have put objects $i$, $j$, and $k$ into the same class. Corresponding to $\rho_{ij}$, the degree of dyadic distinguishability is defined as $\delta_{ij}=m-\rho_{ij}$, and corresponding to $\rho_{ijk}$ we have the *degree of triadic distinguishability* $\tau_{ijk}=m-\rho_{ijk}$.

For future reference, and to justify why Daws' model fits into a framework of presence–absence variables, we note that the count $\rho_{ijk}$ may be written as

$$\rho_{ijk} = \sum_r \sum_{s \in L_r} e_{irs} e_{jrs} e_{krs}, \tag{10}$$

where $e_{irs}$ is a presence–absence variable indicating whether individual $r$ has classified object $i$ in class $s$, and where the index $s$ ranges over the index list $L_r$, enumerating the classes of individual $r$. We now show that $m - \rho_{ijk}$ satisfies the model properties of a triadic distance.

THEOREM 2. *Let $\rho_{ijk}$ be the number of times that the triplet $(i, j, k)$ is contained in one of the classes generated by $m$ partitions. Then the mapping $T$ defined as $t_{ijk} = m - \rho_{ijk}$ is a triadic distance function.*

*Proof.* A straightforward proof follows from the observation that[2]

$$m - \rho_{ijk} = \sum_r \max[\delta_{ij}^{(r)}, \delta_{ik}^{(r)}, \delta_{jk}^{(r)}], \tag{11}$$

where the sum extends over all $m$ individuals $r$. For each individual $r$, $\delta_{ij}^{(r)}$ denotes the corresponding characteristic function of the partition associated with $r$. By Theorem 1 (for $p = \infty$), all summands in (11) constitute triadic distances, and so does their sum. ∎

*Similarities and Dissimilarities Derived from Bipartitions*

Consider a data collection procedure resulting in a list of $m$ binary signals for each individual $i$ from $O_1 = \{a_1, a_2, ..., a_n\}$, indicating the presence or absence of each of $m$ attributes from some collection $O_2 = \{b_1, b_2, ..., b_m\}$. Let the data be collected in an $n \times m$ matrix $\mathbf{U}$, with elements $u_{ir}$ $(r = 1, ..., m)$, defined as $u_{ir} = 1$ if attribute $b_r$ is present in individual $a_i$, and $u_{ir} = 0$ if it is absent. Formally, this situation is a special case of Daws' situation, in which the role of individuals and attributes is reversed, and where each partition has two classes, forming a *bipartition*. Since in a partition the classes are mutually exclusive and exhaustive, the variable $e_{irs}$ satisfies $\sum_{s \in L_r} e_{irs} = 1$ $\forall i, r$, and hence if we restrict our attention to bipartitions, we may identify $e_{ir1} = u_{ir}$ and $e_{ir2} = 1 - u_{ir}$. Inserting this change of variables in (10) we obtain

$$\rho_{ijk} = \sum_r u_{ir} u_{jr} u_{kr} + \sum_r (1 - u_{ir})(1 - u_{jr})(1 - u_{kr}), \tag{12}$$

which underlines the fact that positives matches (first term) and negative matches (second term) enter into $\rho_{ijk}$ in an equal way.

[2] This elegant method of proof was pointed out in different forms by several of the referees.

A large number of dyadic (dis)similarity indices exist that aim to measure the resemblance (or the lack of it) between individuals with a differentiation of positive matches, negative matches, and mismatches. Detailed studies of the most important of such indices and their properties may be found in Fichet and Le Calvé (1984) or Gower and Legendre (1986). As convincingly argued by Cox *et al.* (1991), it may very well be that individuals are equi-(dis)similar when compared two at a time, while they are clearly different when compared three at a time. The parallel phenomenon, possible irreducibility of multivariate association between categorical variables, that is, the occurrence of interaction between attributes taken three at a time, has been well known in statistics ever since Barlett (1935), but triadic resemblance lags behind. Our treatment will be such that all of these indices—classically defined on pairs of objects—can be considered as two-way restricted versions of more general indices defined on triples of objects.

Since, typically, the "present" category of an attribute is given a different role compared to the "absent" category, it is convenient to have a somewhat more elaborate notation. We define

$$n_{\sigma_i \sigma_j \sigma_k}^{i\,j\,k} = \sum_r u_{ir}^{\sigma_i}(1 - u_{ir})^{1-\sigma_i} u_{jr}^{\sigma_j}(1 - u_{jr})^{1-\sigma_j} u_{kr}^{\sigma_k}(1 - u_{kr})^{1-\sigma_k}, \tag{13}$$

where $\sigma_i$ is a binary variable that takes value 0 or 1 according to whether attribute $i$ is absent or present, which allows us to code all eight cells of the $2 \times 2 \times 2$ table, while still identifying the selected triad. In this notation, $\rho_{ijk}$ in (12) becomes $n_{111}^{i\,j\,k} + n_{000}^{i\,j\,k}$, where

$n_{111}^{i\,j\,k}$ = number of positive matches between $i$, $j$, and $k$ (*count of attributes present in all three*),

$n_{000}^{i\,j\,k}$ = number of negative matches between $i$, $j$, and $k$ (*count of attributes absent in all three*),

and the degree of triadic distinguishability $m - \rho_{ijk}$ is the number of mismatches between $i$, $j$, and $k$ (count of attributes different in one of three). Extension of (13) to quadruples gives quantities like $n_{1110}^{i\,j\,k\,l}$, the number of attributes shared by $i$, $j$, and $k$, but not by $l$, and $n_{0001}^{i\,j\,k\,l}$, the number of attributes absent in $i$, $j$, and $k$, but present in $l$. Reduction of (13) to dyads gives quantities like $n_{10}^{i\,j}$, the number of attributes present in $i$, but not in $j$. The relationships

$$n_{\sigma_i \sigma_j}^{i\,j} = n_{\sigma_i \sigma_j 0}^{i\,j\,k} + n_{\sigma_i \sigma_j 1}^{i\,j\,k}, \tag{14}$$

$$n_{\sigma_i \sigma_j \sigma_k}^{i\,j\,k} = n_{\sigma_i \sigma_j \sigma_k 0}^{i\,j\,k\,l} + n_{\sigma_i \sigma_j \sigma_k 1}^{i\,j\,k\,l}, \tag{15}$$

$$\rho_{ijk/l} = n_{1110}^{i\,j\,k\,l} + n_{0001}^{i\,j\,k\,l}, \tag{16}$$

are easily verified, where $\rho_{ijk/l}$ denotes the count of quadruples with $(i, j, k)$ in one class and $l$ in the other. It is also important to observe that counts on the diagonal planes satisfy $\sum_r u_{ir} u_{ir} u_{jr} = \sum_r u_{ir} u_{jr}$, that is, the marginals of a given type of count are contained in the higher-order tables. This relationship is shared by all other quantities defined above. With these notational preliminaries, we now turn to a number of specific (dis)similarity measures.

*Triadic Hamming dissimilarity.* The dyadic Hamming dissimilarity is equal to the degree of distinguishability $m - \rho_{ij}$ discussed in the previous section, specialized to bipartitions. Thus, it counts the total number of mismatches across all attributes. Generalized to the triadic case, we obtain by subtraction of the number of matched triples from the maximal count

$$T_{\text{Ham}}(i, j, k) = m - \rho_{ijk} = m - n_{111}^{ijk} - n_{000}^{ijk}. \tag{17}$$

Normed by the number of variables, the corresponding similarity index $(m - T_{\text{Ham}}(i, j, k))/m$ becomes the triadic version of the Sokal–Michener similarity index. Since the $m$ bipartitions considered here are a special case of the $m$ partitions considered in Daws' model, it is an immediate corollary of Theorem 2 that $T_{\text{Ham}}(i, j, k)$ is a triadic distance. Moreover, a stronger result holds.

PROPOSITION 3. *Given the dyadic distance function $D$ defined as*

$$d_{ij} = \sum_r |u_{ir} - u_{jr}|,$$

*the triadic Hamming dissimilarity $T_{\text{Ham}}(i, j, k) = m - n_{111}^{ijk} - n_{000}^{ijk}$ is a semi-perimeter distance.*

*Proof.* We first write $d_{ij}$ in terms of counts as

$$\begin{aligned} d_{ij} &= \sum_r |u_{ir} - u_{jr}| = \sum_r (u_{ir} - u_{jr})^2 \\ &= \sum_r u_{ir}(1 - u_{jr}) + \sum_r u_{jr}(1 - u_{ir}) = n_{10}^{ij} + n_{01}^{ij}. \end{aligned}$$

The assertion would be true if the semi-perimeter distance

$$\begin{aligned} t_{ijk} &= \tfrac{1}{2}(d_{ij} + d_{ik} + d_{jk}) \\ &= \tfrac{1}{2}(n_{10}^{ij} + n_{01}^{ij}) + \tfrac{1}{2}(n_{10}^{ik} + n_{01}^{ik}) + \tfrac{1}{2}(n_{10}^{jk} + n_{01}^{jk}) \end{aligned}$$

is equal to $m - n_{111}^{ijk} - n_{000}^{ijk}$. The latter quantity can be re-expressed as

$$m - n_{111}^{ijk} - n_{000}^{ijk} = n_{100}^{ijk} + n_{010}^{ijk} + n_{001}^{ijk} + n_{110}^{ijk} + n_{101}^{ijk} + n_{011}^{ijk},$$

since the total count over all eight possible patterns must be equal to $m$. Repeated use of (14) shows that these two sums are indeed equal. ∎

*Triadic Rogers–Tanimoto dissimilarity.* The dyadic Rogers–Tanimoto similarity index is $m - (n_{10}^{ij} + n_{01}^{ij})$ relative to $m + (n_{10}^{ij} + n_{01}^{ij})$. Thus if there are no mismatches, it is equal to one, and it becomes equal to zero only if there are neither positive matches nor negative matches. One minus the Rogers–Tanimoto similarity equals

$$\begin{aligned} 1 - [m - (n_{10}^{ij} + n_{01}^{ij})]/[m + (n_{10}^{ij} + n_{01}^{ij})] \\ = 2(n_{10}^{ij} + n_{01}^{ij})/[m + (n_{10}^{ij} + n_{01}^{ij})]; \end{aligned}$$

that is, it can be regarded as a monotone increasing transformation of the dyadic Hamming dissimilarity $(n_{10}^{ij} + n_{01}^{ij})$. For the triadic case, we define

$$T_{\text{R-T}}(i, j, k) = 2T_{\text{Ham}}(i, j, k)/[m + T_{\text{Ham}}(i, j, k)].$$

Since $T_{\text{Ham}}(\cdot)$ is a triadic distance, application of Proposition 1 directly leads to the next result.

THEOREM 3. *The triadic Rogers–Tanimoto dissimilarity is a triadic distance function.*

*Triadic Jaccard dissimilarity.* The dyadic Jaccard similarity index equals the number of positive matches $n_{11}^{ij}$ divided by $m - n_{00}^{ij}$. Again using one minus similarity as the transformation to obtain a dissimilarity, the triadic Jaccard dissimilarity is defined here as

$$T_{\text{Jac}}(i, j, k) = 1 - n_{111}^{ijk}/[m - n_{000}^{ijk}]. \tag{18}$$

Providing that $T_{\text{Jac}}(\cdot)$ possesses all the properties of the triadic distance function is not as direct as in the previous cases.

THEOREM 4. *The triadic Jaccard dissimilarity is a triadic distance function.*

*Proof.* It is evident that $J_{\text{Jac}}(\cdot)$ is nonnegative and satisfies three-way symmetry. We have $\{n_{111}^{ijk} = m - n_{000}^{ijk}\} \Leftrightarrow i = j = k$; and hence $T_{\text{Jac}}(i, j, k) = 0 \Leftrightarrow i = j = k$. The diagonal plane equality (1d) holds, because both $n_{000}^{ijk}$ and $n_{111}^{ijk}$ have this property. To show that the tetrahedral inequality (3b) holds, note that the actual number of partitions can be counted by enumerating the complete set of partitions (or subset patterns), as

$$m = \rho_{ijkl} + \rho_{ijk/l} + \rho_{ikl/j} + \rho_{jkl/i} + \rho_{ijl/k} + \rho_{ij/kl} + \cdots,$$

where $\rho_{ij/kl}$ denotes the count of $(i, j)$ in the same class but separated from $k$ and $l$. Addition of $2\rho_{ijk/l}$ to both sides of this equation, and dropping the relevant nonnegative terms at the right-hand side yields

$$m + 2\rho_{ijk/l} - \rho_{ijkl} - \rho_{ikl/j} - \rho_{jkl/i} - \rho_{ijl/k} \geqslant 3\rho_{ijk/l}.$$

Using the relationships $\rho_{ijk/l} = \rho_{ijk} - \rho_{ijkl}$ and $\rho_{ijk} = m - T_{\text{Ham}}(i, j, k)$ on the left-hand side, and substituting (16) on the right-hand side, we obtain

$$\begin{aligned} T_{\text{Ham}}(i, k, l) + T_{\text{Ham}}(j, k, l) + T_{\text{Ham}}(i, j, l) - 2T_{\text{Ham}}(i, j, k) \\ \geqslant 3n^{ijkl}_{1110} + 3n^{ijkl}_{0001}. \end{aligned} \tag{19}$$

From (18) we deduce the relationship $T_{\text{Ham}}(i, j, k) = [m - n^{ijk}_{000}]\,\tau_{ijk}$, where $T_{\text{Jac}}(i, j, k)$ is denoted by $\tau_{ijk}$. Substituting this equation into inequality (19), using (15), and simplifying, we find

$$\begin{aligned} &(m - n^{ijkl}_{0000})\{\tau_{ikl} + \tau_{jkl} + \tau_{ijl} - 2\tau_{ijk}\} \\ &\quad \geqslant \{3n^{ijkl}_{1110} + n^{ijkl}_{0100}\tau_{ikl} + n^{ijkl}_{1000}\tau_{jkl} + n^{ijkl}_{0010}\tau_{ijl}\} \\ &\quad\quad + n^{ijkl}_{0001}(3 - 2\tau_{ijk}). \end{aligned}$$

Since $(m - n^{ijkl}_{0000}) \geqslant 0$, and $(3 - 2\tau_{ijk}) > 0$ because $\tau_{ijk}$ is bounded above by 1, we conclude that $\tau_{ijk}$ satisfies the tetrahedral inequality. ∎

*Triadic Fichet–Gower dissimilarity.* Consider a family of indices that includes a positive parameter $\theta$ modifying the contribution of the number of positive matches in the triadic index:

$$T_{\text{F–G}}(i, j, k \mid \theta) = T_{\text{Ham}}(i, j, k)/[\theta n^{ijk}_{111} + T_{\text{Ham}}(i, j, k)]. \tag{20}$$

Metric properties of the dyadic version of the family defined in (20) were discussed by Fichet (1986) and Gower (1986); hence the name coined here. The Fichet–Gower dissimilarity generalizes the indices of Sokal & Sneath and Anderberg ($\theta = 1/2$), Jaccard ($\theta = 1$), and Czekanowski & Dice ($\theta = 2$). To prove under what conditions $T_{\text{F–G}}(i, j, k \mid \theta)$ is a triadic distance, it turns out to be useful to work with a reparametrization.

THEOREM 5. *The triadic Fichet–Gower dissimilarity is a triadic distance if and only if $\theta \leqslant 1$.*

*Proof.* For $\theta = 1$, the triadic Jaccard dissimilarity (18) is obtained, as can be seen by inserting (17) into (20) and simplifying. So $T_{\text{F–G}}(i, j, k \mid \theta = 1)$ is a triadic distance by Theorem 4. For $0 < \theta < 1$, we reparametrize by choosing a strictly positive real number $\alpha$ so that $\theta = \alpha/(\alpha + 1)$. Upon substitution in (20), and dividing both numerator and denominator by $n^{ijk}_{111} + T_{\text{Ham}}(i, j, k)$,

$$\begin{aligned} &T_{\text{F–G}}(i, j, k \mid \theta) \\ &\quad = (\alpha + 1)\, T_{\text{Ham}}(i, j, k)/[\alpha n^{ijk}_{111} + (\alpha + 1)\, T_{\text{Ham}}(i, j, k)] \\ &\quad = (\alpha + 1)\, T_{\text{F–G}}(i, j, k \mid \theta = 1)/[\alpha + T_{\text{F–G}}(i, j, k \mid \theta = 1)]. \end{aligned}$$

Since $T_{\text{F–G}}(i, j, k \mid \theta = 1)$ is a triadic distance, the result follows from Proposition 1. ∎

It should be noted that, for $\theta > 1$, $T_{\text{F–G}}(i, j, k \mid \theta)$ in general is not necessarily a triadic distance. For example, with $\theta = 2$ and 4 individuals with attribute patterns $a_1 = (1\,1\,0\,1)$, $a_2 = (0\,1\,1\,1)$, $a_3 = (1\,0\,1\,1)$, and $a_4 = (1\,1\,1\,1)$, and writing $\tau_{ijk}$ for short, we find $\tau_{123} = 3/5$, $\tau_{124} = 1/3$, $\tau_{134} = 1/3$, and $\tau_{234} = 1/3$. So we have $2\tau_{123} = 6/5 > 1/3 + 1/3 + 1/3$.

*Triadic Russel–Rao dissimilarity.* The dyadic Russel–Rao similarity index is the number of positive matches divided by the number of attributes; thus for $i \neq j \neq k$, one minus the triadic version of this index is

$$T_{\text{R–R}}(i, j, k) = (m - n^{ijk}_{111})/m,$$

where it should be noted that the definition must include $T_{\text{R–R}}(i, j, k) = 0$ if $i = j = k$ and if $i \neq j = k$ (note that this proviso was not necessary in the previous cases).

THEOREM 6. *The triadic Russel–Rao dissimilarity is a triadic distance function.*

*Proof.* It will only be proven that the tetrahedral inequality holds. Writing $\tau_{ijk}$ for $T_{\text{R–R}}(i, j, k)$, and using Eq. (15), we see that

$$\begin{aligned} &m\{\tau_{ikl} + \tau_{jkl} + \tau_{ijl} - 2\tau_{ijk}\} \\ &\quad = (m - n^{ikl}_{111}) + (m - n^{jkl}_{111}) + (m - n^{ijl}_{111}) - 2(m - n^{ijk}_{111}) \\ &\quad = m - (n^{ijkl}_{1111} + n^{ijkl}_{1011} + n^{ijkl}_{0111} + n^{ijkl}_{1101}) + 2n^{ijkl}_{1110}. \end{aligned}$$

The right-hand side is nonnegative, because the term in brackets cannot be larger than $m$. ∎

## APPROXIMATE REPRESENTATIONS OF TRIADIC DISSIMILARITIES

In a broad sense of the term, multidimensional scaling is concerned with the problem of finding an embedding or representation of the set $O$ as $n$ points in some prechosen metric space, in such a way that the associated distances $D$ approximate the given dissimilarities $\Delta$, where the quality of the approximation is measured by a badness-of-fit function (De Leeuw and Heiser, 1982). The representation is a Euclidean space if the chosen distance is Euclidean, a hierarchical tree if it is ultrametric, or an additive tree if it is quadrangular (an additive tree distance). Since in general the dissimilarities cannot be expected to be exactly representable by any of these type of distances, approximate representations are obtained by minimizing the badness-of-fit function.

In this section we propose multidimensional scaling methods for triadic dissimilarity data denoted as $\{\tau_{ijk}\}$. The focus will be on two special cases of the Minkowski-$p$ model: the perimeter model and the generalized Euclidean model. For the perimeter model, we will first discuss the most general case in which the constituent distances that make up the perimeter merely satisfy the metric axioms and then introduce the additional constraint that the constituent dyadic distances be Euclidean. While the general perimeter model is coordinate free, both the Euclidean perimeter model and the generalized Euclidean model represent the elements of $O$ by $n$ points in a space of given dimensionality $q$, with coordinates collected in the $n \times q$ matrix $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_i, ..., \mathbf{x}_n)'$. The ordinary (dyadic) Euclidean distance is defined as

$$d_{ij}^2(\mathbf{X}) = \operatorname{tr} \mathbf{X}'\mathbf{E}_{ij}\mathbf{X}, \tag{21}$$

with the matrix $\mathbf{E}_{ij} = (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)'$, in which $\mathbf{e}_i$ is the $i$th column of the identity matrix. Since $\mathbf{e}'\mathbf{E}_{ij} = \mathbf{0}$, with $\mathbf{e}$ the unit element vector and $\mathbf{0}$ the vector of zeroes, we may assume without loss of generality that the matrix $\mathbf{X}$ is centered, i.e., $\mathbf{e}'\mathbf{X} = \mathbf{0}'$. A triadic distance is called Euclidean if it can be written as a function of $d_{ij}(\mathbf{X})$, $d_{ik}(\mathbf{X})$, and $d_{jk}(\mathbf{X})$; this is expressed in the notation by writing it as $t_{ijk}(\mathbf{X})$. Least squares triadic multidimensional scaling finds a representation $\mathbf{X}$ by minimizing the (weighted) sum of squared deviations of $t_{ijk}(\mathbf{X})$ with respect to $\tau_{ijk}$. Weights $w_{ijk}$ will be introduced in the badness-of-fit functions for greater generality. They are assumed to be given in advance and can be used to accommodate missing data (by setting $w_{ijk} = 0$ if $\tau_{ijk}$ is missing), to control the influence of the residuals as a function of the estimated standard error of each dissimilarity, or to mimic the behavior of other badness-of-fit functions (Heiser, 1988). It is also assumed that the weights have all the properties of triadic dissimilarities, that is, nonnegativity (1a), three-way symmetry (1b), minimality (1c), and diagonal-plane equality (1d).

*Approximation with the General Perimeter Model*

The general perimeter model accounts for triadic dissimilarity data by the sum of dyadic distances merely satisfying the metric axioms. Thus we are interested in the following optimization problem: given a set of triadic dissimilarities $\{\tau_{ijk}\}$, find a set of dyadic distances $D = \{d_{ijk}\}$ that minimizes the least squares criterion

$$\sigma_A^2(D) = \sum_{(i,\,j,\,k) \in L} w_{ijk}(\tau_{ijk} - d_{ij} - d_{ik} - d_{jk})^2. \tag{22}$$

The summation in (22) is over the index list $L$, containing the $\binom{n}{3}$ off-diagonal triplets $(i, j, k)$ with $i < j < k$ and the $\binom{n}{2}$ diagonal-plane triplets $(i, i, j)$ with $i < j$. This list is sufficient if $\{\tau_{ijk}\}$ is a proper set of triadic dissimilarities. For suppose that the latter is not the case; then a simplification is possible, as shown by Proposition 4, where the weights are omitted since they are not essential.

Proposition 4. *Let $\{\delta_{ijk}\}$ denote a general three-way array of dissimilarities that only satisfy nonnegativity. The least squares loss function defined as*

$$\sigma^2(D) = \tfrac{1}{6} \sum_i \sum_j \sum_k (\delta_{ijk} - d_{ij} - d_{ik} - d_{jk})^2,$$

*with summation over the whole range of i, j, and k, can be decomposed into these six components*:

$$\begin{aligned}
\sigma^2(D) = {} & \tfrac{1}{6} \sum_{i=j=k} \delta_{ijk}^2 \qquad (\textit{lack of minimality}) \\
& + \tfrac{1}{6} \sum_{i<j<k} \sum_{\pi} (\delta_{\pi(i)\,\pi(j)\,\pi(k)} - \hat{\tau}_{ijk})^2 \\
& \quad (\textit{lack of three-way symmetry}) \\
& + \tfrac{1}{6} \sum_{i \neq j} \sum_{\kappa} (\delta_{i\kappa(i)\,\kappa(j)} - \hat{\tau}_{iij})^2 \\
& \quad (\textit{lack of diagonal-plane equality}) \\
& + \tfrac{1}{4} \sum_{i<j} (\hat{\tau}_{iij} - \hat{\tau}_{jji})^2 \\
& \quad (\textit{lack of symmetry within diagonal planes}) \\
& + \sum_{i<j<k} (\hat{\tau}_{ijk} - d_{ij} - d_{ik} - d_{jk})^2 \\
& \quad (\textit{off-diagonal loss}) \\
& \sum_{i<j} (\hat{\delta}_{ij} - 2d_{ij})^2. \\
& \quad (\textit{diagonal-plane loss})
\end{aligned}$$

*Here $\pi$ indexes the six permutations of $(i, j, k)$, $\kappa$ indexes the three combinations $(i, j)$, $(j, i)$ and $(j, j)$, and the quantities with a hat are defined as*

$$\hat{\tau}_{ijk} = \tfrac{1}{6} \sum_{\pi} \delta_{\pi(i)\,\pi(j)\,\pi(k)},$$

$$\hat{\tau}_{iij} = \tfrac{1}{3} \sum_{\kappa} \delta_{i\kappa(i)\,\kappa(j)},$$

$$\hat{\delta}_{ij} = \tfrac{1}{2}(\hat{\tau}_{iij} + \hat{\tau}_{jji}).$$

*Proof.* Only the main ingredient of the proof will be given. The decomposition follows from repeated application of the Cartesian decomposition (Halmos, 1958, p. 136) on the residuals in $\sigma^2(D)$. Given an asymmetric pair $(a_{ij}, a_{ji})$ for which $a_{ij} \neq a_{ji}$, the Cartesian decomposition is the unique representation of the form $a_{ij} = b_{ij} + c_{ij}$, where $b_{ij}$ is symmetric ($b_{ij} = b_{ji}$) and $c_{ij}$ is skew symmetric ($c_{ij} = -c_{ji}$). The symmetric component is $b_{ij} = \frac{1}{2}(a_{ij} + a_{ji})$, the skew-symmetric component is $c_{ij} = \frac{1}{2}(a_{ij} - a_{ji})$, and they have the property $\frac{1}{2}(a_{ij}^2 + a_{ji}^2) = b_{ij}^2 + c_{ij}^2$. ∎

Proposition 4 shows that only the off-diagonal loss and the diagonal-plane loss are dependent upon the dyadic distances, and therefore the summation in (22) only needs to involve triplets in the list $L$, as indicated. The quantities $\{\hat{\tau}_{ijk}\}$, $\{\hat{\tau}_{iij}\}$, and $\{\hat{\delta}_{ij}\}$ are the closest approximations of the relevant part of $\{\delta_{ijk}\}$ that satisfy three-way symmetry, diagonal-plane equality, and two-way symmetry, respectively, in the least squares sense. Several pairs of quantities in the decomposition are orthogonal: the off-diagonal residuals $(\delta_{\pi(i)\,\pi(j)\,\pi(k)} - \hat{\tau}_{ijk})$ with $(\hat{\tau}_{ijk} - d_{ij} - d_{ik} - d_{jk})$, the diagonal-plane residuals $(\delta_{i\kappa(i)\,\kappa(j)} - \hat{\tau}_{iij})$ with the asymmetric averages $\hat{\tau}_{iij}$, and the skew-symmetric residuals $(\hat{\tau}_{iij} - \hat{\tau}_{jji})$ with the symmetric residuals $(\hat{\delta}_{ij} - 2d_{ij})$.

The minimization of (22) is a quadratic problem in $\mathbf{d}$ of the type $\min(\mathbf{t} - \mathbf{Ad})'(\mathbf{t} - \mathbf{Ad})$, under linear inequality constraints. It may be shown that $\mathbf{A}'\mathbf{A}$ is positive definite, so there is a unique solution. The problem may be solved by Uzawa's method (see Ciarlet, 1989). Dykstra's (1983) algorithm is efficient for a constraint set defined by closed half-spaces, so that it also applies.

*Approximation with the Euclidean Perimeter Model*

In the least squares function $\sigma_A^2(D)$ defined in (22) the model has the form $t_{ijk} = d_{ij} + d_{ik} + d_{jk}$. The Euclidean perimeter model poses the restriction $t_{ijk}(\mathbf{X}) = d_{ij}(\mathbf{X}) + d_{ik}(\mathbf{X}) + d_{jk}(\mathbf{X})$, so the problem becomes one of minimizing

$$\sigma_B^2(\mathbf{X}) = \sum_{(i,j,k)\in L} w_{ijk}(\tau_{ijk} - d_{ij}(\mathbf{X}) - d_{ik}(\mathbf{X}) - d_{jk}(\mathbf{X}))^2. \tag{23}$$

As before, the minimization only needs to regard the off-diagonal loss and the diagonal-plane loss, involving the residuals listed in the list $L$, containing off-diagonal triplets $(i, j, k)$ with $i < j < k$ and diagonal-plane triplets $(i, i, j)$ with $i < j$, since Proposition 4 applies regardless of the Euclidean restriction. Developing expression (23), we obtain

$$\sigma_B^2(\mathbf{X}) = SSQ_\tau + \alpha(\mathbf{X}) + \gamma(\mathbf{X}) - 2\beta(\mathbf{X}),$$

where

$$SSQ_\tau = \sum_{i<j<k} w_{ijk}\tau_{ijk}^2 + \sum_{i<j} w_{iij}\tau_{iij}^2,$$

$$\alpha(\mathbf{X}) = \sum_{i<j} w_{ij\cdot}\, d_{ij}^2(\mathbf{X}),$$

$$\beta(\mathbf{X}) = \sum_{i<j} \delta_{ij\cdot}\, d_{ij}(\mathbf{X}),$$

$$\gamma(\mathbf{X}) = \sum_{i<j} \sum_k w_{ijk} d_{ij}(\mathbf{X})[d_{ik}(\mathbf{X}) + d_{jk}(\mathbf{X})],$$

with the dyadic quantities $w_{ij\cdot}$ and $\delta_{ij\cdot}$ defined as

$$w_{ij\cdot} = \sum_k w_{ijk},$$

$$\delta_{ij\cdot} = \sum_k w_{ijk}\tau_{ijk}.$$

The components $\alpha(\mathbf{X})$, $\beta(\mathbf{X})$, and $\gamma(\mathbf{X})$ are all convex functions of $\mathbf{X}$, since $d_{ij}(\mathbf{X})$ is convex in $\mathbf{X}$, and nonnegative linear combinations of convex functions again satisfy convexity. Thus $\sigma_B^2(\cdot)$ is the difference between two convex functions, and can be minimized by the iterative majorization (IM) approach. This computational strategy was developed in the ordinary multidimensional scaling context by De Leeuw and Heiser (1980) under the same SMACOF, and since then it has been successfully applied in several other approximation and equilibrium problems (Heiser, 1995). To characterize the stationary points of $\sigma_B^2(\cdot)$, and to show how IM can be applied to find them, it is convenient to express $\sigma_B^2(\cdot)$ in matrix form. The following lemmas are stated without proof.

LEMMA 3. *Let* $\mathbf{V}$ *be defined as the order-n symmetric matrix with elements*

$$v_{ij} = -w_{ij\cdot} \qquad \text{if } i \neq j,$$

$$v_{ii} = \sum_{l\neq i} w_{il\cdot}\,.$$

*Then* $\alpha(\mathbf{X}) = \operatorname{tr} \mathbf{X}'\mathbf{V}\mathbf{X}$ *and the matrix of partial derivatives is* $\nabla\alpha(\mathbf{X}) = 2\mathbf{V}\mathbf{X}$.

LEMMA 4. *Let* $\mathbf{B} = B(\mathbf{X})$ *be the order-n symmetric matrix-valued function defined as*

$$\begin{aligned} b_{ij}(\mathbf{X}) &= 0 && \text{if } i \neq j \text{ and } d_{ij}(\mathbf{X}) = 0 \\ b_{ij}(\mathbf{X}) &= -\delta_{ij\cdot}/d_{ij}(\mathbf{X}) && \text{if } i \neq j \text{ and } d_{ij}(\mathbf{X}) \neq 0 \\ b_{ii}(\mathbf{X}) &= -\boldsymbol{\Sigma}_{l\neq i}\, b_{il}(\mathbf{X}). \end{aligned}$$

*Then* $\beta(\mathbf{X}) = \operatorname{tr} \mathbf{X}'B(\mathbf{X})\,\mathbf{X}$ *and the matrix of partial derivatives is* $\nabla\beta(\mathbf{X}) = B(\mathbf{X})\,\mathbf{X}$.

LEMMA 5. *Let* $\mathbf{C} = C(\mathbf{X})$ *be the order-n symmetric matrix-valued function defined as*

$$c_{ij}(\mathbf{X})=0 \qquad \text{if} \quad i\neq j \text{ and } d_{ij}(\mathbf{X})=0$$

$$c_{ij}(\mathbf{X})=-\sum_k w_{ijk}[d_{ik}(\mathbf{X})+d_{jk}(\mathbf{X})]/d_{ij}(\mathbf{X}) \qquad \text{if} \quad i\neq j \text{ and } d_{ij}(\mathbf{X})\neq 0$$

$$c_{ii}(\mathbf{X})=-\sum_{l\neq i} c_{il}(\mathbf{X}).$$

*Then* $\gamma(\mathbf{X})=\operatorname{tr}\mathbf{X}'C(\mathbf{X})\,\mathbf{X}$ *and the matrix of partial derivatives is* $\nabla\gamma(\mathbf{X}=2C(\mathbf{X})\,\mathbf{X}$.

Using Lemmas 3–5, and defining the matrix $A(\mathbf{X})=\mathbf{V}+C(\mathbf{X})$, we can re-express $\sigma_B^2(\cdot)$ as

$$\sigma_B^2(\mathbf{X})=SSQ_\tau+\operatorname{tr}\mathbf{X}'A(\mathbf{X})\,\mathbf{X}-2\operatorname{tr}\mathbf{X}'B(\mathbf{X})\,\mathbf{X}. \tag{24}$$

From setting the partial derivatives equal to zero it follows that a stationary point $\hat{\mathbf{X}}$ must satisfy

$$A(\hat{\mathbf{X}})\,\hat{\mathbf{X}}=B(\hat{\mathbf{X}})\,\hat{\mathbf{X}}. \tag{25}$$

The stationary equation in (25) implies that $\operatorname{tr}\hat{\mathbf{X}}'A(\hat{\mathbf{X}})\,\hat{\mathbf{X}}=\operatorname{tr}\hat{\mathbf{X}}'B(\hat{\mathbf{X}})\,\hat{\mathbf{X}}$, and hence that $\sigma_B^2(\hat{\mathbf{X}})=SSQ_\tau-\operatorname{tr}\hat{\mathbf{X}}'A(\hat{\mathbf{X}})\,\hat{\mathbf{X}}$. Recalling that $\operatorname{tr}\mathbf{X}'A(\mathbf{X})\,\mathbf{X}=\alpha(\mathbf{X})+\gamma(\mathbf{X})$ contains the two parts of the total sum of squares of the triadic distances $t_{ijk}(\mathbf{X})$, we arrive at the result stated in the next proposition.

PROPOSITION 5. *Given a set of triadic dissimilarities, suppose that the configuration* $\hat{\mathbf{X}}$ *satisfies the nonlinear equation* $\mathbf{V}\hat{\mathbf{X}}+C(\hat{\mathbf{X}})\,\hat{\mathbf{X}}=B(\hat{\mathbf{X}})\,\hat{\mathbf{X}}$ *for the Euclidean perimeter model, where the matrices* $\mathbf{V}$, $C(\cdot)$, *and* $B(\cdot)$ *are as defined in Lemmas* 3, 5, *and* 4, *respectively. Then the total weighted sum of squares of the triadic dissimilarities* $SSQ_\tau$ *can be decomposed as*

$$SSQ_\tau=\sum_{(i,j,k)\in L} w_{ijk}\tau_{ijk}^2=\sigma_B^2(\hat{\mathbf{X}})+\sum_{(i,j,k)\in L} w_{ijk}t_{ijk}^2(\hat{\mathbf{X}}).$$

Proposition 5 shows a basic optimality property of weighted least squares approximation with the Euclidean perimeter model. It allows evaluation of the stationary point, by division of both terms with $SSQ_\tau$, in terms of a *relative loss* and a percentage of *dispersion accounted for* (%DAF). The latter quantity is analogous to the diagnostic *variance accounted for* in regression analysis. Thus, the relative loss is $100\times\sigma_B^2(\hat{\mathbf{X}})/SSQ_\tau$, and the %DAF is $100\times\operatorname{tr}\hat{\mathbf{X}}'A(\hat{\mathbf{X}})\,\hat{\mathbf{X}}/SSQ_\tau$.

To arrive at an iterative algorithm for solving the nonlinear equation (25), it is useful to know some properties of the matrix $A(\mathbf{X})$. Using the same arguments as in De Leeuw and Heiser (1980), it can be verified that this matrix is positive semi-definite and has rank $n-1$ and that the eigenvector corresponding to the vanishing eigenvalue is the unit vector $\mathbf{e}$. From these properties it follows that $A(\mathbf{X})$ has generalized inverse $A^+(\mathbf{X})=[A(\mathbf{X})+\mathbf{ee}'/\mathbf{e}'\mathbf{e}]^{-1}-\mathbf{ee}'/\mathbf{e}'\mathbf{e}$, which can be easily shown to satisfy the Moore–Penrose conditions. The basic IM algorithm for this case is to build up a sequence of configurations $\mathbf{X}_0, \mathbf{X}_1, ..., \mathbf{X}_s$ by applying the mapping

$$\mathbf{X}_{s+1}=A^+(\mathbf{X}_s)\,B(\mathbf{X}_s)\,\mathbf{X}_s=G_B(\mathbf{X}_s). \tag{26}$$

A fixed point $\mathbf{X}_*=G_B(\mathbf{X}_*)$ of the mapping $G_B(\cdot)$ defined in (26) will solve (25), and hence it will be a stationary point of $\sigma_B^2(\cdot)$. As in all fixed-point algorithms, the initial configuration $\mathbf{X}_0$ could be selected arbitrarily, but it is preferable to start the process close to a local minimum, e.g., by ordinary MDS of the diagonal plane dissimilarities $\tau_{iij}$. Not only does a fixed point exist, but IM guarantees that the sequence defined by $G_B(\cdot)$ is monotonically convergent.

THEOREM 7. *The sequence* $\{\sigma_B^2(\mathbf{X}_s)\,|\,s=0,\ 1,\ 2,\ \ldots\}$ *decreases monotonically and converges.*

*Proof.* The proof of this theorem is based upon two inequalities:

$$\beta(\mathbf{X})\geqslant\operatorname{tr}\mathbf{X}'B(\mathbf{Y})\,\mathbf{Y} \quad \forall\mathbf{X},\mathbf{Y}; \tag{27}$$

$$\gamma(\mathbf{X})\leqslant\operatorname{tr}\mathbf{X}'C(\mathbf{Y})\,\mathbf{X} \quad \forall\mathbf{X},\mathbf{Y}. \tag{28}$$

Suppose that these inequalities do hold (this will be verified in a short while), and consider the family of quadratic functions $\mu_B(\mathbf{X}\,|\,\mathbf{Y})$, indexed by $\mathbf{Y}$ and defined as

$$\mu_B(\mathbf{X}\,|\,\mathbf{Y})=SSQ_\tau+\operatorname{tr}\mathbf{X}'\mathbf{V}\mathbf{X}+\operatorname{tr}\mathbf{X}'C(\mathbf{Y})\,\mathbf{X} -2\operatorname{tr}\mathbf{X}'B(\mathbf{Y})\,\mathbf{Y}. \tag{29}$$

It follows from (27) and (28) that each member of the family (29) satisfies, $\forall\mathbf{X}$,

$$\sigma_B^2(\mathbf{X})\leqslant\mu_B(\mathbf{X}\,|\,\mathbf{Y}),$$
$$\sigma_B^2(\mathbf{Y})=\mu_B(\mathbf{Y}\,|\,\mathbf{Y}),$$
$$\nabla\sigma_B^2(\mathbf{Y})=\nabla\mu_B(\mathbf{Y}\,|\,\mathbf{Y});$$

that is, $\mu_B(\cdot\,|\,\mathbf{Y})$ *majorizes* the loss function, coincides with it at $\mathbf{Y}$, and has the same partial derivatives at the point of coincidence. From setting $\mathbf{Y}=\mathbf{X}_s$ it appears that the mapping $G_B(\cdot)$ in (26) defines $\mathbf{X}_{s+1}$ as the argument that minimizes the majorizing function $\mu_B(\cdot\,|\,\mathbf{X}_s)$, and therefore

$$\sigma_B^2(\mathbf{X}_{s+1})\leqslant\mu_B(\mathbf{X}_{s+1}\,|\,\mathbf{X}_s)=\min\mu_B(\mathbf{X}\,|\,\mathbf{X}_s). \tag{30}$$

But since $\mathbf{X}_{s+1}$ is a minimizer, we also have

$$\mu_B(\mathbf{X}_{s+1}\,|\,\mathbf{X}_s)\leqslant\mu_B(\mathbf{X}_s\,|\,\mathbf{X}_s)=\sigma_B^2(\mathbf{X}_s), \tag{31}$$

with equality if and only if $\mathbf{X}_{s+1}=\mathbf{X}_s$, in which case we may conclude that we have found a configuration satisfying $\nabla\sigma_B^2(\mathbf{X}_s)=\nabla\mu_B(\mathbf{X}_s\,|\,\mathbf{X}_s)=0$, that is, a stationary point, and we stop. It is clear from combining (30) and (31) that the sequence is monotonically decreasing; because it is also bounded below by zero, it must converge.

To finish the proof, we have to verify (27) and (28); the first of these follows from the Cauchy–Schwarz inequality written in the form $d_{ij}(\mathbf{X})\, d_{ij}(\mathbf{Y} \geqslant \operatorname{tr} \mathbf{X}'\mathbf{E}_{ij}\mathbf{Y}$ (De Leeuw and Heiser, 1977). Both sides of this inequality can be divided by $d_{ij}(\mathbf{Y})$ if it is nonzero, while if $d_{ij}(\mathbf{Y})=0$, the inequality is replaced by $d_{ij}(\mathbf{X})\geqslant 0$; then the result is obtained by multiplying both sides with $\delta_{ij\cdot}$ and using the definition of $\beta(\mathbf{X})$ in Lemma 4. To derive expression (28) we expand, following an idea first used in Heiser (1987, 1991) and extended in Groenen and Heiser (1991), the inequality

$$[d_{ij}(\mathbf{X})\, d_{ik}(\mathbf{Y})-d_{ij}(\mathbf{Y})\, d_{ik}(\mathbf{X})]^2\geqslant 0,$$

which is an equality if $\mathbf{X}=\mathbf{Y}$. Using (21) and assuming that $d_{ij}(\mathbf{Y})\neq 0$ and $d_{ik}(\mathbf{Y})\neq 0$, we have

$$\begin{aligned} d_{ij}(\mathbf{X})\, d_{ik}(\mathbf{X}) \leqslant \tfrac{1}{2} \operatorname{tr} \mathbf{X}'[\{d_{ik}(\mathbf{Y})/d_{ij}(\mathbf{Y})\}\ \mathbf{E}_{ij} \\ +\{d_{ij}(\mathbf{Y})/d_{ik}(\mathbf{Y})\}\ \mathbf{E}_{ik}]\ \mathbf{X} \end{aligned}$$

(whenever $d_{ij}(\mathbf{Y})<\varepsilon$, it is replaced by $\varepsilon$). Addition of a similar expression for $d_{ij}(\mathbf{X})\, d_{jk}(\mathbf{X})$ yields

$$\begin{aligned} & d_{ij}(\mathbf{X})[d_{ik}(\mathbf{X})+d_{jk}(\mathbf{X})] \\ &\quad \leqslant \tfrac{1}{2} \operatorname{tr} \mathbf{X}'[\{[d_{ik}(\mathbf{Y})+d_{jk}(\mathbf{Y})]/d_{ij}(\mathbf{Y})\}\ \mathbf{E}_{ij}]\ \mathbf{X} \\ &\quad\quad +\tfrac{1}{2} \operatorname{tr} \mathbf{X}'[\{d_{ij}(\mathbf{Y})/d_{ik}(\mathbf{Y})\}\ \mathbf{E}_{ik} \\ &\quad\quad +\{d_{ij}(\mathbf{Y})/d_{jk}(\mathbf{Y})\}\ \mathbf{E}_{jk}]\ \mathbf{X}. \end{aligned} \tag{32}$$

After multiplying (32) with $w_{ijk}$, and summing over $i<j$ and all $k$, we obtain two equal terms on the right-hand side of the summation, because, due to the symmetry of all quantities involved,

$$\begin{aligned} & \sum_{i<j}\sum_{k} w_{ijk}\,\{[d_{ik}(\mathbf{Y})+d_{jk}(\mathbf{Y})]/d_{ij}(\mathbf{Y})\}\ \mathbf{E}_{ij} \\ &\quad =\sum_{i<j}\sum_{k} w_{ijk}\{d_{ij}(\mathbf{Y})/d_{ik}(\mathbf{Y})\}\ \mathbf{E}_{ik} \\ &\quad\quad +\sum_{i<j}\sum_{k} w_{ijk}\{d_{ij}(\mathbf{Y})/d_{jk}(\mathbf{Y})\}\ \mathbf{E}_{jk}. \end{aligned}$$

Thus both terms are equal to $\tfrac{1}{2}\operatorname{tr}\mathbf{X}'C(\mathbf{Y})\,\mathbf{X}$, which establishes the correctness of (28). ∎

Theorem 7 shows an important characteristic of the IM algorithm based on (26), but it should be noted that it does not guarantee convergence of the sequence of configurations itself (we would need to establish properties of $\|\mathbf{X}_{s+1}-\mathbf{X}_s\|$ for this purpose, which would lead us outside the scope of this paper). The major difference with the standard SMACOF algorithm is the fact that the matrix $A^+(\mathbf{X}_s)$ is variable, rather than fixed, during the iterations, as it depends on the previous update $\mathbf{X}_s$. Thus $A(\mathbf{X}_s)$ must be inverted each iteration, which burdens the computations; note, however, that we have a three-way problem that can be solved by a two-way process.

### *Approximation with the $M_2$ Model*

The generalized Euclidean model, obtained by setting $p=2$ in the Minkowski-$p$ model (8), poses the restriction $t_{ijk}(\mathbf{X})=[d_{ij}^2(\mathbf{X})+d_{ik}^2(\mathbf{X})+d_{jk}^2(\mathbf{X})]^{1/2}$. So, due to the square root, and in contrast to the perimeter model, the $M_2$ model is not additive in the dyadic distances. Defining $\mathbf{E}_{ijk}=\mathbf{E}_{ij}+\mathbf{E}_{ik}+\mathbf{E}_{jk}$, we can write the square of the $M_2$ distance as

$$t_{ijk}^2(\mathbf{X})=\operatorname{tr}\mathbf{X}'\mathbf{E}_{ij}\mathbf{X}+\operatorname{tr}\mathbf{X}'\mathbf{E}_{ik}\mathbf{X}+\operatorname{tr}\mathbf{X}'\mathbf{E}_{jk}\mathbf{X}=\operatorname{tr}\mathbf{X}'\mathbf{E}_{ijk}\mathbf{X}. \tag{33}$$

The approximation problem becomes one of minimizing

$$\sigma_C^2(\mathbf{X})=\sum_{(i,j,k)\in L} w_{ijk}(\tau_{ijk}-[\operatorname{tr}\mathbf{X}'\mathbf{E}_{ijk}\mathbf{X}]^{1/2})^2 \tag{34}$$

over $\mathbf{X}$, with $\mathbf{E}_{ijk}$ having the required additive form, and under the same symmetry assumptions as before. Developing expression (34), we obtain

$$\sigma_C^2(\mathbf{X})=SSQ_\tau+\alpha(\mathbf{X})-2\phi(\mathbf{X}),$$

where $SSQ_\tau$ and $\alpha(\mathbf{X})$ are as defined earlier, and the cross-product term $\phi(\mathbf{X})$ becomes

$$\phi(\mathbf{X})=\left\{\tfrac{1}{3}\sum_{i<j}\sum_{k\neq i,j} w_{ijk}\tau_{ijk}t_{ijk}(\mathbf{X})\right\}+\sum_{i<j} w_{iij}\tau_{iij}t_{iij}(\mathbf{X}).$$

Since $t_{ijk}(\mathbf{X})$, and hence $\phi(\mathbf{X})$, is convex in $\mathbf{X}$, we still have a difference between two convex functions. The matrix form of $\phi(\cdot)$ is again stated in a lemma.

LEMMA 6. *Let the quantities $h_{ijk}(\mathbf{X})$ be defined as*

$$\begin{aligned} h_{ijk}(\mathbf{X})&=0 && \text{if } t_{ijk}(\mathbf{X})=0, \\ h_{ijk}(\mathbf{X})&=w_{ijk}\tau_{ijk}/t_{ijk}(\mathbf{X}) && \text{if } t_{ijk}(\mathbf{X})\neq 0, \end{aligned}$$

*and let $\mathbf{F}=F(\mathbf{X})$ be the order-$n$ symmetric matrix-valued function defined as*

$$\begin{aligned} f_{ij}(\mathbf{X})&=-\sum_{k} h_{ijk}(\mathbf{X}) \quad \text{if } i\neq j, \\ f_{ii}(\mathbf{X})&=-\sum_{l\neq i} f_{il}(\mathbf{X}). \end{aligned}$$

*Then* $\phi(\mathbf{X}) = \operatorname{tr} \mathbf{X}' F(\mathbf{X})\, \mathbf{X}$ *and the matrix of partial derivatives is* $\nabla F(\mathbf{X}) = F(\mathbf{X})\, \mathbf{X}$.

*Proof.* By the definition of $t_{ijk}(\mathbf{X})$ and $h_{ijk}(\mathbf{X})$, and using symmetry, we may rewrite $\phi(\mathbf{X})$ as

$$\phi(\mathbf{X}) = \sum_{i<j} \left\{ \tfrac{1}{3} \sum_{k \neq i, j} h_{ijk}(\mathbf{X}) \right\} t_{ijk}^2(\mathbf{X}) + \sum_{i<j} \left\{ \sum_{k = i, j} h_{ijk}(\mathbf{X}) \right\} d_{ij}^2(\mathbf{X}).$$

The matrix expression follows from inserting the additive decomposition (33), substitution of (21), and simplifying. At configurations $\mathbf{X}$ where $t_{ijk}(\mathbf{X}) = 0$ for some $i, j$, and $k$, the partial derivatives are not defined, but the definition of $F(\mathbf{X})$ is such that $\nabla F(\mathbf{X})$ satisfies the definition of a subdifferential (Rockafellar, 1970). ▮

Thus the matrix expression for the loss function of the $M_2$ model becomes

$$\sigma_C^2(\mathbf{X}) = SSQ_\tau + \operatorname{tr} \mathbf{X}'\mathbf{V}\mathbf{X} - 2 \operatorname{tr} \mathbf{X}' F(\mathbf{X})\, \mathbf{X}, \qquad (35)$$

and the stationary equation for an $M_2$ representation is

$$\mathbf{V}\hat{\mathbf{X}} = F(\hat{\mathbf{X}})\, \hat{\mathbf{X}}. \qquad (36)$$

Analogous to a similar implication for the perimeter model, the stationary configuration $\hat{\mathbf{X}}$ will satisfy $\operatorname{tr} \hat{\mathbf{X}}'\mathbf{V}\hat{\mathbf{X}} = \operatorname{tr} \hat{\mathbf{X}}' F(\hat{\mathbf{X}})\, \hat{\mathbf{X}}$, and therefore a result similar to Proposition 5 can be obtained.

PROPOSITION 6. *Given a set of triadic dissimilarities, suppose that the configuration* $\hat{\mathbf{X}}$ *satisfies the nonlinear equation* $\mathbf{V}\hat{\mathbf{X}} = F(\hat{\mathbf{X}})\, \hat{\mathbf{X}}$ *for the* $M_2$ *model, where the matrices* $\mathbf{V}$ *and* $F(\cdot)$ *are as defined in Lemmas* 3 *and* 6, *respectively. Then the total weighted sum of squares of the triadic dissimilarities* $SSQ_\tau$ *can be decomposed as*

$$SSQ_\tau = \sum_{(i,j,k) \in L} w_{ijk} \tau_{ijk}^2 = \sigma_C^2(\hat{\mathbf{X}}) + \sum_{(i,j,k) \in L} w_{ijk} t_{ijk}^2(\hat{\mathbf{X}}).$$

The $M_2$ model may be fitted to a given set of triadic dissimilarities by an IM algorithm consisting of repeatedly applying the mapping

$$\mathbf{X}_{s+1} = \mathbf{V}^+ F(\mathbf{X}_s)\, \mathbf{X}_s = G_C(\mathbf{X}_s). \qquad (37)$$

A fixed point $\mathbf{X}_* = G_C(\mathbf{X}_*)$ of the mapping $G_C(\cdot)$ defined in (37) will solve (36), and hence it will be a stationary point of $\sigma_C^2(\cdot)$. Comparing the mappings $G_B(\cdot)$ and $G_C(\cdot)$ for the Euclidean perimeter model and the $M_2$ model, respectively, we see two differences. First, instead of recalculating the generalized inverse $A^+(\mathbf{X}_s)$ in each iteration, we calculate $\mathbf{V}^+$ once (where $\mathbf{V}$ has the same formal matrix properties as $A(\mathbf{X})$, and hence the calculation of $\mathbf{V}^+$ is easy). Second instead of premultiplying $\mathbf{X}_s$ in $G_B(\cdot)$ with a correction matrix $B(\mathbf{X}_s)$ having generic element $b_{ij}(\mathbf{X}_s) = \{\sum_k w_{ijk} \tau_{ijk}\}/ d_{ij}(\mathbf{X}_s)$, in $G_C(\cdot)$ we premultiply the previous configuration with a matrix $F(\mathbf{X}_s)$ having generic element $f_{ij}(\mathbf{X}_s) = -\sum_k \{w_{ijk} \tau_{ijk} / t_{ijk}(\mathbf{X}_s)\}$. Thus, in $G_B(\cdot)$ the positions of points $i$ and $j$ are corrected on the basis of the disparity of the current dyadic distance $d_{ij}(\mathbf{X}_s)$ with the total dissimilarity involving $i$ and $j$, while in $G_C(\cdot)$ it is the total disparity of the current triadic distance $t_{ijk}(\mathbf{X}_s)$ with each separate dissimilarity that determines the correction.

A convergence theorem for the $G_C(\cdot)$ mapping can be proven along the same lines as the proof of Theorem 3; the only adjustment needed would be the replacement of inequality (27) by

$$\phi(\mathbf{X}) \geqslant \operatorname{tr} \mathbf{X}' F(\mathbf{Y})\, \mathbf{Y} \quad \forall \mathbf{X}, \mathbf{Y}. \qquad (38)$$

That this inequality holds can be derived from the Cauchy–Schwarz inequality written in the form

$$[\operatorname{tr} \mathbf{X}'\mathbf{E}_{ijk}\mathbf{X}]^{1/2} \, [\operatorname{tr} \mathbf{Y}'\mathbf{E}_{ijk}\mathbf{Y}]^{1/2} \geqslant \operatorname{tr} \mathbf{X}'\mathbf{E}_{ijk}\mathbf{Y}, \qquad (39)$$

multiplying both sides of (39) with $h_{ijk}(\mathbf{Y})$ defined in Lemma 6, summing over $i < j$ and $k$, and simplifying.

## EXAMPLES

We now turn to some examples of triadic model fitting. The first two examples concern the approximation of a one-mode triadic dissimilarity table with an $M_2$ distance model. In the third example, a triadic unfolding method is proposed; i.e., a method analyzing a three-mode table, in which the (dis)similarities are defined on $O_1 \times O_2 \times O_3$, the Cartesian product of three different sets of objects. It is shown how this analysis can be embedded within the algorithmic framework presented, and the approach is illustrated for the Euclidean perimeter model.

### *Comparison of $M_2$ Model with Hayashi's Surface Area Model*

Hayashi (1972) studied an example of three-mode data indicating the unproductivity of teams of three individuals as a function of their composition (Table 2). Note that the values on the diagonal plane are not given. In view of the small number of objects ($n = 6$), this example serves only to illustrate a key difference between the $M_2$ model and Hayashi's model, in which each dissimilarity $\tau_{ijk}$ is represented as the *surface area* of the triangle formed by the points $i, j$, and $k$. It should be noted that this quantity does not satisfy the requirements of an triadic distance; a major drawback of the area model would seem to be that three

**TABLE 2**

**Unproductivity of Triadic Teams (Hayashi, 1972)**

| | | | |
|---|---|---|---|
| $\tau_{123} = 1$ | | | |
| $\tau_{124} = 7$ | | | |
| $\tau_{125} = 6$ | | | |
| $\tau_{126} = 9$ | | | |
| $\tau_{134} = 7$ | $\tau_{234} = 8$ | | |
| $\tau_{135} = 6$ | $\tau_{235} = 7$ | | |
| $\tau_{136} = 9$ | $\tau_{236} = 9$ | | |
| $\tau_{145} = 4$ | $\tau_{245} = 6$ | $\tau_{345} = 3$ | |
| $\tau_{146} = 9$ | $\tau_{246} = 8$ | $\tau_{346} = 5$ | |
| $\tau_{156} = 6$ | $\tau_{256} = 7$ | $\tau_{356} = 3$ | $\tau_{456} = 1$ |

collinear points get surface area zero even when they are arbitrarily far apart. Figure 2 gives the configuration as obtained by Hayashi, and Fig. 3 the configuration obtained from fitting the $M_2$ model by minimizing $\sigma^2_C(\mathbf{X})$ in (34) with $\mathbf{X}$ two dimensional, which gives an optimal DAF of 96.6%. The $M_2$ configuration in Fig. 3 clearly falls into two groups, $\{1, 2, 3\}$ and $\{4, 5, 6\}$, accounting quite well for the data in Table 2, which has $\tau_{123} = \tau_{456} = 1$ and the highest values for triads combining one object from the first group with object 6. Once we realize that the area model predicts small dissimilarity for points on a line, the two groups are apparent in Fig. 1 as well, but in a less natural way.

*Free Sorting of Kinship Terms*

Our next example concerns free sorting data collected by Rosenberg (1982).[3] The stimuli in this experiment were 15 kinship terms, representing the most common genetic relationships; 85 undergraduates classified these terms by similarity. Neither the number of classes nor the number of elements within classes was restricted, so the data consist of 85 free partitions of 15 objects. A classical two-dimensional multidimensional scaling analysis of the degree of distinguishability $m - \rho_{ij}$ between pairs of the kinship terms, calculated across subjects, accounts for 67% of the dispersion and tightly groups them into three clusters, at about equal distance: {grandfather, grandmother, grandson, granddaughter}—kins two generations removed from the self, {brother, sister, father, mother, son, daughter}—the nuclear family, and {nephew, niece, uncle, aunt, cousin}—collaterals. Using Daws' triadic index of distinguishability $m - \rho_{ijk}$, we fitted a $M_2$ distance model in three dimensions (see Fig. 4), which accounts for 99.14% of the dispersion. From Fig. 4b it is clear that the third dimension distinguishes the terms by sex, combined with collaterality, while dimensions one and two (Fig. 4a) both capture the remoteness of the genetic relationship in terms of the same three clusters as described above. Since the first two axes are globally redundant, detailed relations between the kinship terms can best seen in the plot of dimensions one and three (Fig. 4b). A subdivision of the three main clusters is formed by the same-sex pairs {grandfather, grandson}, {grandmother, granddaughter}, {father, son}, {mother, daughter}, {nephew, uncle}, and {niece, aunt}, which are close together in the whole space.

[3] The free-sorting data were kindly made available to us by Dr. John Daws of the Department of psychology, Columbia University, New York.

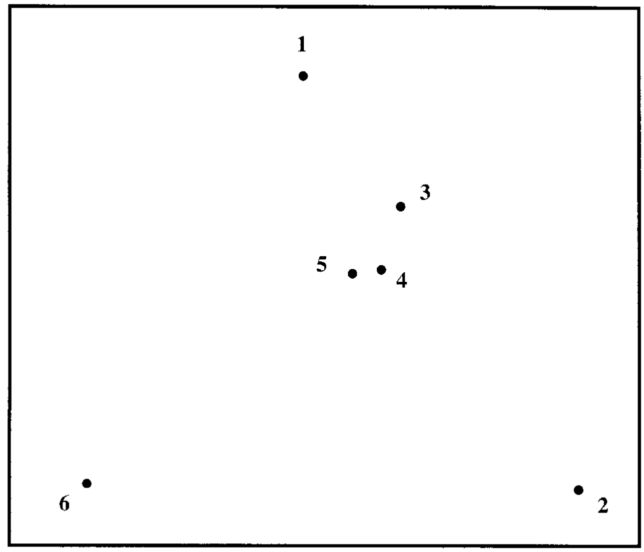


**FIG. 2.** Two-dimensional representation of the Hayashi (1972) data under the surface area model.

*Cross-Modality Similarities*

Joly and Le Calvé (1995) reported a sensory experiment, in which 60 undergraduates were asked to associate a color with a taste and a sound; that is, subjects had to choose triadic combinations from three different stimulus sets $O_1$, $O_2$, and $O_3$, each one associated with one sense modality. Every modality was represented by four terms: $O_1 =$ {white, red, green, black}, $O_2 =$ {salt, sweet, bitter, sour}, and $O_3 =$ {silent, shrill, strident, deep}. The subjects were instructed to select the 5 combinations that to them seemed to be most similar. If the observations are collected in a binary $12 \times 300$ table ($3 \times 4$ rows for the stimuli and $60 \times 5$

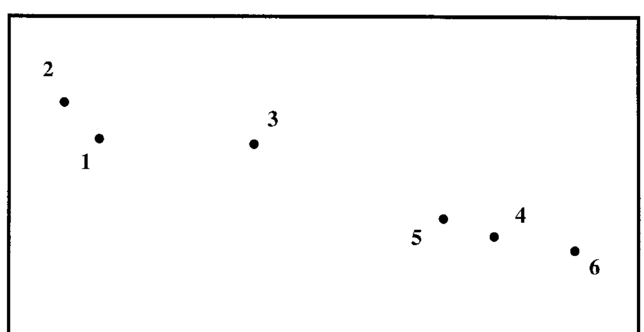


**FIG. 3.** Two-dimensional generalized Euclidean solution for the Hayashi (1972) data.

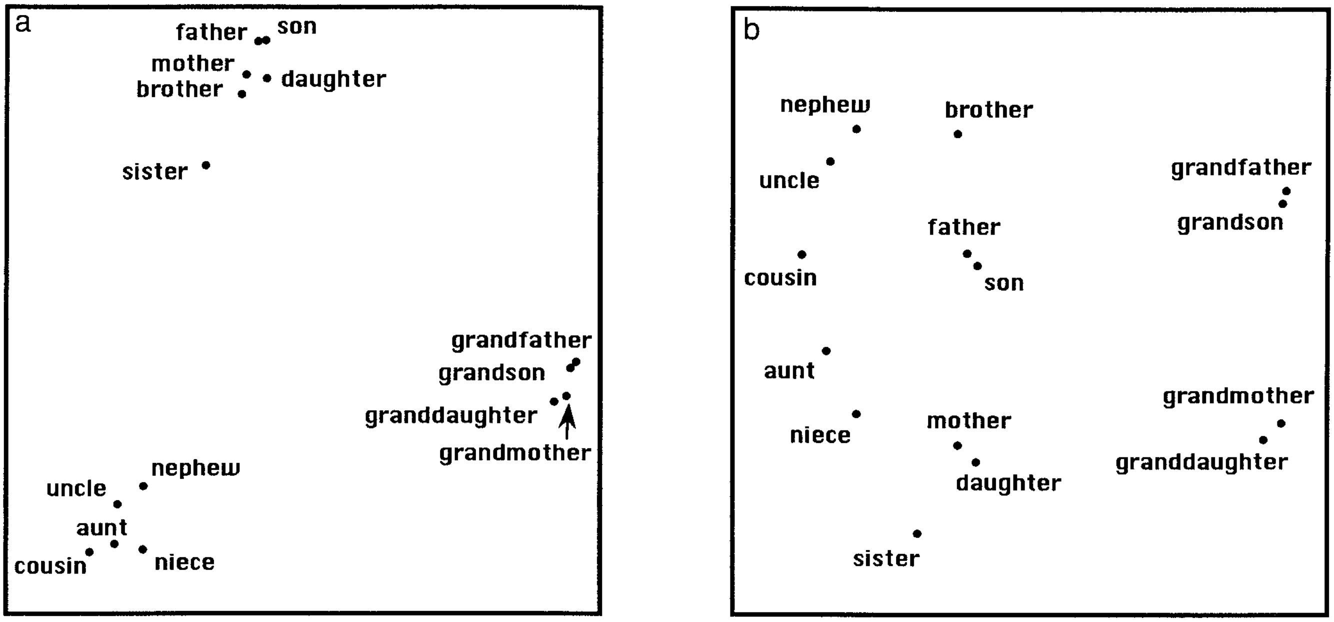


**FIG. 4.** Three-dimensional generalized Euclidean solution for the Rosenberg (1982) data; at the left, axis 2 against axis 1 (a), at the right, axis 3 against axis 1 (b).

columns for the subjects, were each entry in a column indicates membership of a triad), we may calculate any of the triadic indices discussed earlier, but note that the design is such that only a $4 \times 4 \times 4$ subtable of the entire $12 \times 12 \times 12$ table contains empirical information. Each three-way, three-mode table can be regarded as a subtable of a larger three-way, one-mode table, in which the elements of the single mode are the union of the three original modes. Then any triadic one-mode method may be used to find a Euclidean representation of the three-mode table, provided that a weighted algorithm is available in which we set $w_{ijk} = 1$ if $(i, j, k) \in O_1 \times O_2 \times O_3$ and $w_{ijk} = 0$ otherwise.

For the Joly–Le Calvé data we used the Russel–Rao dissimilarity, that is, a constant minus the number of positive matches, and fitted a Euclidean perimeter representation in three dimensions. The resulting configuration,

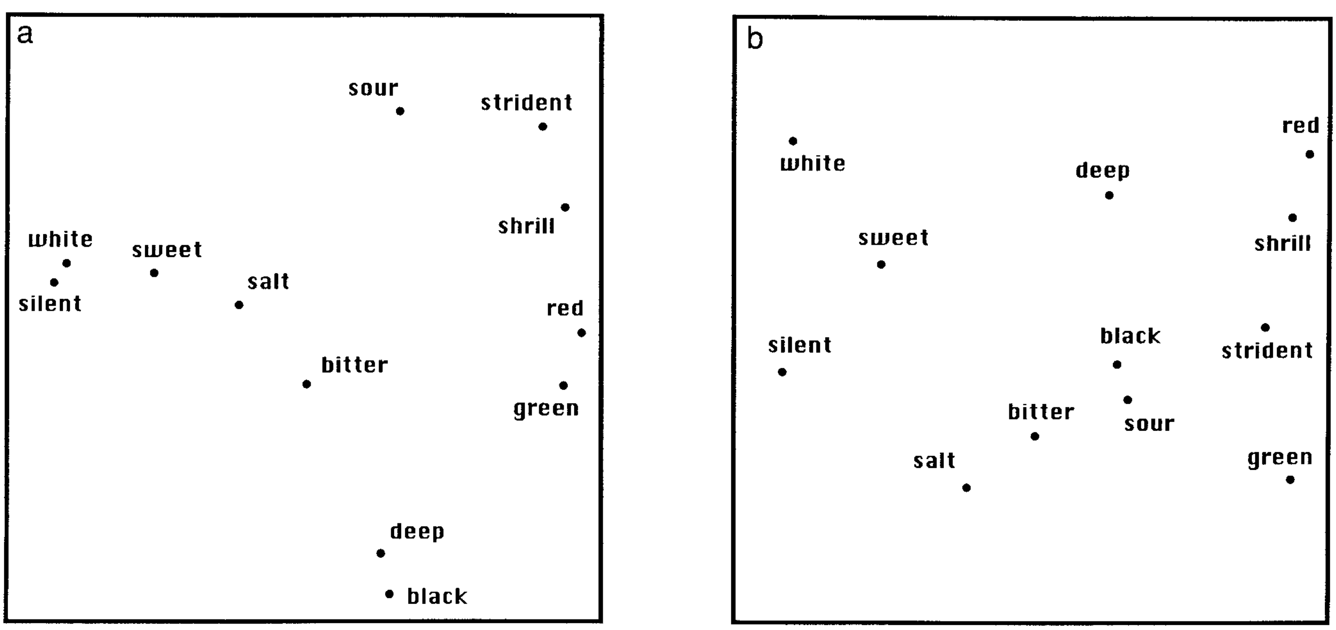


**FIG. 5.** Three-dimensional solution under the Euclidean perimeter model for the Joly & Le Calvé (1995) data; at the left, axis 2 against axis 1 (a), at the right, axis 3 against axis 1 (b).

shown in Fig. 5, accounted for 96.3% of the dispersion. Comparing terms from different sets, the triad {sweet, white, silent} has the smallest perimeter in the three-dimensional configuration as a whole, followed by {bitter, black, deep}. There are other triads, such as {sour, black, strident}, that are closer together in Fig. 5b, but father apart in Fig. 5a, indicating a less strong association. Generally, the color and sound elements are more spread out in space than the taste elements. The first dimension contrasts {red, green shrill, strident} with {white, silent}, that is, activity versus rest. In the second dimension, the most characteristic difference appears to be between {black, deep} and {sour, strident} and could be interpreted as heaviness–lightness. The third dimension is a contrast between {white, red, deep, sweet} and {salt, green, bitter}, which seems to indicate attraction or satisfaction versus aversion or danger.

## DISCUSSION

Several triadic dissimilarity and distance models have been brought together by this paper in a common theoretical framework. Triadic dissimilarity was characterized by three-way symmetry and diagonal-plane equality, two properties which clearly distinguish this concept from the more usual three-way, two-mode dyadic proximities. To define triadic distance, we proposed two additional requirements: definiteness and the tetrahedral inequality. The axiomatization proposed by Joly and Le Calvé (1995) is rather similar to the present one, but based on differently parametrized inequalities:

$$t_{iij} \leqslant \lambda t_{ijk},$$
$$\max(t_{ijk}, t_{ijl}) \leqslant \nu(t_{ikl} + t_{jkl}),$$
$$2t_{ijk} \leqslant \mu(t_{ikl} + t_{jkl} + t_{ijl}).$$

From the proof of Lemma 2 we see that the Joly–Le Calvé model requires $\lambda = 1$, $\mu = \frac{4}{3}$, $\nu = 1$, while the triadic model has $\lambda = 2$, $\mu = 1$, $\nu = 1$. In spite of the apparent similarity, it should be noted that the $M_p$ model as defined in (8) generally does not satisfy $t_{iij} \leqslant \lambda t_{ijk}$ with $\lambda < 2^{1/p}$.

Most models and algorithms presented in this paper can easily be generalized to the $K$-adic case. A natural context in which $K$-adic dissimilarity functions have been studied is the order theoretic framework of hierarchical clustering (Hubert, 1977; Bandelt and Dress, 1994). In principle, hierarchical or additive clustering structures could be fitted within the present framework as well (Heiser, 1996). Compared to the Cox *et al.* (1991) algorithm, which is based on the gradient method, the present algorithm for the $M_2$ model has two advantages: it includes weights and it is guaranteed to converge. The weights can be used in iteratively reweighted least squares procedures to adapt to different assumptions about the residuals. To our knowledge, there was no least squares procedure available for the perimeter model.

Whether or not fitting triadic models really adds a lot to what dyadic models have to tell us is a difficult question. From simulation work of Cox *et al.* (1991) with the Jaccard index it appears that dyadic MDS is best at picking out pairs of individuals, triadic MDS is best at picking out triples, and so on for higher-way analyzes. Pan and Harris (1991) concluded in their example that the results of a dyadic analysis and a triadic analysis were globally similar, although the latter highlighted some interesting triple associations. However, we note that, if the two analyses would essentially give identical information, this circumstance should not be regarded as a negative result, since it implies that a single spatial configuration accounts for a lot more data than is usual. In a statistical analogy, it implies that the three-way interaction is a simple function of the two-way interactions. The three-mode analysis for triadic proximity data, as illustrated in the third example, certainly adds something valuable, since it extends our experimental repertoire with the possibility to study relationships with a whole extra set of points compared to a standard unfolding paradigm.

## REFERENCES


Andreae, T., & Bandelt, H.-J. (1995). Performance guarantees for approximation algorithms depending on parametrized triangle inequalities. *SIAM Journal of Discrete Mathematics*, **8**, 1–16.

Bandelt, H.-J., & Dress, A. W. M. (1994). An order theoretic framework for overlapping clustering. *Discrete Mathematics*, **136**, 21–37.

Barlett, M. S. (1935). Contingency table interactions. *Journal of the Royal Statistical Society*, *Supplement*, **2**, 248–252.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an $n$-way generalization of Eckart–Young decomposition. *Psychometrika*, **35**, 283–319.

Ciarlet, P. J. (1989). *Introduction to numerical linear algebra and optimisation.* Cambridge: Cambridge Univ. Press.

Coppi, R., & Bolasco, S. (Eds.) (1989). *Multiway data analysis.* Amsterdam: North-Holland.

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 coocurrences. *Journal of Classification*, **13**, 57–80.

De Leeuw, J., & Heiser, W. J. (1977). Convergence of correction matrix algorithms for multidimensional scaling. In J. C. Lingoes, E. E. Roskam, & I. Borg (Eds.), *Geometric representations of relational data* (pp. 735–752). Ann Arbor, MI: Mathesis Press.

De Leeuw, J., & Heiser, W. J. (1980). Multidimensional scaling with restrictions on the configuration. In P. R. Krishnaiah (Ed.), *Multivariate analysis* (Vol. V, pp. 501–522). Amsterdam: North-Holland.

De Leeuw, J., & Heiser, W. J. (1982). Theory of multidimensional scaling. In L. Kanal & P. R. Krishnaiah (Eds.), *Handbook of Statistics* (Vol. II, pp. 285–317). Amsterdam: North-Holland.

Dykstra, R. L. (1983). An algorithm for restricted least squares regression. *Journal of the American Statistical Association*, **78**, 839–842.

Fichet, B. (1986). Distances and Euclidean distances for presence-absence characters and their application to factor analysis. In J. de Leeuw, W. J. Heiser, J. J. Meulman, & F. Critchley (Eds.), *Multidimensional data analysis* (pp. 23–46). Leiden: DWSO Press.

Fichet, B., & Le Calvé, G. (1984). Structure géométrique des principaux indices de dissimilarité sur signe de présence-absence. *Statistique et Analyse des Données*, **9**, 11–64.

Gower, J. C. (1986). Euclidean distance matrices. In J. de Leeuw, W. J. Heiser, J. J. Meulman, & F. Critchley, (Eds.), *Multidimensional data analysis* (pp. 11–22). Leiden: DWSO Press.

Gower, J. C., & Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification*, **3**, 5–48.

Groenen, P., & Heiser, W. J. (1991). *An improved tunneling function for finding a decreasing series of local minima in MDS* (Research Report RR-91-06). Leiden, The Netherlands: University of Leiden, Department of Data Theory.

Halmos, P. R. (1958). *Finite-dimensional vector spaces* (2nd ed.). New York: Van Nostrand–Reinhold.

Hayashi, C. (1972). Two dimensional quantification based on the measure of dissimilarity among three elements. *Annals of the Institute of Statistical Mathematics*, **24**, 251–257.

Hayashi, C. (1989). Multiway data matrices and method of quantification of qualitative data as a strategy of data analysis. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 131–142). Amsterdam: North-Holland.

Heiser, W. J. (1987). Correspondence analysis with least absolute residuals. *Computational Statistics and Data Analysis*, **5**, 337–356.

Heiser, W. J. (1988). Multidimensional scaling with least absolute residuals. In H.-H. Bock (Ed.), *Classification and related methods of data analysis* (pp. 455–462). Amsterdam: North-Holland.

Heiser, W. J. (1991). A generalized majorization method for least squares multidimensional scaling of pseudodistances that may be negative. *Psychometrika*, **55**, 7–27.

Heiser, W. J. (1995). Convergent computation by iterative majorization: Theory and applications in multidimensional data analysis. In W. J. Krzanowski (Ed.), *Recent advances in descriptive multivariate analysis* (pp. 157–189). Oxford: Oxford Univ. Press.

Heiser, W. J. (1996). *Fitting graphs and trees with multidimensional scaling methods.* Paper presented at the meeting of the International Federation of Classification Societies, Kobe, Japan.

Hubert, L. (1977). A set-theoretical approach to the problem of hierarhical clustering. *Journal of Mathematical Psychology*, **15**, 70–88.

Joly, S., & Le Calvé, G. (1995). Three-way distances. *Journal of Classification*, **12**, 191–205.

Law, H. G., Snyder, C. W., Hattie, J. A., & McDonald, R. P. (Eds.) (1984). *Research methods for multimode data analysis*. New York: Praeger.

Le Calvé, G. (1985). Distances à centre. *Statistique et Analyse des Données*, **10**, 29–44.

Pan, G., & Harris, D. P. (1991). A new multidimensional scaling technique based upon associations of triple objects Pijk and its application to the analysis of geochemical data. *Mathematical Geology*, **23**, 861–886.

Rockafellar, R. T. (1970). *Convex analysis*. Princeton, NJ: Princeton Univ. Press.

Rosenberg, S. (1982). The method of sorting in multivariate research with applications selected from cognitive psychology and person perception. In N. Hirschberg & L. G. Humphreys (Eds.), *Multivariate applications in the social sciences* (pp. 117–142). Hillsdale, NJ: Erlbaum.

Tucker, L. R. (1964). The extension of factor analysis to three-dimensional matrices. In N. Fredericksen & H. Gulliksen (Eds.), *Contributions to mathematical psychology* (pp. 109–127). New York: Holt, Rinehart, and Winston.

Received: April 24, 1995