Food Quality and Preference 3 (1991/2) 67–87

THE INTERPRETATION OF GENERALIZED PROCRUSTES ANALYSIS AND ALLIED METHODS

G. B. Dijksterhuis[a,b] & J. C. Gower[b]

[a] Oliemans Punter and Partners, PO Box 14167, 3508 SG Utrecht, The Netherlands
[b] Department of Data Theory, University of Leiden, PO Box 9555, 2300 RB Leiden, The Netherlands

(Received 23 *August* 1991; *accepted August* 1992*)*

ABSTRACT

We discuss various issues surrounding the use and interpretation of Generalized Procrustes Analysis and related methods. Included are considerations that have to be made before starting an analysis, how to handle different dimensionalities of data, when to consider fitting scaling factors and when not to, and the distinction between the number of dimensions that are needed to give an adequate fit and the number of dimensions needed for graphical representation. The distinction between signal and noise plays an important part in explaining how different methods are suitable for exploring different aspects of the data, rather than being viewed as competing methods with the same general objectives. Explanations are largely set in a geometrical context, thus keeping technical mathematics to a minimum; a common Analysis of Variance framework allows all the methods to be considered in a unified way and suggests some new ways in which these kinds of data may be analysed. The whole is illustrated by example analyses.

Keywords: Generalized Procrustes Analysis, Analysis of Variance, sensory data.

1 INTRODUCTION

This is an expository paper about how to interpret Generalized Procrustes Analysis (Gower, 1975) and related methods. It seems that such a paper is timely as several issues are of current interest, including how to cope with different dimensionalities among the data-sets concerning different individuals and how to get, and especially how to interpret, optimal representations in a few dimensions. The situation is complicated by a tendency to regard essentially different methods, that happen to have superficial similarities, as variant algorithms for solving the same substantive problem. More usually, different methods give information on different aspects of the data, so may be used to help answer different questions. It is therefore our principal aim to distinguish clearly between objectives so that practitioners can make an informed choice among methods appropriate to their needs. To complete the picture, but very much as a secondary aim, we also briefly discuss in section [A.5] the roles of loss-functions/goodness-of-fit criteria and computing algorithms. We do not explain how to do the computations or discuss the availability of software.

Generalized Procrustes Analysis (GPA) is one of the family of methods that are concerned with the analysis of data arising from several *individuals* (see, for example, Arabie *et al.*, 1987). Among other things, the individuals may be judges of food quality, or they may represent different laboratories where measurements and other observations have been made on a set of objects, or they may represent replications of the same information made on different occasions. We are concerned with how the individuals differ, and equally to what extent they may agree, in their perceptions of the same things, so it is essential that each individual, however defined, presents data on the same *objects* (our preferred term, used throughout the following), samples, stimuli, or substances. The actual data provided by individuals may be in one of several forms and herein lies a potential area of confusion. There are two major classes of data: Type I where each individual records a set of properties for each object and Type II, where each individual records his perception of the *difference* between each pair of objects. With Type I, properties may be assessed by recording qualitative observations, measurements, or by judgements given as a ranking of the objects. Although it is essential that each individual reports on the same objects; individuals need not record the same properties nor even the same number of properties. Free choice profiling (FCP) (Williams & Langron, 1984; Arnold & Williams, 1985) is one way in which different individuals record different properties (see [A.1.2]). In some ways GPA is best suited to the analysis of data-sets where the individuals report on different properties.

As well as being directly recorded, there are several ways in which Type II information may be calculated from Type I. When individuals do provide Type II data, then essentially each is giving a symmetric distance-matrix. Just as distances can be derived from observations on the properties of the objects, so can Type I data be derived from distances. Further, Type I data may benefit from initial transformation, and distances may be transformed to give new distances which perhaps can be better represented in few dimensions than can the original distances, so leading to further variants of GPA. The links between Types I and II are one source of ambiguities in concepts of dimensionality that are at the root of issues, addressed below, of how best to combine individual data-sets.

Type II data may be asymmetric. For example, each individual might provide a confusion matrix whose cells give the number of times that each pair of objects had been confused in a tasting trial, where the order of tasting mattered. Models for analysing this kind of asymmetric data are beyond the scope of this paper, except to note that one way of proceeding is to separate the confusion matrices into their symmetric and skew-symmetric component parts (Gower, 1977; Constantine & Gower, 1978), analysing the symmetric parts by GPA.

The above preamble should give some idea of the potential flexibility of GPA and dispel and tendency to regard it as a rather rigid analysis of quantitative data, usually in the form of scores.

In this paper, our plan is that each section should be self-contained; however, we use the notation [1.2.3] to point to sections where relevant comments, discussion, amplifications and justifications may be found for assertions or for the steps taken in an analysis. Section [2] gives an informal exposition of GPA, simplifying the mathematics so far as is possible, by relying on geometrical explanations. This is followed, in [3], by a typical GPA and related analyses. Section [4] concludes the paper. *An appendix may be viewed as containing an extended set of footnotes*, though we intend that the account may also be read as a connected discussion of the many issues that surround GPA, some of them of a contentious nature; the Appendix sections are referred to by [A.1.2]. The Appendix starts by discussing, in more detail than is desirable in these introductory remarks, some general considerations concerning the forms of data used in GPA and some of the decisions that have to be made before starting an analysis. This is followed by a discussion of several interrelated issues concerning dimensionality, which naturally leads to the consideration of variant forms of analysis, emphasising how these may be useful in answering questions different from those that GPA is designed to answer. We conclude [4] by providing a check-list of the major items covered.

2 AN INFORMAL EXPOSITION OF GENERALIZED PROCRUSTES ANALYSIS

In this section our aim is to convey the general ideas and assumptions used in GPA. Some notation seems unavoidable but we have kept this to a minimum and, for the most part, rely on geometrical explanations. We assume that there are K individuals, each of whom provides information on the same N objects. With Type I data we have K matrices $\mathbf{X}_k$ ($k = 1, \ldots, K$), each with N rows and P columns, which may be thought of as giving, usually, numerical measurements or scores for each object on each of P variables. If some $\mathbf{X}_k$ initially have fewer than P columns, then it is assumed that these matrices have additional zero columns appended to make them compatible; the justification for this practice of *padding* is discussed below in some detail [A.3.1]. With Type II data we assume that the K ($N \times N$) symmetric distance matrices $\mathbf{D}_k$ ($k = 1, \ldots, K$) have already been replaced by generating coordinate matrices $\mathbf{X}_k$, by using multidimensional scaling methods ([A.1.1], [A.1.2]).

We denote the value obtained by the kth individual for the jth variable on the ith object by x_{ijk}. In the following we rely heavily on the notion that $(x_{i1k}, x_{i2k}, \ldots, x_{iPk})$ can be considered to give the coordinates relative to P axes, of a point A_{ik} representing the P measurements on the ith object as recorded by the kth individual, thus giving a total of $N \times K$ points for all object/individual combinations (*see* Fig. 2 for a representation with $N = 4$, $K = 3$, and $P = 2$). We term such a representation a *configuration*.

2.1 The algebra and geometry of GPA

Only two simple algebraic identities are needed to understand the basics of GPA. These are:

$$\sum_{i<j=1}^{N} (x_i - x_j)^2 = \sum_{i=1}^{N} (x_i - \bar{x})^2 \qquad (1)$$

and

$$\sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} (x_i - \bar{x})^2 + N\bar{x}^2 \qquad (2)$$

where $\bar{x}$ is the mean, and is given by

$$\bar{x} = \sum_{i=1}^{N} x_i / N$$

Equation (1) says that the sum-of-squares of all pairwise distances between a set of points is the same as N times the sum-of-squares of the distance from their centroid. This relates the sum of the squared distances on the left-hand side of eqn (1) to the sum-of-squares of deviations from the mean on the right-hand side, and hence gives a useful distance interpretation of Analysis of Variance that is much used in the following. In

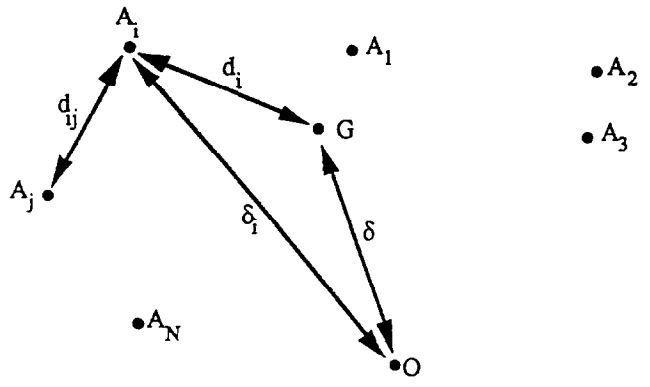


Fig. 1. A configuration of N points $A_1, A_2, \ldots, A_N$ with centroid G and an arbitrary origin O. The distance between two typical points is written in two equivalent forms: $A_iA_j = d_{ij}$. Other distances are written typically $A_iG = d_i$ and $A_iO = \delta_i$ with $GO = \delta$, the distance between the centroid and the origin.

geometric terms (see Fig. 1, where the notation is defined),

$$\sum_{i<j=1}^{N} d_{ij}^2 = N\sum_{i=1}^{N} d_i^2$$

so that when $N = 4$ we have

$$d_{12}^2 + d_{13}^2 + d_{14}^2 + d_{23}^2 + d_{24}^2 + d_{34}^2 = 4(d_1^2 + d_2^2 + d_3^2 + d_4^2)$$

Equation (2) is Huygens's (1629–95) formula, usually stated in terms of moments of inertia but, in the current context, better interpreted as one of the foundations of Analysis of Variance. It states that the sum-of-squares of a set of points about any origin is the same as the sum of their squares about their centroid plus N times the square of the distance between the chosen origin and the centroid of the points. Geometrically (see Fig. 1 for the notation),

$$\sum_{i=1}^{N} \delta_i^2 = \sum_{i=1}^{N} d_i^2 + N\delta^2$$

Although it may not be true of all analyses of variance, it is certainly true that most, including those associated with GPA and allied methods, may be built up from successive uses of Huygens's formula.

2.2 The geometry of configurations

With these results, we are now in a position to demonstrate GPA from a geometrical point of view. Figure 2 establishes some notation and shows a geometrical representation of matrices $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ given by three ($K = 3$) individuals, each representing $N = 4$ objects in $P = 2$ dimensions, here taken to be synonymous with two variables [A.1.2].

In Fig. 2 the configurations for the three individuals are shown with different centroids, G_1, G_2 and G_3, so implicitly we are assuming for the time-being that, unlike with FCP, it is valid to compare the values of variables between individuals [A.1.2]. The configuration for the second individual is contained entirely within the configuration of the first; clearly, these two persons are using different portions of the full ranges of possible scores and some scaling will be needed to account for this [2.4]. However, their centroids G_1 and G_2 are close, so they agree closely on their average scores. Despite the difference in size, these two configurations have similar shapes and orientations. However, the configuration for the third individual has not only a different shape from the other two but also a different orientation, so even after any scaling that might be used to adjust for size differences, this individual perceives the relationships

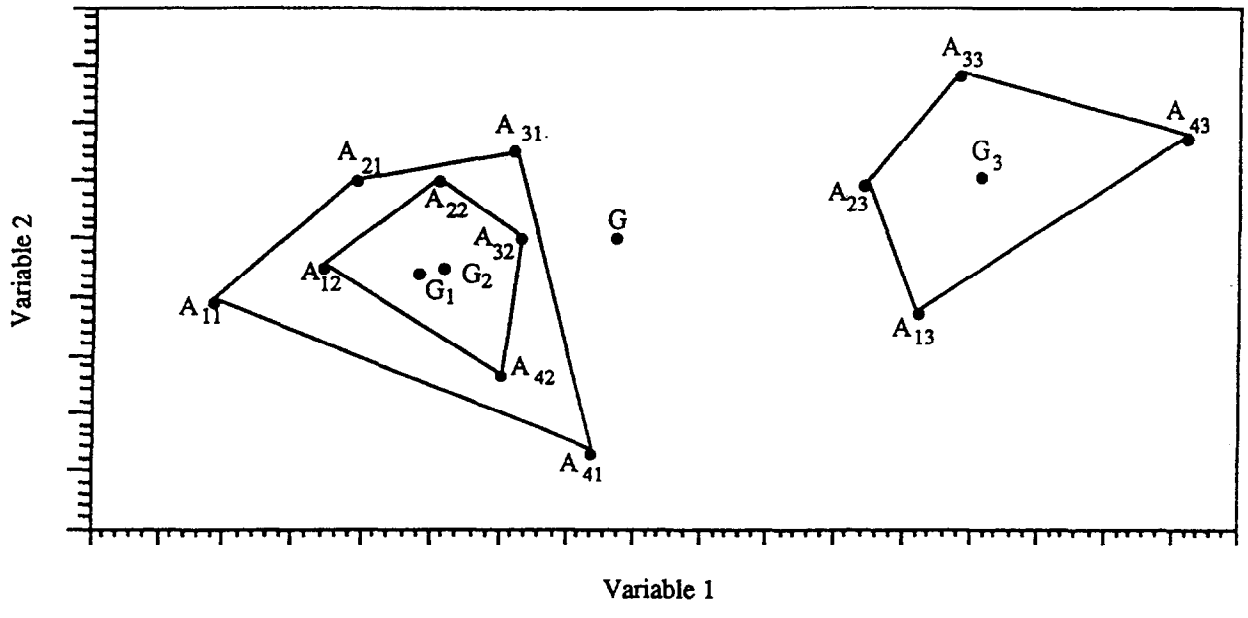


Fig. 2. A_{ik} is the point representing object i for individual k. The configurations for three individuals, with centroids G_1, G_2 and G_3, are shown as quadrilaterals. G is the overall centroid. The scaled axes are meant to indicate that the two dimensions of the diagram refer to substantive measured variables.

between the objects in a different way from the other two individuals. Furthermore, the centroid G_3 of this configuration is remote from the other two centroids, so this individual is using a different range of scores. The purpose of GPA is to formalise statements of these kinds.

2.3 The basic analysis of variance

Placing the origin, for convenience, at G, the centroid of the whole set, and denoting the distance between points A and B by $\overline{AB}$, Huygens's formula (2) may be used for the points of the kth configuration to give

$$\sum_{i=1}^{N} \overline{GA}_{ik}^2 = \sum_{i=1}^{N} \overline{G_kA}_{ik}^2 + N\overline{GG}_k^2$$

Summing over all K configurations to get the total sum-of-squares gives

$$\sum_{k=1}^{K}\sum_{i=1}^{N} \overline{GA}_{ik}^2 = \sum_{k=1}^{K}\sum_{i=1}^{N} \overline{G_kA}_{ik}^2 + N\sum_{i=1}^{N} \overline{GG}_k^2 \qquad (3)$$

The term on the extreme right-hand side of eqn (3) is N times the sum-of-squares of the differences between the centroids. By eqn (1), it may be written alternatively as

$$K\sum_{k=1}^{K} \overline{GG}_k^2 = \sum_{h<k=1}^{K} \overline{G_hG_k}^2$$

showing clearly how pairs of coincident centroids like G_h and G_k contribute zero terms to the total. This term gives the sum-of-squares attributable to differences between the means of Individuals. The contribution on the left-hand side of eqn (3) is the Total sum-of-squares and the remaining contribution, which represents the sum-of-squares of the Objects within Individuals (i.e. object scores about the individual means) may be obtained by subtraction and the whole expressed as the Analysis of Variance, given in tabular form in Table 1.

As explained above, statements about the relative positions of the centroids of the configurations for the individuals make sense only when the two dimensions used in the representation refer to the same substantive variables. When the matrices have arisen from multidimensional scaling or from FCP, the positions of the centroids are arbitrary. This merely acknowledges, for example, that it is not valid to compare one individual's rating of odour with another's rating of texture. In such cases, it remains valid to superimpose configurations, similar to those shown in Fig. 2, to have a common centroid C (say), as shown in Fig. 3, where what were G, G_1, G_2 and G_3 all now coincides at C.

TABLE 1. Analysis of Variance Separating the Translation Term M from the Total Sum-of-Squares T, the Difference Being Attributable to the Variance of the Object Scores about Their Mean Values for Each Individual

Source of variation	Sums of squares
Individuals	$N\sum_{k=1}^{K} \overline{GG}_k^2 = M$ (say)
Objects within individuals	By subtraction = $T - M$
Total	$\sum_{k=1}^{K}\sum_{i=1}^{N} \overline{GA}_{ik}^2 = T$ (say)

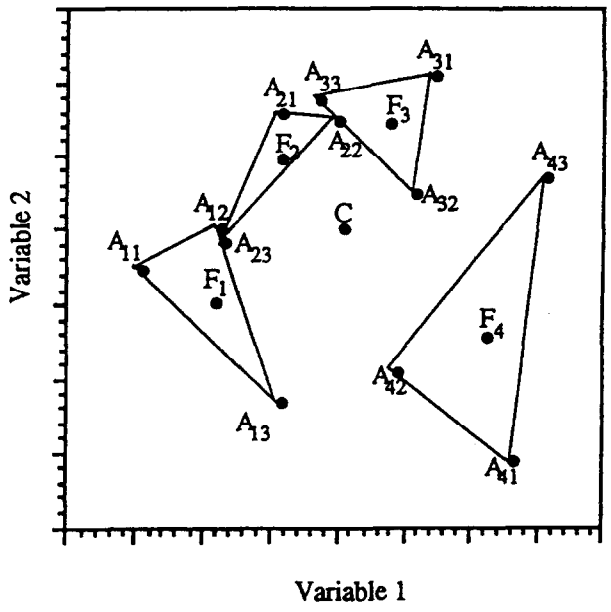


Fig. 3. The configurations for the individuals of Fig. 2 translated to a common centroid C, without any change of orientation. The triangles link the three points representing the same objects as scored by the three individuals with centroids F_1, F_2, F_3 and F_4. Because of the translation, the axes may be used only for within individual comparisons, for which distances between pairs of points, measured in either direction, have meaning.

In Fig. 3, the configurations for the individuals are shown with the same orientations as in Fig. 2. This is acceptable for the present, but it should be borne in mind that the orientations too are arbitrary, except when the axes refer to substantive variables. Shortly, we exploit this indeterminanacy of orientation. For the present, we note that the sum-of-squares about C in Fig. 3 is precisely the term $T - M$ obtained by subtraction in Table 1. Just as the *individuals* of Fig. 2 each have their own centroid, so do the points labelling the *objects* of Fig. 3, where they have been labelled F_1, F_2, F_3 and F_4. Again, we may use Huygens's formula (eqn (2)) to partition the sum-of-squares about C of Fig. 3 to give

$$\sum_{i=1}^{N}\sum_{k=1}^{K} \overline{CA}_{ik}^2 = K\sum_{i=1}^{N} \overline{F_iC}^2 + \sum_{i=1}^{N}\sum_{k=1}^{K} \overline{F_iA}_{ik}^2$$

or

$$T - M = S + (T - M - S)$$

where the term S represents the sum-of-squares for objects. Because all the centroids G_k now coincide in the single point C (see Fig. 3), the left-hand side of this expression represents Objects within Individuals, as before. The values of the two components on the right-hand

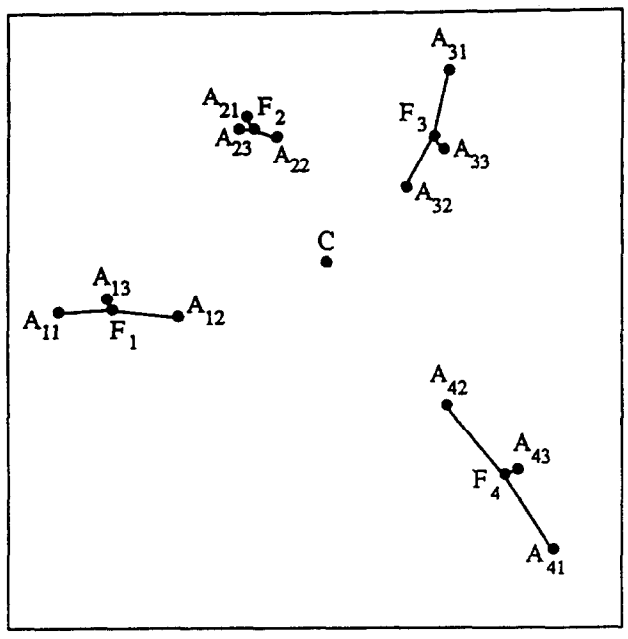
TABLE 2. Analysis of Variance Separating the Translation Term M from the Total Sum-of-Squares T, and Terms Representing Object-Means S and Individual Deviations within Objects, $T - M - S$

Source of variation	Sums of squares
Individuals	$N \sum_{k=1}^{K} \overline{GG}_k^2 = M$
Orientations:	
Object group average	$K \sum_{i=1}^{N} \overline{F_i C}^2 = S$ (say)
Individual deviations	By subtraction = $T - M - S$
Total	$\sum_{k=1}^{K} \sum_{i=1}^{N} \overline{GA}_{ik}^2 = T$

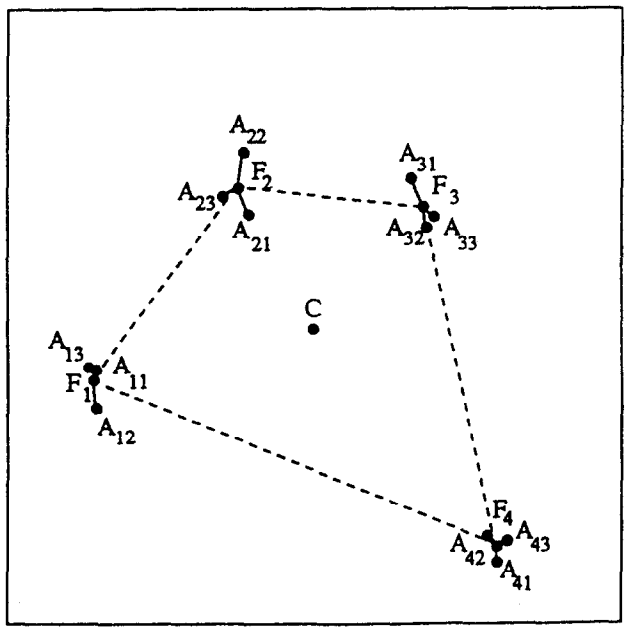
side depend on the orientations and, hence, this decomposition is collectively labelled Orientations. The terms of the decomposition attributable to the sum-of-squares of the rotated object means, i.e. the Group Average [2.5] with the remaining term $T - M - S$ (Individual Deviations), arise from the deviations of the rotated individual scores from the rotated average object scores. Thus the Analysis of Variance of Table 1 can now be refined to give the Analysis of Variance exhibited in Table 2.

It has already been noted that the total sum-of-squares of Fig. 3 does not depend on the orientation of the configurations for the individuals. *The Analysis of Variance of Table 2 is valid for any set of orientations. Although the value of the contribution S varies with orientation, the final two lines of the body of the table (Orientations) will maintain a constant sum* $T - M$. The preceding statement has been set in italics, for it becomes of central importance when we discuss alternative, but related, types of analysis ([A.4], [A.5]). It is an obvious question to ask whether there might be some orientation, or orientations, that are more advantageous than others. Generalized Procrustes Analysis chooses to find the orientation that minimises the sum-of-squares T − M − S for Individual Deviations; that is, GPA maximises the agreement between individuals on each object. Because T − M is a constant for all orientations, minimising T − M − S is the same as maximising both S and the ratio S/(T − M − S); this remark will be useful when we come to compare GPA with other methods ([A.4.1], [A.4.4]). Figure 4(a) shows the result of GPA when applied to Fig. 3; the quantities whose sum-of-squares are minimised are shown as solid lines and in some contexts may be referred to as *residuals* and their sum-of-squares as the residual sum-of-square.

Every term in the Analysis of Variance is formed from summing the squares of separate distances like $\overline{GG}_k$ and $\overline{CF}_i$. These are worth tabulating and examining to see if certain components contribute excessively to the total variation. This policy is especially valuable for the term $\sum_{i=1}^{N} \sum_{k=1}^{K} \overline{F_i A}_{ik}^2$ obtained by subtraction in Table 2, which



(a)



(b)

Fig. 4. In 4(a) the configurations have been rotated to best fit by minimising the sum-of-squares of the marked residuals joining the object points with their respective centroids F_1, F_2, F_3 and F_4. Because of the arbitrary orientations, the axes have no meaning and are not presented. In 4(b) the three individual configurations have also been scaled to reduce even further the residual sum-of-squares. The GPA group average configuration is indicated by the dotted lines.

has $N \times K$ components, such as the 4×3 components highlighted in Fig. 4. The quantities $\overline{F_i A}_{ik}^2$ may be tabulated as a two-way table whose column and row marginal totals give the contributions to the residual sum-of-squares $T - M - S$ attributable to the separate individuals and separate objects respectively [3.2].

2.4 Optimal isotropic scaling

With the example of Figs 2 and 4 it is clear that the substantive variables (rather than merely dimensions) are used differentially by the individuals, so that it is appropriate to estimate a factor ρ_k to scale $\mathbf{X}_k$ ($k = 1, 2, \ldots, K$) isotropically, i.e. all directions are treated equally. To avoid the unhelpful result that a perfect agreement can always be obtained by scaling every $\mathbf{X}_k$ to have zero size, the estimation of optimal factors ρ_k is adjusted so that the sum-of-squares, $T - M$, of the configurations about C remains the same before and after scaling. In this sense the total size of all configurations remains constant, but some configurations will be made bigger and others smaller. Estimating isotropic scaling factors for Fig. 4(a) gives the improved fit of Fig. 4(b). Comparing Figs 4(a) and (b) shows that the orientations that result from minimising $T - M - S$ depend on any isotropic scaling factors that might be associated with each individual's use of scales. Thus, optimal scaling compared with no scaling will give a smaller value of $T - M - S$ and, consequently, a larger value of S. This improvement is no valid reason for scaling automatically [A.2.1]. When size differences are not reasonably attributable to differential use of measurement scales but, say, are merely the effects of the arbitrary size scaling given to configurations produced by some multidimensional scaling computer program, it is sufficient to prescale [A.2] each $\mathbf{X}_k$ to have unit sum-of-squares about its centroid.

Whenever isotropic scaling is used, it will be assumed in the following that the configurations plotted, and the resulting analyses of variance, incorporate the optimal estimate of the scaling factors ρ_k (Gower, 1975; ten Berge, 1977). This avoids having to make unhelpful distinctions between scaled and unscaled forms of analysis; it is understood that the estimated values $\rho_1, \rho_2, \ldots, \rho_k$ will be recorded and used in any interpretation. When each $\mathbf{X}_k$ is initially scaled to have unit sum-of-squares, the term S of Table 2 may be written $S = \rho_1^2 + \rho_2^2 + \ldots, \rho_k^2$ (see Gower, 1975).

2.5 The group average

The configuration of the centroids $F_1, F_2, \ldots, F_N$, associated with the objects provides a valuable summary of how the individuals view the relationships between the objects on the average. This was termed the *consensus configuration* by Gower (1975), but this terminology is somewhat unfortunate because the configuration certainly does not represent a consensus of views of the individual ([A.4.4] (see also Lingoes & Borg, 1978, p. 459). Rather, it represents an average that may conceal a wide range of differences of viewpoint. The term *group average* used in individual scaling [A.4.5] is better, but its unqualified adoption would be confusing. Perhaps all methods that provide an average configuration of this kind should adopt the term 'group average' but precede it by the name of the method used. Thus we could have *GPA group average, STATIS group average* (Lavit, 1988) and even *INDSCAL group average*, though INDSCAL [A.4.5] might be omitted in acknowledgement of priority for the term in that method, rather as the name of the country is omitted from British postage stamps. Alternative terminologies refer to the *common space* of the analysis or to the *centroid configuration*. The GPA group average is shown on Fig. 4(b) and consists of the points F_i ($i = 1, \ldots, 4$). It has greater sum-of-squares (and therefore by eqn (1), it is bigger in size) than all group averages given by other sets of orientations not obtained by GPA.

2.6 The exhibited and unexhibited parts of the analysis

So far everything has been described in terms of two dimensions. It should be clear that however many dimensions there may be, everything follows through, except for the problem of exhibiting high dimensional configurations in the confines of two-dimensional paper. Gower (1975) suggested using Principal Component Analysis of the GPA group average configuration to give a low-dimensional (hopefully two-dimensional) approximation to that configuration and then representing everything relative to these principal axes. Also, this gives a convenient unique final orientation for all the configurations which does no violence, because the orientation of the combined configuration (the multidimensional generalization of Fig. 4(b) is otherwise arbitrary. Thus, the component analysis merely chooses an optimal joint orientation that gives a best configuration for display, without changing any of the inter-point distances given by the GPA. With component analysis, as with other ways for handling problems of dimensionality to be discussed shortly [A.4], the Analysis of Variance may be divided into two independent parts, one representing the part of the total configuration that is exhibited, in the R (say) most effective dimensions, and the remaining unexhibited part, in $P - R$ dimensions; this is shown in Table 3, where we have now adopted a shortened terminology for the sources of variation. As is usual in

TABLE 3. Analysis of Variance as in Table 2 but Partitioning Every Sum-of-Squares into a Part that is Exhibited in R Dimensions and an Independent Non-Exhibited part in $P - R$ Dimensions

Source of variation	Sums-of-squares		
	Exhibited (R dimensions)	**Unexhibited ($P - R$ dimensions)**	**Total (P dimensions)**
Individuals	M_R	$M_{(P-R)}$	M
Orientations:			
Group average	S_R	$S_{(P-R)}$	S
Deviations	$T_R - M_R - S_R$	$T_{(P-R)} - M_{(P-R)} - S_{(P-R)}$	$T - M - S$
Total	T_R	$T_{(P-R)}$	T

a component analysis, 'the proportion of variance accounted for' in R dimensions is S_R/S.

This particular breakdown into exhibited and unexhibited dimensions is designed to give the best representation in R dimensions of the GPA group average. In mathematical terms, this gives a particular projection [A.4] of the P-dimensional space onto R dimensions. There are other possibilities. For example, if one wishes to be even-handed and get a best representation of all the data, then all $N \times K$ P-dimensional coordinates (after orientations and any scaling) should be used for the component analysis [3.2]. Other projections are discussed in [A.4].

2.7 Relating the properties to the reoriented configurations

An important aid to interpretation is to see how the coordinate axes for representing the properties appear in the final reoriented configurations. These axes will have been rotated, scaled and projected in precisely the same way as the data. It follows that the axes will appear as straight lines through the centroid but, because of the projections, they will no longer be at right angles as they were originally. Further, unit length on some axes will project into much smaller lengths than on other axes; such foreshortened axes contribute little to the display in the chosen projection, normally the exhibited space, and should be ignored. To help identify insignificant axes, unit lengths of the original scales of measurement should be indicated on the projected axes [A.5]. Each individual has his or her own axes in FCP and one of the advantages of displaying axes is to get a feel for how descriptions of properties given by the different individuals may be related.

The above is a valid description when the $\mathbf{X}_k$ relate to the original data matrices, perhaps with some prescaling and centering. However, with the more elaborate transformations of multidimensional scaling type, the situation is more complicated. Nevertheless, axes may still be displayed, although they will now be non-linear. The methodology required is beyond the scope of this article (see Gower & Harding, 1988). An example which includes both qualitative and quantitative axes is given by Gower and Dijksterhuis (1992).

3 EXAMPLES

Throughout this section, there is a considerable amount of forward referencing, including analysis of data using methods related to GPA that are explained in the Appendix. The data discussed in this section are of Type I [1] and stem from an experiment in which seven assessors tasted eight different kinds of yoghurt; thus $K = 7$ and $N = 8$. These data are a subset of the data used by Dijksterhuis and Punter (1990). The experiment used Free Choice Profiling in which the assessors themselves were allowed to choose the variables on which to rate the yoghurts. As a result, the seven individual data matrices $\mathbf{X}_k$ ($k = 1, \ldots, 7$) (Table 4) have different numbers of columns—in this case, ranging from four to nine—but all have eight rows.

TABLE 4. The Raw Data Used in the Analysis in this Paper, $K = 7$ Individual Sets ($k = 1, \ldots, 7$), $N = 8$ Products ($i = 1, \ldots, 8$), P_k attributes ($4 \leq P_k \leq 9$)

k	i									k	i								
1	1	21	84	26	11	24				5	1	69	17	64	83				
	2	9	73	75	9	8					2	9	88	27	19				
	3	67	84	40	9	32					3	91	11	77	14				
	4	41	75	55	9	24					4	71	24	87	32				
	5	10	67	31	9	77					5	9	58	15	87				
	6	10	67	84	8	16					6	11	87	13	16				
	7	71	89	14	54	47					7	87	16	66	76				
	8	8	86	11	20	37					8	13	80	75	50				
2	1	37	15	82	65	59	22	32	34	6	1	71	16	26	9	45	87		
	2	22	19	14	24	15	16	25	37		2	88	33	72	10	12	66		
	3	79	58	65	32	46	22	25	40		3	49	21	69	84	54	71		
	4	70	85	42	73	86	25	57	56		4	24	77	62	10	17	52		
	5	22	14	30	20	50	27	42	32		5	86	10	76	77	78	67		
	6	9	10	14	24	21	33	22	47		6	88	20	81	13	15	49		
	7	91	54	90	69	40	27	78	76		7	34	74	63	13	84	20		
	8	15	10	33	62	36	25	25	58		8	87	10	66	89	63	73		
3	1	71	70	75	28	42	34	12	90	43	7	1	42	16	68	59	0	45	28	8
	2	90	90	8	7	90	74	89	27	8		2	12	7	33	37	76	16	63	15
	3	34	24	88	27	64	52	91	41	13		3	11	16	57	42	25	16	12	8
	4	38	31	61	80	30	18	0	46	19		4	70	19	15	24	25	62	40	20
	5	91	95	70	16	75	60	56	16	29		5	10	7	17	50	33	51	57	90
	6	98	91	42	60	47	82	73	23	8		6	80	9	44	33	27	23	43	30
	7	19	9	91	80	22	53	8	75	52		7	32	40	65	51	20	64	32	28
	8	89	81	86	8	9	85	11	9	74		8	61	9	14	46	71	72	57	75
4	1	56	81	20	21	20														
	2	10	11	46	26	81														
	3	88	82	21	20	26														
	4	59	49	60	28	75														
	5	9	83	49	40	49														
	6	9	19	22	23	74														
	7	75	47	21	21	22														
	8	16	80	48	17	29														

Table 5 presents the means of the attributes for each individual. From Table 5 it can be seen that there are large differences in mean scores between attributes. Because the data in this example is of FCP type, the means cannot be compared over individuals in terms of the attributes [A.1.2]. Thus, in Table 5, the first value of the first set (29·6) cannot be compared with the first value of the second set (43·1) because they refer to different attributes. The blanks in Table 5 result from the sets having differing numbers of attributes. $\mathbf{X}_3$ has nine columns and $\mathbf{X}_2$ and $\mathbf{X}_7$ each have eight columns, so we have a situation discussed in [A.3.1] of excess dimensionality. These matrices may be reduced to seven columns by performing a PCA on the raw data in the sets. This initial PCA preserves distances and so does not result in any loss of information [A.3.1]. Theoretically

TABLE 5. Mean Attribute Scores (Translation Terms)

Individual									
1	29·6	78·1	42·0	16·1	33·1				
2	43·1	33·1	46·3	46·1	44·1	24·6	38·3	47·5	
3	66·3	61·4	65·1	38·3	47·4	57·3	42·5	40·9	30·8
4	40·3	56·5	35·9	24·5	47·0				
5	45·0	47·6	53·0	47·1					
6	65·9	32·6	64·4	38·1	46·0	60·6			
7	39·8	15·4	39·1	42·8	34·6	43·6	41·3	34·3	

these initial PCAs are unnecessary but they may help with computer implementations and they ensure the size of the computed rotation matrices is minimal, so improving computer efficiency. Sets with less than seven columns are padded with additional zero columns in order to make all the matrices of the same size. Thus we arrive at $P = 7$.

3.1 Analysis

The same data set has been analysed using three different methods, corresponding to a 'typical' GPA analysis [2], and two 'projecting' analyses, Peay's (1988) [A.4.1] and Green and Gower's (1979) [A.4.4]. ten Berge and Knol (1984) give details of how to generalize the Green and Gower method to include more than two sets. Because FCP has been used, translation has no meaning, so this term is omitted from subsequent analyses of variance and the centroids of all seven configurations are immediately translated to a common origin. In doing this, it is recognized implicitly that the only potentially useful information in this kind of data resides in distances between pairs of objects. Because the data consist of a set of limited scale values [A.2] obtained from an FCP experiment, so called P_k-scaling [A.2.1, Table A1] has been applied to the data, except with the Green and Gower analysis [A.4.4].

To keep the numbers to a manageable size the total sum-of-squares is scaled to 100. This also results in the Analysis of Variance tables containing percentages of the total variance (see also Dijksterhuis & Punter, 1990). This scaling has no influence on the solution but it is convenient to be able to work in terms of percentages. All analyses are carried out with an optimal isotropic scaling step [2.4].

There is insufficient space to give exhaustive analyses, so many of the possibilities described in the main text are not explored in these examples. Only summary information is given for individuals (e.g. Table 7). For example, we illustrate only group averages and not configurations giving detailed information on individuals, such as in Fig. 4. For variables comparable across individuals [A.2.1], information on variables could be given concisely with the group-average configuration, but with FCP data either separate diagrams would be needed for each individual or all $K \sum_{k=1}^{K} P_k$ variables could be represented with the group average. We recommend this but have insufficient space to illustrate it here. The GPA [3.2] is more complete than the other analyses; for example, tables corresponding to Table 7 are not confined to GPA.

3.2 Classical GPA

The first analysis to be presented is a classical Generalized Procrustes Analysis, as described in [2] and in more detail in Gower (1975). The criterion minimised is $T - M - S$, which measures the squared distances between the corresponding points of the $\mathbf{X}_k$ (see Fig. 4 and Table 3); its minimisation is the same as maximising S. The analysis of the yoghurt data-set gave $T - M - S = 22{\cdot}3\%$ and a group-average sum-of-squares $S = 77{\cdot}7\%$. S expresses the variance lost in the Procrustes transformations of rotating and scaling. Table 6 presents the Analysis of Variance for this and the other two analyses to be described below. As described in [A.2.1], each set has been initially scaled to have a constant sum of squares so that estimates of the isotropic scaling factors ρ_k are relative to this initialisation.

The method defined as Generalized Procrustes

TABLE 6. Analysis of Variance for the Three Analyses, Partitioned in an Exhibited Part in $R = 2$ Dimensions and an Unexhibited Part in $P - R = 5$ Dimensions[a]

Source	Individual deviations			Group average			Total		
	R	$P-R$	P	R	$P-R$	P	R	$P-R$	P
GPA[1]	12·8	9·5	22·3	61·4	16·3	77·7	74·2	25·8	100
GPA[2]	13·1	9·2	22·3	61·3	16·4	77·7	74·4	25·6	100
Green and Gower	0·8	30·4	31·2	6·8	62·0	68·8	7·6	92·4	100
Peay	15·9	11·0	26·9	63·0	10·1	73·1	78·9	21·1	100
Terms in Table 3:	$T_R - M_R - S_R$	$T_{(P-R)} - M_{(P-R)} - S_{(P-R)}$	$T - M - S$	S_R	$S_{(P-R)}$	S	T_R	$T_{(P-R)}$	T

[a] GPA is presented in two forms labelled GPA[1] and GPA[2]. GPA[1] and GPA[2] are essentially the same analysis but a different space is found for exhibiting the display. In GPA[1] the exhibited R-dimensional space is relative to the principal components of the $N \times P$ matrix of group-average coordinates. In GPA[2] the exhibited R-dimensional space is relative to the principal components of the $(N \times K) \times P$ matrix of all the combinations of individuals with objects [2.6]. The values in the table can be interpreted as percentages since the totals are 100.

Analysis in [2] performs its transformations (rotation and scaling) in the full-dimensional space—in this case in seven dimensions, since this is the maximum dimensionality of the sets. The analysis results in one high-dimensional group average configuration of eight objects in seven dimensions. As described in [2.6] a PCA may be used to represent the group average relative to its principal components. The seven principal components explain, respectively, 55, 24, 7·4, 6·1, 4·2, 2·2 and 1·1% of variance of the group average. As was shown in Table 3, the Analysis of Variance can be divided into two parts, one exhibited R-dimensional part and an unexhibited $P - R$-dimensional part. Thus, the row labelled GPA[1] in Table 6 shows that the group average accounts for 77·7% of the total variance and with $R = 2$, $S_R = 61{\cdot}4\%$ in the exhibited space of the group average with 16·3% left in the unexhibited space.

Table 7 shows the individual deviations partitioned over objects and over individuals. The numbers in the table are the squared distances $\overline{F_i A}_{ik}$ from [2.3] and Fig. 4. The grand total is equal to $T - M - S = 22{\cdot}3$. The row margins give the residuals for the objects. In this case, the objects numbered 1, 3 and 4 have the largest residuals, which means that they do not fit as well as the other objects; the clusters around F_1, F_3 and F_4 will be looser than the other clusters. These residuals pertain to the full seven-dimensional configurations [3] and will not necessarily be visible in the low-dimensional exhibited space, such as Fig. 5. Tables like Table 7 can help one avoid over-interpretation when the exhibited configuration is not a very good approximation. It can then be useful to partition the table into exhibited and unexhibited residuals, but we have not done this here. The column margins give the residual sum-of-squares for the individuals. It can be seen that individuals 2 and 3 have the best fitting configurations, since their residual sum-of-squares is among the lowest; these individuals are close to the group average. Individuals 6 and 7 are furthest from the group average. Indeed, the intersection of the worst object (3) and with worst individual (6) pinpoints the worst residual (1·430).

TABLE 7. Individual Squared Deviations partitioned over Objects and Over Individuals (GPA)

Objects	Individuals							Sum
	1	2	3	4	5	6	7	
1	0·539	0·237	0·181	0·412	0·560	0·735	0·594	3·26
2	0·228	0·150	0·390	0·434	0·212	0·417	0·611	2·44
3	0·117	0·086	0·480	0·931	0·735	1·430	0·379	4·16
4	0·706	0·687	0·365	0·489	0·349	0·427	0·485	3·51
5	0·330	0·353	0·337	0·129	0·333	0·278	0·153	1·91
6	0·309	0·462	0·101	0·200	0·288	0·123	1·102	2·58
7	0·336	0·224	0·162	0·266	0·290	0·273	0·328	1·88
8	0·391	0·353	0·297	0·136	0·393	0·595	0·390	2·55
Sum	2·95	2·55	2·31	3·00	3·16	4·28	4·04	22·29

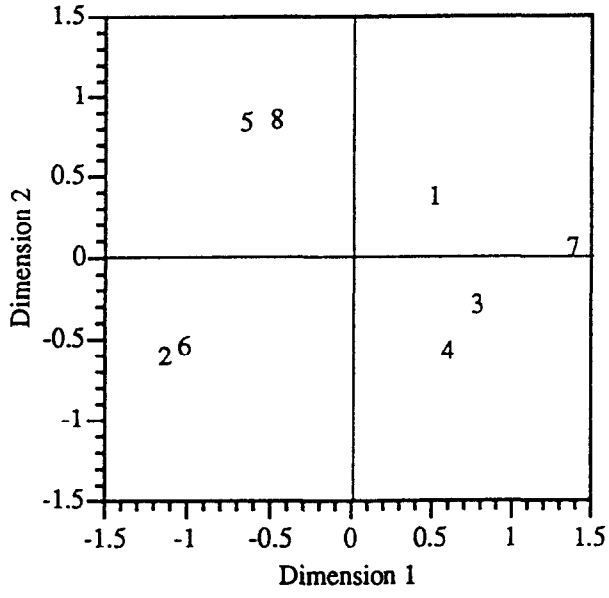


Fig. 5. First two principal components of a PCA of the GPA group average, showing the eight kinds of yoghurt.

The first two principal components of the group average give the two-dimensional approximation exhibited in Fig. 5. From the figure it is clear that there are two pairs of yoghurts (2 and 6, 5 and 8), each of which is judged similar by the panel. In fact, yoghurts 5 and 8 are the same products, blindly presented twice to the panel. Numbers 2 and 6 are, respectively, light (low fat) and normal variants of the same brand of yoghurt. It seems as if the panel were unable to discriminate between the two variants.

Another result of a typical GPA is the set of isotropic scaling factors used to shrink or stretch the sets during the matching process. A configuration $\mathbf{X}_k$ is multiplied by a factor ρ_k that represents shrinking when $0 < \rho_k < 1$ and stretching when $1 < \rho_k$. Table 8 presents these factors for all three analyses performed This table shows that the data of the fifth individual needed shrinking by a factor 0·79, and those of the first individual stretching by a factor 1·26, implying unusual use of the scale ranges. The other isotropic scaling factors do not differ much from unity.

To illustrate a different partition of the exhibited and unexhibited spaces we have also chosen the first

TABLE 8. Scaling Factors ρ_k for the Three Methods (GPA and Peay with P_k scaling, Green and Gower without P_k scaling [A.4.4])

Individual	GPA	Green and Gower	Peay
1	1·26	0·84	1·14
2	1·17	1·14	1·24
3	0·92	0·98	1·01
4	1·09	0·99	1·08
5	0·79	0·55	0·78
6	0·90	1·49	0·79
7	1·11	1·18	1·24

two principal components of the $N \times K$ P-dimensional coordinates to define the exhibited space [2.6]. This gives the analysis labelled GPA^2 in Table 6 in which the P-dimensional sum-of-squares are as before, because it is only the partition into exhibited and unexhibited components that differ. The difference between the two versions of exhibiting the GPA is trivial, and is mainly to be seen in minor changes in the contributions from Individual Deviations; the group average is hardly affected. This variant gives a diagram (not shown) very similar indeed to that of Fig. 5. When outliers are present, the GPA^1 and GPA^2 analyses will differ more and the group average of GPA^2 could then show a much poorer representation in the exhibited space.

3.3 Method of Green and Gower [A.4.4]

The Analysis of Variance associated with this analysis is also presented in Table 6. A value of $R = 2$ was chosen for the exhibited space. The criterion minimised is (see Table 3) $T_R - M_R - S_R = 0{\cdot}8\%$, only 6·8% is captured in the two-dimensional group average. This phenomenon is almost certainly due to the fact that the high dimensions are projected to cast 'shadows' which are of similar shape [A.4.4], but this happens irrespective of the shadows' sizes. It seems as if, at least with this data-set, rather small shadows are closely matched, suggesting that there is little 'consensus' between the individuals.

The unexhibited $(P - R)$-dimensional space contains parts of the individual configurations with arbitrary orientations [A.4.3]. These configurations were analysed using GPA and the resulting sum-of-squares for the individual deviations and the group average are also shown in Table 6.

Figure 6 shows the two-dimensional group average that results from the Green and Gower analysis of the yoghurt data. The resemblance between the pair of yoghurts (5, 8) which was clear from Fig. 5 is absent in Fig. 6; the other similar pair (2, 6) is visible in both configurations. Because the representation of Fig. 6 does not capture much variance (just 6·8%, see Table 6) it is not surprising that this configuration does not match the GPA one (and the Peay one described in [3.4]) very much. The scaling factors (Table 8) for individuals 1 and 5 are smaller than those given by the other two methods. The data of individual 6 were stretched with this analysis and shrunk with the other two analyses; contrariwise, for individual 1. The scaling factors for individuals 3 and 4 approximate unity, as with the other analyses. However, the use of P_k-scaling [A.2.1] complicates the direct comparison of this method with the other two.

3.4 Method of Peay [A.4.1]

Table 6 also contains the result of Peay's method of analysis of the yoghurt data. This method maximises S_R (Table 3), the sum-of-squares contained in the exhibited part of the group average. While maximising the percentage captured in the two-dimensional group average ($S_R = 63{\cdot}0\%$), this method shows the greatest deviations between the individuals in the two-dimensional space ($T_R - M_R - S_R = 15{\cdot}9\%$).

The group average given by Peay's method (Fig. 7) is very similar to that given by GPA (Fig. 5). This is to be expected because the group average sum-of-squares in both the P-dimensional spaces and the R-dimensional exhibited spaces are quite close. In this case at least, it matters little whether the $(P - R)$-dimensional space

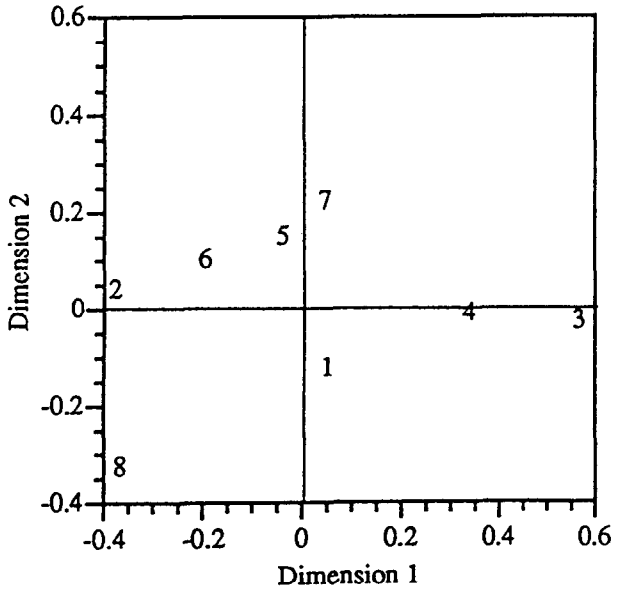


Fig. 6. Group average of the Green and Gower analysis showing the eight kinds of yoghurt.

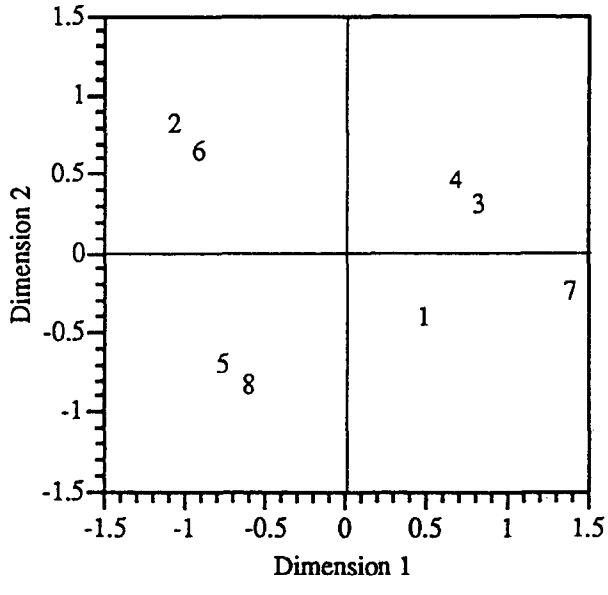


Fig. 7. Group average of the Peay analysis showing the eight kinds of yoghurt.

given by Peay (10·1%) represents pure noise, pure signal or a mixture of both [A.4.2]; other cases may show larger differences.

The Peay group-average sum-of-squares in the exhibited space (63·0%) has been maximised and is, as it must be, greater than the corresponding GPA value (61·4%). However, in the full P-dimensions the reverse is true [A.4.1] (73·1% and 77·7%), even though S_{P-R} has been maximised, conditional on the maximisation of S_R [A.4.3]. Further the P-dimensional signal-to-noise ratios are 3·48 (GPA), 2·72 (Peay) and 2·21 (Green and Gower); recall that this ratio is maximised by GPA [2.3]. It is worth noting that the signal-to-noise ratios in the exhibited space are 4·80 (GPA), 3·69 (Peay) and 8·50 (Green and Gower), but we have not attempted to maximise this ratio to get its optimal value [A.4.2]. Green and Gower does best but note that in that case both the exhibited signal and exhibited noise are small and one cannot expect that the unexhibited parts represent only noise, underlining the importance of deciding what is signal and what is noise.

Figure 7 shows the group average of the Peay analysis. Apart from a reflection in the first dimension, this configuration is very much like that of GPA. The scaling factors (Table 8) have the same pattern as those from GPA. For this data-set, it seems there is not much difference between the solutions of these two methods.

4 CONCLUSIONS

We shall summarise the above discussion in the form of a checklist.

(1) Decide whether data is of Type I or II [1] and, if the former, whether it is in the form of FCP or whether the variables are comparable across individuals [A.1.2]. Decide which, if any, transformations may be needed or be worth exploring [A.1.1]. As part of this process decide if some form of initial scaling is desirable [A.2.1]. Do any preliminary multidimensional scaling or component analyses [A.1.1] perhaps to separate what appears to be signal from what appears to be noise [A.4.2].

(2) Has initial scaling taken care of size differences or is it sensible to fit isotropic scaling factors? Do not fit isotropic scaling merely to improve the fit ([2.4], [A.2.1]).

(3) Eliminate individual means. Do these means contain useful information? If there is a cross-classification, decide whether the analysis is to be focused on individuals or on objects; remove the appropriate means [A.1.2]. Consider a MANOVA.

(4) Decide on what aspects of the data you wish to explore. Recall that GPA gives the best overall group-average for noiseless data [2.3]; Peay gives a method for separating signal from noise [A.4.2], but this is perhaps better achieved by an initial multidimensional scaling or component analysis [A.4.2]; and that Green and Gower's method gives a method for detecting consensus or agreement [A.4.4].

(5) Decide on the dimensionality R of the signal space and the dimensionality R_1 of the exhibited space. Several different values of R may need to be examined. Plot the group-average space together with information on individual differences [2.5] and on the variables [2.7]. Try to interpret this analysis in the light of what you know about the objects, the individuals, the variables and the experimental procedures that generated the data. Make proper allowance for the degree of approximation with respect to the full R-dimensional signal space [A.4].

(6) Record the Analysis of Variance [2]. Record any scaling factors that may have been estimated [2.4]. Examine the contributions to the residual sum-of-squares, breaking this down by both individuals and by objects [2.3]. Record the numerical values of the individual rotated/projected configurations and of the group average configuration.

(7) Decide whether it is worth analysing the part of the data that is in $P - R$ dimensions [A.4.3]. If so, return to (4).

(8) If you must, consider possible reification of directions (not necessarily orthogonal) in the group-average space. The original data and the rotation matrices [A.5] might be helpful in this process—also, correlations between variables and fitted dimensions.

If there are any conclusions to be drawn from this discussion, they are that data analysis is not a straightforward matter of using the right computer package and that, even in the restricted area under discussion, one cannot recommend a single best method. We hope that as a result of this article, researchers will be in a position to appreciate what questions the methods might help answer and hence do better and more relevant analyses.

We conclude with a quote from Hurley and Cattell (1962) who coined the term Procrustes Analysis. They write: 'To publish widely a program which permits any tyro, by pressing a computer button, to seem to verify any theory, is as irresponsible as loosing opium on the open market. That computers and their programs can be a real danger to proper values and directions in research must already be evident in several fields.' Thirty years later things have got much worse—computers are now commonplace and software for a confusing variety of methods is readily available.

REFERENCES

Arabie, P., Carroll, J. D. & De Sarbo, W. S. (1987). *Three-way Scaling and Clustering.* Sage University Papers on Quantitative Applications in the Social Sciences, No. 07–065. Sage Publications, Beverly Hills, California.

Arnold, G. M. & Williams, A. A. (1985). The use of generalized Procrustes analysis in sensory analysis. In *Statistical Procedures in Food Research,* ed. J. R. Piggot. North Holland, Amsterdam.

Carroll, J. D. & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalisation of 'Eckart–Young' decomposition. *Psychometrika,* **35**, 283–319.

Commandeur, J. J. F. (1991). *Matching Configurations.* DSWO Press, Leiden.

Constantine, A. C. & Gower, J. C. (1978). Graphical representation of asymmetry. *Applied Statist.,* **27**, 297–304.

Constantine, A. C. & Gower, J. C. (1982). Models for the analysis of inter-regional migration. *Environment and Planning 'A',* **14**, 477–97.

Davis, A. W. (1978). On the asymptotic distribution of Gower's m^2 goodness-of-fit criterion in a particular case. *Ann. Inst. Statist. Math.,* **30**, 71–9.

Dijksterhuis, G. B. & Punter, P. H. (1990). Interpreting generalized Procrustes analysis 'Analysis of Variance' tables. *J. Food Qual. Pref.,* **2**, 255–65.

Gifi, A. (1990). *Non-Linear Multivariate Analysis.* J. Wiley and Sons, Chichester, UK.

Gittins, R. (1984). *Canonical Analysis.* Springer Verlag, Berlin.

Gower, J. C. (1975). Generalized Procrustes analysis. *Psychometrika,* **40**, 33-51.

Gower, J. C. (1977). The analysis of asymmetry and orthogonality. In *Recent Developments in Statistics,* ed. J. Barra *et al.,* North Holland, Amsterdam, 109–23.

Gower, J. C. (1984). Multivariate analysis: ordination, multidimensional scaling and allied topics. In *Handbook of Applicable Mathematics: Vol. VI, Statistics,* ed. E. H. Lloyd. J. Wiley and Sons, Chichester, 727–81.

Gower, J. C. & Banfield, C. F. (1975). Goodness-of-fit criteria for hierarchical classification and their empirical distributions. In *Proceedings of the Eighth International Biometrics Conference,* ed. L. C. A. Corsten and T. Postelnicu. Editura Acadamiei Republicii Socialiste Romania, Bucharest, pp. 347–61.

Gower, J. C. & Dijksterhuis, G. B. (1992). Coffee images: a study in the simultaneous display of multivariate quantitative and categorical variables for several assessors. Research Report: RR-92-06, Department of Datatheory, University of Leiden.

Gower, J. C. & Harding, S. A. (1988). Nonlinear biplots. *Biometrika,* **75**, 445–55.

Gower, J. C. & Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *J. Classification,* **3**, 5–48.

Green, B. F. & Gower, J. C. (1979). A problem with congruence. Paper presented at the Annual Meeting of the Psychometric Society, Monterey, California.

Hand, D. J. & Taylor, C. C. (1987). *Multivariate Analysis of Variance and Repeated Measures.* Chapman and Hall, London.

Healy, M. J. R. (1978). Is statistics a science? *J. Roy. Statist. Soc. A,* **141**, 385–93.

Hurley, F. R. & Cattell, R. B. (1962). The Procrustes program: producing direct rotation to test a hypothesized factor structure. *Behavioural Sci.,* **7**, 258–62.

Kruskal, J. B. & Wish, M. (1978). *Multidimensional Scaling.* Sage University Papers on Quantitative Applications in the Social Sciences, No. 07–011. Sage Publications, Beverly Hills, California.

Langeheine, R. (1982). Statistical evaluation of measures of fit in the Lingoes–Borg Procrustean individual differences scaling. *Psychometrika,* **47**, 427–42.

Langron, S. P. & Collins, A. J. (1985). Perturbation theory for generalized Procrustes analysis. *J. Roy. Statist. Soc. B,* **47**, 277–84.

Lavit, C. (1988). *Analyse Conjointe de Tableaux Quantitatifs.* Masson, Paris.

Lingoes, J. C. & Borg, I. (1978). A direct approach to individual differences scaling using increasingly complex transformations. *Psychometrika,* **43**, 491–519.

Peay, E. R. (1988). Multidimensional rotation and scaling of configurations to optimal agreement. *Psychometrika,* **53**, 199–208.

Sibson, R. (1978). Studies in the robustness of multidimensional scaling: Procrustes statistics. *J. Roy. Statist. Soc. B,* **40**, 234–8.

Siegel, A. F. & Benson, R. H. (1982). A robust comparison of biological shapes. *Biometrics,* **38**, 341–50.

ten Berge, J. M. F. (1977). Orthogonal Procrustes rotation for two or more matrices. *Psychometrika,* **42**, 267–76.

ten Berge, J. M. F. & Knol, D. L. (1984). Orthogonal rotations to maximal agreement for two or more matrices with different column orders. *Psychometrika,* **49**, 49–55.

van Buuren, S. & Dijksterhuis, G. B. (1988). Procrustes analysis of discrete data. In *The Many Faces of Multivariate Analysis, Proceedings of the SMABS-88 Conference,* Vol. 1, ed. M. G. H. Jansen and W. H. van Schuur. RION, Institute for Educational Research, Groningen, 53–66.

Verboon, P. & Heiser, W. J. (1991). Resistant orthogonal Procrustes Analysis. *J. Classification,* in press.

Williams, A. A. & Langron, S. P. (1984). The use of free choice profiling for the evaluation of commercial ports. *J. Sci. Food Agric.,* **35**, 558–68.

APPENDIX: IMPORTANT ISSUES CONCERNING GPA AND RELATED METHODS

In this section, several issues surrounding the use of GPA and related methods are discussed. To get the most out of these methods, and to avoid misinterpretations it is important that these issues are appreciated. Fortunately much can be understood without having to master the mathematical and algorithmic details that are the concerns of methodological research workers. It will become evident that many issues cannot be resolved in a clear-cut way. Data-analysis is part art, part science and part technology (Healy, 1978); this is certainly true of Procrustes methods.

A.1 Considerations concerning the data of GPA

This subsection considers decisions that have to be taken *before* embarking on a GPA. Potentially here we could discuss much of general data-analytical interest but we have singled out considerations of particular importance in the current context. One topic of major importance that cannot be covered here is the design of methods for collecting (sensory) data and the associated sampling problems.

A.1.1 Raw data, distances, and derived coordinates

We have written x_{ijk} to denote the value obtained by the kth individual for the jth variable on the ith object. This apparently simple statement conceals much that needs discussion. Variables might be quantitative, ordinal or categorical (nominal) whereas the methods under discussion operate only on quantitative values. The values of categorical variables may be replaced by numerical scores (Gifi, 1990) which may be analysed in the usual way. Alternatively and more generally, we may proceed as follows. For each individual k ($k = 1, \ldots, K$) and each pair of objects i_1, i_2 ($i_1 = i, \ldots, N$, $i_2 = 1, \ldots, N$) estimate a distance or dissimilarity $d_{i_1 i_2 k}$; this may be done in many ways (see, for example, Gower & Legendre, 1986). These distances may be assembled into K matrices $\mathbf{D}_k$ which may be treated just as if Type II data had been directly supplied [1]. These initial processes of transformation and distance-matrix computation are valid not only for categorical or ordinal variables but also when all, or some of, the variables are originally in quantitative form.

When transformed versions of the $\mathbf{X}_k$ or $\mathbf{D}_k$ are analysed (see Fig. A1), the estimation of appropriate transformations (including optimal monotonic, spline-monotonic, ordinal and polynomial transformations) may be based on optimising the Procrustes statistic in a given number of dimensions (see, for example, van Buuren & Dijksterhuis, 1988; Gifi, 1990). Another way of deriving coordinate matrices from distance matrices is to do a cluster analysis and then calculate coordinates that generate the induced ultrametric distances—such coordinates are known to be embeddable in Euclidean space and may be calculated easily (see, for example, Gower & Banfield, 1975). The GPA of ultrametrics is useful for comparing different cluster analyses, in precisely the same way that GPA of ordinary distances may be used to compare different multidimensional scalings. Whether $\mathbf{D}_k$ is observed directly, or is computed from raw data on observations of basic variables, it may be analysed by some form of multidimensional scaling (Kruskal & Wish, 1978; Gifi, 1990) to give numerical coordinates $\mathbf{X}_k$ for each individual.

A.1.2 Variables and dimensions

There are essential differences between an $\mathbf{X}_k$ directly observed and one calculated from a distance matrix. These differences stem from the distinction between a jth variable and a jth dimension. When the jth columns of the different data-sets $\mathbf{X}_k$ refer to dimensions (rather than variables) or, as with FCP, to *different* variables, then there is no information in the mean value of any dimension that is useful for comparisons between sets; i.e. for individuals k_1 and k_2, x_{ijk_1} may not be validly compared with x_{ijk_2} unless there is a sense in which the jth coordinate axis of the one configuration matches the jth axis of the other. Even when direct comparisons between variables or dimensions are invalid, distances may be comparable [A.2.2]. Joint rotations and translations—the so-called rigid body motions of applied mathematics—do not affect distances between points of the configuration. The mathematical expression of this simple fact is fundamental to GPA and therefore cannot be avoided. The notion of an orthogonal matrix $\mathbf{H}_k$ may be thought of as specifying a rotation of the kth configuration, although it may also accommodate reflections; translation is specified by a column-vector $\mathbf{m}_k$ giving the displacement of a new origin relative to the current origin. The combined effect of both transformations on a data-set $\mathbf{X}_k$ generated from distances $\mathbf{D}_k$ will give a new set of coordinates $\mathbf{X}_k\mathbf{H}_k + \mathbf{1}\mathbf{m}_k'$ that generates the same distances $\mathbf{D}_k$. GPA is concerned with exploring this distance-preserving model [A.5].

When the jth columns of the different individuals refer to direct measurements on the *same* variable, not only are comparisons valid but they are likely to contain useful information, not available with FCP and with configurations derived from distance matrices, that should certainly be considered in any analysis of how the levels of measurement may differ between individuals and between objects. In the above model, $\mathbf{m}_k$ represents the kth individual's average scores for each variable, the average being taken over objects, and so gives a measure of the *level* of his scoring for each variable. The remaining term represents object scores, not in terms of averages but in terms of rotations. The model permits the individuals' perceptions of the distances between objects to be similar even though objects may have different average scores for different individuals. Note, however, that similar distances do not necessarily imply similar perceptions [A.2.2].

With the same P variables observed by all individuals, a two-way $N \times K$ cross-classified table $\mathbf{Y}_j$ ($j = 1, 2, \ldots, P$) may be constructed for each variable, with entries y_{ijk}; we shall denote the entries in any one of these tables by y_{ik}. The procedure described in the previous paragraph amounts to removing the column-means $y_{\cdot k}$ from this table. If one had preferred to isolate the object means averaged over all individuals, then one would remove the row-means $y_{i\cdot}$. Depending on which method is used, what remains are differences between the means of the objects or individuals (respectively), as well as any rotational differences. Thus the GPA of cross-classified data does not use all the available information about the

structure of the data. For example, when working in deviations from the individual means $\mathbf{m}_k(y_{\cdot k})$ and a rotational term $\mathbf{H}_k$, one should recognise that the distances analysed include additive object means $y_{i\cdot}$. Removing both means would leave for each variable a table of residuals $y_{ik} - y_{i\cdot} - y_{\cdot k} + y_{\cdot\cdot}$, where $y_{\cdot\cdot}$ is the general mean. This is the form of a conventional additive statistical analysis which would be summarised by the marginal means themselves, perhaps associated with a MANOVA (Hand & Taylor, 1987) of all P variables, itself supporting a variety of significance tests concerning certain mathematically derived linear combinations of the observed variables. Generalized Procrustes Analyses *could* be done on residual matrices, focusing either on the row-differences or the column-differences, but because it is not clear to us how these might be interpreted, we would then prefer a conventional analysis that modelled all differences as differences between means rather than in terms of orientations.

Although dimensions should never be treated as variables, an observed, or transformed, variable that is common to several individuals, may be treated as a common dimension if one finds this useful. Figure A1 shows the relationships between some of the different ways in which coordinate matrices may be derived and transformed.

A.2 Scaling and distance

There are two places where scaling has to be considered in GPA. Firstly, there is the possibility, already mentioned in [2.4], of estimating optimal isotropic scaling factors ρ_k associated with each of the K datasets. Secondly, as part of the initial transformations of data [A.1.1], scaling may be required to make the datasets commensurable; the form this may take will depend on whether the data is of Type I or Type II and, if of Type I, different considerations will apply to FCP and situations where Variables are Comparable across Individuals [A.1.2]. Further, different considerations will apply to variables with a well defined Limited set of Scale-values, say line-scales and category-scales, and those with Open-Ended Scales, say measurements of weight or length. The issues underlying initial scaling (also termed prescaling) can be very complex and only an introduction can be given here. Also included in this section is a short discussion of some of the implications of using distance as the underlying interpretive tool for GPA.

A.2.1 Scaling factors

A matter that needs attention is whether or not the coordinate matrices $\mathbf{X}_k$ to be used in GPA call for any kind of scaling. Gower (1975) described how the estimation of a set of optimal isotropic scaling factors ρ_k could be incorporated into a GPA [2.4]. The factor ρ_k scales the matrix $\mathbf{X}_k$, thus allowing for innate

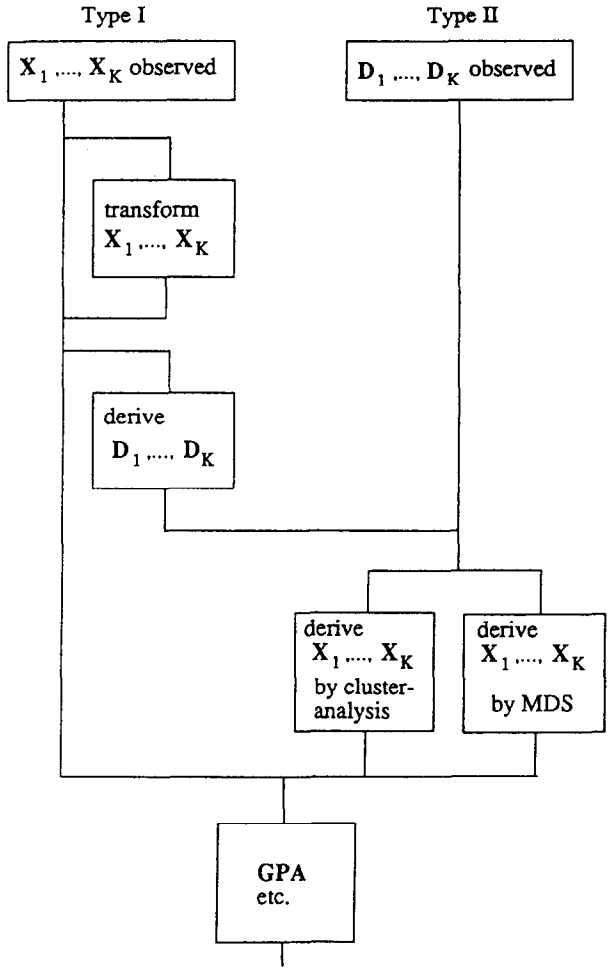


Fig. A1. Initial adjustments (prescaling) to the data preparatory to a GPA.

differences in sizes. This procedure is commonly used but can only be justified when the coordinates contain size information attributable to different behaviour among the individuals. For example, when the variables represent judgements on some set of scales, then some individuals may use the full range of the scale and others only a limited range of values. Using an isotropic scaling factor is then valid provided that this propensity extends, at least approximately, equal to all measurement scales. Such scaling would continue to be justifiable when the observed variables had undergone initial transformations that preserve size information. It would also be acceptable when observed $\mathbf{X}_k$ generate matrices $\mathbf{D}_k$ which inherit size factors attributable to individual differences and which are passed on to new $\mathbf{X}_k$ generated for analysis by GPA (Fig. A1). Generally, these considerations would exclude the use of isotropic scaling factors for coordinates derived from forms of non-metric scaling (see, for example, Kruskal & Wish, 1978) (which are invariant to size information) but allow them for coordinates derived from forms of metric scaling (which retain size information). When scaling factors cannot be given substantive interpretations,

their estimation is questionable and, in particular, to use scaling solely to increase the apparent goodness-of-fit is to be deplored, because without interpretation the apparent gain is illusory. Whenever non-metric multidimensional scaling is used, the derived coordinate matrices $\mathbf{X}_k$ will normally have arbitrary size and rather than adjusting for this by estimating optimal scaling factors ρ_k, which is one possibility, it is simpler and better to ensure equal size by putting the origin at the centroid and adopting an initial scaling such that the sum-of-squares of the elements of $\mathbf{X}_k$ is a constant for all values of k; the obvious choice for this constant is unity, but the value is arbitrary. These operations are indicated in the lower part of Table A1.

Even when there is a plausible reason for it, isotropic scaling can give only a very crude treatment. It does not handle situations where two individuals both use the full range of a measurement scale but one uses it uniformly while the other rarely uses the extremes of the scale. Also it does not accommodate those individuals who treat different variables in different manners. The latter would require an anisotropic scaling factor ρ_{jk} giving the scaling for the kth individual on the jth variable and one now has to consider whether the scaling is applied to the original variables before transformation or to the dimensions after rotation. These generalisations considerably complicate the conceptual framework, interpretation and computation ([A.4.5], Lingoes & Borg, 1978; Commandeur, 1991). Other forms of anisotropic scaling occur in biological applications where the elements of distance matrices $\mathbf{D}_k$ represent distances between landmark points on a (once) living organism, the parts which grow or develop in a differential manner (Siegel & Benson, 1982). Further remarks on isotropic scaling are given in [A.4.5].

When the columns of the $\mathbf{X}_k$ represent variables rather than dimensions [A.1.2], initial scaling may be desirable to adjust for incommensurability of measurement scales. When all variables have limited scales [A.2], then an initial scaling is inbuilt and further scaling may not be required. Whenever the number P_k of variables associated with each of the K individuals is not a constant, as with FCP, a spurious overall size-difference will be induced. This is easily eliminated by dividing each variable for the kth individual by the square root of P_k, thus putting the size of the matrix for each individual on a 'per variable' basis; this we term P_k-scaling. For FCP variables with open-ended scales [A.2], or with different scale ranges, each variable might first be adjusted to have either unit range or unit sum-of-squares about its mean; in the other cases, additional P_k-scaling will ensure that the sum-of-squares for each individual configuration is unity. The aim here is to give all individuals the possibility of generating the same distance matrix. Even after these steps it may be evident that different individuals may be using their own adjusted measurement scales differently from the way other individuals are using their adjusted scales and then it may be justifiable to include the estimation of isotropic scale-factors ρ_k in the GPA of FCP data. All this is summarised in the Free Choice Profile rows of Table A1.

When variables are comparable across individuals [A.1], similar initial scaling operations to the FCP case can be considered, but with variables measured on ratio scales one simple action is to take logarithms, which ensures that distances are invariant to the choice of alternative measurement scales for the same variables (e.g. centimetres and inches). There is the possibility of scaling each variable globally. One useful method of

TABLE A1. Possibilities for Initial and Isotropic Scaling of the Data-Sets $\mathbf{X}_k$[a]

Variables	**Type I data, Observed $\mathbf{X}_k$**	
	Measurement scales	**Scaling allowed**
Variables comparable	Limited scale	ρ_k
	Open-ended scale	Replace $\mathbf{X}_k$ by canonical scores, ρ_k
Free choice profiling	Limited scale	P_k, ρ_k
	Open-ended scale	diag $(\mathbf{X}_k'\mathbf{X}_k) = \mathbf{I}$, P_k, ρ_k
	Type II data $\mathbf{X}_k$ derived from observed $\mathbf{D}_k$ by multidimensional scaling	
	Type of multidimensional scaling	**Scaling allowed**
	Metric	ρ_k
	Non-metric	Trace $(\mathbf{X}_k'\mathbf{X}_k) = 1$

[a] The terminology is explained in [A.2.1].
ρ_k: isotropic scaling is a possibility.
P_k: needs P_k-scaling (see [A.2.1]).
Trace $(\mathbf{X}_k'\mathbf{X}_k) = 1$: scale the sum-of-squares of each centred data-set to be unity.
diag $(\mathbf{X}_k'\mathbf{X}_k) = \mathbf{I}$: scale the sum-of-squares of every column of each centred data-set to be unity.

global scaling is to regard the individuals as defining groups in a canonical variate analysis (see, for example, Gittins, 1984). The original $\mathbf{X}_k$ may then be replaced by their canonical variate values and analysed as if they had been produced by any other metric-scaling method. This procedure has the advantage of making an allowance for the correlations between variables, but as with the other methods for handling comparable variables, it does not adjust for object means [A.1.2]; other MANOVA methods are also worth considering here (Hand & Taylor, 1987). The foregoing remarks about initial scaling refer principally to GPA; in [A.4.4] we shall see that with other models other considerations may be relevant.

A.2.2 Distance

When working with distances between objects given by, or calculated for, each individual separately (Type II), all information is lost on original variable values and hence on how these may differ between individuals. What remains are the differences between the differences between objects; i.e. for individuals k_1 and k_2 we can compare $d_{i_1 i_2 k_1}$ with $d_{i_1 i_2 k_2}$, giving a measure of how the two individuals view the comparison between the same pair (i_1, i_2) of objects. If the perceptions of two individuals are in some sense the same, then their distance matrices would be expected to agree; however, the converse does not follow. Thus, although it is impossible to know in absolute terms how individuals k_1 and k_2 may perceive objects i_1 and i_2, at least we may investigate whether they agree on the degree of difference. When the differences are small it may be inferred that the individuals *may* have similar perceptions, but they could also be diametrically opposed (think of an individual who scores 1, ..., 10 in that order for 10 objects, while a second individual scores 10, ..., 1; both would give the same distances).

GPA and related methods rely heavily on distance interpretations. The rationale for these methods is to search for configurations representing individuals that substantially agree in their inter-object distances, the group average giving an overall measure. What can it mean when two configurations agree although, as with FCP data, based on different sets of measurements? The obvious answer is that perhaps both configurations indicate measures of the same underlying qualities, rather as the latent variables of factor analysis are imagined as inducing structure in measurable variables. GPA does not assume common distance matrices but it does allow the commonality and the degree of departures from commonality to be investigated.

A.3 Considerations of unequal dimensionalities in the data

The practice [2] in GPA of appending zero columns to FCP, or other column-deficient data, has been treated with misgivings in some quarters, and other methods (for example, Peay, 1988) have been seen as circumventing a perceived problem. It is our view that the problem is largely non-existent, arising in the main from a confusion between different possible interpretations of the standard GPA model and partly from misinterpretations about other related types of model. These issues deserve examination and are explored in the following subsection.

A.3.1 Free choice profiling and 'zero columns'

The 'problem' is most likely to arise with FCP when it is rare for every individual to nominate the same number of profiles; that is, unless the individuals have been instructed to nominate a specified number of profiles. As was discussed in [1], even with a constant number of profiles, FCP always carries with it the difficulty that profiles cannot be compared from one individual to another and cannot be regarded as shared. Although, except for translation to a common centroid, the translation part of the analysis is invalid for FCP, the orientation part remains useful. Indeed, orientation is about the only thing that can be exploited when analysing free profiles; even the use of isotropic scaling factors requires special care [A.2.1]. What can be done is to evaluate a distance matrix for each individual and the coordinate matrices $\mathbf{X}_k$ $(k = 1, \ldots, K)$ which generate them, just as was described in [1] and discussed in [A.2.2]. With most choices of distance, any differences in dimensionality will vanish in this process. The initial transformations of [A.2.1] are subsumed in the possible transformation steps shown in Fig. A1, where one route leaves the dimensions of the original data-sets unchanged; hence, when they were of different dimensions before transformation they will remain so. Then zero columns should be appended so that all $\mathbf{X}_k$ occupy the same space as the largest set. This gives the correct mathematical result for rotating the individual configurations to optimal fit in the GPA sense. There is nothing wrong in rotating, say, a linear configuration to fit a two-dimensional one and it is just a generalisation of this that the process accomplishes. Because the initial configurations are arbitrarily oriented, all initial orientations of a lower-dimensional configuration in a larger space are equivalent. Most orientations will lead to non-zero values in all P dimensions but nevertheless give identical GPA results; however, the orientation with zero columns is the most simple. Arbitrary initial orientation can be hidden in non-zero, but linearly related, sets of columns, which have the same effect as the more obvious column-deficiency of adding zero columns. Those concerned about adding zero columns do not seem to be concerned by such unnoticed relationships.

A related issue is when an individual freely chooses so many profiles that his data-set has more variables than objects, in a sense having an *excess* of dimensions. No special action need be taken, provided the remaining data-sets are padded out by zero columns. Alternatively,

for any data-set whose number of chosen profiles exceeds $N - 1$, a principal component analysis may be used to generate a new set with no more than $N - 1$ columns. The two methods give the same distances and therefore are completely equivalent, provided the components are extracted from the sums-of-squares-and-products matrix and not from the correlation matrix. Note that data-sets with N profiles are in $(N - 1)$ dimensions when expressed in deviations from their means; when a component analysis is not used, this gives yet another way in which column deficiency can occur without showing overt zero columns. When plotting the property axes [2.7], a consequence of adding zero columns is that there will be axes corresponding to null-properties. This is also sometimes regarded as a deficiency of the padding process. However, as has already been explained, it is not only the directions of axes but also their lengths which matter. Axes corresponding to null properties have zero lengths, so vanish and cause no problem. Padding with zero columns is a convenient device with no harmful side-effects.

A.4 Dimensionality of the presentation and projections

As well as considerations concerning the dimensionality of the data [A.3], there are also considerations concerning the dimensionality of what we have termed the R-dimensional exhibited space of the GPA group average (consensus) ([2.5], [2.6]). The two concepts of the dimensionality of the data and the dimensionality of the presentation are often confused with one another. In GPA, we have taken the view that they are totally unrelated and that the full GPA solution exists in P dimensions, the R-dimensional space being used only to give an approximation that has visual and presentational conveniences but which should only be used with caution if the approximation is a poor one. We believe that this view often closely represents the true state of affairs, but in the following we discuss how alternative interpretations of the $(P - R)$-dimensional unexhibited space influence thinking and lead to the consideration of different models.

A.4.1 Optimising projections—Peay's model

Recalling [2.3] that GPA minimises the residual sum-of-squares $T - M - S$ and, equivalently, maximises S of Table 2, an alternative criterion that has been considered by Peay (1988) is to find the orientations and, if required, the isotropic scaling factors, that maximise S_R of Table 3; this is equivalent to minimising $T - M - S_R$. Thus, the Peay group average is the one that is best approximated in the exhibited space. There is no formal partition of sums-of-squares in the unexhibited part of the space and, there, the orientation of the configurations is arbitrary, giving different values of $S_{(P-R)}$ and $T_{(P-R)} - M_{(P-R)} - S_{(P-R)}$ with a constant sum. A unique partition is not necessary but, if required, can be found in many ways; for illustrative purposes, and because it maximises $S_{(P-R)}$ conditional on the Peay S_R, we shall imagine that the unexhibited space is partitioned by GPA [3.4]. Because the GPA group average maximises the group average fitted in the maximal space, the total Peay group average fit must be poorer in this respect (whatever partition is used for the unexhibited space) and the residual variation 'Individual Deviations' will be greater than that of GPA. Two reasons have been put forward for favouring the Peay fit. The first is that S_R obtained by Peay is maximised, and hence is undeniably greater than S_R obtained by GPA in an exhibited space of the same number (R) of dimensions. As we have just explained, this 'improved' fit pertains only to the exhibited part of the group average and is poorer in total. It might be acceptable to maximise S_R if one were prepared to regard the unexhibited $(P - R)$ dimensional part of the analysis as arising from noise or random variation, arguing that Peay's criterion might be regarded as maximising the signal. We discuss this in [A.4.2]. The second reason for favouring the Peay criterion is that it is said to give an improved method for handling differing dimensionalities in the data. We have already explained [A.3.1] why we believe that appending zero columns is a satisfactory procedure, provided due attention is paid to any induced size differences [A.2.1]; the latter is something which also deserves serious consideration with the Peay criterion. The way the Peay process handles the dimensional problem, if it is one, is to choose R to be no more than the dimensionality π of the smallest data-set $\mathbf{X}_k$. When all data-sets are of the same size, then $\pi = P$, and we may set $R = P$ and, hence, $P - R = 0$. This use of Peay's method is identical to GPA but with no partition into exhibited and unexhibited components. In general, Peay's criterion may be thought of as first rotating and then projecting the higher-dimensional configurations onto R dimensions and doing a GPA in the smaller space, all of which can be expressed in terms of K projection matrices $\mathbf{R}_k$ $(k = 1, 2, \ldots, K)$ with P_k rows and π columns [A.5]. The trick is to find the projection onto R dimensions of the group average in P dimensions that maximises S_R. From these remarks it should be clear that the geometry of [2] remains valid and that Peay's process operates in P dimensions just as much as does GPA. Mathematically, all methods that satisfy the Analysis of Variance decomposition of Table 3 operate in the same P-dimensional space. They differ in the models they fit and what criterion is adopted to judge best fit [A.5].

Projections may be thought of in terms of casting shadows (sun at mid-day) from a high-dimensional space onto a low-dimensional space. Every set of orientations in the maximal space will have a group average and Peay's criterion ensures that this has the biggest shadow possible in R dimensions. In our opinion, a

disadvantage is that provided the variation in the unexhibited $(P - R)$-dimensional space is not attributable to noise [A.4.2], it could be of at least as equal, and possibly of more, interest than that in the exhibited space though, of course, an Analysis of Variance equivalent to that given in Table 3 would help indicate if this were so. The difference between the two approaches is that GPA with component analysis exhibits the best R-dimensional fit of the best group average while Peay's analysis exhibits the best R-dimensional group average as part of a generally poorer fit, both in terms of an increased residual sum-of-squares and a poorer P-dimensional group average [3.4].

A.4.2 Signal and noise

It seems to us that although Peay's method is mathematically sound, there is little justification for its use in the types of problem where the R-dimensional and $(P - R)$-dimensional spaces both represent substantive effects, but this is not to say that there are no situations where it can be used. One case where it was worth considering is when the unexhibited space is attributable to noise, although Peay's method and GPA are then likely to give similar results. Peay's model then may be regarded as a basis for noise removal but, on the grounds that there seems to be little reason for expecting any signal to be π-dimensional, one would probably wish to explore a range of values of R less than π, and perhaps even some greater values. However, in such cases it seems better to remove the noise at the outset. Indeed when the initial configurations have been derived via multidimensional scaling (see [A.2.1] and Fig. A1) the noise should already have been removed. A simple method that should suffice for removing noise would be to replace each $\mathbf{X}_k$ by scores on its first principal components. This suggestion was also made by Peay (1988), but is dismissed with the remark, 'However, it does not appear likely to attract a great deal of empirical interest'. We believe that the suggestion has genuine empirical interest. Once noise has been eliminated or made small compared to the signal, there seems to be little justification for maximising S_R in fewer dimensions than the signal. Thus our suggestion is first to remove noise, and then to analyse the signal by GPA.

We have already pointed out that Peay's analysis coincides with a GPA of the configurations projected into the R-dimensional space. This space, regarded now as the space containing the signal, may itself be high dimensional and, with advantage, be partitioned as previously into exhibited and unexhibited parts. This consideration suggests a three-column breakdown in the Analysis of Variance, partitioning $R = R_1 + R_2$ (see Table A2), with R_1 dimensions for the exhibited space, R_2 dimensions attributable to the non-exhibited systematic effects, as before, and the remaining $P - R$ dimensions for noise (random variation). Among other possibilities, the partition could be based on the R_1 principal components of the R-dimensional group average, or on all R-dimensional coordinates. All entries in the column for the $P - R$ dimensions and all entries in the residual row, labelled $T - M - S$, should be attributable to random variation, as may some of the entries in the R_2-dimensional space (see below). In principle, and on appropriate distributional assumptions, significance tests can be devised to examine the equality of such terms (Davies, 1978; Sibson, 1978; Langeheine, 1982; Langron & Collins, 1985).

An alternative partition of the Peay (or any other) R-dimensional projection is into R_1 dimensions, also obtained by a Peay projection, and R_2 dimensions, representing what is left over. This has the disadvantages that we have already discussed of giving a sub-optimal view of the substantive signal. However, when the R_1 dimensions are intended for display, it might be worth maximising the signal-to-noise ratio $S_{R_1}/(T_{R_1} - M_{R_1} - S_{R_1})$.

The notions of signal and noise have some appeal but it must be recognised that, with most forms of Procrustes analysis, only rarely are there objective methods for distinguishing signal from noise. Ideally to do this would require individuals to replicate object scores. Without replication, one has to fall back on assuming that the higher dimensions are attributable to noise. This is far from satisfactory but is analogous to

TABLE A2. Analysis of Variance, as in Table 3, but Partitioning Every Sum-of-Squares into a Part that is Exhibited in R_1 Dimensions, a Systematic (Signal) but Unexhibited Part in R_2 dimensions, and a Random (noise) Part in the Remaining $P - R$ Dimensions, where $R = R_1 + R_2$.

Source of variation	Sums-of-squares		
	Exhibited (R_1 dimensions)	Unexhibited (R_2 dimensions)	Noise ($P - R$ dimensions)
Individuals	M_{R_1}	M_{R_2}	$M_{(P-R)}$
Orientations:			
Group average	S_{R_1}	S_{R_2}	$S_{(P-R)}$
Deviations	$T_{R_1} - M_{R_1} - S_{R_1}$	$T_{R_2} - M_{R_2} - S_{R_2}$	$T_{(P-R)} - M_{(P-R)} - S_{(P-R)}$
Total	T_{R_1}	T_{R_2}	$T_{(P-R)}$

the situation in those designed experiments with no replication, where the higher-order interaction effects are regarded as estimates of error. Our Analysis of Variance framework is easily extended to cope with the extra information given by replication. All that is required is to add an additional line to Tables 3 and A2 labelled 'Replication', indicated by E (for error) with sums-of-squares E_{R_1}, E_{R_2}, E_{P-R}. To proceed we need the degrees of freedom g, d and e corresponding to Group Average, Deviations and Replications. Then, if the mean-square for, say, the deviations in the unexhibited space $(T_{R_2} - M_{R_2} - S_{R_2})/d$ is of a similar size to E_{R_2}/e, this would give an informal indication that the corresponding dimensions referred to noise rather than to signal. Similar considerations would apply to the group-average term S_{R_2} and to any column purporting to represent noise in $P - R$ dimensions. In this way we have a rough means of deciding how to partition our space between signal and noise.

Instead of totally discarding random dimensions arising from any initial component analyses, or other forms of multidimensional scaling used to eliminate noise, they might be combined with the terms for the $P - R$ dimensions attributable to error in the final column of the Analysis of Variance of Table A2, or even be presented as a separate fourth column whose entries could be compared with other 'random' entries.

A.4.3 Other criteria

Before considering further criteria, it is worth noting, as a general remark, that if the optimisation criterion operates within the R-dimensional space, however that is defined, then distances in the remaining part of the configuration in $P - R$ dimensions can be generated by sets of coordinates and analysed independently (an example of this was given in [A.4.2]). Thus, for example, having minimised $T_{R_1} - M_{R_1} - S_{R_1}$, we may then minimise $T_{R_2} - M_{R_2} - S_{R_2}$, conditional on the first minimisation. A similar remark applies to any further columns there may be in the Analysis of Variance. One implication is that, unless these further optimisations are done, there is a degree of arbitrariness in some of the terms; but this is no problem when these terms are attributable to error, as will usually be the case. We do not pursue it further here but, nevertheless, it should be borne in mind as one of the options available with Peay's criterion as well as with other criteria in this general class of analyses.

A.4.4 Optimising projections—Green and Gower

A criterion for fitting sets of configurations, possibly of different dimensionalities, has been suggested by Green and Gower (1979) and Gower (1984) and discussed by ten Berge and Knol (1984) and Peay (1988), who refers to it as 'Procrustes'. It is not the same as GPA and was not intended as a fix for GPA to handle different dimensionalities. In the notation of Table 3, this method minimises $T_R - M_R - S_R$ so, like Peay's method, operates in the space of the exhibited configuration. Again, one seeks projections from the higher-dimensional spaces onto an R-dimensional space, but now a GPA in the smaller space minimises the above-mentioned residual sum-of-squares. This criterion was found useful by Constantine and Gower (1982). In [A.4.1] we have used the imagery of shadows to get a feeling for projections. The Green and Gower criterion explores to what extent the configurations for the different individuals can be orientated to give similarly sized and oriented shadows (projections) in R dimensions; note the difference between this and having big shadows. A group average of similar shadows might be interpreted as a true consensus, arising from features of the objects on which there is a high degree of common agreement; this is not to say that the features on which individuals disagree might not form a major part of their perceptions. This point may be illustrated by considering data-sets $(\mathbf{X}_1, \mathbf{x}, \mathbf{y})$, $(\mathbf{X}_2, \mathbf{x}, \mathbf{y})$, ..., $(\mathbf{X}_k, \mathbf{x}, \mathbf{y})$, all having common observations on each of two variables $\mathbf{x}$, $\mathbf{y}$, but otherwise of possibly different dimensionalities. Although variables $\mathbf{x}$, $\mathbf{y}$ are envisaged as having the same numerical values, we remain in the context of FCP, so there is absolutely no guarantee that they represent the same substantive variables. Projection onto the two-dimensional consensus space containing $\mathbf{x}$, $\mathbf{y}$ will always give a perfect fit for $R = 2$, which may be taken as interesting evidence of some underlying agreement between putatively incommensurable variables, notwithstanding how trivial these variables may be compared with those contained in $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k)$, which represent disagreement. Perhaps, after all, there may be some sense in which texture and odour ratings (say) are both measures of some common underlying basic property of the objects [A.2.2]. The R-dimensional space could now reasonably be labelled the consensus space rather than the exhibited space and the group average in that space is a much more plausible measure of consensus than that given either by GPA or Peay. Having fitted the consensus space, one might wish to go on to explore differences in the $(P - R)$-dimensional 'disagreement' space, regarded as signal rather than noise, perhaps by doing a GPA or Peay's analysis in that space [A.4.3]. We must now modify the remarks on initial scaling given in [A.2.1], for dividing by the square root of P_k as recommended there will give different scalings for each $(\mathbf{x}, \mathbf{y})$ pair, depending on how many variables there are in each $\mathbf{X}_k$. In the current context this will tend to obscure the consensus signal, but this would be restored by estimating optimal isotropic scaling factors ρ_k; alternatively, it might be better to eliminate their need for isotropic scaling by omitting scaling altogether.

A.4.5 Individual differences scaling

Another method which deserves mention here is Individual Differences Scaling (INDSCAL) of Carroll

and Chang (1970). Like GPA, this handles sets of data for individuals and essentially operates on distance matrices defined for each individual. Like GPA, a central feature is the calculation of a group-average matrix; unlike GPA, the axes associated with the R dimensions of the group-average matrix are not arbitrary. Associated with each axis is a weight, one for each individual. The axes may be regarded as representing latent variables, which are often reified, and the weights as the scaling each individual gives to each latent variable. These weights can be compared with the anisotropic scaling ρ_{jk} of [A.2.1], that the kth individual might give to the jth substantive variable. Directions in the space of the GPA group average are also sometimes reified, but in these cases the directions are arbitrary and there is certainly no need to use only the principle axes, or indeed any orthogonal axes, as is so often done. It seems to be a general feature of models with anisotropically weighted latent variables, not just the INDSCAL model, that they induce unique axes (Commandeur, 1991).

A.5 Criteria, models and algorithms

The important thing to emphasise about all the methods described above is that just because they handle similar kinds of data and do superficially similar things, they should not be regarded merely as variant forms of analyses. Not only may they use different kinds of goodness-of-fit criteria, or loss functions, as they are often termed, but more important they are trying to expose different features of the data. We have already explained what these different features are ([2], [A.4.1], [A.4.4], [A.4.5]). In a more traditional statistical context it would be said that each is fitting a different model, rather than that all are fitting the same model by different methods. The basic idea in all the models considered here is that each data-set $\mathbf{X}_k$ may be partitioned into signal and noise and the signal modelled in terms of a group average and deviations from it. The essential component of the model(s) has already been written down in [A.1.2] and may be more fully treated as follows. Consider

$$\mathbf{X}_k\mathbf{H}_k = \mathbf{X}_k(\mathbf{R}_k, \mathbf{Q}_k) = (\mathbf{X}_k\mathbf{R}_k, \mathbf{X}_k\mathbf{Q}_k) \qquad \text{for } k = 1, 2, \ldots, K,$$

where $\mathbf{H}_k$ is an orthogonal matrix, which expresses that it is only the distances generated by $\mathbf{X}_k$ that are being modelled. The partition $\mathbf{H}_k = (\mathbf{R}_k, \mathbf{Q}_k)$ gives a part $\mathbf{R}_k$ representing projections onto an R-dimensional space and a part $\mathbf{Q}_k$ representing the complementary projections onto the remaining $(P_k - R)$-dimensions. The first term in the partition is modelled by

$$\mathbf{X}_k\mathbf{R}_k = \mathbf{X} + \mathbf{E}_k \qquad \text{(A1)}$$

where $\mathbf{X}$ is the group average in the space of the projections and $\mathbf{E}_k$ represents the deviations of the projected data-set for the kth individual from the group average. All the models share this common framework; the differences lie in the interpretation of the terms in the models and, in particular, what terms are regarded as noise. In GPA there is no term $\mathbf{X}_k\mathbf{Q}_k$, and $\mathbf{E}_k$ alone takes on the role of noise; in Peay's model, both $\mathbf{E}_k$ and $\mathbf{X}_k\mathbf{Q}_k$ are implicitly regarded as noise; and in the Green and Gower model, only $\mathbf{E}_k$ is regarded as noise while $\mathbf{X}$ (agreement or consensus) and $\mathbf{X}_k\mathbf{Q}_k$ (disagreement) model different aspects of the signal. The term $\mathbf{X}_k\mathbf{R}_k$ may be partitioned into exhibited and unexhibited dimensions $\mathbf{X}_k\mathbf{R}_k\mathbf{R}$ and $\mathbf{X}_k\mathbf{R}_k\mathbf{Q}$, say, with corresponding group averages $\mathbf{XR}$ and $\mathbf{XQ}$ ([2.6], [A.4.2]). Here $\mathbf{R}$ and $\mathbf{Q}$ are matrices projecting into independent subspaces, so that $\mathbf{R'Q} = 0$. In GPA, $\mathbf{R}$ and $\mathbf{Q}$ are normally taken to be the principal components of $\mathbf{X}$, but $\mathbf{R}$ and $\mathbf{Q}$ may be chosen in other ways and their use is not confined to GPA. Further, $\mathbf{X}_k\mathbf{Q}_k$ may itself be modelled similarly to eqn (A1). As described in [2.4], isotropic scaling is included in the models by replacing $\mathbf{X}_k$ elsewhere above by $\rho_k\mathbf{X}_k$. An alternative model for scaling that does not seem to have been considered, but which sometimes might be more realistic, is to attach ρ_k only to the signal $\mathbf{X}_k\mathbf{R}_k$.

For plotting axes [2.7], the coordinates given by the pth row of $\rho_k\mathbf{H}_k$, $\rho_k\mathbf{R}_k$ and $\rho_k\mathbf{R}_k\mathbf{R}$, whatever may be the projection under consideration, represent one unit of the scale of measurement for the pth property of $\mathbf{X}_k$ as recorded by the kth individual. Axes joining these points to the origin may be plotted and the unit point(s) marked; axes should not be extended in either direction much beyond the range of measurement occurring in the data [A.3.1].

Writing $\|\mathbf{A}\|$ to denote the sum-of-squares of all the elements of the matrix $\mathbf{A}$, it turns out that with model eqn (A1), we have that $\sum_{k=1}^{K}\|\mathbf{E}_k\|$ is minimised when $\sum_{k=1}^{N}\mathbf{X}_k\mathbf{R}_k/K$, and then

$$\sum_{k=1}^{K}\|\mathbf{X}_k\| = K\|\mathbf{X}\| + \sum_{k=1}^{K}\|\mathbf{E}_k\| + \sum_{k=1}^{K}\|\mathbf{X}_k\mathbf{Q}_k\| \qquad \text{(A2)}$$

which is the basic Analysis of Variance, Total s.s. = Group Average s.s. + Individual Deviations s.s + Noise s.s., as shown in Tables 3 and A2, where the additional breakdown into exhibited and unexhibited parts is also shown.

In all the above, the optimally criterion used to fit the models, is least-squares, where the sum of one term or more on the right-hand side of eqn (A2) is minimised, depending on the model being fitted (e.g. minimising $T - M - S$ of Table 2). Other types of criteria may be considered: for example, weighted least-squares, in which differences are weighted by quantities provided, or even computed iteratively from the data (Verboon & Heiser, 1991). Two examples are (i) when the elements of $\mathbf{X}_k$ are means of replicates so that the number of replicates associated with each provides a natural system of weights, and (ii) when some objects are missing, which therefore may be given zero weight compared with unit weight for non-missing objects (Commandeur, 1991).

Another criterion, which itself has several variants, is to maximise some form of the correlation between $\mathbf{X}_k$ and the model $\mathbf{Y}_k$, say, to be fitted. The inner-product criterion of ten Berge (1977) (see also ten Berge & Knol, 1984) is of this kind. $\mathbf{Y}_k$ and $\mathbf{X}_k$ may be regarded as being strung out of vectors, and then the criterion is to maximise the sum-of-products of corresponding elements. We have no fundamental objection to the use of this type of criterion but personally prefer the least-squares criterion with the attractive geometric simplicity of its distance interpretation. For us, this preference is reinforced when the fitted configurations are interpreted through distance concepts and summarised through an Analysis of Variance—recall that Table 3 and A2 remain valid, however the configurations are derived. Forms of analysis which mix different optimality criteria have an element of inconsistency.

Just as the above brief account of the different criteria that have, or might have, been considered, is useful for giving a general appreciation of how GPA fits into a wider framework of related types of data-analysis, similarly it is useful to have a general appreciation of some of the issues involved in computation. Some criteria and some models may lead to more simple computations than others; some may lead to faster computations than others; some may have better convergence properties than others; some may have a tendency to find sub-optimal solutions; some may be easier than others to program; and some may use fewer computing resources. Even with a fixed criterion, one numerical process for its optimisation may have advantages relative to other numerical processes with identical objectives. Thus among methodologists and numerical analysts, there is a constant search for improved algorithms. Finally, different people programming precisely the same thing may write computer programs of differing utilities because of the effects of variant methods of organising the calculations and the degree of user-friendliness built into their programs and documentation. Good statistical methods can get a bad reputation merely because they are supported by poor software; bad methods may become widely accepted merely because they are well marketed; good methods not supported by any software are unlikely to gain much currency. The subject is too vast for us to give general advice here. Suffice it to say that GPA leads to reliable and acceptably efficient algorithms that are simple to program, and whose results are interpretable to the informed layman.