Journal of Classification 12:243-263 (1995)

# Additive Two-Mode Clustering: The Error-Variance Approach Revisited

Boris Mirkin

Rutgers University

Phipps Arabie

Rutgers University

Lawrence J. Hubert

University of Illinois

---

The research was supported by the Office of Naval Research under grant number N00014-93-1-0222 to Rutgers University. The authors are indebted both to Fionn Murtagh, who served as Acting Editor, and to anonymous Referees for thoughtful and constructive reviews.

Authors' addresses: Boris Mirkin, DIMACS, Rutgers University, P.O.Box 1179, Piscataway, NJ 08855-1179 USA and Central Economics-Mathematics Institute, Moscow, Russia; Phipps Arabie, Faculty of Management, Rutgers University, 180 University Avenue, Newark NJ 07102-1895 USA; and Lawrence J. Hubert, Department of Psychology, University of Illinois, 603 E. Daniel St., Champaign IL 61820 USA.

**Abstract:** The additive clustering approach is applied to the problem of two-mode clustering and compared with the recent error-variance approach of Eckes and Orlik (1993). Although the schemes of the computational algorithms look very similar in both of the approaches, the additive clustering has been shown to have several advantages. Specifically, two technical limitations of the error-variance approach (see Eckes and Orlik 1993, p. 71) have been overcome in the framework of the additive clustering.

**Keywords:** Two-mode clustering; Additive clustering; Correspondence analysis; Addition/deletion algorithm.

## 1. Introduction

Two-mode clustering has appeared in the literature rather frequently over the last twenty years (see, for example, Hartigan 1972, 1976; DeSarbo 1982; DeSarbo and De Soete 1984; Arabie, Schleutermann, Daws, and Hubert 1988; Packer 1989; Braverman, Kiseleva, Muchnik, and Novikov 1974, and Mirkin and Rostovtsev 1978, among Russian references). Recently, Eckes and Orlik (1991, 1993; also see Eckes 1993) have produced a new approach to the problem, constructing two-mode clusters sequentially using a type of variance criterion, combined with certain additional heuristic approaches. Those authors classify two-mode clustering strategies into three general categories: (a) direct clustering, applied to two-mode data without any preliminary transformations or the use of one-mode distances or similarities; (b) fitting tree structures to two-mode data; (c) additive clustering, representing the proximities between pairs of entities as combinations of discrete and possibly overlapping properties. Eckes and Orlik (1993) view their procedure as combining the advantages of the first two categories of methods.

A goal of the present paper is to demonstrate that such a division of clustering methods should be considered rather arbitrary. Specifically, additive clustering (c) should not be considered as being very different from "direct clustering" (a).

We show a straightforward extension of an additive clustering model (Shepard and Arabie 1979; Mirkin 1987) for two-mode row/column unconditional data in a framework similar to the cluster-by-cluster approach of Eckes and Orlik. An additive two-mode clustering strategy (referred to as Box Clustering) is presented. As a by-product, that strategy resolves two problems noted by Eckes and Orlik (1993, p. 71) in their approach: the possible change of the "standard" proximity value (defined as the maximum value in a proximity matrix) used in forming the clusters and the option of post hoc creation of possibly overlapping clusters. Additionally, our own approach allows estimation of the contributions of the individual clusters to the total sum of squares of the input data. We believe that emphasizing a distinction between

methods based on explicit clustering models versus other "ad hoc" algorithms provides a taxonomy of two-mode clustering techniques that is more general than the three-category classification given by Eckes and Orlik (1993).

The paper is organized as follows. Section 2 presents a formal analysis of the error-variance technique originally proposed by Eckes and Orlik (1991, 1993; also see Eckes 1993). Section 3 introduces an additive ("Box Clusters") clustering model and provides two additive box clustering algorithms for fitting the model. They differ in the value of the "standard" proximity used in forming clusters: the maximum data value, as proposed by Eckes and Orlik, and the least-squares optimal value derived for the Box Clustering model. We discuss some properties of that model, as well as of the associated algorithms and the criteria they use. In Section 4, the box clustering methodology is applied to contingency data analyzed earlier with a Correspondence Analysis approach (see Mirkin 1993). In Section 5, the additive two-mode clustering algorithms are applied to two data tables as considered by Eckes and Orlik (1993). Finally, some features of the box clustering approach are reviewed.

## 2. Error-Variance Approach

### 2.1. The Error-Variance Criterion

Eckes and Orlik's procedure can be described as follows. Let the data be presented as a matrix $\mathbf{X} = (x_{ij})$, $i \in I$, $j \in J$, where $I$ and $J$ are sets of indices corresponding to the two modes of entities and $x_{ij}$ are two-mode proximity data keyed as similarities; the data analytic goal is to reveal the major associations that are present between members of these two sets as represented by the values $x_{ij}$. A two-mode cluster concept (later referred to as a *box cluster*, or just *box*) is used for this purpose and is defined as the Cartesian product $V \times W$ of subsets $V \subseteq I$ and $W \subseteq J$. Any box $V \times W$ is associated with submatrix $\mathbf{X}(V,W) = (x_{ij})$, $i \in V$, $j \in W$. The quality of the box cluster $V \times W$ is measured by the criterion

$$MSD(V,W) = \frac{1}{|V|\,|W|} \sum_{i \in V, j \in W} (x_{ij} - \mu)^2 , \tag{2.1}$$

where $\mu$ is the maximum entry in the input matrix $X$, and $|V|$ and $|W|$ denote the respective cardinalities of the sets $V$ and $W$. The criterion resembles conventional badness-of-fit measures, except for only being defined locally, not for all the data, but only for the proximities relevant to the given cluster.

Eckes and Orlik (1993 p. 57) argued that their choice of $\mu$ helped maintain homogeneity of the selected entries $x_{ij}$, $(i,j) \in V \times W$. This approach can be extended to the context of additive clustering (see Section 3.2)

Those authors' algorithm starts with singleton entries $V = \{i\}$ and $W = \{j\}$ for a box cluster, corresponding to the maximal $x_{ij}$; rows or columns are then added iteratively, based on certain heuristic "subcriteria" (see formulae (5) to (8), p. 58, in Eckes and Orlik 1993). The most general of them, (8), is defined for the case when two arbitrary two-mode clusters are merged. It equals the average of the squared errors for the newly added proximities between the elements of these clusters, which seems to us a rather indirect evaluation of the change in criterion MSD. In our opinion, the subcriteria should be based directly on the changes in criterion (2.1) as the box clusters are iteratively formed by the algorithm.

More explicitly, if a box cluster $V_1 \times W_1$ is to be augmented by subsets $V_2 \subset I$ and $W_2 \subset J$, where $V_1 \cap V_2 = \emptyset$ and $W_1 \cap W_2 = \emptyset$, in the larger box $(V_1 \cup V_2) \times (W_1 \cup W_2)$, the increment to the criterion (2.1) equals

$$\Delta = MSD(V_1 \cup V_2, W_1 \cup W_2) - MSD(V_1, W_1) .$$

For example, the case of adding row $k$ to box $V \times W$ corresponds to the following situation: $V_1 = V$, $W_1 = W$, $V_2 = \{k\}$, $W_2 = \emptyset$. It is easy to show in this case that

$$\Delta = \frac{1}{(|V| + 1)} \left( \frac{1}{|W|} \sum_{j \in W} (x_{kj} - \mu)^2 - MSD(V,W) \right) , \qquad (2.2)$$

which implies that adding a row to the box $V \times W$ requires minimizing

$$d(k) = \frac{1}{|W|} \sum_{j \in W} (x_{kj} - \mu)^2 , \qquad (2.3)$$

to make the increment in criterion (2.1) as small as possible. This formulation corresponds to Case II of Eckes and Orlik (1993, p. 58, Eqs. 6 and 7), but not to those authors' more general Case III (1993, p. 58, Eq. 8).

## 2.2 Error-Variance Box Clustering

Based on the considerations above, we propose the following algorithm as a potential improvement over Eckes and Orlik's (1993) approach.

**Error-Variance Box Algorithm**

1. Find a pair $(i,j) \in I \times J$ maximizing $x_{ij}$ and define $V = \{i\}$, $W = \{j\}$, and $\mu = max\, x_{ij}$.
2. For any row $k \notin V$ and for any column $l \notin W$, calculate the change in criterion (2.1) based on formula (2.2) or its counterpart version for columns, and find the (row or column) entity that minimizes the change in the criterion. If the change is not large (Eckes and Orlik 1993, p. 59 used a specially introduced measure of "centroid effect ratio" [CER]) in the cluster that is not supposed to fall below 80%), add the entity to the box cluster, and repeat Step 2 from the beginning. Else go to Step 3.
3. Redefine sets $I$ and $J$ respectively as sets $I - V$ and $J - W$ to find the next box cluster, and go to Step 2 if both sets are not empty. If either of the sets is empty, END.

After the solution is found it could be used as is, or followed up by one of two options: hierarchical agglomeration of the clusters into a dendrogram (Step 4′ below), or an augmentation from the initial sets $I$ and $J$ to produce possibly overlapping clusters (Step 4″ below). Thus, if necessary, after the END, go to either 4′ or 4″.

4′. **Hierarchy construction:** the clusters obtained are merged pairwise, based on some criterion. In contrast to Eckes and Orlik's (1993) indirect criterion, we suggest an alternative based on the increment in the original criterion (2.1) which can be expressed as follows:

$$\Delta = MSD(V_1 \cup V_2, W_1 \cup W_2) - MSD(V_1, W_1) - MSD(V_2, W_2).$$

4″. **Overlapping follow-up:** Any box cluster is augmented elementwise by adding a row or a column, as done in Step 2, this time from the initial sets $I$ and $J$.

Eckes and Orlik (1993, p. 71) emphasized two targets of opportunity for improving the box clustering approach: (a) a strategy for changing the value $\mu$ when all large values of $x_{ij}$ are to be taken into account, and (b) obtaining overlapping clusters from the beginning, rather than as a follow-up to the determination of disjoint box clusters.

## 3. Additive Box Clustering

### 3.1. Additive Box Model

The additive clustering technique proposed here is closely related to the algorithm considered in the previous section.

Consider a set of $m$ box clusters, $V_1 \times W_1, \ldots, V_m \times W_m$, along with corresponding intensity weights $\lambda_1, \ldots, \lambda_m$. These clusters are referred to as additive box clusters if they fit the raw data $\mathbf{X}$ according to the following model (cf. the model considered in Shepard and Arabie 1979, and Mirkin 1987):

$$x_{ij} = \sum_{t=1}^{m} \lambda_t v_{it} w_{jt} + e_{ij} , \tag{3.1}$$

with "small" residuals $e_{ij}$, $i \in I$, $j \in J$. Boolean vectors $\mathbf{v}_t, \mathbf{w}_t$ correspond to the boxes $V_t \times W_t$ by the common rule: $v_{it} = 1$ iff $i \in V_t$ and $w_{jt} = 1$ iff $j \in W_t$, $t = 1, \ldots, m$.

To relate model (3.1) to the error-variance clustering discussed in the section immediately preceding, we apply the sequential fitting procedure devised by Mirkin (1987). It is based on a doubly greedy optimization strategy: first, clusters are obtained sequentially (cluster-by-cluster) rather than simultaneously, and second, each cluster is formed incrementally, with an element-by-element augmentation strategy.

Specifically, find initially only one box cluster $V \times W$ to minimize the following least-squares criterion based on model (3.1):

$$L^2 = \sum_{i \in I, j \in J} (x_{ij} - \lambda v_i w_j)^2 . \tag{3.2}$$

For any $\lambda$ (for example, equal to the maximal $x_{ij}$, as $\mu$ in (2.1) or, to the average value of submatrix $\mathbf{X}(V,W)$), criterion (3.2) clearly can be rewritten as follows:

$$AB(V,W) = \sum_{i \in V} \sum_{j \in W} (x_{ij} - \lambda)^2 + \sum_{(i,j) \notin V \times W} x_{ij}^2 . \tag{3.3}$$

This last criterion resembles the error-variance criterion (2.1) in that both express the intuitive idea of closeness of the elements of submatrix $\mathbf{X}(V,W)$ to the same value, either $\mu$ or $\lambda$, although in different ways. Two advantages of the additive clustering criterion (3.3) should be pointed out. First, it is connected to the explicit clustering model (3.1). Second, its

behavior is not monotone in the traditional sense of measures of badness-of-fit. Consider, for example, its increment when a row $k \notin V$ is added to $V$:

$$\Delta = AB(V \cup \{k\}, W) - AB(V,W) = \sum_{j \in W} (x_{kj} - \lambda)^2 - \sum_{j \in W} x_{kj}^2 . \qquad (3.4)$$

This value can be either negative or positive depending on the closeness of the subset of row $k$ corresponding to $W$ to $\lambda$ or 0. If $\Delta$ is negative, $k$ must be added to $V$ because doing so decreases the value of criterion $L^2$ in (3.2); if $\Delta > 0$, $k$ may not be added to $V$ because the value of $L^2$ increases with $k$ added. Moreover, the sign of $\Delta$ does not depend on what was added during the previous steps; this feature offers a natural termination for adding the elements to the box — when the change of $L^2$ becomes positive for any external row $k$ (or, symmetrically, column $l$ to add). In contrast, formula (2.2) shows that the increment to criterion (2.1) is always positive if $k$ with minimal $d(k)$ in definition (2.3) were added in previous steps; this property necessitates the heuristic selection of a threshold value.

Consider the additive clustering criterion (3.2) more closely. Clearly, (3.3) equals

$$\begin{aligned} AB(V,W) &= \sum_{i \in V} \sum_{j \in W} (x_{ij} - \lambda)^2 + \sum_{i \in I, j \in J} x_{ij}^2 - \sum_{i \in V} \sum_{j \in W} x_{ij}^2 \\ &= \sum_{i \in I, j \in J} x_{ij}^2 + \sum_{i \in V} \sum_{j \in W} [(x_{ij} - \lambda)^2 - x_{ij}^2] . \end{aligned}$$

Since in the final expression the first term is constant and the contents of the brackets in the last term can be transformed using the elementary formula $a^2 - b^2 = (a - b)(a + b)$, the criterion (3.3) equals that constant term $\Sigma_{i \in I, j \in J} x_{ij}^2$ minus $g(V,W,\lambda)$, where

$$g(V,W,\lambda) = \sum_{i \in V} \sum_{j \in W} \lambda(2x_{ij} - \lambda) . \qquad (3.5)$$

Thus, to minimize (3.3), criterion (3.5) must be maximized. Criterion (3.5) offers a better interpretation of the optimality condition based on the change of sign of (3.4) from negative to positive when $V \times W$ is optimal. Indeed, the change in (3.5) when $k \in I$ is added to $V$ (leaving $W$ invariant) equals:

$$\Delta_g (V,W,k) = \sum_{j \in W} \lambda(2x_{kj} - \lambda) . \qquad (3.6)$$

For the sake of the simplicity, assume $\lambda$ is positive. In this case, $\Delta_g(V,W,k)$ is negative (so that the maximized criterion (3.5) is actually decreased) when

the average value

$$\bar{x}(k,W) = \sum_{j \in W} x_{kj} / |W| \tag{3.7}$$

is less then $\lambda/2$. An analogous condition holds for column objects. This demonstration shows rather clearly the real meaning of the value of $\lambda$ in additive box clustering. Specifically, the requirement of Eckes and Orlik (1993) that $\lambda$ equals the maximal value of $x_{ij}$ implies that the box obtained, $V \times W$, must include only those objects $k \in V$ and $l \in W$ that have their average proximities $\bar{x}(k,W)$, as defined in (3.7), and $\bar{x}(V,l)$, symmetrically defined, to the rest of the box at least as large as half that maximal value. This observation lends support to the choice of the maximal value of $\mu$ for maintaining the large-valued proximities as boxes are being formed.

When the optimal value of $\lambda$ is used in the additive clustering, the construction above can be described as follows. Obviously, the optimal value of $\lambda$ for the criterion minimized in (3.2) for a given box cluster $V \times W$ equals the average internal proximity

$$\lambda = \bar{x}(V,W) = \sum_{i \in V} \sum_{j \in W} x_{ij} / |V||W| . \tag{3.8}$$

With the optimal value (3.8) of $\lambda$ substituted into the criterion $g(V,W,\lambda)$ of (3.5), the criterion equals

$$g(V,W) = (\sum_{i \in V} \sum_{j \in W} x_{ij})^2 / |V||W| = \bar{x}^2(V,W)\,|V||W| . \tag{3.9}$$

This form of the criterion (3.5) does not contain $\lambda$ (which can be determined afterward from formula (3.8)) and can be easily adjusted for the case when the optimal $\lambda$ is negative.

### 3.2 Additive Box Algorithms

The following algorithms to fit model (3.1) by sequentially finding single boxes with criterion (3.2) are analogous to the Error-Variance Box Algorithm described in Section 2.2.

**Additive Box Algorithm 1 (for the case of the maximal intensity weight)**

1. Find a pair $(i,j) \in I \times J$ maximizing $x_{ij}$, and set $V = \{i\}$, $W = \{j\}$, and $\lambda = max\, x_{ij}$.

2. For any row $k \in I$ and for any column $l \in J$, calculate the change in criterion (3.5) based on (3.3) or its counterpart version for columns, and find the row or column minimizing the change. (Alternatively, the simpler formula (3.6) could be used, based on criterion (3.5); in this case, the object-by-object merging process could be interpreted as a version of the average-link method). If the change is negative (positive when (3.6) is considered), add the row/column to the box cluster, and repeat Step 2 from the beginning. Else go to Step 3.

3. Calculate the residuals $x_{ij}' = x_{ij} - \lambda v_i w_j$; replace the the matrix $(x_{ij})$ with those residuals and go to Step 1 if either (a) the cumulative contribution of the obtained box clusters to the total sum of squares of the initial data, $\Sigma_{i \in I, j \in J} x_{ij}^2$, is judged not large enough or (b) the number of clusters found is not sufficient. Else END.

When the box clustering problem is considered for the optimal value (3.8) of $\lambda$ rather than the maximal one, the algorithm above must be modified as follows: (a) to allow the value of $\lambda$ to vary when the box is changed (considering only positive values of $\lambda$, if necessary), (b) to permit the deletion as well as the addition of objects from the box cluster being constructed, for better agreement between the box cluster and the variable $\lambda$. These considerations lead to the following modified algorithm.

**Additive Box Algorithm 2 (for the case of the optimal intensity weight)**

1. Find a pair $(i,j) \in I \times J$ maximizing $x_{ij}$ (when only positive values of $\lambda$ are sought) or criterion (3.9) which equals $x_{ij}^2$ for singleton boxes $\{i\} \times \{j\}$ (when $\lambda$ is permitted to be negative) and set $V = \{i\}$, $W = \{j\}$, and $\lambda = x_{ij}$ for the corresponding $i,j$.

2. For any row $k \in I$ and for any column $l \in J$, calculate the change in the criterion as expressed in formula (3.5) (having $\lambda$ equal to $\bar{x}(V,W)$ given in (3.7)) or in (3.9) because of the modified state of the object in its relation to $V$ or $W$ (that is, adding $k$ to $V$ if $k \notin V$, or removing $k$ from $V$ if $k \in V$, and similarly for $l$), and find the maximum of those changes. If the change is positive, add the row or column to the box cluster, and repeat Step 2 from the beginning. Else go to Step 3.

3. Calculate the residuals $x_{ij}' = x_{ij} - \bar{x}(V,W) v_i w_j$; replace the matrix $(x_{ij})$ with those residuals and go to Step 1 if either (a) the cumulative contribution of the obtained box clusters to the total sum of squares of the initial data, $\Sigma_{i \in I, j \in J} x_{ij}^2$, is judged to be not large enough, or (b) the number of clusters found is not sufficient. Else END.

### 3.3. Comments on the Additive Box Algorithms

Although we already have pointed out two features of the approach, there are several other distinctions to be made between the algorithms for additive box clustering and the error-variance box algorithm of Section 2.2:

1. The residuals are now used, thus allowing recalculation of the maximal element of the data after taking into account the previous ones (after Step 3 and for any box cluster, the calculations start from Step 1 rather than Step 2). This feature avoids problem (a) of the error-variance approach mentioned above in Section 2.2, as discussed by Eckes and Orlik (1993).

2. The additive box clusters are constructed so as to allow possible overlap from the beginning, in contrast to the more restrictive structures arising from the error-variance approach.

3. As shown above (when (3.5) was derived), the maximized value of $g(V,W,\lambda)$ equals the contribution of the cluster to the total sum of squares, $\Sigma_{i\in I,j\in J} x_{ij}^2$, which allows use of the former value (single or cumulative) in an analysis of the comparative importance of the clusters obtained.

4. The results of the Additive Box Algorithm above depend on the origin of the scale of measurement for the data $x_{ij}$: the results could differ using the transformation $(x_{ij} - a)$ for varying $a$, in contrast to those from the error variance-criterion (2.1). This drawback necessitates careful selection of a preliminary transformation of the data. Usually, the data should be centered (that is, $a$ equal to the average of all the $x_{ij}$ should be subtracted from the data values). From a different perspective, this arbitrariness also allows for the adjustment of an additional parameter and the search for a variety of solutions.

5. The additive clusters obtained do not generate any hierarchical structure (which is not necessary in the authors' opinion for successful data analyses; see a similar point of view discussed in Eckes 1993). But a modification analogous to the one described in the error-variance box clustering algorithm as Step 4′ (in Section 2.2) could be considered. This modification requires initially obtaining nonoverlapping clusters by either of the additive box clustering Algorithms 1 or 2 of the previous Section 3.2. To construct the latter clusters, it is sufficient in the merging process of the algorithms (Steps 1 and 2) to use the reduced sets $I$ and $J$ (removing the elements of the clusters already formed). Then, the same merging process as in Step 4′ of the Error-Variance Box Clustering Algorithm (Section 2.2) is carried out as follows:

*Hierarchy construction.* The clusters obtained are merged pairwise, based on the increment of the original criterion (3.2) or the derived criterion (3.9) (for optimal $\lambda$), which can be expressed as follows:

$$\Delta = F(V_1 \cup V_2, W_1 \cup W_2) - F(V_1, W_1) - F(V_2, W_2),$$

where $F$ is either $AB$ in (3.3) or $g$ in (3.9) based on the maximized criterion (3.5), so that

$$\Delta_g = \lambda \left( \sum_{i \in V_1} \sum_{j \in W_2} (2x_{ij} - \lambda) + \sum_{i \in V_2} \sum_{j \in W_1} (2x_{ij} - \lambda) \right).$$

## 4. Correspondence-wise Additive Clustering

### 4.1. The Model

A specific additive box clustering strategy can be developed for two-way contingency tables using the methodology of correspondence analysis (Nishisato 1980, 1994; Greenacre 1984; Lebart, Morineau, and Warwick 1984; Greenacre and Blasins 1994). Two features of correspondence analysis, in particular, distinguish it from principal components analysis (see, for example, Carlier and Kroonenberg 1993, Lebart and Mirkin 1993, and Mirkin 1993). The first is that the raw contingency data $\mathbf{P} = (p_{ij})$, where $p_{ij}$ is the probability or frequency or proportion of the entities corresponding to the pair of categories $i \in I$ and $j \in J$, are first transformed into values

$$x_{ij} = \frac{p_{ij} - p_i p_j}{p_i p_j} = (p(i \mid j) - p(i)) / p(i), \tag{4.1}$$

where $p_i$ and $p_j$ are the marginal frequencies. The last expression above shows the meaning of the value $x_{ij}$ as the relative change of the probability of $i$ when $j$ is known.

The second feature is that the model for correspondence analysis can be expressed in the form of model (3.1) fitted with a modified least-squares criterion, where the squared residuals are weighted by the products of corresponding marginal frequencies.

Both these features can be easily taken into account in the framework of additive box clustering. First, consider the model (3.1) as applied to the matrix $\mathbf{X}$ defined by contingency table $\mathbf{P}$ with (4.1). Second, consider the following weighted least-squares criterion instead of (3.2):

$$\Phi^2(V,W) = \sum_{i \in I, j \in J} p_i p_j (x_{ij} - \lambda v_i w_j)^2 , \tag{4.2}$$

where $x_{ij}$ is defined as in (4.1).

Consider only the case when $\lambda$ is defined as the value minimizing (4.2) for any fixed $V$ and $W$. It is not difficult to derive (by setting the derivative of $\Phi^2(V,W)$ with respect to $\lambda$ equal to zero) that the optimal $\lambda$ has the same meaning as the relative change of probability in (4.1), this time for $V$ and $W$. That is, the optimal $\lambda$ equals:

$$\lambda(V,W) = x_{VW} = \frac{p_{VW} - p_V p_W}{p_V p_W} , \tag{4.3}$$

where the aggregate frequencies are defined traditionally as: $p_{VW} = \Sigma_{i \in V} \Sigma_{j \in W} p_{ij}$, $p_V = \Sigma_{i \in V} p_i$, $p_W = \Sigma_{j \in W} p_j$.

This observation can be considered as a legitimization of the use of the weighted least-squares criterion in correspondence analysis, though our result has arisen somewhat unexpectedly in the context of clustering.

Substituting this value of $\lambda$ into (4.2), the criterion could be expressed as follows:

$$\Phi^2(V,W) = \sum_{i \in I, j \in J} p_i p_j x_{ij}^2 - \lambda^2(V,W) p_V p_W . \tag{4.4}$$

The last form of the criterion shows that its final term must be maximized to define a correspondence-wise additive box. By substituting expression (4.3) for optimal $\lambda$ into that term, the criterion can be written:

$$f(V,W) = (\sum_{i \in V} \sum_{j \in W} p_i p_j x_{ij})^2 / (p_V p_W) . \tag{4.5}$$

Now an algorithm can be formulated analogously to those considered above, i.e., the Error-Variance Box Algorithm (Section 2.2) and Additive Box Algorithms 1 and 2 (Section 3.2).

### 4.2. Correspondence-wise Box Algorithm

1. Find a pair $(i,j) \in I \times J$ maximizing $x_{ij}^2 p_i p_j$, and set $V = \{i\}$, $W = \{j\}$, and $\lambda = x_{ij}$.
2. For any row $k \notin V$ and for any column $l \notin W$, calculate the change in criterion (4.5), and find the entity (row or column) maximizing the change. If it is positive, add the row/column to the box cluster, and repeat Step 2

from the beginning. Else go to Step 3.

3. Calculate the residuals $x'_{ij} = x_{ij} - \lambda(V,W)v_i w_j$; replace the matrix $(x_{ij})$ with those residuals and go to Step 1 if either (a) the cumulative contribution of the constructed box clusters to the total sum of squares of the initial data is not large enough, or (b) the number of the clusters found is judged not sufficient. Else End.

### 4.3. Comments on the Correspondence-wise Box Algorithm

The following characteristics of the algorithm above should be pointed out:

1. The value of $\Sigma_{i\in I, j\in J} p_i p_j x_{ij}^2$ equals the Pearson contingency coefficient $\Phi^2$. Thus, the criterion maximized is just the contribution of the box cluster to the value of the contingency coefficient.
2. The box constructed can show a rather deviant behavior in the sense of the relative change of the frequencies; that is, for any row $i$ or column $j$ outside the constructed box $V \times W$, the absolute values of the relative changes $x_{Vj}$ and $x_{iW}$ are half or less than the absolute value of the relative "internal" change $x_{VW}$ as defined in (4.3) (see Mirkin 1993).
3. The preliminary transformation of the contingency data through formula (4.1) resembles the one recommended by DeSarbo and De Soete (1984) and used by Eckes and Orlik (1993). Here, the transformation is used as part of the correspondence-wise clustering approach. Moreover, an interesting quantitative characteristic of any additive box cluster $V \times W$ (in the model (3.1)) obtained with the approach is that its intensity weight $\lambda$ has the same form as expressed in (4.3); that is, obtained with the same transformation, this time applied to the aggregate probability $p_{VW} = \Sigma_{i\in V} \Sigma_{j\in W} p_{ij}$ of the box.
4. The algorithm works with the preliminary transformation of the data uniquely defined through formula (4.1). This feature distinguishes this algorithm from the other Additive Box Clustering algorithms (see Comment 4 in Section 3.3).

## 5. Some Empirical Results

We reconsider two examples of data analyzed earlier by Eckes and Orlik (1993), using the additive two-mode clustering approach as such, without any supplementary agglomeration of the clusters. The basic motivation is to find which classes of the columns correspond to specific classes of

the rows; the question arises in substantive research (see, for example, Price 1974; Hartigan 1976; Eckes 1993).

In the first example, the data are a $15 \times 15$ table of proximities between pairs formed from 15 kinds of situations and 15 kinds of human behavior, based on the appropriateness of the behavior to the situation (judged on a scale from 0 to 9 by fifty-two subjects in an experiment by Price and Bouffard 1974), as reported in Eckes and Orlik (1993, p. 66). The data in Table 1 are deviations of the raw proximities from their grand mean.

The results of applying the Additive Box Algorithms to the data are described in Tables 2 and 3. The first presents the results obtained with Algorithm 1 with maximal $\mu$ (from the residuals matrix) as the intensity weight for any box cluster. The resulting clusters seem reasonable. But at the same time, the following peculiarity of the results seems to be forced by defining the intensity weight $\mu$ for any cluster as the maximal proximity: all the row object subsets $V_s$ in the four box clusters having maximal $\mu$s are mutually exclusive (the first four such row clusters cover 14 situations with only Church, which has the smallest proximities, not covered) because after the intensity weight was subtracted from the corresponding data, the other entries in the rows were not sufficiently large for the corresponding objects to be selected again. The other peculiarity, that the column object sets in the initial clusters are proper subsets of the column set of the first box, seems to reflect the actual characteristics of the data.

The clusters shown in Table 2 account for 39.7% of the total sum of squares of the data in Table 1, or of the variance of those data (for these specific data, these measures are equivalent since the mean value of the proximities in Table 1 equals zero, and thus, the sum of squares is proportional to the variance for the results shown in both Tables 2 and 3). Such a fit might seem unacceptably low, compared to higher levels usually obtained with regression or factor analyses. But recall that our solution uses only Boolean (not quantitative) factors for clusters. The low contribution we obtained suggests that the clusters shown in the Table 2 reflect an aggregate picture of the data but that there are many local peculiarities in the data which will be reflected in the less substantial and smaller clusters omitted here. Also, in this particular approach, the intensity weight equals the maximal proximity in the matrix, which could be rather far from the optimal intensity weight (equal to the average), and this discrepancy decreases the amount of explained variance (although not dramatically so, as will be seen from the results given below in Tables 3 and 4). The Eckes-Orlik approach does not seek to minimize any least-squares function, though their criterion is a "local" goodness-of-fit, and we therefore cannot compare our goodness-of-fit to those authors' even for the same data.

Table 1: Situation-Behavior Centered Proximities.

| Situation | Behavior | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Run | Talk | Kiss | Write | Eat | Sleep | Mumb | Read | Fight | Belch | Argue | Jump | Cry | Laugh | Shout |
| Class | -1.99 | 1.70 | -2.41 | 3.66 | -0.28 | -0.91 | -0.89 | 2.76 | -3.30 | -2.74 | 0.82 | -2.72 | -2.30 | 1.72 | -2.57 |
| Date | 0.49 | 4.05 | 4.22 | -0.89 | 3.28 | -0.74 | -1.39 | -1.63 | -0.93 | -2.28 | -0.01 | -0.09 | -1.47 | 3.49 | -0.72 |
| Bus | -3.07 | 3.57 | -0.24 | 0.36 | 0.97 | 2.53 | 0.66 | 2.66 | -2.99 | -2.36 | -0.34 | -1.39 | -1.43 | 2.59 | -1.51 |
| FDinner | -1.95 | 4.01 | 0.41 | -1.93 | 3.93 | -2.22 | -1.97 | -0.55 | -2.84 | -2.01 | -1.26 | -2.22 | -1.30 | 2.62 | -2.55 |
| Park | 3.43 | 3.91 | 3.20 | 2.49 | 3.62 | 1.12 | 0.89 | 3.26 | -1.45 | 0.49 | 0.55 | 2.91 | 0.70 | 3.59 | 2.41 |
| Church | -3.13 | -1.22 | -2.13 | -1.66 | -3.13 | -2.74 | -0.99 | -0.93 | -3.89 | -3.09 | -2.59 | -2.80 | -1.38 | -1.91 | -3.18 |
| JInterv | -2.57 | 3.95 | -3.43 | 0.34 | -2.78 | -3.76 | -3.20 | -2.03 | -3.47 | -3.30 | -2.68 | -3.03 | -3.14 | 1.37 | -2.86 |
| Sidewalk | 1.07 | 3.68 | 0.24 | -1.13 | 0.32 | -3.05 | 0.45 | 0.30 | -3.05 | -1.70 | -0.43 | -0.97 | -0.80 | 2.89 | 0.37 |
| Movies | -2.05 | 0.47 | 1.70 | -1.78 | 2.97 | -0.43 | -0.38 | -2.78 | -3.14 | -1.93 | -2.80 | -2.20 | 2.64 | 3.43 | -2.09 |
| Bar | -2.55 | 3.74 | 0.66 | 0.87 | 3.16 | -1.61 | 1.70 | 0.20 | -2.61 | 0.53 | -0.20 | -0.76 | -1.07 | 3.72 | -0.38 |
| Elevator | -2.88 | 2.89 | 0.28 | -1.47 | 0.59 | -3.20 | 0.61 | -0.03 | -2.93 | -1.97 | -1.93 | -2.39 | -1.03 | 2.26 | -2.78 |
| Restroom | -1.68 | 2.74 | -1.70 | -1.05 | -2.16 | -1.68 | 0.53 | 0.24 | -2.74 | 0.61 | -1.03 | -0.86 | 0.28 | 1.39 | -0.99 |
| Own room | 1.64 | 4.07 | 4.01 | 3.78 | 3.43 | 4.34 | 3.16 | 4.07 | -0.26 | 2.30 | 3.01 | 2.22 | 3.49 | 3.66 | 1.93 |
| DLounge | -0.11 | 3.37 | 2.03 | 3.22 | 2.68 | 1.57 | 0.99 | 4.05 | -2.11 | -0.51 | 0.37 | 0.07 | -0.63 | 3.24 | -0.91 |
| FBGame | -0.39 | 3.57 | 0.57 | 0.05 | 3.53 | -1.53 | 0.72 | -0.82 | -2.47 | -0.66 | 0.47 | 2.61 | -0.20 | 3.39 | 3.43 |

Table 2: Additive boxes obtained with Additive Box Algorithm 1.

| Box | Rows | Columns | $\mu$ | Contrib.,% |
|---|---|---|---|---|
| 1 | Park, Own room, Dorm lounge | Talk, Kiss, Write, Eat, Sleep, Read, Laugh | 4.34 | 16.9 |
| 2 | Date, Family dinner, Movies, Bar, Football game | Talk, Kiss, Eat, Laugh | 4.22 | 10.4 |
| 3 | Bus, Job interview, Sidewalk, Elevator, Restroom | Talk, Laugh | 3.95 | 5.0 |
| 4 | Class | Write, Read | 3.66 | 1.7 |
| 5 | Own room | Mumble, Belch, Argue, Jump, Cry, Shout | 3.49 | 3.3 |
| 6 | Park, Football game | Run, Jump, Shout | 3.43 | 2.4 |

Table 3: Additive boxes obtained with Algorithm 2.

| Box | Rows | Columns | $\lambda$ | Contrib.,% |
|---|---|---|---|---|
| 1 | Date, Bus, Park, Sidewalk, Family dinner, Bar, Elevator, Movies, Own room, Dorm lounge, Football game | Talk, Kiss, Eat, Laugh | 2.68 | 26.5 |
| 2 | Class, Bus, Park, Own room, Dorm lounge | Write, Sleep, Read | 2.60 | 8.5 |
| 3 | Class, Date, Job interview, Bar, Park, Restroom, Own room, Football game | Talk, Laugh | 1.46 | 2.8 |
| 4 | Park, Own room | Run, Mumble, Read, Belch, Argue, Jump, Cry, Shout | 1.96 | 5.1 |
| 5 | Football game | Jump, Shout | 3.02 | 1.5 |
| 6 | Movies, Own room | Cry | 2.09 | 0.7 |

Table 3 gives a pattern obtained with the Additive Box Algorithm 2 applied when the optimal intensity weights $\lambda_s$ are required to be positive. The six boxes presented account for 45.1% of the variance of the data in Table 1; additional boxes fitted had successively smaller intensity weights and also very small contributions to the variance, and thus are omitted from Table 3. The solution here is somewhat different from that presented in Table 2, although some clusters are common to the two solutions; specifically, Boxes 2 and 3 from Table 2 correspond to Boxes 1 and 3 from Table 3, respectively, in such a way that their column-sets coincide, and the row-sets of the boxes from Table 3 contain the row-sets of the corresponding boxes from Table 2. Another feature shared by the Tables 3 and 4 is that some clusters contribute to the variance less than do those that follow (see Clusters 4 and 5 in Table 2; 3 and 4 in Table 3); this result contradicts the goal of the sequential fitting procedure which seeks a maximal contribution at each step. That contradiction is generated by the local nature of the Algorithms 1 and 2 in Section 3.2: any cluster is created with a sequential addition/deletion of only one of the entities, beginning with a pair of maximally proximal entities. The results shown in the tables demonstrate that such a greedy procedure generally does not lead to the global maximum of the cluster contribution.

In general, the boxes in Table 3 seem to be reasonable both according to their content and the coverage of the raw proximities. For example, in the first of the boxes, the proximities in submatrix $X(V_1, W_1)$ are much higher than the other values of the corresponding columns, Kiss, Eat, Laugh, with the only two exceptions occurring in the column Talk: the proximity 3.95 to Job Interview (see Table 1) is taken into account in another box (the third), and the low proximity 0.47 to Movies, which is still positive and much greater than all the other proximities in the row Movies (excluding the column Cry, taken into account in Box 6). There is one entry in the first box cluster, Bus/Kiss, which has a rather small proximity value, -0.24 (similarly, the proximity for the entry Class/Sleep in the third cluster equals -0.91). Although it may seem unnatural to have that entry in the first cluster, the presence of Bus/Kiss can be explained by the column Kiss being connected to the other rows of the cluster more tightly than it is disconnected from Bus, so the exclusion of either Bus or Kiss from Cluster 1 will decrease the value of the criterion maximized in (3.9). It appears that our least-squares estimation strategy is heavily dependent on the value of the threshold ($\lambda(V,W)/2$, in this case) and sometimes allows the inclusion of marginal proximity values in the best-fitting solution. In support of this explanation, we note that for the data in Table 3, the inclusion of Kiss in Box 1 gives a better fit compared to its exclusion. Also the Bus/Kiss proximity value of -0.24 in Table 1 is still considerably greater than any of the proximities between Kiss and the row-items not included in the first box. Similar explanations could be offered for other

Table 4: Correspondence-wise boxes with their relative change of probability (RCP) values.

| Box | Rows | Columns | λ(RCP) | Contrib.,% |
|---|---|---|---|---|
| 1 | Job interview | Talk, Write, Laugh | 0.81 | 7.9 |
| 2 | Class | Write, Read, Argue | 0.74 | 7.3 |
| 3 | Park, Sidewalk, Football game | Run, Jump, Shout | 0.41 | 7.4 |
| 4 | Date, Family dinner, Movies | Kiss, Eat | 0.43 | 6.3 |

Table 5: Switching (from row to column product) data on soft drinks from Bass, Pessemier, and Lehmann (1972); estimates of potential movers among loyal (diagonal) consumers are in parentheses.

| Drinks | Coke | 7-Up | Tab | Like | Pepsi | Sprite | DPepsi | Fresca |
|---|---|---|---|---|---|---|---|---|
| Coke | 188(35) | 33.0 | 3.0 | 10.0 | 41.0 | 17.0 | 4.0 | 11.0 |
| 7-Up | 32.0 | 77(14) | 1.0 | 11.0 | 24.0 | 17.0 | 2.0 | 8.0 |
| Tab | 2.0 | 3.0 | 4(1) | 9.0 | 2.0 | 1.0 | 2.0 | 2.0 |
| Like | 4.0 | 7.0 | 4.0 | 7(1) | 11.0 | 2.0 | 6.0 | 5.0 |
| Pepsi | 47.0 | 35.0 | 2.0 | 8.0 | 137(25) | 20.0 | 7.0 | 10.0 |
| Sprite | 8.0 | 13.0 | 2.0 | 5.0 | 11.0 | 23(4) | 2.0 | 6.0 |
| DPepsi | 4.0 | 2.0 | 8.0 | 4.0 | 5.0 | 4.0 | 11(2) | 5.0 |
| Fresca | 17.0 | 7.0 | 4.0 | 8.0 | 11.0 | 8.0 | 5.0 | 15(3) |

possible anomalies in the obtained solution (e.g., the Class/Sleep entry in Box 3 of Table 3).

The boxes shown in Table 3 could be also compared with the results from the analysis provided by Eckes and Orlik (1993) in their Table 5 (see Eckes and Orlik 1993, p. 70), which contains four box clusters, A, B, C, and D, in two forms: as obtained originally without overlap and then as followed up with those authors' overlapping procedure. The clusters refer to the augmented raw data table with a "negative" behavior added for each of the original behaviors, by using complementary values, $9 - x_{ij}$, for the proximities. Only two of the clusters include actual activities not resulting from that augmentation of the data: C and D. From those, Cluster C coincides exactly with Box 1 in Table 3 (and also includes the "inconsistent" Bus/Kiss entry discussed above), and Cluster D almost coincides with Box 4 in Table 3. Still other boxes in Table 3 reveal other connections between situations and behavior.

More sharply delineated segments of the data in Table 1 are found with correspondence-wise box clustering (see Table 4 where the clusters contributing more than 5% to the Pearson contingency coefficient are presented). These boxes correspond to data most deviant from the average values, but the deviance is not great: the connection of the row objects to the column objects in those boxes ranges from 41% (Box 3) to 81% (Box 1) higher than average.

The correspondence-wise box clustering was also applied to data collected by Bass, Pessemier, and Lehmann (1972), which have served to illustrate many data analytic techniques (see, for example, DeSarbo 1982; Arabie, Schleutermann, Daws, and Hubert 1988). The data are usually considered in the form of conditional probabilities. The original contingency table form of these data is presented here (see Table 5). The correspondence-wise box clustering applied ad hoc to the matrix produced only one non-trivial box (that is, a box that was not a dyad of the same brand with itself) consisting of three diet cola drinks: Tab, Like, and Diet Pepsi, both as the row and the column clusters. Although this cluster was repeatedly found in the analyses just cited, the main principle of organization among the drinks was generated by the contrasts of cola versus non-cola and diet versus non-diet. The use of box clustering techniques allows us to display yet another aspect of the data. Brand switching data typically have large values for the elements in the principal diagonal, corresponding to brand-loyal consumers. Colombo and Morrison (1989) have argued that the influence of the diagonal entries should be mitigated when emphasizing the information in the off-diagonal entries. Our normalization is based on a type of statistical independence hypothesis (as in Colombo and Morrison 1989) that replaces the observed diagonal values with those in parentheses (see Table 5). The hypothesis suggests that any nondiagonal entry $p_{ij}$ $(i \neq j)$ can be expressed as $p_{ij} = \alpha f_i g_j$ where $f_i$ is the probability of switching from brand $i$, $g_j$ is the probability of switching to brand $j$, and $\alpha$ is the probability of potential switching behavior (applied to both loyal and nonloyal purchasers). After the unknown values of $f_i$, $g_j$, and $\alpha$ are obtained (with least-squares techniques, for example), the proportion of the potential switchers for any given principal diagonal entry $p_{ii}$ is estimated as $\beta = (\alpha - \Sigma_{i \neq j} p_{ij}) / \Sigma p_{ii}$. These proportions $\beta p_{ii}$ are presented in Table 5 in parentheses beside the diagonal entries.

The results show that there are only three important boxes, accounting for 52.7% of the total sum of weighted squares of the data. These boxes are: Diet Pepsi × Tab (RCP = 4.9, contribution 24.6%), Tab × Like (RCP = 3.58, contribution = 19.05%), and Like × {Tab, Diet Pepsi} (RCP = 1.85, contribution = 9.05%). The same three drinks found here are contained in the nontrivial box obtained using the original data, but the boxes now give more interpretable information: obviously, the consumers do not like any of these drinks, and keep changing the selection as if ever hoping for an acceptable diet entry to appear.

## 6. Conclusion

A model-based approach to two-mode clustering is provided and its theoretical and practical advantages have been shown. Two versions of the

additive box clustering model in (3.1) have been considered: direct (Section 3) and correspondence-wise (Section 4), and a special technique has been developed for fitting this kind of clustering model. The technique is based on a doubly greedy optimization strategy: first, clusters are obtained sequentially (cluster-by-cluster) rather than simultaneously, and second, each cluster is formed incrementally, with an element-by-element addition/deletion strategy; each greedy step maximizes the explained part of the total sum of squares. Such a technique is rather simple computationally, and also allows a formal analysis of its properties. Because of these formal underpinnings, our approach has some technical advantages over the one presented by Eckes and Orlik (1993):

1. The parameters of the algorithms can be set according to theoretical considerations, not heuristically (compare, for example, the arbitrary 80% threshold stopping rule of Eckes and Orlik 1993, with the model-based stopping criterion in Section 3). Although such a theoretical criterion also uses thresholds, such as 1% of the contribution to the variance, these thresholds have a traditional statistical meaning;

2) Interpretation of the clusters becomes much easier (compare, for example, the precise meaning of the RCP intensity weight in (4.3) or the contribution to the total sum of squares
of the data with the very imprecise meaning of the parameter $\mu$ considered by Eckes and Orlik (1993));

3) The problems Eckes and Orlik (1993) noted, namely an arbitrary definition of $\mu$ and the post hoc construction of possibly overlapping clusters, are resolved rather easily in our model-based strategy;

4) The analyses presented in Section 5 also appear more direct and complete in comparison to those of Eckes and Orlik (1993). The additive clustering allows: first, treating the raw data without any special augmentation techniques; second, obtaining a larger number of clusters because they are allowed to overlap from the very beginning. Also, in those specific examples, additive clustering techniques have found the same clusters obtained by Eckes and Orlik (1993) with their error-variance approach, and some additional clusters have been identified, with several other connections between situations and behavior presented (like jumping and shouting at a football game in Table 3), but these advantages might be less evident for other examples.

## References

ARABIE, P., and HUBERT, L. (1992), "Combinatorial Data Analysis," *Annual Review of Psychology, 43*, 169-203.

ARABIE, P., SCHLEUTERMANN, S., DAWS, J., and HUBERT, L. J. (1988), "Marketing Applications of Sequencing and Partitioning of Nonsymmetric and/or Two-mode Matrices," in *Data, Expert Knowledge and Decisions*, Eds., W. Gaul, and M. Schader, Berlin: Springer-Verlag, 215-224.

BASS, F. M., PESSEMIER, E. A., and LEHMANN, D. R. (1972), "An Experimental Study of Relationships between Attitudes, Brand Preference, and Choice," *Behavioral Science, 17*, 532-541.

BRAVERMAN, E.M., KISELEVA, N.E., MUCHNIK, I.B., and NOVIKOV, S.G. (1974), "Linguistic Approach to the Problem of Processing Large Bodies of Data", *Automation and Remote Control, 35*, (11, part 1), 1768-1788.

CARLIER, A., and KROONENBERG, P. M. (1993), "Biplots and Decompositions in Two-way and Three-way Correspondence Analysis," in *Publications du Laboratoire de Statistique et Probabilités*, Université Paul Sabatier, Toulouse (France), 01-93, 1-31.

COLOMBO, R. A., and MORRISON, D. G. (1989), "A Brand Switching Model with Implications for Marketing Strategies," *Marketing Science, 8*, 89-99.

DESARBO, W. S. (1982), "GENNCLUS: New Models for General Nonhierarchical Clustering Analysis," *Psychometrika, 47*, 446-449.

DESARBO, W. S., and DE SOETE, G. (1984), "On the Use of Hierarchical Clustering for the Analysis of Nonsymmetric Proximities," *Journal of Consumer Research, 11*, 601-610.

ECKES, T., (1993), "A Two-mode Clustering Study of Situations and Their Features", in *Information and Classification*, Eds., O. Opitz, B. Lausen, and R. Klar, New York: Springer-Verlag, 510-517.

ECKES, T., and ORLIK, P. (1991), "An Agglomerative Method for Two-mode Hierarchical Clustering," in *Classification, Data Analysis, and Knowledge Organization*, Eds., H.-H. Bock and P. Ihm, Berlin: Springer-Verlag, 3-8.

ECKES, T., and ORLIK, P. (1993), "An Error Variance Approach to Two-mode Hierarchical Clustering," *Journal of Classification, 10*, 51-74.

GREENACRE, M. J. (1984), *Theory and Applications of Correspondence Analysis*, London: Academic Press.

GREENACRE, M., and BLASIUS, J. (Eds.) (1994), *Correspondence Analysis: Recent Developments and Applications*, New York: Academic Press.

HARTIGAN, J.A. (1972), "Direct Clustering of a Data Matrix", *Journal of American Statistical Association, 67*, 123-129.

HARTIGAN, J. A. (1976), "Modal Blocks in Dentition of West Coast Mammals", *Systematic Zoology, 25*, 149-160.

LEBART, L., and MIRKIN, B. (1993), "Correspondence Analysis and Classification," in *Multivariate Analysis: Future Directions 2.*, Amsterdam: Elsevier, 341-357.

LEBART, L., MORINEAU, A., and WARWICK, K. M. (1984), *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices* (E. M. Berry, Trans.), New York: Wiley. (Original work published 1977).

MIRKIN, B. (1987), "Additive Clustering and Qualitative Factor Analysis Methods for Similarity Matrices," *Journal of Classification, 4*, 7-31; (1989) Erratum, *6*, 271-272.

MIRKIN, B., "Clustering for Contingency Tables: Boxes and Partitions", *Statistics and Computing*, in press.

MIRKIN, B., and ROSTOVTSEV, P. S. (1978), "Method to Reveal Associated Subsets of the Variables", in *Models for Socioeconomic Data Aggregation*, Ed., B. Mirkin, Novosibirsk: Institute of Economics Press, 107-112.

NISHISATO, S. (1980), *Analysis of Categorical Data: Dual Scaling and Its Applications*, Toronto: University of Toronto Press.

NISHISATO, S. (1994), *Elements of Dual Scaling*, Hillsdale, NJ: Erlbaum.

PACKER, C.V. (1989), "Applying Row-Column Permutation to Matrix Representations of Large Citation Networks", *Information Processing & Management, 25*, 307-314.

PRICE, R. H. (1974), "The Taxonomic Classification of Behaviors and Situations and the Problem of Behavior-Environment Congruence," *Human Relations, 27*, 567-585.

PRICE, R. H., and BOUFFARD, D. L. (1974), "Behavioral Appropriateness and Situational Constraint as Dimensions of Social Behavior," *Journal of Personality and Social Psychology, 30*, 579-586.

SHEPARD, R. N., and ARABIE, P. (1979), "Additive Clustering: Representation of Similarities as Combinations of Discrete Overlapping Properties," *Psychological Review, 86*, 87-123.