RN 009-68

ROTATIONS TO OPTIMIZE CONTINUITY

Jan de Leeuw

September 1968

PSYCHOLOGICAL INSTITUTE - UNIVERSITY OF LEYDEN - NETHERLANDS

CONTENTS

SUMMARY

In this paper we have investigated some methods to rotate the results of a principal component analysis of multiple time series. In Chapter I we review the decomposition and least squares approximation of rectangular matrices. In the second chapter we formulate a tentative definition of continuity in terms of polynomials and we investigate the index η, the quotient of two unbiased variance estimates. In Chapter III the sucessive difference techniques are extended to all $\binom{n}{2}$ differences. In the final chapter a numerical example is analyzed in some different ways. Computerprograms DIFTAB, KAPPA, SSSSAX and PPPPAX are described.

I: INTRODUCTION

-1.1-

Suppose we have an ordered set Φ of functions

$$\Phi = \left\{ \phi_1, \phi_2, \ldots, \phi_n \right\} \tag{1}$$

Each ϕ_i maps the m-element ordered set R of real numbers into the reals. The values of these functions can be collected in the n x m matrix X

$$X = \left\{ x_{ij} \right\} = \left\{ \phi_i(r_j) \right\} \tag{2}$$

In the sequel we shall assume that $\det(X'X) \neq 0$, which implies that $n \geqslant m$. If X is an nxm matrix and $\det(X'X) \neq 0$, then there exist an nxm matrix K, an mxm matrix L, and an mxm diagonal matrix Λ, in such a way that K, L, and Λ obey the relationships

$$K'K = LL' = L'L = I \tag{3}$$

$$K\Lambda L' = X \tag{4}$$

It is understood that all these matrices are real. Proof: Because $\det(X'X) \neq 0$, X'X can be written as

$$X'X = P\psi P' \tag{5}$$

In (5), both P and ψ are mxm matrices, $P'P = PP' = I$, ψ is diagonal, and $\psi_{ii} > 0$ for all i (cf Gantmacher 1959, vol I, p 285). Take

$$L = P \tag{6}$$

$$\Lambda = \psi^{\frac{1}{2}} \tag{7}$$

$$K = XL\Lambda^{-1} = XP\psi^{-\frac{1}{2}} \tag{8}$$

Then

$$K\Lambda L' = XP\psi^{-\frac{1}{2}}\psi^{\frac{1}{2}}P' = XPP' = X \tag{9}$$

and

$$K'K = \psi^{-\frac{1}{2}}P'X'XP\psi^{-\frac{1}{2}} = \psi^{-\frac{1}{2}}\psi\psi^{-\frac{1}{2}} = I \tag{10}$$

Note: the condition that $\det(X'X) \neq 0$ is not quite necessary, but if $n \geqslant m$ the case in which $\det(X'X) = 0$ is only of theoretical importance, as is the case in which X'X has multiple eigenvalues. It is quite obvious that the decomposition of X given by (4) into the product of an nxm and an mxm matrix is not unique. If T is any regular mxm matrix, then

$$K\Lambda^{p}(T')^{-1}T'\Lambda^{(1-p)}L' = K\Lambda L' = X \tag{11}$$

Conversely, it is quite easy to see that is there is a decomposition of X into the nxm matrix A and the mxm regular matrix B, i.e.

$$X = AB' \tag{12}$$

then there is a regular transformation T, in such a way that

$$T'B' = L' \tag{13}$$

and

$$A(T')^{-1} = K\Lambda \tag{14}$$

-1.2-

If it is required to find an nxr matrix A and an mxr matrix B, with $1 \leqslant r \leqslant m \leqslant n$, in such a way that

$$F = \mathrm{Tr}\left\{(X-AB')'(X-AB')\right\} \tag{15}$$

is a minimum, then these matrices can be found by taking the r columns of $K\Lambda$ and L corresponding with the r greatest elements of Λ. Proof: (for a more complete development of this theorem and the theorem in the previous section see Eckart & Young (1936), Schönemann, Bock & Tucker (1965), Johnson (1963))

$$\begin{aligned} F &= \mathrm{Tr}(X'X) - 2\mathrm{Tr}(BA'X) + \mathrm{Tr}(BA'AB') = \\ &= \mathrm{Tr}(XX') - 2\mathrm{Tr}(AB'X') + \mathrm{Tr}(AB'BA') \end{aligned} \tag{16}$$

Symbolic differentiation with respect to all elements of A and B simultaneously (cf Dwyer & McPhail 1948, Wrobleski 1963, Schönemann 1965, Dwyer 1967)

$$\frac{\partial F}{\partial B} = -2X'A + 2BA'A \tag{17}$$

$$\frac{\partial F}{\partial A} = -2XB + 2AB'B \tag{18}$$

Equating all partial derivatives to zero and solving for A and B gives

$$B = X'A(A'A)^{-1} \tag{19}$$

$$A = X\,B(B'B)^{-1} \tag{20}$$

II: METHODS BASED ON SUCCESSIVE DIFFERENCES

-2.1-

In his article on the factor analytic treatment of learning curves, Tucker (1966) uses the coefficient

$$\alpha = (n-1)\delta^2 = \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2 \qquad (1)$$

to evaluate the smoothness of the resulting reference curves. Tucker uses α in a rather primitive way, classifying components with α -values greater than a certain amount as unsatisfactory and others as satisfactory. He seems to have been unaware of the work of Von Neumann et al (1941), who investigated the distribution of δ^2 in the normal case. The first two moments about the origin in samples of size n from a normal population with variance σ^2 are

$$\mu_1' = 2\sigma^2 \qquad (2)$$

$$\mu_2' = \frac{4(n^2+n-3)}{(n-1)^2}\sigma^4 \qquad (3)$$

The variance of δ^2 is consequently given by

$$\sigma^2_{\delta^2} = \frac{4(3n-4)}{(n-1)^2}\sigma^4 \qquad (4)$$

It follows that δ^2 offers an unbiased estimate of the variance with relative efficiency (as compared with the sample variance)

$$RE = \frac{2(n-1)}{3n-4} \qquad (5)$$

which means that the ARE equals 2/3. More detailed information about the distribution of δ^2 can be found in the article of Von Neumann a.o. (1941), and in several more recent papers of Kamat (1953a, 1955). Other coefficients, based on absolute differences and squared or absolute second differences, are discussed by Kamat (1953b, 1954, 1958).

-2.2-

If we evaluate curves by computing the value of δ^2 and if the goodness of fit of the least squares approximation is unaffected by non_singular transformations in general, and by (orthogonal) rotations in particular, it seems quite logical to rotate the components in such a way, that the columns of the rotated matrix have values of δ^2 that are as small as possible. More formally: we shall assume that the domain of the functions $\phi_1, \phi_2, \ldots, \phi_n$ mentioned in section 1.1 is the set $\{1, \ldots, m\}$, and that X is decomposed as follows

$$X_{(nxm)} = A_{(nxr)}B'_{(rxm)} \qquad (6)$$

The first column of B is the eigenvector of X'X associated with the dominant eigenvalue, the second column is the next eigenvector, etc. Evidently

$$B'B = I \qquad (7)$$

$$A'A = \mathrm{diag}(A'A) \qquad (8)$$

These columns are the reference curves, in the sense that the first column of B is the most representative for all functions $\phi_1, \ldots, \phi_n$ simultaneously, and so on. The problem can be posed quite simply now: find an rxr rotation matrix K, with K'K = KK' = I, in such a way that, if δ_i^2 is the value of (1) for the i-th column of

$$Y = BK \qquad (9)$$

then $\sum_{i=1}^{r} \delta_i^2$ is a minimum.

-2.3-

In optimization problems of this kind it is already a very important accomplishment if the function that is to be maximized or minimized can be written down as a quadratic or bilinear form, or as some simple function or quadratic and/or bilinear forms. A matrix formulation of the problem tends to facilitate treatment. In dealing with differences, it is very helpful to define the differencing matrices $S^{(k)}$. In this paper $S^{(k)}$ will be a (k-1) x k matrix, with

$$s_{ij}^{(k)} = \delta^{i+1,j} - \delta^{ij} \qquad (10)$$

where superscripted δ is the Kronecker operator. If $\sum_{i=1}^{k-1} s_{ij}^{(k)} = 0$ for each $j=1,\ldots,k$, then it is quite easy to see that $x_1,\ldots,x_{k-1}=0$. This means that $S^{(k)}$ is of rank $(k-1)$, and that the kxk matrix

$$D^{(k)} = S^{(k)\prime}S^{(k)} \qquad (11)$$

has one and only one zero eigenvalue. Von Neumann (1941) proves, that the eigenvalues of $D^{(k)}$ are equal to

$$\mu_i = 2 - 2\cos\frac{i\pi}{k} = 4\sin^2\frac{i\pi}{2k} \qquad i=1,\ldots,k-1 \qquad (12)$$

Of course $\mu_k = 0$.

--2.4--

If y is an n-element vector of real numbers, it follows that the (n-1)-element vector

$$\Delta^{(1)} = S^{(n)}y \qquad (13)$$

has as its general element

$$\Delta_i^{(1)} = \sum_{j=1}^{n} s_{ij}^{(n)}y_j = y_{i+1} - y_i \qquad (14)$$

By recursion we define

$$\Delta^{(0)} = y \qquad (15)$$

$$\Delta^{(k)} = S^{(n-k+1)}\Delta^{(k-1)} \qquad (16)$$

In the calculus of finite differences (jordan 1965), it is proved that

$$\Delta_i^{(k)} = \sum_{j=0}^{k} (-1)^j\binom{k}{j} y_{i+j} \qquad (17)$$

For our purposes, however, the following theorem (Hamming, 1962, p 8) is more important: If

$$y_i = a_0 + a_1 i + a_2 i^2 + \ldots + a_p i^p \qquad (18)$$

then

$$\Delta_i^{(p)} = a_p p! \qquad (19)$$

$$\Delta_i^{(p+1)} = 0 \qquad (20)$$

for all $i=1,\ldots,n$. This is called the fundamental theorem of the difference calculus. If we translate this back into matrix algebra, it means that if y_i is a polynomial of degree p in i, then

$$y'S^{(n)\prime}S^{(n-1)\prime}\ldots S^{(n-p)\prime}S^{(n-p)}\ldots S^{(n-1)}S^{(n)}y = 0 \qquad (21)$$

-2.5-

If B is the mxr matrix that must be rotated, and k is a vector of direction cosines (the first column of the rotation matrix), we maximize the "zero-degree-polynomiality" by choosing k in such a way, that

$$\lambda_1 = k'B'D^{(m)}Bk \tag{22}$$

is minimized. Here, $D^{(m)}$ is defined by (11). We impose the condition that $k'k = 1$. The second column ℓ must be found in such a way, that

$$\lambda_2 = \ell'B'D^{(m)}B\ell \tag{23}$$

is minimized, under the conditions $\ell'\ell = 1$, and $k'\ell = 0$. The matrix

$$C^{(0)} = B'D^{(m)}B \tag{24}$$

is of rank min(r,m-1), because $D^{(m)}$ is of rank m-1 (section 2.3). We suppose in this section that $r \leq m-1$, which means that $(C^{(0)})^{-1}$ exists, and that the solution for k is the eigenvector associated with the dominant eigenvalue of $(C^{(0)})^{-1}$; ℓ is the second eigenvector, and so on. This procedure finds an rxr rotation matrix K, with $K'K = KK' = I$, in such a way that the columns of $Y = BK$ are as constant as possible. The next step is to maximize first-degree-polynomiality, which can be done quite simply by computing the eigenvectors K of

$$C^{(1)} = B'S^{(m)}{}'S^{(m-1)}{}'S^{(m-1)}S^{(m)}B \tag{25}$$

The columns of $Y = BK$ are maximally linearly related to i. Observe, however, that a zero degree polynomial is a special case of a first degree polynomial, or, in other words, if $y'S^{(p)}{}'S^{(p)}y = 0$, then a forteriori $y'S^{(p)}{}'S^{(p-1)}{}'S^{(p-1)}S^{(p)}y = 0$. Addition of another S-matrix means, so to speak, a further relaxation of the continuity requirements. In general, the matrix $C^{(p)}$ is of rank min(r,m-p-1). Evidently, its order exceeds its rank whenever $m-p-1 < r$, or $p > m-r-1$. In that case there is a vector $k \neq 0$, such that $k'C^{(p)}k = 0$, which implies that the elements of the vector $y = Bk$ are a perfect p-degree polynomial function of $i=1,\ldots,m$.

-2.6-

If r=m, then the discussion in the previous section shows that $C^{(0)}$, $C^{(1)}$, ... , $C^{(m-1)}$ are all singular. Let $k^{(p)}$ be a r-element vector in such a way that

$$k^{(p)\prime}C^{(p)}k^{(p)} = 0 \qquad (26)$$

and

$$k^{(p)\prime}C^{(p-1)}k^{(p)} \neq 0 \qquad (27)$$

Moreover, let $y^{(p)} = Bk^{(p)}$. Then $y^{(p)}$ is a p-degree polynomial in i. If the r vectors $k^{(p)}$ are collected in the m x m matrix K, then K'K = KK' = I. If B'B = I (cf section 1.2), then Y'Y = K'B'BK = K'K = I, which means that the polynomials in Y are orthonormal. Actually this is a special case of the following, rather trivial, theorem: If K and L are square orthonormal matrices, i.e. K'K = KK' = LL' = L'L = I, then there exists another square orthonormal matrix of the same order, in such a way that KM = L. Proof: take M=K'L, then KM = KK'L = L, M'M = L'KK'L = I, and MM' = K'LL'K = I. Observe moreover that LM' = LL'K = K.

-2.7-

Our first definition of continuity concentrated on equally spaced data points, and a function was called continuous if it resembled a polynomial of sufficiently low degree. This is not a new use of the term "continuity" in psychometric literature. When feeded with small sets of dissimilarities the Kruskal MDSCAL, GL-SSA and McGee programs for multidimensional scaling tend to produce rather jagged distance functions. The step-like bends in the Shepard diagrams do come back in an entirely different way in the analysis of another (similar) data set. Especially when we are interested in the shape of the distance function (as with stimulus generalization data) this is a nuisance. Shepard (1964) tried to correct this by requiring that the distance functions be polynomials of a sufficiently low degree, which did result in distance functions that were more smooth and more stable. The approach to maximize smoothness outlined in the previous sections uses a similar definition of continuity. It has two obvious disadvantages. In the first place it is restricted to equally spaced data points. By using divided differences instead (Jordan 1965, p 18) we could do away with this restriction, but the relation with polynomiality becomes much more complicated. Secondly, the expected value of δ^2 depends on the variance of the population from which the data are sampled. We have seen that $\frac{1}{2}\delta^2$ is an unbiased estimate of the variance. The sample variance s_y^2, defined as

$$s_y^2 = \frac{1}{m} \sum_{i=1}^{m} (y_i - \bar{y})^2 \qquad (28)$$

can also be defined in terms of the sum of squares of all differences (Kendall & Stuart, I, p 47)

$$s_y^2 = \frac{1}{2m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} (y_i - y_j)^2 \qquad (29)$$

Clearly the estimate of σ^2 provided by δ^2 will be close to s_y^2, if the successive differences are representative for the whole set of $\frac{1}{2}m(m-1)$ differences. This representativeness will be destroyed if there is a trend in the data, and this trend does not have to be linear. For these reasons Von Neumann a.o. (1941) proposed

$$\eta = \frac{\delta^2}{s_y^2} = \frac{m}{m-1} \frac{\sum_{i=1}^{m-1} (y_{i+1} - y_i)^2}{\sum_{i=1}^{m} (y_i - \bar{y})^2} \qquad (30)$$

as a measure of trend. The distribution of η was investigated by Williams (1941), Young (1941), Von Neumann (1941, 1942). They found, among other things, for a normal parent population

$$\mu_\eta = \frac{2m}{m-1} \qquad (31)$$

$$\sigma_\eta^2 = \frac{4m^2(m-2)}{(m+1)(m-1)^2} \qquad (32)$$

The η-measure was used by Carroll explicitely as a quantitative measure of continuity for the case of equally spaced data points. Percentage points of the distribution of η in the normal case were tabulated by Hart (1942 a,b).

-2.3-

As with δ^2 we have the advantage that η can be written as a simple function of quadratic forms. If y is an m-element vector of real numbers, then

$$\eta = \frac{m}{m-1} \frac{y'D^{(m)}y}{y'[I_{mm} - \frac{1}{m} J_{mm}]y} \tag{33}$$

In (33) J_{mm} is an mxm matrix with $j_{kl} = 1$ for each (k,l), and I_{mm} is the mxm identity matrix. Define

$$E^{(m)} = \frac{m-1}{m} I_{mm} - \frac{m-1}{m^2} J_{mm} \tag{34}$$

then

$$\eta = \frac{y'D^{(m)}y}{y'E^{(m)}y} \tag{35}$$

In the context of rotation of an mxr matrix B with a vector of direction cosines k

$$\eta = \frac{k'B'D^{(m)}Bk}{k'B'E^{(m)}Bk} \tag{36}$$

Define

$$F^{(m)} = B'E^{(m)}B \tag{37}$$

then clearly $F^{(m)}$ is proportional to the variance-covariance matrix of the variates in B, and

$$\eta = \frac{k'C^{(0)}k}{k'F^{(m)}k} \tag{38}$$

Because no confusion is possible, we will drop the superscripts of $C^{(0)}$ and $F^{(m)}$ in the sequel and write simply C and F. Finding the extreme values of (38) means solving the generalized eigen-problem (Wilkinson 1965; p 337-340; Gantmacher 1959, I, p 310-326). We assume that both C and F are nonsingular, and we find k and η in such a way, that

$$Ck = \eta Fk \tag{39}$$

Evidently the value of k that maximizes η is the eigenvector associated with the dominant eigenvalue of the matrix $C^{-1}F$, that is, we solve the asymmetric eigen-problem

$$C^{-1}Fk = \eta^{-1}k \tag{40}$$

The eigenvector of $C^{-1}F$ associated with the dominant eigenvalue maximizes η^{-1}, which means that it minimizes η. Observe that the second eigenvector ℓ of the system (39) does hot in general obey the relation $k'\ell = 0$, but $k'F\ell = 0$. This means that the vectors are orthogonal in a generalized sense. To be more explicit, they are orthogonal in the skew coordinate system whose axes are defined by the equation for the unit sphere $x'Fx = 1$ (Gantmacher, 1959, I, p 315). The r eigenvectors of K can be scaled in such a way that

$$K'CK = H \tag{41}$$

$$K'FK = I \tag{42}$$

In (41) H is a diagonal matrix of η values. If $Y = BK$, then $Y'Y = K'B'BK$. Evidently a sufficient condition for $Y'Y$ to be diagonal is that the columns of B are centered. In that case

$$F = B'E^{(m)}B = \frac{m-1}{m} B'B - (m-1)m^{-2} B'JB = \frac{m-1}{m} B'B \tag{43}$$

and

$$Y'Y = K'B'BK = \frac{m}{m-1} K'FK = \frac{m}{m-1} I \tag{44}$$

III: METHODS BASED ON ALL DIFFERENCES

-3.1-

The use of successive differences in constructing a measure of smoothness seems logical enough in the analysis of equally spaced data points, because in that case successive differences are equal (or at least proportional) to the successive divided differences. The mathematical concept of continuity is translated into finite-difference terminology as follows: a unit step in the independent variable must have only a small effect on the dependent variable. By replacing the word "unit" by the word "small" we have a much more general requirement, which is not restricted to equally spaced data points, and which can take all differences into account. Accordingly, Carroll & Chang (1964) proposed the index

$$K_A = \frac{1}{s_y^2} \sum_{i} \sum_{\neq j} u_{ij}(y_i - y_j)^2 \qquad (1)$$

where u_{ij} is a decreasing function of $|x_i - x_j|$. A similar coefficient was used by Shepard & Carroll (1966):

$$K_B = \frac{1}{s_y^2} \sum_{i} \sum_{\neq j} \left[\frac{y_i - y_j}{x_i - x_j}\right]^2 v_{ij} \qquad (2)$$

Of course (2) is a special case of (1). In this paper we shall use

$$K = \frac{1}{m^2 s_y^2} \sum_{i} \sum_{< j} w_{ij}\,(y_i - y_j)^2 \qquad (3)$$

with

$$w_{ij} = |x_i - x_j|^{-p} \qquad (4)$$

These weights will be scaled in such a way that

$$\sum_{i} \sum_{< j} w_{ij} = \tfrac{1}{2}m(m-1) \qquad (5)$$

If the parameter p equals zero, then $w_{ij} = 1$ for each (i,j). Evidently in this case the value of (3) is unity (cf formula 2.29). If p=2 then (3) is a special case of (2) with v_{ij} constant for all (i,j).

-3.2-

Again we use differencing matrices. Let (i,j) be a pair of natural numbers with $i < j \leq m$. There are, of course, $\frac{1}{2}m(m-1)$ different pairs. With each pair corresponds a row of the $\frac{1}{2}m(m-1) \times m$ matrix $T^{(m)}$. Let row p correspond with pair (i,j). Then

$$t_{pq} = \delta^{iq} - \delta^{jq} \tag{6}$$

If y is an m-element vector of reals, then the vector $d = T^{(m)}y$ contains the $\frac{1}{2}m(m-1)$ differences of the elements of y. In particular the element d_p equals $y_i - y_j$. Clearly

$$T^{(m)\prime}T^{(m)} = mI_{mm} - J_{mm} \tag{7}$$

If Q is the diagonal matrix of order $\frac{1}{2}m(m-1)$, with values of w_{ij} on the appropriate places, i.e. $q_{pp} = w_{ij}$, then (writing T for $T^{(m)}$)

$$\kappa = \frac{y'T'QTy}{y'T'Ty} \tag{8}$$

If $w_{ij} = 1$ for all (i,j), then $Q = I$, and consequently $\kappa = 1$.

-3.3-

The quantity κ can be minimized quite simply by solving the pencil problem

$$B'T'QTBk = \kappa B'T'TBk \tag{9}$$

or, equivalently, the determinantal equation

$$B'T'(Q - \kappa I)BT = 0 \tag{10}$$

by methods similar to those in section 2.8.

IV: NUMERICAL EXAMPLES

The data we will use for our numerical examples were collected in a paired associate word-learning experiment. The value x_{ij} was the number of correct answers of subject i on trial j ($i=1,\ldots,n=20$; $j=1,\ldots,m=10$). We computed principal components of the matrix X'X. The sum of the first six eigenvalues, divided by Trace(X'X), gave a value of .9992. We analized the data with r=6. The diagonal elements of the matrix Λ_{66} were

λ_1	269.8496
λ_2	33.4711
λ_3	13.4725
λ_4	10.2097
λ_5	6.2902
λ_6	5.9026

The corresponding eigenvectors from L_6 are shown in figure I a-f. The elements of the eigenvalues are plotted as a function of trial number. A PL/I program called DIFTAB computed difference tables for each of the components. The difference tables themselves are not shown, but in table II we haven given the values

$$v_k = \frac{\sum_{i=1}^{n-k} (\Delta_i^{(k)})^2}{(n-k)\ \binom{2k}{k}} \qquad (1)$$

The reasons to use this particular scaling of the sum of squares can be found in Kendall & Stuart, III, p 384-393. The value of η is also given in table II. From the tables of Hart we see that under normal theory assumptions the probability $p(\frac{\delta^2}{s^2} < \eta)$ is less than 10^{-5} for the first two components, less than .01 for the third, and greater than .05 for the rest of the components. In table III we have collected the values of K for different values of p, and for each of the six components. (they were computed by the PL/I program KAPPA). In figure IV these K-values are portrayed as a function of the number of the component. Clearly, an increase in p results in more extreme κ-values, and the pattern converges rather quickly for $p \rightarrow \infty$. Compare also figure V, in which the values of η are plotted as a function of component number.

-1.2-

For the rotations discussed in chapter II, we wrote a PL/I program called PPPPAX (the program uses the FORTRAN subroutines NROOT and EIGEN from the IBM SYSTEM/360 SSP-series). Most important for the program are the parameters ITYP and JTYP. These parameters are used in the computation of two matrices P and Q. If JTYP = -1 then

$$Q = I \tag{2}$$

If JTYP = 0 then

$$Q = m^{-2}\left[B'\left\{mI - J\right\}B\right] \tag{3}$$

If JTYP = $k > 0$ then

$$Q = (m-k)^{-1}\, C^{(k-1)} \tag{4}$$

The parameter ITYP is always greater than zero, and if ITYP = ℓ, then

$$P = (m - \ell)^{-1}\, C^{(\ell - 1)} \tag{5}$$

The next step of PPPPAX is to find the vector k that minimizes

$$f = \frac{k'Pk}{k'Qk} \tag{6}$$

Clearly, η is a special case of f, with ITYP = 1, and JTYP = 0. Minimizing the mean square of the k-th differences means putting ITYP = k, and JTYP = -1. The statistical aspects of these more general coefficients were investigated by Kamat (1958). It may indeed prove to be rewarding to minimize coefficients like

$$f = \frac{\delta_2^2}{\delta^2} = \frac{m-1}{m-2}\;\frac{\sum_{i=1}^{m-2}(y_{i+2}-2y_{i+1}+y_i)^2}{\sum_{i=1}^{m-1}(y_{i+1}-y_i)^2} \tag{7}$$

which means something as maximizing second degree polynomiality, while keeping first degree polynomiality constant. Clearly minimizing a coefficient with ITYP = k and JTYP = ℓ means maximizing another coefficient with ITYP = ℓ and JTYP = k. Because we find all stationary values of the eigen-problem the solutions are identical, which means that we only have to consider cases with ITYP > JTYP. If ITYP = JTYP, then p = 1 and f is uniformly one, so there is nothing to minimize (cf K with p = 0). The six rotated components with ITYP = 1, JTYP = -1 are plotted in figure VI a-f. The components for ITYP = 2, JTYP = -1 are very similar. The six eigenvalues are presented in table VII. Some of the other results of PPPPAX: in figure VIII a and b the first two rotated components are shown with ITYP =1, JTYP = 0. Results for ITYP = 2, JTYP = 0 and ITYP = 2, JTYP = 1 are essentially identical, with one exception: the second component in the last case (figure VIII c). The components III-VI are not interesting. They all look like VIIIb amd are even somewhat flatter. The eigenvalues are collected in table IX.

-4.3-

For the rotational problem analyzed in chapter III a PL/I program called SSSSAX was written (using the same FORTRAN IV SSP's). The most important parameter was, of course, the value of p (formula 3.1.4). In figure X the results of SSSSAX with p = 1 are plotted. The results with p = 2,3,4 are very similar, and from the configurations it can be seen that they converge rather rapidly to a final configuration if $p \to \infty$. The eigenvalues are shown in table XI. Observe that the p-parameter of formula 1.1.11 is constantly taken to be unity. The more logical case with $p = \frac{1}{2}$ and the "opposite" case with p = 0 were not investigated.

V: LITERATURE

Carroll, J.D. Functional learning: the learning of continuous functional mappings relating stimulus and response continua
Educational Testing Service, Res. Bull. RB-63-26, Princeton, New Yersey.

Carroll, J.D. & Chang, J.J. A general index of nonlinear correlation and its application to the interpretation of multidimensional scaling solutions (abstract)
Am. Psychologist, 1964, 19, 540

Dwyer, P.S. Some applications od matrix derivatives in multivariate analysis
J. Am. Statist. Ass., 1967, 62, 607-625

Dwyer, P.S. & McPhail, M.S. Symbolic matrix derivatives
Ann. Math. Statist., 1948, 19, 517-534

Eckart, C. & Young, G. The approximation of one matrix by another of lower rank
Psychometrika, 1936, 1, 211-218

Gantmacher, F.R. Matrix Theory (2 volumes)
New York, Chelsea, 1959

Hamming, R. Numerical methods for scientists and engineers
New York & Tokyo, McGrawHill-Kogakusha, 1962

Hart, B.I. Tabulation of the probabilities for the ratio of the mean square successive difference to the variance
Ann. Math. Statist., 1942, 13, 207-214

Hart, B.I. Significance levels for the ratio of the mean square succesive difference to the variance
Ann. Math. Statist., 1942, 13, 445-447

Johnson, R.M. On a theorem by Eckart and Young
Psychomatrika, 1963, 28, 259-263

Jordan, C. Calculus of finite differences
New York, Chelsea, 1965

Kamat, A.R. On the mean successive difference and its ratio to the root mean square
Biometrika, 1953, 40, 116-127 (b)

Kamat, A.R. Some properties of estimates for the standard deviation based on deviations fro the mean and variate differences J. Roy. Statist. Soc., B, 1953, 15, 233-240 (a)

Kamat, A.R. Distribution theory of two estimates for standard deviation based on second variate differences Biometrika, 1954, 41, 1-11

Kamat, A.R. Modified mean square successive difference with an exact distribution Sankhya, 1955, 15, 295-302

Kamat, A.R. Contributions to the theory of statistics based on the first and second successive differences Metron, 1958, 19, 97-118

Kendall, M.G. & Stuart, A. The advanced theory of statistics (3 volumes) London, Griffin, 1963-1967

Sathe, Y.S. & Kamat, A.R. Approximations to the distributions of some measures of dispersion bsed on successive differences Biometrika, 1957, 44, 349-359

Schönemann, P.H. On the formal matrix differentiation of traces and determinants Univ. of North Carolina, Psychometric Laboratory, Research Memorandum No 27, 1965

Schönemann, P.H., Bock, R.D., and Tucker, L.R. Some notes on a theorem by Eckart and Young Univ. of North Carolina, Psychometric Laboratory, Research Memorandum No 25, 1965

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

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

Tucker, L.R. Learning theory and multivariate experiment: illustration by dtermination of generalized learning curves In: Cattell, R.B. (ed): Handbook of multivariate experimental psychology. New York, Rand-McNally, 1966, p 476-501

Von Neumann, J. — Distribution of the ratio of the mean square successive difference to the variance
Ann. Math. Statist., 1941, 12, 367-395

Von Neumann, J. — A further remark on the distribution of the ratio of the mean sqaure successive difference to the variance
Ann. Math. Stattist., 1942, 13, 86-88

Von Neumann, J., Kent, R.H., Bellinson, H.R., Hart, B.I.
The mena square successive difference
Ann. Math. Statist., 1941, 12, 153-162

Wilkinson, J.H. — The algebraic eigenvalue problem
Oxford, Clarendon, 1967

Williams, J.C. — Moments of the ratio of the mean square successive difference to the mean square difference in samples from a normal universe
Ann. Math. Statist., 1941, 12, 239-241

[illegible], J.R. — Extension of the Dwyer-McPhail matrix derivative calculus with applications to estimation problems involving errors-in-variables and errors-in-equations
Technical Report, Univ. Michigan, 1963

[illegible], L.C. — On randomness in ordered sequences
Ann. Math. Statist., 1941, 12, 293-300

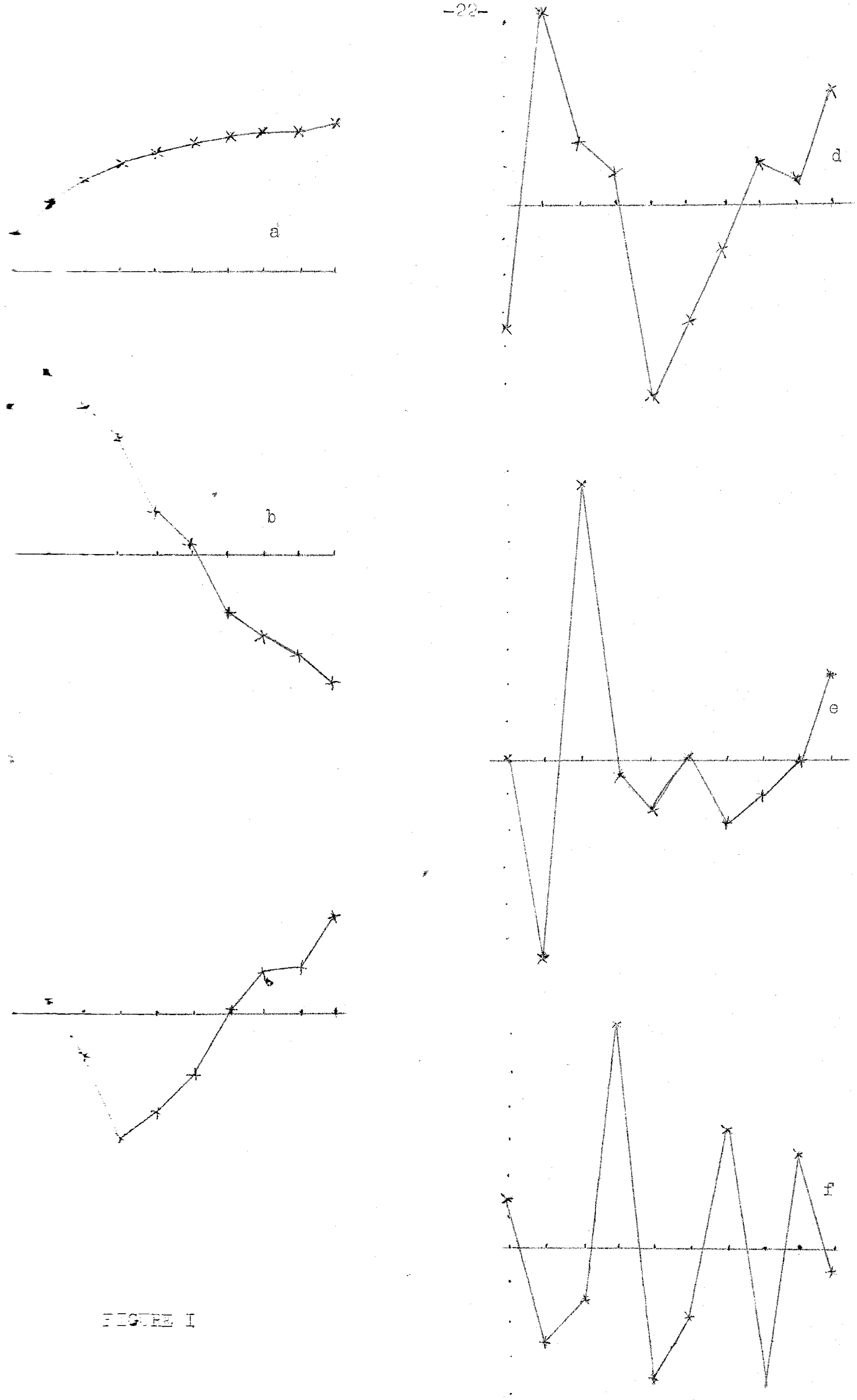

FIGURE I

	I	II	III	IV	V	VI
[illegible]	.099958	.099968	.099974	.099968	.099976	.099975
[illegible]	.000768	.006621	.040862	.087123	.156572	.172177
[illegible]	.000041	.001419	.009904	.059635	.182637	.201069
[illegible]	.000018	.001063	.004880	.043871	.179197	.213442
[illegible]	.000013	.001168	.003794	.041095	.143998	.217813
[illegible]	.000009	.001410	.003145	.042962	.097135	.206953
[illegible]	.000008	.001691	.002707	.045135	.056371	.171575
[illegible]	.000008	.001954	.002457	.046190	.027973	.116980
[illegible]	.000008	.002161	.002361	.046561	.010612	.061761
[illegible]	.000008	.002285	.002384	.047324	.001370	.023970
	.200887	.141012	.830605	1.743136	3.132356	3.444443

table II

	I	II	III	IV	V	VI
[illegible]	1.4991	0.4868	0.7451	0.9474	1.1366	1.2265
[illegible]	1.2363	0.2116	0.5439	0.8704	1.2680	1.3973
[illegible]	1.1409	0.1126	0.4446	0.8220	1.3442	1.4840
[illegible]	1.1093	0.0810	0.4035	0.8000	1.3799	1.5215
[illegible]	1.0962	0.0704	0.3867	0.7909	1.3959	1.5373

table III

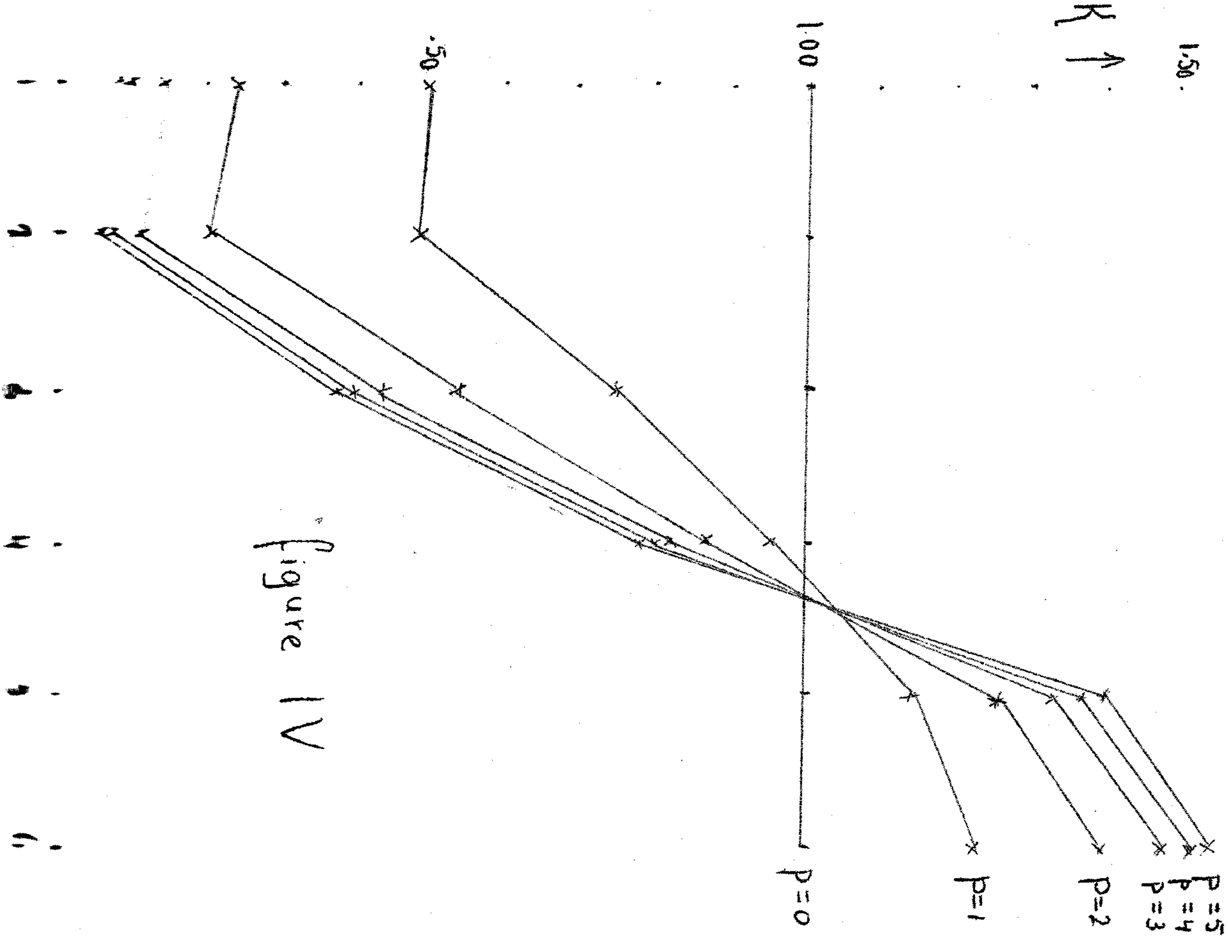

figure IV

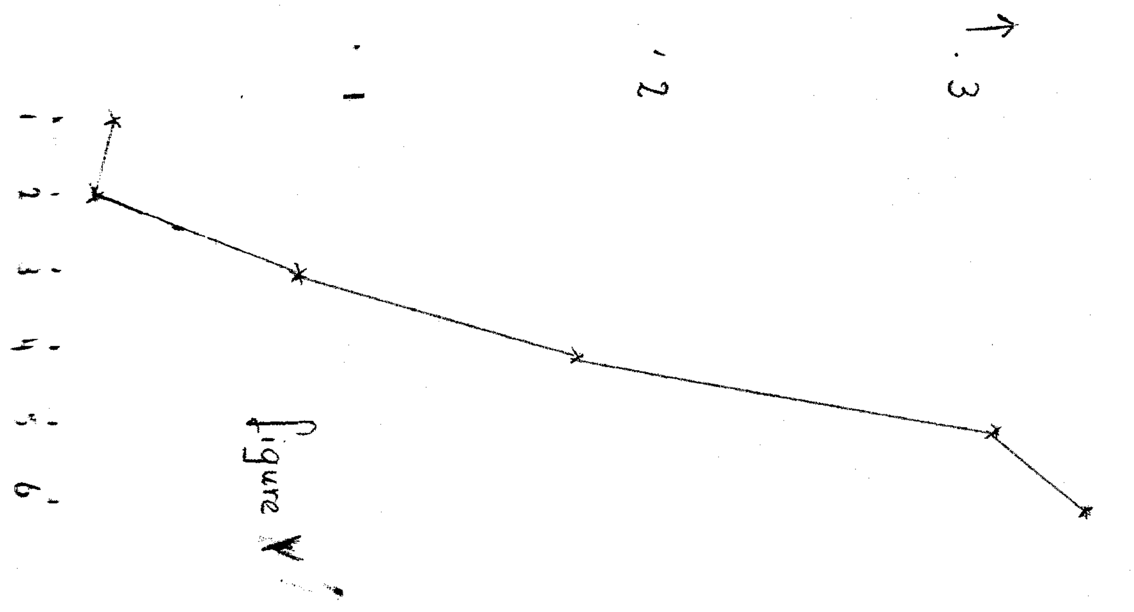

figure V

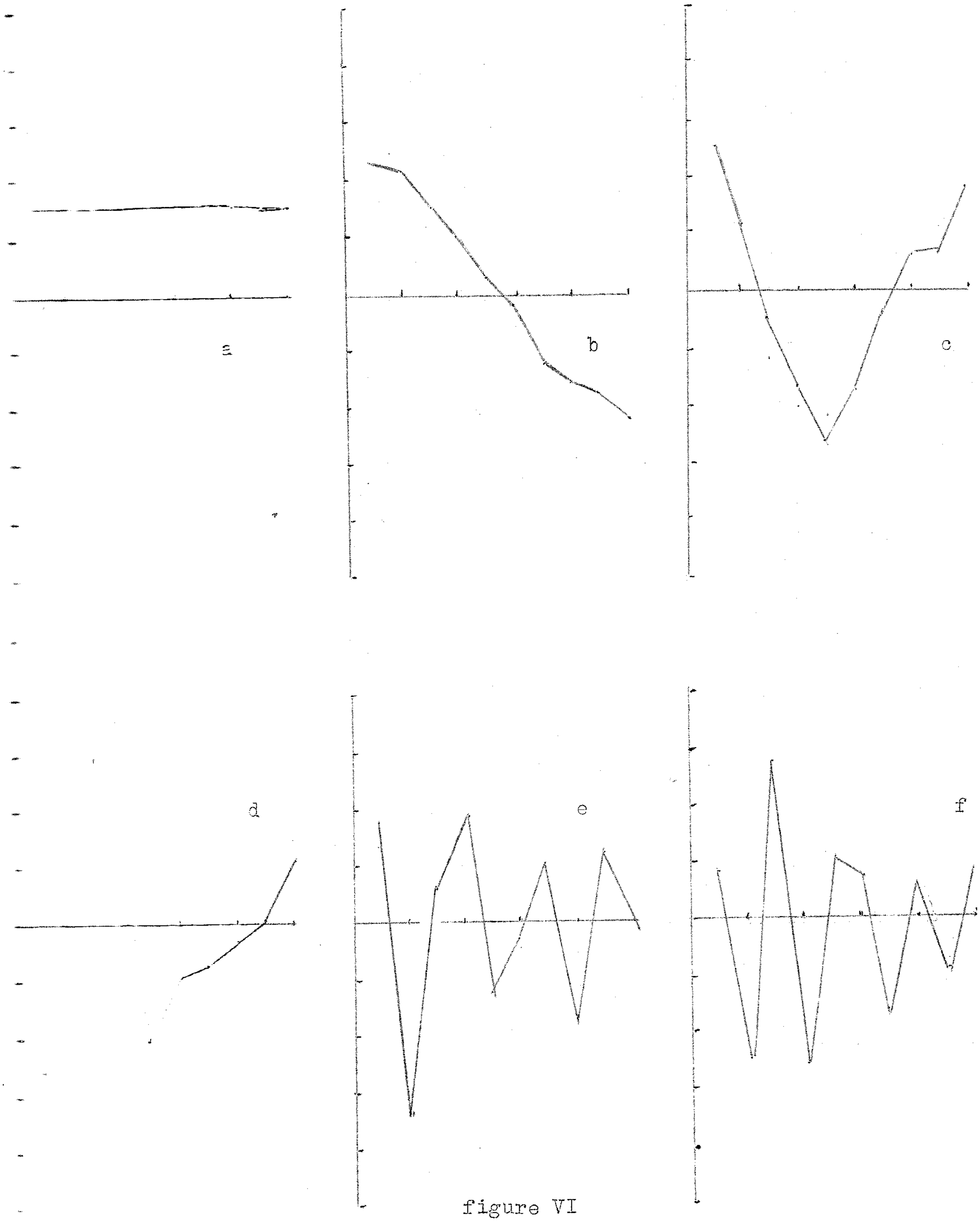

figure VI

[illegible]	I	II	III	IV	V	VI
[illegible]	.000003	.011969	.053363	.146885	.324609	.391418
[illegible]	.000008	.001725	.023032	.201799	.973554	1.528120

table VII

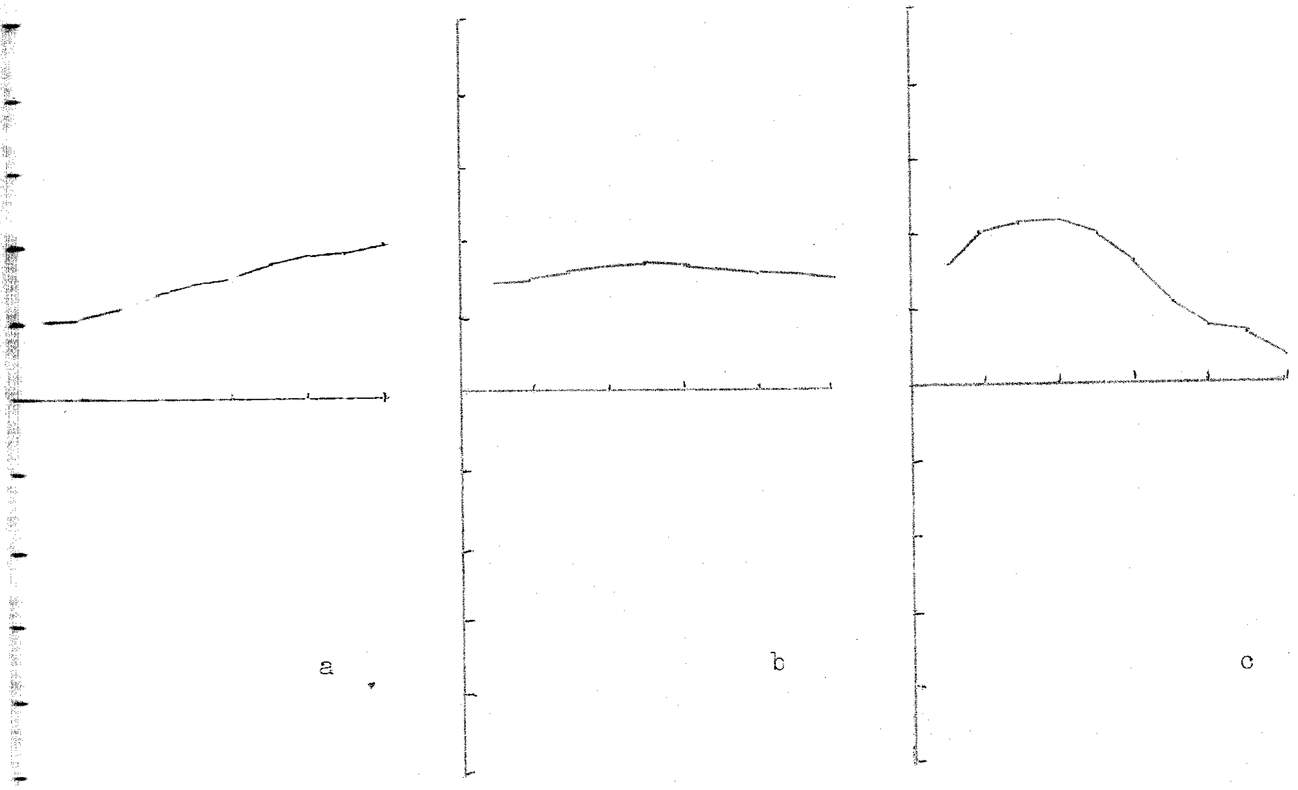

figure VIII

[illegible]	[illegible]	I	II	III	IV	V	VI
[illegible]	[illegible]	.119024	.520475	1.468319	2.475301	3.479711	3.928900
[illegible]	[illegible]	.012949	.215795	1.818509	6.803179	12.675313	15.356659
[illegible]	[illegible]	.077580	.515257	1.299623	2.789002	3.837963	3.926927

table IX

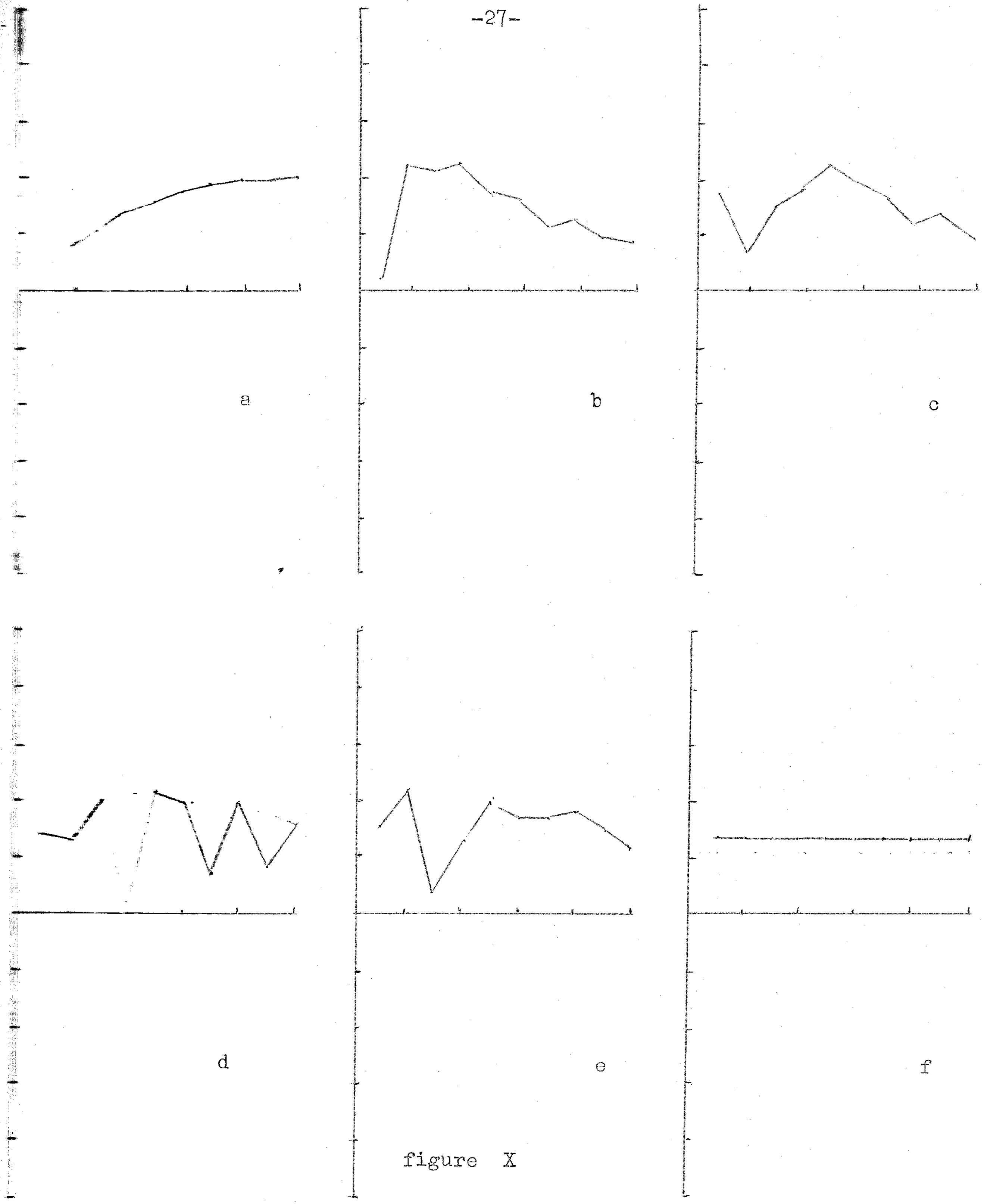

figure X

[illegible]	I	II	III	IV	V	VI
1	.001346	.147522	.409446	1.12295	2.18256	1261.77
2	.000278	.039314	.318547	2.19151	4.65534	721.503
3	.000088	.016289	.240889	3.04824	6.66214	783.228
4	.000034	.013684	.237871	3.54370	7.79844	873.143

table XI