

## Chapter 1

### Introduction

## 1.1 Preliminaries

In the past few decades a large class of data analysis techniques have emerged resulting in a spatial representation of data. The general aim of such a spatial representation is to aid the researcher in understanding the relationships between the observational units, or objects, about which information has been gathered, especially when the underlying dimensions are not known. The objects, which can be of any kind, are represented as a geometric configuration of points, usually (but not necessarily) in a Euclidean space of limited dimensionality. Each object corresponds to one point, and similarity between objects is reflected in the distance between points. Not only does this result in quantified objects (since the coordinates of a point quantifies the corresponding object), but the ordering of the objects on each dimension can also be used to give an interpretation to those dimensions.

An important subclass of techniques aimed at spatial representation of data is known as multidimensional scaling (MDS). For recent textbooks on MDS see, for instance, Coxon (1982), Davison (1983), Young and Hamer (1987), and Green, Carmone, and Smith (1989). Some of the techniques known as multivariate data analysis (MVA) also result in a configuration of points. Examples of such configurations are the object scores in principal components analysis and homogeneity analysis, and the factor scores in factor analysis. For recent textbooks on MVA we refer to Van de Geer (1986), Tatsuoka (1988), and Gifi (1990).

It is not our intention to discuss the numerous ways in which data can be spatially represented. We merely want to emphasize that configurations have become an increasingly important tool in helping to understand the structure of data. The present study is concerned with the situation where several configurations are already available, and where the interest lies in the *relationships between these configurations*. Suppose, for example, that a number of researchers have collected data on the same topic and submitted their data to one or another form of multidimensional scaling or multivariate analysis. Each researcher thereby produces an  $m$ -dimensional 'map' or configuration of the stimuli that he or she is studying. The question immediately arises whether their results, that is, their 'maps', are the same or not. Another example is the situation where someone has collected  $n$  (dis)similarity matrices and decides to analyse each (dis)similarity matrix separately with MDS as a first step, instead of performing a simultaneous analysis of the complete three-way data matrix using, for instance,

INDSCAL (cf., Horan, 1969; Carroll and Chang, 1970). Only afterwards she or he starts investigating the possible relations between the  $n$  configurations. This procedure may seem a rather roundabout way to analyse (dis)similarity data, but it has the advantage that also other transformations (more simple as well as more complex ones) than in INDSCAL can be investigated, as will be discussed below. Another example is the situation where a researcher has analysed one and the same data matrix using different measurement levels and/or different approaches to ties in the data and/or even different analysis techniques, and is interested to find out whether the resulting configurations are the same, and, if not, whether there exists a systematic pattern in the differences.

The problem in these situations can be expressed more formally as follows: given a set of  $n$  ( $n \geq 2$ ) configurations  $X_j$  ( $j = 1, \dots, n$ ) of order ( $p \times m$ ), each configuration containing the coordinates of the same  $p$  stimuli in  $m$  dimensions, what is the similarity or *match* between these configurations? This immediately confronts us with two other questions: what are the uniqueness properties of the configurations, that is, which transformations are they allowed to undergo, and what criterion should be used to evaluate the match? To provide an answer to these questions, the *PINDIS models* of LINGOES and BORG (1978) (see also BORG, 1979; BORG and LINGOES, 1987) seem particularly appropriate. PINDIS is an acronym for Procrustean INDividual Differences Scaling, and refers both to a series of six increasingly complex models, as well as to the algorithms used to estimate the model parameters, and the computer program in which these algorithms were collectively implemented. In all six PINDIS models certain linear transformations are applied to the given configurations  $X_j$ , and the well-known least squares criterion is used to evaluate the match between the optimally transformed configurations. We will briefly discuss each PINDIS model in turn; for a more detailed description of these models and of the transformations involved we refer to LINGOES and BORG (1978), and to the next three chapters of this book.

## 1.2 The PINDIS models of Lingoes and Borg

The first and most simple PINDIS model was called 'the similarity transformation model' by Lingoes and Borg, but since it is much better known in the literature under the name Generalized Procrustes Analysis (GPA) as originally developed by Gower (1975), we will refer to it as the *GPA model* throughout this monograph. Let  $\mathbf{R}_j$  be an unknown orthonormal matrix of order  $(m \times m)$ ,  $s_j$  an unknown central dilation, and  $\mathbf{u}_j$  an unknown translation vector of order  $(m \times 1)$ . Further defining  $\mathbf{Z}$  as a group or centroid configuration of order  $(p \times m)$ ,  $\mathbf{E}_j$  as a  $(p \times m)$  matrix of residuals, and  $\mathbf{1}$  as a  $(p \times 1)$  vector containing ones, the GPA model can be written as

$$s_j(\mathbf{X}_j - \mathbf{1}\mathbf{u}_j')\mathbf{R}_j = \mathbf{Z} + \mathbf{E}_j, \quad \text{for } j = 1, \dots, n. \quad (1.1)$$

Orthonormal transformations, translations, and central dilations do not affect the relative distances between the stimulus points of a configuration. Since the configurations obtained from MDS and MVA are often generated according to the Euclidean distance model, and because such a configuration has the property of containing information that is unchanged by distance preserving transformations, it is only natural to start the analysis of  $n$  configurations with the GPA model.

If a generalized Procrustes analysis does not result in a satisfactory match, this may be due to the fact that the relations between the configurations are of a more complex nature. In PINDIS this can be investigated by evaluating the effect of more complex transformations which are still linear but no longer preserve relative distances between stimulus points.

In the second and third PINDIS model a transformation called *dimension weighting* is introduced, where dimensions are allowed to be stretched or shrunk (i.e., weighted) differently when matching the  $n$  configurations. Following the name of the transformation, the second and third model of Lingoes and Borg are called the *dimension weighting models*. Let  $\mathbf{Q}_j$ ,  $\mathbf{S}$ , and  $\mathbf{S}_j$  denote unknown orthonormal matrices of order  $(m \times m)$ ,  $\mathbf{g}_j$  and  $\mathbf{h}_j$  unknown translation vectors of order  $(m \times 1)$ , and  $\mathbf{W}_j$  an unknown diagonal matrix of order  $(m \times m)$ . Also, let  $\mathbf{Z}$  be the centroid configuration obtained in model (1.1), then the second PINDIS model can be written as

$$(\mathbf{X}_j - \mathbf{1}\mathbf{g}_j')\mathbf{Q}_j = (\mathbf{Z} - \mathbf{1}\mathbf{h}_j')\mathbf{S}\mathbf{W}_j + \mathbf{E}_j, \quad \text{for } j = 1, \dots, n, \quad (1.2)$$

while the third PINDIS model takes the following form:

$$(\mathbf{X}_j - \mathbf{1g}_j)\mathbf{Q}_j = (\mathbf{Z} - \mathbf{1h}_j)\mathbf{S}_j\mathbf{W}_j + \mathbf{E}_j, \quad \text{for } j = 1, \dots, n. \quad (1.3)$$

The diagonal matrices  $\mathbf{W}_j$  in (1.2) and (1.3) have the effect of differentially weighting the axes of the (optimally rotated) group configuration  $\mathbf{Z}$  obtained from GPA. The essential difference between the two models is that  $\mathbf{Z}$  is only rotated once by  $\mathbf{S}$  in (1.2), whereas it is rotated differently by  $\mathbf{S}_j$  for each configuration  $j$  in (1.3).

Assuming that the individual configurations represent  $n$  subjects, the psychological implication of model (1.2) is that the subjects agree on the underlying dimensions that structure the stimuli under investigation (that is: the axes of  $\mathbf{ZS}$ ), but differ in the importance they attach to the dimensions on which the stimuli are ordered. Model (1.3) can be interpreted as meaning that the subjects not only differ in the importance they attach to the dimensions, but also have their own views on *which* dimensions of  $\mathbf{Z}$  are to be considered as important. These idiosyncratic dimensions are represented by the axes of  $\mathbf{ZS}_j$ . In terms of interpretation, therefore, the two dimension weighting models in PINDIS are related to the INDSCAL and IDIOSCAL models of Carroll and Chang (1970, 1972). These PINDIS models are less parsimonious than the GPA model since they require the estimation of more parameters than in GPA model (1.1). This means that the dimension weighting models always fit the data equally well as or better than the GPA model.

The fourth and fifth PINDIS model proposed by Lingo and Borg are characterized by a transformation that we will call *stimulus weighting* in this monograph. In stimulus weighting the stimulus points of a configuration are connected with the origin of  $m$ -dimensional space, and the thus generated vectors are allowed to be shortened or lengthened differently. Stimulus weighting gives rise to two different models: the *stimulus weighting models*. Let  $\mathbf{g}_j$ ,  $\mathbf{h}$ , and  $\mathbf{h}_j$  be unknown translation vectors of order  $(m \times 1)$ ,  $\mathbf{T}_j$  an unknown orthonormal matrix of order  $(m \times m)$ , and  $\mathbf{V}_j$  an unknown diagonal matrix of order  $(p \times p)$ , then the fourth PINDIS model can be written as

$$(\mathbf{X}_j - \mathbf{1g}_j)\mathbf{T}_j = \mathbf{V}_j(\mathbf{Z} - \mathbf{1h}') + \mathbf{E}_j, \quad \text{for } j = 1, \dots, n, \quad (1.4)$$

and the fifth PINDIS model is

$$(\mathbf{X}_j - \mathbf{1g}_j)\mathbf{T}_j = \mathbf{V}_j(\mathbf{Z} - \mathbf{1h}_j') + \mathbf{E}_j, \quad \text{for } j = 1, \dots, n. \quad (1.5)$$

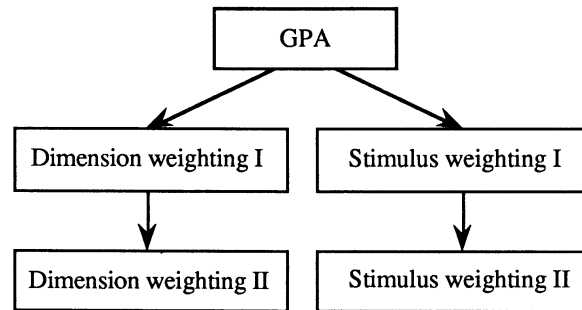
In both models the centroid configuration  $\mathbf{Z}$  is assumed to be fixed, and is equal to the optimal group configuration obtained in the GPA model. The diagonal matrices  $\mathbf{V}_j$  in (1.4) and (1.5) have the effect of differentially weighting the vectors obtained by connecting the stimulus points in  $(\mathbf{Z} - \mathbf{1h}')$  or in  $(\mathbf{Z} - \mathbf{1h}'_j)$  with the origin of  $m$ -dimensional space.

A little thought should convince the reader that the *location* of a configuration is crucial in stimulus weighting. Imagine an arbitrary two-dimensional configuration of points, all points being connected with the origin. Changing the location of this configuration, which is achieved by a translation, will also change its position relative to the coordinate axes. Therefore, the set of vectors obtained by connecting the stimulus points of the *relocated* configuration with the origin will be quite different from the original set: both the lengths of the vectors as well as the angles between the vectors will have been changed. This explains the essential difference between the two stimulus weighting models:  $\mathbf{Z}$  is only translated once in (1.4) while it is translated differently for each configuration  $j$  in (1.5).

Assuming that the  $\mathbf{X}_j$  represent subjects, model (1.4) can be interpreted as follows. All subjects share one common point of view in the group configuration  $\mathbf{Z}$  (represented by the origin from which the stimulus vectors in  $(\mathbf{Z} - \mathbf{1h}')$  emanate). But, since weighting the vectors representing the stimulus points in  $(\mathbf{Z} - \mathbf{1h}')$  has the effect of either pushing the stimulus points further away towards the periphery, or pulling them closer towards the centre or origin of  $m$ -dimensional space, the subjects differ in the 'centrality' they assign to the stimuli. In model (1.5) the subjects not only differ in the centrality they assign to the stimuli, but they also literally take different points of view in the group configuration because  $\mathbf{Z}$  is translated differently for each subject. Since the configurations usually contain more rows than columns (i.e., more stimuli than dimensions), and because one optimal weight is estimated for each row of each configuration  $\mathbf{X}_j$ , the stimulus weighting models are the most 'expensive', and thus also the least parsimonious of the PINDIS models.

Lingoes and Borg (1978) even discussed a sixth model where the combined effects of dimension weights *and* stimulus weights are investigated. This 'double weighting' model has such dazzling implications for a possible psychological interpretation, however, that we will not consider it in the present monograph.

The relations between the PINDIS models (1.1) through (1.5) are shown in Figure 1.1. As the figure illustrates the five models in fact form two hierarchies: the first



**Figure 1.1** Scheme of hierarchy in the PINDIS models.

enabling one to assess the gain in fit obtained by dimension weighting compared to GPA, the second allowing one to establish the gain in fit by stimulus weighting compared to GPA. In both hierarchies the GPA model plays the role of a 'yardstick' or reference model for the more complex models. Within each hierarchy the first weighting model (whether it is dimension or stimulus weighting) further acts as a reference model for the second weighting model.

A nice feature of these two hierarchies is that they give one the opportunity to find out whether more complex transformations have anything new to add above more simple transformations. We already noted that if someone has collected  $n$  (dis)similarity matrices there may be reasons to analyse the matrices separately with MDS, and only then to investigate the relations between the resulting  $n$  configurations using the models of Lingoes and Borg. The advantage of this procedure is that, if the GPA model fits the configurations well, a more complex model involving dimension weighting, for example, becomes superfluous. By directly submitting the  $n$  (dis)similarity matrices to an INDSCAL analysis, in general one would not find out that distance preserving transformations are sufficient to uncover a common structure in the data.

Moreover, if weights have anything new to add above transformations that do not distort relative distances, they are not restricted to be applied in just one way (by which we imply: the popular choice for INDSCAL). Instead, one can investigate and compare the results of four different weighting models.

Hence, the PINDIS models offer an attractive way to investigate the relationships between  $n$  configurations containing information about the same  $p$  stimuli.

Unfortunately, quite a number of objections can be raised against the algorithms developed by Lingoes and Borg for the estimation of the unknown parameters in these models, as well as against the PINDIS program itself.

### 1.3 Problems with PINDIS

Lingoes and Borg obtained estimates of the unknown parameters in the PINDIS models by minimizing the sum of the squared residuals corresponding to models (1.1) through (1.5), that is, by minimizing  $\sum \text{tr } \mathbf{E}_j' \mathbf{E}_j$ . They found that none of the models can be solved analytically, and resorted to estimate the parameters by using iterative algorithms. However, only for model (1.1) did their estimation procedures yield a convergent algorithm. The PINDIS algorithms corresponding to model (1.2) through (1.5) are not guaranteed to converge.

Specifically, the procedure used in PINDIS for the estimation of the unknown rotation matrices  $\mathbf{Q}_j$  in models (1.2) and (1.3), and of the unknown rotation matrices  $\mathbf{T}_j$  in models (1.4) and (1.5) can be easily improved. Lingoes and Borg based their estimation procedures for these matrices on results derived for the two-dimensional case. Consequently, in higher dimensional spaces they were forced to use an iterative procedure by performing planar rotations for all  $\binom{m}{2}$  combinations of the  $m$  dimensions. They could not prove that this procedure must converge. Lingoes and Borg remarked on this subject (1978, p.506):

It would be possible, however, to use an alternating method analogous to the one used in INDSCAL and ALSCAL, for example. If the spaces were higher than two-dimensional, this would have the advantage that the rotation matrices would be computed directly rather than in the plane-wise manner, which cannot be proved to converge necessarily. ... In evaluation of the entire matrix approach, one could say that it has clear advantages over the scalar solutions previously discussed whenever the transformational analyses are carried out in higher dimensional spaces.

Since the estimation of the rotation matrices  $\mathbf{Q}_j$  and  $\mathbf{T}_j$  is the well-known orthonormal Procrustes problem for which an analytical solution is available in any number of dimensions (see, e.g., Schönemann, 1966), the latter procedure is to be preferred.

A second objection that can be raised against PINDIS is that the results of the dimension and stimulus weighting models all depend upon the centroid configuration  $\mathbf{Z}$  obtained in GPA model (1.1). Although  $\mathbf{Z}$  is optimal in GPA, this is not necessarily the case in models (1.2) through (1.5).



Another problem specific for the dimension weighting algorithms implemented in PINDIS is that the rotation matrices  $S_j$  in model (1.3) are estimated in an ad hoc way. For each  $S_j$ , and in each plane that can be formed by the  $\binom{m}{2}$  combinations of the  $m$  dimensions, the centroid configuration  $Z$  is rotated in steps of  $5^\circ$  over an interval of  $\pm 45^\circ$ . Then a  $\pm 5^\circ$  sector around the 'best' rotation angle in the previous interval is searched in intervals of  $1^\circ$ . The best angle in the latter search process is picked as the optimal rotation of  $Z$ . This procedure is repeated for all pairwise combinations of the  $m$  dimensions. Then, for want of a better solution, the 'optimal'  $S$  in model (1.2) is determined as follows:  $\bar{Z}$  is defined as the mean of the  $n$  matrices  $ZS_j$  obtained in model (1.3), and  $S$  is identified as the orthonormal matrix bringing  $Z$  in maximal agreement with  $\bar{Z}$  (cf. Lingoes and Borg, 1978, p.500). A specific problem of stimulus weighting model (1.4) in PINDIS is that the 'optimal' translation vector  $h$  in model (1.5) is simply taken as the average of the optimal translation vectors  $h_j$  obtained in model (1.5) (cf. Lingoes and Borg, 1978, p.502). Clearly, all these procedures can only provide rough approximations of the parameters at issue.

Also, the formulas numbered (36) and (51) in the Lingoes and Borg article are not correct. Using our notation these formulas should be

$$h_j = \frac{[V_j Z - (X_j - 1g_j)T_j]'V_j 1}{1'V_j^2 1}, \quad (1.6)$$

and

$$S_j'Z'(X_j Q_j - ZS_j W_j)W_j = \text{symmetric}, \quad (1.7)$$

respectively. Anyone interested in programming the algorithm of Lingoes and Borg corresponding to model (1.5) will find that his or her algorithm behaves erratically if using formula (36) in the 1978 article to update  $h_j$ , while (1.6) gives the expected results. Moreover, formula (51) in the Lingoes and Borg article suggests that the estimation of  $S_j$  in model (1.3) is a simple orthonormal Procrustes problem. But the problem is much more complicated, as is witnessed by (1.7).

We finally note that the output of the PINDIS program available in the MDS(X) series by Davies and Coxon (1983) is very obscure and difficult to understand, and that a number of options which, according to the accompanying manual, should give the user control of the program do not work properly, or do not work at all.

#### 1.4 MATCHALS as an alternative

Clearly, the PINDIS program contains a number of serious shortcomings. The purpose of the present study is to provide better and more general alternating least squares algorithms than those implemented in PINDIS. In the following chapters we will develop convergent algorithms, generalizing all estimation procedures to the  $m$ -dimensional case. Moreover, a number of features will be added that are not available in the PINDIS models. First of all, the five models and the corresponding algorithms will be generalized to the case of incomplete configurations, that is, the case where information about some stimuli in some configurations is missing in any arbitrary pattern. Next, new optimal group configurations will be estimated in the dimension and stimulus weighting models instead of borrowing the group configuration  $\mathbf{Z}$  from the GPA model. Finally, all models will be supplemented with an analysis of variation allowing one to assess the relative contributions of configurations, stimuli and dimensions to the total fit of each model. In this respect we have been inspired by the analysis of variation proposed by Gower (1975) for the GPA model.

Specifically, the incorporation of missing data in GPA yields the following model:

$$\mathbf{M}_j \mathbf{S}_j (\mathbf{X}_j - \mathbf{1} \mathbf{u}_j') \mathbf{R}_j = \mathbf{M}_j \mathbf{Z} + \mathbf{M}_j \mathbf{E}_j, \quad \text{for } j = 1, \dots, n, \quad (1.8)$$

where  $\mathbf{M}_j$  is a given diagonal matrix of order  $(p \times p)$ . This matrix contains ones on the diagonal if the corresponding rows of configuration  $\mathbf{X}_j$  are not missing and zeroes elsewhere. PINDIS model (1.2) will be replaced by the following model

$$\mathbf{M}_j (\mathbf{X}_j - \mathbf{1} \mathbf{g}_j') \mathbf{Q}_j = \mathbf{M}_j (\mathbf{Y} - \mathbf{1} \mathbf{h}_j') \mathbf{W}_j + \mathbf{M}_j \mathbf{E}_j, \quad \text{for } j = 1, \dots, n, \quad (1.9)$$

and we will use

$$\mathbf{M}_j (\mathbf{X}_j - \mathbf{1} \mathbf{g}_j') \mathbf{Q}_j = \mathbf{M}_j (\mathbf{Y} - \mathbf{1} \mathbf{h}_j') \mathbf{S}_j \mathbf{W}_j + \mathbf{M}_j \mathbf{E}_j, \quad \text{for } j = 1, \dots, n, \quad (1.10)$$

instead of model (1.3). The matrix  $\mathbf{Y}$  of order  $(p \times m)$  in (1.9) and (1.10) represents an unknown group configuration. We refer to Chapter 3 for an explanation why matrix  $\mathbf{S}$  contained in model (1.2) no longer needs to be considered in (1.9). The stimulus weighting models (1.4) and (1.5) will be replaced by

$$\mathbf{M}_j (\mathbf{X}_j - \mathbf{1} \mathbf{g}_j') \mathbf{T}_j = \mathbf{M}_j \mathbf{V}_j \mathbf{Y} + \mathbf{M}_j \mathbf{E}_j, \quad \text{for } j = 1, \dots, n, \quad (1.11)$$

and

$$\mathbf{M}_j(\mathbf{X}_j - \mathbf{1}\mathbf{g}_j')\mathbf{T}_j = \mathbf{M}_j\mathbf{V}_j(\mathbf{Y} - \mathbf{1}\mathbf{h}_j') + \mathbf{M}_j\mathbf{E}_j, \quad \text{for } j = 1, \dots, n, \quad (1.12)$$

respectively. Again, matrix  $\mathbf{Y}$  in the latter two models is assumed to be an unknown group configuration. For the reason why translation vector  $\mathbf{h}$  in model (1.4) has disappeared in (1.11) we refer to Chapter 4 of this monograph.

We have baptized the computer program used to estimate the parameters in (1.8) through (1.12) the MATCHALS program, an acronym for MATCHing configurations by Alternating Least Squares.

### 1.5 Restrictions

In the present monograph we do not generalize the models to the case where the  $n$  configurations have unequal numbers of columns. In other words, we assume throughout that the configurations have the same dimensionality (although they may have different (and thus also deficient) column ranks). When confronted with configurations having unequal numbers of columns, and letting  $m = \max(m_j)$ , where  $m_j$  is the number of columns in configuration  $j$ , we advise to append  $(m - m_j)$  zero columns to each configuration  $\mathbf{X}_j$ . But we are well aware of the fact that this procedure may lead to suboptimal solutions (see, e.g., Ten Berge & Knol, 1984).

Moreover, the present book heavily emphasizes the algorithmic aspects of the models. Consequently, hardly any real life examples are presented, while the statistical aspects of the models are only very briefly touched upon (in the final, fifth, chapter). Before applying the algorithms to empirical configurations which will usually be contaminated with error, we thought it important first to investigate whether the algorithms would succeed in recovering an a priori known structure in the data. This is the reason why we primarily use geometrically constructed data sets to test the algorithms presented in this book.

## 1.6 Overview

The book is organized as follows. In Chapter 2 the efficient algorithm of Ten Berge (1977) for the estimation of the unknown parameters in the GPA model is generalized to the case of missing data. Chapter 3 is devoted to the development of the dimension weighting models (1.9) and (1.10). In Chapter 4 the stimulus weighting models (1.11) and (1.12) are discussed. Chapter 5 summarizes the most important results found in the previous chapters, and provides recommendations for further research.

The book abounds with geometrical figures like squares, pentagons, octagons, cubes, tetrahedrons, dodecahedrons, etc. These figures play such a predominant part because they lend themselves perfectly to illustrate the properties of the MATCHALS models. As for the notation, a bold upper case letter will always denote a matrix (e.g.,  $\mathbf{X}$ ), a bold lower case letter a column vector ( $\mathbf{x}$ , for example), and a plain lower case letter a scalar (e.g.,  $x$ ). For a recent textbook on the matrix algebra used in the present study we refer to Magnus and Neudecker (1988).