RN 010-68

NON-METRIC MULTIDIMENSIONAL SCALING

Jan de Leeuw

November 1968

PSYCHOLOGICAL INSTITUTE - UNIVERSITY OF LEYDEN - NETHERLANDS

CONTENTS

0: SUMMARY

0.1

In this paper we shall be interested in the mapping of a finite set A into p-dimensional euclidean space. The data consist of some kind of information about the (dis)similarity of the elements of A with respect to each other. The existing forms of multidimensional scaling can be classified as follows:

DATA	ERROR THEORY	REPRESENTATION	REFERENCES
ratio scale	metric	metric	Young & Householder (1938) Ekman (1963)
interval scale	metric	metric	Messick & Abelson (1956) Torgerson (1958)
weak order	metric	metric	Shepard (1962 a,b) Kruskal (1964 a,b)
weak order over subsets of $A \times A$	metric	metric	Gleason (1967) Guttman (1967) Roskam (1968)
partial order	non-metric	non-metric	Coombs (1964)

We did not include (multidimensional) unfolding, because it seeks to find a representation of a subset of A:

partial orders over a fixed $A' \subset A$	non-metric	non-metric (A' only)	Coombs (1950) Bennett & Hays (1960) Hays & Bennett (1960)

It is quite easy to see from this scheme that the term 'non-metric' multidimensional scaling is ambiguous, to say the least. In most cases it refers to the data-part of the scheme, indicating which properties of the data are used in the procedure. A non-metric error theory means that all errors are treated as equally serious, and that the size of the errors does not affect the goodness-of-fit coefficient. A metric error theory takes the size of the errors into account. A non-metric (euclidean) representation gives us p partial orders of projections on the axes of

method is the only existing form of multidimensional scaling that may be called non-metric without any ambiguity. Observe, however, that the last meaning of non-metric (as a property of the representation) is a weakness of the method. It simply indicates that the method does not make an optimal use of its data (Shepard 1966). In this paper we shall develop some other algorithms. In the first place algorithms with

data: partial order

error theory: metric

representation: metric

The methods differ in the degree in which the error theory depends on the size of the errors. Some of them are less non-metric than the others, so to speak. We shall also discuss a method with

data: partial order

error theory: non-metric

representation: metric

The principal points we shall pay attention to are the organisation of the data, treatment of ties and inconsistencies, error theory, and minimization algorithms. The relevant computer programs are discussed in some detail and examples are given. In the first section a fairly complete overview of the literature on aspects of multidimensional scaling not covered in this report is given.

0.2

The data in a non-metric multidimensional scaling experiment can be described as a set of quadruples of indices. Let A be the set of stimuli. The fact that a quadruple (i,j,k,l) is in the data set L roughly means that the dissimilarity between the stimuli a_i and a_j is greater than the dissimilarity of a_k and a_l. We now want to map A into the euclidean p-space X with metric d in such a way that

$$(i,j,k,l) \in L \Rightarrow d_{ij} - d_{kl} \geqslant 0 \qquad (1)$$

One algorithm is simply to maximize the sum of the right hand terms, i.e.

the sum of squares of all these SBP-differences, and SS_{min} the sum of squares of the nagative ones. Our next algorithm minimizes SS_{min} under the condition that SS_{all} is some constant value. The final algorithm minimizes the number of negative SBP-differences.

1: INTRODUCTION

1.1

In this paper we shall develop some new algorithms for non-metric multidimensional scaling, compare them with existing algorithms, and present some examples. We shall not be concerned with measurement theoretical problems (Tversky 1965, Beals & Krantz 1967, Krantz 1967, Beals, Krantz & Tversky 1968), with existence theorems (Guttman 196- a,b), with properties of the metric used (Cross 1965, Coombs 1966, Wender 1968, Roskam 1968 ch III), with psychological aspects of the similarity concept (Attneave 1950, Shepard 1964, Torgerson 1965, Handel 1967, Ellis & Manning 1967, Wender 1968), with theoretical approaches to the distance function (Shepard 1957, Luce 1961, Krantz 1966), with generalizations to more general metric spaces (Restle 1959, Hays 1964, Harary 1964, Johnson 1967, Hartigan 1967), with statistical models (Benzécri 1964, Suppes & Zinnes 1963 p 60-62), with continuity considerations (Shepard 1964, Shepard & Carroll 1967), and with facet theoretical (Guttman 1959, 1967[b], Foa 1965) and data theoretical sides of the problem (Coombs 1964). For all these other aspects of multidimensional scaling we refer the reader to the literature mentioned above.

1.2

In non-metric multidimensional scaling (NMS) the data consist of a subset L of the cartesian product $M \times M \times M \times M$. If the set of stimuli A has n elements, then $M = \{1,\ldots,n\}$. This means that L can be made to correspond with a binary relation Q over $A \times A$:

$$(i,j,k,l) \in L \Leftrightarrow Q\{(a_i,a_j),(a_k,a_l)\} \qquad (1)$$

L can be constructed in several ways. The most complete type of data arises if we show the subject all n^4 possible combinations of pairs of stimuli and ask him whether stimulus a_i and a_j are less alike (similar) then stimuli a_k and a_l. Let $(i,j,k,l) \in L$ iff he responds in the affirmative, and $(k,l,i,j) \in L$ otherwise. In other cases the raw data

consist of a mapping δ of $A \times A$ into a set Δ, strictly ordered by a relation $\rangle$. We may define L by the rule

$$(i,j,k,l) \in L \Leftrightarrow \delta(a_i,a_j) \rangle \delta(a_k,a_l) \qquad (2)$$

for all $a_i,a_j,a_k,a_l \in A$. Because the order relation is asymmetric and irreflexive, L will have $\binom{n^2}{2} = \frac{1}{2}n^2(n+1)(n-1)$ elements. In most cases however there will be some symmetry in the data, such that

$$\delta(a_i,a_j) = \delta(a_j,a_i) \rangle \delta(a_i,a_i) = \delta(a_j,a_j) \qquad (3)$$

for all $a_i,a_j \in A \wedge i \neq j$. In that case we only have to consider the mapping of a subset C of $A \times A$ into Δ, with $(a_i,a_j) \in C$ iff $j \rangle i$. Because of the irreflexivity and asymmetry it follows that L contains $\binom{\binom{n}{2}}{2} = \frac{1}{8}n(n-1)(n-2)(n+1)$ elements. In the sequel we shall write D_n for this number. Formally, the rule is

$$\begin{aligned}(i,j,k,l) \in L \Leftrightarrow\ & \delta(a_i,a_j) \rangle \delta(a_k,a_l) \quad \wedge \\ & j \rangle i \wedge l \rangle k \quad \wedge \\ & k + \tfrac{1}{2}l(l-1) \rangle i + \tfrac{1}{2}j(j-1)\end{aligned} \qquad (4)$$

In other cases there are mappings of the sets $\{a_i\} \times A$ into the strictly ordered sets $\langle \Delta_i, \rangle_i \rangle$ for all $i=1,\ldots,n$. Evidently this produces $n\binom{n}{2} = \frac{1}{2}n^2(n-1)$ elements in L. In the method of triads all $\binom{n}{3}$ possible sets of 3 stimuli are strictly ordered. This produces in a similar way $3\binom{n}{3} = \frac{1}{2}n(n-1)(n-2)$ elements in L. The number of elements in L will be denoted by N. Moreover λ is a one-one mapping of L onto $\{1,2,\ldots,N\}$.

1.3

In the previous section we have used the strict order $\rangle$ for ease of presentation. Actually we do, in most cases, use a weak order $\rangle\!\!\rangle$, or even a weak partial order. The reasons for this will become clear after the discussion of the problem of ties in this section, and the algorithmic problem in section 1.4. As in most non-metric problems (cf Kendall 1962 ch III) the appearance of ties is a nuisance for NMS-algorithms. We may interprete ties in two different ways (cf Kruskal 1964 a, b

Roskam 1968 p 39-40). Consider the case in which δ maps A X A into a subset of the reals with their usual ordening, and let ε_1 be a non-negative real number. The first rule we shall discuss is

$$(i,j,k,l) \in L \iff \delta(a_i,a_j) - \delta(a_k,a_l) \geq \varepsilon_1 \quad (1)$$

If $\varepsilon_1 = 0$ and if $\delta(a_i,a_j) = \delta(a_k,a_l)$ then both (i,j,k,l) and $(k,l,i,j) \in L$. If $\varepsilon_1 > 0$ and $\delta(a_i,a_j) = \delta(a_k,a_l)$ then both (i,j,k,l) and $(k,l,i,j) \notin L$. The interpretation is clear: ε_1 is an estimate of the precision with which we have measured our dissimilarities and $\varepsilon_1 = 0$ is only a special case. If $\varepsilon_1 > 0$ then a tie is interpreted as a lack of precision in the subject's discrimination, and both corresponding quadruples are eliminated from L. The interpretation is reasonable, but the choice of ε_1 is in most cases rather arbitrary. Another rule

$$(i,j,k,l) \in L \iff \delta(a_i,a_j) - \delta(a_k,a_l) \geq -\varepsilon_1 \quad (2)$$

In this case, if $-\varepsilon_1 \leq \delta(a_i,a_j) - \delta(a_k,a_l) \leq +\varepsilon_1$, then evidently both (i,j,k,l) and $(k,l,i,j) \in L$. This is an entirely different interpretation: all values between the boundaries are included as actual ties. The numerical responses of the subject suggest more discrimination capacity than he really has. Both quadruples are included in L. The treatment of ties is especially important in the analysis of adjacency matrices of graphs (as in sociometry). The entries of the matrix $\Delta = \{\delta(a_i,a_j)\} =_{def} \{\delta_{ij}\}$ are either zero or one, i.e. $\delta_{ij} = 1$ iff subject a_i has 'chosen' subject a_j, otherwise $\delta_{ij} = 0$. Evidently in this case the rule must be

$$(i,j,k,l) \in L \iff \delta_{ij} = 0 \wedge \delta_{kl} = 1 \wedge i = k \wedge l > j \quad (3)$$

because the matrix is a conditional proximity matrix, and $\delta_{ij} = \delta_{il} = 1$ means only that they are both not greater than a certain 'critical' dissimilarity δ_{ε_i}, which is characteristic for subject i. In a similar way $\delta_{ik} = 0$ actually means (in the distance model) $\delta_{ik} \geq \delta_{\varepsilon_i}$.

-1.4-

The algorithmic problem can now be stated as follows: we want to find a metric $d : A \times A \longrightarrow Re^{+}$. This mapping must be found in such a way that

$$(i,j,k,l) \in L \Rightarrow d(a_i,a_j) \geqslant d(a_k,a_l) \qquad (1)$$

For $d(a_i,a_j)$ we shall write d_{ij} in the sequel. A more severe set of requirements would be

$$\begin{cases} (i,j,k,l) \in L \Rightarrow d_{ij} \geqslant d_{kl} & (2a) \\ (i,j,k,l) \in L \wedge (k,l,i,j) \notin L \Rightarrow d_{ij} > d_{kl} & (2b) \end{cases}$$

but we shall see that some of our algorithms allow for cases with

$$(i,j,k,l) \in L \wedge (k,l,i,j) \notin L \wedge d_{ij} = d_{kl} \qquad (3)$$

The same thing is true for the Kruskal-algorithm (cf Roskam 1968 p 43-45, p 122-126) and the Guttman-Lingoes rank-image principle (Guttman 1967, p 14). Both Kruskal and Roskam require

$$\delta_{ij} \geqslant \delta_{kl} \Leftrightarrow d_{ij} \geqslant d_{kl} \qquad (4)$$

which is equivalent to (2a) and (2b) in the cases they consider (weak orders that connect $A \times A$). Their algorithms, however, do not try to solve this problem (cf Guttman 1967 p 42). Guttman is much more careful in these matters. Moreover, the discussion in the previous section implies that cases should be possible with

$$\delta_{ij} = \delta_{kl} \wedge d_{ij} \neq d_{kl} \qquad (5)$$

(for example in the case of graphs). Observe that our algorithmic requirements (1) and (2) both imply

$$(i,j,k,l) \in L \wedge (k,l,i,j) \in L \Rightarrow d_{ij} = d_{kl} \qquad (6)$$

In combination with tie-rule (1.3.2)

$$-\varepsilon_1 \leq \delta_{ij} - \delta_{kl} \leq +\varepsilon_1 \Rightarrow d_{ij} = d_{kl} \qquad (7)$$

which is quite a strong requirement. In combination with tie-rule (1.3.1) we have

$$\delta_{ij} - \delta_{kl} = \varepsilon_1 \Rightarrow d_{ij} = d_{kl} \qquad (8)$$

If $\mathcal{E}_1 = 0$ both (7) and (8) reduce to

$$\int_{ij} = \int_{kl} \Rightarrow d_{ij} = d_{kl} \qquad (9)$$

-1.5-

Suppose ω is a ma pping of A into $Re^p = Re \times Re \times \ldots \times Re$ (p times). We write $\omega(a_i) = \bar{x}_i = (x_{i1},\ldots,x_{ij},\ldots,x_{ip})$, where the x_{ij} are real numbers. Define

$$d_{ij} = \left[\sum_{s=1}^{p} |x_{is} - x_{js}|^r\right]^R \qquad (1)$$

If $r \geqslant 1$ and $R = r^{-1}$ then d is a metric by Minkovski's inequality. If $0 < r \leqslant 1$ and $R = 1$ then d also is a metric (Hardy, Little wood & Polya 1952, p 30-32). Because the (positive) power R clearly does not affect the ordening of the distances this specification of the metric makes it possible to replace the algorithmic requirements $(1.4.1)_p$ by

$$(i,j,k,l) \in L \Rightarrow \sum_{s=1}^{p}|x_{is}-x_{js}|^r - \sum_{s=1}^{p}|x_{ks} - x_{ls}|^r \geqslant 0 \qquad (2)$$

with $r > 0$.

2: EUCLIDEAN MAXIMAL SUM ALGORITHM (NMSEMS)

-2.1-

Our first algorithm is already (faultly) described in De Leeuw (1968b, p 28-29). It is a program in the canonical discriminant analysis of relational data or CDARD-series, a slight modification and further development of some ideas of Guttman (1941,1946,1959). It provides non-metric solutions by computing principal components. The technique is surprisingly general (cf also De Leeuw 196- a). If L contains N elements, then define the N $\times$ n matrices S and T in the following way: if $\lambda(i,j,k,l) = m$ then

$$s_{mq} = \delta^{qi} - \delta^{qj} - \delta^{qk} + \delta^{ql} \tag{1}$$

and

$$t_{mq} = \delta^{qi} - \delta^{qj} + \delta^{qk} - \delta^{ql} \tag{2}$$

In (1) and (2) the superscripted δ's denote the Kronecker symbol. The function λ is defined in section 1.2. If x is an n-element vector, then the N-element vectors Sx and Tx contain, respectively, the elements $(x_i - x_j) - (x_k - x_l)$ and $(x_i - x_j) + (x_k - x_l)$ on the appropriate places. Define $f(x) = x'T'Sx$, then

$$f(x) = \sum_{(i,j,k,l) \in L} (x_i - x_j)^2 - (x_k - x_l)^2 \tag{3}$$

The relation with (1.5.2) is obvious. Moreover

$$\begin{aligned} x'T'Sx &= x'\left\{(T'S + S'T)/2 + (T'S - S'T)/2\right\}x = \\ &= x'\left\{(T'S + S'T)/2\right\}x \end{aligned} \tag{4}$$

because T'S - S'T is skew symmetric and hollow, which implies that $x'(T'S - S'T)x = 0$. Define the Lagrangian function

$$\underset{\sim}{\Phi}_A(x) = x'T'Sx - \mu(x'x - 1) \tag{5}$$

Symbolic differentiation with respect to all elements of x simultaneously gives

$$\frac{\partial \underline{\underline{\phi}}(x)}{\partial x'} = (S'T + T'S)x - 2\mu x \tag{6}$$

De Leeuw (1968 b) erroneously states that the conditions for an extreme value are $T'Sx = \mu x$, but finding the extreme values of $x'T'Sx$ means solving the eigenproblem

$$\left\{(S'T + T'S)/2\right\} x = \mu x \tag{7}$$

Of course, by (4), we could also have defined

$$\underline{\underline{\phi}}_B(x) = x'\left\{(S'T + T'S)/2\right\} x - \mu(x'x - 1) \tag{8}$$

with exactly the same result. It is quite easy to see that S and T are composed of two N × n matrices, say S_1 and S_2, with $S = S_1 - S_2$ and $T = S_1 + S_2$. Evidently

$$S'T = S_1'S_1 - S_2'S_1 + S_1'S_2 - S_2'S_2 \tag{9a}$$

$$T'S = S_1'S_1 - S_1'S_2 + S_2'S_1 - S_2'S_2 \tag{9b}$$

so

$$\tfrac{1}{2}(S'T + T'S) = S_1'S_1 - S_2'S_2 \tag{10}$$

Moreover, because both S_1 and S_2 have two non-zero elements in each row, one of them equal to +1, the other to -1, it follows that

$$\text{trace}(S_1'S_1) = \text{trace}(S_2'S_2) = 2N \tag{11}$$

and if

$$Q =_{\text{def}} S_1'S_1 - S_2'S_2 \tag{12}$$

then trace(Q) = 0. Because the row sums of both S_1 and S_2 disappear, $S_1'S_1$ and $S_2'S_2$ are doubly centered, and consequently so is Q. All eigenvectors of Q are centered, and the sum of the eigenvalues equals zero. At least one of them actually equals zero, because Q is singular. We may order the eigenvalues according to

$$\mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_k = 0 \geqslant \cdots \geqslant \mu_n \tag{13}$$

and define

$$f_p = \sum_{s=1}^{p} \mu_s =$$

$$= \sum_{(i,j,k,l) \in L} \left[\sum_{s=1}^{p} (x_{is}-x_{js})^2 - \sum_{s=1}^{p} (x_{ks}-x_{ls})^2 \right] \quad (14)$$

which makes the relation with (1.5.2) even more obvious. Moreover f_p will increase as long as we add new dimensions with positive eigenvalues. This technique can be used to obtain an initial configuration for the iterative programs discussed in the next chapters. It is related to a technique of Guttman (1967 p 36-41). As Guttman's technique it may help us to obtain a practical estimate of the upperbound of the dimensionality: the number of positive eigenvalues of Q. An evident advantage of a very good initial configuration is that the local minimum problem is much less urgent (we solve for the absolute maximum of the eigenproblem). In general this initial approximation is indeed extremely good, moreover it is unique and considerations of degeneracy and triviality do not apply. It requires less than ten percent of the computer time used for a complete iterative solution. It is easily generalized to an NMS-solution that is optimal for a number of subjects at the same time (CDARD7-program, cf De Leeuw 196- a).

-2.2-

Define

$$d_{ij}^2 = \sum_{s=1}^{p} (x_{is} - x_{js})^2 \quad (1)$$

An obvious generalization of (2.1.14) gives us

$$f_p = \sum_{s=1}^{p} \mu_s = \sum_{(i,j,k,l) \in L} w_{\lambda(i,j,k,l)} (d_{ij}^2 - d_{kl}^2) \quad (2)$$

where the $w_{\lambda(i,j,k,l)}$ are M non-negative weights. Consider the following algorithm: take $w_{\lambda(i,j,k,l)} = 1$ for all $(i,j,k,l) \in L$. Maximizing f_p gives us an n X p matrix X. Define $t_{\lambda(i,j,k,l)} = d_{ij}^2 - d_{kl}^2$. We now want to maximize the number of negative elements in t, i.e. maximize

$$\sum_{(i,j,k,l) \in L} \text{sign}(t_{\lambda(i,j,k,l)}) =$$

$$= \sum_{(i,j,k,l) \in L} \frac{d_{ij}^2 - d_{kl}^2}{|d_{ij}^2 - d_{kl}^2|} \qquad (3)$$

provided, of course, that $d_{ij}^2 \neq d_{kl}^2$. Our next step to achieve this is to define new weights:

$$w_{\lambda(i,j,k,l)}^{(i+1)} = \left\{ |d_{ij}^2 - d_{kl}^2| - 1 \right\}^i \qquad (4)$$

And again f_p is maximized with these new weights. As soon as

$$w_{\lambda(i,j,k,l)}^{(i)} \doteq \left\{ |d_{ij}^2 - d_{kl}^2| - 1 \right\}^i \qquad (5)$$

the algorithm has converged, and we have maximized (3) in p dimensions under the condition that $d_{ij}^2 \neq d_{kl}^2$ for all $(i,j,k,l) \in L$. Cf also section 3.5. The results of an application of this algorithm to a set of data are given in table I and figure II. The proximity matrix was the sociogram in table XVIII, which generated 292 inequalities. In table I we have given the three largest eigenvalues and the number of violated inequalities in iteration 1-10, 15, 20, and 22. In figure II all 12 eigenvalues are plotted for iteration 1, 5, 10, 20. Clearly the algorithm fails to do what we wanted it to do.

3: POSITIVE ORTHANT ALGORITHM (NMSPOM)

-3.1-

The algorithmic problem can be reformulated as follows: find a solution to the system of N homogeneous inequalities defined by (1.5.2). And if such a solution does not exist, find a best solution in some well-defined sense. In the case of linear inequalities De Leeuw (1968a) used a two-phase alternating least squares algorithm. The procedure described below is a generalization of this algorithm. A more extensive discussion and a comparison with the general Shepard-Kruskal-Roskam and Guttman-Lingoes algorithms can be found in De Leeuw (196- b). A closely related procedure is used in Phase One of some mathematical programming routines (cf Wolfe 1967 p 105-106). To find a point in the permissible region we have to find a solution to the N inequalities

$$\sum_{s=1}^{p} \left| x_{is} - x_{js} \right|^{r} - \sum_{s=1}^{p} \left| x_{ks} - x_{ls} \right|^{r} \geqslant 0 \qquad (1)$$

For a particular choice of the x_{ij} $(i=1,\ldots,n;\ j=1,\ldots,p)$, let t_m denote the values on the left hand side of these inequalities $(m=1,\ldots,N)$. The value of

$$F(x) = \sum_{m=1}^{N} \left[\max(0,-t_m) \right]^{q} \qquad (2)$$

(with $q > 0$) is our 'raw' goodness of fit measure. $\min_x F(x)$ is defined as the best solution of our N inequalities (1). If the system (1) is consistent, then evidently $\min_x F(x) = 0$ and our solution lies in the positive orthant of the N-dimensional space defined by the values of t_m. In a similar way sets of inequalities can be defined for other measurement models. Programs for non-metric factor analysis and additive conjoint measurement are in preparation.

-3.2-

To obtain more practical goodness of fit coefficients observe that

$$\sum^{N} \left[\max(0,-t_m) \right]^{q} = 2^{-q} \sum^{N} \left(\left| t_m \right| - t_m \right)^{q} \qquad (1)$$

To exclude the trivial solution with $x_{ij} = 0$ for all i, j we define

$$\Gamma_1 = \frac{\sum_{m=1}^{N} (|t_m| - t_m)^q}{2^q \sum_{m=1}^{N} |t_m|^q} \tag{2}$$

If $t_m \geqslant 0$ for all m and $t_m > 0$ for at least one m, then $\Gamma_1 = 0$. If $t_m \leqslant 0$ for all m and $t_m < 0$ for at least one m, then $\Gamma_1 = 1$. In the sequel we will only consider the case q = 2. Then

$$\Gamma_1 = \frac{\sum (|t_m| - t_m)^2}{4 \sum t_m^2} \tag{3}$$

And

$$1 - 2\Gamma_1 = \frac{\sum t_m |t_m|}{\sqrt{\sum t_m^2 \sum |t_m|^2}} \tag{4}$$

It follows that $0 \leqslant \Gamma_1 \leqslant 1$ for all t, and $\Gamma_1 = \frac{1}{2}$ iff $\sum t_m |t_m| = 0$ and $t_m \neq 0$ for at least one m. If we change the sign of all elements in t, then the sum of the two corresponding Γ_1's equals unity. An alternative coefficient can be defined as

$$\Gamma_2 = \frac{\sum (|t_m| - t_m)^2}{\sum (|t_m| + t_m)^2} \tag{5}$$

Evidently $0 \leqslant \Gamma_2 < +\infty$ and if t lies in the negative orthant of the t-space Γ_2 is not defined. The difference between Γ_1 and Γ_2 is obvious: in Γ_1 the errors are weighted according to their size, in Γ_2 the $t_m \geqslant 0$ are weighted according to their size too. It is, however, easy to prove that

$$\Gamma_1 = \frac{\Gamma_2}{\Gamma_2 + 1} \tag{6}$$

and, inversely,

$$\Gamma_2 = \frac{\Gamma_1}{1 - \Gamma_1} \tag{7}$$

If follows that

$$\Gamma_2 - \Gamma_1 = \frac{\Gamma_1^2}{1 - \Gamma_1} = \Gamma_1 \Gamma_2 \tag{8}$$

and consequently $\Gamma_2 \geqslant \Gamma_1$ for all t. Moreover

$$\Gamma_1^{-1} - \Gamma_2^{-1} = 1 \tag{9}$$

By (4) and (7) $\Gamma_2 = 1$ iff $\Sigma t_m |t_m| = 0$ and $t_m > 0$ for at least one m. If we change the sign of all elements in t then the product of the two corresponding Γ_2's equals unity.

-3.3-

In this section we shall repeat our goodness of fit formulae and list some of the relevant derivatives. Define

$$S = \Sigma (|t_m| - t_m)^2 \tag{1}$$

$$T_1 = 4 \Sigma t_m^2 \tag{2}$$

$$T_2 = \Sigma (|t_m| + t_m)^2 \tag{3}$$

$$\Gamma_1 = \frac{S}{T_1} \tag{4}$$

$$\Gamma_2 = \frac{S}{T_2} \tag{5}$$

If y is an arbitrary variable then

$$\frac{\partial \Gamma_i}{\partial y} = \frac{1}{T_1} \left\{ \frac{\partial S}{\partial y} - \Gamma_i \frac{\partial T_i}{\partial y} \right\} \qquad (i=1,2) \tag{6}$$

$$\frac{\partial S}{\partial y} = 4 \Sigma (t_m - |t_m|) \frac{\partial t_m}{\partial y} \tag{7}$$

$$\frac{\partial T_1}{\partial y} = 8 \Sigma t_m \frac{\partial t_m}{\partial y} \tag{8}$$

$$\frac{\partial T_2}{\partial y} = 4 \Sigma (t_m + |t_m|) \frac{\partial t_m}{\partial y} \tag{9}$$

$$\frac{\partial \Gamma_1}{\partial y} = \frac{4}{T_1} \left[\Sigma \left\{ (1-2\Gamma_1) t_m - |t_m| \right\} \frac{\partial t_m}{\partial y} \right] \tag{10}$$

$$\frac{\partial \Gamma_2}{\partial y} = \frac{4}{T_2} \left[\Sigma \left\{ (1-\Gamma_2) t_m - (1+\Gamma_2) |t_m| \right\} \frac{\partial t_m}{\partial y} \right] \tag{11}$$

By using (3.2.6) and (3.2.7) it follows that (if $\Gamma_2 \neq 0$)

$$\frac{\partial \Gamma_2}{\partial y} = \frac{4}{T_2} \cdot \frac{\Gamma_1}{\Gamma_2} \left[\sum \left\{ (1-2\Gamma_1) t_m - |t_m| \right\} \frac{\partial t_m}{\partial y} \right] =$$

$$= \frac{T_1}{T_2} \cdot \frac{\Gamma_1}{\Gamma_2} \cdot \frac{\partial \Gamma_1}{\partial y} \tag{12}$$

Because $\Gamma_1 = 0$ iff $\Gamma_2 = 0$ and $\Gamma_1 = 0$ implies $\frac{\partial \Gamma_1}{\partial y} = 0$ clearly

$$\frac{\partial \Gamma_1}{\partial y} = 0 \iff \frac{\partial \Gamma_2}{\partial y} = 0 \quad (*) \tag{13}$$

Remember that

$$\lambda^{-1}(m) = (i,j,k,l) \Longrightarrow t_m = d_{ij}^{1/R} - d_{kl}^{1/R} \tag{14}$$

$$d_{ij}^{1/R} = \sum_{q=1}^{p} |x_{iq} - x_{jq}|^r \tag{15}$$

Then

$$\frac{\partial t_m}{\partial x_{gh}} = r \left[\operatorname{sign}(x_{ih} - x_{jh}) |x_{ih} - x_{jh}|^{r-1} (\delta_{ig} - \delta_{jg}) - \right.$$

$$\left. \operatorname{sign}(x_{kh} - x_{lh}) |x_{kh} - x_{lh}|^{r-1} (\delta_{kg} - \delta_{lg}) \right] \tag{16}$$

By (1)-(5), (10) and (11) $t_m = |t_m|$ for all m and $t_m \neq 0$ for at least one m is a sufficient condition for all derivatives to vanish.

-3.4-

The minimization of the goodness of fit measures can be carried out by one of the descend methods. We are not sure yet which one of them is the best for the NMS-problem. So far we have investigated the optimal gradient method (Curry 1944), the method of parallel tangents (Shah, Buehler, & Kempthorne 1964), and a simple steepest descend method with step-size procedure based on the cosine between successive gradients (cf Kruskal 1964b, Spang 1962). This simple procedure seems to do a rather satisfactory job, because both partan and optimal gradients use too much time to compute the optimal step-size, even if we use rather efficient interpolation techniques. For the sake of completeness we list some of the formulae that can be used in optimal gradient

(*) Of course this also follows from (3.2.6) and (3.2.7) since Γ_2

and partan methods. Superscripts indicate iteration numbers.

$$x_{iq}^{(s+1)} = x_{iq}^{(s)} - \Theta g_{iq} \tag{1}$$

$$\left\{ d_{ij}^{(s+1)} \right\} 1/R = \sum \left| x_{iq}^{(s+1)} - x_{jq}^{(s+1)} \right| r =$$

$$= \sum \left| (x_{iq}^{(s)} - x_{jq}^{(s)}) - \Theta(g_{iq} - g_{jq}) \right| r \tag{2}$$

$$t_m^{(s+1)} = \left\{ d_{ij}^{(s+1)} \right\} 1/R - \left\{ d_{kl}^{(s+1)} \right\} 1/R =$$

$$= \sum \left| (x_{iq}^{(s)} - x_{jq}^{(s)}) - \Theta(g_{iq} - g_{jq}) \right| r -$$

$$\sum \left| (x_{kq}^{(s)} - x_{lq}^{(s)}) - \Theta(g_{kq} - g_{lq}) \right| r \tag{3}$$

$$\frac{\partial t_m^{(s+1)}}{\partial \Theta} = -r \sum \Big[\left| (x_{iq}^{(s)} - x_{jq}^{(s)}) - \Theta(g_{iq} - g_{jq}) \right| r-1$$

$$\operatorname{sign} \left\{ (x_{iq}^{(s)} - x_{jq}^{(s)}) - \Theta(g_{iq} - g_{jq}) \right\} (g_{iq} - g_{jq}) -$$

$$\left| (x_{kq}^{(s)} - x_{lq}^{(s)}) - \Theta(g_{kq} - g_{lq}) \right| r-1$$

$$\operatorname{sign} \left\{ (x_{kq}^{(s)} - x_{lq}^{(s)}) - \Theta(g_{kq} - g_{lq}) \right\} (g_{kq} - g_{lq}) \Big]$$

In the future we want to try out the modified gradient methods of Fletcher & Powell (1963) and Fletcher & Reeves (1964), using the fast FORTRAN-subroutines FMFP and FMCG from IBM's System/360 SSP-series. For the NMS-problem the Kruskal-type simple steepest descend process may, however, very well turn out to be the most efficient.

-3.5-

An obvious generalization of Γ_1 gives us

$$\Gamma_1 = \frac{\sum (w_m(\left| t_m \right| - t_m))^2}{4 \sum (w_m t_m)^2} \tag{1}$$

The weights w_m can be used in various ways. If $\lambda(i,j,k,l) = m$ and $\delta : A \times A \longrightarrow \mathrm{Re}$, then we may define $w_m \doteq \delta_{ij} - \delta_{kl}$. Minimizing this generalized Γ_1 cannot properly be called non-metric multidimensional scaling, because we use the numerical values of the dissimilarities.

In some cases, however, we do have more information than the ordening of the dissimilarities alone, and minimizing Γ_1 is still a rather weak procedure (compared with, for instance, maximizing the product moment correlation coefficient between δ_{ij} and d_{ij}). An entirely different use of the weights is in a form of simultaneous MDS. Suppose we have a set of K subjects. Each subject S_q, of course, defines a set $L_q \subseteq M \times M \times M \times M$. If $t_m = d_{ij}^{1/R} - d_{kl}^{1/R}$, then

$$w_m = \sum_{q=1}^{K} \textcircled{1}_q(i,j,k,l) \tag{2}$$

The function $\textcircled{1}_q : R_q \longrightarrow \{0,1\}$ is, of course, the indicator of R_q. Using these weights is a positive orthant analogue of CDARD7: (2.1.12) can be generalized by defining $Q = S_1'WS_1 - S_2'WS_2$ (cf De Leeuw 196- a, where also more sophisticated extensions of MDS are discussed, cf also McGee 1968). Let $t^{(min)}$ denote the value of t that minimizes Γ_1, and suppose $t_m^{(min)} \neq 0$ for all $m = 1,\ldots,N$. If (by some lucky choice of the weights for example)

$$w_m = (t_m^{(min)})^{-1} \tag{3}$$

then, by substituting (3) into (1),

$$\Gamma_1 = \frac{\sum(\text{sign}(t_m^{(min)}) - 1)^2}{4N} = \frac{N_-}{N} \tag{4}$$

where N_- denotes the number of negative elements in $t^{(min)}$. For Γ_2 we obtain in a similar case

$$\Gamma_2 = \frac{N_-}{N_+}$$

Observe that (3.2.6)-(3.2.9) are still valid, because $N_+ + N_- = N$. This suggests the following algorithm. Take $w_m = 1$ for all m and minimize Γ_1 (or Γ_2). This gives us elements $t_m^{(min)}$. For the next cycle of iterations take $w_m = (t_m^{(min)})^{-1}$ and minimize Γ_1 (or Γ_2) using these weights. Continue the process until the values in $t^{(min)}$ are proportional to the weights in the same cycle of iterations. In that case (4) garantoos that

$\Gamma_1 = N_- N^{-1}$ (and (5) garantees that $\Gamma_2 = N_- N_+^{-1}$), and clearly the requirements for this algorithm would be

$$(i,j,k,l) \in L \Rightarrow d_{ij} > d_{kl} \tag{6}$$

Of course the procedure is only reasonable if the data structure satisfies for each $a_i, a_j, a_k, a_l \in A$

$$(i,j,k,l) \in L \Rightarrow (k,l,i,j) \notin L \tag{7}$$

This algorithm was tried out on some relatively small examples. There were some indications of convergence (the points $t_m^{(min)}$ and w_m come to lie on the hyperbola $t_m^{(min)} w_m = C$), but the program used an enormous amount of machine time, and the experiments were stopped (cf the discussion in section 2.2).

5: MAXIMUM SOLVABLE SUBSET ALGORITHMS (NMSMSS)

-5.1-

A trivial solution is obtained when in a considerable number of cases

$$(i,j,k,l) \in L \wedge (k,l,i,j) \notin L \wedge d_{ij} = d_{kl} \qquad (1)$$

If the stimuli can be partitioned for example into two sets A_1 and A_2, and dissimilarities between elements of the same set are always less than between elements of different sets, then a one-dimensional solution with perfect fit is always possible: give all stimuli in set A_1 a coordinate value x and all stimuli in set A_2 a coordinate value $y \neq x$. Coombs, Kruskal, Shepard, Torgerson, and Guttman all state that the order relations contain not enough information. In a certain sense this is true: the distance between the two clusters can be made arbitrarily large without even violating strict monotonicity. But in a more important sense it is not true: the fact that the stimuli within the cluster can be collapsed into one single point is a peculiarity of the algorithms, because they essentially require (1.4.1) and not (1.4.2), which makes cases like (1) legitimate. For a much more subtle example of such a trivialized solution see Roskam (1968 p 61-64). The remedy is simple: use an algorithm that requires monotonocity in the sense of (1.4.2). This was already tried (unsuccessfully) with the algorithms in sections 2.2 and 3.5. In the next sections we shall try out another approach.

-5.2-

In fact the algorithms of sections 2.2 and 3.5 are rather complicated attempts to minimize the functions

$$\phi_1 = \frac{N_-}{N_- + N_+} \qquad (1)$$

and

$$\phi_2 = \frac{N_-}{N_+} \qquad (2)$$

under the condition that $N_- + N_+ = N$, i.e. that there are no zero elements in t. These coefficients also satisfy equations like (3.2.6)-(3.2.9). The reason that such complicated algorithms were tried out at all was simply that the functions Φ_1 and Φ_2 do not have derivatives at a considerable number of points. At all other points the derivatives vanish. To illustrate this, consider the following example: 4 points in one-dimensional space, r=R=1, $x_2 = 0$, $x_3 = 4$, $x_4 = 3$, the requirements are $d_{12} \geqslant d_{13} \geqslant d_{14} \geqslant d_{23} \geqslant d_{24} \geqslant d_{34}$, so N = 15. The coordinate value x_1 varies, making Γ_1, Γ_2, Φ_1, and Φ_2 functions of one variable. Some values of these functions are given in table III. Clearly the functions Φ_1 and Φ_2 are discontinuous step functions, while Γ_1 and Γ_2 vary continuously. In this case

$$\lim_{x_1 \to \infty} \Gamma_1 = \lim_{x_1 \to \infty} \Gamma_2 = 0 \tag{3}$$

and

$$\lim_{x_1 \to \infty} \Phi_1 = 0.0667 \tag{4}$$

$$\lim_{x_1 \to \infty} \Phi_2 = 0.0714 \tag{5}$$

The fact that Φ_1 and Φ_2 cannot be minimized by the standard gradient methods was the reason that the complicated processes discussed above were tried out. Another attempt to exclude ties seems to be minimizing

$$\Phi_p = \frac{1}{4} \sum \left[\frac{(|t_m| - t_m)^2}{t_m^2} \right] \tag{6}$$

But it is easy to see that $\Phi_p = N_- = N\Phi_1$, which means that it does not solve anything.

-5.3-

Moreover the problem is not to exclude ties, but we are concerned with a more general problem: to get ties on the places where we want them to occur and nowhere else (formula 1.4.2). Minimizing Φ_1 (or Φ_2) means

a more direct approximation of the algorithmic requirements (1.4.1), and (5.2.3)-(5.2.5) show that it will less readily produce trivial solutions, although it does not exclude them because

$$\Gamma_1 = 0 \Leftrightarrow \Gamma_2 = 0 \Leftrightarrow \phi_1 = 0 \Leftrightarrow \phi_2 = 0 \tag{1}$$

$$\Gamma_1 = 1 \Leftrightarrow \phi_1 = 1 \tag{2}$$

In the sequel we shall concentrate mainly on ϕ_1 and Γ_1.

-5.4-

It is possible to compare the coefficients ϕ_1 and Γ_1 (and consequently ϕ_2 and Γ_2) in yet another way. In this section we shall derive a very interesting inequality which is considerably more precise than the one that follows from the a priori bounds on ϕ_1 and Γ_1:

$$|\Gamma_1 - \phi_1| \leqslant 1 \tag{1}$$

Observe that

$$\Gamma_1 = \frac{\sum t_i^2 (\mathrm{sign}(t_i) - 1)^2}{4 \sum t_i^2} \tag{2}$$

$$\phi_1 = \frac{\sum (\mathrm{sign}(t_i))^2 (\mathrm{sign}(t_i) - 1)^2}{4 \sum (\mathrm{sign}(t_i))^2} \tag{3}$$

It follows that, if y is the N-element vector with

$$y_i = \frac{1}{4} (\mathrm{sign}(t_i) - 1)^2 \tag{4}$$

then both Γ_1 and ϕ_1 are weighted (arithmetical) means of the elements of y. Moreover for all i

$$0 \leqslant y_i \leqslant 1 \tag{5}$$

This is already a very interesting result. Let

$$\Gamma_1 = \sum v_i y_i \tag{6}$$

with

$$v_i = \frac{t_i^2}{\sum t_i^2} \tag{7}$$

and

$$\Phi_1 = \sum w_i y_i \tag{8}$$

with

$$w_i = \frac{(\text{sign}(t_i))^2}{\sum (\text{sign}(t_i))^2} \tag{9}$$

and

$$\beta =_{\text{def}} \Gamma_1 - \Phi_1 = \sum c_i y_i \tag{10}$$

with

$$c_i = v_i - w_i \tag{11}$$

Then β is a contrast of the values of y, because $\sum c_i = 0$. Consider the maximum possible value of β. By (5) and (10) this value is reached when

$$c_i \geqslant 0 \Leftrightarrow y_i = 1 \tag{12a}$$

$$c_i < 0 \Leftrightarrow y_i = 0 \tag{12b}$$

In this case β is equal to the sum of the positive c-values, and it is easy to see that $\beta_{\max} = \frac{1}{2}\sum |c_i|$. By exactly the same reasoning the minimum value of β is reached if

$$c_i < 0 \Leftrightarrow y_i = 1 \tag{13a}$$

$$c_i \geqslant 0 \Leftrightarrow y_i = 0 \tag{13b}$$

and $\beta_{\min} = -\frac{1}{2}\sum |c_i|$. It follows that

$$|\Gamma_1 - \Phi_1| \leqslant \tfrac{1}{2}\sum |c_i| \tag{14}$$

As a matter of fact this is a special case of a more general inequality for contrasts, that does not use the special features of our y. Let the vector c be arranged in such a way that $c_i \geqslant 0$ for i=1,...,k and $c_i < 0$ for i=k+1,...,N. Then

$$\sum_{i=1}^{N} c_i y_i = \sum_{i=1}^{k} |c_i|\, y_i - \sum_{i=k+1}^{N} |c_i|\, y_i \leqslant$$

$$\max_{i=1}^{k}(y_i) \sum_{i=1}^{k} |c_i| - \min_{i=k+1}^{N}(y_i) \sum_{i=k+1}^{N} |c_i| =$$

$$\tfrac{1}{2}\left(\max_{i=1}^{k}(y_i) - \min_{i=k+1}^{N}(y_i)\right)\sum_{i=1}^{N}|c_i| \leq$$

$$\tfrac{1}{2}\left(\max_{i=1}^{N}(y_i) - \min_{i=1}^{N}(y_i)\right)\sum_{i=1}^{N}|c_i| = \tfrac{1}{2}\max_{i,j=1}^{N}|y_i - y_j|\sum_{i=1}^{N}|c_i| \quad (15)$$

Moreover, by Minkovski's inequality, in our case

$$\tfrac{1}{2}\Sigma|c_i| = \tfrac{1}{2}\Sigma|v_i - w_i| \leq \tfrac{1}{2}\Sigma\left[|v_i| + |w_i|\right] = \tfrac{1}{2}\left[\Sigma v_i + \Sigma w_i\right] = 1$$

Equality iff there are r,s with $r \neq s$ in such a way that $i \neq r$ implies $v_i = 0$ and $j \neq s$ implies $w_j = 0$. Because $t_i = 0$ iff $\text{sign}(t_i) = 0$ this case is impossible, i.e.

$$|\Gamma_1 - \phi_1| \leq \tfrac{1}{2}\Sigma|c_i| < 1 \quad (17)$$

Of course the impossibility of having, at the same time, $\Gamma_1 = 0$ and $\phi_1 = 1$ or $\Gamma_1 = 1$ and $\phi_1 = 0$ follows directly from their definition too (cf 5.3.1 and 5.3.2). Observe that $\frac{1}{2}\sum|c_i|$ only depends on the absolute values of the t_i. Consider the example in table IV. No matter in what way we choose the signs in the t vector in that example, we always have $|\Gamma_1 - \phi_1| \leq .34$. In table V we have enumerated all $2^5 = 32$ cases. Clearly the lower half of this table is redundant.

-5.5-

Define an N-element vector b in which each element corresponds with an element of L:

$$(i,j,k,l) \in L \wedge (k,l,i,j) \notin L \Leftrightarrow {}^{b}\lambda_{(i,j,k,l)} = 1 \quad (1)$$

$$(i,j,k,l) \in L \wedge (k,l,i,j) \in L \Leftrightarrow {}^{b}\lambda_{(i,j,k,l)} = 0 \quad (2)$$

It follows that

$${}^{b}\lambda_{(i,j,k,l)} = 0 \Leftrightarrow {}^{b}\lambda_{(k,l,i,j)} = 0 \quad (3)$$

Now define

$$S = \sum_{(i,j,k,l) \in L} {}^{b}\lambda_{(i,j,k,l)}\,\text{sign}(d_{ij} - d_{kl}) \quad (4)$$

$$T = \sum_{(i,j,k,l) \in L} \left[\text{sign}(d_{ij} - d_{kl})\right]^2 \quad (5)$$

$$U = \sum_{(i,j,k,l) \in L} b^2 \lambda(i,j,k,l) =$$

$$= \sum_{(i,j,k,l) \in L} b \lambda(i,j,k,l) \qquad (6)$$

$$\Upsilon = \frac{S}{\sqrt{TU}}$$

Evidently $-1 \leq \Upsilon \leq +1$, and Υ is a generalization of Kendall's familiar coefficient to partial orders. Maximizing Υ seems a rather attractive algorithm because in the first place the requirements clearly are strict monotonicity with ties on the $b\lambda(i,j,k,l) = 0$ places. Observe that if we code b with zeroes and ones it is not neccessary to include both (i,j,k,l) and (k,l,i,j) in L. In the second place it seems likely that Υ behaves more nearly as a differentiable function as ϕ_1 or ϕ_2.

-5.6-

Suppose $A = A_1 \cup A_2$ and $A_1 \cap A_2 = \emptyset$. Moreover (for p=1, q=2 or q=1, p=2)

$$a_i, a_j, a_k \in A_p \wedge a_l \in A_q \Rightarrow (i,j,k,l) \in L \qquad (1)$$

In this case a trivial solution in one dimension is possible with $\Gamma_1 = \phi_1 = 0$ (as indicated in section 5.1):

$$a_i \in A_1 \Rightarrow x_i = x \qquad (2a)$$

$$a_i \in A_2 \Rightarrow x_i = y \qquad (2b)$$

with $x \neq y$. Suppose all $\binom{n}{2}$ distances were strictly ordered, suppose that A_1 has n_1 elements and A_2 has n_2 elements. Then

$$\hat{\Upsilon} = \sqrt{1 - \frac{D_{n_1} + D_{n_2}}{D_n}} \qquad (3)$$

If n = 10 for example, the possibilities are given in table VI. For n = 20 the results are given in table VII. Clearly Υ provides us with a measure of degeneracy, while Γ_1 and ϕ_1 do not.

-5.7-

In the NMSEMS-algorithm ties play aa rather peculiar role. Because we always have

$$(d_{ij}^2 - d_{kl}^2) + (d_{kl}^2 - d_{ij}^2) = 0 \tag{1}$$

it evidently follows that

$$\mu = \sum_{(i,j,k,l) \in L} (d_{ij}^2 - d_{kl}^2) =$$

$$= \sum_{(i,j,k,l) \in L} {}^{b}\lambda(i,j,k,l)(d_{ij}^2 - d_{kl}^2) \tag{2}$$

where the b are defined as in 5.5. By (2) and (5.5.4)

$$\frac{\mu}{S} = \frac{\sum {}^{b}\lambda(i,j,k,l)\,\mathrm{sign}(d_{ij}^2 - d_{kl}^2)\,|d_{ij}^2 - d_{kl}^2|}{\sum {}^{b}\lambda(i,j,k,l)\,\mathrm{sign}(d_{ij}^2 - d_{kl}^2)} \tag{3}$$

It follows that

$$\frac{\mu}{S} \leq \max_{(i,j,k,l) \in L} |d_{ij}^2 - d_{kl}^2| \tag{4}$$

Combining this with

$$S \leq N\tau \tag{5}$$

we obtain, because $\mathrm{sign}(S) = \mathrm{sign}(\tau)$,

$$\gamma = {}_{\mathrm{def}}\frac{\mu}{\tau} \leq N \max_{(i,j,k,l) \in L} |d_{ij}^2 - d_{kl}^2| \tag{6}$$

Inequality (6) is rather wasteful compared with (5.4.17). In the case described in tables IV and V we have $\gamma \leq 35$. The value of μ, τ, and γ is also given in table V .

-5.8-

In the same way as in section 5.4 it is possible to prove some other inequality relationships. Define

$$\delta = \frac{\sum |t_i|\,(\mathrm{sign}(t_i) - 1)^2}{4\sum |t_i|} = \frac{\sum (|t_i| - t_i)}{2\sum |t_i|} =$$

$$= \tfrac{1}{2}\left(1 - \frac{\mu_1}{\sum |t_i|}\right) \tag{1}$$

This new coefficient is another special case of (3.2.2), and clearly it also is a weighted mean of the elements of y, defined by (5.4.4). Let

$$z_i = \frac{|t_i|}{\sum |t_i|} \tag{2}$$

and

$$d_i = v_i - z_i \tag{3}$$

$$e_i = w_i - z_i \tag{4}$$

where v_i and w_i are defined by (5.4.7) and (5.4.9). Again

$$|\Gamma_1 - \delta| \leq \tfrac{1}{2} \sum |d_i| \tag{5}$$

and

$$|\phi_1 - \delta| \leq \tfrac{1}{2} \sum |e_i| \tag{6}$$

Substituting (1) into these inequalities

$$1 - \sum |d_i| \leq 2\Gamma_1 + \frac{\mu}{\sum |t_i|} \leq 1 + \sum |d_i| \tag{7}$$

$$1 - \sum |e_i| \leq 2\phi_1 + \frac{\mu}{\sum |t_i|} \leq 1 + \sum |e_i| \tag{8}$$

In our example (values of z_i, d_i, e_i in table IV) we obtain

$$.72 \leq 2\Gamma_1 + \frac{\mu}{20} \leq 1.28 \tag{9}$$

$$.60 \leq 2\phi_1 + \frac{\mu}{20} \leq 1.40 \tag{10}$$

In figure VIII we have plotted the sixteen values of μ and Γ_1 and we have drawn the (parallel) lines corresponding with (9) and (10):

$$\mu = 25.60 - 40\,\Gamma_1 \tag{11}$$

$$\mu = 14.40 - 40\,\Gamma_1 \tag{12}$$

Again the boundaries are quite satisfactory, again there is equality for two points.

-5.9-

The obvious similarity of Γ_1, ϕ_1, and δ can be explained quite simply. Define:

$$f(q) = \frac{\sum (|t_i| - t_i)^q}{2^q \sum |t_i|^q} \tag{1}$$

And suppose the t_i are N constants. Then

$$f(1) = \delta \tag{2}$$

$$f(2) = \Gamma_1 \tag{3}$$

$$\lim_{q \to 0} f(q) = \Phi_1 \tag{4}$$

Without loss of generality we may assume that only the first k elements of t are negative. Then

$$f(q) = \frac{\sum_{i=1}^{k} |t_i|^q}{\sum_{i=1}^{N} |t_i|^q} \tag{5}$$

And clearly f(q) satisfies (3.2.6)-(3.2.9) for all q. Another interesting case is

$$\lim_{q \to \infty} (f(q))^{1/q} = \frac{\max(|t_i| - t_i)}{2 \max(|t_i|)} \tag{6}$$

If $t_i \geqslant 0$ for all i and $t_i > 0$ for at least one i then this function vanishes. If the first k elements are negative, then

$$\max_{i=1}^{N}(|t_i| - t_i) = \max_{i=1}^{k}(|t_i| - t_i) = 2 \max_{i=1}^{k}(|t_i|) \tag{7}$$

so

$$\lim_{q \to \infty} (f(q))^{1/q} = \frac{\max_{i=1}^{k}(|t_i|)}{\max_{i=1}^{N}(|t_i|)} \tag{8}$$

Of course

$$f(q) = \frac{\sum |t_i|^q (1 - \mathrm{sign}(t_i))^q}{2^q \sum |t_i|^q} \tag{9}$$

which means that f(q) is a weighted arithmetical mean of the N numbers

$$y_i = 2^{-q}(1 - \mathrm{sign}(t_i))^q \tag{10}$$

-5.10-

In the previous section we have shown that Γ_1, ϕ_1, and δ are all members of the family of weighted means f(q). In this section we shall discuss the family

$$g(q) = \frac{\sum |t_i|^q \operatorname{sign}(t_i)}{\sum |t_i|^q} \tag{1}$$

In practice we have in most cases ${}^b\lambda_{(i,j,k,l)} = 1$ for all $(i,j,k,l) \in L$ because we prefer ties approach (1.3.1) with $\varepsilon_1 > 0$. In this case γ reduces to

$$\gamma_A = \frac{\sum \operatorname{sign}(t_i)}{N_+ + N_-} = \frac{\sum (\operatorname{sign}(t_i))^2 \operatorname{sign}(t_i)}{\sum (\operatorname{sign}(t_i))^2} = \lim_{q \to 0} g(q) \tag{2}$$

while

$$\tilde{\mu} = \frac{\mu}{\sum |t_i|} = \frac{\sum |t_i| \operatorname{sign}(t_i)}{\sum |t_i|} = g(1) \tag{3}$$

If follows that

$$|\tilde{\mu} - \gamma_A| \leq \tfrac{1}{2} \sum |z_i - w_i| \tag{4}$$

For our example this means that the difference is always less than .20.

For all q we have

$$-1 \leq g(q) \leq +1 \tag{5}$$

And for all $q > 0$

$$f(q) = \frac{1 - g(q)}{2} \tag{6}$$

$$g(q) = 1 - 2f(q) \tag{7}$$

This can be proved quite rigorously by using the binomial expansion of $(1 - \operatorname{sign}(t_i))^q$, but it is easier to note (by checking all three possible values of $\operatorname{sign}(t_i)$) that

$$2^{-q} \sum |t_i|^q (1 - \operatorname{sign}(t_i))^q = 2^{-r} \sum |t_i|^q (1 - \operatorname{sign}(t_i))^r \tag{8}$$

for all $q,r > 0$. This implies

$$f(q) = \frac{\sum |t_i|^q (1 - \operatorname{sign}(t_i))}{2 \sum |t_i|^q} \tag{9}$$

If $b_{\lambda(i,j,k,l)} = 1$ for all $(i,j,k,l) \in L$ then

$$\gamma_A = \lim_{q \to 0} g(q) = \frac{N_+ - N_-}{N_+ + N_-} = \frac{(N_+ + N_-) - 2N_-}{N_+ + N_-} =$$

$$= 1 - 2\phi_1 = 1 - 2 \lim_{q \to 0} f(q) \qquad (10)$$

In this sense (6) and (7) are also true for $q \longrightarrow 0$. We already obtained

$$1 - 2f(2) = 1 - 2\Gamma_1 = \frac{\Sigma |t_i| t_i}{\sqrt{\Sigma t_i^2 \Sigma |t_i|^2}} =$$

$$= \frac{\Sigma |t_i|^2 \text{sign}(t_i)}{\Sigma |t_i|^2} = g(2) \qquad (11)$$

(cf 3.2.4). In summary: if $q > 0$ it does not matter whether we choose for our minimization procedure one of the $g(q)$ (the maximum sum coefficients) or one of the $f(q)$ (the positive orthant coefficients). If $q \longrightarrow 0$ they both reduce to the maximum solvable subset coefficients ϕ_1 and γ_A. If we use ties approach (1.3.1) with $\varepsilon_1 > 0$ or if there are no ties at all in the data set, then $\tau_A = \gamma$.

-5.11-

Define (with ε_2 a non-negative real number)

$$d_{ij} =_1 d_{kl} \iff |d_{ij} - d_{kl}| \leq \varepsilon_2 \qquad (1)$$

$$d_{ij} >_1 d_{kl} \iff d_{ij} - d_{kl} > \varepsilon_2 \qquad (2)$$

$$d_{ij} \geqslant_1 d_{kl} \iff d_{ij} >_1 d_{kl} \vee d_{ij} =_1 d_{kl} \iff$$

$$|d_{ij} - d_{kl}| \leq \varepsilon_2 \vee d_{ij} - d_{kl} > \varepsilon_2 \iff$$

$$d_{ij} - d_{kl} \geqslant -\varepsilon_2 \qquad (3)$$

It is possible to prove that

$$d_{ij} \geqslant_1 d_{kl} \wedge d_{kl} \geqslant_1 d_{ij} \iff d_{ij} - d_{kl} \geqslant -\varepsilon_2 \wedge$$

$$d_{kl} - d_{ij} \geqslant -\varepsilon_2 \iff -\varepsilon_2 \leq d_{ij} - d_{kl} \leqslant +\varepsilon_2 \iff$$

$$|d_{ij} - d_{kl}| \leqslant \varepsilon_2 \iff d_{ij} =_1 d_{kl} \qquad (4)$$

The introduction of $\mathcal{E}_2$ means that we redefine a tie in the numerical representation, which may be more realistic because of error of computation, etc. It is important to observe that $=_1$ in general is not an equivalence relation. Although it is reflexive and symmetric, it is transitive iff $\mathcal{E}_2 = 0$. Consequently if $\mathcal{E}_2 \neq 0$ then $\geqslant_1$ is not anti-symmetric which means that it is not an order relation. It is reflexive, but it is transitive iff $\mathcal{E}_2 = 0$. The relation $>_1$ however is both transitive:

$$\begin{aligned} & d_{ij} >_1 d_{kl} \wedge d_{kl} >_1 d_{pq} \iff \\ & d_{ij} - d_{kl} > \mathcal{E}_2 \wedge d_{kl} - d_{pq} > \mathcal{E}_2 \implies \\ & d_{ij} - d_{pq} > 2\mathcal{E}_2 \geqslant \mathcal{E}_2 \end{aligned} \tag{5}$$

and asymmetric. Evidently is a strict order. Moreover it is connected because the real numbers d_{ij} and d_{kl} can always be compared. Our algorithmic requirement (1.4.1) becomes

$$(i,j,k,l) \in L \implies d_{ij} \geqslant_1 d_{kl} \tag{6}$$

It follows from (4) and (6) that we also require

$$(i,j,k,l) \in L \wedge (k,l,i,j) \in L \implies d_{ij} =_1 d_{kl} \tag{7}$$

The definitions of ϕ_1 and $\hat{\Gamma}$ can easily be adapted to the case of $\mathcal{E}_2 \neq 0$.

-5.12-

Minimization of functions such as ϕ_1, ϕ_2, $\hat{\Gamma}$, and $\hat{\Gamma}_A$ must be done by methods that do not use derivatives. Such methods have been devised by Rosenbrock (1960), Smith (1962), Powell (1962). Moreover we have the relaxation methods (Spang 1962) that change one variable at the time and interpolate for the minimum. Up to now we have only tried out the simplex method that was originally devised by Spendley et al (1962) and improved upon by Nelder & Mead (1964).

6. ALTERNATING LEAST SQUARES ALGORITHM (NMSALS)

-6.1-

In this chapter we shall discuss a project that was abandoned about two years ago and that shows the superiority of gradient methods to intuitively satisfying 'direct' techniques. The alternating least squares solution resulted from a suggestion of Van De Geer about three years ago to solve the Ross-Cliff (1964) problem iteratively. Let D_s be the off-diagonal sub-matrix of distances between elements from two different (non-overlapping) sets, say $D_s = \left\{d_{ij}^{(s)}\right\}$, i=1,...,n; j=n+1,...,n+m. If n and m are not too small then the elements of the diagonal submatrices can be located rather precisely by using

$$\max_{k=n+1}^{n+m} \left| d_{ik}^{(s)} - d_{jk}^{(s)} \right| \leqslant d_{ij} \leqslant \min_{k=n+1}^{n+m} (d_{ik}^{(s)} + d_{jk}^{(s)}) \quad (i,j<n) \qquad (1a)$$

$$\max_{k=1}^{n} \left| d_{ki}^{(s)} - d_{kj}^{(s)} \right| \leqslant d_{ij} \leqslant \min_{k=1}^{n} (d_{ki}^{(s)} + d_{kj}^{(s)}) \quad (i,j>n) \qquad (1b)$$

The index k varies over the known elements of D_s. By taking the average of the two bounds in (1a) and (1b) we obtain an initial estimate $d_{ij}^{(0)}$ of the diagonal submatrices. Complete the matrix D, apply the Young-Householder (Y-H) process (Young-Householder 1938, Torgerson 1958), use the first k principal components to compute a new matrix of (euclidean) distances $\tilde{D}$, compute

$$\beta^{(0)} = \sqrt{\frac{\sum_{i=1}^{n} \sum_{j=n+1}^{n+m} (d_{ij}^{(0)} - \tilde{d}_{ij}^{(0)})^2}{nm}} \qquad (2)$$

Replace the off-diagonal submatrices of $\tilde{D}$ by D_s and repeat the Y-H part. As soon as

$$\left| \beta^{(i)} - \beta^{(i+1)} \right| \leqslant \varepsilon_A \qquad (3)$$

the process has converged. If

$$\beta^{(i)} \leqslant \varepsilon_B$$

then we are ready, otherwise k=k+1 and restart. In general this process

-6.2-

Another process, already used by Torgerson (personal communication during NUFFIC summer course on Psychological Measurement Theory to L.J.Th. Van Der Kamp, who communicated it to me) is combined with this process. We start our discussion of it with a complete, symmetric matrix of dissimilarities, say $\Delta = \{\delta_{ij}\}$. Applying the Y-H process to this matrix gives us k components, compute distances d, Kruskal's $\hat{d}$-values and stress, reapply Y-H on $\hat{D}$, etc. The initial configuration in fairly good (although the eigenvalues must not be used to estimate dimensionality: the smallest ones may very well be negative). Convergence is rather fast in the beginning, but in the neighboorhood of the minimum the value of stress begins to rise very slowly. Moreover, the sum of the k eigenvalues begins (equally slowly) to exceed the trace of the pseudo-scalar-product matrix, indicating negative residual eigenvalues. Nevertheless, the solutions were essentially identical with Kruskal's solutions, and the program was used on some data sets. We even went through the trouble of generalizing it to conditional matrices (apply the d-$\hat{d}$-process on each row or column and make the resulting $\hat{D}$-matrix symmetric by averaging: a third least squares process), to off-diagonal matrices (combine it with the iterative Van-De-Geer-process discussed in 6.1). When the programs became more complicated, however, the asymptotic behaviour and the computer time used became less and less satisfactory. Therefore the project was stopped. The principal lessen is a warning against ALS-algorithms in general. As a matter of fact the processes are comparable with the direct MINRES-method tested by Harman & Jones (recomputing communality estimates after each principal component analysis) and to the Rao and Lawley iterative procedures for maximum likelihood factor analysis, investigated by Jöreskog (1967). We know that these procedures converge very slowly, and in some cases they even diverge. Another ALS-

algorithm is the one used by De Leeuw (1968 a) in non-metric discriminant analysis. Further numerical experience indicates that it does not minimize the normed coefficient Γ_1. De Leeuw also used an ALS-algorithm for additive conjoint measurement (combine Kruskal's d-d-process with row and column means as best least squares estimates of scale values). It diverged. For the same reason it is not unreasonable to distrust the Lingoes-Guttman (1967) algorithm for non-metric factor analysis.

-6.3-

In this section we shall discuss an ALS-version of NMSPOM, which clearly illustrates the idea of three alternating least squares processes. The vector t can be written as

$$t = Sd \tag{1}$$

where S is an N $\times$ q matrix with in each row one element equal to +1 one equal to -1, and the rest equal to zero, and d is a q-element vector with the values of $d^{1/R}$ in the appropriate order. The matrix S is constant throughout the iterations. Suppose we have a configuration $X^{(i)}$. Compute $t^{(i)}$ and $d^{(i)}$. The first least squares problem: find a vector $\tilde{t}^{(i)}$ such that

$$\tilde{t}_m^{(i)} \geqslant 0 \qquad \text{for all } m \tag{2}$$

and

$$F = \sum (\tilde{t}_m^{(i)} - t_m^{(i)})^2 \tag{3}$$

is a minimum. Evidently

$$\tilde{t}_m^{(i)} = t_m^{(i)} \qquad \text{if } t_m^{(i)} \geqslant 0 \tag{4}$$

$$\tilde{t}_m^{(i)} = 0 \qquad \text{if } t_m^{(i)} < 0 \tag{5}$$

In general

$$\tilde{t}_m^{(i)} = \max(0, t_m^{(i)}) = \tfrac{1}{2}(t_m^{(i)} + |t_m^{(i)}|) \tag{6}$$

Substitution into (3) shows that

$$F = \tfrac{1}{4}\sum (|t_m^{(i)}| - t_m^{(i)})^2 \tag{7}$$

Note the similarity with Γ_1. The second least squares problem: find a vector $\tilde{d}^{(i)}$ such that $S\tilde{d}^{(i)}$ is maximally like $\tilde{t}^{(i)}$. Obviously there may arise some trouble here, because S is singular. We add a row of ones to S (the $N+1^{th}$ row), and we add the element $\sum_{k=1}^{q} d_k^{(i)}$ to $\tilde{t}^{(i)}$ (the $N+1^{th}$ element), and we solve the second least squares problem with these enlarged matrices (the trick is simply to use an additional requirement: the $\tilde{d}^{(i)}$ must be in the same scale as the $d^{(i)}$). Evidently

$$\tilde{d}^{(i)} = (S'S)^{-1}S'\tilde{t}^{(i)} \tag{8}$$

In the case of a weak order over the upper-diagonal elements of A $\times$ A, we have for the enlarged S-matrix

$$S'S = qI \tag{9}$$

So (8) simplifies to

$$\tilde{d}^{(i)} = \frac{1}{q} S'\tilde{t}^{(i)} \tag{10}$$

If j is a q-element vector with $j_i = 1$ for all i, then

$$\sum_{k=1}^{q} \tilde{d}_k^{(i)} = j'\tilde{d}^{(i)} = \frac{1}{q} j'S'\tilde{t}^{(i)} = \frac{1}{q} (C \ldots 0 \vdots q)\, \tilde{t}^{(i)} =$$

$$= \tilde{t}_{N+1}^{(i)} = \sum_{k=1}^{q} d_k^{(i)} \tag{11}$$

which is exactly what we wanted. The final least squares problem is finding a p-dimensional representation of A, in such a way that the euclidean distances between the points maximally resemble the corresponding $\tilde{d}_k^{(i)}$ elements. This problem is solved by the Y-H process, which gives us $X^{(i+1)}$. Again, this ALS-procedure is intuitively appealing (at least to me). It is straightforward, the process is neatly separated into three distinct parts, and each time we have three unique solutions, corresponding with absolute minima. But again, although it arrives at a satisfactory solution from a practical point of view, as an algorithm for minimizing Γ_1 it diverges (cf however section 8.8).

7: COMPUTER PROGRAMS

-7.1-

Although the relevant computer programs are still in the test phase, we shall give a short description of thei basic features in this chapter. The first program is called NMSPOM. The current version is written in PL/I. We plan to test another version with the main program (input and output statements, array declarations) written in PL/I, while the computing is done in FORTRAN4 (using IBM-developed subroutines such as EIGEN, FMFP, and FMCG). The initial configuration can be either read in, created in the manner of Kruskal (1964b) or computed. In the last case we use the algorithm of chapter II, and the resulting estimate is called canonical. The minimization technique can be either optimal gradient, partan or cosine method (cf section 3.4). The function to be minimized can be either Γ_1 or Γ_2. At each iteration the program outputs the current values of Γ_1 and Γ_2, the number of negative elements in the vector t, the time, and, if $G^{(s)}$ is the matrix of partial derivatives at iteration s, and $X^{(s-1)}$ and $X^{(s)}$ are two successive configurations,

$$\alpha_1^{(s)} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}(g_{ij}^{(s)})^2}{np}} \tag{1}$$

$$\beta_1^{(s)} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}(x_{ij}^{(s-1)} - x_{ij}^{(s)})^2}{np}} \tag{2}$$

$$\alpha_2^{(s)} = \max_{i,j} \left| g_{ij}^{(s)} \right| \tag{3}$$

$$\beta_2^{(s)} = \max_{i,j} \left| x_{ij}^{(s-1)} - x_{ij}^{(s)} \right| \tag{4}$$

After each iteration the value of $\alpha_1^{(s)}$ is compared with a predetermined stopping criterion α_c. If $\alpha_1^{(s)} \leqslant \alpha_c$ (or if s equals the predetermined maximum s_{max}) then the program outputs $X^{(s)}$, $G^{(s)}$, the distances $D^{(s)}$ and

(optionally) the vector $t^{(s)}$. Other important parameters (whose use is self-evident) are the values of r, and two values p_{min} and p_{max} that bound the values of p that are to be tried out. If the initial estimate $X^{(0)}$ is canonical, then the program also outputs $X^{(0)}$, $D^{(0)}$, and the spectrum of the matrix Q (formula 2.1.12).

-7.2-

The other fundamental program we shall discuss is called NMSMSS. The current version is written mainly in PL/I, but the subroutine that evealuates the function that is to be minimized is written in FORTRAN4. Moreover the program uses the IBM-written FORTRAN sobroutine RANDU to generate np+1 (random) initial configurations. The parameters p_{min}, p_{max}, r are similar to those of NMSPOM. There is another important parameter that indicates which function must be minimized (either Γ_1, Γ_2, Φ_1, Φ_2, or Υ). The np+1-dimensional simplex always has np+1 function values associated with it, let they be denoted by y_k. After each iteration (i.e. either a contraction, a reflexion, or an expansion) we compute the standard deviation of the y_k, say s_y. At iteration s we output s, $\min_k y_k$, $\max_k y_k$, and s_y. If s_y is less than some cutoff value s_c, iterations are terminated and the program prints the configuration with the lowest y-value and the corresponding distances.

-7.3-

The NMSEMS-program uses the technique outlined in chapter 2. It is written in PL/I, but uses the SSP EIGEN. There is only one parameter, pmax. The program computes all eigenvectors, and outputs those associated with the p_{max} greatest eigenvalues. Moreover for each added dimension the program outputs the value of μ, Φ_1, and Γ_1 and the (cumulative) distances.

-7.4-

The data of NMSEMS, NMSMSS and NMSPOM consist of quadruples of indices.

In general the number of quadruples is considerable, and this coding of the data (in terms of the number of punched cards for example) is highly inefficient in the cases in which we start with numerical matrices. Therefore we have written a PL/I program SMPUNCH that analyzes numerical matrices and punches the quadruples of indices. The parameters indicate whether the matrix is off-diagonal, conditional and/or symmetric. Moreover we specify in advance whether the diagonal elements should be included, and whether ties approach (1.3.1) or (1.3.2) should be used. In both cases the value of ε_1 is another parameter. Finally we specify a value λ that takes care of incomplete matrices. If $\delta_{ij} = \lambda$ then δ_{ij} is ignored. The only printed output is the value of N when all quadruples are punched. The punched output of SMPUNCH can be used directly in NMSEMS, NMSPOM, and NMSMSS, as well as in NFAEMS, NFAPOM, and NFAMSS, the analogous series that interpretes the dissimilarities as scalar products (i.e. performs a non-metric factor analysis).

8: COMPARISON WITH OTHER ALGORITHMS

-8.1-

In this chapter we shall compare our algorithms with the most important existing algorithms. We shall only discuss the Kruskal-algorithm and its generalizations (by Gleason (1967) to conditional proximity matrices and by Roskam (1968) to all four types of data matrices of the Coombsian classification). The other important series is the Guttman-Lingoes SSA-SSAR-SSAP-SSARP-series (Guttman 1967, 1967b. For short program descriptions of Lingoes articles in Behavioural Science 1965-68). This means that we do not discuss the algorithms of Shepard (1962 a,b), Young-Torgerson-Schulman (Torgerson & Schulman 1962, Young & Torgerson 1968, Young 1968 a,b,c), Sydow (1968), and McGee (1966,1967). A demonstration of the striking algorithmic similarities between the Kruskal, Shepard and Guttman approaches in the euclidean case has already been given by Gleason (1967), who used a simple geometric model for this purpose. A discussion of the different transformation principles used by these three authors has been given by Guttman (1967 p 42-45).

-8.2-

The results of Shepard's and Young's Monte Carlo studies on the determinateness of non-metric multidimensional scaling solutions seem to be quite satisfactory, although they investigated the most tightly constrained case (Shepard 1966, Sherman& Young 1968, Young 1968). If we want to map n stimuli into a metric space with one of our programs, then we must compare n(L), the number of elements in L, with the maximum, in most cases D_n (cf section 1.2). We may define the completeness of a set of comparisons as

$$\eta = \frac{n(L)}{D_n} \qquad (1)$$

For the method of triads we find for example that $\eta = 4(n+1)^{-1}$. If A is partitioned into two subsets A_1 with n_1 elements and A_2 with n_2 elements (for example attributes and concepts), and the subject has to select for each element of A_1 the k elements of A_2 that are most like that particular element of A_1, then

$$\eta = \frac{n_1 k(n_2-k)}{D_{n_1+n_2}} \qquad (2)$$

which will in most cases be very small. The value of η affects the determinateness of the solution, and clearly we have to take the value of η into account to know for which value of Γ_1 or Φ_1 we may consider our solution satisfactory. Compare conditional proximity matrices with complete matrices. For complete matrices $\eta = 1$, for conditional matrices $\eta = 4n(n-2)^{-1}(n+1)^{-1}$ (both do not have ties). The definition of a satisfactory solution with our programs must take these relative values of η into account. Of course ultimately the question whether a particular solution should be called satisfactory is a statistical problem, but statistical work on the goodness-of-fit coefficients is rather primitive (cf McGee 1966, 1967). The only way out seems to be sampling experiments, but clearly they should be repeated for each data set under consideration (and for all values of r and p).

-8.3-

The principal difference between the NMSPOM-algorithm and the Kruskal and Guttman-Lingoes series is the generality of the first program. It can analyze all kinds of data matrices, partial orders, and even general binary relations over $(A \times A) \times (A \times A)$. It also covers the MSA-case (Lingoes 1968) for the analysis of categorical data (using a point representation of both individuals and categories as in MSA-II, while the classification boundaries have the same form as the isosimilarity contours of the L_p-norm used). The NMSEMS and NMSMSS-programs can handle

exactly the same data. This means that data collected by the method of triads, for example, do not have to be blown up to obtain a weak order by using a CDARD2-type technique (as is done by Levelt, Plomp & Van De Geer (1966), De Gruyter (1967)), or by averaging them over subjects (as is done in the classical Torgerson approach). In general the NMSPOM-program is much more subtle is much more subtle than the existing ones, because we can weight each inequality separately, count the number of unsatisfied inequalities, etc.

-8.4-

The NMSPOM and NMSMSS programs are also capable to handle a very general type of metric, more general than those of Kruskal (who requires $r \geqslant 1$, $R = 1/r$) and more general than Guttman (who uses euclidean norms in his programs, although his theoretical discussion covers a much more general class of pseudo-metrics). From an increasing number of studies it appears, however, that the size of r does not considerably affect the final configuration, although the values of the goodness of fit coefficients may differ. A comparison of stresses with different values of r, as is done by Kruskal (1964a) and Wender (1968) for example, presupposes that the sampling distributions are independent of r. Moreover we already know by Jensen's theorem, which states that

$$F(r) = \left[\sum_{k=1}^{p} | x_{ik} - x_{jk} |^{r} \right]^{1/r} \qquad (1)$$

(with $r \geqslant 1$) is an increasing function of r, that the distances are measured on a diefferent scale. The NMSEMS program essentially concentrates on the euclidean case, because this is the only case in which matrix theory and eigenvalue machinery can be used.

-8.5-

Even if $r \neq 2$ the NMSEMS solution gives a very good initial configuration. As mentioned in chapter II, it is a generalizationof Guttman's initial

configuration. Of course it is related to the initial configuration used for example by Gleason, who correlates rows or columns of proximity matrices and solves for the eigenvectors and eigenvalues of the resulting correlationmatrix. Moreover it is related to the initial configuration used in the Young-Torgerson program and in NMSALS. The better the starting point, the less chance of arriving at a local minimum. In this respect an automatic application of the Kruskal program, starting with an arbitrary configuration, may be quite dangerous. It means that it is probably a long way to the minimum, and there may be some valleys on that way. With NMSMSS, simplex method, the situation is more complicated, because we start with np+1 different configurations. It may prove to be rewarding however to include a NMSEMS solution as one of these, and even to create the other np configurations by superimposing random noise on the NMSEMS-solution. This is a matter for further investigation.

-8.6-

The NMSMSS-program is the most general program available, because it can minimize either Γ_1, Γ_2, Φ_1, Φ_2, and Υ. As a consequence the computer time used is considerably greater as the time used for NMSPOM. Because NMSPOM accepts a much more greater generality of data than the Kruskal or Guttman-Lingoes programs, it takes much more time. From a practical point of view (costs, results) the optimal rule seems to be: if the Kruskal and Guttman-Lingoes pregrams can be applied at all (without irresponsible use of the data) they must be preferred to NMSMSS and NMSPOM. If the data are partial order (possibly even inconsistent), or if a simultaneous analysis for a group of subjects must be carried out, use NMSPOM. If a trivial solution arises, try NMSMSS. With NMSEMS the advice is somewhat more difficult. The computer time used is only a tiny fraction of the time used by the other programs, the solution is

in general not less satisfactory. Trivial solutions are impossible, even the distance between the clusters is regulated automatically and cannot be made arbitrary large. Although a NMSEMS solution is in general the least satisfactory in terms of the number of violated inequalities, from a data reduction point of view it is superior to all other algorithms, especially when a set of subjects is analyzed at the same time.

-8.7-

For NMSEMS there is a clear-cut rule for the maximum number of dimensions, although that rule most certainly not be terated as a substitute for a statistical test. For the other coefficients we can only hope that in the future each experienced investigator will build up some 'subjective' sampling distribution, that tells him when to be satisfied and when not. Sampling experiments such as those of Young (1968) may be a great help.

-8.8-

A conceptual advantage of NMSPOM and NMSMSS is the fact that the minimization is a single process. Kruskal's and Guttman's processes are essentially two-phase processes: each iteration consists of the computation of $\hat{d}$ or d^* (first least squares process), and the taking of a gradient step (second process). Kruskal's $\hat{d}$ can be written as an explicit (although not differentiable) function of d (Van Eeden 1958). Nevertheless Kruskal's differentiating implies that he takes the $\hat{d}_i$ as temporary constants. At each iteration he differentiates a different function. A convergence proof is unlikely. Guttman distinguishes the two phases much more clearly and even uses two indices to indicate iteration numbers. In the NMSPOM-case we have a single process, because $|t|$ is a relatively simple function of t, and we are sure of convergence to at least a local minimum if we use the optimum gradient method (Curry 1944). In the two-phase algorithms we are not quite sure about the effects of the interactions of the two processes. The ALS-algorithms find non-iteratively the absolute minimum at each phase. Nevertheless they diverge as algorithms

for minimizing normed coefficients as Kruskal's stress, Guttman's nomalized Φ, and our Γ's. The effect of norming the configuration on the convergence of two-phase algorithms is not clear at all. Lack of convergence of NMSALS may even be due to norming. There is evidence that without norming ALS-algorithms converge to the proper solution if the system of inequalities is consistent. If the system in inconsistent then they converge to the solution with $t_m = 0$ for all m, and the more inconsistent the system is, the faster this takes place.

9: NUMERICAL EXAMPLES

-9.1-

The first examples are taken from analyses with the NMSALS-series of programs that use the d-$\hat{d}$ algorithm. We constructed an arificial example of ten points in two dimensions. Distance function $\delta_{ij} = \frac{1}{6} d_{ij}^2$. In figure IX the value of stress is portrayed as a function of iteration number. In figure X we have indicated the movements of the points from the first iteration to their final position (40^{th} iteration) by arrows. In figure XIa and b the value's of the recovered distances in the s^{th} iteration $d_{ij}^{(s)}$ are plotted against the true distances d_{ij}. For camparison purposes we have drawn in the curve $d_{ij}^{(s)} = \frac{1}{6} d_{ij}^2$ that gives the relation of $\delta_{ij} = d_{ij}^{(0)}$ to d_{ij}. The data distinctly indicate (as do similar data on examples in which zero stress can be reached) that $d_{ij}^{(\infty)} = \lim_{s \to \infty} d_{ij}^{(s)} = d_{ij}$ if there is a sufficient number of points. A second example with real-world data shows the non-convergence of NMSALS-algorithms as algorithms for minimizing normalized coefficients. Data (collected by Leo J. Th. Van Der Kamp) consisted of a symmetric proximity matrix of dissimilarities between nine Dutch political parties. The lower diagonal part of this matrix is given in table XII. Forty iterations were performed, both in one and two dimensions. The development of stress is sketched in figure XIII. The arrows indicate on which iteration the lowest value was reached. The final configuration is shown in figure XIV. Figure XV shows the way stress decreases with a 'trivial' solution. In table XVI for some values of s the values of stress and $\frac{\lambda_1 + \lambda_2}{\text{trace}}$ are shown. The solution was obtained by the program for the conditional case and is shown in table XVII.

-9.2-

The sociogram in table XVIII was obtained by defining $\delta_{ij} = 1$ iff

subject i knew the first name of subject j. Otherwise $\delta_{ij} = 0$. The data consist of 292 inequalities. In figure XIX we have plotted the two best dimensions according to the NMSEMS solution. In table XX we have given eigenvalues μ for the successive dimensions. The values of Γ_1 and Φ_1 are computed by using cumulative distances. Observe that this data set would give rise to a trivial solution if NMSPOM was used, that it would give rise to a much more valuable and constrained NMSEMS solution if subject 8 was ignored and that a NMSMSS solution would almost certainly be superior to all other solutions if subject 8 is included. In table XXI a conditional proximity matrix is given with dissimilarities between Dutch political parties. They were obtained by summing the nine similarity rank-orders (each party as standard once) of eleven subjects. The NMSEMS result is given in figure XXII, the μ's and cumulative Γ_1's and Φ_1's in table XXIII. The number of inequalities was 250.

-9.3-

The same political example used in the previous section was analyzed by NMSPOM in several ways. In figure XXIV the optimal solution with r = 2, p = 2 is given. In the case of r = 1, starting with an arbitrary configuration ($\Gamma_1^{(1)} = .554443$, $\Phi_1^{(1)} = .528$), the program used k = 47 iterations and t = 2.31 minutes to reach convergence criteria and $\Gamma_1^{(k)} = .000327$, $\Phi_1^{(kk)} = .096$. For r = 2 the corresponsing figures were $\Gamma_1^{(1)} = .559160$, $\Phi_1^{(1)} = .544$, k = 34, t = 2.14 minutes, $\Gamma_1^{(k)} = .000005$, $\Phi_1^{(k)} = .048$, For r = 3 we have $\Gamma_1^{(1)} = .555447$, $\Phi_1^{(1)} = .544$, k = 44, t = 3.00 minutes, $\Gamma_1^{(k)} = .000031$, $\Phi_1^{(k)} = .060$. The three configurations were almost identical, the minimum $\Phi_1^{(min)}$ reached while iterating was equal to .056 for r = 1, to .032 for r = 2, and to .056 for r = 3. The nine distance functions for r = 2 are given in figure XXVa-i. The $\binom{3}{2} = 3$ plots showing the relations of the three different sets of distances

in figure XXVIa-c, the plot of the distances of the NMSEMS solution against the NMSPOM solution with r = 1 and r = 2 respectively in figure XXVIIa,b.

-9.4-

At this time the results with the simplex algorithm for NMSMSS in minimizing Φ_1 are far from impressing. We analyzed the same example as in the previous section with N = 250, r = 2, p = 2, and we already know that $\Phi_1^{(min)} \lesssim .032$. The simplex however lands on a platform much too soon, as can be seen from the results of some different runs (using random initial configurations and somewhat different iteration parameters each run). In the first run $s_y^{(1)} = .040153$, $y_{max}^{(1)} = .608$, $y_{min}^{(1)} = .412$ and after k = 95 iterations and t = 13.30 minutes $s_y^{(k)} = .000001$, $y_{max}^{(k)} = .140$ and $y_{min}^{(k)} = .136$. For the second run $s_y^{(1)} = .048925$, $y_{max}^{(1)} = .632$, $y_{min}^{(1)} = .404$, k = 61, t = 13.17 minutes, $s_y^{(k)} = 0$, $y_{max}^{(k)} = y_{min}^{(k)} = .148$. In the third run the simplex landed on a step with $y_{max} = y_{min} = .156$ after 12.83 minutes and 66 iterations. A final run produced $\Phi_1 = .288$ after 76 iterations. This is rather sad and we do not want to end this report without mentioning some research currently in preparation which may give a more satisfactory NMSMSS solution. Define

$$f(q) = \frac{\sum (|t_i| - t_i)^q}{2^q \sum |t_i|^q} \qquad (1)$$

We have shown that (section 5.9)

$$\lim_{q \to 0} f(q) = \Phi_1 \qquad (2)$$

NMSPOM minimizes f(2), but by making q smaller and smaller f(q) approaches Φ_1 and, of course,

$$\lim_{q \to 0} \min f(q) = \min \Phi_1 \qquad (3)$$

The proposition is to choose q for example .001 and minimize f(q). We h[illegible]

no idea about the optimal choice of q, whether this value can be used directly or must be approximated e.g. by halving q when min f(q) < min f(2q). Moreover the possibility of making q very small depends on the type of computer used. Of course minimizing

$$g(q) = \frac{\sum |t_i|^q \operatorname{sign}(t_i)}{\sum |t_i|^q} \tag{4}$$

is an equivalent procedure, because in general

$$g(q) = 1 - 2f(q) \tag{5}$$

and in particular

$$\tau_A = \lim_{q \to 0} g(q) = 1 - 2 \lim_{q \to 0} f(q) = 1 - 2\Phi_1 \tag{6}$$

Excluding trivial solutions may be done by maximizing

$$h(q) = \frac{\sum b_i |t_i|^q \operatorname{sign}(t_i)}{\sqrt{\sum b_i^2 \sum |t_i|^{2q}}} \tag{7}$$

with q very small, because

$$\lim_{q \to 0} h(q) = \tau \tag{8}$$

and using h(q) with q not very small does not make sense (still we have no idea how small very small is). The trick is, of course, that Φ_1, τ_A, and τ are discontinuous, non-differentiable step-functions, while f(q), g(q) and h(q) remain differentiable and continuous if $q \neq 0$ although they become more and more like step functions. Results of these new investigations (if successful) will be reported in later papers.

10: Literature

Attneave, F. Dimensions of similarity
American J. of Psychology, 1950, 63, 516-566

Beals, R. & Krantz, D.H. Metrics and geodesics induced by order relations
Math. Zeitschr., 1967, 101, 285-298

Beals, R., Krantz, D.H. & Tversky, A.N. Foundations of multidimensional scaling
Psychol. Rev., 1968, 75, 127-142

Benzécri, J.P. Analyse factorielle des proximités
Publ. de l'institut de statistique de L'université de Paris, I, 1964, 13, 235-282; II, 1965, 14, 65-80

Coombs, C.H. A theory of data
New York, Wiley, 1964

Coombs, C.H. General Mathematical Psychology, chapter Scaling and Data Theory
Unpublished manuscript, Univ. Michigan, May 1966

Cross, D.V. Metric properties of multidimensional stimulus control
Unpublished Ph.D. thesis, Univ. Michigan, 1965

Curry, H.D. The method of steepest descend for non-linear minimization problems
Q. J. Appl. Math., 1944, 2, 258-260

De Gruyter, D.N.M. The cognitive structure of the Dutch political parties in 1966
University of Leyden, Department of Experimental Psychology, Report E019-67, 1967

De Leeuw, J. Non-metric discriminant analysis
University of Leyden, Department of data theory for the social sciences, Research note RN006-68 (1968 a)

De Leeuw, J. Canonical discriminant analysis of relational data
University of Leyden, Department of data theory for the social sciences, Research Note RN007-68 (1968 b)

De Leeuw, J. Canonical Analysis of relational data (In preparation) (a)

De Leeuw, J. Algorithms for the numerical analysis of ordinal data structures (In preparation) (b)

Ellis, N.L. & Manning, N.H. Rote learning as a function of similarity structure and task sequence J. Exp. Psychol., 1967, 75, 292-297

Fletcher, R. & Powell, M.J.D. A rapidly convergent descend method for minimization Computer J., 1963, 6, 163-168

Fletcher, R. & Reeves, C.M. Fuction minimization by conjugate gradients Computer J., 1964, 7, 149-154

Foa, U.G. New developments in facet design and analysis Psych. Rev., 1965, 72, 262-275

Gleason, T.C. A general model for nonmetric multidimensional scaling Michigan Mathematical Psychology Program, MMPP67-3

Guttman, L. The quantification of a class of attributes: a theory and method of scale construction In Horst, P (ed): The prediction of personal adjustment New York, Social Science Research Council, 1941, p 253-312

Guttman, L. An approach for quantifying paired comparisons and rank order Ann. Math. Statist., 1946, 17, 144-163

Guttman, L. An introduction to facet design and analysis In: Proceedings of the 15th international congress of psychology Brussels, 1959, pp 130-132 (abstract)

Guttman, L. A general nonmetric technique for finding the smallest euclidean space for a configuration of points Jerusalem, Israel Institute for Applied Social

Guttman, L. A definition of dimensionality and distance for graphs (in preparation) (a)

Guttman, L. Existence theorems for smallest space analysis (In preparation) (b)

Handel, S. Classification and similarity of multidimensional stimuli Perceptual and motor skills, 1967, 24, 1191-1203

Harary, F. A graph theoretic approach to similarity relations Psychometrika, 1964, 29, 143-151

Hardy, G.H., Littlewood, J.E. & Polya, G. Inequalities Cambridge University Press, 1934, 1952

Harman, H.H. & jones, W.H. Factor analysis by minimizing residuals Psychometrika, 1967, 32, 443-482

Hartigan, J.A. Representation of similarity matrices by trees J. Am. Statist. Ass., 1967, 62, 1140-1158

Hays, D. Lattice models of similarity Acta Psychologica, 1964, 23, 347 (abstract)

Johnson, S.C. Hierarchical clustering schemes Psychometrika, 1967, 32, 241-254

Jöreskog, K.G. Some contributions to maximum likelihood factor analysis Psychometrika, 1967, 32, 443-482

Kendall, M.G. Rank correlation methods London, Griffin, 1962

Krantz, D.H. Rational distance functions for multidimensional scaling Michigan Mathematical Psychology Program, MMPP66-3

Krantz, D.H. A survey of measurement theory Michigan Mathematical Psychology Program MMPP67-4

Kruskal, J.B. Multidimensional scaling by optimizing goodness of fit to an nonmetric hypotheses Psychometrika, 1964, 29, 1-27 (a)

Kruskal, J.B. Nonmetric multidimensional scaling: a numerical method

Levelt, W.J.M., Van De Geer, J.P., & plomp, R. Triadic comparisons of musical intervals
Brit. J. Math. Statist. Psychol., 1966, 19, 163-179

Lingoes, J.C. The multivariate analysis of qualitative data
Multivariate behavioural research, 1968, 3, 61-94

Lingoes, J.C, & Guttman, L. Nonmetric factor analysis: a rank reducing alternative to linear factro analysis
Multivariate behavioural research, 1967, 2, 485-505

Luce, R.D. A choice theory analysis of similarity judgments
Psychometrika, 1961, 26, 151-163

McGee, V.E. The multidimensional analysis of 'elastic' distances
Brit. J. Math. Statist. Psychol., 1966, 19, 181-196

McGee, V.E. A reply to some criticisms of elastic multidimensional scaling
Brit. J. Math. Statist. Psychol., 1967, 20, 243-248

McGee, V.E. Multidimensional scaling of N sets of similarity measures: a nonmetric individual differences approach
Multivariate behavioural research, 1968, 3, 233-248

Nelder, J.A. & Mead, R. A simplex method for function minimization
Computer j., 1964, 7, 308-313

Powell, M.J.D. An efficient method for finding the minimum of a function of several variables without calculating derivatives
Computer j., 1964, 7, 155-161

Restle, F. A metric and an ordening on sets
Psychometrika, 1959, 24, 207-220

Roskam, E.E. Ch.I. Metric analysis of ordinal data in psychology
Voorschoten (netherlands), VAM, 1968

Rosenbrock, H.H. An automatic method of finding the greatest or least value of a function
Computer J., 1960, 3, 175-

Ross, J. & Cliff, N. A generalisation of the interpoint distance model
Psychometrika, 1964, 29, 167-176

Shah, B.V., Buehler, R.J. & Kempthorne, O. Some algorithms for minimizing a function of several variables
J. SIAM, 1964, 12, 74-92

Shepard, R.N. Stimulus and response generalization: a stochastic model relating generalization to distnace in psychological space
Psychometrika, 1957, 22, 325-345

Shepard, R.N. The analysis of proximities: multidimensional scaling with an unknown distance function I
Psychometrika, 1962, 27, 125-140

Shepard, R.N. The analysis of proximities: multidimensional scaling with an unknown distance function II
Psychometrika, 1962, 27, 219-246

Shepard, R.N. Polynomial fitting in the analysis of proximities
Acta Psychologica, 1964, 23, 345-346 (abstarct)(a)

Shepard, R.N. Attention and the metric structure of the stimulus space
J. Math. Psychol., 1964, 1, 54-87 (b)

Shepard, R.N. Metric structures in ordinal data
J. Math. Psychol., 1966, 3, 287-315

Shepard, R.N. & Carroll, J.D. Parametric representation of nonlinear data structures
In: Krishnaiah, P.R. (ed): Multivariate Analysis
New York, Academic Press, 1966, p 561-592

Sherman, C.F. & Young, F.W. Nonmetric multidimensional scaling: a Monte Carlo study
Proc. 76th annual convention APA, 1968, p 207-208

Smith, C.S. The automatic computation of maximum likelihood estimates
N.C.B. Scientific Dep., Report No SC 846/MR/40

Spendley, W., Himsworth, F.R. & Hext, G.R. The sequential application of simplex designs in optimisation and evolutionary operation
Technometrics, 1962, 4, p 441-

Spang, H.A. A review of minimization techniques for nonlinear functions
SIAM review, 1962, 4, 343-

Suppes, P. & Zinnes, J.L. Basic measurement theory
In: Luce, R.D., Bush, R.R., & Galanter, E. (eds) Handbook of mathematical psychology, vol I, p 1-76
New York, Wiley, 1963

Sydow, H. Ähnlichkeitsskalierung mittels allgemeiner Minkovskiräume
Paper read at the 2nd annual meeting of the Gruppe Mathematische Psychologie in Marburg/lahn, April 1968

Torgerson, W.S. Theory and methods of scaling
New York, Wiley, 1958

Torgerson, W.S. Multidimensional scaling of similarity
Psychometrika, 1965, 30, 379-393

Torgerson, W.S. & Schulman, G. Informal notes on Torgerson and Meuser's IBM 7090 program for multidimensional scaling
Department of psychology, John Hopkins University, Mimeographed report, 1962

Tversky, A.N. The dimensional representation and the metric structure of similarity data
Mimeographed, Center for Cognitive Studies, Harvard University, 1965

Van Eeden, C, Testing and estimating ordered parameters of probability distributions
Unpublished doctoral dissertation, University of Amsterdam, 1958

Wender, K. Interpretation of Minkowski metrics in ordinal multidimensional scaling
Paper read at the 2nd annual meeting of the Gruppe Mathematische Psychologie in Marburg/lahn April 1968

Wolfe, P. Methods of nonlinear programming
In: Abadie, J. (ed): Nonlinear programming
Amsterdam, North Holland Publishing Company, 1967

Young., F.W. A Fortran Iv program for nonmetric multidimensional scaling
The L.L. Thurstone psychometric laboratoty,
Univ. of North Carolina, Tech. Rep. No 56, 1968 (a)

Young, F.W. TORSCA-9: An IBM 360/75 FORTRAN IV program for nonmetric multidimensional scaling
Journal of marketing research, 1968, 5 (b)

Young, F.W. TORSCA-9: a FORTRAN IV program for nonmetric multidimensional scaling
Behavioural science, 1968, 13, 334-344 (c)

Young, F.W. Nonmetric multidimensional scaling: development of an index of metric determinacy
The L.L. Thurstone psychometric laboratory
Univ. of North Carolina, Tech. Rep. No 68, 1968 (d)

Young, F.W. & Torgerson, W.S. TORSCA: a FORTRAN IV program for Shepard-Kruskal multidimensional scaling analysis
Behavioural science, 1967, 12, 498

Young, G. & Householder, A,S. Discussion of a set of points in terms of their mutual distances
Psychometrika, 1938, 3, 331-333

And I forgot to include in thid list:

Guttman, L. The development of nonmetric space analysis, letter to professor John Ross
Multivariate behavioural research, 1967, 3, 71-82 (b)

n	μ_1	μ_2	μ_3	N_-
1	63.3600	45.6359	42.2761	52
2	136.2813	74.6515	15.4814	164
3	288.9875	32.7266	16.4776	101
4	225.1909	42.9258	2.3513	121
5	143.4240	51.1973	36.4876	79
6	145.7715	67.0146	48.3260	106
7	146.5766	87.9850	17.9323	106
8	184.6664	118.0972	20.8070	142
9	171.2913	105.4877	55.4907	87
10	172.3293	143.1977	1.0462	129
15	398.8083	2.1988	0.0749	132
20	449.6423	7.9416	0.0210	138
22	357.6350	0.0026	0.0000	122

table I

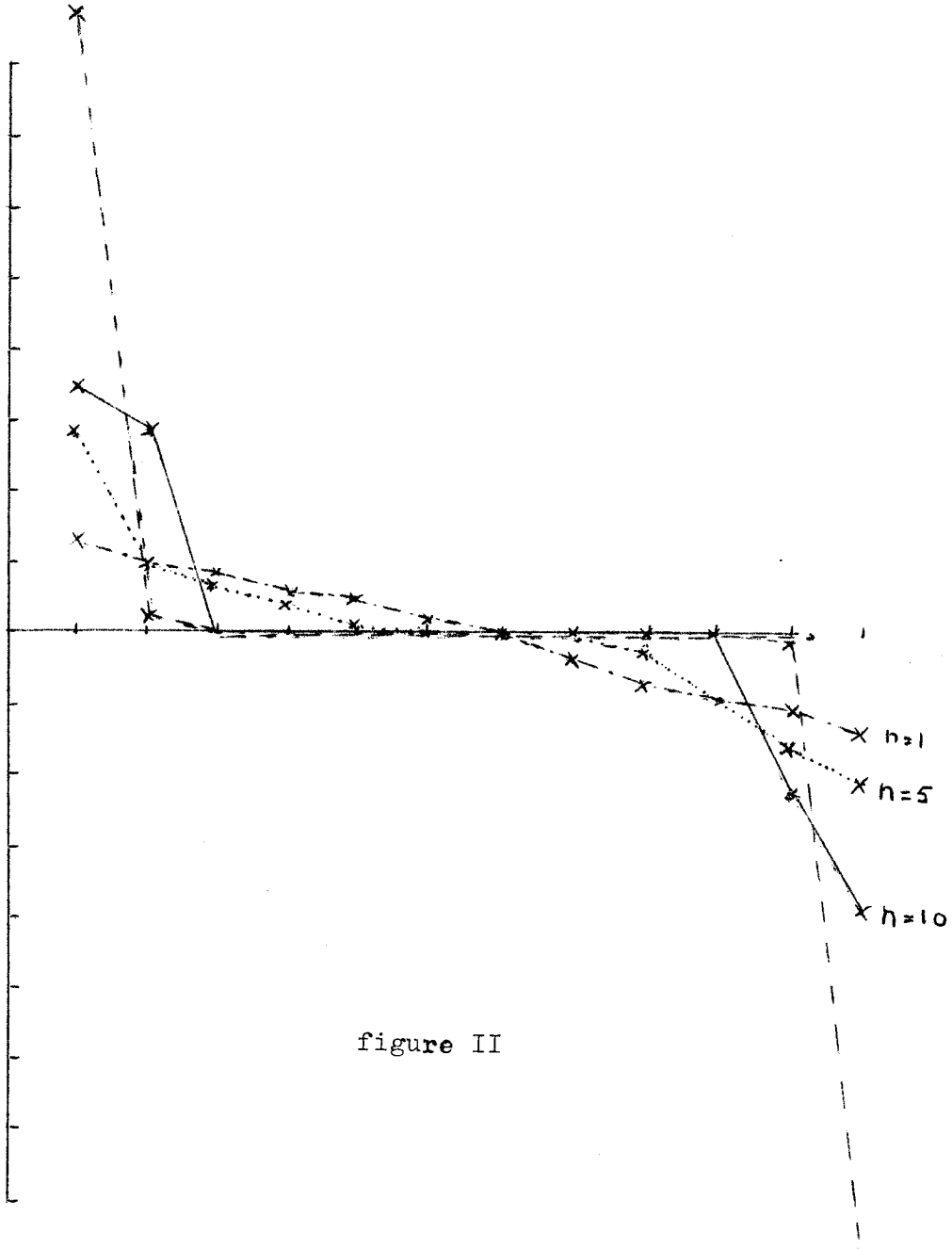


figure II

x_1	Φ_1	Φ_2	Γ_1	Γ_2
0	.4000	.8571	.6420	1.7931
1	.4667	1.6667	.5455	1.2000
2	.4000	.8571	.5610	1.2778
3	.4000	.8571	.4878	.9524
4	.4000	.8571	.4494	.8163
5	.3333	.5556	.2375	.3113
6	.2667	.4000	.0787	.0854
7	.1333	.1818	.0172	.0175
8	.0667	.0769	.0062	.0063
9	.0667	.0714	.0045	.0045
10	.0667	.0714	.0033	.0033

table III

$\lvert t_i \rvert$	v_i	w_i	z_i	c_i	d_i	e_i
3	.09	.20	.15	-.11	-.06	.05
4	.16	.20	.20	-.04	-.04	.00
7	.49	.20	.35	.29	.14	-.15
5	.25	.20	.25	.05	.00	-.05
1	.01	.20	.05	-.19	-.04	.15

table IV

n_1	n_2	$D_{n_1} + D_{n_2}$		$D_{n_1+n_2} = D_{10} = 990$
0	10	990	.000	
1	9	630	.603	
2	8	379	.786	
3	7	213	.886	
4	6	120	.937	
5	5	90	.953	

				$D_{n_1+n_2} = D_{20} = 17955$
0	20	17955	.000	
1	19	14535	.436	
2	18	11629	.594	
3	17	9183	.699	
4	16	7155	.776	
5	15	5505	.833	
6	14	4200	.875	
7	13	3213	.906	
8	12	2523	.927	
9	11	2115	.939	
10	10	1980	.943	

table VI & VII

signs of elements vector t					Γ_1	ϕ_1	δ	$\lvert\beta\rvert$	μ	τ	γ
+1	+1	+1	+1	+1	.00	.00	.00	.00	20	1.00	20.00
−1	+1	+1	+1	+1	.09	.20	.15	.11	14	.60	23.33
+1	−1	+1	+1	+1	.16	.20	.20	.04	12	.60	20.00
+1	+1	−1	+1	+1	.49	.20	.35	.29	6	.60	10.00
+1	+1	+1	−1	+1	.25	.20	.25	.05	10	.60	16.67
+1	+1	+1	+1	−1	.01	.20	.05	.19	18	.60	30.00
−1	−1	+1	+1	+1	.25	.40	.35	.15	6	.20	30.00
−1	+1	−1	+1	+1	.58	.40	.50	.18	0	.20	0.00
−1	+1	+1	−1	+1	.34	.40	.40	.06	4	.20	20.00
−1	+1	+1	+1	−1	.10	.40	.20	.30	12	.20	60.00
+1	−1	−1	+1	+1	.65	.40	.55	.25	−2	.20	−10.00
+1	−1	+1	−1	+1	.41	.40	.45	.01	2	.20	10.00
+1	−1	+1	+1	−1	.17	.40	.25	.23	10	.20	50.00
+1	+1	−1	−1	+1	.74	.40	.60	.34	−4	.20	−20.00
+1	+1	−1	+1	−1	.50	.40	.40	.10	4	.20	20.00
+1	+1	+1	−1	−1	.26	.40	.30	.14	8	.20	40.00
−1	−1	−1	+1	+1	.74	.60	.70	.14	−8	−.20	40.00
−1	−1	+1	−1	+1	.50	.60	.60	.10	−4	−.20	20.00
−1	−1	+1	+1	−1	.26	.60	.40	.34	4	−.20	−20.00
−1	+1	−1	−1	+1	.83	.60	.75	.23	−10	−.20	50.00
−1	+1	−1	+1	−1	.59	.60	.55	.01	−2	−.20	10.00
−1	+1	+1	−1	−1	.35	.60	.45	.25	2	−.20	−10.00
+1	−1	−1	−1	+1	.90	.60	.80	.30	−12	−.20	60.00
+1	−1	−1	+1	−1	.66	.60	.60	.06	−4	−.20	20.00
+1	−1	+1	−1	−1	.42	.60	.50	.18	0	−.20	0.00
+1	+1	−1	−1	−1	.75	.60	.65	.15	−6	−.20	30.00
−1	−1	−1	−1	+1	.99	.80	.95	.19	−18	−.60	30.00
−1	−1	−1	+1	−1	.75	.80	.75	.05	−10	−.60	16.67
−1	−1	+1	−1	−1	.51	.80	.65	.29	−6	−.60	10.00
−1	+1	−1	−1	−1	.84	.80	.80	.04	−12	−.60	20.00
+1	−1	−1	−1	−1	.91	.80	.85	.11	−14	−.60	23.33
−1	−1	−1	−1	−1	1.00	1.00	1.00	.00	−20	−1.00	20.00

table V

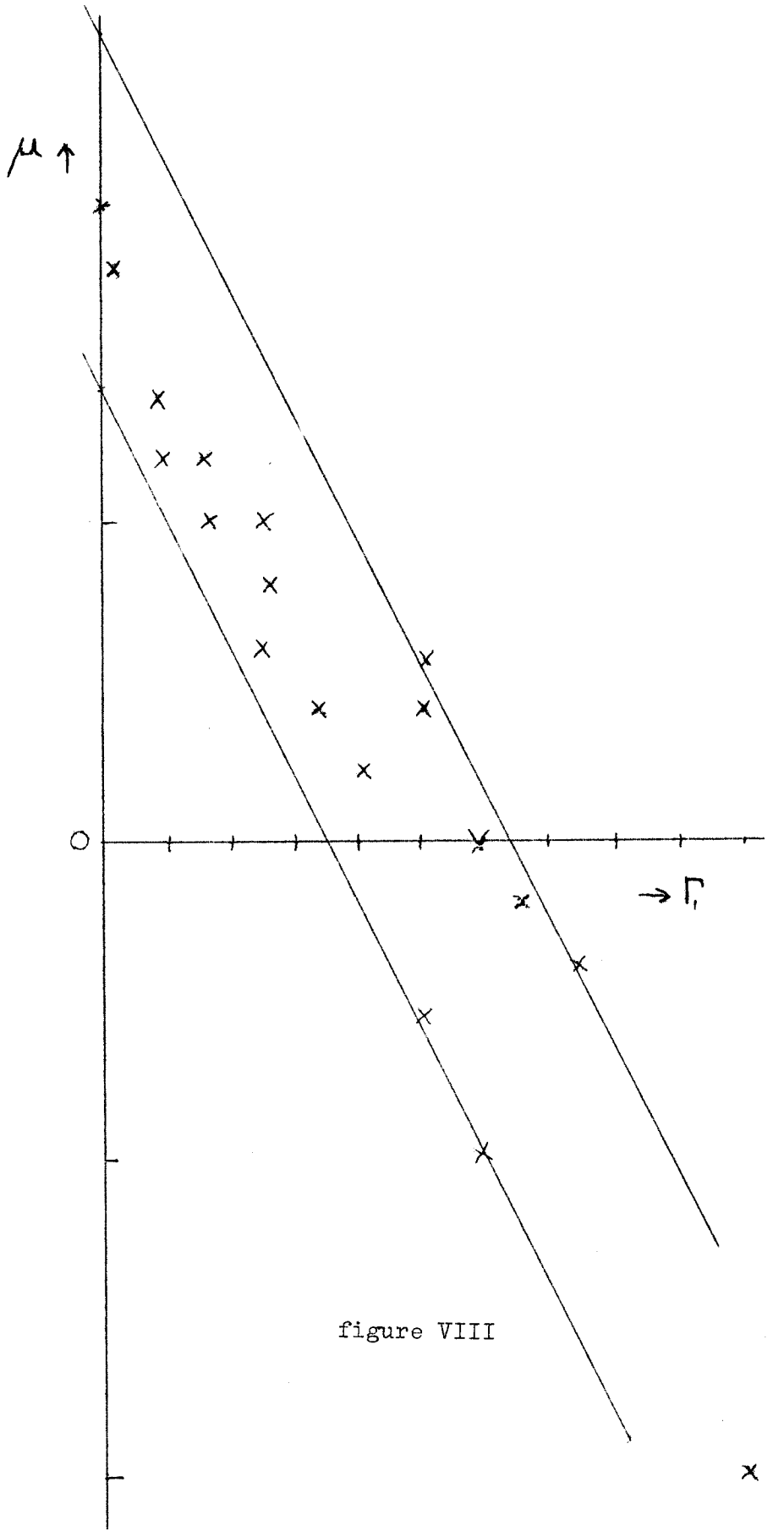


figure VIII

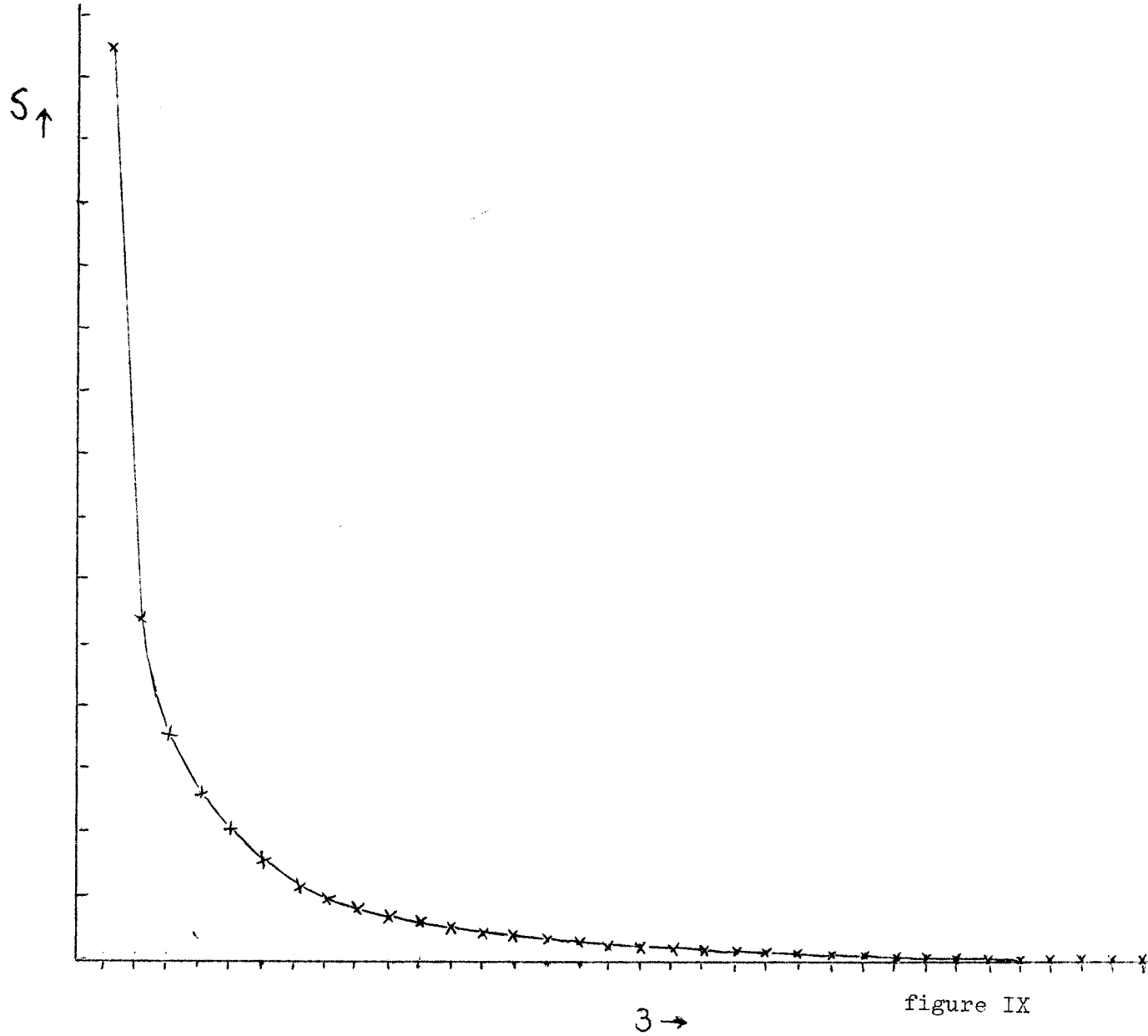
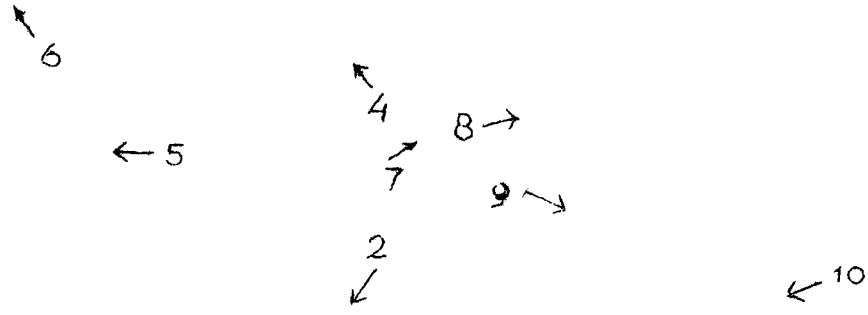


figure IX



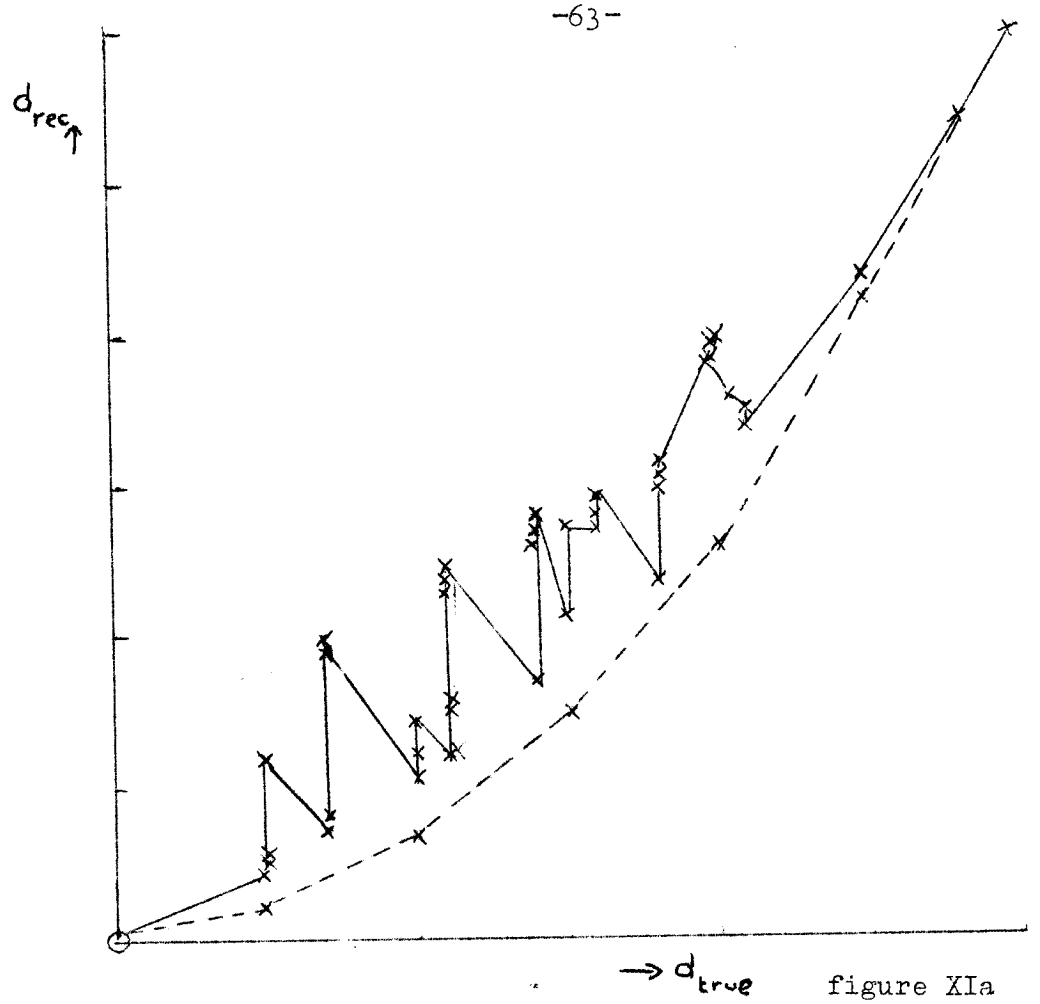


figure XIa

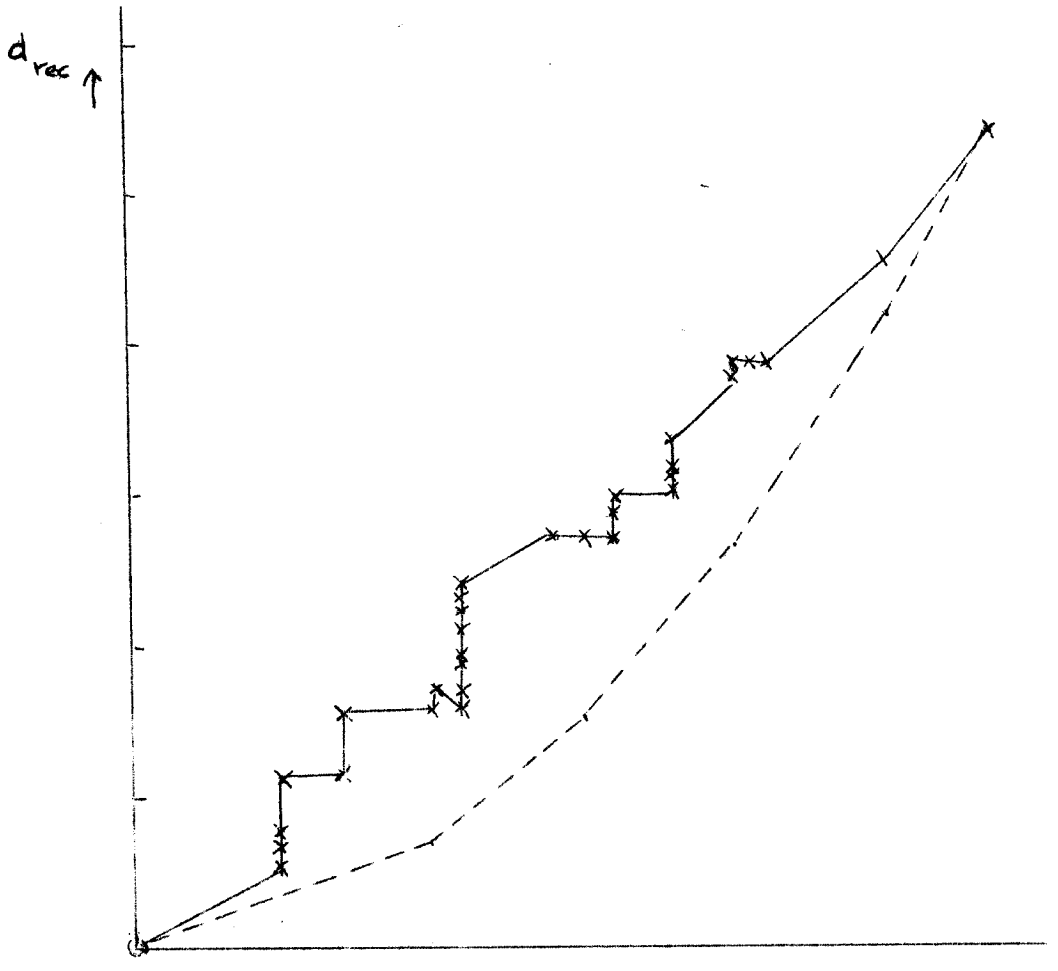


figure XIb

K.P	0.00							
PvdA	5.63	0.00						
VVD	5.27	6.72	0.00					
ARP	4.60	5.64	5.46	0.00				
CHU	4.80	6.22	4.97	3.20	0.00			
CPN	7.54	5.12	8.13	7.84	7.80	0.00		
PSP	6.73	4.59	7.55	6.73	7.08	4.08	0.00	
BP	7.18	7.22	6.90	7.28	6.96	6.34	6.88	0.00
D'66	6.17	5.47	4.67	6.13	6.04	7.42	6.36	7.36	0.00

table XII

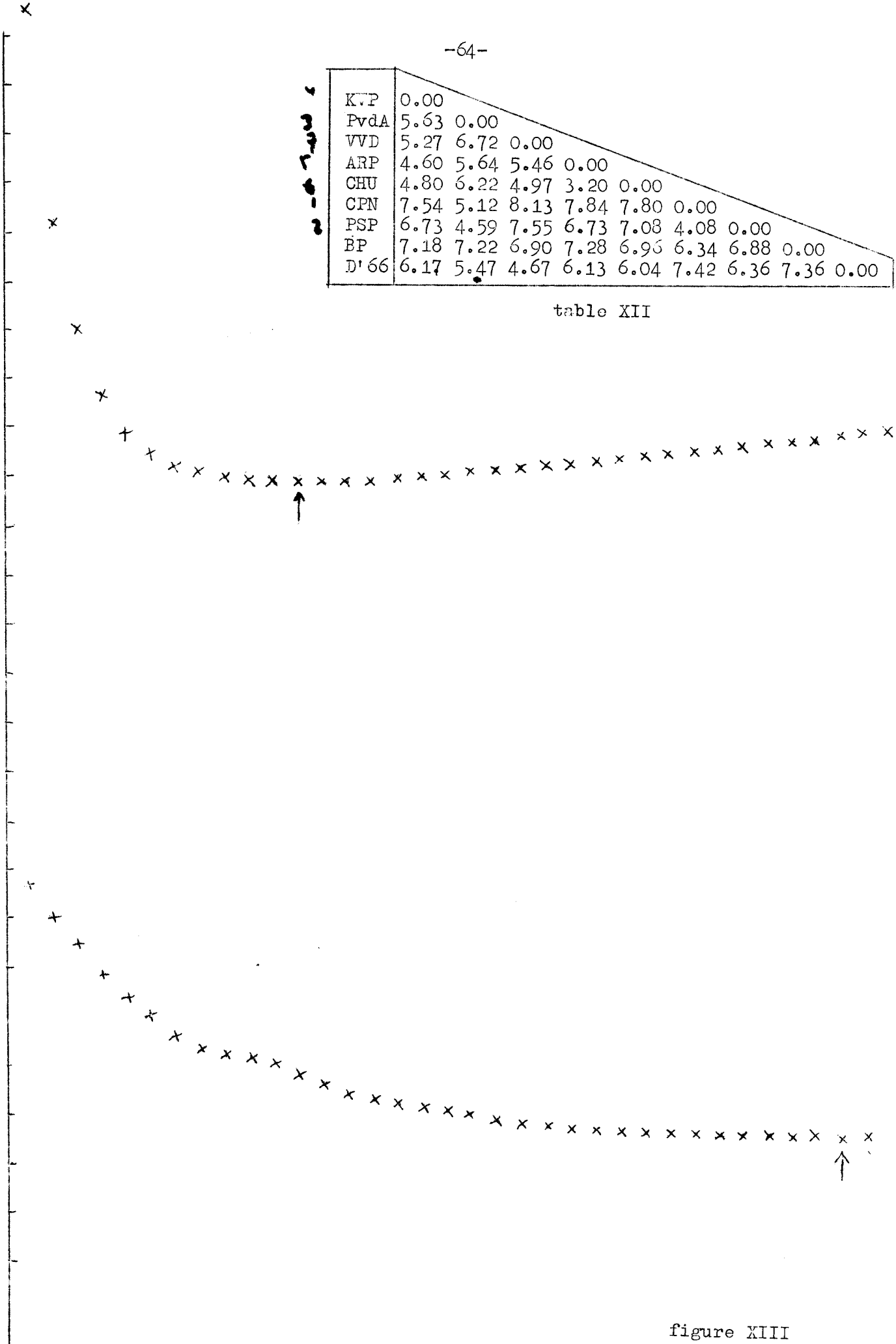


figure XIII

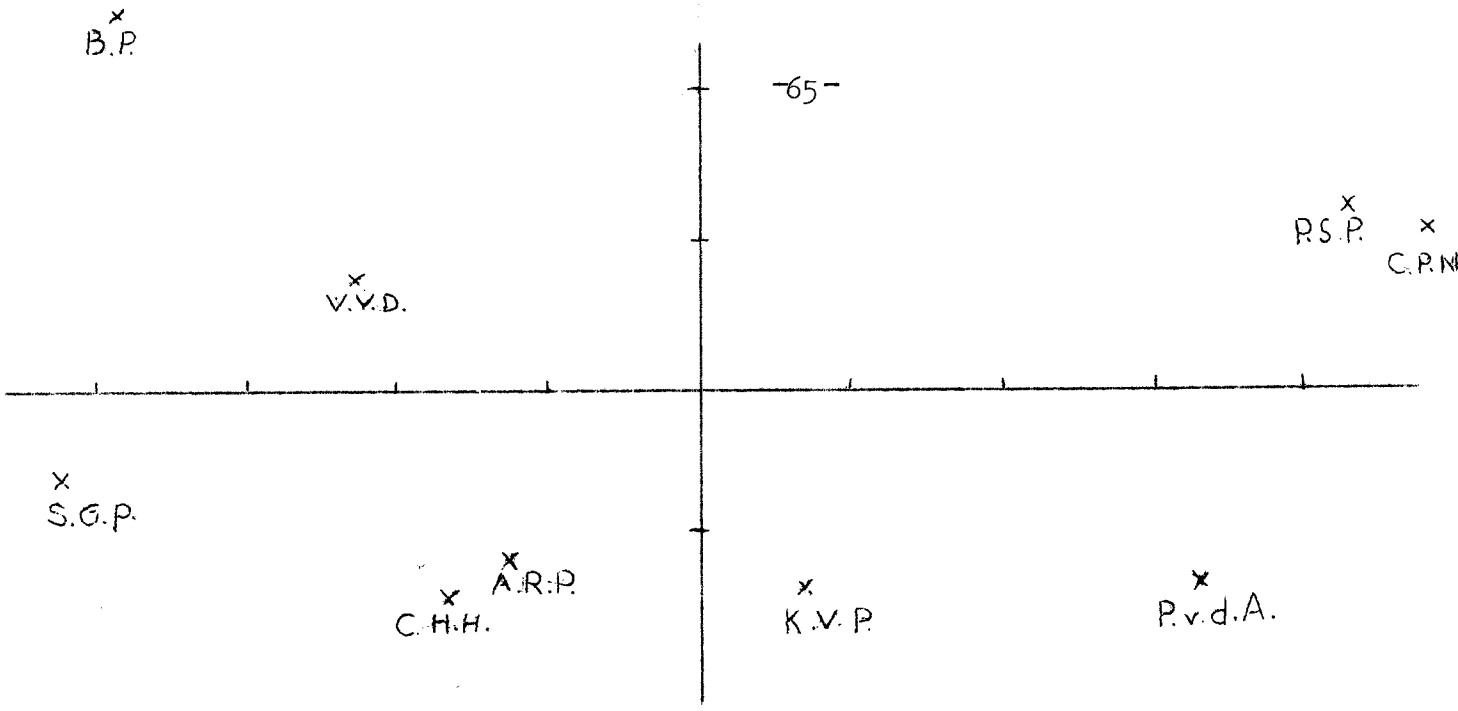


figure XIV

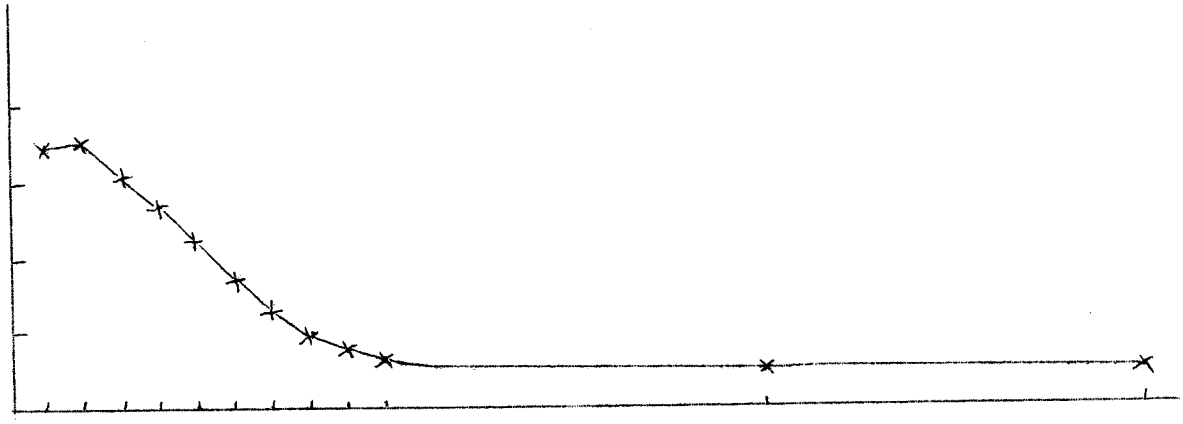


figure XV

	I	II
1	1.6034	-0.3076
2	1.6061	-0.3090
3	1.6073	-0.0890
4	1.5453	-0.0790
5	1.5811	-0.2683
6	1.4983	0.3530
7	1.4983	0.3534
8	1.4986	0.3536
9	-12.4389	-0.0073

table XVII

n	Stress	Perc.
1	.3459	.80
2	.3487	.53
10	.0668	.99
20	.0556	.99
30	.0555	1.00

table XVI

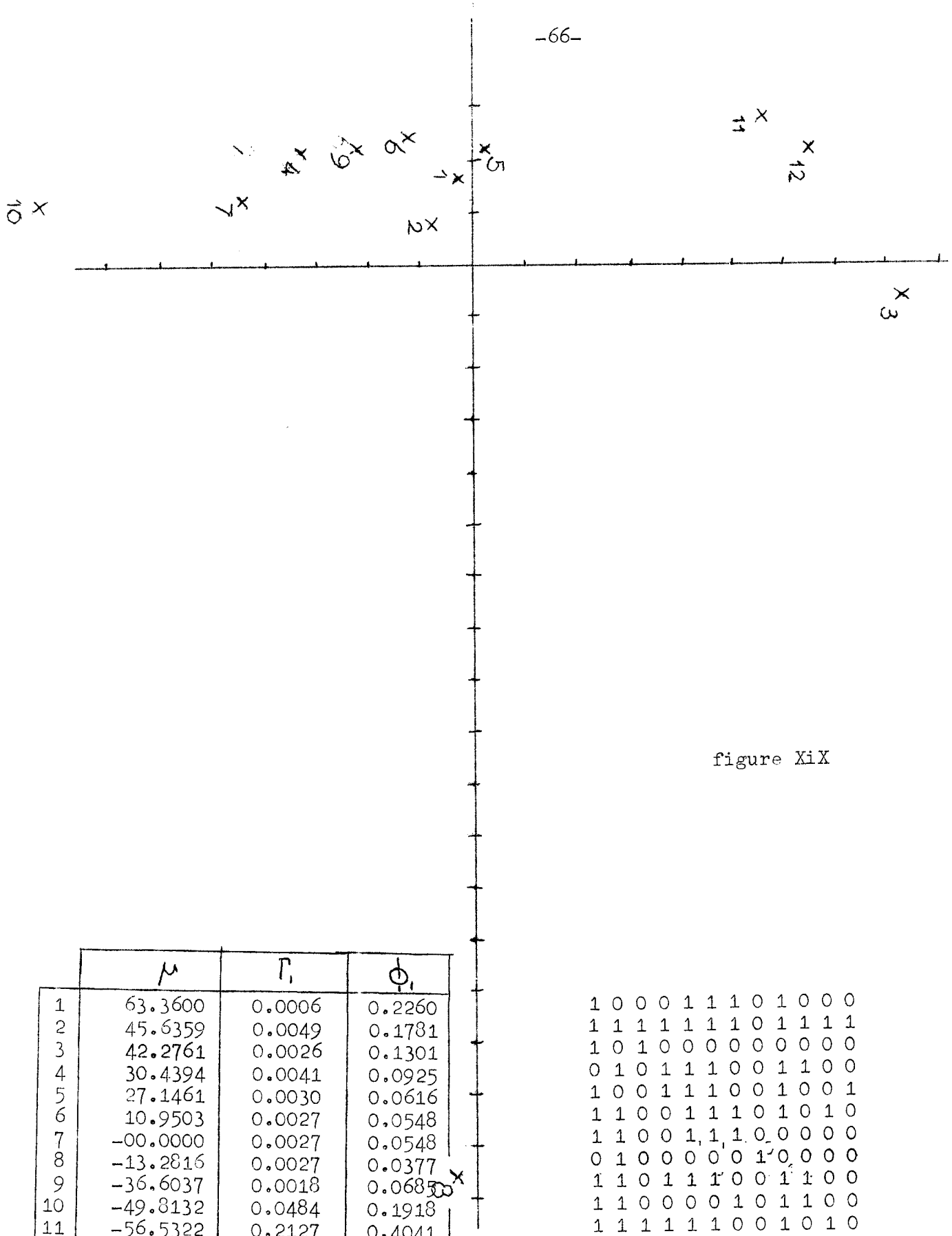


figure XiX

	μ	Γ_i	Φ_i
1	63.3600	0.0006	0.2260
2	45.6359	0.0049	0.1781
3	42.2761	0.0026	0.1301
4	30.4394	0.0041	0.0925
5	27.1461	0.0030	0.0616
6	10.9503	0.0027	0.0548
7	-00.0000	0.0027	0.0548
8	-13.2816	0.0027	0.0377
9	-36.6037	0.0018	0.0685
10	-49.8132	0.0484	0.1918
11	-56.5322	0.2127	0.4041
12	-63.5775	0.6632	0.6336

table XX

1	0	0	0	1	1	1	0	1	0	0	0
1	1	1	1	1	1	1	0	1	1	1	1
1	0	1	0	0	0	0	0	0	0	0	0
0	1	0	1	1	1	0	0	1	1	0	0
1	0	0	1	1	1	0	0	1	0	0	1
1	1	0	0	1	1	1	0	1	0	1	0
1	1	0	0	1	1	1	0	0	0	0	0
0	1	0	0	0	0	0	1	0	0	0	0
1	1	0	1	1	1	0	0	1	1	0	0
1	1	0	0	0	0	1	0	1	1	0	0
1	1	1	1	1	1	0	0	1	0	1	0
1	1	0	0	1	0	0	0	1	0	1	1

table XVIII

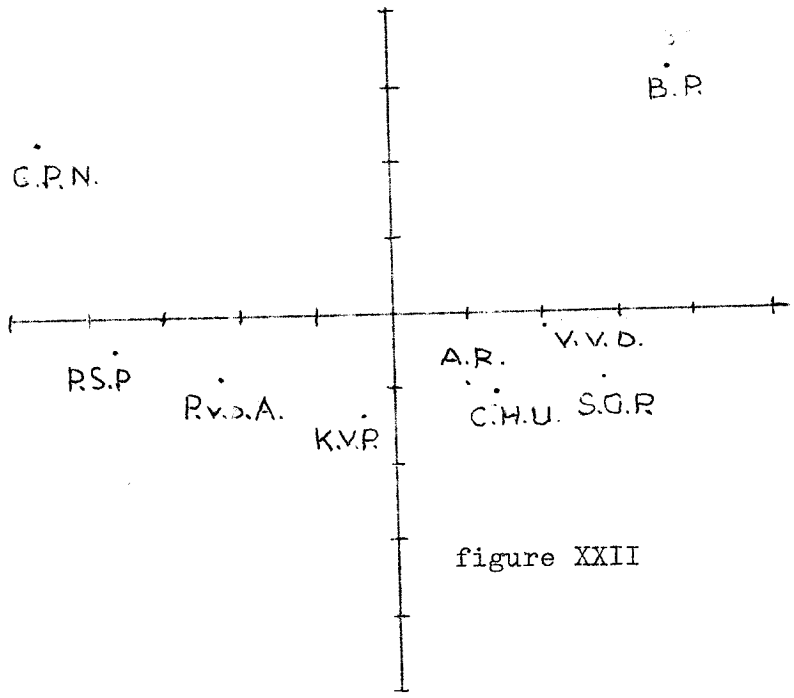


figure XXII

PvdA	11	38	90	86	40	54	66	34	76
PSP	33	11	87	83	31	60	66	49	75
BP	79	83	11	36	83	53	56	61	33
SGP	76	90	48	11	92	35	33	55	55
CPN	30	31	88	91	11	58	68	49	69
AR	62	80	73	45	94	11	28	44	54
CHU	71	85	65	54	94	28	11	46	42
KVP	39	76	83	71	81	33	47	11	54
VVD	75	83	49	53	96	44	30	54	11

table XXI

1	68.7758	0.0018	0.0720
2	21.1703	0.0042	0.0800
3	9.3872	0.0064	0.0720
4	2.4913	0.0056	0.0560
5	0.0000	0.0056	0.0560
6	- 4.5008	0.0049	0.0520
7	-21.5820	0.0107	0.1000
8	-32.9345	0.0793	0.1960
9	-42.8071	0.2583	0.3760

table XXIII

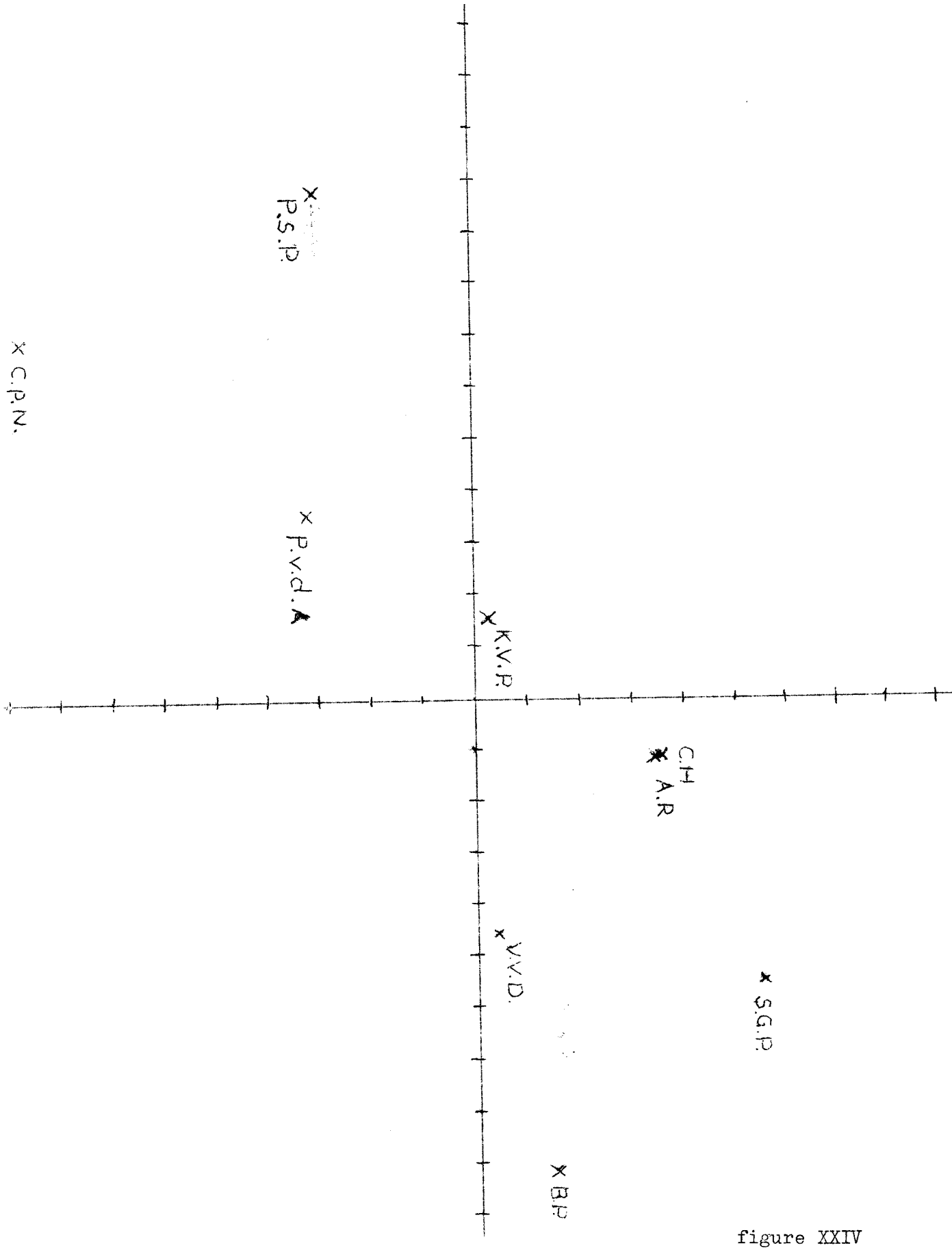


figure XXIV

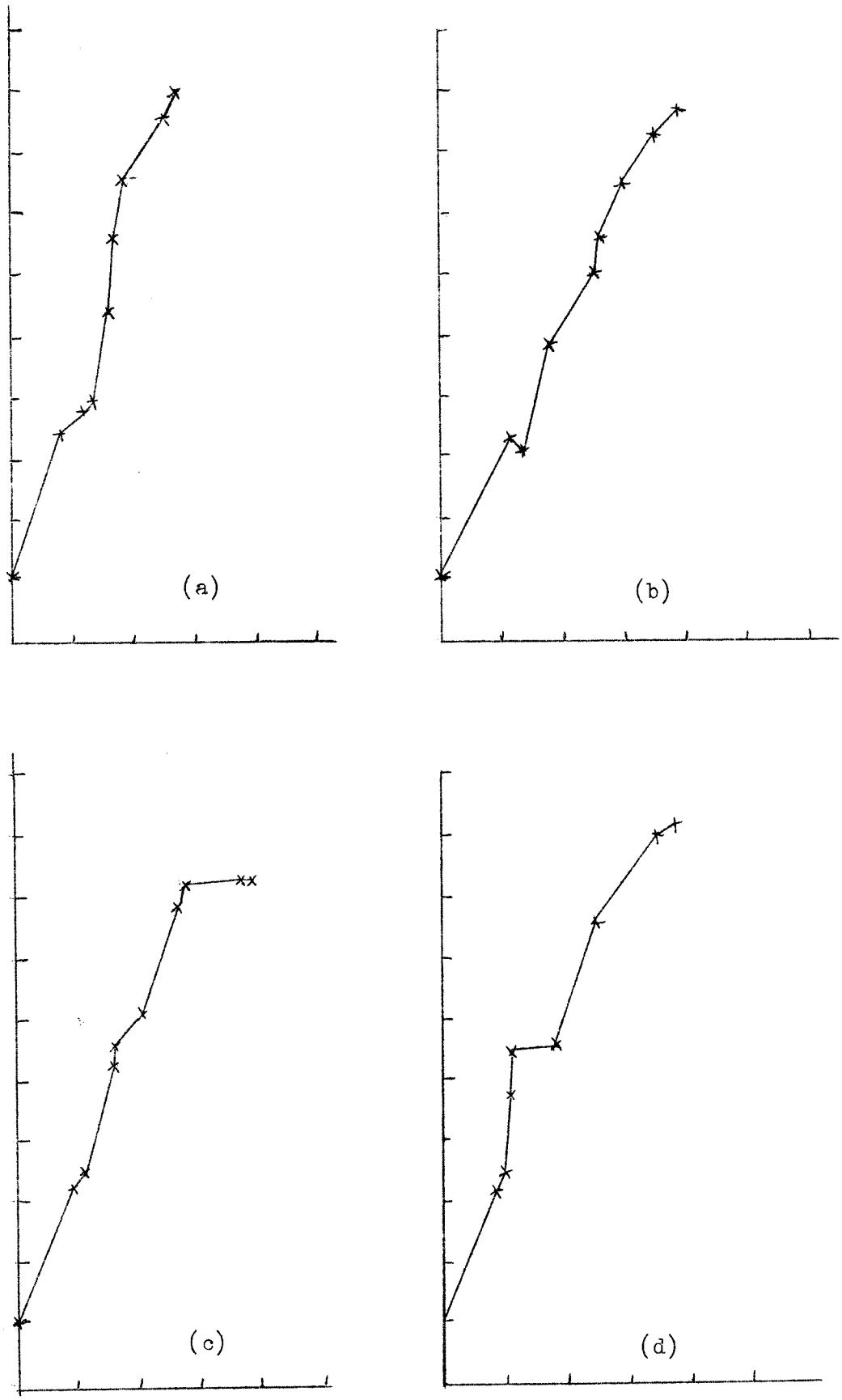


figure XXV

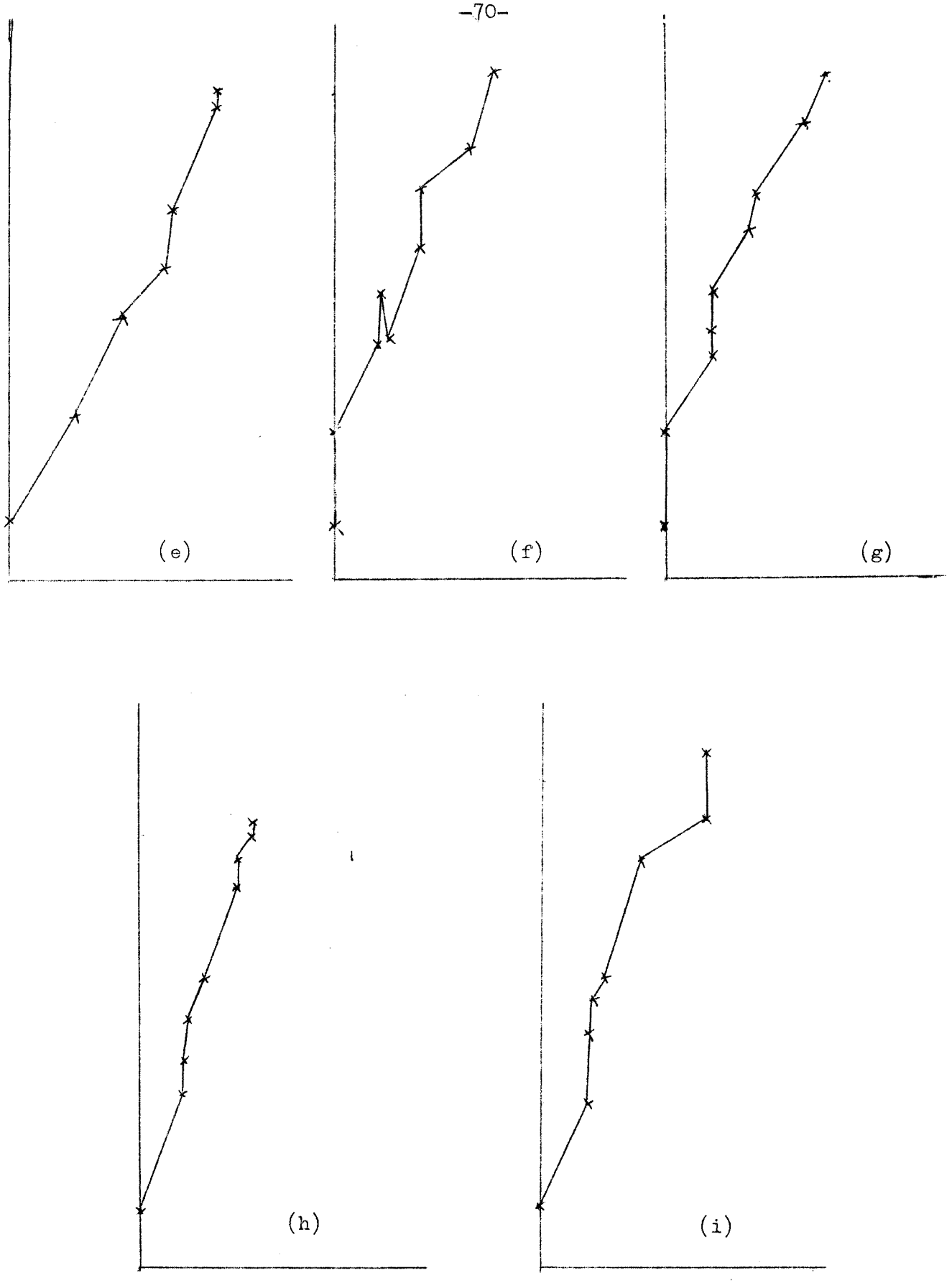


figure XXV

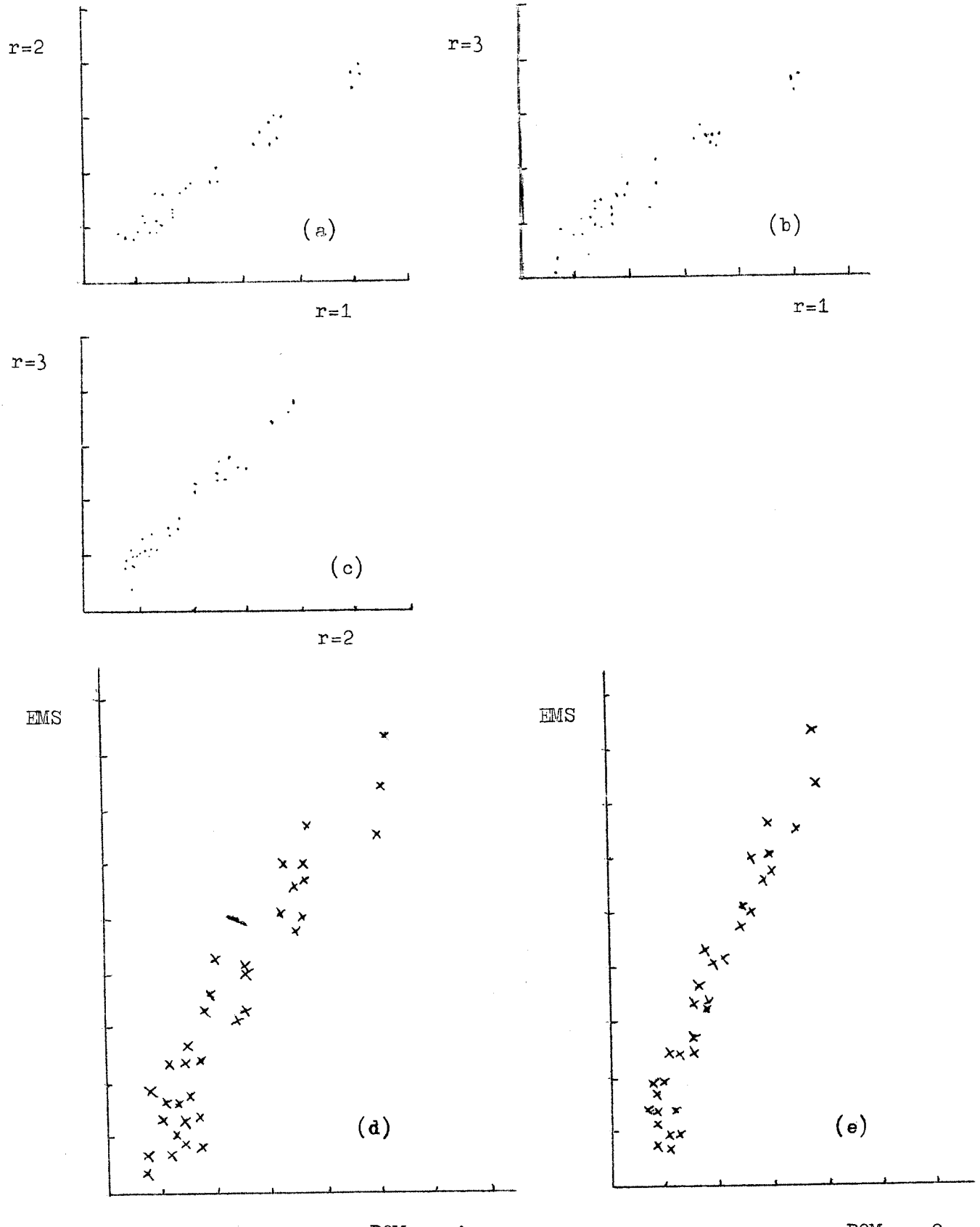


figure XXVI