DOCUMENT RESUME

ED 045 696 TM 000 172

AUTHOR Horst, Paul
TITLE Scale Free Reduced Rank Image Analysis.
SPONS AGENCY Office of Naval Research, Washington, D.C. Personnel and Training Research Programs Office.
PUB DATE Jun 70
NOTE 26p.

EDRS PRICE EDRS Price MF-$0.25 HC-$1.40
DESCRIPTORS *Correlation, *Factor Analysis, Mathematical Models, *Research Methodology, *Statistical Analysis
IDENTIFIERS Image Analysis

ABSTRACT

In the traditional Guttman-Harris type image analysis, a transformation is applied to the data matrix such that each column of the transformed data matrix is the best least squares estimate of the corresponding column of the data matrix from the remaining columns. The model is scale free. However, it assumes (1) that the correlation matrix is basic and (2) that the data matrix is free of measurement errors. In this paper a more generalized model is developed that does not require these two assumptions. Computational procedures are suggested. (Author)

SCALE FREE REDUCED RANK IMAGE ANALYSIS

Paul Horst

This research was sponsored by the Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, under Contract 477(33), Project NR 151-143.

Principal Investigator: Paul Horst
University of Washington

June 1970

SCALE FREE REDUCED RANK IMAGE ANALYSIS

Suppose we have given an $N \times n$ data matrix X of rank ρ. For convenience, we assume that the origins and scalings of the n variables are such that the $n \times n$ correlation matrix R is given by

$$R = X'X \tag{1}$$

To begin with, we make no assumptions about the rank of X or the relative values of N and n. In particular, we may have $N < n$. We let D be an $n \times n$, basic diagonal, scaling matrix, otherwise at present unspecified, and define

$$W = XD \tag{3}$$

We let

$$G = W'W \tag{4}$$

Hence from (1), (3), and (4)

$$G = DRD \tag{5}$$

The general problem of approximating the matrices X and W and their corresponding covariance matrices R and G have been extensively treated over the years in books and articles on factor analysis techniques and applications, and the literature is too voluminous and well known to require specific references. However, a distinction has been drawn and given major emphasis by some investigators between principal component analysis and factor analysis. Unfortunately, the distinction has for the most part been discussed with reference to the covariance matrices rather than the data matrices and, as a

consequence, considerable confusion and controversy has resulted with reference to the implications of the distinction for the data matrices. Horst (1969) has recently attempted to unify the various proposed models of principal component and factor analysis by showing that they may be regarded as special cases of a more general approach which utilizes variable parameters for scaling and loss functions. As a procedure for reconciling the various proposed matrix approximation models, our generalized factor analysis appears to have some merit. However, it does fail to include as a special case an important model introduced by Guttman (1953) and elaborated by Harris (1962) known as image analysis. The Guttman model has fundamental and important implications from the prediction point of view and, if one takes the position that prediction is the ultimate goal of science, these implications assume overriding importance. Perhaps the most important feature of the Guttman-Harris model has to do with the particular transformation that is applied to the data matrix. This transformation is such that each column of the transformed matrix is the best least squares estimate of the corresponding column of the data matrix as estimated from the remaining columns.

As Harris has shown, the model can be generalized so that it is scale free, and this scale free model has interesting invariant properties with reference to the matrix of transformed variables. The model, however, has two serious limitations. First, it assumes that the correlation or covariance matrix is basic. A necessary though not sufficient condition for this assumption to hold is that $N > n$, or that the number

of entities is greater than the number of variables. In many important cases of actual data, this assumption is not satisfied.

In the second place, the model assumes there are no errors of measurement in a data matrix that samples some domain of entities and attributes.

We shall develop a more generalized model of the image analysis type that is free of these two assumptions. We let v and u be $n \times m$ basic matrices where $m \leq \rho$ and define an $n \times n$ matrix B by

$$B = vu' \tag{6}$$

We let

$$\Delta = (D_B + I) \tag{7}$$

where

$$D_B = \text{diag}\ (B) \tag{8}$$

Also we define

$$Z = WB \tag{9}$$

and

$$Y = W(B - D_B) \tag{10}$$

From equations (6) and (9) it is clear that the rank of Z cannot be greater than m. From equation (10) it is clear that each column of Y is independent of the corresponding column of W and hence also of X. This property of Y corresponds to that of the transformed data matrix in the traditional image analysis model. In that model, however, the B matrix is taken as basic.

We next define an $N \times n$ residual matrix e by

$$e = W - Y \tag{11}$$

From equations (7), (10), and (11)

$$e = W(\Delta - B) \tag{12}$$

We let

$$\phi = \mathrm{tr}\ e'e \tag{13}$$

and

$$\psi = \phi + \mathrm{tr}\Delta d \tag{14}$$

where d is a diagonal matrix of Lagrangian multipliers. We wish to minimize ϕ with the constraint (7). Therefore we take the symbolic derivative of ψ with respect to the matrix v and equate to zero, viz.

$$\text{(14a)} \quad \frac{\partial \psi}{\partial v} = 0$$

Now from (6) and (7)

$$\text{(15)} \quad \mathrm{tr}\Delta d = \mathrm{tr}\ (vu'd) + \mathrm{tr}\ d$$

From (4), (6), (12), (13), (14), and (15)

$$\text{(16)} \quad \psi = \mathrm{tr}(vu'Guv' - vu'G\Delta - \Delta Guv' + \Delta^2 + vu'd + d)$$

From (14a) and (16)

$$\text{(17)} \quad u'Guv' = u'(G\Delta - d)$$

From (17)

$$\text{(18)} \quad v' = (u'Gu)^{-1}u'(G\Delta - d)$$

From (6) and (18)

$$\text{(19)} \quad B = u(u'Gu)^{-1}u'(G\Delta - d)$$

Let

$$(20) \quad S = u(u'Gu)^{-1}u'$$

From (19) and (20)

$$(21) \quad B = SG\Delta - Sd$$

Suppose now we assume that the scaling matrix D in (3) and (5) can be chosen so that B is symmetrical, that is,

$$(22) \quad B = B'$$

A sufficient condition that (22) be satisfied is obviously that SG is symmetrical and that Δ and d be scalar matrices, since by (20) S is symmetric. Suppose we indicate the basic structure of G by

$$(23) \quad G = Q_m \delta_m Q_m' + Q_s \delta_s Q_s'$$

where

$$(24) \quad m + s = \rho$$

A sufficient condition that SG be symmetrical is that

$$(25) \quad u = Q_m c$$

where c is any $m \times m$ basic matrix. That (25) is also necessary is the case if we let δ_m contain any m nonvanishing roots of G, and Q_m the corresponding vectors. However, we shall take these to be the m largest and later justify this choice. From (20), (23), and (25)

(26) $S = Q_m \delta_m^{-1} Q_m'$

(27) $SG = Q_m Q_m'$

From (21), (26), and (27)

(28) $B = Q_m Q_m' \Delta - Q_m \delta_m^{-1} Q_m' d$

Now so far we have put no restrictions on D in (3) and (5). Let us assume that D can be determined so that

(29) $\Delta = If$

where f is a scalar quantity. Because of (22), (28), and (29) we must also have

(30) $d = Ia$

where a is a scalar. From (28), (29), and (30)

(31) $B = Q_m Q_m' f - Q_m \delta_m^{-1} Q_m' a$

From (7) and (29)

(32) $D_B = (f - 1)I$

We shall now determine a as a function of f. From (31) and (32) we have

(33) $(f - 1)n = mf - a \operatorname{tr} \delta_m^{-1}$

From (33)

$$(34) \quad a = \frac{n - (n - m)f}{\operatorname{tr} \delta_m^{-1}}$$

To determine f we first write from (29), (31), and (34)

$$(35) \quad (\Delta - B) = (I - Q_m Q_m')f + \frac{n - (n - m)f}{\operatorname{tr} \delta_m^{-1}} Q_m \delta_m^{-1} Q_m'$$

From (4), (12), and (13)

$$(36) \quad \psi = \operatorname{tr}(\Delta - B')G(\Delta - B)$$

From (22) and (36)

$$(37) \quad \phi = \operatorname{tr}(G(\Delta - B)^2)$$

From (35)

$$(38) \quad (\Delta - B)^2 = (I - Q_m Q_m')f + \left(\frac{n - (n - m)f}{\operatorname{tr} \delta_m^{-1}}\right)^2 Q_m \delta_m^{-2} Q_m'$$

From (23) and (38)

$$(39) \quad G(\Delta - B)^2 = (G - Q_m \delta_m Q_m')f + \left(\frac{n - (n - m)f}{\operatorname{tr} \delta_m^{-1}}\right)^2 Q_m \delta_m^{-1} Q_m'$$

From (5), (37), and (39)

$$(40) \quad \phi = (\operatorname{tr} D^2 - \operatorname{tr} \delta_m)f^2 + \frac{(n - (n - m)f)^2}{\operatorname{tr} \delta_m^{-1}}$$

For convenience we constrain D so that

$$(41) \quad \operatorname{tr} D^2 = n$$

From (5), (23), and (41)

$$(41a) \quad \operatorname{tr} \delta_m + \operatorname{tr} \delta_s = n$$

Since we wish to minimize ϕ, we set

$$(42) \quad \frac{d\phi}{df} = 0$$

From (40) and (41)

$$(43) \quad 0 = (n - \operatorname{tr} \mathcal{S}_m)f - \frac{(n - m)(n - (n - m)f)}{\operatorname{tr} \mathcal{S}_m^{-1}}$$

From (43)

$$(44) \quad f = \frac{n(n - m)}{(n - \operatorname{tr} \mathcal{S}_m) \operatorname{tr} \mathcal{S}_m^{-1} + (n - m)^2}$$

Now let

$$(44a) \quad \alpha = \frac{n - \operatorname{tr} \mathcal{S}_m}{n - m}$$

It is clear then that α is the mean of the $n - m$ smallest roots of G. From (44) and (45)

$$(44b) \quad f = \frac{n}{\alpha \operatorname{tr} \mathcal{S}_m^{-1} + (n - m)}$$

It is probably intuitively obvious that the solution (44) for f yields a minimum. For this to be the case we should have

$$(45) \quad \frac{d^2\phi}{df^2} > 0$$

From (43)

$$(46) \quad \frac{d^2\phi}{df^2} = (n - \operatorname{tr} \mathcal{S}_m) + (n - m)^2 \,/\, \operatorname{tr} \mathcal{S}_m^{-1}$$

Except for the limiting case of $n = m$, which we shall consider later, (46) does satisfy (45) since, because of (5), (23), and (41), we must have $n > \operatorname{tr} \mathcal{S}_m$.

If now we substitute (44) in (34), we get

$$(47) \quad a = \frac{n(n - \operatorname{tr} \mathcal{S}_m)}{(n - \operatorname{tr} \mathcal{S}_m) \operatorname{tr} \mathcal{S}_m^{-1} + (n - m)^2}$$

From (44a), (44b), and (47)

$$(48) \quad a = \alpha f$$

Substituting (48) in (31) we get

$$(48a) \quad B = Q_m(I - \alpha_s \mathcal{S}_m^{-1})Q_m' f$$

If we let

$$(48b) \quad d_m = (I - \alpha \mathcal{S}_m^{-1})$$

we have from (48a) and (48b)

$$(49) \quad B = Q_m d_m Q_m' f$$

It will now be of interest to find the covariance matrices involving the matrices Z, Y, and e, given by equations (9), (10), and (11), respectively.

From (4) and (9)

$$(50) \quad Z'Z = B'GB$$

From (4), (9), and (10)

$$(51) \quad Z'Y = B'G(B - D_B)$$

From (4), (9), and (11)

$$(52) \quad Z'e = B'G(\Delta - B)$$

From (4) and (10)

$$(53) \quad Y'Y = (B' - D_B)G(B' - D_B)$$

From (4), (10), and (11)

$$(54) \quad Y'e = (B' - D_B)G(\Delta - B)$$

From (4) and (11)

$$(55) \quad e'e = (\Delta - B')G(\Delta - B)$$

Since B is symmetrical, it can be proved that B, $(B - D_B)$, and $(\Delta - B)$, together with G, are all commutative for multiplication. Therefore we have from (50) through (55) respectively

$$(58) \quad Z'Z = GB^2$$

$$(59) \quad Z'Y = GB(B - D_B)$$

$$(60) \quad Z'e = GB(\Delta - B)$$

$$(61) \quad Y'Y = G(B - D_B)^2$$

$$(62) \quad Y'e = G(B - D_B)(\Delta - B)$$

$$(63) \quad e'e = G(\Delta - B)^2$$

Now from (32) and (49)

$$(64) \quad B - D_B = (Q_m \, d_m \, Q_m' - \frac{f - 1}{f} I)f$$

From (29) and (49)

$$(65) \quad \Delta - B = (I - Q_m \, d_m \, Q_m')f$$

Now from (49)

$$(66)\quad B^2 = Q_m\, d_m^{\,2}\, Q_m'\, f^2$$

From (49) and (64)

$$(67)\quad B(B - D_B) = Q_m\, d_m(d_m - \frac{f-1}{f} I)Q_m'\, f^2$$

From (49) and (65)

$$(68)\quad B(\Delta - B) = Q_m\, d_m(I - d_m)Q_m'\, f^2$$

From (64)

$$(69)\quad (B - D_B)^2 = (Q_m\, d_m(d_m - 2\,\frac{f-1}{f} I)Q_m' + (\frac{f-1}{f})^2 I)f^2$$

From (64) and (65)

$$(70)\quad (B - D_B)(\Delta - B) = -(Q_m\, d_m(d_m - \frac{2f-1}{f} I)Q_m' + \frac{f-1}{f} I)f^2$$

From (65)

$$(71)\quad (\Delta - B)^2 = (I + Q_m\, d_m(d_m - 2I)Q_m')f^2$$

We may now write the covariance matrices given by (58) though (63) as functions of G and of the corresponding right sides of equations (66) through (71). We have from these two sets of equations and equation (23)

$$(72)\quad Z'Z = Q_m\, \delta_m\, d_m^{\,2}\, Q_m'\, f^2$$

$$(73)\quad Z'Y = Q_m\, \delta_m\, d_m(d_m - \frac{f-1}{f} I)Q_m'\, f^2$$

$$(74)\quad Z'e = Q_m\, \delta_m\, d_m(I - d_m)Q_m'\, f^2$$

(75) $Y'Y = (Q_m \delta_m d_m(d_m - 2(\frac{f-1}{f}) I)Q_m' + (\frac{f-1}{f})^2 G)f^2$

(76) $Y'e = -(Q_m \delta_m d_m(d_m - \frac{2f-1}{f} I)Q_m' + \frac{f-1}{f} G)f^2$

(77) $e'e = (G + Q_m \delta_m d_m(d_m - 2I)Q_m')f^2$

To further simplify the notation, we let

(78) $g = Z'Z$

(79) $E = Q_m \delta_m d_m Q_m'$

Then substituting (78) and (79) in (72) through (77)

(80) $Z'Y = g - f(f - 1)E$

(82) $Z'e = -g + f^2E$

(83) $Y'Y = g - 2f(f - 1)E + (f - 1)^2G$

(84) $Y'e = -g + f(2f - 1)E - f(f-1)G$

(85) $e'e = g - 2f^2E + f^2G$

We may now regard g as the rank m approximation to G, the covariance matrix of W, which is the rescaled data matrix X. Then if we let

(86) $A = Q_m \delta_m^{\frac{1}{2}}(I - \alpha \delta_m^{-1})f$,

we have from (72), (78), and (86)

(87) $g = AA'$

and A may be regarded as the factor loading matrix.

The matrix A may be compared with the conventional basic structure form

$$(88) \quad A = Q_m \delta_m^{\frac{1}{2}}$$

where A is the basic structure solution in our generalized factor analysis with variable scaling and loss function parameters, special cases of which have been shown to closely resemble, if not be identical to, a number of current factor analytic models.

We may now evaluate the Z, Y, and e matrices directly. We let the basic structure of W be given by

$$(89) \quad W = \begin{bmatrix} P_m , P_s \end{bmatrix} \begin{bmatrix} \delta_m^{\frac{1}{2}} & 0 \\ 0 & \delta_s^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} Q_m' \\ Q_s' \end{bmatrix}$$

From (9), (49), and (89)

$$(90) \quad Z = P_m \delta_m^{\frac{1}{2}} d_m Q'f$$

From (10), (48b), (64), and (89)

$$(91) \quad Y = P_m(\delta_m^{\frac{1}{2}} - f \alpha \delta_m^{-\frac{1}{2}})Q_m' - P_s \delta_s^{\frac{1}{2}} Q_s'(f - 1)$$

From (12), (48b), (65), and (89)

$$(92) \quad e = (P_m \alpha \delta_m^{-\frac{1}{2}} Q_m' + P_s \delta_s^{\frac{1}{2}} Q_s')f$$

It is clear from (90) that Z is of rank m, while from (91) and (92) we see that Y and e are of rank n if X is basic. It is also clear from (89) that

$$(93) \quad P_m = WQ_m \delta_m^{-\frac{1}{2}}$$

We may regard P_m as yet another type of score matrix.

To summarize, we have then five types of score matrices as follows:

(1) The W matrix is the X matrix rescaled so that the B matrix calculated from the roots and vectors, as indicated by (49), has the diagonal

(94) $D_B = (f - 1) I$

where f is given by (44b).

(2) The Z matrix is the rank m approximation to the W matrix by the B transformation on W.

(3) The Y matrix is the best least square approximation to the W matrix calculated from the Z matrix where each Y vector is independent of the corresponding W vector. It is important to note that B has been determined so as to optimize this approximation in the least square sense.

(4) The e matrix is the matrix of residuals between W and Y, and B is determined so as to minimize the trace of its product moment.

(5) The P_m matrix is analogous to the principal axis factor score matrix. It is in fact the principal axis factor acore matrix for the rescaled data matrix W. The width of this matrix is obviously only m whereas the other four matrices are all of width n. This matrix is perhaps of most practical interest among the five types.

To return to the covariance matrices (80) through (85), to date their properties have been only briefly investigated. Perhaps the most that can be said as of now concerns the traces of $Y'Y$, $Y'e$, and $e'e$. By straightforward but somewhat tedious manipulation, it can be proved that

(95) $\operatorname{tr} Y'Y = n - \alpha fn$

(96) $\operatorname{tr} Y'e = 0$

(97) $\operatorname{tr} e'e = \alpha fn$

It is clear therefore from (5), (41), (95), (96), and (97) that

(98) $\operatorname{tr} Y'Y + \operatorname{tr} e'e = \operatorname{tr} G$

which is as it should be.

We may now consider the case where $n = m$ and R is basic. In this case we must return to equations (31) and (32). Equation (31) becomes simply

(99) $B = fI - Q_m \delta_m^{-1} Q_m' a$

From (32) and (99)

(100) $B - D_B = I - Q\delta^{-1} Q'a$

From (5) and (23)

(101) $Q\delta^{-1}Q' = D^{-1} R^{-1} D^{-1}$

From (100) and (101)

(102) $0 = I - D^{-1} D_{R^{-1}} D^{-1}a$

From (102)

(103) $D^2 = D_{R^{-1}} a$

From (41) and (103)

(104) $\sqrt{a} = \sqrt{n/\operatorname{tr} R^{-1}}$

From (103) and (104)

$$(105) \quad D = D_{R^{-1}}^{\frac{1}{2}} \sqrt{n/\mathrm{tr}\, R^{-1}}$$

Equation (105) is the same result obtained by Harris's scale free modification of Guttman's image analysis model except for the scaling factor $\sqrt{a}$.

If now R is of rank m where $m < n$, we have from (44)

$$(106) \quad f = \frac{n}{n - m}$$

and from (34) and (106)

$$(107) \quad a = 0$$

From (31), (106), and (107)

$$(108) \quad B = Q_m Q_m' \frac{n}{n - m}$$

From (32), (106), and (108)

$$(110) \quad \frac{m}{n - m} I = D_{Q_m Q_m'} \frac{n}{n - m}$$

From (110)

$$(111) \quad D_{Q_m Q_m'} = \frac{m}{n} I$$

It is seen therefore from (111) that if the R matrix is not basic and m is taken as the rank of the matrix, then the scaling of R must be such that the row vectors of the vertical basic orthonormal Q_m are all of equal length. That this is always possible has not been proved, although it has been proved for special cases. In any event, it appears to date that this case is of more theoretical than practical interest.

Of more general and practical interest is the case where m is less than the rank of R, whether or not R is basic. Thus far we have ignored the question of how to determine m. This of course is the question of how many factors to solve for and although many answers have been proposed, none of them has gained universal acceptance. The various tests of statistical significance leave much to be desired. In our opinion, these tests are founded on assumptions that are irrelevant or inappropriate for most important practical situations.

One of the simplest and most appealing criteria is the one proposed by Kaiser (1958) and accepted by many as a good rule-of-thumb procedure. It is the number of roots greater than unity in the correlation matrix. This rule may be generalized to the number of roots greater than unity in the matrix G as defined in the foregoing developments. The smallest root in δ_m of (23) would, according to this criterion, be greater than unity and the largest root in δ_s of (23) should not exceed unity. As a first approximation one might start with the Kaiser criterion, namely, the number of roots greater than unity in the correlation matrix.

It may be of interest to examine the generalization of the Kaiser criterion to the case of the G matrix. First let us return to equation (97). From this we get

$$(112) \quad \frac{\operatorname{tr} e'e}{n} = \alpha f$$

From (98) and (112) we get

$$(113) \quad \frac{\operatorname{tr} Y'Y}{n} = 1 - \alpha f$$

From (44a) and (44b)

$$(114)\quad \alpha f = \frac{n(n - \operatorname{tr} \delta_m)}{(n - \operatorname{tr} \delta_m)\operatorname{tr} \delta_m^{-1} + (n - m)^2}$$

To gain better insight into the properties of (114), we let

$$(115)\quad \mu = \frac{\operatorname{tr} \delta_m}{m}$$

That is, μ is simply the mean of the m values in δ_m. We let

$$(116)\quad \nu^2 = \frac{\operatorname{tr} \delta_m^2}{m} - \mu^2$$

so that ν is simply the variance of the values in δ_m. It can be proved that

$$(117)\quad \operatorname{tr} \delta_m^{-1} = \frac{m}{\mu}\left(1 + \frac{\nu^2}{\mu^2} + \epsilon\right)$$

where ϵ is a positive quantity which tends to increase as ν^2 increases. We let

$$(118)\quad \gamma = \frac{\nu^2}{\mu^2} + \epsilon$$

so that γ also increases as ν^2 increases. From (114), (115), (117), and (118)

$$(119)\quad \alpha f = \frac{n(n - m\mu)\mu}{m(n - m\mu) + (n - m)^2\mu + m(n - m\mu)\gamma}$$

We let

$$(120)\quad \Gamma = \frac{(n - m\mu)\gamma}{n}$$

From (119) and (120)

$$(121)\quad \alpha f = \frac{\left(\frac{n}{m} - \mu\right)\mu}{1 + \left(\frac{n}{m} - 2\right)\mu + \Gamma}$$

From (113), (120), and (121)

$$(122)\quad \frac{\operatorname{tr} Y'Y}{n} = \frac{(\mu - 1)^2 + \Gamma}{1 + (\frac{n}{m} - 2) + \Gamma}$$

From (112), (115), and (122)

$$(123)\quad \frac{\operatorname{tr} e'e}{\operatorname{tr} Y'Y} = \frac{(n - \operatorname{tr} \delta_m) \operatorname{tr} \delta_m}{(\operatorname{tr} \delta_m - m)^2 + (n - \operatorname{tr} \delta_m)\eta}$$

where

$$(124)\quad \eta = \frac{m^2}{n} \gamma$$

Since for a given m the ratio in (123) should be as small as possible, the numerator term should be small and the denominator term large. As all the roots in δ_m approach equality, η in (123) approaches zero. For a given m the ratio in (123) would then approach a maximum if these roots were the m largest roots of G. That this would also be the case if the m largest roots were not all equal is also probably true but we have not found a mathematical proof.

Perhaps an argument for choosing m as the number of roots of G greater than unity is suggested by (92). Neglecting the scaling factor f on the right, the right hand term in the parentheses is precisely the basic structure form of the residual matrix in the principal axis type solution. The term on the left involves the remaining parts of the basic orthonormals of P and Q of W in (89). The basic diagonal of this matrix is $\alpha \delta_m^{-\frac{1}{2}}$. We wish to suppress the component of the e matrix, given by the first term on the right of (92), as much as possible. In particular, we wish to guarantee that

$$(124a)\quad \alpha < 1$$

and

(124b) $\mathcal{S}_m^{-1} < I$

where (124b) implies that the inequality holds for each of the diagonal elements. Obviously, because of (41a) and (44a), a necessary and sufficient condition for both (124a) and (124b) to hold is that $\mathcal{S}_m$ consist of the roots of G greater than unity.

To see how we might calculate the D scaling matrix, we proceed as follows. We let

(125) $V = DU$

(126) $\sigma = V(V'RV)^{-1}V'$

From (1) through (5), and (20), (125), and (126)

(127) $S = D^{-1}\sigma D^{-1}$

(128) $GS = D^{-1}\sigma RD$

(129) $\operatorname{tr} \mathcal{S} = \operatorname{tr}(V'RV)^{\frac{1}{2}}$

(130) $\operatorname{tr} \mathcal{S}^{-1} = \operatorname{tr}(V'RV)^{-\frac{1}{2}}$

From (31), (127), and (128)

(131) $B = D^{-1}\sigma RDf - D^{-1}\sigma D^{-1}a$

From (7), (32), and (131)

(132) $D^2(f - 1) = D_{\sigma R} D^2 f - D_\sigma a$

From (48) and (132)

(132a) $D^2(f - 1) = (D_{\sigma R} D^2 - \alpha D_\sigma)f$

Let

(133) $F = \frac{f - 1}{f}$

From (132) and (133)

$$(134) \quad D^2 = (D_{\sigma R} - FI)^{-1} \alpha D_\sigma$$

From (41) and (133)

$$(135) \quad n - tr(D_{\sigma R} - F)^{-1} \alpha D_\sigma = 0$$

From (44b) and (133)

$$(136) \quad F = \frac{m - \alpha \, tr \, \delta_m^{-1}}{n}$$

Let

$$(137) \quad D_m = D_{Q_m Q_m'}$$

From (27), (128), (134), and (137)

$$(138) \quad D^2 = (D_m - FI)^{-1} \alpha D_\sigma$$

and from (135) and (138)

$$(139) \quad n - tr(D_m - FI)^{-1} \alpha D_\sigma = 0$$

Without loss of generality, assume that the D_m values are in descending order of magnitude. It can be proved that one and only one F exists lying between each of the $(n - 1)$ adjacent pairs of D_m which satisfies (139). But (139) has n roots. Because of the left side of (138) and since D_σ is positive definite, the matrix in parentheses must also be positive definite. This obviously cannot be the case for any F greater than the smallest value in D_m. The remaining root must lie outside the range of the values in D_m. It cannot be greater than the largest value in D_m since any such value

could not satisfy (139). It must therefore be smaller than the smallest value in D_m and can therefore satisfy both (138) and (139). Also because of (136), F must be greater than zero.

Next we may write from (23) and (25)

$$(140) \quad u = Gu(u'Gu)^{-\frac{1}{2}}h$$

where h is any square orthonormal matrix. From (5) and (125)

$$(141) \quad V = D^2RV(V'RV)^{-\frac{1}{2}}h$$

Equation (141), together with (126), (129), (130), (134), and (135), provides the basis for the suggested iterative computing algorithms to solve for the D scaling matrix, which in turn provides the basis for all other computations. We let

$$(142) \quad {}_iw = R\,{}_iV$$

$$(143) \quad {}_it\,{}_it' = {}_iV'\,{}_iw$$

$$(144) \quad {}_i\sigma = {}_iV\,{}_it'^{-1}\,{}_it^{-1}\,{}_iV'$$

$$(145) \quad {}_i\alpha = n - tr({}_it)$$

$$(146) \quad n - tr((D\,{}_i\sigma R - {}_iFI)^{-1}\,{}_i\alpha\, D\,{}_i\sigma) = 0$$

$$(147) \quad {}_iD^2 = (D\,{}_i\sigma R - {}_iFI)^{-1}\,{}_i\alpha\, D\,{}_i\sigma$$

$$(148) \quad {}_{i+1}V = {}_iD^2\,{}_iw\,{}_it'^{-1}$$

We begin with the basic structure solution of R given by

$$(149) \quad R = q_m \, d_m \, q_m' + q_s \, d_s \, q_s'$$

where m is the number of roots in R greater than unity. We let

$$(150) \quad {}_0V = q_m$$

Then

$$(151) \quad {}_0\sigma = q_m \, d_m^{-1} \, q_m'$$

$$(152) \quad {}_0\sigma R = q_m \, q_m'$$

$$(153) \quad {}_0\alpha = \frac{n - \operatorname{tr} d_m}{n - m}$$

$$(154) \quad n - \operatorname{tr}((D_{q_m q_m'} - {}_0FI)^{-1} \, {}_0\alpha \, D_{q_m d_m^{-1} q_m'}) = 0$$

$$(155) \quad {}_0D^2 = (D_{q_m q_m'} - {}_0FI)^{-1} \, {}_0\alpha \, D_{q_m d_m^{-1} q_m'}$$

$$(156) \quad {}_1V = {}_0D^2 \, Q_m \, d_m^{\frac{1}{2}}$$

Beginning then with $i = 1$ and using (156), one would iterate with the computations (142) through (148) until hopefully D stabilized.

It is possible that better procedures for solving for D could be formulated. For example, in equation (128) GS is symmetric. We might solve iteratively for the solution for D which in the least square sense makes each approximation to GS most nearly symmetric. For example, let

$$(157) \quad M = {}_i\sigma \, R$$

Let

(158) $M' - M = \epsilon$

(159) $D1 = V_D$

(160) $\left[(D_{M'M} - M' \cdot M) - \phi I\right] V_D = 0$

where the dot means elemental multiplication. It can be shown that the V_D corresponding to the smallest root in (159) is proportional to the D which minimizes the trace of $\epsilon'\epsilon$ in (158). What methods will be most efficient in solving for D must await actual computational research.

ADDENDUM

Recently Jöreskog* has presented a model for Image Factor Analysis, together with computational procedures for estimating the parameters of the model. The relationships among the objectives and end results of his approach and ours is not yet clear.

* Jöreskog, K. G. Efficient estimation in image factor analysis. Psychometrika, 34, 51-75, 1969.

REFERENCES

Guttman, Louis. "Image Theory for the Structure of Quantitative Variates," Psychometrika, 18, 277-296, 1953.

Harris, Chester W. "Some Rao-Guttman Relationships", Psychometrika, 27, 247-264, 1962.

Horst, Paul. "Generalized Factor Analysis," Technical Report, Office of Naval Research, April 1969.

Kaiser, Henry F. "The Varimax Criterion for Analytic Rotation in Factor Analysis," Psychometrika, 23, 187-200, 1958.