

GENERALIZED PROCRUSTES ANALYSIS

BY

J. C. GOWER

ROTHAMSTED EXPERIMENTAL STATION,
HARPENDEN, HERTS.

Suppose $P_j^{(i)}$ ($i = 1, 2, \dots, m, j = 1, 2, \dots, n$) give the locations of mn points in p -dimensional space. Collectively these may be regarded as m configurations, or scalings, each of n points in p -dimensions. The problem is investigated of translating, rotating, reflecting and scaling the m configurations to minimize the goodness-of-fit criterion $\sum_{i=1}^m \sum_{j=1}^n \Delta^2(P_j^{(i)}G_j)$, where G_j is the centroid of the m points $P_j^{(i)}$ ($i = 1, 2, \dots, m$). The rotated positions of each configuration may be regarded as individual analyses with the centroid configuration representing a consensus, and this relationship with individual scaling analysis is discussed. A computational technique is given, the results of which can be summarized in analysis of variance form. The special case $m = 2$ corresponds to Classical Procrustes analysis but the choice of criterion that fits each configuration to the common centroid configuration avoids difficulties that arise when one set is fitted to the other, regarded as fixed.

Suppose X_i ($i = 1, 2, \dots, m$) is a matrix with n rows and p_i columns whose j th row gives the coordinates of a point $P_j^{(i)}$ referred to p_i orthogonal axes. Typical practical situations are when each X_i is an observed data-matrix or has been obtained as multidimensional scales of n stimuli or as factor loadings or scores etc. Figure 1 illustrates the geometry for $m = 3$, $n = 4$, $p_1 = p_2 = p_3 = 2$ and where the three configurations agree quite well.

In the following, it is assumed that the m points $P_j^{(i)}$ ($i = 1, 2, \dots, m$) all refer to the same j th entity. For example, each X_i may have been derived from different sets of variables observed for the same samples in every case (j refers to samples) or each X_i may have been obtained as different scalings of the same observational data (j refers to observations) or each X_i may have arisen from one type of scaling of the same stimuli as perceived by different individuals (j refers to stimuli). It is also assumed that $p_i = p$, a constant. This simplifies the exposition without loss of generality, because it is sufficient to choose $p = \text{Max}(p_i)$ and append zero columns to every X_i that initially has fewer than p columns.

Problems are common where it is desired to study the relationships amongst the m sets and frequently some kind of combined analysis is desirable. Examples are Individual Scaling Analysis [Carroll & Chang, 1970] or comparing factor loadings derived from different studies. Schönemann & Carroll

[1970] and Gower [1971] have discussed how two such matrices X_1 and X_2 can be fitted, allowing the motions of translation, rotation, reflection and estimation of a homogeneous scaling factor, defining best fit as that which minimizes the least squares criterion $m_{12}^2 = \sum_{i=1}^n \Delta^2(P_i^{(1)}, P_i^{(2)})$, where $\Delta(A, B)$ is the Euclidean distance between the pair of points A and B . This problem has an analytical solution. For best fit, the centroids of X_1 and X_2 should be superimposed and the rotation matrix H such that X_2H best fits X_1 is given as $H = V'U$ where $X_1'X_2 = U'\Gamma V$, is the Eckart-Young [1935] (or singular value) decomposition with U and V orthogonal and Γ diagonal. With reflection Γ has no negative element and without reflection Γ has no more than one negative element. The least-squares estimate of the scaling factor obtained by fitting X_2 to X_1 is $\rho = \text{tr}(X_2HX_1')/\text{tr}(X_2X_2')$ which is clearly not the inverse of that obtained by fitting X_1 to X_2 . To overcome this difficulty Schönemann & Carroll [1970] propose a symmetric scale measure; an alternative solution occurs as a special case of the work described below.

Without translation and scaling this problem is known as Procrustes rotation. In this paper the term Procrustes is extended to include all the classical rigid-body motions and also the possibility of uniform scaling (either stretching or shrinking). This extended terminology seems more in the spirit of the Greek tale of the innkeeper Procrustes who stretched or lopped off traveller's limbs so that they would fit his bed.

Procrustes rotation provides one way of analyzing individually scaled data. Given a pair of individual scales X_u and X_v , the first may be rotated to best fit the second, giving a least-squares value m_{uv}^2 (say). With scaling factors, in general, $m_{uv} \neq m_{vu}$ but without scaling factors $m_{uv} = m_{vu}$ and it is shown below that elements of the $(m \times m)$ symmetric matrix of all such comparisons, form a metric. This follows from considering three configurations X_1 , X_2 and X_3 and regarding all np values of each matrix (ordered by columns, say) as being the coordinates of a single point, so that X_1 , X_2 and X_3 are represented by points Q_1 , Q_2 and Q_3 respectively, in a Euclidean np -space. The distance between any two points Q_u , Q_v of this representation is the square root of the sum of squares (not necessarily minimal) measuring the goodness of fit between X_u and X_v . If X_1 and X_2 have been rotated to best fit with X_3 , m_{13} and m_{23} are merely the distances $\Delta(Q_1, Q_3)$ and $\Delta(Q_2, Q_3)$. Because this representation is Euclidean, and hence metric,

$$\Delta(Q_1, Q_2) \leq m_{13} + m_{23}.$$

But $\Delta(Q_1, Q_2) \geq m_{12}$, because X_1 and X_2 cannot be rotated to a better fit than that with sum of squares m_{12}^2 . Hence the metric inequality $m_{12} \leq m_{13} + m_{23}$ must hold. Analysis of several sets of data suggests further that, provided reflections are excluded, except perhaps for an initial reflection of some of the X_i , the m_{uv} form a matrix of Euclidean distances, but I have not been able to prove this. The $m \times m$ matrix of all comparisons may be

analyzed by metric or non-metric scaling methods, as exemplified by Gower [1971], Krzanowski [1972] and Banfield and Harries [1974]. These studies show that the quantities m_{uv} are of interest in their own right and are not to be regarded as merely measures of goodness of fit.

This paper is concerned with another form of analysis of m sets of matrices. Instead of examining all possible pairs, the Procrustes idea is generalized so that all m sets are simultaneously translated, rotated, reflected and scaled so that a goodness of fit criterion is optimised. The criterion adopted is to minimize the sum-of-squares between each cluster of m like-points $P_i^{(i)}$ ($i = 1, 2, \dots, m$) and their centroid G_i , summed over all n clusters. The centroids are shown in Figure 1. The mn lengths $\Delta(P_i^{(i)}, G_i)$ are termed *residuals*. The residual sum-of-squares S_r is therefore

$$(1) \quad S_r = \sum_{i=1}^n \sum_{i=1}^m \Delta^2(P_i^{(i)}, G_i).$$

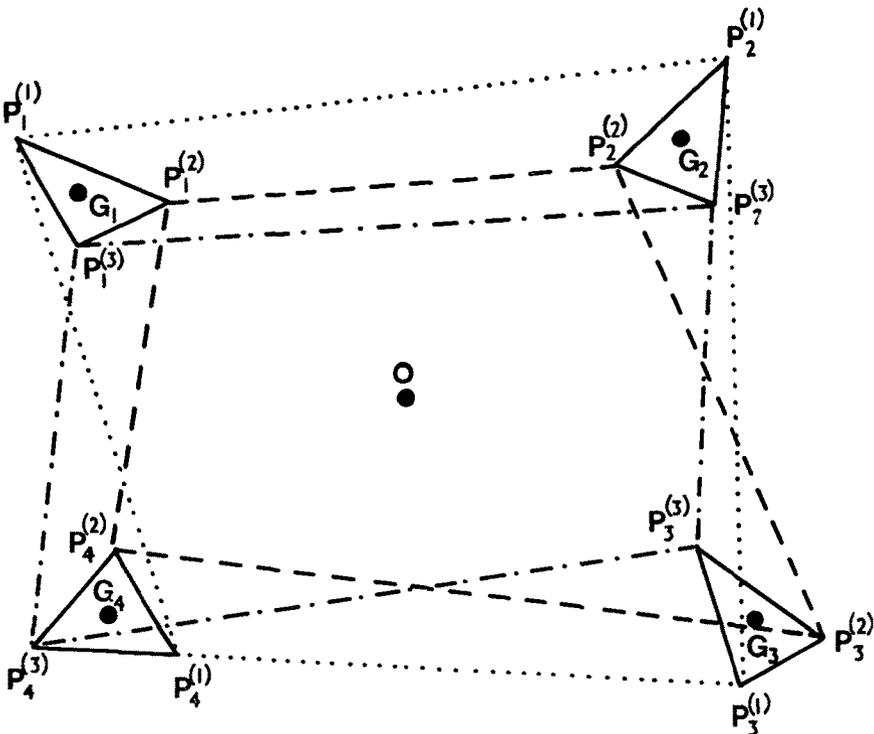


FIGURE 1

Geometrical representation of three configurations ($m = 3$) each with two dimensions ($p_1 = p_2 = p_3 = 2$) and four vertices ($n = 4$) referring to the same four entities. Each entity therefore gives rise to a triangular cluster of vertices whose centroids G_j ($j = 1, 2, 3, 4$) are marked. The centroid of the whole system is at O .

However because of the identity

$$(2) \quad \sum_{u < v}^m \Delta^2(P_i^{(u)}, P_i^{(v)}) \equiv m \sum_{u=1}^m \Delta^2(P_i^{(u)}, G_i)$$

it is simpler to work in terms of $S = m S_r$ given by

$$(3) \quad S = \sum_{i=1}^n \sum_{u < v}^m \Delta^2(P_i^{(u)}, P_i^{(v)}).$$

Rather than minimize a sum-of-squares of residuals, other criteria could be considered which might have desirable properties. As usual the least-squares criterion leads to tractable algebra and straight-forward computation. The special case of minimizing (3), but without translation, reflection and scaling considerations, has been considered by Kristof & Wingersky [1971], who derive what is essentially the result (10), below, by an ingenious algebraic argument. The alternative Lagrange multiplier approach used here is more pedestrian, but copes easily with the additional parameters being estimated in this work.

Algebraic Development

Although the geometrical discussion is useful to establish motivation, and also for suggesting the analysis of variance described later, it does not help with the minimization problem, which must be treated algebraically.

We are given $m(n \times p)$ matrices $X_i (i = 1, 2, \dots, m)$ where the j th row of X_i is interpreted as giving the coordinates of a point $P_i^{(j)}$ in Euclidean space. Rotating the configurations of n points given by X_i is equivalent to post-multiplying X_i by an orthogonal matrix H_i , and uniform scaling is expressed by a multiplicative constant ρ_i . Translation to a new origin is achieved by adding the same $(1 \times p)$ row-vector t_i to every row of X_i . Writing T_i for the $(n \times p)$ matrix, all of whose rows are t_i , then scaling, rotating and translating are expressed algebraically by the transformation

$$X_i \rightarrow \rho_i X_i H_i + T_i.$$

It is required to determine $\rho_i, H_i, T_i (i = 1, 2, \dots, m)$ so that the "residual sum-of-squares" (3) is minimum.

Thus the algebraic problem is to minimize

$$(4) \quad S \equiv \text{tr} \sum_{i < j}^m [(\rho_i X_i H_i + T_i) - (\rho_j X_j H_j + T_j)] \cdot [(\rho_i X_i H_i + T_i) - (\rho_j X_j H_j + T_j)]'.$$

Before proceeding to minimize S , some preliminary discussion is necessary.

Clearly S is minimized by choosing all ρ_i to be zero. Apart from this trivial solution a single matrix (X_1 , say) could be selected, ρ_1 set to unity

and all other matrices scaled relative to that one. This gives the difficulties of classical orthogonal Procrustes rotation ($m = 2$) where fitting X_1 to X_2 does not give the inverse scaling to that of fitting X_2 to X_1 . A more satisfactory method is to estimate all m parameters ρ_i subject to some constraint and the constraint $\sum_{i=1}^m \rho_i^2 \text{tr}(X_i X_i') = \sum_{i=1}^m \text{tr}(X_i X_i')$ has been chosen. This constraint expresses that the final sum-of-squares about the origin of the rotated and scaled configurations is unchanged from that given. We shall also see that it insures that the residual sum-of-squares S_r remains unchanged whether or not the parameters ρ_i are included. Of course setting the final sum-of-squares to any positive constant other than $\sum_{i=1}^m \text{tr}(X_i X_i')$ would only affect the solution found, by a proportionate factor. Constraints other than on sums-of-squares might be considered but these are likely to lead to more awkward algebra and computation.

Because translation terms are being included, the original means of each X_i are immaterial, and it is convenient to take them all to be at the origin; that is all the column sums of each X_i are assumed to be zero.

As S is invariant to orthogonal rotations of the whole system of nm points, no unique solutions for H_i can be found. Here again X_1 , say, could be fixed and all rotations determined relative to it. However, a (non-unique) solution giving all m rotation matrices is preferable, because it keeps the algebraic equations in a symmetric form. A unique solution can be determined as a final step by referring all final coordinates to the principal axes of the set of centroid-points $G_i (i = 1, 2, \dots, m)$. We are therefore lead to minimize (4) subject to the constraints

$$(5) \quad \sum_{i=1}^m \rho_i^2 \text{tr}(X_i X_i') = \sum_{i=1}^m \text{tr}(X_i X_i')$$

$$(6) \quad \text{Each } H_i \text{ is orthogonal, } i.e.$$

$$\sum_{k=1}^p h_{uk}^{(i)} h_{vk}^{(i)} = \delta_{uv}, \text{ the Kronecker-}\delta, \text{ for } u \leq v, v = 1, 2, \dots, p.$$

Associate with (5) the Lagrange multiplier μ and with (6) the $\frac{1}{2}mp(p+1)$ Lagrange multipliers $\lambda_{uv}^{(i)}$. It will be convenient to consider these as arranged in m symmetric matrices $\Lambda_i (i = 1, 2, \dots, m)$ with general elements $\lambda_{uv}^{(i)}$ ($u \neq v$) and $2\lambda_{uv}^{(i)}$ on the diagonal. Thus finally we have to minimize

$$(7) \quad S + \mu \left(\sum_{i=1}^m \rho_i^2 \text{tr} X_i X_i' - \sum_{i=1}^m \text{tr} X_i X_i' \right) + \sum_{r=1}^m \sum_{i \leq j}^p \lambda_{ij}^{(r)} \left(\sum_{k=1}^p h_{ik}^{(r)} h_{jk}^{(r)} - \delta_{ij} \right).$$

Translation

The only terms involving T_i occur in S , (4), and are

$$(8) \quad \underline{S_T} \equiv \text{tr} \left[(m-1)(\rho_i X_i H_i + T_i)(\rho_i X_i H_i + T_i)' \right. \\ \left. - 2 \left\{ \sum_{i \neq i}^m (\rho_i X_i H_i + T_i) \right\} T_i' \right].$$

Differentiating (8) w.r.t. the elements of the row vector t_i gives $\partial S_T / \partial t_i \equiv n m t_i$, (the vector of column sums of $\sum_{i=1}^m T_i$). Thus at the minimum, $t_1 = t_2 = \dots = t_m$. This implies that all m configurations should be translated to have the same centroid, which without loss of generality is conveniently chosen to be the origin. The terms of (4) in T_i ($i = 1, 2, \dots, m$) may now be dropped from further consideration.

Rotation

Differentiating (7) w.r.t. $h_{ik}^{(r)}$ gives

$$\rho_r \{ \rho_1 (X_r' X_1 H_1)_{ij} + \dots + \rho_{r-1} (X_r' X_{r-1} H_{r-1})_{ij} + \rho_{r+1} (X_r' X_{r+1} H_{r+1})_{ij} + \dots \\ + \rho_m (X_r' X_m H_m)_{ij} \} - \sum_{k \neq i}^2 \lambda_{ik}^{(r)} h_{ik}^{(r)} - 2 \lambda_{ij}^{(r)} h_{ij}^{(r)}.$$

Equating this to zero and expressing the results in matrix terms gives

$$(9) \quad \rho_r X_r' (mY - \rho_r X_r H_r) = \Lambda_r H_r, \quad r = 1, 2, \dots, m$$

where $Y = 1/m \sum_{i=1}^m \rho_i X_i H_i$ are the coordinates of the centroid of the group, or consensus configuration, after rotation and scaling. Post-multiplying by H_r' and rearranging gives

$$(10) \quad (\rho_r X_r' Y) H_r' = (\rho_r^2 X_r' X_r + \Lambda_r) / m.$$

Writing $\rho_r X_r' Y$ in its singular value (or Eckart-Young) form: $\rho_r X_r' Y = U_r' \Gamma_r V_r$, where U_r and V_r are orthogonal and Γ_r is diagonal, and noting that the right-hand-side of (10) is symmetric, gives

$$(11) \quad H_r = U_r' V_r.$$

The essential step in deriving this result is proving that $\rho_r X_r' Y H_r'$ is symmetric. As pointed out earlier, this result was established by Kristof & Wingersky [1971] using an algebraic argument that did not rely on Lagrange multipliers. However, their approach assumes that certain singular values are positive and therefore their proof may require modification when rotations without reflections are required, a problem considered in the next paragraph.

As in the two-group Procrustes problem, the reduction in sums-of-squares due to rotation increases with $\text{tr}(\Gamma_r)$, so is maximized by choosing signs attached to the columns of U_r and V_r so that Γ_r has no negative elements. Such a choice of signs may include reflection as well as rotations. R. W. M. Wedderburn (personal communication) points out that for rotation *with*

reflection $|H_r| = -1$ so that the best rotation *without* reflection is obtained by first arranging that Γ_r has no negative elements and then determining $|H_r|$. If this is positive a true rotation has been found, but if negative the sign must be changed of the column of either U_r or V_r that corresponds to the smallest non-zero element of Γ_r . This may make the smallest singular value negative but ensures that $\text{tr}(\Gamma_r)$ is maximized subject to $|H_r| = 1$. Gower [1975] gives an algorithm for determining the sign of the determinant of an orthogonal matrix.

Equation (11) does not give an immediate method for calculating H_r , because the centroid matrix Y is not known. However, it gives a fairly obvious basis for an iterative computational method discussed below.

Scale

Differentiating (7) with respect to ρ_i and equating the result to zero gives:

$$(m - 1)\rho_i \text{tr} X_i X_i' - \text{tr} \left(X_i H_i \sum_{\substack{j=1 \\ j \neq i}}^m \rho_j H_j' X_j' \right) + \mu \rho_i \text{tr} X_i X_i' = 0.$$

or

$$(m - 1)\rho_i \text{tr} X_i X_i' - \text{tr} \left(X_i H_i \sum_{j=1}^m \rho_j H_j' X_j' \right) + \rho_i \text{tr} X_i X_i' + \mu \rho_i \text{tr} X_i X_i' = 0.$$

Finally

$$(12) \quad m\rho_i \text{tr} X_i X_i' - m \text{tr} (X_i H_i Y') + \mu \rho_i \text{tr} X_i X_i' = 0.$$

Multiplying (12) by ρ_i , summing over $i = 1, 2, \dots, m$ and recalling the constraint (5) yields:

$$(m + \mu) \sum_{i=1}^m \text{tr} X_i X_i' = m^2 \text{tr} Y Y'$$

and hence

$$(13) \quad \rho_i = \text{tr} (X_i H_i Y') \sum_{i=1}^m \text{tr} (X_i X_i') / m \text{tr} (X_i X_i') \text{tr} (Y Y').$$

The alternative form

$$(14) \quad \rho_i^2 = \text{tr} (\rho_i X_i H_i Y') \sum_{i=1}^m \text{tr} (X_i X_i') / m \text{tr} (X_i X_i') \text{tr} (Y Y').$$

will be useful for computational purposes (see below).

Because Y itself involves the scaling factors, the above formulae do not give a direct method for calculating the ρ_i , but have to be used iteratively (see below). However, (12) is the same equation for determining the ρ_i as

when given $X_i H_i$ are to be scaled to fit *any* configuration Y and (13) and (14) still follow but with $\text{tr}(YY')$ replaced by $\text{tr}(\sum_{i=1}^m \rho_i X_i H_i Y')/m$.

This completes the algebraic derivation of the estimates of the parameters that minimize the residual sum-of-squares (4).

Analysis of Variance

The foregoing analysis can be expressed as an analysis of variance that is useful for identifying the relative importance of the items that go to make up the total sum-of-squares. In the usual way the total sum-of-squares can be split into between-groups and within-groups components, where the term groups refers to the m sets or configurations X_i ($i = 1, 2, \dots, m$). The between-group component represents the contribution of the translation terms. The within-group component splits into consensus and residual elements through the identity

$$(15) \quad \text{tr} \sum_{i=1}^m (\rho_i^2 X_i X_i') \equiv m \text{tr}(YY') + S_r$$

which is merely the multivariate form of the well-known univariate identity

$$\sum_{i=1}^m y_i^2 \equiv m\bar{y}^2 + \sum_{i=1}^m (y_i - \bar{y})^2.$$

The left-hand-side of (15) is the within-group sum-of-squares after scaling and rotation, and because of the constraint (5) this is the same as the within-groups sum-of-squares before transformation.

The i th term is the contribution of the i th configuration to the total within-group sum-of-squares and shows the reduction, or increase, due to scaling. If the configurations have been initially standardized to unit sum-of-squares, *i.e.* $\text{tr}(X_i X_i') = 1$, the contribution becomes ρ_i^2 , the square of the scaling coefficient. On the right-hand-side of (15), $m \text{tr}(YY')$ is a term representing the contribution of the group average, or consensus, configuration. Geometrically (see Figure 1) this is $m \sum_{i=1}^m \Delta^2(O, G_i)$. The residual S_r can be partitioned in two separate ways. Firstly, we can sum the squares of the residuals for each group, giving terms $\sum_{i=1}^m \Delta^2(P_i^{(i)}, G_i)$ ($i = 1, 2, \dots, m$) and secondly, we can sum the squares for each stimulus giving terms $\sum_{i=1}^m \Delta^2(P_i^{(j)}, G_j)$ $j = 1, 2, \dots, n$. These considerations give the tabular form of analysis of variance expressed geometrically in Table 1.

Computation

The first step is to compute the between-groups sums of squares and to deal with translation by centering each set X_i at the origin. To allow for different magnitudes of data, it is convenient to scale each X_i uniformly so that $\sum_{i=1}^m \text{tr}(X_i X_i') = m$. To return to the original units, results should be multiplied by the scaling factor. Thereafter the process becomes iterative.

TABLE 1
Analysis of Variance

Source	Method of Calculation and Interpretation		
Between Groups	Translation		
Within Groups By Stimulus	Consensus	Residual	Total (Within-Groups)
1	$m\Delta^2(OG_1)$	$\sum_{i=1}^m \Delta^2(P_1^{(i)}G_1)$	$\sum_{i=1}^m \Delta^2(OP_1^{(i)})$
2	$m\Delta^2(OG_2)$	$\sum_{i=1}^m \Delta^2(P_2^{(i)}G_2)$	$\sum_{i=1}^m \Delta^2(OP_2^{(i)})$
⋮	⋮	⋮	⋮
n	$m\Delta^2(OG_n)$	$\sum_{i=1}^m \Delta^2(P_n^{(i)}G_n)$	$\sum_{i=1}^m \Delta^2(OP_n^{(i)})$
Stimuli	$m \operatorname{tr} (YY')$	S_r	$\sum_{i=1}^m \operatorname{tr} (X_iX_i')$
By Group	Consensus	Residual	Total (Within-Groups)*
1		$\sum_{i=1}^n \Delta^2(P_i^{(1)}G_i)$	$\sum_{i=1}^n \Delta^2(OP_i^{(1)}) = \rho_1^2 \operatorname{tr} (X_1X_1')$
2		$\sum_{i=1}^n \Delta^2(P_i^{(2)}G_i)$	$\sum_{i=1}^n \Delta^2(OP_i^{(2)}) = \rho_2^2 \operatorname{tr} (X_2X_2')$
⋮		⋮	⋮
m		$\sum_{i=1}^n \Delta^2(P_i^{(m)}G_i)$	$\sum_{i=1}^n \Delta^2(OP_i^{(m)}) = \rho_m^2 \operatorname{tr} (X_mX_m')$
Groups		S_r	$\sum_{i=1}^m \operatorname{tr} (X_iX_i')$

Total = Between + Within

$$= \text{Between} + \sum_{i=1}^m \rho_i^2 \operatorname{tr} (X_iX_i')$$

$$= \text{Between} + [m \operatorname{tr} (YY') + S_r]$$

* With standardization $\operatorname{tr}(X_iX_i') = 1$ ($i = 1, 2, \dots, m$) these terms simplify to ρ_i^2 with a total of m .

Initial rotations for each set can be found by using ordinary 2-set Procrustes rotations [Gower 1971, Schönemann & Carroll 1970]. First X_2 is rotated to fit X_1 and then the matrix Y giving the centroids of the n pairs

of vertices $P_i^{(1)}, P_i^{(2)}$ ($i = 1, 2, \dots, n$) given by the current positions of X_1 and X_2 is found. X_3 is rotated to fit Y and Y recomputed. Similarly subsequent X_i are fitted to the current value of Y . This gives initial values of each X_i and their mean Y and the initial residual sum-of-squares $S_r = m(1 - \text{tr}(YY'))$.

Next keep Y fixed and rotate each X_i in turn to fit Y . After all m sets have been rotated, recompute Y to give the new centroid Y^* of the current X_i , and the residual sum-of-squares $S_r^* = S_r - m \text{tr}(Y^*Y^{*'} - YY')$. If scaling is not required and S_r has changed by less than some specified tolerance, iteration is completed, otherwise repeat this process until it converges. Rotating each X_i to fit Y cannot increase, and will usually reduce, the residual sums-of-squares about Y . When Y is replaced by Y^* , the true current mean, the residual sum of squares is further reduced. S_r is bounded below by zero and during iteration it is monotonically decreasing. This is insufficient to prove convergence, but it seems that the process converges satisfactorily in practice.

The preceding steps correspond to those outlined by Kristof & Wingersky [1971]. The following additional steps are required to estimate the scale factors ρ_i . Dropping the asterisk notation so that the current configurations are now again written X_i ($i = 1, 2, \dots, m$) with mean Y , then the least squares estimates of scaling coefficients ρ_i such that $\rho_i X_i$ best fit Y are given by (13) or its equivalent (14). $X_i H_i$ there may be replaced by X_i without loss of generality, for this merely implies that the current rotated positions of the X_i configurations are being regarded as starting configurations. If X_i currently has scale ρ_i , then its new scale ρ_i^* is estimated numerically from the following formula derived from (14).

$$(16) \quad \rho_i^{*2} = \frac{\text{tr}(\rho_i X_i Y')}{\text{tr}(X_i X_i')} \bigg/ \frac{m \text{tr}(Y Y')}{\sum \text{tr}(X_i X_i')}.$$

The initial scaling $\sum_{i=1}^m \text{tr}(X_i X_i') = m$ simplifies (16), which may be rearranged to give

$$(17) \quad \frac{\rho_i^{*2}}{\rho_i^2} = \frac{\text{tr}(\rho_i X_i Y')}{\text{tr}(\rho_i^2 X_i X_i') \text{tr}(Y Y')}.$$

This form implies that the current scaled versions $\rho_i X_i$ may replace X_i throughout the computation and (17) be used to adjust the scaling to

$$(18) \quad \rho_i^* X_i = \frac{\rho_i^*}{\rho_i} (\rho_i X_i).$$

New values $\rho_i^* X_i$ are found for each configuration $\rho_i X_i$ and a new mean Y^* calculated, where $Y^* = \sum \rho_i^* X_i / m$. The new residual sum of squares is given by $S_r^* = S_r - m \text{tr}(Y^* Y^{*'} - YY')$, the same form as after the rotational step. The forms (16) and (17) ensure that

$$(19) \quad \sum \rho_i^{*2} \text{tr} (X_i X_i') = \sum \text{tr} (X_i X_i') = m$$

so that the constraint (5) on the scaling coefficients is satisfied throughout the iteration. Because Y is the mean of current $\rho_i X_i$, S_r^* can never be greater than S_r , so the process always reduces the residual sum-of-squares.

Convergence may now be tested in the same way as when no scaling is required.

This computational process seems to converge rapidly in practice, but the possibility that a different mix of rotational and scaling steps (say two scaling steps to one rotational step) might be superior has not been investigated. The residual sum-of-squares is reduced even when the mean Y is not updated after rotation, but whether or not the minor saving in not doing so is outweighed by requiring more iterations is unknown.

After satisfactory convergence, a unique representation of the results is obtained by referring Y and each $\rho_i X_i$ to the principal axes of Y .

Summary of computation

1. Evaluate between-groups sum-of-squares (*i.e.* translation term in A.O.V.).
2. Center X_i ($i = 1, 2, \dots, m$) and scale each X_i by λ so that

$$\sum_{i=1}^m \lambda \text{tr} (X_i X_i') = m.$$

3. Set $Y = X_1$ (initial setting of matrix of means). For $i = 2, 3, \dots, m$ rotate X_i to fit Y , re-evaluate Y as mean of (X_1, X_2, \dots, X_i) . Evaluate initial residual sum-of-squares $S_r = m(1 - \text{tr} (YY'))$ and set $\rho_i = 1$ ($i = 1, 2, \dots, m$).
4. For $i = 1, 2, \dots, m$ rotate the current matrix $\rho_i X_i$ to fit Y giving $X_i' = \rho_i X_i H_i$. Compute Y^* and $S_r^* = S_r - m \text{tr}[Y^* Y^{*'} - YY']$. Set $S_r^{**} = S_r^*$.
5. If scaling is not required go to step 7.
6. For $i = 1, 2, \dots, m$ evaluate ρ_i^*/ρ_i from (17), scale $X_i^{**} = (\rho_i^*/\rho_i) X_i^*$ and set $\rho_i = \rho_i^*$. Compute new mean Y^{**} and new residual sum-of-squares $S_r^{**} = S_r^* - m \text{tr}[Y^{**} Y^{**'} - Y^* Y^{*'}] = S_r - m \text{tr}[Y^{**} Y^{**'} - YY']$.
7. If $S_r - S_r^{**} > \text{tolerance}$, set $S_r = S_r^{**}$ and go to step 4, else go to next step.
8. Iteration is complete. Calculate and print analysis of variance.
9. Find principal axes of Y , *i.e.* compute the orthogonal latent-vector matrix H satisfying $(Y'Y)H = HM$ where M is the diagonal matrix of latent roots.
10. Refer all sets of coordinates to these principle axes and print final results $YH, \rho_i X_i H, \rho_i$ ($i = 1, 2, \dots, m$).

This computation was programmed in Genstat (see Nelder et al [1973]). The whole process of program-writing and development was completed in two days. A convergence tolerance of .0001 was found satisfactory.

A Special Case

Sometimes the data-sets $X_i (i = 1, 2, \dots, m)$ will clearly be incommensurate and some initial standardization will be essential. An example is when each X_i uses a different measurement scale. The simplest standardization is to arrange that $\text{tr}(X_i X_i') = 1 (i = 1, 2, \dots, m)$. This simplifies the elements $\rho_i^2 \text{tr}(X_i X_i')$, representing the within-groups variance for the i th group in the analysis of variance of Table 1, to ρ_i^2 which can be identified immediately as the contribution to the total variance induced by scaling the i th group.

With this scaling and only two groups ($m = 2$), (12) gives:-

$$(20) \quad (2 + \mu)\rho_i = 2 \text{tr}(X_i Y') \quad i = 1, 2$$

where X_i now represents the final rotated version of the original data-sets. Substituting $Y = \frac{1}{2}(\rho_1 X_1 + \rho_2 X_2)$ into (20) gives

$$(2 + \mu)\rho_1 = \rho_1 + \rho_2 \text{tr}(X_1 X_2')$$

$$(21) \quad \text{and}$$

$$(2 + \mu)\rho_2 = \rho_2 + \rho_1 \text{tr}(X_2 X_1').$$

Hence $\rho_1 = \rho_2$ and because the constraint (5) becomes $\sum_{i=1}^2 \rho_i^2 = 2$, we must have $\rho_1 = \rho_2 = 1$.

This shows that with two-groups, the estimation of scaling is equivalent to an initial standardization of each data-set to have the same, not necessarily unit, sums-of-squares. If it is desired to preserve the scaling constraint (5), initial standardization should adjust both sets to have sum-of-squares $[\text{tr}(X_1 X_1') \pm \text{tr}(X_2 X_2')]/2$. With two groups, because

$$2S_r = \sum_{i=1}^n \Delta^2(P_i^{(1)} P_i^{(2)}),$$

the use of residuals from the means G_i is equivalent to the criterion of classical Procrustes analyses and then the iterative form of computation needed for the general case is unnecessary. With more than two groups initial standardization is not equivalent to scaling.

Example

The example illustrated here uses part of the data from a much larger investigation, in which a number of judges were required to rank given beef carcasses with respect to several characters. The main interest is to see how the judge's views of the same carcasses differ, rather than examine differences between the carcasses themselves. These data have been examined in many ways [Pomeroy, et al. 1974a, b; Banfield & Harries 1974] and clearly an INDSCAL analysis might also prove interesting although it has not been

tried. There is insufficient space to discuss these data in detail here or to present a full generalized Procrustes analysis. The objectives are to illustrate the main features of such an analysis and to provide an example which can be used to check any computer program others might write for this analysis.

For these limited purposes the data for 3 judges, 7 characters and 9 carcasses were used, and are shown in Table 2. Thus we have $m = 3$, $p = 7$, $n = 9$. The scores were obtained by marking an appropriate point on a linear scale.

TABLE 2

The Scores Given by Three Judges for Seven Characters Observed on Nine Beef Carcasses

Judge 1							
Character	1	2	3	4	5	6	7
Carcass 1	47	44	49	38	35	40	40
2	72	45	41	77	72	73	35
3	61	49	40	58	58	62	30
4	66	56	45	55	53	46	30
5	37	72	50	27	30	33	25
6	76	76	53	81	79	75	45
7	64	59	51	72	61	66	40
8	21	70	43	27	22	26	20
9	71	70	34	72	72	71	35
Judge 2							
Character	1	2	3	4	5	6	7
Carcass 1	31	39	33	29	48	38	42
2	30	60	36	22	36	34	39
3	27	55	30	18	28	22	42
4	48	52	53	27	21	30	31
5	20	55	28	22	33	27	35
6	21	42	31	46	76	33	42
7	30	52	53	35	44	30	44
8	5	57	53	12	13	6	31
9	55	63	53	77	79	57	49
Judge 3							
Character	1	2	3	4	5	6	7
Carcass 1	43	46	44	22	53	44	29
2	53	79	75	79	73	52	27
3	22	85	83	19	27	17	22
4	28	89	78	13	29	20	24
5	75	86	85	34	75	55	38
6	53	79	82	72	78	74	38
7	15	85	85	46	75	52	35
8	5	95	95	3	20	2	24
9	27	78	85	89	92	81	41

As all scores are comparable, there were no preliminary transformations. The total between- and within-judges sums of squares are shown in Table 4. The between-judge sum of squares, which is substantial, shows that there are large differences between the mean scores of the three judges and hence that the translation of each configuration to a common origin removes much of the variation between judges.

There remain other differences between the judges that can be discerned from the generalized Procrustes analysis. These differences are those arising (a) from any consistent eccentricities in the use of the measurement scales, such as a tendency to use an unusually small or large part of the scale, and (b) from the different orientations of the multivariate spaces that might allow all judges to perceive different scores that nevertheless reflected the same relationships between each pair of carcasses. Table 3 shows the consensus scores referred to their first two principal axes together with the scores of the three judges referred to the same axes. Table 4 shows the associated analysis of variance, for all seven dimensions. In Tables 3 and 4 all values have been scaled so that the total within-groups sum-of-squares is m , *i.e.*

TABLE 3

First two components of consensus fit and associated values for the three individual judges. The figure after each judge name is his estimated scaling factor ρ .

	Consensus		Judge 1 (1.034)	
	Component 1	Component 2	Component 1	Component 2
Carcass 1	0.088	0.180	0.244	0.126
2	-0.145	0.016	-0.269	-0.014
3	0.143	-0.006	-0.049	-0.006
4	0.150	-0.045	0.020	0.009
5	0.150	0.094	0.372	0.020
6	-0.318	0.065	-0.385	0.002
7	-0.094	-0.057	-0.171	0.021
8	0.464	-0.124	0.501	-0.061
9	-0.436	-0.121	-0.262	-0.097
	Judge 2 (1.105)		Judge 3 (0.913)	
	Component 1	Component 2	Component 1	Component 2
Carcass 1	-0.052	0.154	0.071	0.259
2	0.069	0.024	-0.236	0.038
3	0.184	0.065	0.294	-0.078
4	0.129	-0.109	0.301	-0.035
5	0.144	0.089	-0.067	0.171
6	-0.269	0.145	-0.299	0.047
7	-0.028	-0.080	-0.084	-0.113
8	0.442	-0.112	0.448	-0.199
9	-0.620	-0.176	-0.428	-0.091

TABLE 4
Analysis of Variance

Between Judges	53254.87*		
Within Judges	22114.80*		
By Carcass	Consensus	Residual	Total
1	.162	.066	.228
2	.114	.084	.198
3	.087	.079	.166
4	.126	.067	.193
5	.158	.129	.287
6	.361	.038	.399
7	.059	.026	.085
8	.712	.032	.744
9	.622	.078	.700
Carcasses	2.401	.599	3.000
By Judge			
1		.240	.930
2		.177	1.031
3		.182	1.039
Judges		.599	3.000

* These terms are in the units (squared) of the data given in Table 2. The sums of squares in the remainder of this table should be multiplied by 7371.6 to return to the same basis.

trace $\sum_{i=1}^3 (\rho_i^2 X_i X_i') = 3$; to return to the scale of the data in Table 2, multiply by 7371.6 for sums-of-squares, and 85.86 for scores. Table 3 is given pictorially in Figure 2.

From Table 3 the scales ρ_i are much the same, so there is little evidence of using different ranges of the scales. The analysis of variance of Table 4 shows that Judge 1 has the biggest residual and therefore differs most from the consensus opinion, while carcass number 5 is the one about which the judges were least agreed. Carcasses 2, 3 and 9 also had rather high residuals. These facts can be seen also in Figure 2, but it must be remembered that the two dimensional configurations give only an approximation to the true residuals. In this example the approximation is quite good as the first two components of the consensus configuration account for 86% of the variance. Although these two dimensions do not span the spaces of the first principal components of the individual judges, they nevertheless account for 86%, 83% and 86% of the variance for the three judges, respectively.

Table 5 gives the successive values of the residual sum-of-squares for this data, both when scaling coefficients are being estimated and when not. This table will be useful for those developing their own generalized Procrustes

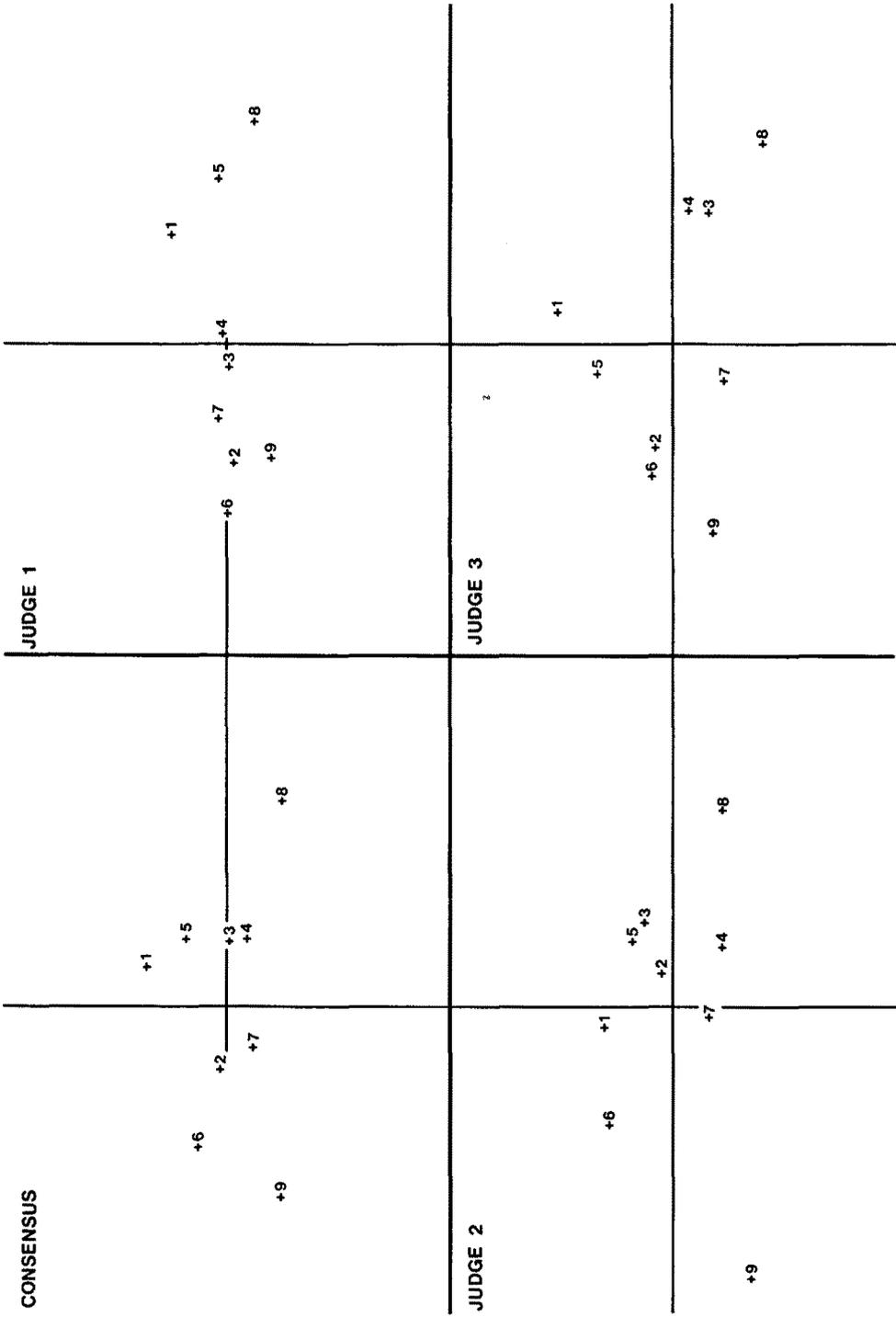


FIGURE 2

Consensus results referred to their first two principal axes with rotated and scaled judges' scores referred to the same axes.

TABLE 5
Values of S_r for the Data of Table 2 Both With and Without Scaling

Iteration Number	With Scaling		Without Scaling
	Rotation Step	Scaling Step*	Rotation Step*
Initial setting		.661438	.661438
1	.657312	.616714	.657312
2	.616620	.604215	.657158
3	.604201	.600456	.657137
4	.600452	.599322	
5	.599322	.598978	
6	.598980	.598875	
7	.598877	.598842	

* Convergence is assessed from changes in the criterion after the indicated step.

analysis program. Iteration is stopped when successive values of S_r differ by less than .0001, the test being done after the scaling step, when scaling is required. Clearly this level of tolerance is near its limit for the numerical process as programmed, for S_r actually increased very slightly after the rotation steps of iterations 6 and 7. As S_r should theoretically decrease, this discrepancy must reflect numerical inaccuracy.

Discussion

Clearly the method discussed above is related to the class of methods that analyze the multivariate behavior of individuals and contrast these behaviors with some kind of average behavior. The best-known procedure of this class is Individual Scaling Analysis, INDSCAL, [Carroll & Chang, 1970]. Generalized Procrustes Analysis is not intended as a rival to INDSCAL but provides a complementary analysis giving different information.

In Generalized Procrustes Analysis differences between the mean scores of individuals may be examined in the translation terms of the analysis and in the "between groups sum-of-squares". After eliminating differences between means, the residual scores of two individuals may differ in the ranges of the measurement scales being used and in orientation. These differences can be examined in detail in the general analysis.

The analysis of variance gives a useful summary of the differences between individuals and how the different stimuli contribute to these differences. As yet no attempt has been made to provide degrees of freedom or to work out any sampling theory for the observed sums-of-squares; clearly improvements can be expected here. The rotational contribution to an analysis of variance is of general interest as it seems to be a truly multivariate aspect that has no counterpart in univariate analyses. Its value may be limited when the columns of the data matrices $X_i (i = 1, 2, \dots, m)$ are measurements on known scales

but when the X_i have been derived from some scaling analysis of association-data (distances, similarities, proximities etc.), the initial orientation of the X_i configurations is arbitrary and some rotational analysis seems essential if spurious differences are to be eliminated from the data; the danger is that spurious similarities might be introduced.

When dealing with data for more than two individuals it is natural to consider rotations with respect to the common centroid or consensus configuration. The same approach for $m = 2$ proves advantageous in overcoming a difficulty in previous approaches, where one of the two configurations is regarded as fixed and the other rotated to fit it.

Schönemann [1968] has considered two-sided Procrustes problems in which orthogonal matrices T and S are required such that $T'X_2S$ best fits X_1 or, when X_1 and X_2 are symmetric, $T'X_2T$ best fits X_1 . Gruvaeus [1970] considers oblique transformations where T is not restricted to be orthogonal but X_2T best fits X_1 subject to $\text{diag}(T'T)^{-1} = I$, expressing that $(T'T)^{-1}$ gives cosines between transformed axes. Clearly such problems can be generalized along the lines discussed above to deal with more than two matrices but potential applications would be required for this to be worthwhile.

REFERENCES

- Banfield, C. F. & Harries, J. M. A technique for comparing judges' performance in sensory tests. *J. Food Technology* (In press).
- Carroll, J. D. & Chang, J. J. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 1970, **35**, 283-320.
- Eckart, V. & Young, G. The approximation of one matrix by another of lower rank. *Psychometrika*, 1935, **1**, 211-218.
- Gower, J. C. Statistical methods of comparing different multivariate analyses of the same data. In *Mathematics in the Archaeological & Historical Sciences*. Hodson, F. R., Kendall, D. G., & Tautu, P. Edinburgh: Edinburgh Univ. Press, 1971, 138-149.
- Gower, J. C. The determinant of an orthogonal matrix. Algorithm AS 82. *Applied Statistics*, (In press) **24**.
- Gruvaeus, G. T. A general approach to Procrustes pattern rotation. *Psychometrika*, 1970, **35**, 493-505.
- Kristof, W. & Wingersky, B. Generalization of the orthogonal Procrustes rotation procedure to more than two matrices, 1971. Proceedings, 79th Annual Convention, A. P. A., 81-90.
- Krzanowski, W. J. A comparison of some distance measures applicable to multinomial data, using a rotational fit technique. *Biometrics*, 1971, **27**, 1062-1068.
- Nelder, J. A. and Members of the Rothamsted Statistics Department. *Genstat Reference Manual*, Inter-University/Research Councils Series, Report No. 3, Second Edition. Edinburgh: Program Library Unit, Edinburgh Regional Computing Centre, 1973.
- Pomeroy, R. W., Williams, D. R., Harries, J. M. & Ryan, P. O. Material, measurements, jointing and tissue separation. *Journal of Agricultural Science*, (In press).
- Pomeroy, R. W., Williams, D. R., Harries, J. M. & Ryan, P. O. The use of regression equations to estimate total tissue components from observations on intact and quartered sides and partial dissection data. *Journal of Agricultural Science*, (In press).

- Schönemann, P. H. On two-sided Procrustes problems. *Psychometrika*, 1968, **33**, 19-34.
- Schönemann, P. H. & Carroll, R. M. Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika*, 1970, **35**, 245-256.

Manuscript received 3/11/74

Revised manuscript received 8/8/74