RN 004-69

SOME CONTRIBUTIONS TO THE ANALYSIS OF CATEGORICAL DATA

Jan de Leeuw

august 1969

PSYCHOLOGICAL INSTITUTE / UNIVERSITY OF LEIDEN / THE NETHERLANDS

Chapter I: The data

1.1 Categorical data

In this paper we shall be concerned with the analysis of categorical data. These data arise, for example, if a number of subjects fill in a multiple choice test, or if one single subject judges a set of stimuli on a number of multicategory items. More generally: we deal with a finite number of partitionings of a finite set A. In the first example mentioned above the set A is a set of subjects, in the second example it is a set of stimuli. The partitionings correspond with items, the subsets of a particular partitioning with categories of the corresponding item. Unless otherwise indicated we shall assume that the categories are unordered, that the partitionings are proper (in the sense that the subsets corresponding with the categories are nonvoid, exhaust A, and are pairwise disjoint), that all partitionings consist of an equal number of subsets (i.e. all items have the same number of categories). A has n elements a_i, there are m partitionings Π_j, each partition Π_j contains l subsets A_{js}. Thus

$$A_{js} \cap A_{jt} = \emptyset, \qquad \forall\, j,\ s \neq t \tag{1}$$

$$\bigcup_{s=1}^{l} A_{js} = A. \qquad \forall\, j \tag{2}$$

Unless otherwise indicated the ranges of the subscripts will always be $s,t = 1,\ldots,l$; $i,k = 1,\ldots,n$; $j = 1,\ldots,m$. The corresponding index sets are denoted by L, N, M. Sets consisting of one single element will sometimes, if no confusion is possible, be written as points, i.e. x in stead of $\{x\}$.

1.2 Derived measures and binary matrices

Techniques for analyzing categorical data with unordered categories are relatively new and unknown, notwithstanding the fact that these data are very common. In the past it was necessary to construct dichtomous variates on which an order relation could be defined (yes-no, right-wrong). For

these 0-1 variates there is a large number of analytic techniques available, varying from techniques based on derived measurement (principal component analysis of phi-coefficients or tetrachorics) to more fundamental procedures (such as Coombs-Kao nonmetric factor analysis, Guttman scaling, and Lingoes multiple scalogram analysis). Another possibility is to use the categorical variables in conjunction with measured variates. This is done in the analysis of variance and covariance, and in (canonical) discriminant analysis. But the only tool for people who were confronted with purely nominal data has for a long time been the cross table and the associated chi-square (or some other measure of independence).

1.3 The ICP-matrix

In all basic papers on the multivariate analysis of qualitative data (Guttman 1941, Burt 1950, Lingoes 1968, De Leeuw 1968) the fundamental data matrix is the same. Each item corresponds with an 1 x n matrix X^j, with $x^j_{si} = 1$ iff $a_i \in A_{js}$. Otherwise $x^j_{si} = 0$. On the assumption that we have mutually exclusive and exhaustive categories

$$x^j_{si} x^j_{ti} = 0, \qquad i,\ s \neq t \tag{3}$$

$$\sum_s \sum_i x^j_{si} = n. \tag{4}$$

Of course (3) implies that $(X^j)(X^j)'$ is a diagonal matrix. The elements on the diagonal of this matrix are denoted by d^j_s. Thus d^j_s is the number of elements in the set A_{js}, and

$$\sum_s d^j_s = n. \tag{5}$$

The basic data matrix E is a supermatrix obtained by placing all m matrices X^j beneath each other. In Lingoes (1968) this matrix is called the attribute or trait matrix, in De Leeuw (1968) it is called the Indicator of Cartesian Product or ICP-matrix.

1.4 Belonging and order

There is a close correspondence between the relation of belonging and the order relation. In the first place all items can be reduced in a natural way to binary items, by splitting up

$$\Pi_j = \{A_{j1},\ldots,A_{jl}\} \tag{6}$$

into

$$\begin{aligned} \Pi_{j1} &= \{A_{j1},\ A \mid A_{j1}\}, \\ \Pi_{j2} &= \{A_{j2},\ A \mid A_{j2}\}, \\ &\vdots \\ \Pi_{jl} &= \{A_{jl},\ A \mid A_{jl}\}, \end{aligned} \tag{7}$$

where $X \mid Y$ denotes the complement of Y relative to X, i.e.

$$X \mid Y = X \cup \bar{Y}. \tag{8}$$

On each row of the associated binary ICP matrix E (of order ml x n) a partial order $\gtrdot_q$ can be defined by

$$(\Pi_q,\ a_i) \gtrdot_q (\Pi_q, a_k) \text{ iff } e_{qi} = 0 \wedge e_{qk} = 1, \tag{9}$$

with q=1,...,ml. On the original ICP-matrix the order relation would be

$$(A_{js}, a_i) \gtrdot_{js} (A_{js}, a_k) \text{ iff } x^j_{si} = 0 \wedge x^j_{sk} = 1. \tag{10}$$

Although the data define both kinds of ICP-matrices uniquely, it is in general not possible to reconstruct the structure of items and alternatives from a given ml x n ICP-matrix. By just considering the ICP-matrix we lose information and the information we lose is exactly the fact that some of the rows correspond to (mutually exclusive and exhaustive) categories of an item. We neglect the fact that E is a supermatrix.

Chapter II: Separation by continuous functions

2.1 Introduction

The algorithms we are interested in in this paper make it possible to represent the set A in some kind of space. The key concept in the analysis of categorical data is separation. In a purely set theoretical context two sets can be considered separated iff they are disjoint, but if we want a particular spatial representation more stringent conditions are needed. The concept of separation plays an important role in general topology (it can even be used as a single primitive to define the notion of a topological space). Because we assume that most of our readers are not familiar with this disciplin, we review some of the basic concepts and results in the first paragraphs of this chapter. For this review we relied heavily on the books Kelley (1955), Vaidyanathaswamy (1960), and Mamuzić (1963).

2.2 Some concepts and results from general topology

2.2.1 The closure operator

Suppose X is a set, 2^X is the set of all subsets of X, and τ is a single-valued mapping of 2^X into 2^X. Then τ is called a generalized topology for X and the pair (X, τ) is called a generalized topological space. If τ satisfies

$\tau_1)$ $\quad \tau(\emptyset) = \emptyset$.

$\tau_2)$ $\quad A \subset \tau(A), \qquad \forall\, A \subset X.$

$\tau_3)$ $\quad \tau(\tau(A)) = \tau(A), \qquad \forall\, A \subset X.$

$\tau_4)$ $\quad \tau(A \cup B) = \tau(A) \cup \tau(B), \qquad \forall\, A,B \subset X.$

then τ is called a topology for X and the pair (X, τ) is called a topological space. In this case it follows from τ_2 that $\tau(X) = X$, and $\tau(A)$ is called the closure of A. Suppose β is another single valued mapping of 2^X into 2^X with $\beta(A) = X \mid A$. If o denotes composition of mappings then define $\iota(A) = \beta \circ \tau \circ \beta(A)$. Then also $\tau(A) = \beta \circ \iota \circ \beta(A)$.

If τ satisfies $\tau_1 - \tau_4$ then it follows that $\iota(\phi) = \beta \circ \tau \circ \beta(\phi) = \beta \circ \tau(X) = \beta(X) = \phi$, and $\iota(X) = \beta \circ \tau \circ \beta(X) = \beta \circ \tau(\phi) = \beta(\phi) = X$. By similar arguments it follows that ι satisfies in this case

$\iota_1)$ $\quad \iota(\phi) = \phi.$

$\iota_2)$ $\quad \iota(A) \subset A, \qquad \forall A \subset X.$

$\iota_3)$ $\quad \iota(\iota(A)) = \iota(A), \qquad \forall A \subset X.$

$\iota_4)$ $\quad \iota(A \cup B) = \iota(A) \cup \iota(B), \qquad \forall A,B \subset X.$

We call $\iota(A)$ the interior of A. The set $\varepsilon(A) = \iota \circ \beta(A)$ is called the exterior of A. Thus: $a \in X$ is an exterior point of A iff it is an interior point of the complement of A. The interior boundary of A, $\gamma_i(A)$, is defined as $A \setminus \iota(A)$, the exterior boundary, $\gamma_e(A)$, as $\beta(A) \setminus \iota \circ \beta(A)$. The boundary of a set is the union of its interior and exterior boundary. A point a of a topological space is called an accumulation point of $A \subset X$ if $a \in \tau(A \setminus a)$. The set of all accumulation points of A is called the derived set of A. It is denoted by A'. We illustrate these concepts by a few examples.

Example A: Discrete topology.

Let $\tau(A) = A$ for all $A \subset X$. Evidently τ satisfies $\tau_1 - \tau_4$. For $\iota(A)$ we obtain $\iota(A) = \beta \circ \tau \circ \beta(A) = \beta \circ \beta(A) = A$ for all A. Moreover $\gamma_i(A) = \gamma_e(A) = \gamma(A) = \phi$, and $A' = \phi$ for all A (if $a \in A'$ then $a \in \tau(A \setminus a) = A \setminus a$ which is a contradiction).

Example B: Indiscrete topology.

Let $\tau(A) = X$ for all $A \neq \phi$ and $\tau(\phi) = \phi$. Again $\tau_1 - \tau_4$ are satisfied, $\iota(A) = \beta \circ \tau \circ \beta(A) = \beta(X) = \phi$ for all $A \neq X$ and $\iota(X) = X$. Moreover $\gamma_i(A) = A$, $\gamma_e(A) = \beta(A)$, so $\gamma(A) = X$. Finally $A' = X$ for all $A \neq \phi$.

2.1.2 Open and closed sets

Topological spaces can be defined in other (equivalent) ways. Let X be a set, and $\mathcal{T}$ a family of subsets of X satisfying

$f_1)$ $\quad \emptyset \in \mathcal{F}$.

$f_2)$ $\quad X \in \mathcal{F}$.

$f_3)$ $\quad A_1, \ldots, A_n \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{n} A_i \in \mathcal{F}$.

$f_4)$ $\quad \mathcal{F}' \subset \mathcal{F} \Rightarrow \bigcup_{A \in \mathcal{F}'} A \in \mathcal{F}$.

The members of the family $\mathcal{F}$ are called open sets. Observe that f_4 states that the union of all members of any subfamily of $\mathcal{F}$ is in $\mathcal{F}$, while f_3 requires that the intersection of the members of any finite subfamily is in $\mathcal{F}$. Define $\iota(A)$ as the union of all open sets contained in A. This mapping satisfies $\iota_1 - \iota_4$. It follows that $(X, \mathcal{F}, \tau)$ is a topological space if we let $\tau = \beta \circ \iota \circ \beta$. If $\iota(A) = A$ then A is the union of all open sets contained in A and thus, by f_4, A is open. Conversely, if A is open then the union of all open sets contained in A equals A: $\iota(A) = A$ iff A is open. A set is called closed if its complement $\beta(A)$ is open. If $\beta(A)$ is open then $\iota \circ \beta(A) = \beta(A)$, and $\tau(A) = \beta \circ \iota \circ \beta(A) = \beta \circ \beta(A) = A$. If $\tau \circ \beta(A) = \beta(A)$ then $\iota(A) = A$, so A is open and $\beta(A)$ is closed: A is closed iff $\tau(A) = A$. A closed set contains its derived set: $A' = \tau(A) \mid A$ or $\tau(A) = A \cup A'$. If A is closed then $\tau(A) = A$ and thus $A = A \cup A'$, or $A' \subset A$. If A is open then its interior boundary is empty: $\gamma_i(A) = A \mid \iota(A) = A \mid A = \emptyset$. If A is closed then its exterior boundary is empty: $\gamma_e(A) = \beta(A) \mid \iota \circ \beta(A) = \beta(A) \mid \beta(A) = \emptyset$. The sets $\emptyset$ and X are both open and closed.

Example C: The real line in its natural topology

A set A of real numbers is open if for every point $a \in A$ there exists a number $r > 0$ such that the open interval $(a-r, a+r) \subset A$. Moreover $\emptyset$ is open. It follows that open intervals are open, the complements of closed intervals are open and thus closed intervals are closed. If A is the closed interval $[a,b]$ then the interior of A is the open interval (a,b), whose closure again is A. The derived set of both (a,b) and $[a,b]$ is the set $[a,b]$. This

set is also the boundary for both the open and closed interval. The only sets that are both open and closed are $\emptyset$ and X.

2.2.3 Neighborhoods

We have defined topological spaces using as primitives the closure operator, the interior operator, or the family of open sets. There are other alternatives. We may use the concept of a closed set or the boundary operator as primitives. Another possibility is to use the notion of neighborhood. With each point a of the set X we associate a number of subsets Va of X. These subsets are called neighborhoods of a, and they are chosen in such a way that they satisfy

v_1) Each point has at least one neighborhood.

v_2) Each point is contained in every one of its neighborhoods.

v_3) For each pair of neighborhoods Va and Wa of a there is a Ua $\subset$ Va $\cap$ Wa.

v_4) For each Va there exists a Wa, such that Va contains a neighborhood of each point of Wa.

We shall say that b is a contact point of A iff each Vb has a nonempty intersection with A. Let $\tau(A)$ be the set of all contact points of A. Then τ satisfies $\tau_1 - \tau_4$. Each point is an interior point of each of its neighborhoods, and a set is open in the neighborhood topology defined above iff it contains a neighborhood of each of its points. A neighborhood of a set is the union of the neighborhoods of each of its points.

Example D: Lower-limit topology for the real line

Define as neighborhoods of the real number a all half open intervals $[a,b)$, with $b > a$. Postulates v_1 and v_2 are trivially satisfied. If $[a,b_1)$ and $[a,b_2)$ are two neighborhoods of a with $a < b_1 \leq b_2$ then $[a,b_1) \cap [a,b_2) = [a,b_1)$, and the neighborhood $[a,\frac{1}{2}(a+b_1))$ is contained in $[a,b_1)$. This

verifies v_3. Let $[a,b_1)$ be a neighborhood of a, then each point of $[a,\frac{1}{2}(a+b_1))$ has a neighborhood contained in $[a,b_1)$, so v_4 is also true. It follows that b is a contact point of a set $A \subset X$ iff $b \in A$ or b is a lower limit point of A. The closure of A is the union of A and its lower limits. The closure of the open interval (a,b) is $[a,b)$. Closed intervals are closed. An open interval contains a neighborhood of each of its points and is thus open. An upper limit topology for the real line can be defined in the same way.

2.2.4 Continuity

Consider the topological spaces (X,τ) and (Y,σ) and a mapping f of X into Y. The function f is continuous at a point $a \in X$ iff for each neighborhood $Wb = Wf(a) \subset Y$ a neighborhood $Va \subset X$ can be determined such that $f(x) \in Wb$ for every $x \in Va$. The function f is continuous on $A \subset X$ if it is continuous at each $a \in A$. By considering the definition of open and closed sets in neighborhood spaces, it is obvious that a necessary and sufficient condition for f to be continuous is that the inverse image of every open set in Y is open in X (or, equivalently, the inverse image of every closed set is closed). It can be verified that in the case in which X and Y are real lines in their natural topology this definition is equivalent to the familiar definition used in analysis.

Example E: Suppose the infinite sets X and Y are topologized in the following way: the only closed subsets of X and Y are all finite sets, $\emptyset$, and X and Y themselves. If A is an infinite subset of X (or Y) then the closure of A equals the intersection of all closed sets which contain A. But X is the only closed set that contains A, i.e. $\tau(A) = X$ (or Y). If A is finite, then $\beta(A)$ is infinite, and $\iota(A) = \beta \circ \tau \circ \beta(A) = \beta(X) = \emptyset$. If A is infinite, then $\beta(A)$ may be either finite or infinite. If $\beta(A)$ is finite (i.e. A equals X except for a finite number of points) then $\iota(A) = \beta \circ \tau \circ \beta(A) = \beta \circ \beta(A) = A$, if both A and $\beta(A)$ are infinite then

$\iota(A) = \beta \circ \tau \circ \beta(A) = \beta(X) = \emptyset$. The only open sets are $\emptyset$, X, and the sets A that are equal to X except for a finite number of points. The only sets that are both open and closed are $\emptyset$ and X, the sets A for which both A and $\beta(A)$ are infinite are neither open nor closed. Clearly all one-one mappings of X onto Y are continuous.

2.2.5 Separation axioms

Topological spaces can be classified by several conditions called sparation axioms. We list the most important of them.

T_0: For any two distinct points $a,b \in X$ either $a \notin \tau(b)$ or or $b \notin \tau(a)$ or both.

T_1: For any two distinct points $a,b \in X$ both $a \notin \tau(b)$ and $b \notin \tau(a)$.

It follows that $\tau(a) = a$ for all points $a \in X$ if T_1 is satisfied.

T_2: For any two distinct points $a,b \in X$ there exist disjoint neighborhoods Va and Vb.

T_3: For any closed set $A \subset X$ and any point $b \in A$ there exist disjoint neighborhoods V(A) and Wb.

Evidently $T_3 \rightarrow T_2 \rightarrow T_1 \rightarrow T_0$. A topological space that satisfies axiom T_i is called a T_i-space. The remaining two separation axioms are for T_1-spaces only.

T_4: For any two disjoint closed sets A and B of a T_1-space X there exist disjoint neighborhoods V(A) and V(B).

T_5: For any two sets in a T_1-space X with $A \cap \tau(B) = B \cap \tau(A) = \emptyset$ there exist disjoint open neighborhoods O(A) and O(B).

Of course $T_5 \rightarrow T_4$. A T_{13}-space is a space which satisfies both T_1 and T_3. Such a space is called a regular space. A T_{14}-space is called a normal space, and a T_{15}-space a completely normal space, because in a T_{15}-space every subspace is a normal (T_{14}) space.

S_1: Each nonempty closed set $B \subset X$ and each point $a \notin B$ can be

functionally separated, in the sense that there is a continuous function f of (X,τ) into $[0,1]$ with $f(a) = 0$ and $f(b) = 1$ for each $b \in B$.

A T_1S_1-space is called completely regular, $T_1S_1 \to T_{13}$.

S_2: Each pair of disjoint closed sets $A,B \subset X$ can be functionally separated ($\exists$ f: $X \to [0,1]$, f continuous, $f(a) = 0$ $\forall a \in A$, $f(b) = 1$ $\forall b \in B$).

Trivially $T_1S_2 \to T_1S_1$, but also $T_1S_2 \leftrightarrow T_{14}$. We restate this last result as: In a T_1-space every pair of disjoint closed sets can be functionally separated iff the space is normal. Thus every normal space is completely regular. A very thorough investigation of the relation between the different separation axioms has been carried out by Van Est and Freudenthal (1951).

2.2.6 (Pseudo)-metric space

Suppose d is a single-valued mapping of X^{*X} into the positive reals satisfying

d_1: $d(a,a) = 0$, $\forall a \in X$

d_2: $d(a,b) = d(b,a)$, $\forall a,b \in X$

d_3: $d(a,c) \geqslant d(a,b) + d(b,c)$. $\forall a,b,c \in X$

We shall denote the open spheres around a point a by $S(a,r)$. Thus $S(a,r) = \{x \mid d(a,x) < r\}$ where r is a positive real number. The neighborhoods of a are all sets $S(a,r)$. These neighborhoods satisfy v_1-v_4, so (X,d,τ) is a topological space. All functions that satisfy d_1-d_3 are called pseudometrics. Let $D(A,B) = \inf\{d(x,y) \mid x \in A, y \in B\}$. Then it is easy to see that a is a contact point of B (i.e. $a \in \tau(B)$) iff $D(a,B) = 0$. Thus $\tau(a) = \{x \mid d(x,a) = 0\}$. The dismeter δ of a set $A \subset X$ is defined as $\sup\{d(x,y) \mid x,y \in A\}$. In any pseudo-metric space $D(A,B) = D(\tau(A), \tau(B))$, and $\delta(A) = \delta(\tau(A))$. A pseudometric space satisfies T_4. If d satisfies in addition

d_4: $d(a,b) = 0 \Rightarrow a=b$, $\forall a,b \in X$

then d is called a metric, and (X,d,τ) a metric space. In a metric space

$\tau(a) = \{ x \mid d(x,a) = 0 \} = a$, i.e. a metric space is T_{14} or normal. The metric can be restricted to subspaces of X while remaining a metric, and thus a metric space is also a completely normal space.

Example F: Let X be any set, $d(a,b) = 1$ iff $a \neq b$, otherwise $d(a,b) = 0$. Then d satisfies d_1-d_4. For the closure of A we obtain $\tau(A) = \{ x \mid D(x,A) = 0 \}$. Now $D(x,A) = 0$ iff $\exists b \in A \ni d(x,b) = 0$ iff $x \in A$. Thus $\tau(A) = A$, and $\iota(A) = \beta \circ \tau \circ \beta(A) = \beta \circ \beta(A) = A$. All subsets of X are both open and closed. The topology induced by the metric d is equivalent to the topology considered in example A. Or: the topological space in example A is <u>metrizable</u>.

Example G: Let X be the set of all pairs of real numbers: $X = \{ (y,z) \mid y \in Re, z \in Re \}$. Each point x corresponds with a pair (y,z). Define $d(x_i,x_j) = |y_i - y_j|$. Then d is a pseudo-metric for X. In this space the closure of a point x is the set of all points that have the same first coordinate as x. If we let $d(x_i,x_j) = \sqrt{(y_i - y_j)^2 + (z_i - z_j)^2}$ then d is a metric and the closure of x is x. In this metric space circles play the same role as intervals in the natural topology for the real line.

2.3 <u>Separation of sets in metric spaces (binary items)</u>

Suppose we have a representation of our stimulus set A in some kind of topological space (X, τ), and suppose $\overline{\Pi}$ is a partitioning of A into two subsets B and C. We shall say that our representation of the item (partition) is weakly contiguous (WC) iff there is a continuous real valued function ϕ that satisfies

ϕ_1: if $x \in B$ and $y \in C$ then $\phi(x) < \phi(y)$.

A set of items has a WC-representation if all items have one. Suppose (X, τ) is a normal space. Then sets consisting of a single point are closed (by T_1), the union of a finite number of closed sets is closed, so B and C are closed. By T_4 there exists a continuous ϕ with $\phi(x) = 0$ for all $x \in B$ and $\phi(y) = 1$ for all $y \in C$. This means that a sufficient

condition for an item to have a WC-representation is that the space in which we represent A is normal. It also means that requiring a WC-representation poses no constraints on the data, only on the space. In a metric space a WC-representation is always possible. We proceed to construct an explicit expression for ϕ in this case.

Lemma 2.3.1: If A is a nonvoid subset of a pseudo-metric space $(X,d,\mathcal{T})$ then $D(x,A)$ is a continuous function of x.

Proof: Take any two points x and $y \in X$, and let z be a point of A such that $D(y,A) = d(y,z)$. Such a point exists if A is nonvoid, though it may not be unique. We have $d(x,z) \leq d(x,y) + d(y,z) = d(x,y) + D(y,A)$. By definition $D(x,A) \leq d(x,z)$ and thus

$$D(x,A) \leq d(x,y) + D(y,A). \tag{1}$$

In the same way z' is a point of A such that $D(x,A) = d(x,z')$. Then $d(y,z') \leq d(x,y) + d(x,z') = d(x,y) + D(x,A)$, and, a fortiori,

$$D(y,A) \leq d(x,y) + D(x,A). \tag{2}$$

Combining (1) and (2) yields

$$\left| D(x,A) - D(y,A) \right| \leq d(x,y). \tag{3}$$

If y lies in the open sphere $S(x,r)$ then $|D(x,A) - D(y,A)| < r$, and thus $D(y,A)$ lies in the open interval $(D(x,A) - r, D(x,A) + r)$. Consequently $D(x,A)$ is continuous for all (nonvoid) A. Q.E.D.

By using this lemma, the fact that for disjoint B and C the function $d(x,B) + d(x,C)$ never vanishes, and the familiar rules on the continuity of combinations of continuous real valued function we obtain the following

Theorem 2.3.2: The function

$$f(x) = \frac{D(x,B)}{D(x,B) + D(x,C)} \tag{4}$$

is real valued, continuous, and satisfies ϕ_1.

If λ is a real number on the open interval (0,1) and $f_\lambda = \{x \mid f(x) = \lambda\}$, then each continuous path from a point of B to a point of C contains at

least one point of f_λ. In the case of an Euclidian (or, more generally, a normed linear topological space) the line segment $\gamma x + (1 - \gamma)y$ (with $0 \leq \gamma < 1$, $x \in B$, and $y \in C$) intersects f_λ in at least one point. In the examples that follow we have taken (X, d, τ) as the real plane, d as the ordinary Euclidean distance, and $\lambda = \frac{1}{2}$. In each example the set $f_{\frac{1}{2}}$ is drawn. Of course $f(x) = \frac{1}{2}$ iff $d(x,B) = d(x,C)$ and the set $f_{\frac{1}{2}}$ is piecewise linear (a combination of line segments). The examples are shown in figure Ia-i. Some important conclusions are the following: there are other continuous functions besides f that satisfy ϕ_1. The sets B and C can be separated by a straight line, for example, in figures Ia, Id, Ie, and If. Only in the case Ic the function f is a straight line too. On the other hand there may be reasons to prefer this function to the straight line. In example If the deviation from a straight line boundary expresses a significant feature of the representation, in example Ib the line through b_1, c_1, and b_5 'weakly' separates the two sets but there are reasons to prefer our separating function. Define

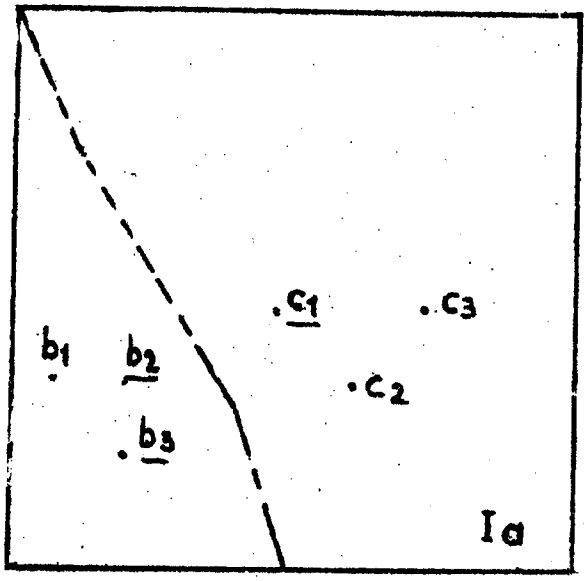
$$S_< = \left\{ x \mid f(x) < \tfrac{1}{2} \right\}, \tag{5a}$$

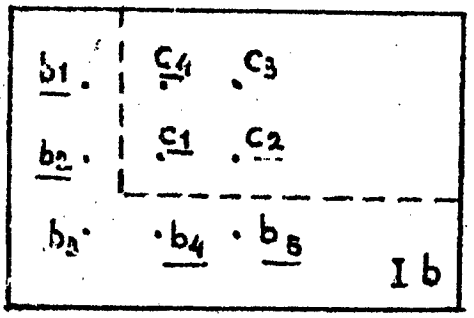
$$S_> = \left\{ x \mid f(x) > \tfrac{1}{2} \right\}. \tag{5b}$$

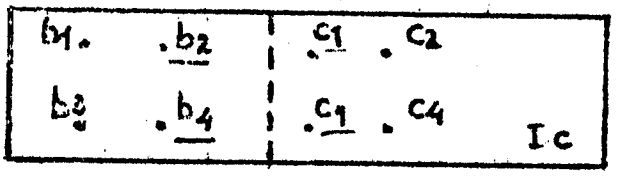
In examples Ia, Ib, Ic, Ie, If, Ih, Ii both $S_<$ and $S_>$ are connected. In example Ig the set $S_<$ is connected, but $S_>$ has two components. In Id both sets are disconnected, $S_<$ has three components and $S_>$ two. Although the function f always satisfies ϕ_1 we are sometimes dissatisfied with its performance as a separator. In later sections we shall try to give a quantitative measure of dissatisfaction and a way to optimize satisfaction. A very interesting attempt to do just this was developed by Guttman. His approach will be discussed in the next section.

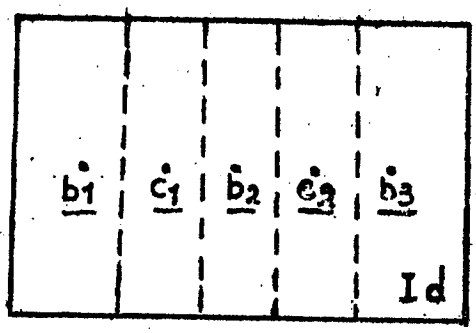
2.4 The MSA-I rationale

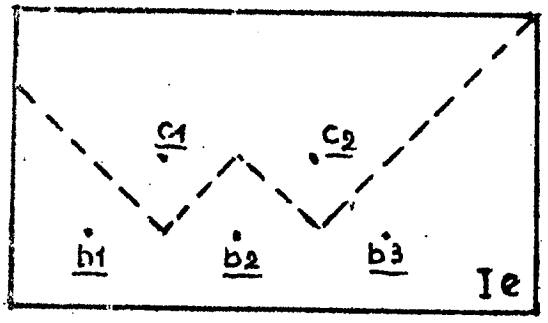
The mathematical rationale of Guttman's multidimensional scalogram analysis (MSA) is still largely unpublished, all we have to work with are some

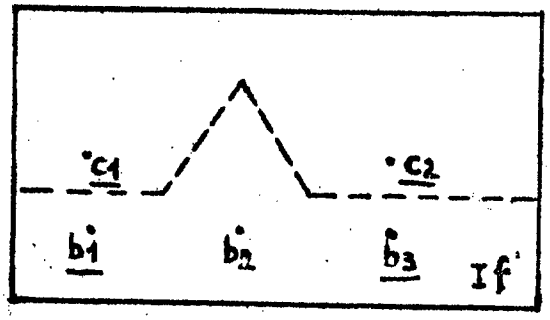

c1
c3
b1
b2
c2
b3
Ia

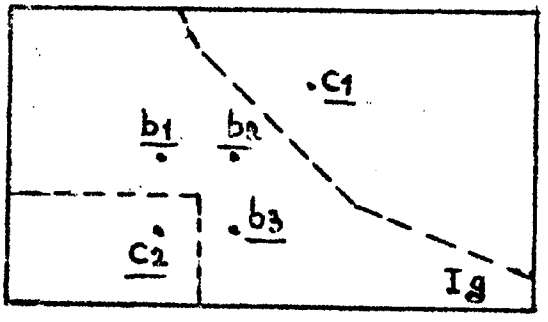

b1
c4
c3
b2
c1
c2
b3
b4
b5
Ib

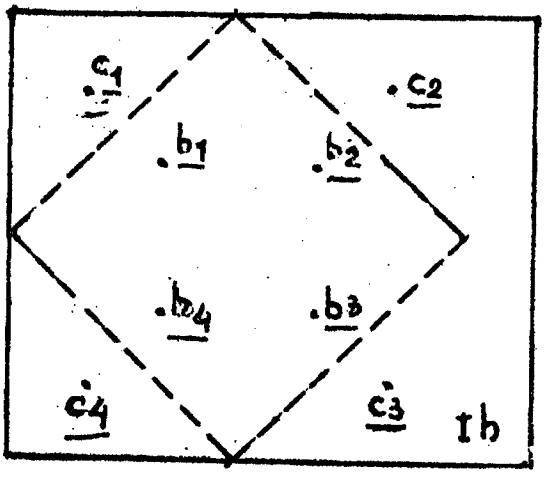

b1
b2
c1
c2
b3
b4
c1
c4
Ic

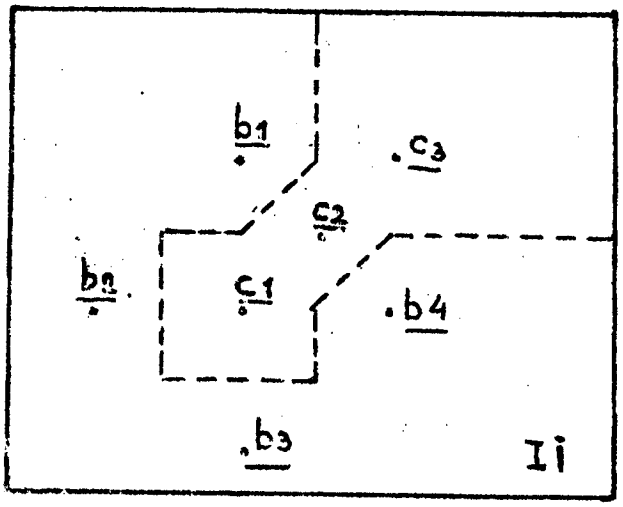

b1
c1
b2
c2
b3
Id


c1
c2
b1
b2
b3
Ie


c1
c2
b1
b2
b3
If


c1
b1
b2
b3
c2
Ig


c1
c2
b1
b2
b4
b3
c4
c3
Ih


b1
c3
c2
b2
c1
b4
b3
Ii

examples (Israel Institute for applied social reseach, mimeographed), and user-oriented descriptions of the computer program and definitional system by Lingoes (1968, 1969). We shall treat the general case: an item with $l \geqslant 2$ mutually exclusive and exhaustive categories $A_1,\ldots,A_l$. We work in a pseudometric space (X,d,τ). Each category is partitioned into two sets: the inner-points A_s^i and the outer-points A_s^o. Outerpoints are defined as follows. Let a be a point of $A \mid A_s$ and let b be the point of A_s such that $d(a,b) = \inf\{d(a,x) \mid x \in A_s\} = D(a,A_s)$. Such a point b is called an outer point of A_s, i.e. $b \in A_s^o$. More formally:

$b \in A_s^o$ iff $b \in A_s \wedge \exists\, a \in A \mid A_s \Rightarrow \forall\, x \in A_s$ it is true that $d(a,b) \leqslant d(a,x)$.

Evidently $A_s^o \neq 0$ for all s (assuming, of course, nonempty categories). A_s^i is defined as $A_s \mid A_s^o$, and A_s^i may very well be empty. We shall say that category s has a MSA^I-contiguous representation iff $\forall\, a \in A_s^i$ it is true that $\exists\, b \in A_s^o \Rightarrow \forall\, c \in \bigcup_{t \neq s}^{l} A_t^o$ it is true that $d(a,b) < d(a,c)$. In words: iff each inner point of category s is closer to some outer point of this category than it is to any outer point of a different category. If the category s has no inner points we shall say that the MSA^I- problem is undecided for that category. An item has a MSA^I-contiguous representation if all its categiries have, a set of items if all items have one. In the examples in figure Ia-i we have underlined the outer points. Items Ia, Ib, and Ic have an MSA^I-contiguous representation, in items Id, Ie, Ig, Ih the MSA^I-problem is undecided for all categories, in examples If and Ii the problem is undecided for one of the two categories while the other category has a contiguous representation. Of course partly our difficulties (having no inner points) are due to our small scale examples with coordinates in the lattice points of the plane.

Because we work in a pseudometric space all items that have a MSA^I-contiguous representation also have a WC-representation. It is easy to construct

representations, however, in which the categories can be separated by a straight line but are not MSA^I-contiguous, and vice versa. In figure IIa,b the outer points again are underlined. In figure IIa there is a straight line separating B and C, but clearly (using the ordinary Euclidean metric) $d(c_2,b_4) < d(c_1,c_2)$ and category $\{C\}$ has no MSA^I-contiguous representation. As a digression, consider the positive symmetric function

$$d(x,y) = \sqrt{|x_1 - y_1|} + \sqrt{|x_2 - y_2|}. \tag{6}$$

This function defines a metric. Proof: d_1, d_2, and d_4 are obviously satisfied. We proof d_3. Triangle inequality for Euclidean metric on the first axis:

$$|x_1 - y_1| \leq |x_1 - z_1| + |y_1 - z_1|, \tag{7}$$

thus, a forteriori,

$$|x_1 - y_1| \leq |x_1 - z_1| + |y_1 - z_1| + 2\sqrt{|x_1 - z_1||y_1 - z_1|}, \tag{8}$$

or

$$\left(\sqrt{|x_1 - y_1|}\right)^2 \leq \left(\sqrt{|x_1 - z_1|} + \sqrt{|y_1 - z_1|}\right)^2, \tag{9}$$

or

$$\sqrt{|x_1 - y_1|} \leq \sqrt{|x_1 - z_1|} + \sqrt{|y_1 - z_1|}. \tag{10}$$

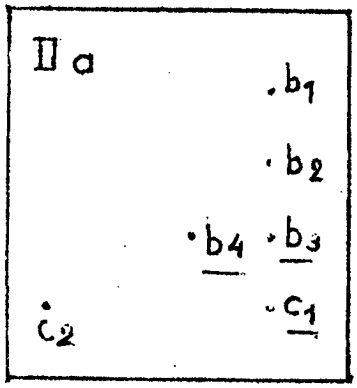
The same reasoning for the second axis gives

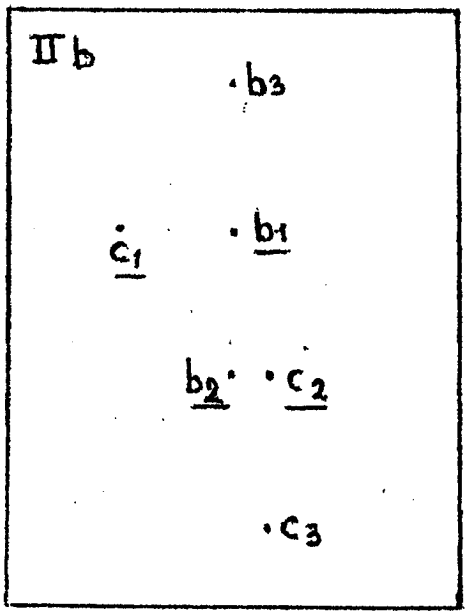
$$\sqrt{|x_2 - y_2|} \leq \sqrt{|x_2 - z_2|} + \sqrt{|y_2 - z_2|}. \tag{11}$$

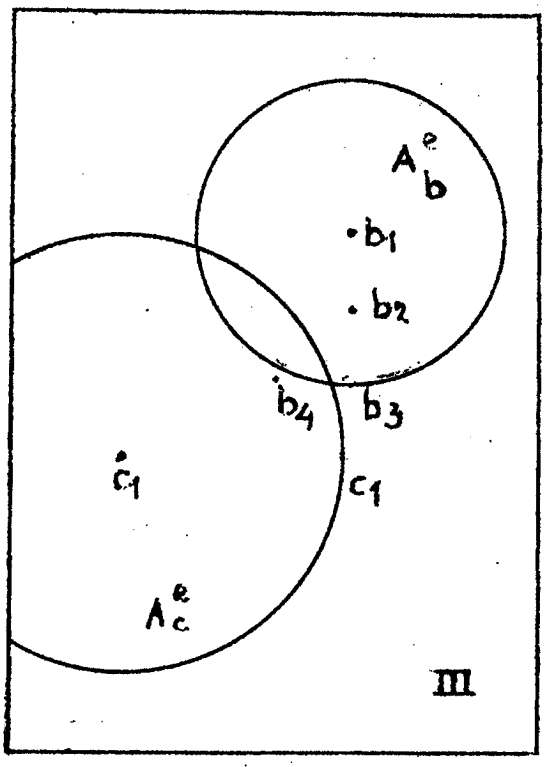
Adding (10) and (11) gives the required result, i.e. d defined by (6) is a metric. The distances, using this metric, are

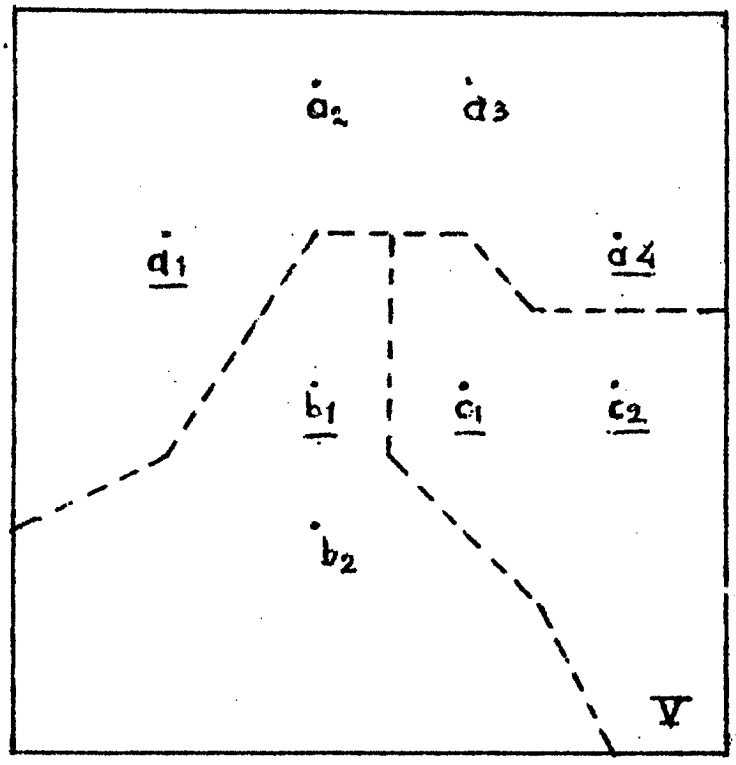
b_1	0	1	$\sqrt{2}$	$1+\sqrt{2}$	$\sqrt{3}$	$2\sqrt{3}$
b_2		0	1	2	$\sqrt{2}$	$\sqrt{2}+\sqrt{3}$
b_3			0	1	1	$1+\sqrt{3}$
b_4				0	2	$1+\sqrt{2}$
c_1					0	$\sqrt{3}$
c_2						0

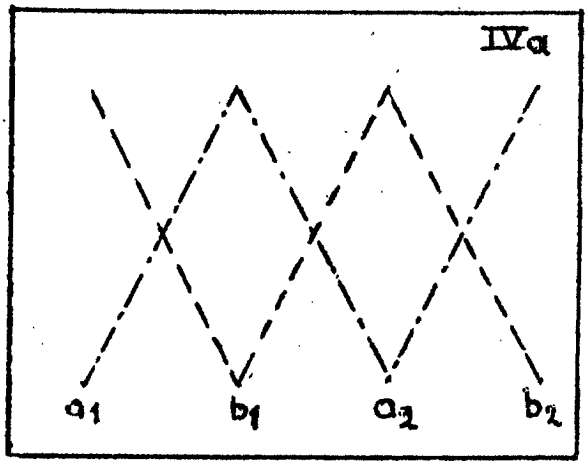
It follows that the outer points, using this metric, are the same, but now $d(b_1,b_4) > d(b_1,c_1)$ and $d(b_2,b_4) > d(b_2,c_1)$, while $\{C\}$ has an MSA^I-contiguous representation. There is little doubt that there exist metrics

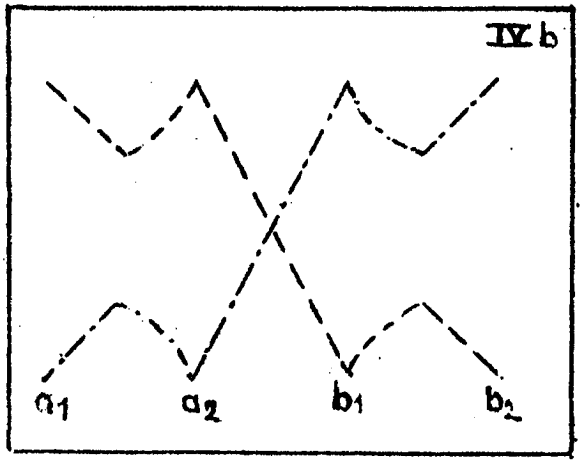

II a
b_1
b_2
b_4
b_3
c_1
c_2


II b
b_3
c_1
b_1
b_2
c_2
c_3


A^e_b
b_1
b_2
b_4
b_3
c_1
c_1
A^e_c
III


a_2
a_3
a_1
a_4
b_1
c_1
c_2
b_2
V


IV a
a_1
b_1
a_2
b_2


IV b
a_1
a_2
b_1
b_2

for which both sets are MSA^I-contiguous and others for which they both are not. In the example of figure IIb the sets B and C cannot be separated by a straight line (a proof is given in a later chapter), but the outer-inner-point structure shows that they are MSA^I-contiguous in the Euclidean case.

Let us suppose that category A_s has inner points. For each $a \in A_s^i$ we find the closest outer point $b \in A_s^o$, and we define the closed sphere $T(a,b) = \mathring{c}(S(a,b)) = \{x \mid d(a,x) \leq d(a,b)\}$. The MSA^I-extension of A_s^i, written as A_s^e, is defined as the closed set

$$A_s^e = \bigcup_{a \in A_s^i} T(a,b). \qquad (12)$$

Observe that $A_s^i \subset \mathring{c}(A_s^e)$ and $\mathring{c}_i(A_s^e) \cap A_s^o \neq \emptyset$. An example is given in figure III. In a metric (T_{14}) space the closed sets A_s^o and $\bigcup_{t \neq s} A_t^o$ can be separated by a continuous function iff they are disjoint. If A_s^e and $\bigcup_{t \neq s} A_t^o$ are disjoint, then A_s^e and $\bigcup_{t \neq s} A_t^i$ are disjoint too (if A_s^e contains an inner point b of another category t, then b must lie in one of the spheres $T(a,b)$; because b is an inner point of t there is an outer point c closer to a then b, which implies that $c \in T(a,b) \subset A_s^e$). The intersection of A_s^e and A_t^e does not have to be empty, necessary and sufficient for MSA^I-contiguity of both categories is that is does not contain any points of $A_s \cup A_t$. Although it is, of course, possible to give a rigourous definition of MSA^I-contiguity and a satisfactory error theory, it is rather difficult to see what MSA^I-contiguity implies in terms of separating functions (or, equivalently, in terms of what regions are considered contiguous). Moreover we have some objections to the MSA^I-algorithm that are similar to our objections to the algorithms in the SSA-series (De Leeuw 1969), and even somewhat more pertinent. Nevertheless the MSA-I program solves a number of problems in a very satisfactory way. For items with binary responses there is perfect recovery of the underlying circumplex or radex. The Coombs-Kao disjunctive-

conjunctive structure with rectangular boundaries comes out perfectly. Although the definition of MSA^I-contiguity is somewhat arbitrary from a purely theoretical point of view, the results are often satisfactory from a practical point of view. Moreover, because of the formulation of the requirements in terms of a particular metric and not of a separating boundary, generalisation from two to more than two categories is obvious.

2.5 Weak contiguity (1-ary items)

The MSA^I-rationale generalises easily to items with $1 > 2$. For the approach that uses the function f the generalization is less immediate. Consider the 1 functions

$$f_s(x) = \frac{D(x,A_s)}{D(x,A_s) + D(x,A \mid A_s)} \qquad (13)$$

separating A_s from $A \mid A_s$, and the associated regions $S_s = \left\{ x \mid f_s(x) < \frac{1}{2} \right\}$. We shall proof the following

Theorem 2.5.1: The 1 regions S_s are disjoint.

Proof: By definition $x \in S_s$ iff $f_s(x) < \frac{1}{2}$ iff $D(x,A_s) < D(x,A \mid A_s)$. Because $D(x,B) = \inf \left\{ d(x,y) \mid y \in B \right\}$ it is always true that if B is partitioned into subsets B_v then $D(x,B) = \inf_v \left\{ D(x,B_v) \right\}$. Thus $D(x,A \mid A_s) = \inf \left\{ D(x,A_t) \mid t \in L \mid s \right\}$. It follows that $x \in S_s$ iff $D(x,A_s) < \inf \left\{ D(x,A_t) \mid t \in L \mid s \right\}$. Suppose that there are $s \neq r \in L$ such that there is an $x \in S_s \cap S_r$. Then $x \in S_s$ implies that $D(x,A_s) < D(x,A_r)$ and $x \in S_r$ implies that $D(x,A_r) < D(x,A_s)$. This contradiction establishes the theorem.

Theorem 2.5.2: Let $T_s = \left\{ x \mid f_s(x) \leq \frac{1}{2} \right\}$. Then the union of the T_s exhausts the whole space.

Proof: Using the proof of the previous theorem: $x \in T_s$ iff $D(x,A_s) = \inf \left\{ D(x,A_t) \mid t \in L \right\}$, where it is possible that for $r \neq s$ $D(x,A_r) = D(x,A_s)$. For each point $x \in X$ compute all values $D(x,A_t)$ and select one $s \in L$ such that $D(x,A_s) \leq D(x,A_t)$ for all $t \in L$. Then $x \in T_s$. Thus all points $x \in X$

can be placed in at least one of the T_s (they are elements of exactly one T_s iff they an element of an S_s). This completes the proof of the theorem. It follows that the sets $S_1,\ldots,S_l, \bigcup_{s=1}^{l} T_s \mid S_s$ constitute a partitioning of X.

Corrolary 2.5.3: If the functions f_s are defined on the Euclidean space X with Euclidean metric d and if they are all hyperplanes, then they are parallel.

2.6 Smooth separating boundaries

In section 2.3 we have expressed our satisfaction with some examples and our dissatisfaction with others. Moreover we have shown in 2.3 and 2.5 that requiring weak contiguity with 1-ary items is trivial in a metric space. Nevertheless the properties of the function f_s proved in the previous section are sufficient important to hold on to it somewhat longer.

Consider the examples IVa and IVb on the real line (with Euclidean metric d). In both functions we have drawn graphs of the functions f_1 and f_2. The separating boundary, where $f_1(x) = f_2(x) = \frac{1}{2}$, is more satisfactory in the second example. It is clear that the degree of jaggedness and connectedness of these boundaries depends on the smoothness of $f_1(x)$ and $f_2(x)$. Quantitative measures of smoothness have been proposed in psychometric literature by Carroll (1963a,b), Carroll & Chang (1964), Shepard (1964), Shepard & Carroll (1966), Tucker (1966), and De Leeuw (1968). The paper De Leeuw (1968) contains an extensive list of references to both the psychometric and the statistical literature.

Our problem thus becomes to find a mapping of A into the metric space $(X,d,\mathcal{T})$ such that all functions $f_s(x)$ are as smooth as possible. This must be true for all categories and all items, i.e. there are ml functions $f_s(x)$. The first remark must be that the value of $f_s(x)$ on the points of A is either zero or one, no matter how smooth or jagged $f_s(x)$ is. In our

definition of smoothness the points of A cannot possibly play a role. Therefor we need a (finite) set of auxilary points Z. Z contains, say, p points $z_1,\ldots,z_p$. The ml x p matrix F contains all values of $f_{js}(z_p)$. Define

$$K_{js} = \sum_{i \neq}^{p} \sum_{j}^{p} \left[\frac{\{f_{js}(z_i) - f_{js}(z_j)\}^2}{d(z_i,z_j)^2} \right] / \left[\sum_{i \neq}^{p} \sum_{j}^{p} \{f_{js}(z_i) - f_{js}(z_j)\}^2 \right], \tag{14}$$

$$K_j = \sum_{s=1}^{l} K_{js}, \tag{15}$$

and

$$K = \sum_{j=1}^{m} K_j, \tag{16}$$

as our inverse measure of smoothness of the representation. The index K is similar to the coefficients used by Carroll & Chang (1964), Shepard & Carroll (1966), and De Leeuw (1968). Our algorithmic problem is to find a representation of A such that, (for fixed Z) K becomes a minimum.

2.7 Separation by functions of degree n (binary items)

The discussion in this section refers exclusively to binary items and to Euclidean multidimensional space. It applies equally well, of course, to items that can be reduced to binary items by the method of section 1.4. Incidentally, this method can be applied whenever the investigator feels that its use will cause no serious loss of information. This may be true, for example, if the items is designed as a short way to ask a number of (binary, yes-no) items at the same time. In stead of using this syntactical (logical) method, we can, of course, also use a sematic method in which we pool several categories on the basis of an external criterion and end up with only two categories (for example right and wrong). We use Weierstras theorem for functions of sevral variables (cf Hobson, 1957, II, p232: 'A continuous function af any number of variables, defined in a given closed

cell, is such that a finite polynomial in the variables exists which differs from the function by less than a prescribed positive number, at all points of the cell'). For our purpose we shall use the following series of (continuous) functions:

$$f^0(x) = d, \tag{17a}$$

$$f^1(x) = \sum c_i x_i + d, \tag{17b}$$

$$f^2(x) = \sum\sum b_{ij} x_i x_j + \sum c_i x_i + d, \tag{17c}$$

$$f^3(x) = \sum\sum\sum a_{ijk} x_i x_j x_k + \sum\sum b_{ij} x_i x_j + \sum c_i x_i + d, \tag{17d}$$

and so on.

The superscripts used in (17) denote the degree of the function (the highest possible power of x). Observe that we may require, without loss of generality,

$$b_{ij} = b_{ji}, \tag{18a}$$

$$a_{ijk} = a_{ikj} = a_{jik} = a_{jki} = a_{kij} = a_{kji}, \tag{18b}$$

and so on.

Suppose item j has categories B and C. We shall say that the representation is contiguous of degree p iff there is a function of degree p such that

$$f^p(x) < 0 \qquad \forall x \in B, \tag{19a}$$

$$f^p(x) > 0 \qquad \forall x \in C. \tag{19b}$$

If an item is contiguous of degree p then it is contiguous of degree r with $r > p$. Moreover, by Weierstrass' theorem, if we take p larger and larger we can approximate the function f of (4) arbitrary close. Consequently for each representation there is a finite integer p such that the representation is contiguous of degree r for all $r \geq p$. Clearly contiguity of degree p implies weak contiguity for all p.

Generalization to l-ary items seems difficult. Of course we can construct l functions f_s of degree p that separate $A \mid A_s$ for all $s \in L$. But the conditions that guarantee that the regions $S_s = \{x \mid f_s^p(x) < 0\}$ are a partitioning of the space X are quite complicated. Moreover there is no

reason why all functions f_s should be of the same degree. This is, of course, also true if we seek a representation of a set of binary items. Therefor a set of binary items is called contiguous of degree $(p_1,p_2,\ldots,p_m)$ if item 1 is contiguous of degree p_1, item 2 of degree p_2, and so on. In general this complicates the problem because in our optimization problems the number $\bar{p} = \frac{1}{m}\sum p_j$ is an additional quantity to be minimized, and as such it is quite difficult to manipulate by purely analytical (non-heuristic, non-enumerative) methods. The ide of minimizing $\sum p_j$ is directly related to requiring 'smooth' separating boundaries (of Shepard 1964). The special case of contiguity of degree $(p_1,p_2,\ldots,p_m)$ with $p_j = 1$ for all j is treated in the following chapter from a different and more general point of view and under a different name. Some special cases of contiguity of degree $(p_1,\ldots,p_m)$ with $p_j = 2$ for all $j \in M$ are discussed in chapter IV. The problem of finding a representation that is contiguous of degree $(p_1,p_2,\ldots,p_m)$ is related to the problem of nonmetric discriminant analysis. In that case, howver, the only unknowns are d, c_i, b_{ij}, a_{ijk} and so on, while the coordinates (the x-values) are known.

Chapter III: Separation by hyperplanes in linear spaces

3.1 Introduction

The most simple separating boundary is a straight line, or, more generally, a hyperplane. In this chapter we develop a theory of contiguity based on separation by hyperplanes. In order to develop these theories more fully we need some basic results on linear topological spaces. The relevant books are Kelley et al (1963), Day (1959), Bourbaki (1953,1955). Moreover we need a number of results on convex sets. These can be found in the books of Bonnesen & Fenchel (1939), Fenchel (1953), Eggleston (1958), and especially Valentine (1964).

3.2 Some basic concepts

3.2.1 Algebraic concepts

3.2.1.1 Real linear spaces

A real linear space is a system $(X,+,.)$ with X a nonempty set, and '+' and '.' binary operations:

$$+ : X \times X \to X, \tag{1a}$$

$$. : Re \times X \to X. \tag{1b}$$

The elements of X are called vectors, + is called addition and . scalar multiplication. The operation + is defined in such a way that $(X,+)$ is an abelian group (i.e. it is closed under addition, addition is commutative and associative, there is a zero element, and each element has an inverse). If no confusion with the real number 0 is possible then the zero of X (the origin) will also be written as 0, otherwise we write 0_X, 0_{Re}, and so on. Scalar multiplication is distributive with respect to addition in X and with respect to addition of real numbers. Moreover it is associative, and $1.x = x$ for all $x \in X$. Thus multiplication by a fixed scalar is an endomorphism of the group $(X,+)$. A subset $A \subset X$ is called linearly independent if every finite linear combination $\sum_{\alpha_i} x_i$ of elements x_i of A equals zero iff all α_i are zero. A subset $A \subset X$ is called a (Hamel) basis

for X iff each element of X can be represented in a unique way as a linear combination of the elements of A. All Hamel bases for a linear space have the same cardinal number, this number is called the dimension of the space. If A and B are subsets of X then A + B denotes $\{x \mid x = a + b, \text{ ac } A, b \in B\}$. The set $\{x\} + A$ is called a translation (translate) of A. A real linear space $(Y,+,.)$ is a linear subspace of $(X,+,.)$ iff $Y \subset X$ and the operations + and . in Y coincide with those in X. Two linear subspaces Y and Z (of course this is an inaccurate short cut notation) are called complementary iff $Y + Z = X$ and $Y \cap Z = 0_X$. The rank (co-dimension, deficiency) of Y in X is the dimension of a subspace of X that is complementary to Y in X. A hyperplane is a subspace of rank one. If X and Y are two real linear spaces and f is a mapping of X into Y, then f is called a linear function (resp. mapping, transformation) iff for all $x,y \in X$ and all pairs of real numbers α, β it is true that $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$. A linear functional is a real valued linear function (i.e. Y is the set of real numbers). If f is a linear functional and t is a real number then the sets $\{x \mid f(x) \leq t\}$ and $\{x \mid f(x) \geq t\}$ are complementary. They are called halfspaces. Y is a hyperplane iff there is a nonidentically zero linear functional f and a real number α such that $Y = f^{-1}(\alpha)$. The null space (kernel) of a linear function is the set $f^{-1}(0_Y)$. The linear function f is a linear isomorphism iff $f^{-1}(0_Y) = 0_X$. Equivalently: a linear function is a linear isomorphism iff it is one-to-one. The space of all linear functionals on X is called the dual of X.

3.2.1.2 Convexity

The line segment joining $x,y \in X$ is defined as the set $\{z \mid z = \lambda x + (1 - \lambda)y, 0 \leq \lambda \leq 1\}$. Its is denoted by $[x:y]$. The open line segment $(x:y)$ is defined in a similar way, with $0 < \lambda < 1$. A set A in a linear space is called convex iff for all pairs of points $x,y \in A$ it is true that $[x:y] \subset A$. The intersection of all convex sets that contain a set A

is called the convex hull(extension, cover, envelope) of A. It is convex, and denoted by (A) Clearly A is convex iff A = (A). If B is the set of all finite linear combinations $\sum \alpha_i x_i$ of elements of A, with $\sum \alpha_i = 1$ and $\alpha_i \geqslant 0$ for all i, then B = (A). A set A is circled iff $\lambda A \subset A$ for all $|\lambda| \leqslant 1$. A circled set is symmetric in the sense that $-A = A$. The smallest circled set that contains A is called the circled extension of A. A set $A \subset X$ is radial at a point x (or: x is a core point of A) iff for each $y \neq x$ there is a $z \neq x$ such that $[x:z] \subset [x:y] \cap A$ (in words: A contains a line segment through x in every direction). The set of all points at which A is radial is the radial kernel (core) of A. If A is convex then the radial kernel of A is convex and it is its own radial kernel. If $A \subset X$ radial at 0_X then the Minkovski functional for A is defined as $p(x) = \inf \left\{ a \mid \frac{1}{a} x \in A,\ a \geqslant 0 \right\}$. It follows that p(x) is nonnegatively homogeneous: $p(\alpha x) = \alpha\, p(x)$ for all $\alpha \geqslant 0$, that p(x) is subadditive: $p(x + y) \leqslant p(x) + p(y)$ if A is convex, and that p(x) is absolutely homogeneous: $p(\alpha x) = |\alpha|\, p(x)$ if A is circled. A nonnegative, abolutely homogeneous, subadditive, real valued function is called a pseudo-norm. It is a norm if p(x) = 0 iff x = 0. For any convex set radial at o the Minkovski functional has the following properties: $\{ x \mid p(x) < 1 \}$ is the radial kernel, and $A \subset \{ x \mid p(x) \leqslant 1 \}$. Conversely, if p is a nonnegative, nonnegatively homogeneous subadditive function on X and $A = \{ x \mid p(x) \leqslant 1 \}$, then A is convex and radial at 0, and p is the Minkovski functional for A. Moreover A is circled iff p is a pseudo-norm. A set A is called star-shaped relative to a point x if for each $y \in A$ the segment $[x:y] \subset A$. If A is radial at x then it is star-shaped relative to x. The kernel of a set A is the set of all points with respect to which A is star-shaped. Because the kernel of an arbitrary set is always convex, it is also called the convex kernel. A set $A \subset X$ is a cone iff $A + A \subset A$ and $\alpha A \subset A$ for all $\alpha \geqslant 0$. Both 0_X and X are cones. The intersection of all cones that contain a set A i called the conal extension of A.

3.2.1.3 Separation

Two subsets A and B of a real linear space can be separated if there are complementary half-spaces that contain A and B respectively. The linear function f separates A and B iff $\sup\{f(x) \mid x \in A\} \leq \inf\{f(x) \mid x \in B\}$. If the inequality is strong then f strongly separates A and B. If A and B are non-empty convex subsets of a real linear space X and A is radial at some point, then there is a linear functional f separating A and B iff B is disjoint from the radial kernel of A. For a proof see Kelley et al (p 22-23), or Valentine (p 24-25). For our purposes the following theorem is more important: if A and B are nonempty, convex, and disjoint subsets of a finite dimensional linear space X, then they can be separated by a hyperplane. This result is proved, for example, in Valentine (p 25). The corresponding theorems on strong separation are: two convex subsets A and B of a real linear space X can be strongly separated by a linear functional iff there is a convex set C which is radial at 0_X such that $(A + C) \cap B = \emptyset$. Proof: Kelley et al, p 23.

3.2.2 Topological concepts

A linear topological space (topological vector space) is a linear space X with a topology such that addition and scalar multiplication are continuous simultaneously in both variables. We shall suppose moreover that all linear topological spaces satisfy T_2 (this causes nearly no loss of generality, cf Kelley et al, p 41). In linear topologcal spaces closures of circled sets are circled, and closures of convex sets are convex. The interior of a circled set is circled iff it contains 0_X. It is interesting to consider the conditions equivalent to continuity of a linear functional. In general, if f is a (not identically zero) linear function of the linear topological space X into the linear topological space Y, f is continuous iff f is continuous at some point of its domain iff the null space of f is closed iff f is bounded on some neighborhood of 0_X. We know from general topology that a subfamily $\mathcal{B}$ of the neighborhood system V_x of a point x is a base iff each member of V_x contain

a member of $\mathcal{B}$. A base for the neighborhood system of 0 is called a local base. A linear topological space is locally convex iff the family of convex neighborhoods of 0 is a local base. A set $A \subset X$ is bounded iff for each neighborhood U of 0_X there is a real number α such that $A \subset \alpha U$. A linear topological space is normable iff it is locally convex and it contains a nonempty bounded open set. All finite dimensional linear topological spaces are normable. A normable linear topological space with a norm is called a normed topological space. If it is also finite dimensional it is called a Minkovski space. Another useful concept from general topology is compactness. A cover of a set A is a family of sets $\{B_i\}$ such that $A \subset \bigcup_i B_i$. A set A is compact iff every cover of A has a finite subcover. In a finite dimensional linear topological space all bounded closed sets are compact. We are now ready to state the relevant results: if A and B are nonvoid convex subsets of a linear topological space X and A has nonempty interior, then there is is a continuous linear functional separating A and B iff B is disjoint from the interior of A. If A and B are nonvoid convex subsets of a locally convex linear topological space, then there is a continuous linear functional strongly separating A and B iff 0_X is not a member of the closure of B - A. If A and B are disjoint nonvoid convex subsets of a locally convex linear topological space X, and A is compact and B is closed, then there is a continuous linear functional strongly separating A and B. For the proofs see Kelley et al p 118-120. In Valentine (p 25) we find: if A and B are nonvoid disjoint compact convex subsets in a finite dimensional linear topological space, then there exists a hyperplane strongly separating A and B. If A and B are nonvoid disjoint convex sets in a finite dimensional linear topological space, then there exists a hyperplane separating A and B. The following two interesting results are also proved in Valentine (p 86-89): If P and Q are two compact collections of points in an n-dimensional linear topological space, then P and Q can be strongly separated by a hyperplane iff for each subset T of n + 2 or fewer

points of $P \cup Q$ there exists a hyperplane strongly separating $T \cap P$ and $T \cap Q$. The previous theorem is also true if we replace the notion of strong separation by that of separation and n+2 by 2n + 2.

3.3 Separation by hyperplanes in Minkovski spaces (binary items)

In this section we apply the results of the previous sections to our problem. A (binary) item with categories B and C is said to have a semi-strong contiguous representation iff there is a continuous linear functional f such that

$$c_1\colon \sup\{f(x) \mid x \in B\} < \inf\{f(x) \mid x \in B\}.$$

A set of items has a SSC-representations iff all items have one. We restrict ourselves to Minkovki spaces.

Theorem 3.3.1: An item $A = \{B,C\}$ has a SSC-representation in a Minkovski-space X iff the convex hulls (B) and (C) are disjoint.

Proof: Sufficiency: A Minkovski space is of course normable, which implies that it is Hausdorff (T_2) and thus T_1. Therefor B and C, being finite, are closed. Moreover they are of finite diameter, and thus bounded. It follows that B and C are compact. In Minkovski spaces the convex hulls of compact sets are compact. Thus (B) and (C) fulfill the conditions of the theorem on strong separation in finite dimensional spaces, and consequently item A has a SSC-representation.

Necessity: If the item has a SSC-representation then there is a real number α such that $f(x) < \alpha$ for $x \in A$ and $f(x) > \alpha$ for $x \in B$. The open half-spaces $S_1 = \{x \mid f(x) < \alpha\}$ and $S_2 = \{x \mid f(x) > \alpha\}$ are disjoint, $A \subset S_1$ and $B \subset S_2$. Moreover S_1 and S_2 are convex, and thus $(A) \subset S_1$ and $(B) \subset S_2$. It follows that (A) and (B) are disjoint. This completes the proof of the theorem.

It is easy to see that requiring SSC is equivalent to requiring the existence of a solution to a finite system of strict linear inequalities.

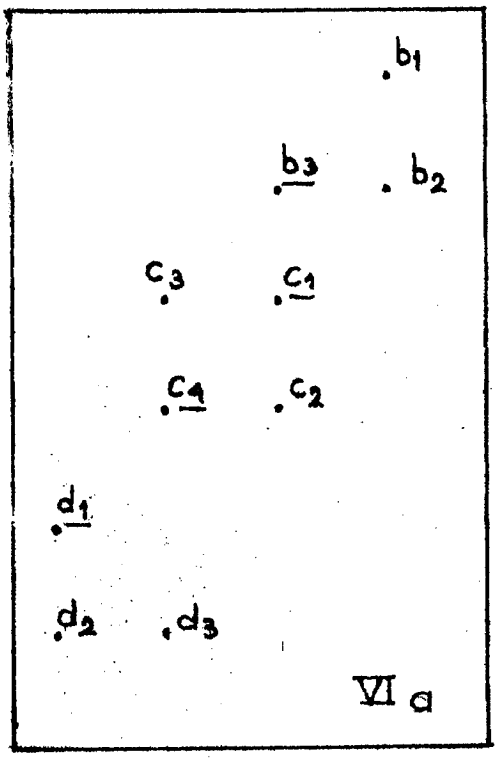
3.4 Separation by hyperplanes (1-ary items)

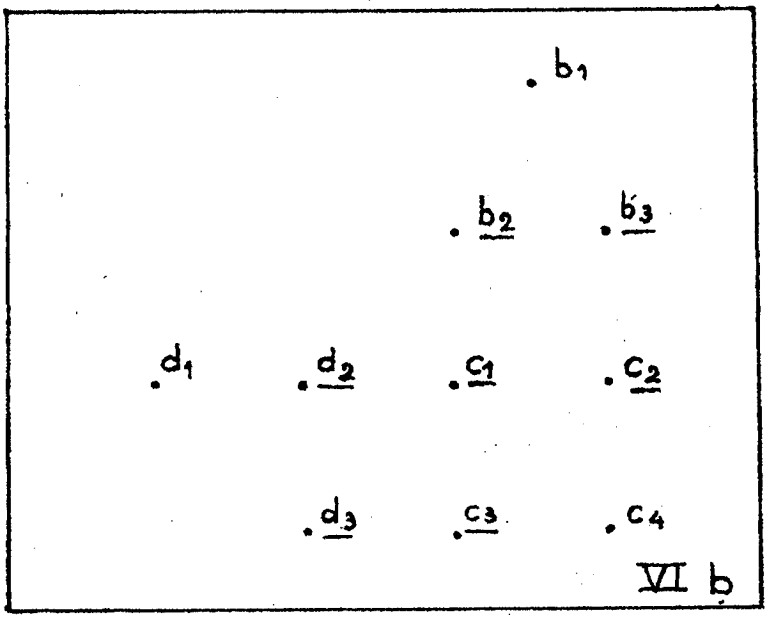
There are several possibilities to define SSC for 1-ary items. We may require

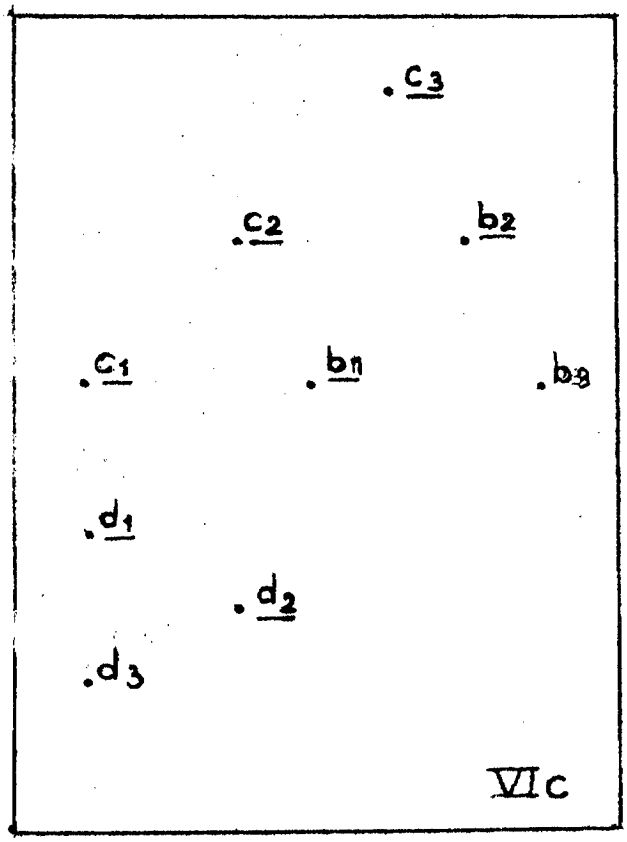
(SSC_I) that the convex hulls of the representations of all categories are pairwise disjoint, or (SSC_{II}) that the convex hulls of A_s and $A \mid A_s$ are disjoint for all $s \in L$. Clearly the latter is the stronger condition: if $(A_s) \cap (A \mid A_s) = \emptyset$ then, because for $t \neq s$ $(A_t) \subset (A \mid A_s)$, also $(A_s) \cap (A_t) = \emptyset$. A third condition is that the separating hyperplanes are parallel (i.e. translates of each other). Again $SSC_P \rightarrow SSC_I$. In the GL-MSA-series the same requirements are used in MSA-III. It is evident that both SSC_I and SSC_{II} can be formulated as requiring a solution to a system of strict linear inequalities (that is: given a particular representation). Moreover if a representation satisfies either SSC_I, SSC_{II}, or SSC_P then any translate of this representation also satisfies the same requirements. For a fourth possibility this is not true. We call it SSC_E and all separating hyperplanes pass through a common point. If we take this point as the origin then a necessary and sufficient condition for an item to have an SSC_E contiguous representation is that the conal extensions of $A_1,\ldots,A_l$ are disjoint. Because the convex hull is a subset of the conal extension we have $SSC_E \rightarrow SSC_I$. But also, less obviously, $SSC_P \rightarrow SSC_E$. This can be seen most easily by taking the origin at a very large distance from A. The hyperplanes emanating from this origin and separating the subsets of A will then be almost parallel, and by taking the origin further away they can be made to be as close to the parallel hyperplanes of SSC_P as we wish. Thus SSC_P is a limiting case of SSC_E, and $SSC_P \rightarrow SSC_E$. Examples are shown in figure VI a-d. Whether or not the representations in these figures are contiguous according to different definitions is scored in the table below.

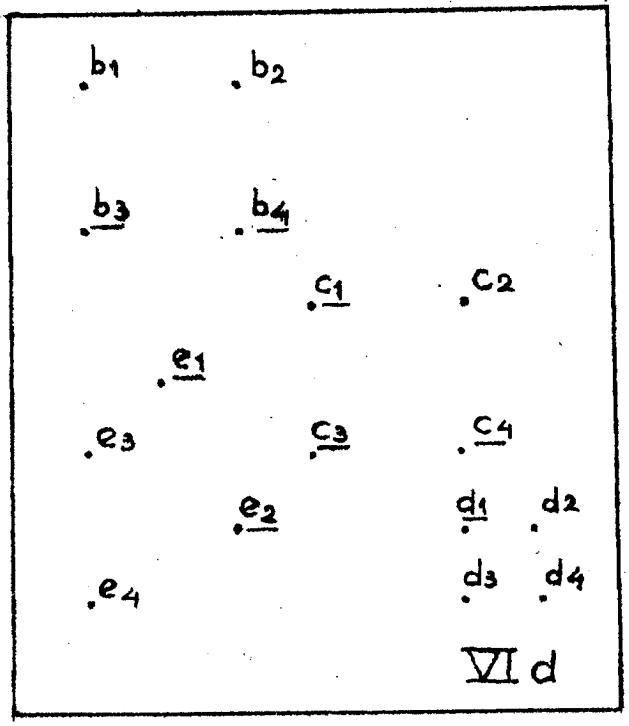
figure	SSC_I	SSC_{II}	SSC_P	SSC_E	MSA^I
VIa	+	-	+	+	+
VIb	+	-	-	+	+
VIc	+	+	-	+	?
VId	+	-	-	-	+

In figure VIc the MSA^I-problem is undecided for category c, the other catego-


b1
b3
b2
c3
c1
c4
c2
d1
d2
d3
VI a


b1
b2
b3
d1
d2
c1
c2
d3
c3
c4
VI b


c3
c2
b2
c1
b1
b3
d1
d2
d3
VI c


b1
b2
b3
b4
c1
c2
e1
e3
c3
c4
e2
d1
d2
d3
d4
e4
VI d

ries do have an MSA^I-contiguous representation.

3.5 The inner product model

Another approach to the study of separation by hyperplanes is to represent the item alternatives and the subjects (the set A) as points in the same space. We assume that this space is a Minkovski space with inner product. The inner product is denoted by (x,y) and satisfies

ip_1: $(x,y) = (y,x)$ $\quad \forall x,y \in X.$

ip_2: $(x,x) > 0$ $\quad \forall x \in X \mid 0_X.$

ip_3: $(x + y,z) = (x,z) + (y,z)$ $\quad \forall x,y,z \in X.$

ip_4: $(\alpha x,z) = \alpha(x,z)$ $\quad \forall x,z \in X,\ \alpha \in Re.$

The norm associated with the inner product is

$$\| x \| = (x,x)^{\frac{1}{2}},$$

and the metric

$$d(x,y) = \| x - y \| .$$

A sequence of points $x_1,x_2,\ldots$ in a metric space is called a Cauchy (or fundamental) sequence if $d(x_m,x_n) \to 0$ if m,n tend independently to ∞. Every convergent sequence is a Cauchy sequence, but the converse is not necessarily true. A metric space is complete iff every Cauchy sequence converges (of course Re^p is complete, because in this space with its usual topology a Cauchy sequence converges by definition). A complete, normed, real linear space is called a real Banach space. If $f(x)$ is a continuous linear functional on a real Banch space X with inner product, then there exists an element $b \in X$ such that $f(x) = (b,x)$. It follows that in Euclidean p-space $(x,y) = \sum x_i y_i$. We shall limit our discussion to this particular inner product space.

Consider item j with l categories. The categories are represented by the points $y_1,\ldots,y_l$, the elements of A by the points $x_1,\ldots,x_n$. We want to find the representation in such a way that

$$(y_s,x_i) > (y_s,x_k) \text{ whenever } x^j_{si} = 1 \wedge x^j_{sk} = 0. \qquad (1)$$

Theorem 3.5.1: An SSC_{II}-contiguous representation for item j can be found in Euclidean p-space iff the inequalities defined by (1) have at least one solution.

Proof: Sufficiency: Suppose that there is a solution x,y to the inequalities. Define (using this solution) the real number $\alpha_s = \frac{1}{2}\left[\inf\left\{(x, y_s) \mid x \in A_{js}\right\}\right.$ $\left. + \sup\left\{(x,y_s) \mid x \in A \mid A_{js}\right\}\right]$. Then $(x,y_s) < \alpha_s$ for all $x \in A_{js}$ and $(x,y_s) > \alpha_s$ for all $x \in A \mid A_{js}$, and thus the function $\left\{x \mid (x,y_s) = \alpha_s\right\}$ separates A_{js} and $A \mid A_{js}$. Moreover this set can be written as $\left\{x \mid \sum_r x_r y_{sr} = \alpha_s\right\}$ and is thus a hyperplane.

Necessity: Simply reverse the argument. Q.E.D.

This theorem makes it clear why we stressed the relation between belonging and order in section 1.4. Consider the nonmetric factor analysis (NFA) model discussed by Shepard (1966), Roskam (1968). We have an n x m data matrix Z, on each of the rows of Z a partial order $\succ_i$ is defined, we want to represent the row stimuli in an n x p matrix Y and the column stimuli in a m x p matrix X by finding a representation in such a way that

$$\sum_{s=1}^{p} x_{is} y_{js} > \sum_{s=1}^{p} x_{is} y_{ks} \quad \text{whenever} \quad z_{ij} \succ_i z_{ik} \tag{2a}$$

$$\sum_{s=1}^{p} x_{is} y_{js} = \sum_{s=1}^{p} x_{is} y_{ks} \quad \text{whenever} \quad z_{ij} =_i z_{ik} \tag{2b}$$

Applying these requirements to our ICP-matrix E is equivalent to requiring SSC_{II}. This also helps to show a weakness of SSC_{II}. We consider each row of E separately, without actually using the fact that some of the rows correspond with (mutually exclusive and exhaustive) categories of an item. If we permute the rows of E the outcome will be the same. This is because we consider each category as a binary categorizer, and not each item as an 1-ary categorizer. Of course for binary items there is only one form of SSC, and this form is also equivalent to NFA and to the compensatory model devised by Coombs and Kao (1955) for binary items.

For SSC_P the situation is much less simple, although it can also be translated into the inner product model.

Theorem 3.5.2: For item j an SSC_P contiguous representation can be found in Euclidean p-space iff there exists a point y_j and l disjoint open intervals $R_1, R_2, \ldots, R_l$ such that the system

$$(y_j, x) \in R_s \quad \text{whenever} \quad x \in A_{js} \tag{3}$$

has at least one solution.

Proof: Sufficiency: Let $R_s = (\alpha_s, \beta_s)$ and suppose, without loss of generality, that the item alternatives are ordered in such a way that $\beta_s < \alpha_t$ for all $s,t \in L$ with $s < t$. Consider the l-1 quantities $\gamma_s = \frac{1}{2}(\alpha_s + \beta_{s+1})$, $s=1,\ldots,l-1$. Then the l-1 hyperplanes $(y_j, x) = \gamma_s$ are parallel and separate the sets A_{js}.

Necessity: Choose y_j such that (y_j, x) is a hyperplane through 0_X and parallel to the separating hyperplanes. Let $\alpha_s = \inf \{(y_j, x) \mid x \in A_s\}$ and $\beta_s = \sup \{(y_j, x) \mid x \in A_s\}$. Then the closed intervals $[\alpha_s, \beta_s]$ are disjoint. Order them in such a way that $\beta_s < \alpha_{s+1}$, $s=1,\ldots,l-1$, and define

$$R_1 = (\alpha_1 - \Delta_1, \tfrac{1}{2}(\beta_1 + \alpha_2)),$$
$$R_2 = (\tfrac{1}{2}(\beta_1 + \alpha_2), \tfrac{1}{2}(\beta_2 + \alpha_3)),$$
$$\vdots$$
$$R_l = (\tfrac{1}{2}(\beta_{l-1} + \alpha_l), \beta_l + \Delta_2),$$

with Δ_1, Δ_2 arbitrary positive numbers. Then the open intervals R_s are disjoint, and $x \in A_{js}$ implies $x \in [\alpha_s, \beta_s]$ implies $x \in R_s$. Q.E.D.

The essential differences with SSC_{II} are clear: in SSC_P each item (and not each category) is represented as a point in the joint space. Moreover a byproduct of the SSC_P representation of an item is a natural ordering of the categories. The reason why the SSC_p requirements cannot be formulated as a simple set of inequalities (as in SSC_{II}) is that we do not know this ordering beforehand. Suppose we define an arbitrary (strict) ordering $\succ_0$ over the

categories and require

$$(y_j,x_i) > (y_j,x_k) \text{ whenever } x_i \in A_{js} \wedge x_k \in A_{jt} \wedge A_{js} \supset A_{jt}. \quad (4)$$

If this set of inequalities has a solution then an SSC_P representation exists, i.e. the condition is sufficient. Of course 1! different strict orderings can be defined over the categories, each ordening has a solution set Q_v (which is possibly void). The conditions $Q_v \neq \emptyset$ (taken separately) are all sufficient for SSC_P but the condition

$$\bigcup_{v=1}^{1!} Q_v \neq \emptyset \quad (5)$$

is both necessary and sufficient. Because $(-y_j,x) = -(y_j,x)$ by ip_4, we only have to consider $\frac{1}{2}1!$ different orderings (delete one of each pair of mirror-images), but in a set of m items this may amount to the prohibitive amount of $(\frac{1}{2}1!)^m$ differents sets of inequalities that have to be investigated. Necessary and sufficient for SSC_P is that at least one of these systems has a solution. Of course using conditions that are only sufficient but not necessary for SSC_P as the rationale and basis for an algorithm means in fact that we require a stronger form of contiguity. It seems clear that SSC_P is a more appropriate theory for unordered 1-ary items. It has the same drawbacks (not knowing the order on the categories beforehand, and thus being forced to use somewhat heuristic methods) as Guttman scaling with unordered categories (which is, of course, one-dimensional).

3.6 Some special structures

Both SSC_{II} and SSC_P are interesting because they have a direct connection with the inner product model and with nonmetric component analysis. For SSC_E the comparable rationale is given by Guttman's circumplex theory, and with elliptic multidimensional scaling (Van De Geer 1969). To give algorithm-oriented necessary and sufficient conditions for SSC_E we construct the sphere S_j with radius 1 and center y_j. Thus $S_j = \{x \mid d(x,y_j) = 1\}$. The half-lines emanating from y_j and passing through x_i are given by

$H = \{ z \mid z = \lambda x_i + (1-\lambda) y_j , \ \lambda \geq 0 \}$. These half-lines pass through the sphere at the point where $d(y_j, z) = 1$. Solving for λ gives

$$\lambda_i = 1/d(x_i, y_j). \tag{6}$$

Consider the points (on the sphere)

$$\mathfrak{x}_i = \lambda_i x_i + (1-\lambda_i) y_j, \tag{7}$$

and let s_{ij} be the cosine of the angle between the two vectors x_i and x_k (seen from y_j). Thus

$$s_{ij} = \frac{(x_i - y_j)'(x_k - y_j)}{d(x_i, y_j) d(x_k, y_j)}, \tag{8}$$

and

$$d^2(\mathfrak{x}_i, \mathfrak{x}_k) = 2 - 2s_{ij}. \tag{9}$$

For the distances measured along the sphere we obtain

$$\delta(\mathfrak{x}_i, \mathfrak{x}_k) = \arccos(s_{ij}), \tag{10}$$

and $0 \leq \delta(\mathfrak{x}_i, \mathfrak{x}_k) \leq 2\pi$.

Theorem 3.5.3: If p=2 then the points x_i can be projected as points $\bar{x}_i$ on the real axis in such a way that $\delta(\mathfrak{x}_i, \mathfrak{x}_k) = d(\bar{x}_i, \bar{x}_k)$.

Proof: Let

$$\bar{x}_i = \arcsin \frac{(x_{i1} - y_{j1})}{d(x_i, y_j)} \tag{11}$$

or

$$\frac{(x_{i1} - y_{j1})}{d(x_i, y_j)} = \sin(\bar{x}_i). \tag{12}$$

Then

$$s_{ik} = \sin \bar{x}_i \sin \bar{x}_k + \cos \bar{x}_i \cos \bar{x}_k = \cos(\bar{x}_i - \bar{x}_k), \tag{13}$$

and thus

$$\delta(\mathfrak{x}_i, \mathfrak{x}_k) = \arccos s_{ij} = \bar{x}_i - \bar{x}_k. \tag{14}$$

This proof is somewhat provisory and not very exact. Some of the difficulties will be clarified in the examples. Nevertheless: Q.E.D.