7 Three-Way and Further Extensions of the Basic Model

There are not three incomprehensibles, nor three uncreated: but one uncreated and one incomprehensible

QUICUNQUE VULT (Creed of St. Athanasius)
Book of Common Prayer, 1662

7.1 Introduction

The remaining programs in the MDS(X) series are either designed for the analysis of three- (and in the case of CANDECOMP, higher-) way data or (as in the case of PREFMAP I and II) are more complex variants of models already encountered in Chapter 6.

The main differentiating characteristic of the programs considered here is the form of the model, or rather models, since both PREFMAP and PINDIS consist of a hierarchy of models of increasing complexity. As in the previous chapter, we shall begin by examining the type of data input to these programs—this provides the best clue to their most fruitful areas of application—and then go on to describe the form of the models employed.

7.1.1 Three- (and higher-) way data

The term three-way data refers to a 'cube' of data (see Figure 7.1, p. 192). Such data occur frequently. The third way usually consists of a set of individuals, occasions, methods, points in time, experimental conditions or geographical locations. Two types of 3-way data are usefully distinguished:

(i) **3-way data which are two-mode,** i.e. consist of a set of ordinary 2-way, one-mode, (dis)similarity matrices, and
(ii) **3-way data which are three-mode,** representing for instance the preferences of a set of subjects (mode 1) for a set of food items (mode 2), where the judgments were made at a number of different occasions (mode 3).

Examples of such 3-way data are:

(a) *Three-way, two-mode data* (A set of pairwise (dis)similarity matrices)

A set of individuals (mode 1) each produce a matrix of pairwise similarity ratings between stimuli (mode 2).

Over a number of weeks (mode 1) the mutual attraction of a set of fraternity members (mode 2) is assessed by averaging the preference scores they give to each other.

A set of individuals rate a set of concepts in terms of a set of semantic differential scales. The ratings between each pair of scales are then correlated. This gives rise to a set of correlation matrices between scales (mode 1), one matrix for each concept (mode 2). (Note that in this case what were originally 3-way, 3-mode data have been reduced by the researcher to 3-way, 2-mode data by aggregating over individuals.)

The frequency with which each pair of plant species (mode 1) co-occurs is tabulated for each of a number of locations (mode 2).

A set of attitude items (mode 1) are rated on a 7-point scale by a set of subjects, and a number of different coefficients of ordinal association (mode 2) are calculated between the items.

A set of five live fish (mode 1) are confronted with different stylised shapes of fish, differing on sexual and other characteristics. Their behaviours are summarised in each case by a matrix of rank correlations representing the similarity of behaviour when presented with stylised fish i as opposed to stylised fish j (mode 2).

(b) *Three-way, three-mode data* (Three distinct sets of entities)

A set of individuals (mode 1) rate a set of automobiles (mode 2) on a set of rating scales (mode 3).

Members of a social group (mode 1) rank each other (mode 2) in terms of emotional closeness. Data are collected on a number of occasions (mode 3).

The input (mode 1)-output (mode 2) matrices between a set of industries is collected for a set of nations (mode 3).

(c) *Higher-way data* (N-way data)

There are examples in the literature of four-way data, e.g. semantic differential experiments on a number of occasions (mode 1), using the same set of individuals (mode 2) to judge the same set of concepts (mode 3) on the same set of rating scales (mode 4). (This is 4-way, 4-mode scaling.)

Each year (mode 1), a (different) set of individuals make pairwise judgments of similarity between a set of names of nations. The investigator wished to distinguish European, North American, Latin American and Third World subjects' judgments, and therefore produced a separate correlation matrix between nations (mode 2) for each sphere of origin (mode 3). This is 4-way, 3-mode data.

In principle, data of any way can be scaled, and the CANDECOMP program accepts up to seven-way data. Users are advised to proceed beyond three-way data with considerable caution. They are in largely uncharted territory.

7.1.2 Organisation of the chapter

The defining characteristics of the MDS(X) programs for analysing three-way and related data are described in Table 7.1. As in previous chapters, characteristics of the *data*, scaling *transformation* and *model* are used to define the programs involved. The models described in this chapter consist mainly of generalised versions of the distance and vector models encountered in Chapter 6. The exact form of the generalisation is specified in Table 7.1 under the headings of dimensional weighting and rotation (in the case of distance models) and vector weighting and translation (in the case of vector models). It will be easier to discuss these increasingly complex transformations in the context of the program(s) where they occur.

Let us first take the programs in the order in which they appear in Table 7.1, an order determined by the type of data they analyse. In the subsequent sections

DATA	MODEL	Specification* T R V W P	TRANSFORM**	MDS (X) PROGRAM	Description
TWO-WAY 2- MODE	DISTANCE	(R W P)	M or L	PREFMAP (I) (PM 1)	Dimensional Salience with Idiosyncratic Orientation; Ideal Point.
		(W P)	M or L	PREFMAP (II) (PM 2)	Dimensional Salience with Ideal Point.
THREE-WAY 2-MODE (CONFIGURATIONS)	DISTANCE	(W)	L	INDSCAL -S	Dimensional Salience
		((R))	Similarity (S)	PINDIS (PO)	Basic: Procrustes Rotation (General Similarity)
	DISTANCE	(W)	S	PINDIS. (P1)	Dimensional Salience
		((R) W)	S	PINDIS (P2)	Dimensional Salience with Idiosyncratic Rotation.
	VECTOR	(V)	S	PINDIS (P3)	Perspective (Fixed Origin)
		(T V)	S	PINDIS (P4)	Perspective with Idiosyncratic Origin
	MIXED	(V W)	S	PINDIS (P5)	Double Weighting
N-WAY N- MODE	VECTOR	(V)	L	CANDECOMP	*N*-way Scalar Products
	Model Specification:* Individual: **Translation of origin **Rotation of axes** **Vector weighting** **Weights (dimensional)** **Point representation of subjects**		***Transform:* **Monotonic** **Linear** **Similarity**		

Table 7.1 *Analysis of 3-way and related data by MDS(X) programs*

of the chapter, by contrast, the order will proceed from the simplest to the more complex models.

We have already encountered PREFMAP as a program for mapping two-way, two-mode data into a user-provided configuration according to the vector and simple distance models (see 6.2.1 and 6.2.4). In this chapter these models are extended to include a weighted distance model (PREFMAP-II) and a rotated and weighted distance model (PREFMAP-I). As before, these are chiefly used to analyse sets of preference or, in general, similarity rankings or ratings when the user wishes to represent both the stimuli and subjects in the same solution.

The most common form of three-way data is two-mode, and the most popular form of analysis is the INDSCAL model. This model interprets differences between the subjects (third-way) as arising from differences in the weights (interpreted as importance or salience) ascribed to the dimensions of a common configuration. Because of its conceptual simplicity it makes a natural starting point for discussing more complex models, and is explained in 7.2.1.

An alternative approach to studying individual differences is to scale each matrix separately as an initial stage and then compare the configurations obtained. The PINDIS hierarchy of models provides a successively complex set of models for comparing configurations. There is no reason why the configurations should be obtained in this way; any set of configurations referring to the same set of objects, however obtained and of whatever dimensionality, may legitimately be input. The PINDIS models are discussed last, in section 7.4.1, due to their greater complexity.

This chapter also deals with three-way, three-mode and higher-way data, which may be analysed using a generalisation of the scalar products or factor models already encountered in the last chapter (e.g. MDPREF). The basic ideas of canonical decomposition, used to implement these models, are discussed in 7.2.2, following the exposition of the INDSCAL model which turns out to be a special case.

7.2 Individual Differences and Dimensional Salience

Three-way, two-mode data appear very frequently in the form of a set of (dis)similarity matrices. A typical example occurs when psychologists have subjects make pair-comparison estimates of the similarity between stimuli (such as colour chips) and wish to examine how individuals differ among themselves in the way they perceive colour. (This is the origin of the acronym: INdividual Differences SCALing.) Sociologists often have correlation matrices between a given set of variables for a number of different survey subgroups, and wish to see how the subgroup matrices differ (see 7.2.1.3). Plant ecologists may have co-occurrence matrices for a number of species, one for each of a number of sites chosen to differ on given criteria, and wish to inspect the differences between the sites.

Each example poses a similar methodological problem of aggregation. If the data for each element of the third-way differ to a substantial degree then there is little communality and it is hard to see how they are to be compared at all. If, by contrast, subjects differ in no systematic way but simply represent minor random fluctuations, then there is no point in making anything of the differences. However, if the data are simply pooled together as a single matrix at the outset then all information about differences—whether systematic or random—is lost.

Drawing on ideas developed by Horan (1969), Carroll and Chang (1970) propose the following way of thinking about such individual differences. Suppose each individual (group, or element in the third-way) makes use of a variety of attributes or dimensions in judging the stimuli (the exposition is easiest in psychological terms, but the model generalises easily to encompass any other sort of third-way element). Then define a master or *group space*, which consists of all the dimensions which the subjects happen to use. Each individual subject's space can now be thought of as a special case of the group space—as a reduction of the group space, since she is using some subset of the total available dimensions. This is termed the subject's 'private space'.

In Horan's original formulation, every individual was simply thought of as either using, or not using, each group space dimension, so each 'private space' could be represented by a sequence of 1s and 0s, indicating whether the subject used (1) or did not use (0) the dimensions of the group space. This 'all or nothing' approach was modified by Carroll and Chang by postulating that each subject attaches a *varying* (positive) weight to each dimension which represents the *degree* of salience (or importance, or attention or relevance or centrality) of that dimension to her judgments. So each individual i can be thought of as having an idiosyncratic set of weights, symbolised by $w_a^{(i)}$: the weight given to dimension a of the group space by individual i. These weights hence represent the way in which the subjects differ in the importance attached to each of the dimensions. An individual who attaches equal importance to each of the dimensions will have a set of weights of the same value, and it is such a subject whom the group space actually represents. Others by contrast will attach different weights to different dimensions of the group space and thus systematically distort the group space into the 'subjective metric' of their own private space.

The INDSCAL model presents a way of interrelating these 'private spaces' and provides one-way of comparing how subjects (or elements of the third-way) differ among themselves, but only, be it noted, by accounting for the individual differences in terms of differing weights being associated with the *same* dimensions: INDSCAL is explicitly a *dimensional* model.

*7.2.1 The INDSCAL model**

The Carroll-Chang model is described in full in their definitive 1970 article. A lucid and extended exposition, relating INDSCAL to other forms of three-way scaling is given in Carroll and Wish (1973) and a wide range of applications is discussed in Wish and Carroll (1974). Elementary treatments are given in chapter 4 of Kruskal and Wish (1978), in Spence (1978) and in the MDS(X) documentation. In this section we shall concentrate chiefly on the basic characteristics of the model and upon the interpretation of an INDSCAL-S solution. Further details of the estimation procedure in INDSCAL-S are contained in Appendix A7.2.

Before using INDSCAL-S or embarking upon interpretation of an INDSCAL solution, it is essential to understand clearly the characteristics of the *group space*, the *subject space*, the *private spaces* and their interrelationships.

*Hereafter, INDSCAL refers to the model and INDSCAL-S to the version of the program in MDS(X).

(i) The *group space* (denoted **X**) consists of a configuration of p stimulus points in a user-chosen number of dimensions r. The orientation of the axes of this space are *uniquely determined* in the sense that any change in their orientation destroys the optimality of the INDSCAL solution. The INDSCAL axes are often found to be readily interpretable.

This group space acts as the 'reference configuration' to which all the subjects' private spaces may be referred and from which they may be all derived. The group space need not in fact describe any actual subject, and the configuration should not itself be interpreted if it turns out that it is simply a compromise between the configurations of groups of subjects with very different patterns of individual weights.

(ii) The *private space of each subject* i (denoted $\mathbf{Y}^{(i)}$) is a configuration of the p points in r dimensions. Within each private space, distances between stimuli are straightforwardly Euclidean.*

(iii) The *subject space* (denoted **W**) is simply a useful graphical way of comparing subjects in terms of their sets of dimensional weights. It has the same dimensions as the group space and each subject is represented by a vector located by the value of the weights on each of the dimensions.

These basic ideas are illustrated in Figure 7.1 by reference to a simple artificial 2-dimensional example (see also Carroll 1972, p. 105 et seq., Carroll and Wish 1973, p. 57 et seq., and Kruskal and Wish 1978, p. 61 et seq. for similar examples). In this expository example, there are 3 objects and 16 subjects, so the data would consist of 16 lower triangular matrices between the 3 objects. The overall 2-dimensional group space configuration, **X**, consists of 3 points which make an equilateral triangle (representing equal distance between the objects). The private spaces, $\mathbf{Y}^{(i)}$, for subjects 1 and 2 are also presented. Note that in the private spaces the configuration of points no longer forms an equilateral triangle but rather an isosceles triangle (two sides remain the same length but the third is foreshortened). Clearly, *the distances between stimulus points are different within each private space.* The two private spaces are nonetheless related: they may be derived from the reference group space by a simple process of differentially stretching or shrinking the axes of the group space by the square root of the subject's 'importance weights'. In other words, the co-ordinates in the private space (say, for subject 1) are simply a weighted version of the group space co-ordinates. To obtain subject i's private space, we take the co-ordinates of the p stimulus points on the 1st dimension of the group space (x_{ji}) and rescale (stretch or shrink) them by the square root of subject i's weight for this dimension ($\sqrt{w_a^{(i)}}$): that is,

$$y_{ja}^{(i)} = \sqrt{w_a^{(i)}}\,x_{ja}$$

Then the distance between the stimuli j and k in subject i's private space will be:

$$d_{jk}^{(i)} = \sqrt{\sum_a \left(\sqrt{w_a^{(i)}}\,x_{ja} - \sqrt{w_a^{(i)}}\,x_{ka}\right)^2}$$

*In INDSCAL the private space of each subject is *estimated* as a distortion of the group space directly from the data. In PINDIS, by contrast, each subject's 2-way data are *first scaled* and then input in the form of configuration co-ordinates into the program.

or, in simplified form (taking the weight outside the squared term):

$$d_{jk}^{(i)} = \sqrt{\sum_a w_a^{(i)} \left(x_{ja} - x_{ka} \right)^2}$$

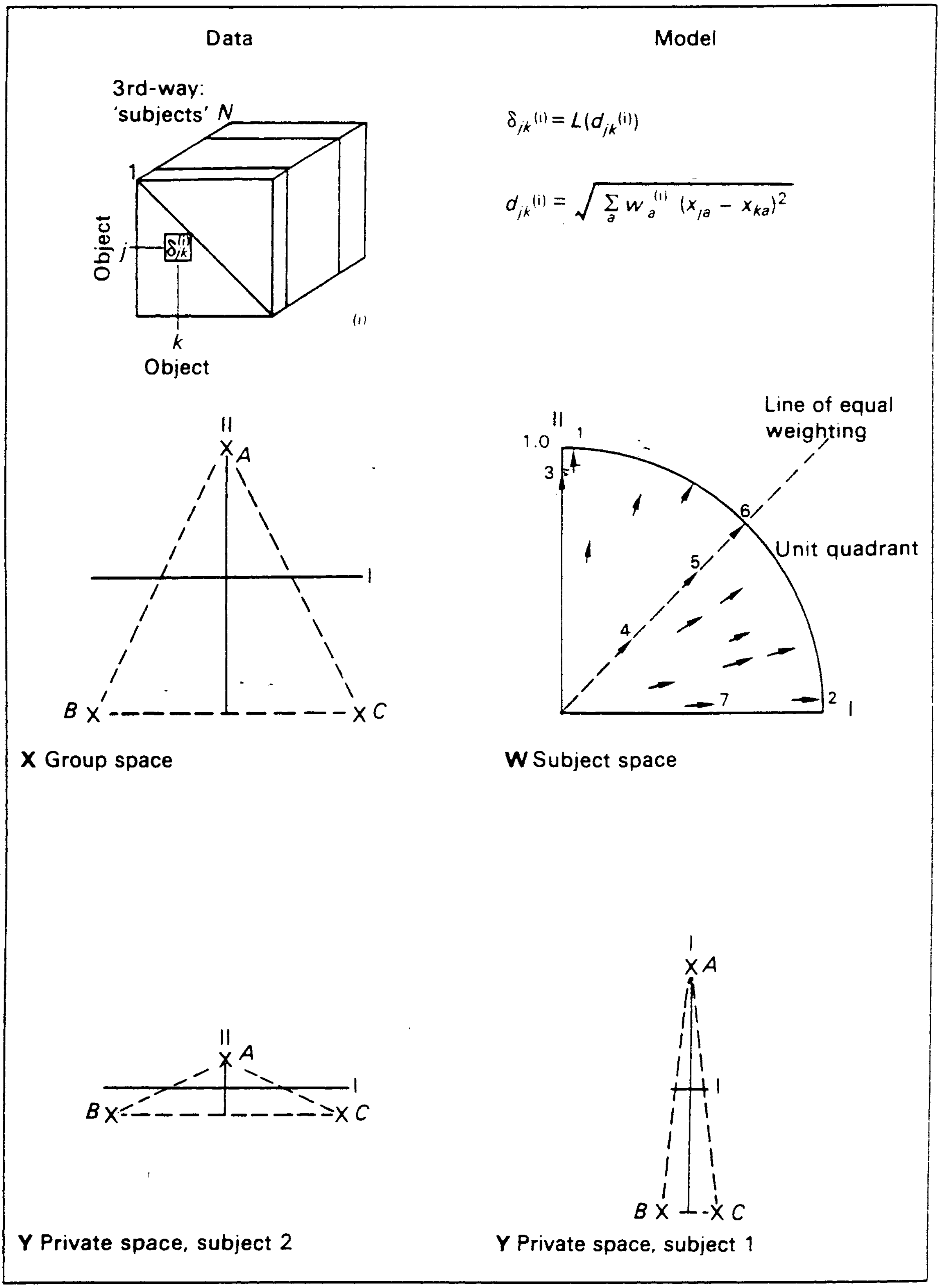


Figure 7.1 *Basic INDSCAL model*

This last equation gives the general form of the INDSCAL *and other weighted distance models:* A 'subject' *i*'s judgment of the (dis)similarity between objects *j* and *k* is taken to be a (linear) function of the overall distance between stimuli *j* and *k* in the group space, after that space has been differentially rescaled (stretched and shrunk) by the subject's set of weights into the 'subjective metric' of the subject concerned.

7.2.1.1 The group stimulus space: its properties and interpretation

The group stimulus space functions as the basic reference configuration from which the private configurations of individual subjects can be derived by differentially shrinking or stretching the dimensions by the (square root) of the corresponding weights.

The INDSCAL dimensions actually represent the (orthogonal) directions where the variation among subjects is the greatest: it is for this reason that they are normally easy to interpret. These dimensions are uniquely identified, in the sense that if the original dimensions are rotated and new subject dimension weights calculated, the resulting solution will explain the subjects' data less well than the original solution.† If it turns out that the extent of individual differences is not great, then such a reduction in explained variance is likely to be small, but in the normal way the reduction is usually fairly substantial. Unless there are compelling reasons of interpretability or little subject variation, the INDSCAL axes should be regarded as fixed.

In most MDS solutions encountered so far, the final configuration is rotated to principal axes—that is, dimensions are chosen which have the statistically convenient property that co-ordinates of the points are not correlated across dimensions. *This is not (generally) true of an* INDSCAL *group space:* the dimensions of greatest subject variation will usually give rise to a configuration where the co-ordinates of the stimulus points are to a greater or lesser extent correlated.‡ The information about the extent of this correlation between pairs of axes is contained in the output from INDSCAL-S in the matrix of scalar products between dimensions ('sums of products') for matrix 2.

The INDSCAL group stimulus space configuration should therefore be interpreted with caution: strictly speaking it represents a subject who weights the dimensions equally, and if a significant number of subjects' weights depart markedly from equality then there is a danger of trying to interpret a configuration which is in no sense representative. That said, methods for external interpretation of INDSCAL *dimensions*—and especially linear property-fitting (see 4.4.1)—are particularly appropriate, since the dimensions are *not* arbitrary and it is important to try to tie down their meaning as accurately as possible. Good examples of the use of

†The unique orientation of axes in the INDSCAL model means that the solution is unique up to permutation of axes, which is equivalent to saying that the only permissible rotation of the dimensions which preserves all significant information is through multiples of 90°. However, the actual size of the configuration is arbitrary, and is therefore normalised so that the variance of the projections on each of the co-ordinate axes is unity and the centroid of the configuration provides the origin (Carroll and Wish 1973, p. 30).

‡An option SOLUTIONS (1) exists in the MDS(X) version to obtain a solution where the axes are as close as possible to being uncorrelated. Such a solution will normally be sup-optimal compared to the ordinary solution.

property fitting to validate or confirm the interpretation of INDSCAL dimensions occur in the classic Carroll and Chang (1970) paper and elsewhere.

INDSCAL-S can also be used in an external mode if the user provides the program with a group stimulus space configuration (which remains fixed in orientation) and the INDSCAL analysis then concentrates entirely upon estimating from the subjects' data the subject weights for this configuration. (External use is achieved in INDSCAL-S using the FIX POINTS (1) option). External analysis of this sort has two main uses: (i) to scale a large number of subjects' data and (ii) to compare a number of different data sets by referring them to a common reference configuration. Thus if the user has, say, 500 matrices for analysis, it is sensible to choose a manageable sample of those matrices and scale them. The resulting group stimulus space can then be fixed, and the subject weights can then be estimated for as many batches of subject matrices as desired.* An example of the second use occurs where a replication has been made of a previous study and the researcher wishes to investigate the extent to which her subjects' data compare to the weights obtained in the earlier study. The original group space configuration is fixed under this option, and the subjects' weights may then be estimated and compared to those of the original study.

7.2.1.2 The subject space: its properties and interpretation

When subjects' data are input to INDSCAL-S they are normalised to have equal weight, which has the effect of giving each subject's data equal influence on the solution. This fact, in conjunction with the normalisation of the group stimulus space described above, gives rise to several nice properties of the subject space which are useful to bear in mind when interpreting an INDSCAL solution:

(i) The subject's weight on a dimension is (approximately) equal to the correlation of the intervals between stimulus co-ordinates on that dimension and the corresponding pairwise dissimilarity values in the subject's data

(ii) Consequently, the *squared* subject's weight on a dimension is (approximately) equal to the proportion of variance in the subject's data that can be accounted for by that dimension (Wish and Carroll 1974, p. 452).

(iii) Therefore, the squared distance from the origin of the subject space to a subject's point in that space is (approximately) equal to the proportion of variance in the subject's data accounted for by the full INDSCAL solution.

If the dimensions of the INDSCAL solution are uncorrelated, then the word 'exactly' replaces the word 'approximately' in the above three sections. Thus in the subject space portrayed in Figure 7.1, subjects 4, 5 and 6 provide an example of subjects who weight the dimensions equally: they differ only in the fraction of their data explained by the model, with the data of subject 6 perfectly accounted for. Similarly, subjects 7 and 2 have the same pattern, giving virtually exclusive salience to dimension I, whilst subject 3 uses only dimension II. Looking at the pattern in terms of goodness of fit, the data of subjects 1, 6 and 2 are totally accounted for, whilst those of subject 4 are very poorly explained.

*See Coxon and Jones 1979, pp. 54–9, and especially T3.17, for an example using a balanced set of 68 matrices to obtain the group space configuration by reference to which 286 subjects' subject weights were estimated.

Note that only *positive* weights are allowed by the INDSCAL model. If, as occasionally happens, a very small negative weight occurs in may be considered as approximation to a zero weight; if it is substantial it can only be interpreted as indicating that the basic model does not hold for the data of the subject concerned.

The significant information in the subject space is contained, then, (1) in the *direction* in which a point is located from the origin, since any points lying on line from the origin have weights in the same ratio, and (2) in the *distance* (of a subject vector) from the origin, representing how well the subject's data are explained by the model.

Before embarking on any systematic analysis of INDSCAL subject weights, it is important to know something of the stability of INDSCAL solutions (see Jones and Waddington 1973; MacCallum 1977).

(i) Simulation studies show that, even in circumstances of high error in the data, recovery of the group space configuration and its dimensional orientation is excellent, but that

(ii) the stability of the subject space is far less stable and much more subject to fluctuation in the presence of error.

The temptation to use cluster analysis on subject weights should be strongly resisted: the separations of subject points are in no sense ordinary distances and their location is far from stable. The question of whether any linear procedures such as ANOVA and its multivariate variants should be used on INDSCAL weights remains contentious. MacCallum (1977) and others often strongly counsel against their use; Carroll and others think that a more lenient approach is called for.

Usually the user will want to compare subjects in terms of the patterns of relative salience given to dimensions. This is best done by concentrating on the angular separation between subject vectors: the smaller the angle of separation, the more similar is the pattern of weights. In the two-dimensional case, it is usually a simple matter to see closely collinear 'sheaves' of subject vectors in the subject space, and such bunching can also be detected visually in three dimensions. Beyond that, statistical analyses of different subject vectors should be used (see Mardia 1972; Coxon and Jones 1979, pp. 128–36 for use in an MDS context). A simple alternative for two-dimensional data is simply to take the ratio of the weights for each subject and, since the distribution of such ratios is usually markedly positively skew, it often makes sense to correct this by taking a logarithmic transformation of the weight ratios.

An alternative to Carroll and Chang's representation of subject weights has been suggested recently by Young (1978). The Young Plot allows the amount of variation explained to be represented independently of the relative salience of the subject weights, and is illustrated in Figure 7.3b (p. 199).

The Young Plot

The Young Plot charts each subject in terms of two things—the relative salience ascribed to one dimension over another (on the horizontal axis) and how well the subject's data are fit by the model (the vertical axis). The first is measured by the ratio of the two-dimensional weights—which can be interpreted trigonometrically as the tangent of the angular separation between a subject's vector and the line of equal weighting in the conventional representation of INDSCAL subject space.* The goodness of fit is given simply by the squared correlation r^2 between the subject's original data and the values predicted by the INDSCAL model. This information is provided separately in an INDSCAL run.

The Young Plot and its construction from a set of subject weights is illustrated in Figure 7.3b and is very simple to read. Subjects located in the centre of the horizontal axis (such as the Labour group of voters in this example) weight the dimensions equally; the more that dimension I dominates over II the further left the subject point is, so the non-voters group has the most dominant weight for dimension I and the Conservative voters group has the most dominant weight for dimension II. The goodness of fit is simply read up the vertical axis. In this example the greatest differentiation is between the 'other parties' group whose data are not well explained (being largely Scottish and Welsh Nationalist party supporters they are presumably dancing to a different piper) and the others.

The most important advantages of the Young Plot are that it gives accurate representation of patterns of dimensional weighting and of goodness of fit independently of dimensional correlation, and concentrates the user's attention onto the angular separation (relative salience) of patterns of subject weights rather than on the proximity of points portrayed somewhat misleadingly in a conventional subject space. The Young Plot can also be modified in various ways—to portray patterns of three-dimensional weights, or to compare relative salience of weights with any other variable of interest (see Coxon and Jones 1980, p. 59 et seq.).

7.2.1.3 An example: political party imagery

Alt et al. (1976) carried out a survey of 2,462 British voters after the 1974 British election. The questionnaire included 20 attitudinal items—political party features (items 1–7), the parties' handling of contentious issues (items 8–10), blame (11–12), taxes and pensions (13–14) and policy positions (15–17). These are reproduced in Table 7.2. Each pair of items was cross-tabulated and the association between them measured by Goodman and Kruskal's gamma, which preserves weak monotonicity of the item categories (see 2.2.2 above). The respondents were divided into five subgroups, viz

- A Conservative voters
- B Labour voters
- C Liberal voters
- D Other voters (principally Scottish and Welsh National Parties)
- E Non-voters

Each of these subgroups were then treated as a 'pseudo-subject', and gamma coefficients were calculated for each subgroup, hence providing a (5 × 20 × 20) array for input to INDSCAL. The group space configuration is given in Figure 7.2 and

*The tangent of the angle which the subject vector makes with the first dimension (tan θ_I) is defined as the ratio of the weight on dimension II to the weight on dimension I. Tan (θ_I − 45°) measures this predominance of dimension II over dimension I as a deflection (angular departure) from the line of equal weighting.

Item No.	*Symbol*	*Title*
1	K	Keeps/breaks promises
2	D	Divides/unites country
3	B	Bloody-minded/reasonable
4	G	Good for one/all classes
5	E	Extreme/moderate
6	Ca	Capable/not capable
7	SF	Stands firm/gives way
8	P	Prices
9	M	Miners' strike
10	S	Strikes
11	PB	Blame for prices
12	MB	Blame for miners' strike
13	T	Taxation
1	Pe	Pensions
15	CM	Common Market
16	N	Nationalisation
17	SS	Social services
18	W	Wage controls
19	C	Communists
20	R	Reliability

Table 7.2 *Items in political party imagery study* (Alt et al. 1976) (Reproduced by permission of the journal *Quality and Quantity*)

the subject weight plots are given in Figure 7.3. Alt et al. identify dimension I as 'image consciousness' (by which they mean an ideologically-based concern with both political style and performance) and dimension II as 'policy consciousness' (concerned primarily with welfare and related policy issues). Note from the shape of the group stimulus space configuration that the two dimensions are clearly positively correlated. The authors do not provide this information, but our estimate is $r_{\text{I, II}} = 0.23$.

Further interpretation of the group space should wait upon inspection of the subject weights (Figure 7.3). Even a cursory examination of the subject space (a) and more obviously of the Young diagram (b), shows very considerable differences in the goodness of fit and in the relative salience of two-dimensional weights between the subgroups.

But just how significant are these relative differences in weights, given what we know of the relative instability of INDSCAL weights? Alt et al. use an unusual form of internal validation. They divide their subjects into a number of pseudo-groups based upon 'irrelevant' factors (such as male/female) and random criteria (exclusive but randomly constructed subgroups and overlapping random subsamples of subjects) and proceed to estimate weights for each group, keeping the reference configuration fixed. Only if the voter subgroup differences exceed the

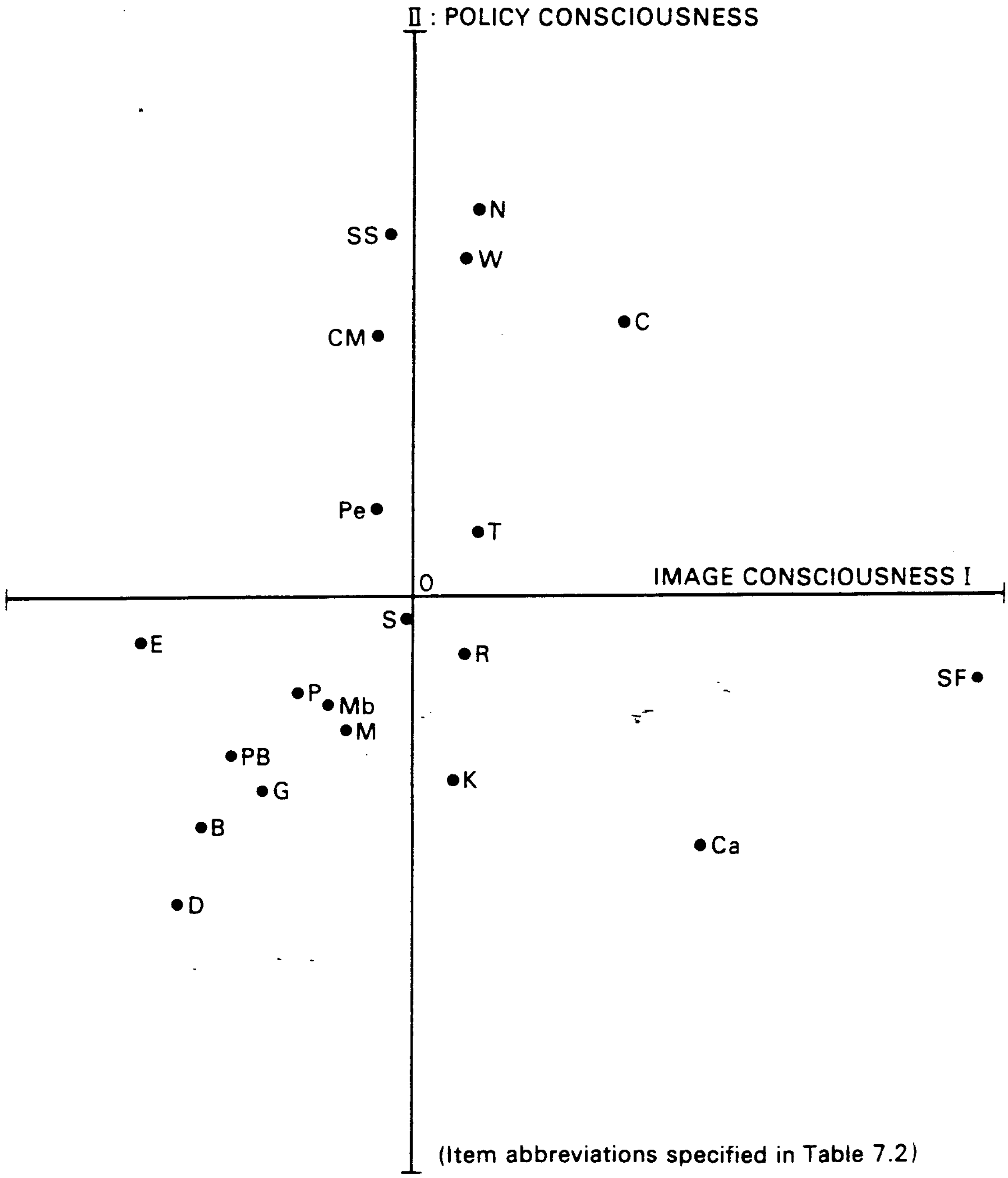
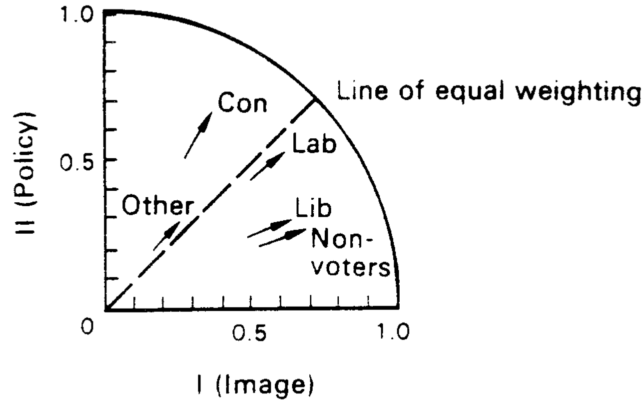


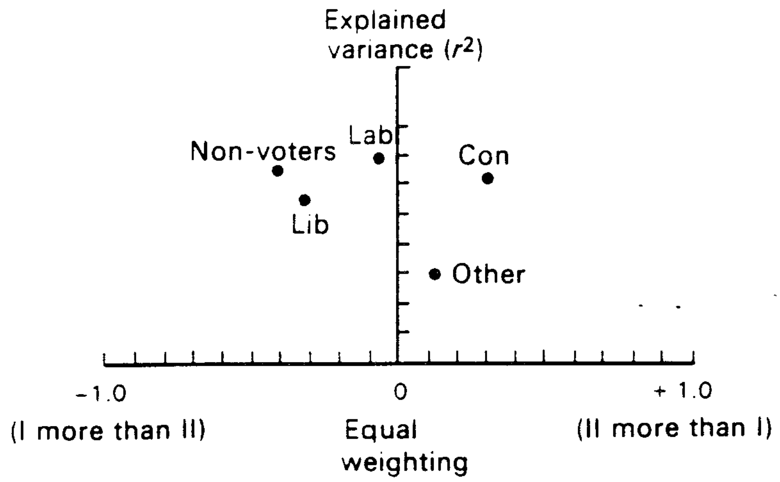
Figure 7.2 *INDSCAL group space: political imagery study*

random differences do they consider them to be sufficient to merit separate treatment. It turns out that the differences among voter subgroups greatly exceed those found for the random groups, especially on the first dimension. The authors then construct the private space for each subgroup (pp. 308–9), and comment:

> The relative unidimensionality of the items for Liberals and Non-voters is apparent. For them, big differences between items only occur between those most clearly reflecting 'style' and 'performance'. In contrast, voters for the two major parties use both dimensions in differentiating items, and the previously mentioned differences between these groups are also evident. Particularly striking is how small the group space looks—how undifferentiated all the items appear—to 'voters for other parties'. These results are substantively not necessarily surprising: the items were, after all, re-scaled as inter-party



(a) INDSCAL Subject Space



(b) Young diagram of (a)

Group	(i) Dimension weights I	II	(ii) Angle w.r.t. dim. I	(iii) Deflection from line of equal weighting*	(iv) Goodness-of-fit (r^2)
A. Conservative	0.32	0.60	61.93°	0.30	0.62
B. Labour	0.57	0.51	41.82°	-0.06	0.69
C. Liberal	0.56	0.28	26.57°	-0.33	0.55
D. Other	0.21	0.26	51.07°	0.11	0.31
E. Non-voter	0.64	0.26	22.11°	-0.42	0.66

*Tan (θ_I − 45°)

Figure 7.3 *Subject weights plots: political imagery study*

comparisons. The clarity and parsimony with which INDSCAL recovers this property of the group space is, nevertheless, impressive.

(Alt et al. 1976, p. 310)

This example of the use of INDSCAL shows well how, with a little initiative and imagination a model developed within an individually-based psychological tradition can be adapted with considerable success to analyse survey data referring to several thousand respondents (see also Coxon and Jones 1977).

7.2.2 Canonical decomposition (CANDECOMP)

Canonical decomposition (CANDECOMP hereafter) is a very general model, of which INDSCAL, two-way scalar products (factor, vector) models and multiplicative conjoint analysis are the better-known special cases. The N-way CANDECOMP model states that an N-way, N-mode array of data ($z_{ijk \ldots n}$) can be decomposed into a separate set of numerical 'effects' for each way, which combine multiplicatively within a dimension, and additively across dimensions:

$$z_{ijk \ldots n} \simeq \sum_a v_{ia} w_{ja} x_{ka} \cdots y_{na}$$

where the v_{ia}, w_{ja}, x_{ka} and y_{na} are the numerical 'effects' or 'weight' parameters to be estimated (one weight per dimension), and $\simeq$ means a least-squares estimation or approximation.*

An example of a 4-way CANDECOMP application would be an investigation of the connotative meaning (Osgood 1965) of a set of political concepts, where the investigator had 60 subjects evaluate 12 concepts on 25 bipolar rating scales on 4 different occasions. In this case, having decided upon an appropriate dimensionality for the analysis, the researcher could use CANDECOMP to estimate the numerical effect or weight to be ascribed to each element in each of these four ways, assuming that the weights combined together multiplicatively (in a manner akin to scalar products in the 2- and 3-way case).

A quite common variant of CANDECOMP occurs when two ways of the data (conventionally known as the second and third ways) refer to the same set of stimuli, and the analysis then becomes a higher-way generalisation of INDSCAL. Continuing the earlier example, the researcher might decide to ignore individual differences and run correlations between the 25 rating scales of each of the 12 concepts for each of the 4 occasions, thus forming a 12 × 4 stack of 2-way symmetric correlation coefficients between the rating scales. Labelling the occasions as the 1st way and the concepts as the 4th way, we now have a (4 × 25 × 25 × 12) data array. Since this is an extension of INDSCAL the researcher would normally require that the effects of the stimuli (rating scales) be constrained to be the same, as in other INDSCAL analyses. (This is effected by stipulating SET (1) on the PARAMETERS card, but unless this equalisation condition is imposed, the weights for each separate way will be estimated.)

7.2.2.1 Normalisation of the data and solutions

In INDSCAL, each subject's data are normalised before the analysis, so that each subject is given the same influence in the solution. This is not the case in CANDECOMP: any modification or normalisation of the data the user wishes to make must usually be performed before input. In many cases the actual unit of measurement of the data is in any case arbitrary, and may vary from way to way in the case of higher-way data. But since CANDECOMP does not equalise data at the outset, these arbitrary aspects will affect the analysis. Modification of data before input in order to control for arbitrary characteristics can be illustrated by a now famous example drawn from two-way data scaling. Coombs (1964, pp. 464–6) analyses the co-citation patterns of psychological journals: how often articles in Journal X cite articles from Journal Y. Now journals differ not only in the extent to which they

*Three-way CANDECOMP is closely allied to Tucker's Three-mode Factor Analysis developed in the mid-1960s and related by him to three-way scaling in his 1972 paper (Tucker 1964, 1972). The interrelationships are also discussed in Carroll and Wish 1973.

make citations, but also in the number of articles they publish—and both of these characteristics may be viewed as irrelevant, arbitrary, aspects of the data if one is interested in studying the pattern of *relative* citation patterns between the journals. He therefore decided to remove both the row effects (representing the differing size of the journals and the conventions about the appropriate number of citations they make) and the column effects (representing the overall popularity or frequency of citations of each journal) by the device of 'double-centring' the matrix. This leaves only the interaction effects to be scaled. (Row and column centring of the input matrix are options within the MDPREF program.) In higher-way data matrices, the same problem arises and a facility is provided for triple- (or higher-) centring the data matrix, using CENTRE (1). This option should normally be chosen, unless the user has reason to keep the original values of the data. (See Gower 1976 and Tagg 1979 for a discussion of this question in the 3-way case.)

The *solution configuration*, by contrast, is normalised in both CANDECOMP and INDSCAL. In CANDECOMP, the solution weights *for all but the first way* are normalised so that the sum of the squares of the weights on each dimension is unity, (Carroll and Wish 1973, pp. 30–1).* Moreover, in CANDECOMP the origin of the spaces is not constrained to be at the centroid, as occurs in the INDSCAL group space, unless the data have been initially centred.

7.2.2.2 *Using* CANDECOMP

Three-way, three-mode CANDECOMP and four-way three-mode INDSCAL are the most obviously useful variants of the N-way model. Beyond these the user is advised to proceed with caution, since very little is known about the properties and stability of the solution, although one example of four-way three-mode INDSCAL has been alluded to in the literature (Carroll and Wish 1973, p. 98), where subjects (1) made pairwise similarity judgments between nations (2 and 3) by a variety of methods (4). The authors themselves counsel users to remain with three-way analysis:

> It is not clear when or whether these higher (than three) way models are appropriate. In their current form they impose strong, and usually unrealistic constraints on the data ... In most cases it is more appropriate to simply concatenate the N-2 nonstimulus modes into a single mode, and then do a three-way analysis. (For example, if there are n_1 subjects and n_2 conditions there would be $n_1 \times n_2$ 'pseudosubjects' defining the third way of the anlysis; there would be one two-way matrix for each of those 'pseudosubjects'. Another alternative would be to do separate three-way analyses for each of the n_2 conditions).
>
> (Carroll and Wish 1973, p. 99)

The effect of this warning has perhaps been too discouraging: few published examples of CANDECOMP as yet exist.

One of the earliest is presented in Green and Rao (1972, pp. 45–8), where they use 3-way CANDECOMP to assess the congruence between a set of 9 configurations derived from the same data, but each obtained by a differing scaling method. They input the 2-dimensional co-ordinates of 9 scaling solutions referring to 15 stimuli

*For this reason (because all differences in the sum of squares are absorbed in the first way) users are advised when using the program to identify the first way with the one which represents for them the most important source of variation.

(breakfast foods), choosing the 9 solutions to be the first way, the 15 stimuli to be the second way and the 2-dimensional co-ordinates of the original solution to be the third way. From an analysis of the 2-dimensional CANDECOMP solution weights for each way, they conclude:

(1) The nine scaling algorithms represent essentially the same configuration after allowing for differential stretching along the axes.
(2) The stimulus configuration is a composite of the original configurations, but one where the axes are interpretable and the pattern of points is very similar to the position in the original configuration.
(3) The weights for the original orientations of the configurations again confirms 'the closeness of the CANDECOMP orientation to the original orientations'.

In a later section we shall see that PINDIS (a program which had not been developed at that time) provides a more precise way of investigating how transformations such as axis-stretching allow the relationship between configurations to be examined more precisely.

Canonical decomposition is clearly a very general and wide-ranging procedure which can also be adapted to estimate the parameters of a wide variety of models, including latent structure analysis.* But its very flexibility and power mean that the user should exercise caution in its use and not let enthusiasm outrun understanding.

7.3 Comparing Configurations

When using MDS, it is only a matter of time before the user wishes to compare two or more configurations. If a study has set out to replicate a previous one, then it is important to know the ways, and the extent, to which the current MDS solution resembles that of the original study. More often, the user has employed more than one variant of MDS on the same data, or has scaled the data of different subgroups of subjects, and the issue of similarity between the resulting configurations once again arises.

Before attempting to compare configurations the user should be sufficiently persuaded that in each case the fit of the solution to the data is good enough to warrant proceeding any further.

The answer to the question, 'How similar are two configurations?' depends on two things:

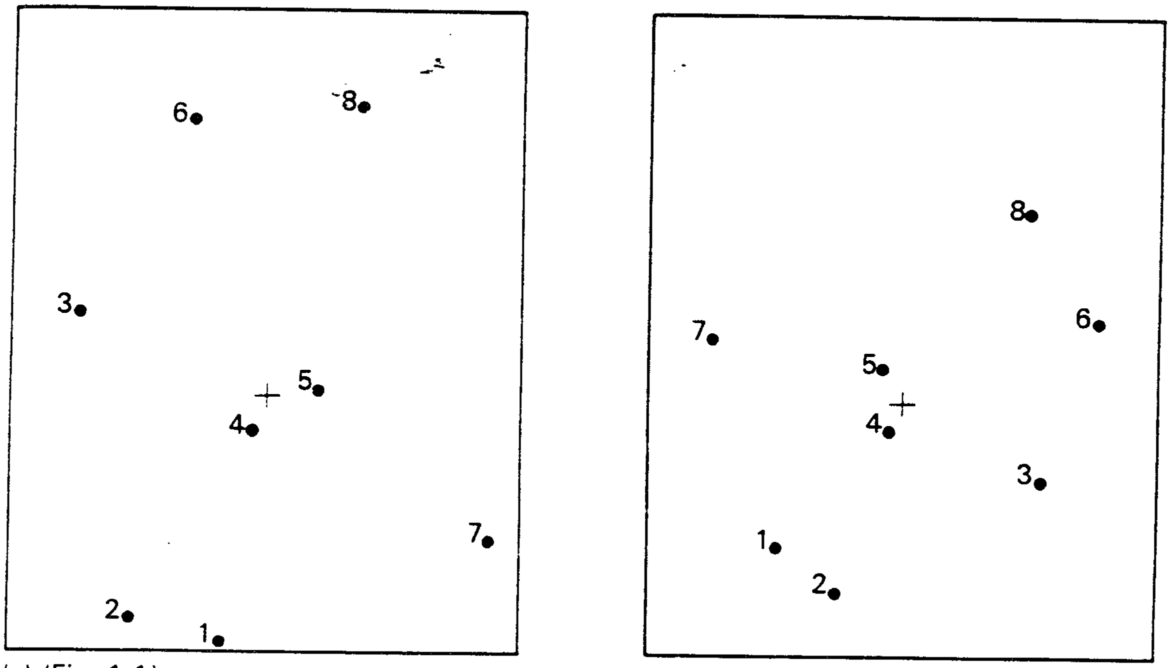
(i) what aspects of the configuration are considered *relevant*, and
(ii) what properties of the configuration are *unique*.

*Latent structure is a family of models developed to analyse sets of dichotomous response patterns (Lazarsfeld and Henry 1968; Fielding 1977). In these models, observed patterns are thought of as arising from the multiplicative combination of the latent probabilities of giving a positive answer to each item, where these probabilities differ according to the position of the subject in the space. In the latent class model, the 'space' consists of a set of partitions or classes. Whilst the ideas of latent class analysis are very appealing, estimation of the model parameters have proved to be very difficult. However, Carroll (1975) shows how the latent class model may be seen as a special case of canonical decomposition, where the dimensionality may be identified with the number of latent classes and the parameters of the model may be estimated in a straightforward manner from the CANDECOMP weights. Paradoxically, the CANDECOMP estimation is better than the procedure suggested by Lazarsfeld and Henry, despite the fact that it is minimising a different and less obviously appropriate badness-of-fit function.

For instance, if one configuration were simply twice the size of the other but in every other way identical, it is unlikely that most researchers would consider this a relevant difference. Hence one would normally wish to compare *relative* rather than absolute distances. If dealing with solutions obtained from a Euclidean distance model, it is unlikely that the orientation of the configuration will be considered relevant since any rigid rotation leaves distances unchanged. Similarly, the origin of the space would normally be treated as arbitrary, unless the model were a vector model (since change of origin alters scalar products and hence alters the angles separating the vectors, cf. Appendix A2.1) or unless a facet analysis had been employed in interpretation and the point chosen as the origin was substantively meaningful.

In actual fact, many users of MDS and factor analysis resort to crude and misleading methods for comparing configurations—such as restricting attention to the first two dimensions of a solution and simply 'eye-balling' the configurations, hoping that salient differences will in some way reveal themselves. Unfortunately, differences which are irrelevant can often make identical configurations *appear* to be very different. This is illustrated in Figure 7.4, where the two-dimensional configuration of seriousness of offences given in Figure 1.1a, is first reproduced as Figure 7.4a, and then submitted to a series of transformations which preserve relative distances. In terms of keeping the same relative distances these two configurations are identical but they certainly *look* different. So appearances are not a reliable guide as to how two configurations are alike. How, then, does one go about comparing them?

Sometimes transparent acetate sheets are used to compare two-dimensional configurations—one configuration is first copied onto a transparent sheet and laid



(a) (Fig. 1.1)

(b): (a) subjected to restricted similarity transform:
(i) clockwise rotation through 40°
(ii) reflection in axis I
(iii) rescaling by factor of 0.75

Figure 7.4 *Two 'identical' configurations* on top of the other configuration. By moving the sheet in a circular manner (rotations) and/or flipping it over (reflections) the one configuration can be moved into maximum apparent congruence with the other. This procedure has its uses, but even this expedient cannot deal with differences of scale, and is obviously restricted to two-dimensional situations. Moreover, we still need some explicitly defined index of configurational similarity if the comparison is to be anything more than approximate or if we need to compare more than two configurations. How, then, should two or more configurations be compared?

(i) First, we must return to the question of what aspects count as significant and unchangeable information and what are irrelevant.
(ii) That decided, a method is needed which will bring the configurations into the closest possible conformity with each other.
(iii) Finally, some indication is needed of how closely the configurations correspond to each other, preferably by using an appropriate measure of goodness of fit.

We shall take up each of these points in turn.

First of all, in this section (and indeed, as far as 7.4.1) we will assume that two configurations are considered identical if they only involve differences of:

(1) *scale* (how large the actual configuration is);
(2) *orientation* (rigid rotation and/or reflection of axes); and
(3) *origin* (the zero point of the space).

Consequently, configurations may be shrunk or expanded at will (1), moved—rigidly rotated—through any angle (2), and may have the origin translated to any point in the space (3) in order to get them into greater conformity with each other. The value of any index of similarity between configurations should remain unchanged whenever these operations are performed.

7.3.1 Geometric transformations of a configuration

The geometry of these three basic operations, which taken together define an 'extended similarity transformation', is described in Appendix A7.1. Users who are unfamiliar with them should read it carefully before proceeding further. Two transformations are particularly important in examining the similarity between configurations and in moving them into closest conformity. Both keep the comparative or relative distances in the configurations unchanged:

Configuration transformations which preserve (Euclidean) distances
1 *Extended similarity:* involves rotation, translation and rescaling
2 *Restricted similarity:* involves rotation and rescaling (no translation of origin)

Usually it will not be possible to transform two (or more) configurations into an identical structure, and it is therefore necessary to define an index of how similar two configurations are. All commonly encountered indices can be thought of as being a function of the *distances* between the points in the two configurations. Gower (1979, 1980) has provided an excellent review of such indices by examining the form of the function relating the distances in the two configurations. One obvious measure for assessing the similarity between two configurations is the product moment correlation* between the distances involved, and this is commonly used. A related measure of similarity between the configurations is called S by Lingoes (Lingoes and Schönemann 1974, p. 436). This has the nice properties that its value depends neither upon the number of points, nor upon the number of dimensions, nor on the scale of the configuration. Consequently it can be used to answer the question, 'Does configuration X match Y better than X matches Z?', which will be a central issue when we come to compare several configurations. An even simpler measure, which is just the squared linear correlation r^2 between the co-ordinates of the configuration X and Y which have been brought into maximum conformity, is also frequently used. It is directly related to S and shares its desirable properties.†

7.3.2 Comparison using Procrustes analysis

> *Procoptes, better known as Procrustes, 'The Smasher', caught (travellers) on the borders of Athens and by stretching or pruning made them an exact fit to his lethal bed.*
>
> *(Kirk 1974, p. 153)*

As originally employed in comparing configurations, simple Procrustes analysis consisted of moving two configuration matrices **X** and **Y** into closest conformity, allowing only rotation and reflection. Procrustes analysis was later extended to include rescaling and translation of origin. This more extended usage, which will be employed here, is termed generalised Procrustes analysis (Gower 1975) and is illustrated in Figure 7.5. Here three configurations (A, B, C) of four points each forming a roughly rectangular shape have been positioned into closest conformity with each other and, since the configurations are not identical, a new configuration **Z** is then produced, defined by the location of the square points in Figure 7.5. This new configuration is an 'average configuration' in the sense that each of its points is a least-squares fit to the corresponding points of the original configurations. This best fitting configuration is called by a variety of names: the compromise, consensus, centroid (as in PINDIS) or group configuration (as in INDSCAL). From now on it is referred to as the centroid configuration, and denoted by **Z**.

The iterative computing procedure for producing the centroid configuration is described in Gower (1975, p. 43).

Procrustes rotation can be used on any number of configurations, and is commonly thus used.

7.4 Procrustes Rotation and Individual Differences Scaling

Procrustes rotation is not only useful for comparing configurations, it can also serve as a basis for individual differences scaling. The basic idea is simple. Given a set of configurations, we begin by moving them into closest conformity by

*Carroll (1972) has suggested the use of linear, monotone, non-linear (continuity) and canonical correlation coefficients for assessing different aspects of goodness of fit between two configurations, according to different criteria for assessing 'fit'.

†$S^2 = 1 - r^2(X, Y)$ is the equation relating the two measures, and both measures remain unchanged when a similarity transformation is performed on X and Y to bring them into maximum conformity.

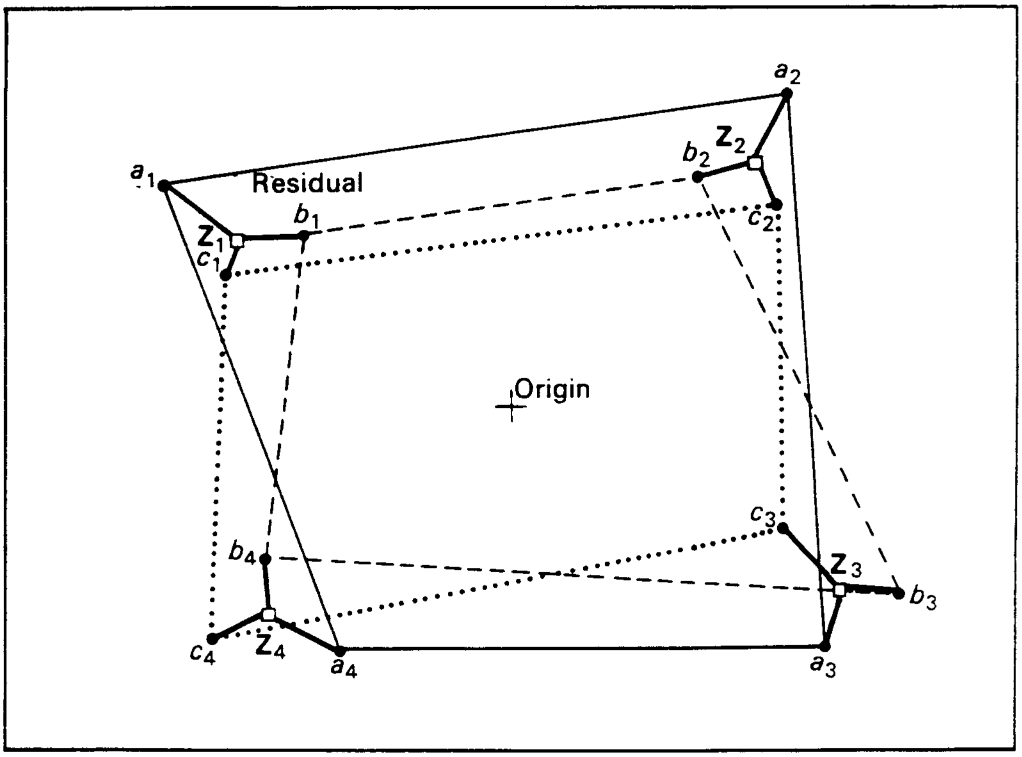
generalised Procrustes analysis, thereby producing the centroid configuration. Next we measure how closely each configuration fits the centroid configuration. Conceptually, the *badness* of fit of each configuration to the centroid can be defined in terms of the sum of squares of the residuals, i.e. $\Sigma (x_i - z_i)^2$ in Figure 7.5, but we shall follow Lingoes' example and use the squared correlation between the 'subject' configuration co-ordinates and those of the centroid configurations as a measure of *goodness* of fit.

7.4.1 The PINDIS models

Lingoes has developed a series of increasingly complex models based on Procrustes rotation, which have affinities with other models we have encountered. His set of models is known collectively as PINDIS (Procrustean INdividual DIfferences Scaling).

The data for PINDIS, unlike other three-way models, consist of a *set of configurations* obtained from previous scaling solutions. The configurations are first moved into maximum conformity by generalised Procrustes rotation. Since this procedure consists entirely of 'admissible' transformations (in the sense that they leave the relative distances unchanged) this basic general similarity 'model' (denoted P0) provides a yardstick or reference point for subsequent models which do *not* leave original relative distances intact.

Beyond P0, the models all involve so-called 'inadmissible' transformations—



(Adapted, with permission, from Gower 1975)

3 configurations of 4 points, with a common origin: A denoted by ———
B denoted by - - - - -
C denoted by ··········

Best fitting Procrustes ('centroid') configuration is given by Z (z_1, z_2, z_3, z_4)
It minimises sum of squares of residuals (denoted by thick lines).

Figure 7.5 *Procrustes analysis of 3 configurations* i.e. operations which change the original relative distances in some systematic way in order to obtain a better fit between the original configuration $\mathbf{X}_i$ and the reference centroid configuration $\mathbf{Z}$. At this point it should be stressed that:

(i) the centroid configuration is defined somewhat differently in each model unless the user wishes to keep it unchanged throughout (in this latter case, PINDIS becomes an 'external' analysis, an option effected in the MDS(X) version by using the READ HYPOTHESIS command);
(ii) at each stage (in each model) the individual configurations are moved into optimal fit to the centroid configuration, using admissible transformations.

The interrelations between the PINDIS models are illustrated in Figure 7.6. A brief simplified resumé will be given here, and separate models are discussed in greater detail in subsequent sections. In the diagram, each model is represented by a box. The top half indicates the usual name of the model, and the bottom half indicates what operations are performed on the centroid matrix (Z) and what 'individual differences' parameters are estimated in the model. Arrows are drawn upward from less general to the more general models. Note that the models do *not* form a strict order, but rather two parallel hierarchies, i.e. the *distance* models (P2, P1, P0) and

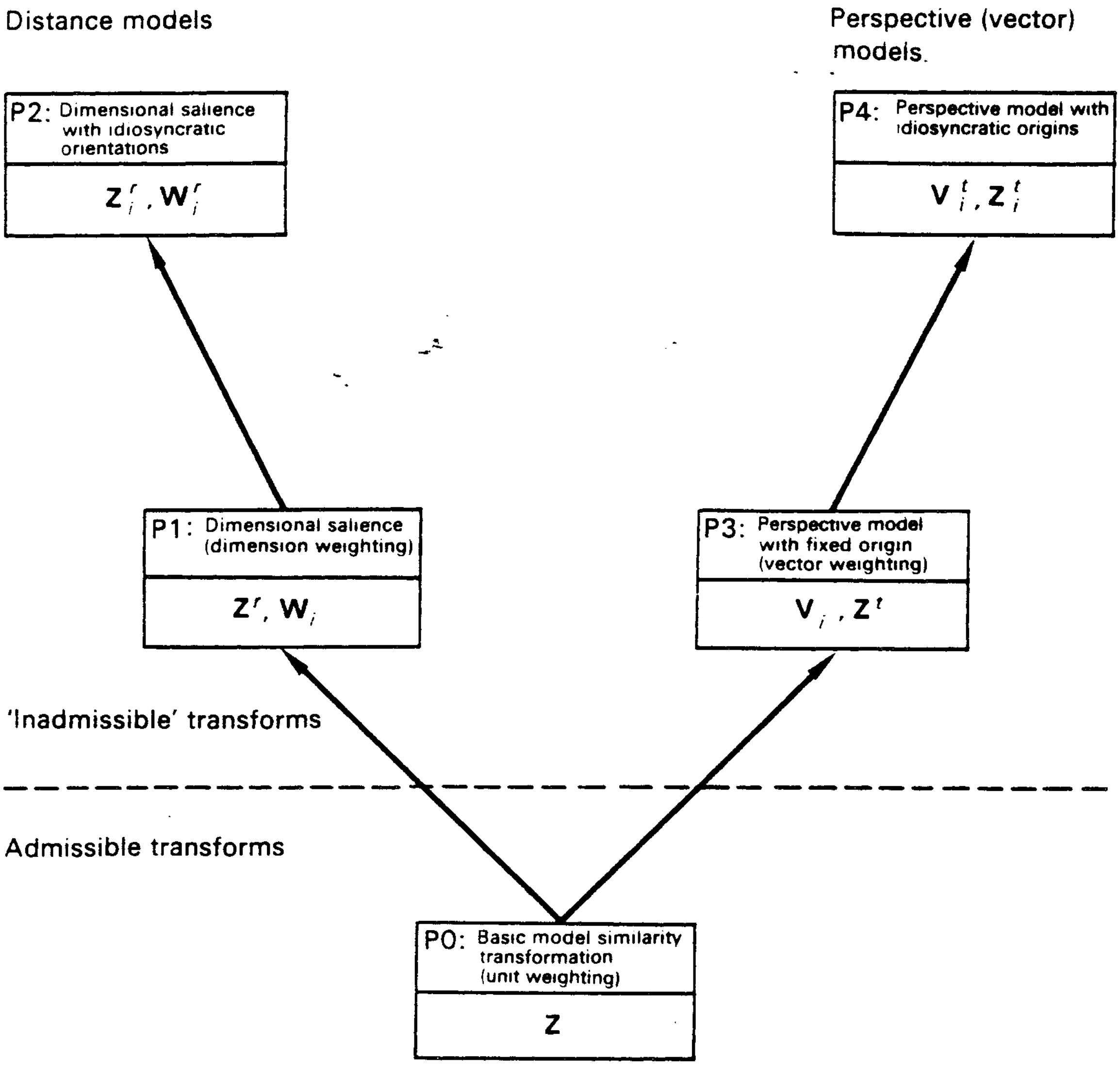


Figure 7.6 *PINDIS models*

the *perspective/vector* models (P4, P3, P0), which share a common basis (P0).*

Before going on to describe the models, some preliminary points need to be made about the form of their specification (in the lower half of the boxes in Figure 7.6).

(i) The parameters of each model are of two sorts—they refer either to what is done to the centroid matrix, **Z**, and to what systematic weights (dimensional weights, **W** or vector weights, **V**) are applied to move the centroid configuration into greater conformity with the original configurations.

(ii) The irritating superscripts and subscripts which bedeck the matrices in the model specification are in fact necessary to distinguish between the models, and repay careful attention. The original average centroid configuration **Z** appears in its pristine, undecorated form in P0. In every other model the centroid configuration is changed in some way.

In the distance models, the superscript *r* signifies that it is rotated. Thus the centroid configuration is rotated in both P1 and P2—to a different orientation for each individual configuration in P2 (signified by $\mathbf{Z}_i^r$) and to a *single* new orientation in P1 (signified by $\mathbf{Z}^r$), indicated by the absence of a subscript. In the vector models, the superscript *t* denotes translation or shift of origin. Hence the centroid configuration is translated to a unique position for each individual in P4 ($\mathbf{Z}_i^t$) and to a single new origin in P3 ($\mathbf{Z}^t$).

The Distance Models

P1 The dimensional salience model is the PINDIS equivalent of INDSCAL. The centroid configuration is first rotated into an optimal position for dimensional weighting and then a set of individual dimension weights are estimated for each input configuration. These weights are entirely analogous to INDSCAL weights, except that in PINDIS they may legitimately be negative, in which case they signify the reflection of the dimension concerned.

P2 The individually rotated dimensional salience model is the PINDIS equivalent of the Carroll-Chang (Carroll and Wish, 1973, pp. 90 et seq.) IDIOSCAL model which allows the axes of the centroid configuration to be rotated to an idiosyncratic orientation for each individual configuration, and then be differentially weighted. In this model, the dimension weights will only be comparable between 'subjects'/configurations if they happen to share the same rotation.

The Vector Models

P3 In the basic simple perspective model, the origin of the centroid configuration is first translated to an optimal position and a vector is then constructed from the origin to each of the constituent stimulus points. Each individual configuration can be thought of as having had a different set of weights applied to each of the vectors. A high vector weight will push a point further out from the origin, and a low vector weight will contract the vector and move a point closer to the origin. A negative

*A further 'double-weighted' model is discussed by Lingoes and is available in the PINDIS program. However, this model is particularly subject to sub-optimal solutions and its use is not generally recommended. Its estimation is suppressed in the MDS(X) version by setting SUPPRESS(1). The double-weighted model is not treated further here.

weight (which is quite permissible) has the effect of 'flipping' the vector in the opposite direction. The effect of a set of vector weights is thus to 'unscramble' a configuration by selectively relocating the stimulus points of the centroid configuration, with the proviso that they can only move in the same direction, towards or away from the origin.

P4 The individually translated perspective model differs from P3 in allowing *each* individual configuration to have its own 'point of view' (origin). Since, as we know, translation of origin changes vector separations, the same set of vector weights may well have markedly different effects on differently centred configurations in this model. Thus P4 permits each individual configuration to have a different origin *and* a different set of weights. It is just as well that Procrustes' imagination was not fired by modern MDS!

Psychologically, the vector models have a certain appeal, since they allow different *categorisations* of stimuli to be related by regular transformations, and allow for such processes as over-compensation—in these models, the Maoist and the Stalinist can in effect share the same political map whilst consigning each other to the fascist camp by means of a single parameter, the negative vector weight!

These increasingly complex transformations bring better fit, but at a cost. In its more complex models PINDIS becomes prolific in its use of parameters, and many users (especially statisticians) are rightly wary of the degrees of freedom consumed. The parsimony of Occam's razor appears not simply to be blunted but to be thrown away with abandon. Assuming that the number of stimuli considerably exceeds the number of dimensions, then the models assume a natural hierarchy defined by the number of parameters (cf. Lingoes and Borg 1978, p. 495).

Fewest free parameters	*Model*	*Parameters per individual configuration*
	P0	(No parameters) (only admissible transformations)
	P1	r dimensional weights
	P2	r dimensional weights and $\binom{r}{2}$ rotation coefficients
	P3	p vector weights
	P4	p vector weights and translation vector of r elements
Most parameters	(P5	r dimension weights and p vector weights)

When deciding which PINDIS model is most appropriate, one will necessarily be trading off the increase in goodness of fit (or explained variance) against the increase in the number of fitting parameters. There are no reliable statistical ways for deciding whether the trade-off is worth it, so the assessment of which PINDIS model is best will always retain a strong subjective element.*

7.4.2 The distance models (P1 and P2)

The two forms of distance (or, more strictly, dimensional) model in PINDIS examine the extent to which a given configuration can be better fitted to the centroid

*Since this was written, Langeheine (1980) has produced the results of his simulation studies of the PINDIS models, which provide expected fit measures and other statistics as a guide to help the user in deciding on the appropriate model. The use of these approximate norms is strongly recommended.

configuration by differential weighting of the axes (P1) or by differential rotation followed by individual weighting in the case of P2. It should be remembered that in neither case are the original relative distances preserved and in this sense the models involve 'inadmissible transformations'.

7.4.2.1 P1, the weighted distance model (with fixed dimensional orientation)
The specification of this model is formally identical to INDSCAL:

$$\delta_{jk}^{(i)} = d(x_j, x_k) = \sqrt{\sum w_a^{(i)}(z_{ja} - z_{ka})^2}$$

i.e. the Euclidean distance between points x_j and x_k in the original configuration $\mathbf{X}_i$ is assumed to be (a similarity transformation of) the distance between points z_j and z_k in the centroid configuration, after each dimension has been differentially weighted. A detailed comparison of P1 and the INDSCAL model is contained in Borg and Lingoes (1978).

The chief advantage which P1 has over INDSCAL is that in the PINDIS hierarchy it is possible to investigate first how much variation can be accounted for in a given set of data by legitimate similarity transformations (i.e. the Procrustes rotation of model P0) *before* having recourse to individual weights. Only if the improvement in explained variance between P0 and P1 is substantial is it worth proceeding to a more complex distance model. P1 differs from INDSCAL in two other significant ways:

(i) in the manner of estimating the group space, and
(ii) in the interpretation of the subject space weights.

In INDSCAL, the group space co-ordinates and subject weights are determined simultaneously, whereas in P1 the centroid space is determined first, then put into optimal orientation for dimensional weighting. Only then are the subject weights calculated as a separate operation. This difference has the effect of improving the properties of the subject space. In particular:

> The squared length of the vector drawn from the origin to the subject space point corresponds *exactly* in the P1 model of PINDIS to the variation in the subject's data (individual configuration) explained by the model and is independent of any orthogonality properties of the 'group space' or centroid configuration. (This, it will be recalled, holds only 'approximately' in INDSCAL, depending on the correlation between the dimensions.) It also means that in P1 it is possible to estimate the contribution of each dimension to the total communality, if the user so desires.
>
> The separation of subject vectors in the subject space correctly represents the correlation of the respective configurations in P1. (This holds only approximately in INDSCAL.)
>
> P1 can accommodate negative weights, which can be interpreted as reversed or reflected dimensions.

In actual practice, the application of P1 and INDSCAL to the same data (after preliminary scaling in the latter case) will lead to very similar results in terms of the centroid configuration, but there will often be significant differences in the subject space. This is well illustrated in the Borg and Lingoes 1978 paper.

An example involving the P1 model is given in 7.4.5.

7.4.2.2 P2, the idiosyncratically rotated and weighted distance model

This model allows individual configurations ('subjects') to differ both in the frame of reference which they adopt (i.e. in the dimensions they choose as significant, so long as they are orthogonal to each other) and in the weights they attach to them. This is also a dimensional model, but one where an idiosyncratic rotation of the axes of the centroid space occurs before the application of weights. The individual differences in rotation could be trivial or substantial. In most cases using PINDIS, it will be found that a slightly different individual orientation of axes will fit an individual configuration somewhat better than the averaged orientation provided by P1. (Indeed, in the PINDIS computational procedure, the individual orientations of axes are estimated first and the orientation for P1 is then averaged from these.)

On the other hand, some subjects may employ a rotation which is clearly different from others. In psychological terms, this often means that some *combination* of the initial P1 dimensions or properties are more salient than the original ones. If, for instance, it turns out that in the judgment of the Irish political candidates two main consensual dimensions are a pro/anti-British and a left/right dimension, it may well be that for a significant fraction of the population a single left-wing, anti-British *vs* right-wing, pro-British dimension is virtually the only one that matters, with small residual variation on the other dimension. In terms of the P2 model, this would mean that for such people a rotation of about 45° with a large weight on the new dimension I and a small weight on dimension II would provide a better frame of reference than the P1 centroid configuration.

The P2 model is very similar to the Carroll-Chang IDIOSCAL model (the acronym stands for Individual Differences In Orientation SCALing) to a variant of Tucker's Three-mode Scaling and to Harshman's (1972) PARAFAC-2. These are discussed in Carroll and Wish (1973, p. 440 et seq.) and the precise relationship between IDIOSCAL and INDSCAL is discussed in Borg (1979, p. 635 et seq.). The P2 and IDIOSCAL models resemble each other in much the same ways as P1 and INDSCAL do, in particular in having separate *vs* simultaneous estimation of the group space and subject parameters, clearer interpretation of the subject parameters in P2, and estimation from the original data (IDIOSCAL) as opposed to ready-scaled data (P2). These characteristics are discussed in Borg and Lingoes (1978) and Borg (1980), and both include a detailed mathematical treatment and a number of illustrative examples.

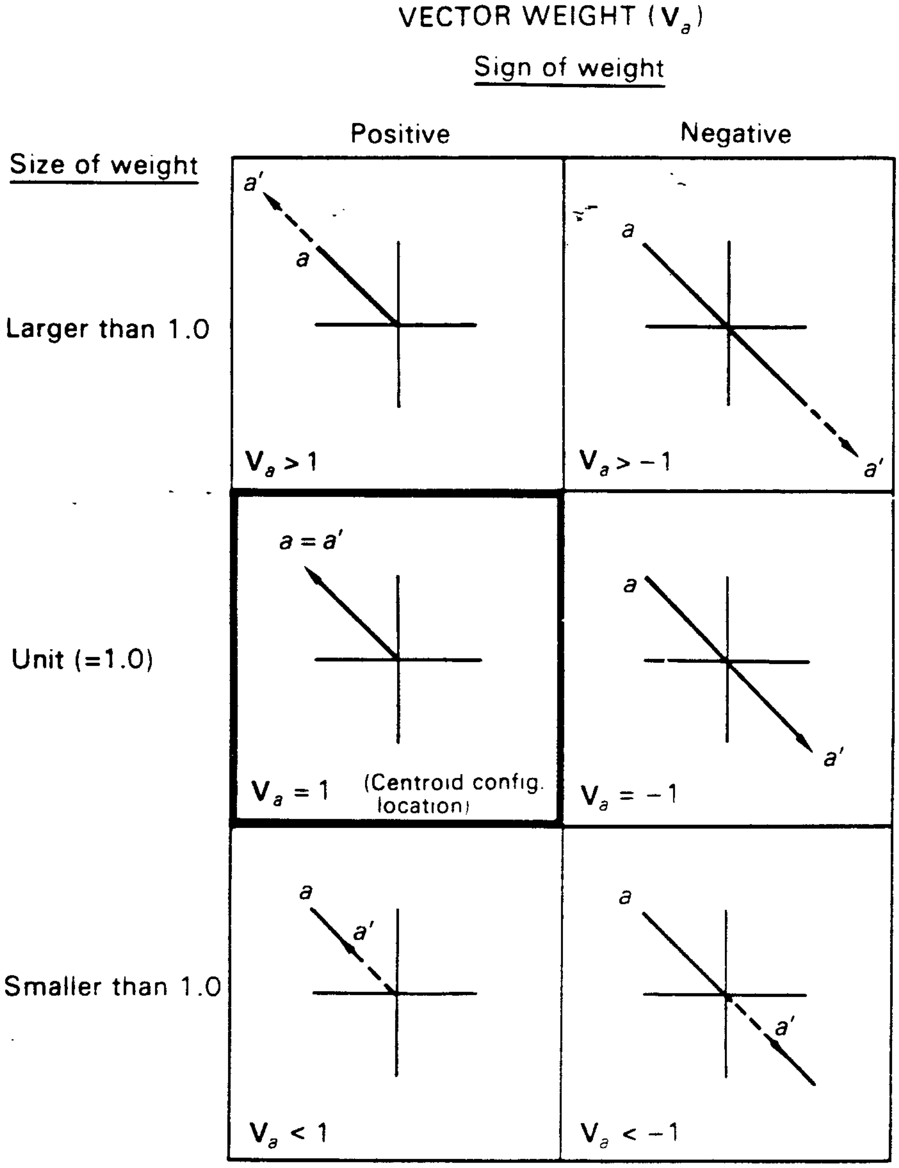
In many ways the P2 IDIOSCAL model is very appealing, but it is a rather complex and relatively ill-understood model, and one for which there are not as yet any very compelling empirical examples. Users are once again cautioned to proceed with care.

7.4.3 The perspective (vector weighting) models (P3 and P4)

The procedure involved in the perspective models is the construction of a vector from the origin to each stimulus point of the centroid configuration, and the weighting of these vectors differentially for each configuration in order to get it into

VECTOR WEIGHT: Size	*Direction* POSITIVE sign (same direction)	NEGATIVE sign (opposite direction)
LARGER (further out)	Point moves further out in same direction	Point moves a greater distance but away from origin
UNIT (same length)	*Point in same position as in Centroid Configuration*	Point is same distance from origin, but in opposite direction compared to Centroid
SMALLER (closer in)	Point moves closer towards origin	Point moves a smaller distance but away from origin

Table 7.3 *Effect of size and direction on point relocation*



Note: *a* denotes location of a point in the centroid configuration; *a'* denotes relocation after applying vector weight (V_a).

Figure 7.7 *Effects of size and sign of vector weight (V_a) on point relocation* better fit with the centroid. Taking the original centroid vectors as the unit, each individual vector weight may be smaller, the same or larger, and it may be positive or negative in sign. These differences and their effects are summarised in Table 7.3 and illustrated in Figure 7.7. Note that whatever value the vector weight has, it can only relocate a point in the same or in the opposite direction.

7.4.3.1 P3, the weighted vector model (fixed origin)

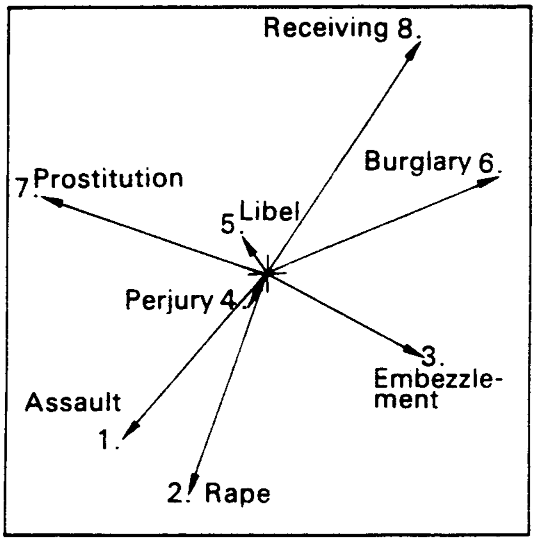
In P3 the main focus of interest is obviously on the set of vector weights which consistently transforms the centroid configuration into as close an approximation to this individual configuration as possible. What should be looked for in comparing individual sets of vector weights? For any configuration, the closer these weights are to +1, the more that individual configuration resembles the centroid and the less useful the vector model is. Within an individual set of weights interest will normally centre upon which are largest and/or which have negative sign, since these imply the greatest relocation compared to the centroid. It can often happen, for instance, that apart from one or two points, the weights are all close to +1, indicating that the significant differences are concentrated in a few points, but that the remaining structure of the configuration closely resembles the centroid. (This incidentally, could never be detected using a dimensional model where all point co-ordinates are *ex hypothesi* systematically weighted.)

In P3 the vector weights are comparable across individual configurations (this is not true of P4) and this provides a second important type of comparison. Presumably, the stimulus points where vector weights vary most from configuration to configuration are the ones which are least stable in the configuration and could be removed from analysis, or alternatively could be given more detailed study. It often happens that the variation in weights for a given stimulus vector is higher in some individual configuration than in others, which suggests that variation is concentrated in a particular area or substructure of the configuration. Once again such a difference cannot be detected using the dimensional models.

In many applications, the simple perspective model P3 shows a dramatic increase in explained variation compared to the basic model (P0), and to P1*, and provides a considerable degree of detail for analysis. The P3 model is further illustrated in Figure 7.8. Here the eight crimes configuration of Figure 7.4b is taken as the centroid configuration. Two sets of vector weights are then applied to produce the 'private spaces' of (b) and (c) in Figure 7.8. As in the INDSCAL and P1 models, the overall shape of the configuration is changed, but in P3 the local structure is also changed, as any cluster analysis would dramatically show!

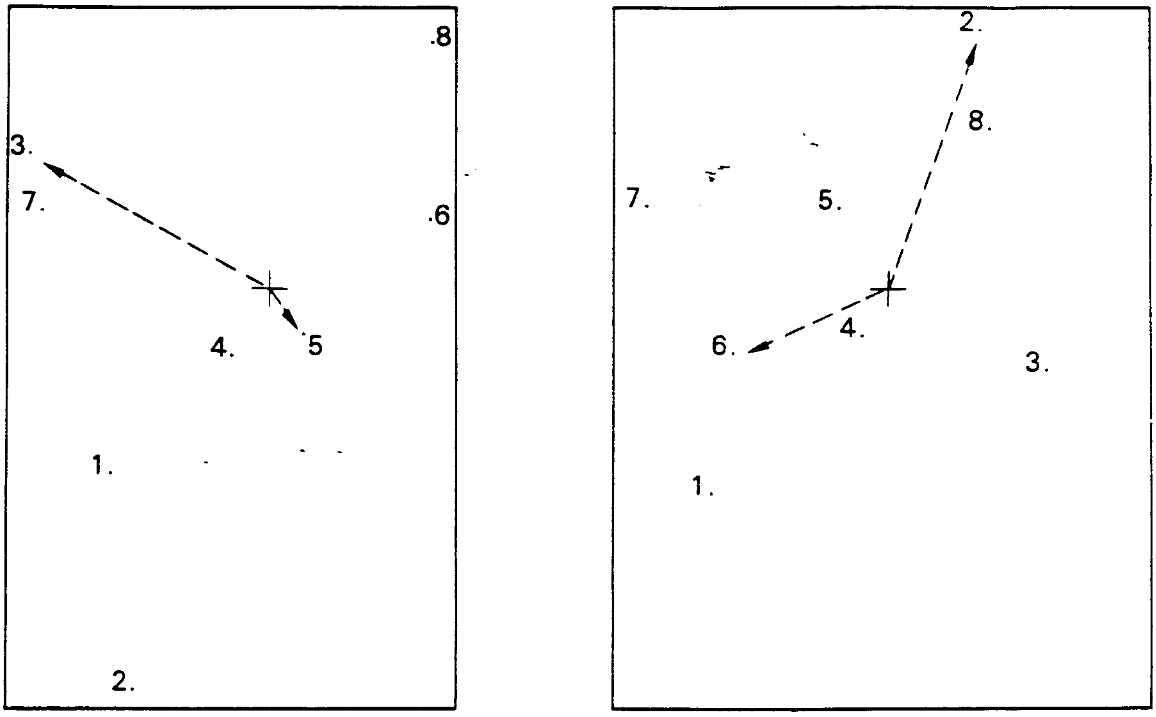
The main differences between (b) and (c) in their *pattern* of weighting are concentrated in the location of Rape (2) and Libel (5). (b) isolates Rape further from assault (moving it away in a south-westerly direction from all the other points) and (c) projects Rape into the opposite direction entirely, to join receiving as its nearest

*Such a result must be treated with some caution, for P3 allows one parameter to be estimated for each stimulus, whereas P1 simply allows one for each dimension. Usually the number of stimuli considerably exceeds the number of dimensions, hence one would expect a better fit for P3 compared to P1. The question is just how dramatic an increase is needed before deciding that it is not simply due to the additional degrees of freedom. See Langeheine (1980) for information relevant to this decision.



(a) Original configuration (Fig 7.4 b)

PRIVATE SPACES (Fixed origin; dotted vector denotes negative weight)



(b) V_i = (0.8,1.3,−1.1,1.5, −0.9,0.5,0.8,0.8)

(c) V_i = (0.9,−1.2,0.7,0.9,1.5, −0.5,0.8,0.5)

Figure 7.8 *P3: perspective model (fixed origin)*

neighbour. However, whilst the size of the weight for libel differs most, its effect is not so marked since libel is located fairly close to the origin in the first place and the composite effect is less dramatic. Nonetheless, (b) now locates libel more in the direction of burglary whereas (c) moves it somewhat closer to prostitution.

This example illustrates rather well the point that, when interpreting the P3 model, it is not simply the size and pattern of vector weights that are relevant but also the multiplicative effect upon the original length of the vector. A massive weight on a point located close to the origin can often move a point a very small distance, whereas it only needs a fairly small weight to move a peripheral point yet further away.

With some justice, the P3 model has been hailed as the major innovation introduced into MDS by PINDIS. It certainly provides a powerful and subtle form of analysis of individual differences and often gives insight into the detail about the source of variation in configurations.

7.4.3.2 P4, the idiosyncratically translated and weighted vector model
In the words of all good detective stories, 'a little thought should convince the reader' that vectors drawn from different origins will alter the pattern and shape of the configuration. The facts of the matter are illustrated in Figure 7.9. Here the centroid configuration consists of three stimulus points (labelled 1, 2 and 3) which form an equilateral triangle centred upon the origin (0, 0). Initially, the vector lengths drawn from the origin of the centroid configuration are all unity. Suppose now that three individuals all happen to employ the *same* vector weighting, namely ($\frac{1}{2}$, -1, $1\frac{1}{2}$). (In the normal way, P4 model individual vector weights can, and will, differ. Making them identical just simplifies the example.)

From the perspective *located at the origin* (labelled *A*), the original equilateral triangle will be deformed by the weights into the (1^a, 2^a, 3^a) triangle joined by the unbroken line. But from the perspective of B at $(-2, 2)$, the configuration (1^b, 2^b, 3^b), denoted by the dotted line, looks quite different, and it looks different again from *C*'s perspective at $(-1, 2)$. Convince yourself that the differences between *A*'s, *B*'s and *C*'s triangles arise purely from the fact that they have different origins. (The configurations would be different again if they did not share the same vector weights.)

For obvious reasons, P4 is often called the 'points of view' model (though this term was originally used by Tucker to refer to a quite different model.) In P4, it is the idiosyncratic origins that can be directly compared and that form the main focus of attention. For instance, the question of whether two subgroups of subjects differ significantly in their perspectives can be readily investigated. But the individual vector weights cannot be directly compared *except* in conjunction with the idiosyncratic shift of origin, which would mean constructing a new set of vectors all emanating from the same origin.

There have been few studies to date which have used this model, and only one where the increase in explained variations is impressive (Lingoes 1977).

7.4.4 Variants and options within PINDIS

As currently programmed, the user may also use PINDIS in an 'external' manner by inputting a hypothesis configuration instead of calculating the centroid. If so requested, this configuration will *not* be differentially rotated (for the dimensional models) and the origin will *not* be translated (for the vector models). The net effect is to suppress P2 and P4 respectively; by a separate user-controlled parameter, P5 can also be suppressed.

The external use of PINDIS is most appropriate when a target configuration is being used as a fixed reference point for other configurations (as in replications and confirmatory studies) and where a known (or hypothesised) structure such as physical properties or a geographical map underlies the data. Note that a hypothesis matrix can be input (to replace the centroid configuration) without also requiring that it be fixed.

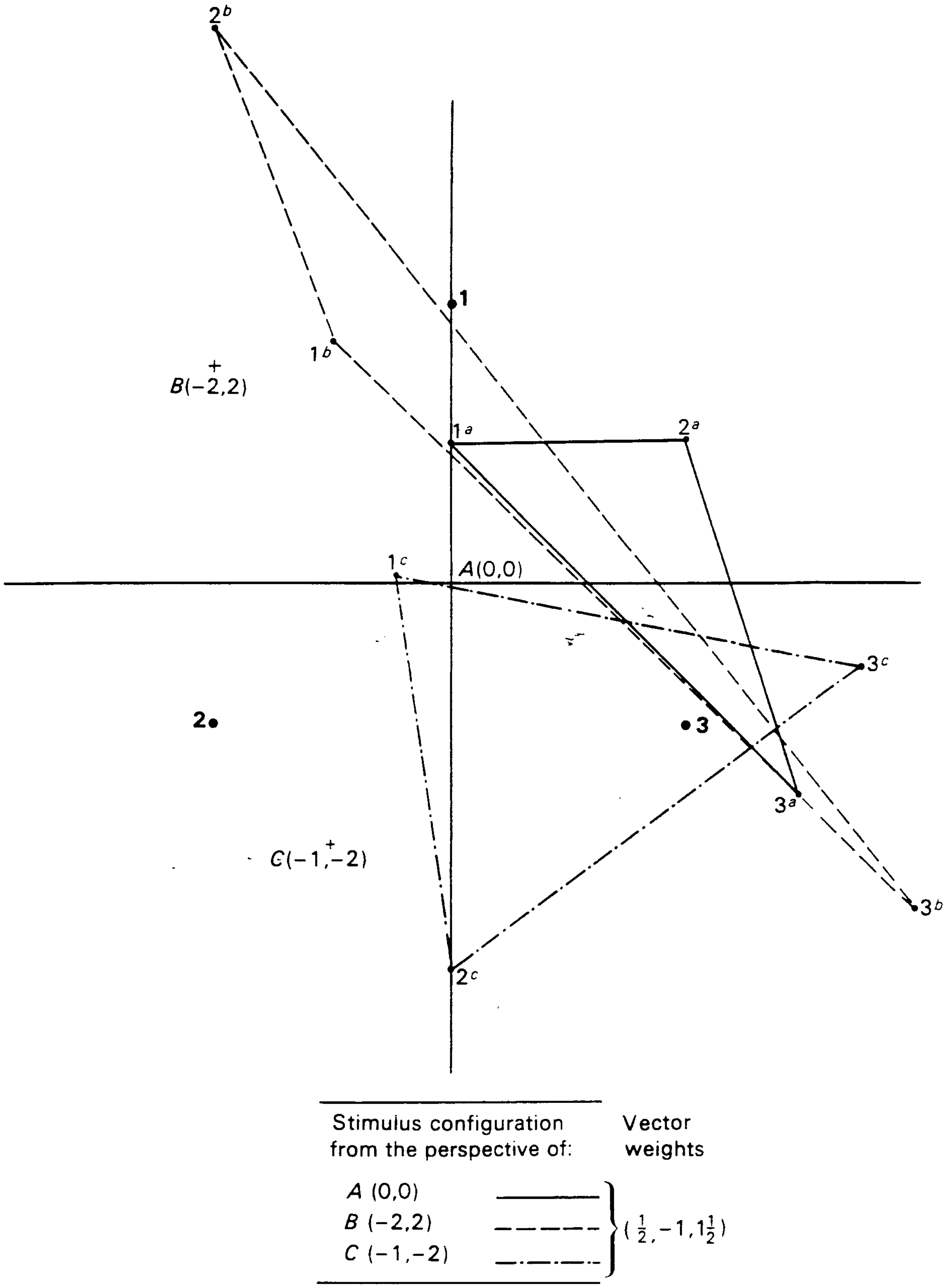


Figure 7.9 *P4: perspective model (idiosyncratic origins)*

Finally, the user may choose to have a specific stimulus location as the origin of the centroid configuration for the vector models. This is especially useful if the configurations being compared have a radex structure, with some stimulus item forming the natural origin. Details of the program parameters used to implement these options are given in MDS(X) *User's Manual* documentation for PINDIS.

7.4.5 An application of PINDIS

A range of published examples of the application of PINDIS is included in Table 7.4. It is interesting to note that the 'optimal' PINDIS model often turns out to be either the simple dimensional model (P1) or the simple vector model (P3). Only the somewhat atypical second example of Lingoes (1977) gives strong support to the double weighting model.

These examples also show that PINDIS can be used in an exploratory or a confirmatory mode, or in combination of both.

A further example, also drawn from the occupational cognition study, illustrates these points. A set of 48 subjects of varying status and occupational backgrounds were asked to rate or rank 16 occupational titles on three criteria frequently employed by sociologists to serve as forms of 'prestige', to which were added two factual criteria concerned with estimated average income and how much the subject thought she knew about the occupations concerned.

The data were scaled internally using MDPREF with the following values of an overall goodness of fit measure (linear root mean square correlations) of the data to the 3-D MDPREF solution. (Data, solutions and further details are contained in Coxon and Jones 1979, pp. 86–105):

Criterion		Goodness of fit RMS, 3-D MDPREF solution with data
1. *Social standing*	(own opinion of general standing in the community)	0.90
2. *Prestige and rewards*	(an occupation ought to get)	0.89
3. *Social usefulness*		0.84
4. *Monthly earnings*	(estimated by subject)	0.87
5. *Cognitive distances*	(how much subject knows about what a job involves)	0.73

We were interested in knowing how these MDPREF solutions differed among themselves, and also how they compared to the INDSCAL group space configuration (obtained by separate estimation of similarity and discussed earlier under 4.6 and illustrated in Figure 4.5). The differences among the MDPREF configurations were first investigated by using a straightforward PINDIS analysis. Their similarity to the 3-D INDSCAL group space was then analysed by using it as a fixed hypothesis configuration in a second run.

The result of the first analysis is summarised in Table 7.5, and the centroid configuration is given in Figure 7.10. The resemblance between this and the INDSCAL configuration of Figure 4.5 is, at least at first sight, very marked indeed. In terms of similarities between the configurations (Table 7.5a), there is quite good internal agreement: on average 69 per cent of variation between the configurations is attributable to admissible transformations, with the two status criteria (1 and 2) being especially well fit and estimated earnings (4) particularly badly fit. Neither of the dimensional models improves this fit in any way at all—a mere 2 per cent increase is involved. On the other hand the vector models do rather better, with 17 per cent improvement for P3 and 24 per cent improvement for P4. Once again, the earnings configuration fits relatively poorly, but its fit is markedly improved by allowing it to have an idiosyncratic origin.

Reference	'Subjects' Configurations	Stimuli	PINDIS models P0	P1	P2	P3	P4	P5	Fit (r^2) P0	'Optimal'	Comments
	(i) 2 species-related, but apparently dissimilar, fish structures. *Lingoes 1977*	29 points defining fish outline	√	√	—	√	—	√	0.72	P5:0.96	Dimensional weighting does better than vector despite fewer parameters. One of the few instances where P5 produces markedly better fit.
	(ii) 3 biological (genetic) 'maps' measuring dermaglyphic, anthropometric and genetic characteristics of the Indian population.	7 Latin American villages	√	√	√	√	—	√	0.63	P3:0.90	Interesting use of target (geographical) configuration which is *not* kept fixed. Clear superiority of P3 over P1 suggests interesting conclusions (q.v.).
Borg 1980 (i) *data* Lingoes et al. (1979)	14 individual subjects (SSAR-1)	6 Political Party × 6 'Closeness' combinations	√	√	√	√	—	√	0.77	P3:0.88	Not a marked improvement.
(ii) *data* Green and Rao (1972)	41 individual subjects (MDSCAL)	15 breakfast foods	√	√	√	√	—	—	0.72	P3:0.85	Lot of individual variability.
(iii) *data* Lingoes et al. (1979)	10 normal (N) and 4 colour-dificient (CD) subjects (Classic Scaling)	10 colour tiles	√	√	√	√	—	—	N:0.98 CD:0.74	(P0) P1:0.88	Z of normals used as fixed target for assessing colour-deficients. As expected, simple dimensional weighting of colour-circle is best model.
Maimon et al. 1980	4 groups of managerial graduates and supervisors. (SSA-1, 2D)	9 items (facet-design) type × area of skills	√	√	√	√	—	√	0.71	P3:0.90	High variability on all other models.
Coxon and Jones 1980	6 individual deviant cases from occupational cognition project (MINISSA, 3D)	16 occupations	√	√	—	√	—	—	0.29	P3:0.70	Used fixed target (INDSCAL) configuration.

Table 7.4 *Applications of* PINDIS

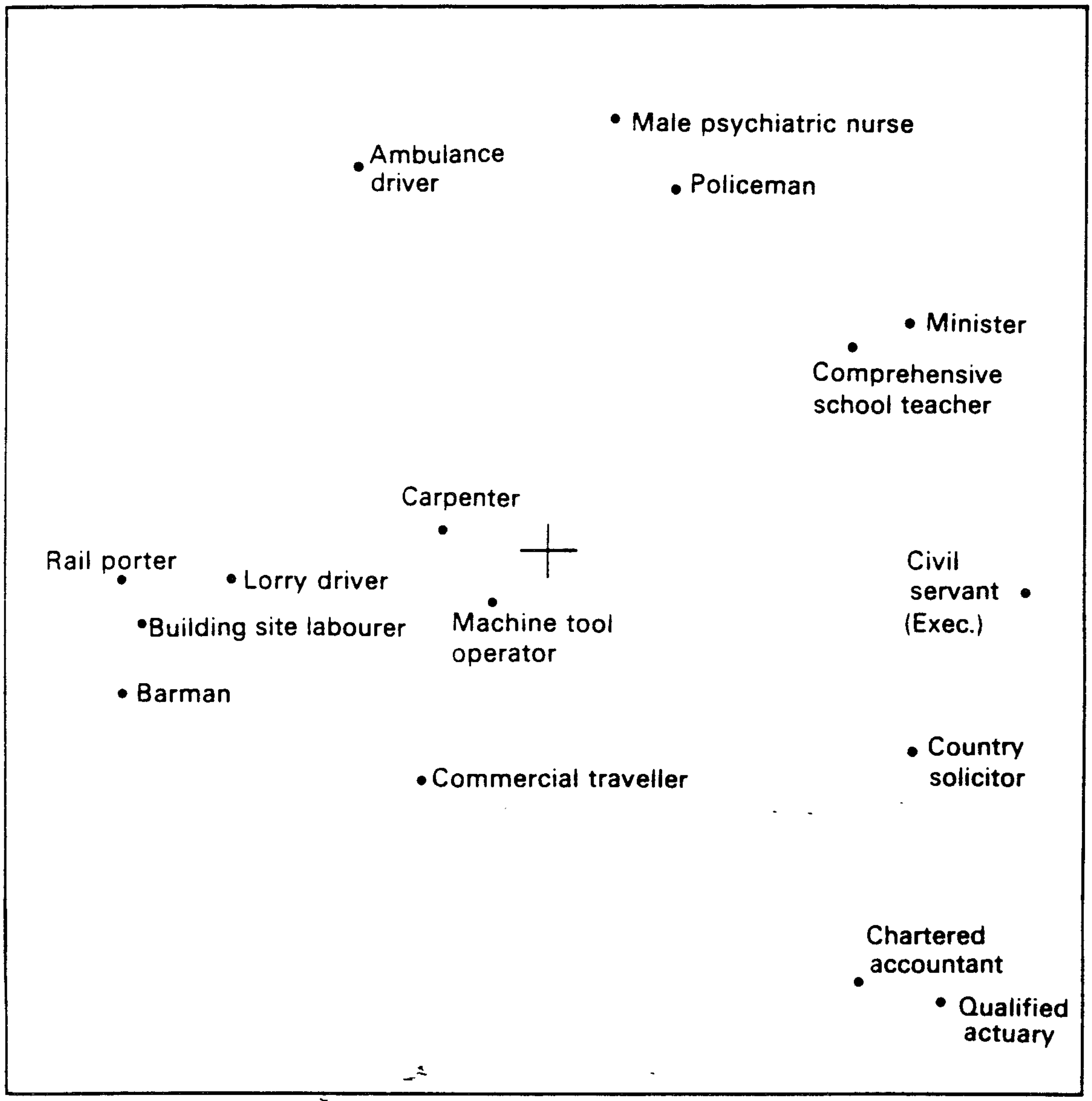


Figure 7.10 *First 2 dimensions of PINDIS centroid configuration derived from 5 MDPREF configurations*

Concentrating now on P3, where are the discrepancies located? An analysis of the vector weights indicates unequivocally that the first three status criteria and the last two ('cognitive') criteria agree very considerably among themselves, and contrast with each other in terms of the relative size and pattern of the weights. Most 'unscrambling' is done with respect to the machine tool operator (the prestige configuration has a vector weight of 2.43, whilst the cognitive distance configuration has one of the few negative weights of −0.17 for this occupation), and to a lesser extent the civil servant (the earnings configuration attaches a weight four or five times that of the other criteria).

In the second run, the configurations were related to the INDSCAL 3-D configuration—which, as we have commented, the PINDIS centroid configuration seems to resemble closely—the degree of fit drops dramatically (see Table 7.5b). Clearly there *are* very considerable differences with respect to the INDSCAL configuration and we have been over-impressed by their surface resemblance. Nor is the fit improved by the dimensional model. Since the INDSCAL configuration is fixed,

Configuration	Communalities (r^2) for PINDIS Transformations					
	P0	P1	P2	P3	P4	P5
	Basic	Dimensional weighting	Dim. weighting and rotation	Vector weighting	Vector weighting and translations	Double weighting
1 Social Standing	0.83	0.83	0.83	0.94	0.99	0.88
2 Prestige and Rewards	0.83	0.84	0.84	0.96	0.92	0.90
3 Social Usefulness	0.77	0.78	0.78	0.86	0.92	0.83
4 Monthly Earnings	0.40	0.42	0.44	0.68	0.89	0.81
5 Cognitive Distance	0.62	0.64	0.64	0.82	0.93	0.73
Average	0.69	0.70	0.71	0.86	0.93	0.83

(a) Full analysis (all simple models, no hypothesised configurations)

	P0	P1	P2	P3	P4	P5
1 Social Standing	0.26	0.27	—	1.00	—	0.64
2 Prestige and Rewards	0.16	0.17	—	0.79	—	0.65
3 Social Usefulness	0.19	0.20	—	0.55	—	0.47
4 Monthly Earnings	0.12	0.12	—	0.40	—	0.57
5 Cognitive Distance	0.25	0.26	—	0.62	—	0.59
Average	0.19	0.20	—	0.67	—	0.58

(b) Analysis with INDSCAL 3-D configuration as hypothesis

Table 7.5 PINDIS *analysis of 5 occupational criteria configurations* the analysis excluded rotation of axes and translation of origin, so the parameters of P2 and P4 were not estimated. Once again, the simple vector model does a dramatic job of improvement, with the social standing configuration now fitting perfectly, and the earnings configuration still being rather poorly fit.

Such a finding is substantively important: it shows for instance that we can conclude that data obtained from judgments of 'general standing' of an occupation give rise to a conclusion that is a systematic transformation of the INDSCAL structure obtained from judgment about 'overall similarity'. At the very least, such a finding strongly contests the received opinion that the latter is a cognitive criterion and is generically distinct from the evaluative status criteria.*

7.5 Preference Mapping (PREFMAP I and II)

The Carroll-Chang (Carroll 1972, p. 114 et seq.) preference mapping models form a hierarchy of models, akin in many ways to PINDIS. They differ principally in that PREFMAP (PM) is designed primarily for *external* scaling (where the user provides a stimulus configuration) and the input data consist of a rectangular matrix of ratings or rankings of p stimuli given by N subjects. The purpose is to map *each subject* into the stimulus space in the form of an ideal point (PM Phases I–III) or as a vector (PM Phase IV) according to a hierarchy of increasingly complex models. (The transformation may be linear/metric or quasi non-metric/ordinal according to the user's choice.) We have already encountered earlier in this book the two simplest models: the *simple distance model* (Phase III) (5.3.3.1 and 6.2.4) and the *vector model* (Phase IV) (5.3.2 and 6.2.2) and so the focus here will be on the more complex models—the rotated and weighted distance model (Phase I, akin to P2 in PINDIS) and the weighted distance model (Phase II, akin to P1 and INDSCAL). But it will be helpful to begin by summarising the full range of the PREFMAP models.

7.5.1 The PREFMAP hierarchy of models

The basic notion of all the models in PREFMAP is that an individual's preference ranking is a function of the distance separating her point of maximum preference (ideal point) and the stimulus points. It is the way in which the distances are defined which differentiates the models.† (PREFMAP can also be seen as extending and generalising Coombs' unfolding model, and as showing that the vector model of preference is a special case of unfolding.)

Briefly, the hierarchy of models is as follows:

I *General Unfolding Model (PM1)*

This is the most general model. Each subject is viewed as having a specific, most-

*When a two-dimensional PINDIS analysis is performed, the results are much the same, but even more marked. Admissible transformations account for 75% of variation in the 'internal' PINDIS analysis and for 14% variation when related to the 'external' INDSCAL configuration. In neither case do the dimensional models add more than a derisory amount (2% in both cases), but the increase due to simple vector weighting is more impressive (18%, and 57% in the external case). But the 'unscrambling' is different in detail: the machine tool operator is still relocated more than any other occupation, but the actuary and accountant are also seen to be relatively unstable in their positioning.

†Technical details of the models are given in Carroll (1972) and in Coxon and Jones (1979, p. 106 et seq.), and a detailed exposition of the computational procedure and output details is given in van Schuur (1977).

preferred ideal point in the space, rotating the axes of the similarity space to his own reference dimensions, and then attaching an evaluative weight to each of them.

II *Weighted Unfolding Model (PM2)*
Subjects are assumed to have an ideal point and to share the same set of reference dimensions (i.e. no individual rotation), but to evaluate or weight the dimensions differentially.

III *Simple Unfolding Model (PM3)*
Subjects are assumed to share the same set of reference dimensions *and* to give the same weighting to them. However, subjects do still differ in terms of where their ideal points are located in the space. This is the external analysis analogue to Coombs' unfolding analysis.

IV *The Vector Model (PM4)*
Carroll (1972b) has shown this to be a special case of Coombs' unfolding, when an ideal point is located far from the origin of the space.* It is for this reason that it features in the hierarchy, since all the other models are straightforward *distance* models. In the vector model, by contrast, subjects are represented as a vector (or line directed toward the region of greatest preference), and a preference ranking is interpreted as the order of the projections of the stimuli points on this line. (The internal analogue of this model is MDPREF.)

One particularly valuable aspect of the hierarchical nature of the PREFMAP models is that it is possible to test whether a more complex model explains significantly more variation in the data than a simpler one. In this way, the model which makes the most parsimonious set of assumptions can be chosen. Moreover, as in PINDIS, there is no reason why one particular level or model should be thought to apply to all subjects—it might well be, for instance, that whilst the simple unfolding model applied to most subjects, the data of the remaining subjects might be far better explained by assuming that they simply differentially weight the dimensions of the space.

All the PREFMAP models make two basic assumptions:

(a) *A common similarity space is assumed to apply to all the subjects included in a* PREFMAP *analysis.* If it turns out that this does not hold (because, for example, a previous INDSCAL analysis of the similarities led to the conclusion that subjects were divided between distinct 'points of view'), then a separate PREFMAP analysis should be run for each subset, using the group space for each subset as the input configuration.

(b) *An individual's preference values for a set of stimuli are assumed to be linearly (or in the non-metric version, monotonically) related to the distance between her ideal*

*Intuitively this can best be seen in terms of isopreference contours (see Figure 7.12). For the basic distance model, all stimuli at a given distance from a subject's ideal point form a circle (isopreference contour) in 2-space, a sphere in 3-space, and so forth. For the vector model all stimuli at a given distance along a subject's vector form a line (projection) in 2-space. If an ideal point is taken further and further out from the origin, the circular isopreference contours come closer and closer to being a line in the vicinity of the stimulus points. See Carroll (1972).

point and the stimulus points. This assumption is only made for the first three models (which are distance models), and is not made for the vector model.

To give these ideas some substance, let us return to the example where a sample of respondents had been asked to assess the general similarity of a set of Irish politicians, and then been asked to rank them in order of personal preference. Let us also assume that the INDSCAL analysis of their similarity data indicated that they had fairly similar perceptions of the politicians and that the two main differentiating axes were left/right orientation and Republican/Unionist. The INDSCAL group space configuration could then be used as the 'independently established' cognitive space for input PREFMAP, and the focus of interest would now shift to explaining differences in the preference rankings of politicians which the respondents gave.

The simplest *vector model* (Phase IV) of PREFMAP assumes (as in MDPREF) that the subjects collapse the multidimensional stimulus space into one dimension (or line) representing the order of preference. Individual differences in preference are expressed by the differing directions which the vectors have in the common space. In this example, it is conceivable that some subjects simply prefer politicians who combine radicalness with support for Irish Republicanism (or conservatism with union with Britain), whilst others evaluate a politician solely in terms of how left-wing she is, or how much she supports Irish Republicanism.

The *simple unfolding (distance) model* (Phase III) assumes, by contrast, that each subject has one most preferred point in the group space, and that this serves as a reference point for evaluating the politicians, according to how close they are to her ideal point. The *weighted unfolding model* (Phase II) assumes that subjects differ considerably in the value they attach to the dimensions of the space—figuratively, that they pay attention to how highly they value or weight each dimension before they decide how close a politician is to their ideal. The effect of this is to 'pull' a politician closer to the subject's ideal point than would be the case in the simple unfolding model, if the politician occupies a high position on a dimension which the subject highly values.

The *general unfolding model* (Phase I) drops the assumption that subjects' evaluations refer to the same fixed set of dimensions. Instead, it allows them to structure the space as they wish by providing their own reference axes (so long as the dimensions they choose are not correlated) and *then* it allows them differentially to evaluate these 'private dimensions'. The effect of this is that subjects can be allowed to place high evaluation on *combinations* of the original dimensions. For instance, if the original axes were rotated anticlockwise through 45°, subjects who chiefly prefer politicians who combine socialism *and* republicanism (but who still differentiate socialist-Unionists from conservative-Republicans) could easily be accommodated.

Turning now to the hypothetical results of a PREFMAP analysis, it might turn out that, on average, the simple unfolding model (II) held best—that is, most subjects evaluated the politicians in terms of highly salient characteristics (dimensions) which were evaluated in the same way, and only differed substantially in what their positions were on the dimensions. But the data of a minority of subjects might be much better explained by assuming, *in addition*, that they valued highly politicians who were socialist-Republicans, cared not at all for conservative-Unionists but still made some (but relatively little) differentiation between socialist-Unionists and conservative-Republicans.

7.5.1.1 PREFMAP *Phases I and II*
Phase I: General unfolding model
Subjects are permitted

(i) to rotate the reference dimensions of the space, and
(ii) *then* to weight them differentially.

In Carroll's terms:

> We allow distinct individuals additional freedom in choosing a set of 'reference axes' . . . and then to weight differentially the dimensions defined by this rotated reference frame, in addition to being permitted an idiosyncratic ideal point.
> (Carroll 1972, p. 120)

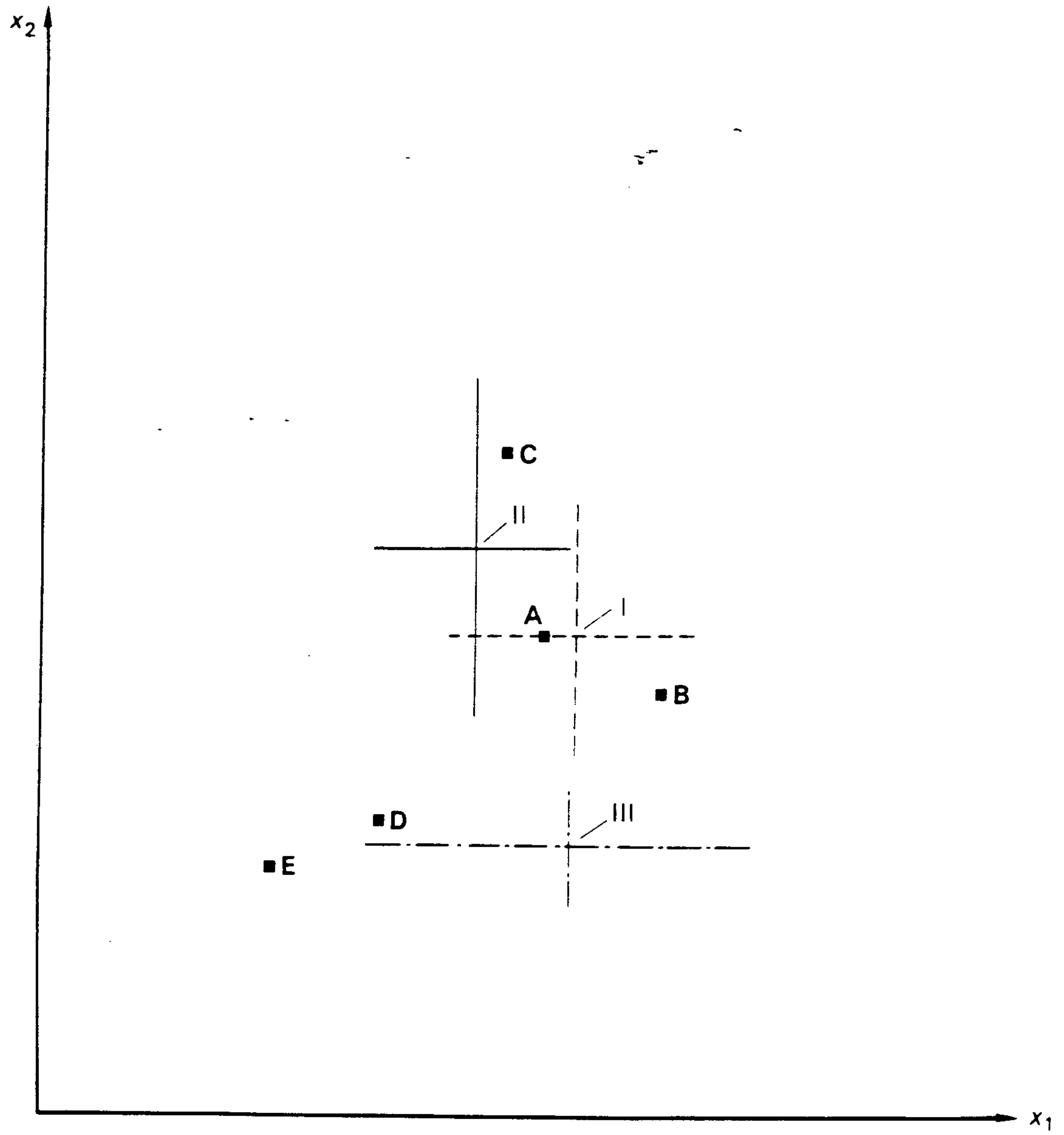


Figure 7.11 *PREFMAP, Phase II (weighted unfolding): subject ideal points and weighted dimensions*

A subject is assumed to apply his own orthogonal rotation to both the stimulus and ideal point co-ordinates, and *then* weight the rotated dimensions. If x^*_{ja} represents the *transformed* stimulus co-ordinates, and y^*_{ia} the transformed ideal points, then

$$s_{ij} = F_i(d^2_{ij})$$

where $d^2_{ij} = \sum_a w_{ia}(y^*_{ia} - x^*_{ja})^2$,

i.e. a Euclidean distance in an individually-rotated and weighted 'private space'.

Phase II: Weighted unfolding model
This is similar in many ways to the INDSCAL and PINDIS P1 model except that in this *preference* model, the weights are interpreted as reflecting the subjects's *evaluation* of the dimension when making an overall preference judgment. To continue the Irish example—two subjects might well entirely agree about, say, the left-right orientation of the politicians. To one subject this may override all other considerations when it came to choosing a candidate, whilst to another it might be considered entirely irrelevant compared to the politician's position on Republicanism.

In this model, a subject is assumed to apply an evaluative weight w_{ia} to each dimension, so that

$$s_{ij} = F_i(d^2_{ij})$$

where now $d^2_{ij} = \sum w_{ia}(y_{ia} - x_{ja})^2$,

i.e. a Euclidean distance in a weighted 'private space'.

The weighted unfolding model (PM2) is illustrated in Figure 7.11 (and see also Figure 7.12). In Figure 7.11 the joint space includes five stimulus points (A to E) and the location of the ideal points of three subjects (I, II and III). In this model, the subjects differ in terms of the location of their ideal points (akin to PINDIS idiosyncratic perspective model (P4)†), and they also attach different evaluative weight or salience to the dimensions. In Figure 7.11 the differential weights are denoted in the form of the arms of a cross. In the case of I, the weights for the dimensions are equal; for II the weight of dimension II is greater than that for dimension I, and for III the reverse is true. Note that in this model the individual axes are all oriented in the same direction, parallel to the reference axes and are therefore directly comparable. 'Private spaces' for each individual can be produced, if so desired, by stretching and shrinking the reference axes. A more convenient representation is illustrated in Figure 7.12.

7.5.2 The *a priori* stimulus space

Since the PREFMAP models are designed to be external in form, the user must normally supply an *a priori* configuration, and similar issues arise as in the case of PINDIS: what may reasonably be used as an *a priori* configuration? How, if at all,

†Compared to the PINDIS hierarchy, PM2 is a hybrid model. PM2 is similar to P4 in the sense of allowing differing points of view, but is similar to P1 in allowing differential weighting of axes.

does the configuration change at different levels?

The source of the *a priori* configuration can be considered under three heads: (i) a previous scaling, (ii) a theoretical or rational configuration and (iii) an 'internally-generated' configuration.

A previous scaling
The configuration may be the result of a previous MDS scaling analysis of similarities data for the same set of stimuli, probably obtained from the same subjects that provided the preference data. This is probably the most common instance in social science applications.

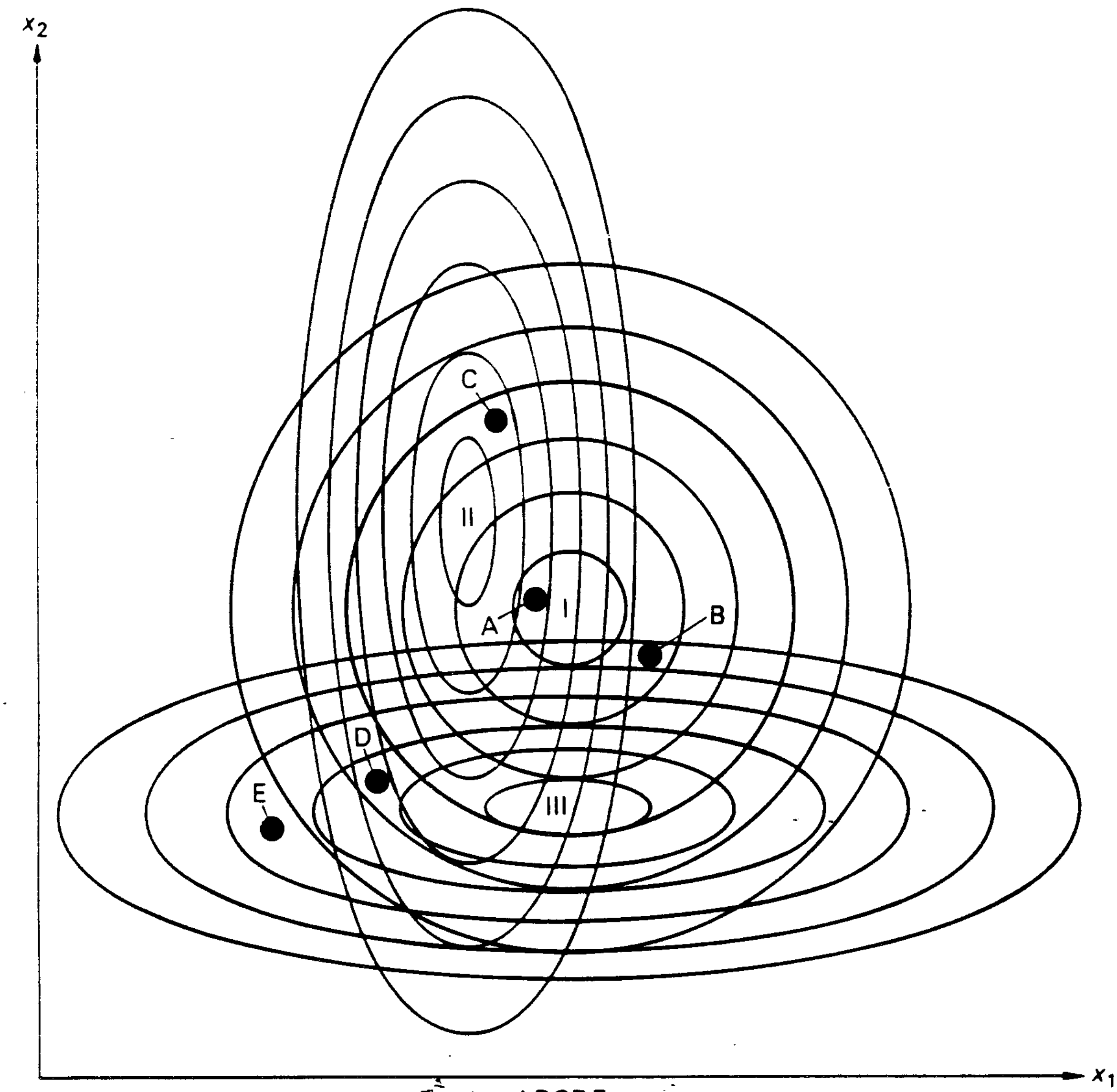
When replicating a study it can be useful to see how well the data from one's own study will fit the configuration obtained by the original investigator.

A theoretical or rational configuration
By a 'rational configuration' is meant one which incorporates the actual characteristics of the stimuli as dimensions of the *a priori* space. This occurs typically in psychophysical applications and in cases where the stimuli are well-defined compositions. Examples include the now infamous 'hypothetical cups of tea' data (collected by Wish and analysed in the Carroll 1972 paper) whose defining characteristics were the *hotness* and the *sweetness* of the brew, and in the Delbeke-Bollen family composition data (cf. Coxon 1974 and section 6.2.2) whose characteristics were the *number of sons* and *number of daughters* making up a family. Other instances might include a facet analysis of the stimuli, the geographical location of stimuli, or some other theoretically expected or physically underlying configuration.

Internally generated configuration
The option also exists in PREFMAP for a stimulus configuration to be constructed from the *same* preference rankings data as are used in the analysis. (It is implemented in the MDS(X) version using the INITIAL parameter.) Such a 'quasi-internal' configuration is constructed by forming the minor (stimulus × stimulus) product moment matrix from the data, (which may have been transformed in some user-specified way*), and then performing a classic basic scaling (see Appendix A5.2) to yield the configuration.

In the MDS(X) version of PREFMAP, the user also has the freedom to keep the stimulus space configuration fixed throughout all phases of analysis, or to allow the program to modify it to obtain a better fit to the data by the use of the KEEP parameter. If the first option is chosen, the original stimulus configuration remains unchanged throughout the analysis. In the second case, the form of modification is slightly different, depending upon the level of the model (phase) at which the user starts analysis (see Carroll and Chang 1967, p. 10 et seq. and Carroll 1972, p. 134 et seq.).

*Options exist within the INITIAL parameter for double centring the product matrix (removing row and column means and replacing the overall mean); removing row means; standardising rows, and for removing row and column effects (see *User Manual*, PREFMAP Report, 2.3.1). If the first option is chosen and the data then analysed under the linear version of PM3, the resulting analysis turns out to be equivalent to an internal metric unfolding analysis. On this and related issues, see Carroll 1970, 1971, Carroll and Chang 1970, and Schonemann 1970.



(Reproduced with permission from Carroll 1972)

Figure 7.12 *Weighted unfolding model (PREFMAP Phase II): isopreference contours for three individuals*

7.5.3 Preference and 'anti preference'

In the distance (unfolding) model of preference, it is assumed that a subject's preference is single peaked and symmetric—that she has one point of maximum preference in the stimulus space and that preference decreases systematically in all directions (along each dimension). One way of representing this is to define a set of contours (at arbitrary but equal intervals) centred upon the ideal point. Each contour represents points at an equal distance from the subject's ideal point, and hence all points will be equally preferred. In preference analysis, these contours are referred to as 'isopreference contours'. In the case of the simple distance model (PM3), where each dimension is equally weighted, these contours will in the two-dimensional case be circular, and in the weighted distance model (PM2) they will be ellipses the length of whose axes will equal the value of the evaluative weight. This is illustrated in Figure 7.12, which presents the isopreference contours for the three subjects of Figure 7.11. Note that subject I has equal weights (and hence circular preference contours) whereas the others have unequal, and hence elliptical contours. Once the preference contours are drawn in, it becomes straightforward to reproduce a subject's preference ranking by following the relevant contours. Thus, for subject II, C is within the second ellipse from the ideal point. A within the third, D within the fifth, B just beyond the sixth and E further beyond it—yielding the I-scale: CADBE.

In the case of the rotated and weighted model (PM1) the ellipses are oriented in different directions, but the same logic applies. In the case of the vector model (PM4) the isopreference contours are straight lines, rather than circles or ellipses, since all points which project onto the same point along the preference vector, wherever they be located in the space, will be equally preferred.

The PREFMAP models specifically allow for negative evaluative weights. How are they to be interpreted? Carroll argues that it is simply a matter of changing the form of the preference function in the distance models, from being single-*peaked* and symmetric to having a single valley (one single point of *minimum* preference) and symmetric. In this case, the contours will decrease towards the subject's 'pessimal', 'anti-ideal' or least preferred point. (Perhaps the simplest example is the preference for the heat of liquids: many people dislike lukewarm tea most of all, but increasingly prefer both hot and iced tea.) It will sometimes happen that a subject's point will include a mixture of negative and positive weights; this situation has been explored in Carroll (1972, pp. 121–3), who shows that in the two-dimensional case the function will be saddle-shaped. Nonetheless, it is often difficult to give meaning to such mixed-sign combinations and some care is needed in their interpretation.*

7.5.4 Goodness of fit

As a nested set of models, it is possible to use analysis of variance statistics as an indicator of how much better one model does than another (or how much more variation one explains) than another. At the end of a PREFMAP run a table is printed containing correlations, F ratios for each subject by each phase/model and between pairs of phases/models, together with root mean squares (RMS) for each phase. The correlations are between the squared distances of the model and the preference data values (or corresponding disparities in the case of the quasi-nonmetric option having been chosen).

The average RMS values usually give an indication of the level of the hierarchy of models which is generally most appropriate, i.e. that level where the marginal increase in RMS values for more complex models is very small. The subject-phase correlations can indicate whether any phase fits a subject's stated preferences, and the extent to which one phase fits better than another. In conjunction, these measures enable the user to decide upon the most appropriate model (phase) of PREFMAP for representing the data, and they also allow subsets of subjects to be allocated to different levels.

Since each model is a special case of the one immediately above it, it is justifiable

*Davison (1976) has proposed and recently implemented (Davison 1980) a variant of preference mapping which allows the evaluative weights to be constrained to non-negativity or non-positivity.

to test for goodness of fit of a model to the data by the usual ANOVA procedures, and test for significant increases in variation explained by means of an F-test between models, although the models are clearly not independent.

Most users enter the PREFMAP program at a higher phase/model and drop down to a lower phase. Carroll (1972, p. 135) argues that the solution for the 'average subject' of the highest phase entered should form the basis for the succeeding phases.

Thus, if S-PHASE (1) and E-PHASE (4) are chosen then the full set of models will be applied to the data. This will mean:

(i) After Phase I (PM1), the rotated reference axes of the average subject form the 'canonical reference frame' for Phase II (PM2).

(ii) The subject's weights and ideal points are fitted in *this* reference frame in Phase II (PM2), rather than in the original input frame.

(iii) In a similar manner, the average subject's dimensional weights are applied to the rotated configuration for analysis in Phase III (PM3).

These operations are not always desirable, and the remedy is either to return to the original input configuration at each phase, by setting KEEP (1) or to enter at a lower phase. For instance:

to keep the *orientation* of the original configuration, enter at PM2
to keep the dimensions unweighted, enter at PM3.

The parameters of the model are estimated through quadratic and linear regression (for the metric version) and subsequent monotone regression (for the non-metric version). The former allows statistical testing, on the assumption that the data and fitted values are linearly related, but the latter only permits approximate statistical tests and they should be used with caution.

In many applications of PREFMAP, the main differences seem usually to occur between the simple distance (unfolding, P3) model and the simple vector (P4) models, with few convincing examples of the necesity of invoking more complex models except for particular subjects.

7.5.5 Uses of PREFMAP

The PREFMAP program can be used in three principal ways:

(i) to map individual ratings or rankings into an independently obtained external configuration (external joint mapping).

(ii) to first estimate the stimulus configuration from the subject's data, and then map the same rankings/ratings into it (quasi-internal mapping).

(iii) to use PREFMAP to extend two-mode scaling to include additional subjects (extending two-mode solutions).

7.5.5.1 External joint mapping

The first is undoubtedly the most common use and the one for which PREFMAP was originally designed. In most cases, the researcher has obtained both pairwise similarities data and preference rankings or ratings from the same subjects. The similarities data will typically have been scaled by INDSCAL, or averaged over all subjects and scaled by MINISSA, and the resulting stimulus configuration (group stimulus space in the case of INDSCAL) is then input as the external configuration.

7.5.5.2 Quasi-internal mapping

So far there are no published examples of quasi-internal preference mapping. However, as mentioned above (see Carroll and Chang 1971), choice of the metric option FIT (0) with a double-centred configuration generated from the preference data GENERATE (0) provides an optimal solution for an internal *metric* unfolding analysis which, given the frequent instability of the non-metric unfolding analysis (implemented by MINIRSA), may well be a desirable option to choose.

7.5.5.3 Extending two-mode solutions (large data sets)

Users often want to scale data sets where the number of subjects is too large for stated program limits. In these circumstances PREFMAP may be used to effect an increase in the number of stimuli or subjects. A reasonable strategy is to proceed as follows:

(i) Sample the maximum permitted number of subjects and include the relevant data in a preliminary run of scaling programs to produce a 'core scaling'.
(ii) Take the output configuration and treat it as an 'external' configuration for PREFMAP analysis.
(iii) Select the options within PREFMAP which match the model and transformation of the original scaling.
(iv) Include the remaining data as input 'preference' data for the PREFMAP run, which will then estimate the positioning of the new subject points (or vectors) within the existing configuration.

Although this procedure is most appropriately used for extending the number of subjects for two-way, two-mode scaling models (such as MDPREF and MINIRSA), it may also be used for basic one-mode scaling models (such as SSA and MRSCAL) if the additional data consist, for each new point, of a set of relative distance estimates between the new point and the points already in the fixed configuration. Kruskal (1972b, pp. 5–6) shows how he used essentially this procedure when faced with scaling a data matrix of 10,000 rows (computer malfunctions) by 657 columns (diagnostic tests) to get a configuration of 10,000 points. In essence, subsets of a manageable size were first used to get a rough estimate of the structure and dimensionality of the data. Then a core set of points (in Kruskal's case, 20) is chosen in such a way that they are as well spaced in the configuration as possible. Finally, the remaining points are positioned into the configuration with respect to the core points, by metric or non-metric scaling, as implemented by PREFMAP. This and other ways of scaling large data sets are discussed in Golledge et al. (1981).

Care should be taken to ensure that the transformation is correctly specified (by the parameter FIT) and that the correct model is chosen—S-PHASE (3) for distance models and S-PHASE (4) for vector models.

7.6 Interrelations of INDSCAL, PINDIS and PREFMAP Models

The more complex models discussed in this chapter have a family resemblance.

INDSCAL, PINDIS and PREFMAP each consists of a set of models of increasing complexity, and the type of complexity is also alike, including distance models involving differential dimensional weighting and rotation, and one or more vector models. The relationship can be portrayed as follows:

	Program/Series		
	(a)	(b)	(c)
Model	INDSCAL	PINDIS	PREFMAP
Rotated, Weighted Distance	(IDIOSCAL)	P2	PM1
Weighted Distance	INDSCAL	P2	PM2
(Simple Distance)	SSA/MRSCAL	P0	PM3
Vector	—	P3, P4	PM4

The parallelism is not exact, and not all of the programs exist in the MDS(X) package but, arrayed in this way, certain common elements become clear:

(a) The INDSCAL series is the most basic, and take their input from one or more (dis)similarity matrices. The most complex model, IDIOSCAL (IDIosyncratic Orientation SCALing) is a fairly predictable generalisation of INDSCAL where each subject is thought of as carrying out an idiosyncratic rotation of the group stimulus space axes, followed by a weighting of those axes. Unfortunately, as a program it has a number of sub-optimal characteristics, and far from convincing examples of its use have been published. The basic references are Carroll and Wish (1973, pp. 90–2 and 1974, pp. 440–1).

The basic metric and non-metric models, MRSCAL and MINISSA, are obviously special cases of INDSCAL—when each subject has identical sets of weights. There also exists in the original MINI series (and elsewhere) variants of the basic vector ('factor analysis') model, (see, for example, Lingoes et al. 1979, pp. 268–9 (SSA-III) and pp. 307–9 (MINI-NFA)). The widespread availability of (metric) factor analysis programs and of the closely related classic scaling procedure in MRSCAL make its inclusion in the MDS(X) series superfluous. In its non-metric form it has rarely been used, since product moment correlations are (inversely) monotonic with distances (see Appendix A2.1), making it a somewhat redundant model.

(b) The PINDIS series distance models correspond closely to their INDSCAL analogues, except that they take individual configurations as input. The PINDIS series also includes the vector series and the hybrid P5 (double-weighting) model.

(c) The PREFMAP series differ from the other two series in taking rectangular (row-conditional) data as basic input, and representing each row element by an ideal point (distance models) or an ideal vector. Otherwise, the models correspond to their analogues in the other series.

It should also be clear by now that the form of transformation chosen is relatively unimportant compared to the form of the model. Indeed, it is rather ironic that after the 'non-metric revolution' has been accomplished it turns out that in many cases the linear (metric) assumption is a good (and considerably less costly) approximation to the monotonic one. That said, it is as well to err on the side of caution and allow the extensive family of monotonic functions to suggest what more regular transformation might be more appropriate.

APPENDIX A7.1 BASIC GEOMETRIC TRANSFORMATIONS (TWO-DIMENSIONAL)

A.7.1.1 Linear mapping or transformation of a point: $P \rightarrow P'$

A point is defined by its co-ordinates on each dimension: $P = (x, y)$. If two points are related by a linear equation, then one is said to be a *linear mapping* of the other:

$$\begin{cases} a_1x + b_1y = x' \\ a_2x + b_2y = y' \end{cases}$$

This defines the mapping of point $P = (x, y)$ into point $P' = (x', y')$. Put in another way, the coefficients a and b can be gathered into a *transformation matrix* **T** which, when applied to P, maps it into P':

$$\underset{\mathbf{T}}{\begin{pmatrix} a_1 b_1 \\ a_2 b_2 \end{pmatrix}} \underset{\mathbf{P}}{\begin{pmatrix} x \\ y \end{pmatrix}} = \underset{\mathbf{P'}}{\begin{pmatrix} x' \\ y' \end{pmatrix}}$$

Example **Q**′ is a linear mapping of **Q** (Figure A7.1)

$$\underset{\mathbf{T}}{\begin{pmatrix} 0.3 & 0.2 \\ 0.1 & 0.4 \end{pmatrix}} \underset{\mathbf{Q}}{\begin{pmatrix} 6 \\ 5 \end{pmatrix}} = \underset{\mathbf{Q'}}{\begin{pmatrix} 2.8 \\ 2.6 \end{pmatrix}}$$

A.7.1.2 Origin and translation

The *origin* of the space is normally defined as the point (0, 0), and all other points are defined by reference to this origin.

It is sometimes useful to move the origin of the space to a new position. This is termed *translation* of the origin (and axes). A particularly common translation is to the centroid—the average co-ordinate or centre of gravity—of a configuration of points. If the origin is moved to (a, b), then the co-ordinates of a point $\mathbf{P} = (x, y)$ relative to the *new* origin are given by: $\mathbf{P} = (x - a, y - b)$.

In the example below, if the origin is translated to (3, 4) then the point **P** has the co-ordinates (2, 1) in terms of the old origin and has the co-ordinates $(-1, -3)$ in terms of the new origin (Figure A7.2).

Note Translation preserves distances (identically).

Translation does *not* preserve scalar products or angular separation between vectors (see Appendix A2.1, Figure A2.3).

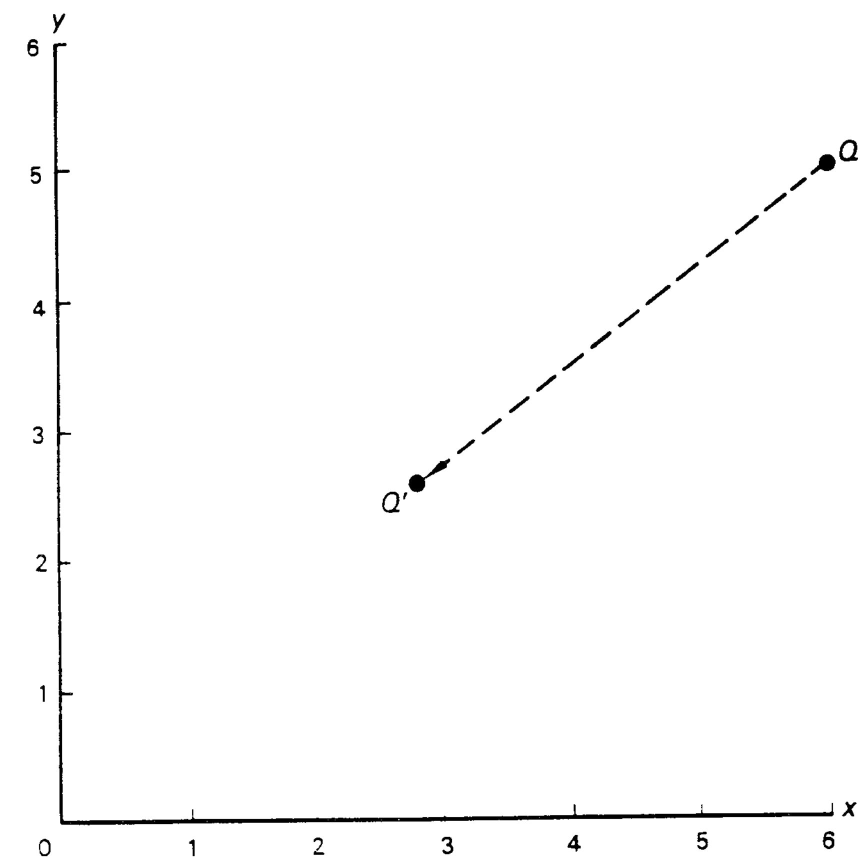


Figure A7.1 *Linear mapping*

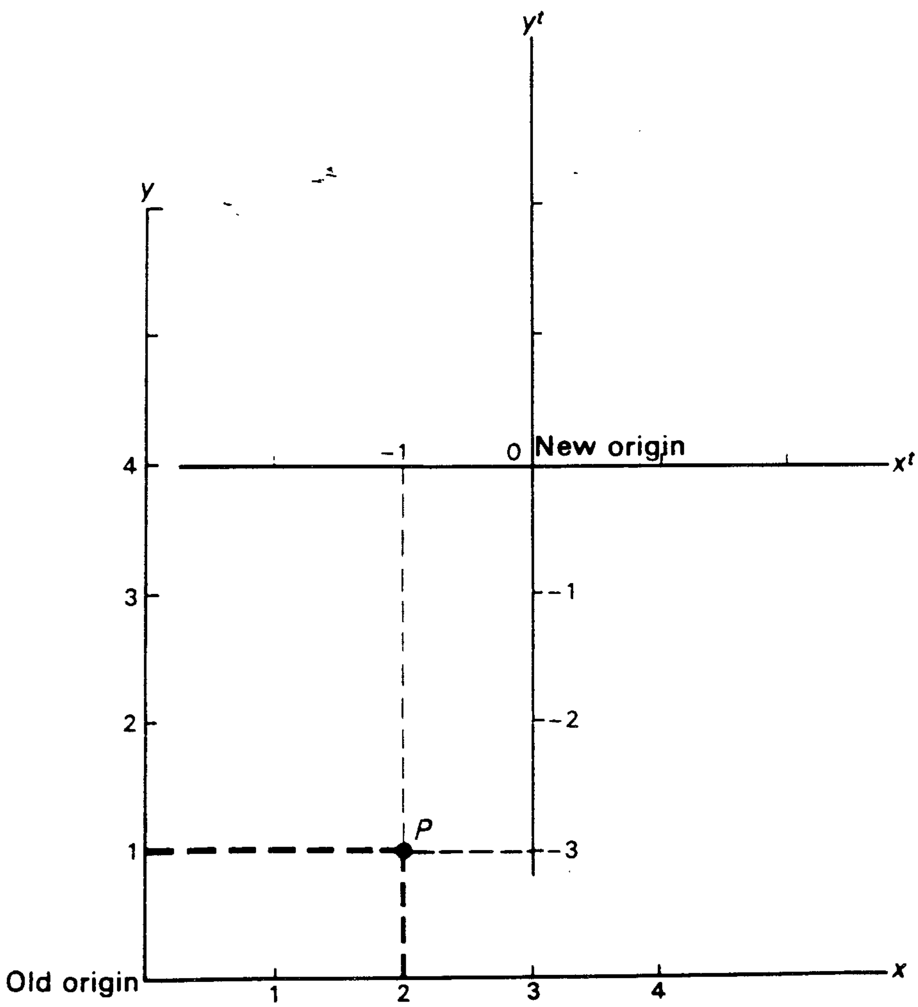


Figure A7.2 *Translation of origin and axes*

A7.1.3 Elementary transformations (null and unit)

Two very basic transformations (which serve as the geometric analogue of zero and unity in algebra) are the *null* and the *identity* transformations:

(i) *The null matrix:* $\mathbf{T} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$

projects all points to the origin (0, 0) of the space and in so doing destroys all information (and obviously preserves neither distances nor scalar products),

e.g. $$\underset{\mathbf{T}}{\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}} \underset{\mathbf{P}}{\begin{pmatrix} 2 \\ 1 \end{pmatrix}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \mathbf{O} \text{ (origin)}$$

(ii) *The unit matrix:* $\mathbf{T} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

simply preserves the exact location of all points and is therefore an identity mapping. It preserves all information (including distances and scalar products),

e.g. $$\underset{\mathbf{T}}{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \underset{\mathbf{P}}{\begin{pmatrix} 2 \\ 1 \end{pmatrix}} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \mathbf{P}'$$

A7.1.4 Diagonal transformations (reflection and rescaling)

A particularly important set of geometric transformations involve diagonal transformation matrices (whose off-diagonal elements are zero). Clearly, the unit matrix is one example; other relevant instances are:

reflection;
uniform rescaling;
differential rescaling (weighting of axes);

(i) *Reflection*

Reflection of points in 2-space occurs when the sign (+, −) of *one* of the co-ordinate axes is changed. The effect is to 'flip' the configuration symmetrically. This is achieved by a diagonal transformation matrix, *one* of whose elements is −1.

Reflection in y-axis

$$\mathbf{T} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

Reflection in x-axis

$$\mathbf{T} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Note Reflection preserves distances and scalar products identically.

Example $P = (2, 1)$ and $Q = (4, 2)$ (see Figure A7.3)

(a) *Reflection in y-axis*

$$\underset{\mathbf{T}}{\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}} \underset{\mathbf{P}}{\begin{pmatrix} 2 \\ 1 \end{pmatrix}} = \underset{\mathbf{P}'}{\begin{pmatrix} -2 \\ 1 \end{pmatrix}} \text{ and } \underset{\mathbf{T}}{\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}} \underset{\mathbf{Q}}{\begin{pmatrix} 4 \\ 2 \end{pmatrix}} = \underset{\mathbf{Q}'}{\begin{pmatrix} -4 \\ 2 \end{pmatrix}}$$

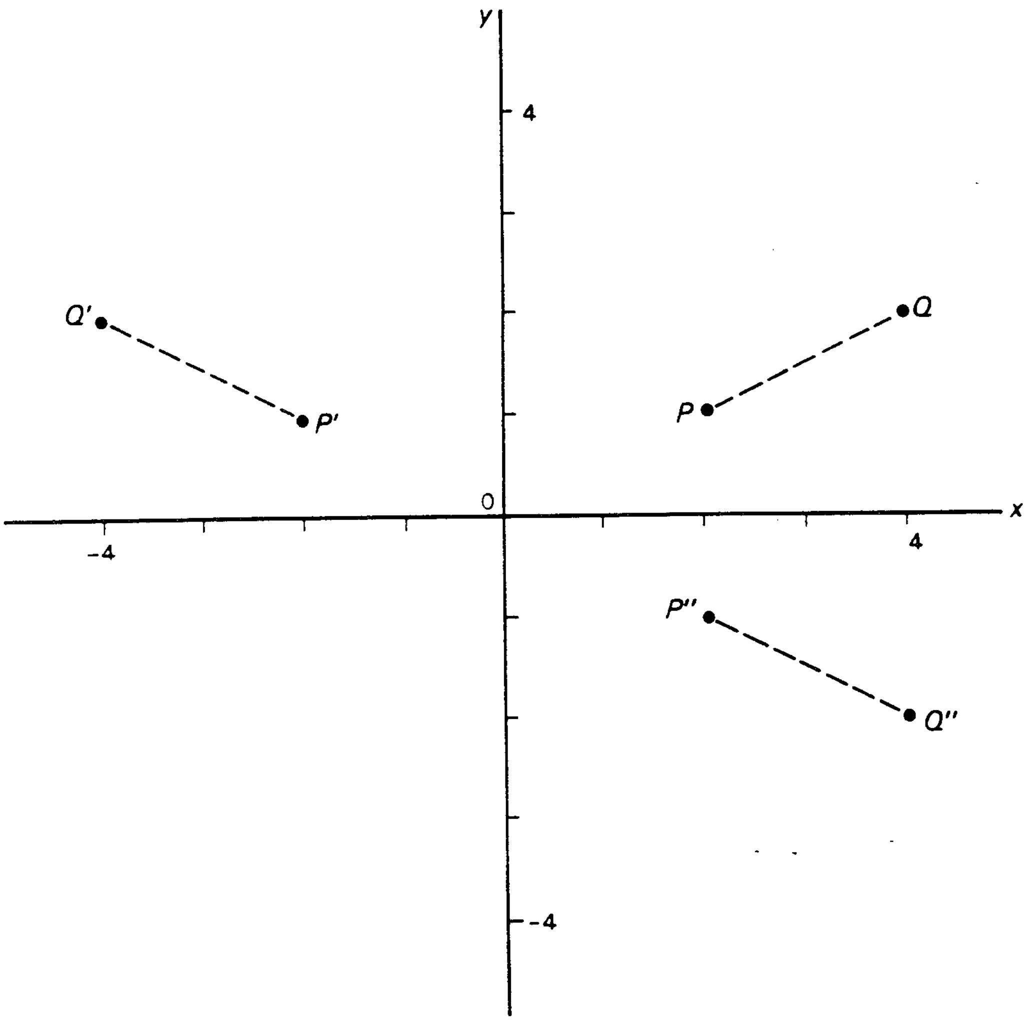


Figure A7.3 *Reflection on x-axis*

(b) *Reflection in x-axis*

$$\begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}\begin{pmatrix}2\\1\end{pmatrix} = \begin{pmatrix}2\\-1\end{pmatrix} \text{ and } \begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}\begin{pmatrix}4\\2\end{pmatrix} = \begin{pmatrix}4\\-2\end{pmatrix}$$

$$\mathbf{T} \quad \mathbf{P} = \mathbf{P}'' \qquad \mathbf{T} \quad \mathbf{Q} = \mathbf{Q}''$$

(ii) *Uniform weighting (rescaling)*
A diagonal transformation matrix *with identical diagonal elements a,*

$$\mathbf{T} = \begin{pmatrix}a & 0\\0 & a\end{pmatrix}$$

produces a uniform stretching ($a > 1$) or shrinking ($0 < a < 1$) of point locations, with each axis weighted by a factor of a. The effect is to rescale a configuration by a factor of a.

Note Uniform weighting/rescaling preserves *relative* distances and scalar products, up to a proportional rescaling by a, i.e. a ratio scale.

Example $\mathbf{T} = \begin{pmatrix}2 & 0\\0 & 2\end{pmatrix}$

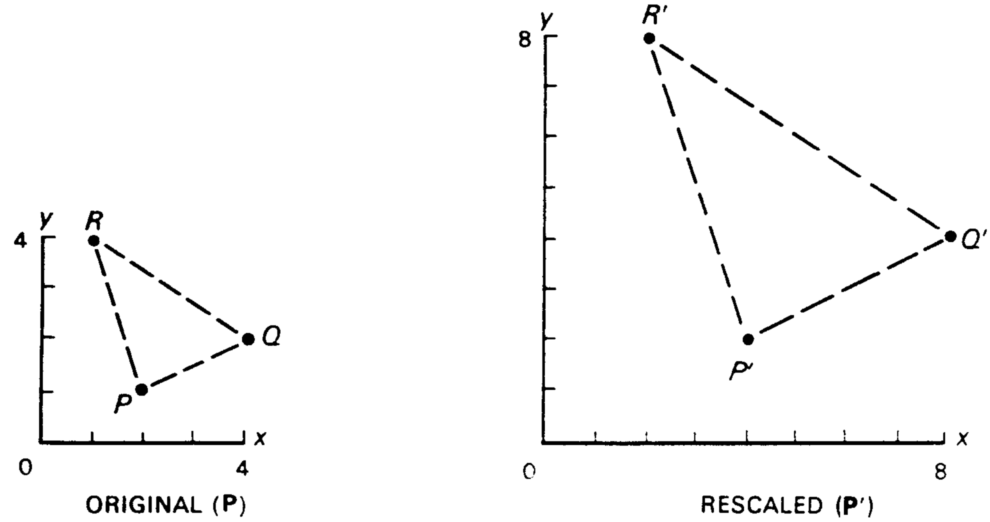


Figure A7.4 *Rescaling with uniform weighting*

and the point co-ordinates for P, Q and R are gathered into a matrix $\mathbf{P}$:

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \overset{\begin{matrix} P & Q & R \end{matrix}}{\begin{pmatrix} 2 & 4 & 1 \\ 1 & 2 & 4 \end{pmatrix}} = \overset{\begin{matrix} P' & Q' & R' \end{matrix}}{\begin{pmatrix} 4 & 8 & 2 \\ 2 & 4 & 8 \end{pmatrix}}$$

$$\mathbf{T} \quad \mathbf{P} \quad = \quad \mathbf{P}'$$

In the distance matrix below, the Euclidean distances in the *original* configuration ($\mathbf{P}$) are given beneath the diagonal, and those in the rescaled configuration ($\mathbf{P}'$) are given above the diagonal. Note that the distances are doubled, since $a = 2$.

$$\mathbf{D} = \begin{pmatrix} & P & Q & R \\ P & 0 & 4.47 & 6.32 \\ Q & 2.24 & 0 & 7.21 \\ R & 3.16 & 3.61 & 0 \end{pmatrix}$$

(iii) *Differential weighting (rescaling)*
A diagonal transformation matrix whose diagonal values are *not* equal

$$\mathbf{T} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \quad \text{with} \quad a \neq b$$

produces a differential elongation or compression of the X and Y axes, with the X-axis weighted by a and the Y-axis by b. The effect is to distort the configuration along the directions of the axes, increasing the co-ordinate values if the weight is greater than 1, and decreasing them if it is less than 1. In the process of differential rescaling the original distances are changed, often dramatically so if the weights are very different.

Note Differential weighting does *not* preserve even relative distances or scalar products. It is not therefore a similarity transformation (see below).

$$\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 3 \end{pmatrix} \overset{\begin{matrix} P & Q & R \end{matrix}}{\begin{pmatrix} 2 & 4 & 1 \\ 1 & 2 & 4 \end{pmatrix}} = \overset{\begin{matrix} P'' & Q'' & R'' \end{matrix}}{\begin{pmatrix} 1 & 2 & \frac{1}{2} \\ 3 & 6 & 12 \end{pmatrix}}$$

$$\mathbf{T} \quad \mathbf{P} \quad = \quad \mathbf{P}''$$

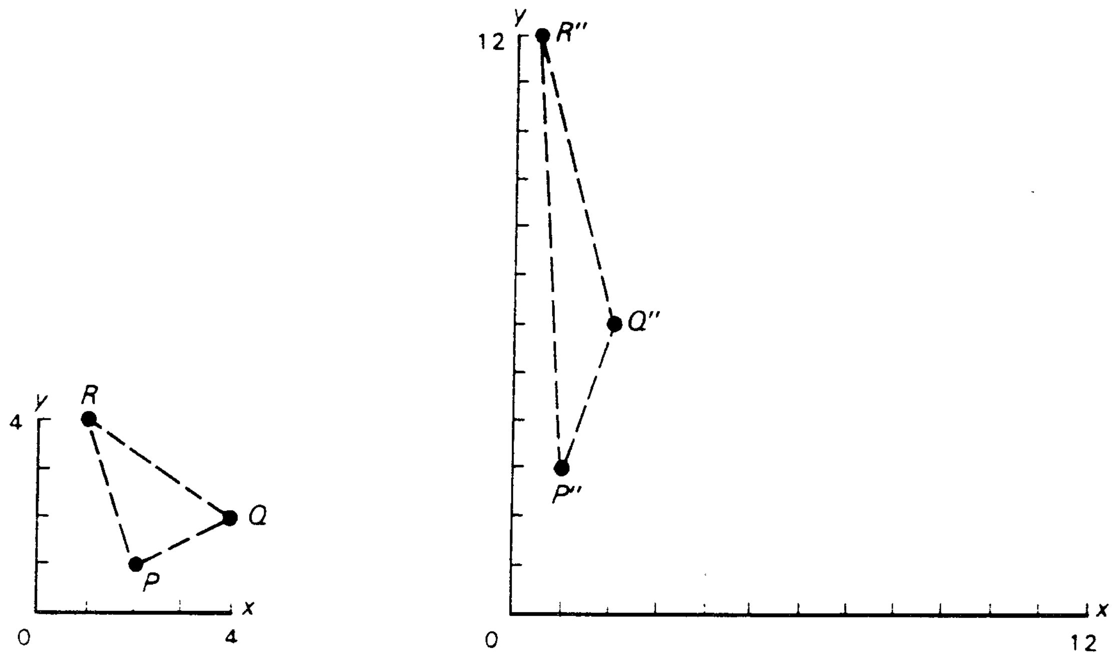


Figure A7.5 *Rescaling with differential weighting*

$$\mathbf{D} = \begin{array}{c} \\ P \\ Q \\ R \end{array} \begin{array}{c} \begin{array}{ccc} P & Q & R \end{array} \\ \begin{pmatrix} 0 & 3.16 & 9.01 \\ 2.24 & 0 & 6.19 \\ 3.16 & 3.61 & 0 \end{pmatrix} \end{array}$$

(original distances below diagonal, rescaled above diagonal)

A7.1.5 (Orthogonal) rotations

Very often the co-ordinate axes of a set of points are arbitrary, especially for Euclidean distance calculations, and another set of co-ordinate axes (such as principal components) may be preferable or have more desirable properties. The move from one set of co-ordinate axes to another is termed a *rotation of axes* about the origin of the space. In the 2-dimensional case, this move is described in terms of the *direction* and the *extent* of the change. (Here we assume a rigid or *orthogonal* rotation, keeping the axes at right angles.) The axes may be moved clockwise, or in an anticlockwise direction, and the extent of the rotation is given by the angle θ through which the axes move.

(i) *Permutation of axes.*

A rigid rotation anticlockwise through 90° is effected by a rotation matrix of the form $\mathbf{T}_a$, (where the subscript a denotes an anticlockwise rotation)

$$\mathbf{T}_a = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

This has the effect of permutating the x into the y axis, or moving the co-ordinate system through a right angle. The rotation matrix, $\mathbf{T}_c$, (where the subscript c denotes a clockwise rotation)

$$\mathbf{T}_c = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

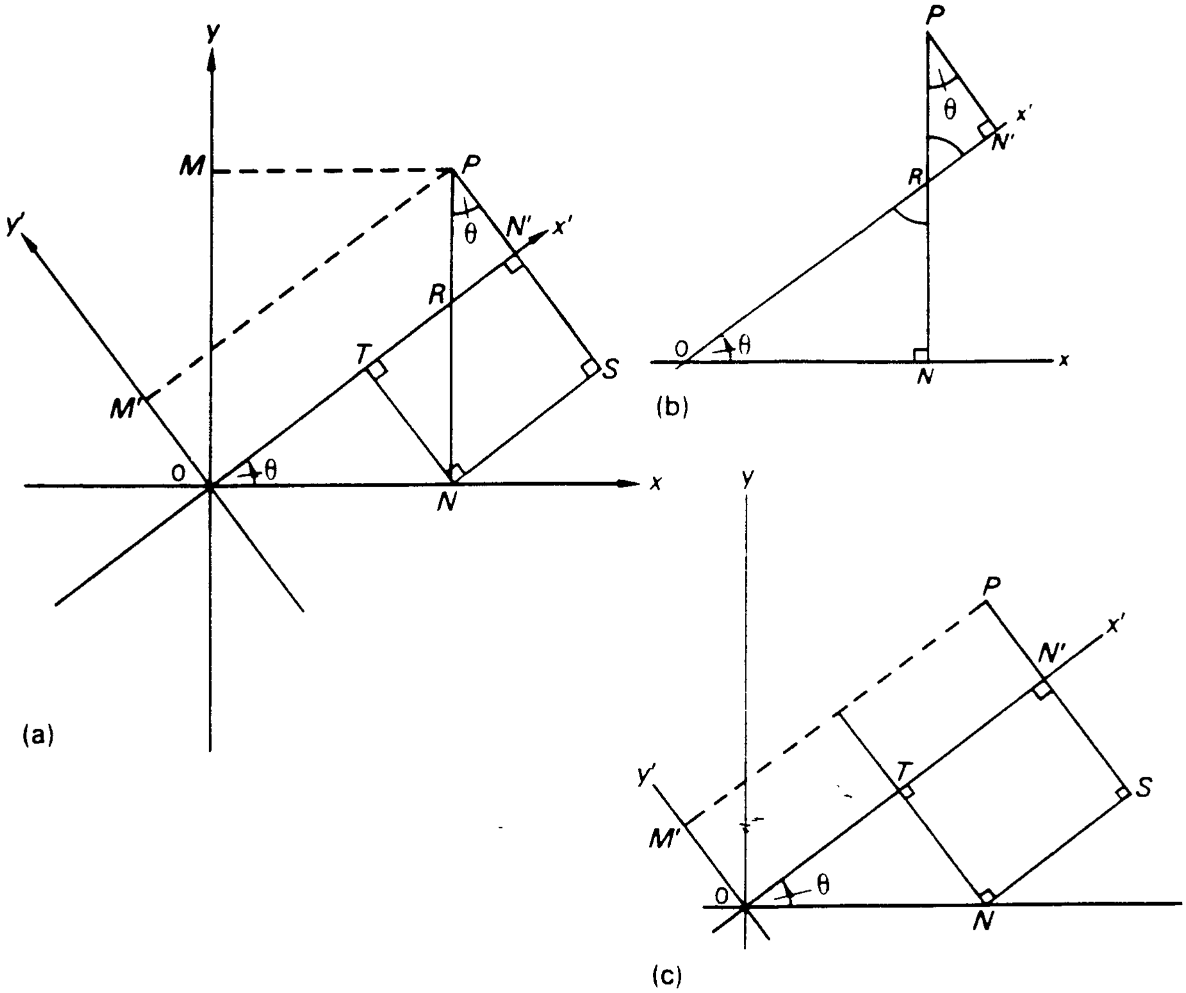


Figure A7.6 *The geometry of rotation of axes*

has the opposite effect; rotating the axes in a clockwise direction through 90°.

(ii) *Rotation through an angle*

For any angle other than 90°, the anticlockwise rotation matrix $\mathbf{R}_a$, taking the original x, y axes into the new x', y' axes has the form:

$$\mathbf{R}_a = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

where θ is the positive angle through which the original reference axes are rotated. (Readers who wish to ignore the trigonometry of this rotation matrix should skip the following section and move to the example, simply noting that if $\theta = 90°$, then $\cos\theta = 0$ and $\sin\theta = 1$, reducing $\mathbf{R}_a$ to the simple form $\mathbf{T}_a$ noted earlier.)

How is the rotation matrix derived? Although not obvious, the procedure is quite simple. You are advised to follow the steps in Figure A7.6 (a)–(c).

First we do some elementary geometry.

(1) We identify a point P with reference to the old axes (x, y) and the new axes (x' y') which have been rotated through an angle of θ degrees.

(2) The co-ordinates of P on x and y are N and M respectively, its co-ordinates on x' and y' are respectively N' and M'.

(3) It is useful to convince yourself at this stage that the length OM = the length PN and, similarly $MP = ON$, $OM' = N'P$ and $ON' = M'P$.

(4) We label the point where PN crosses x' as R for convenience, and consider the two triangles ORN and RPN' (see inset b). We know that the angles of a triangle sum to 180°, and therefore the angle ORN must be $180° - (90° + \theta°) = 90° - \theta°$. The angle PRN' is also $90° - \theta°$, thus the angle RPN' must be $\theta°$. We will make use of this fact later.

(5) We now construct a rectangle $TN'SN$, redrawn for clarity in inset c. Again the reader should be convinced that the length ON' is equal to $OT + TN'$ and ON' *is also equal to* $OT + NS$ (since $TN' = NS$ by construction).

(6) Also, OM' is equal to $PS - SN'$ and OM' *is also equal to* $PS - TN$ (since again $SN' = TN$ by construction).

(7) Returning to the large diagram (inset a) we are now in a position to give expressions for the new co-ordinates of P (N' and M') in terms of the old co-ordinates (N and M), thus

$$ON' = OT + NS$$

and $$OM' = PS - TN$$

(8) Now consider the triangles OTN and PNS. The reader should be convinced that these right-angled triangles are similar, having identical angles (90°, θ and $90° - \theta°$).

(9) We know that in a right-angled triangle the cosine of an angle θ is given by the ratio of the side adjacent to the angle to the hypotenuse. Thus, in triangle OTN the cosine of θ ($\cos\theta$) is OT/ON, and in PNS, $\cos\theta = PS/PN$.

(10) Similarly, we know that the sine of an angle ($\sin\theta$) is given by the ratio of the side opposite the angle to the hypotenuse. Thus, in PNS the sine of θ ($\sin\theta$) is NS/PN, while in OTN, $\sin\theta = TN/ON$.

(11) We are now in a position to do some algebra. Consider the fact that

$$ON' = OT + NS \tag{1}$$

We seek an expression for ON' in terms of O, N, M, P and the angle θ (i.e. an expression which will relate the new axes to the old).

We know that

$$\cos\theta = \frac{OT}{ON}.$$

Cross-multiplying gives

$$OT = ON\cos\theta. \tag{2}$$

Similarly we know

$$\sin\theta = \frac{NS}{PN}.$$

Again, cross-multiplying gives

$$NS = PN\sin\theta. \tag{3}$$

Thus, substituting (2) and (3) back in (1), we obtain

$$ON' = ON\cos\theta + PN\sin\theta. \tag{4}$$

We may treat the expression

$$OM' = PS - TN$$

in precisely the same way.

$$\cos\theta = \frac{PS}{PN}$$

$$PS = PN\cos\theta$$

$$\sin\theta = \frac{TN}{ON}$$

$$TN = ON\sin\theta$$

Therefore $$OM' = PN\cos\theta - ON\sin\theta \qquad (5)$$

Thus in (4) and (5) we have derived the desired result: an expression for the new co-ordinates in terms of the old co-ordinates and the angle of rotation.

Now, letting x stand for ON, y for OM, x' for ON' and y' for OM' we may write

$$\left.\begin{aligned} x' &= x\cos\theta + y\sin\theta \\ y' &= -x\sin\theta + y\cos\theta \end{aligned}\right\}$$

which in matrix form is

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

i.e.

$$\mathbf{x}' = \mathbf{R}_a\mathbf{x}$$

which is the desired result.

An example will illustrate the procedure. In Figure A7.6a. the angle of rotation ,θ, is almost 37°*. In this instance, the position of P with respect to the new axes (X', Y') is obtained as follows:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos 37° & \sin 37° \\ -\sin 37° & \cos 37° \end{pmatrix}\begin{pmatrix} 4.0 \\ 5.2 \end{pmatrix} = \begin{pmatrix} 6.32 \\ 1.75 \end{pmatrix}$$

If we wish to rotate a whole configuration of P points in 2 dimensions, the same rotation matrix is applied to each point. (In more than 2 dimension the rotation procedure is slightly more complicated, but consists of taking all pairs of dimensions and proceeding in the above manner.)

A7.1.6 Transformations: a summary

In discussing the various geometrical transformations employed in MDS, the following are particularly significant and basic:

The *identity or unit* transformation, which leaves a configuration and its absolute and relative distances and scalar products unchanged;

The *(central) dilation or rescaling* transformation, which multiplies each axis by a

*In fact it is 36°52′; the triangle ORN is a 3:4:5 triangle, yielding a cosine of 4/5 and a sine of 3/5.

given scalar value. It preserves relative (but not absolute) distances and scalar products.

The *differential dilation or rescaling* transformation, which multiplies each axis by a different scalar value. It preserves neither relative nor absolute distances (nor scalar products).

The *translation* transformation which removes the origin of the space. It preserves relative and absolute (Euclidean) distances, but changes scalar products.

The *orthogonal rotation* transformation, which preserves Euclidean distances and scalar products identically.

An *extended similarity* transformation (often called a similarity transformation) is a composite transformation, which preserves all relative distances. It may comprise a reflection, a translation of origin, a central dilation and an orthogonal rotation. Taken together these are permissible operations on a distance model configuration.

APPENDIX 7.2 NOTES ON THE ESTIMATION PROCEDURE IN INDSCAL

Full details of the alternating least squares procedure for estimating the parameters of the INDSCAL model are contained in Carroll and Chang (1970), Carroll and Wish (1973) and in the MDS(X) documentation of INDSCAL-S.

What follows here is a brief introduction to the basic method of analysis.

(i) The basic model assumes that the subject's data dissimilarities are a linear function of the distances of the solution

$$\delta_{jk}^{(i)} = L(d_{jk}^{(i)})$$

where the distances refer to the ith individual's (private) space, i.e.

$$d_{jk}^{(i)} = \sqrt{\sum_a (y_{ja} - y_{ka})^2}$$

The first step, as in other metric models, is to *convert the subject's data ('relative distances') into estimates of distances* by calculating an additive constant, as in classic scaling (see 5.2.3.2), which will make the data satisfy the triangle inequality.

(ii) *The data 'distances' are then converted into estimated scalar products*, as in classic metric scaling (see Appendix A5.2), with their origin at the centroid of the points. At this stage, each subject's data are normalised to have equal influence. The relationship between the estimated scalar products (b_{jk}) and the private space co-ordinates is simply:

$$b_{jk}^{(i)} = \sum_a y_{ja}^{(i)} y_{ka}^{(i)} \qquad (1)$$

(iii) The INDSCAL model, stated in its distance form is:

$$\delta_{jk}^{(i)} = \sqrt{\sum_a w_a^{(i)}\left(x_{ja} - x_{ka}\right)^2}$$

and the relationship between the group space co-ordinates (x_{ja}) and the private space co-ordinates (y_{ja}) is:

$$y_{ja}^{(i)} = \sqrt{w_a^{(i)}}x_{ja} \tag{2}$$

Substituting (2) into (1) gives

$$b_{jk}^{(i)} = \sum_a w_a^{(i)} x_{ja} x_{ka} \tag{3}$$

This is the three-way scalar products formulation of the INDSCAL *model.* For notational simplicity, it helps to rewrite (3), putting subject references (i) as subscripts:

$$b_{ijk} = \sum_a w_{ia} x_{ja} x_{ka} \tag{3a}$$

(iv) The estimation of the subject weights (w_{ia}) and group space co-ordinates (x_{ja}) in INDSCAL is performed by a variant of the three-way canonical decomposition model (see 7.2.2), which ensures that the second and third ways (x_{ja} and x_{ka}) are in fact identical.

(v) The INDSCAL model has been shown by Schönemann (1972) to have an exact algebraic solution—for perfect data. In the case of errorful data, an iterative process (which may use Schönemann's method to provide an initial configuration) is employed, using an alternating procedure. It consists of finding a preliminary estimate for the two stimulus weights (x_{ja} and x_{ka}), fixing them, and then estimating (by least squares) the subject weights w_{ia}. Then the x_{ja} are estimated, with the w_{ia} and x_{ka} fixed, and so on.

When a satisfactory approximation to the data is obtained, the process terminates, ways 2 and 3 are set equal, a final estimate of way 1 (subject weights) is made and the weights are then appropriately normalised before being output.

A caution

All variants of alternating least squares estimation procedures are susceptible to a greater or lesser extent to local minimum solutions. In any event, users should be prepared for this eventuality: often ten runs with different starting configurations are necessary before one can be virtually certain that one has an optimal solution. In any event, it would be foolhardy to rely on less than three. In the repeated runs, the group space configurations will probably be very similar *except for slight differences in orientation.* Since subject weights refer directly to a particular orientation and will often change considerably under relatively small rotations of the dimensions, particular attention should therefore be paid to how the group space dimensions change and to the individual and overall goodness of fit measures.