# Chapter 3

# Principal Components Analysis With Nonlinear Optimal Scaling Transformations for Ordinal and Nominal Data

Jacqueline J. Meulman

Anita J. Van der Kooij

Willem J. Heiser

## 3.1. Introduction

This chapter focuses on the analysis of ordinal and nominal multivariate data, using a special variety of principal components analysis that includes nonlinear optimal scaling transformation of the variables. Since the early 1930s, classical statistical methods have been adapted in various ways to suit the particular characteristics of social and behavioral science research. Research in these areas often results in data that are nonnumerical, with measurements recorded on scales having an uncertain unit of measurement. Data would typically consist of qualitative or categorical variables that describe the persons in a limited number of categories. The zero point of these scales is uncertain, the relationships among the different categories is often unknown, and although frequently it can be assumed that the categories are ordered, their mutual distances might still be unknown. The uncertainty in the unit of measurement is not just a matter of measurement error because its variability may have a systematic component.

For example, in the data set that will be used throughout this chapter as an illustration, concerning feelings of national identity and involving 25,000 respondents in 23 different countries all over the world (International Social Survey Programme [ISSP], 1995), there are variables indicating how close the respondents feel toward their neighborhood, town, and country, measured on a 5-point scale with labels ranging from *not close at all* to *very close.* This response format is typical for a lot of behavioral research and definitely is not numerical (even though the categories are ordered and can be coded numerically).

### 3.1.1. Optimal Scaling Transformations

An important development in multidimensional data analysis has been the optimal assignment of quantitative values to qualitative scales. This form of optimal quantification (optimal scaling, optimal scoring) is a very general approach to treat multivariate (categorical) data. Taking the linear regression model as a leading case, we may wish to predict a response variable from a number of predictor variables. This objective is achieved by finding a particular linear combination of the predictor variables that correlates maximally with the response variable. Incorporating optimal scaling amounts to further maximization of this correlation, not only over the regression weights but also over admissible nonlinear functions of the predictor variables. For instance, in the National Identity Study data, we may try to find nonlinear scale values of the response categories of the closeness variables that improve the multiple-correlation coefficient for predicting willingness to move because it may be that some response categories equally predict high willingness, whereas other categories strongly differentiate between small steps in low willingness. These nonlinear functions are called transformations, optimal scalings, scorings, or quantifications. In this chapter, we will use both the terms *nonlinear optimal scaling transformations* and *optimal quantifications.* The optimal scaling process turns qualitative variables into quantitative ones. Optimality is a relative notion, however, because it is always obtained with respect to the particular data set that is analyzed.

The nonlinear optimal scaling transformations of ordered categorical or continuous (ordinal) data can be handled by means of *monotonic* transformations, which maintain the order in the original data. Categorical (nominal) data in which the categories are not ordered will be given an optimal quantification (scoring). Nonmonotonic functions can also be used for continuous (numeric) and ordinal variables when nonlinear relationships among the variables are assumed. In these cases, it is often useful to collapse the data in a limited number of categories (sometimes called *binning*) and find an optimal quantification for the categories (see Section 3.6.2). However, if we do not want to lose the fine gradings, we can also fit a monotonic or nonmonotonic spline. A spline is a function that consists of piecewise polynomials of a low degree that are joined at particular points, called *knots.* Of course, special software is required to simultaneously transform and analyze the data.

### 3.1.2. Software for Nonlinear Principal Components: CATPCA

A state-of-the-art computer program, called CATPCA, that incorporates all the features that will be described in this chapter is available from SPSS Categories 10.0 onwards (Meulman, Heiser, & SPSS, 1999). In CATPCA, there is a large emphasis on graphical display of the results, and this is done in joint plots of objects[1] and variables, also called *biplots* (Gower & Hand, 1996). In addition to fitting points for individual objects, additional points may be fitted to identify groups among them, and graphical display can be in a *triplot,* with variables, objects, and groups of objects. Special attention will be given to particular properties that make the technique suited for data mining. Very large data sets can be analyzed when the variables are categorical at the outset or by binning.

Because CATPCA incorporates differential weighting of variables, it can be used as a "forced classification" method (Nishisato, 1984), comparable to "supervised learning" in machine learning terminology. Objects and/or variables can be designated to be supplementary; that is, they can be omitted from the actual analysis but fitted into the solution afterwards. When a prespecified configuration of points is given, the technique may be used for property fitting (external unfolding), that is, fitting external information on objects, groups, and/or variables into the solution (see Section 3.6.1). The information contained in the biplots and triplots can be used to draw special graphs that identify particular groups in the data that stand out on selected variables.

Summarizing, CATPCA can be used to analyze complicated multivariate data, consisting of nominal, ordinal, and numerical variables. A straightforward spatial representation is fitted to the data, and different groups of objects can be distinguished in the solution without having to aggregate the categorical data beforehand. We will discuss the various aspects of the analysis approach, giving attention to its data-analytical, graphical, and computational aspects.

### 3.1.3. Some Historic Remarks on Related Techniques

Historically, the idea of optimal scaling originated from different sources. On one hand, we find the history of the class of techniques that is nowadays usually called (*multiple*) *correspondence analysis,* a literal translation of Benzécri's *L'analyse des correspondances (multiples)* (Benzécri, 1973, 1992). This history can be traced in the work of Fisher (1948),

1. In the CATPCA terminology, the units of analysis are called objects; depending on the application, these can be persons, groups, countries, or other entities on which the variables are defined.

Guttman (1941), Burt (1950), and Hayashi (1952), among others, and in the rediscoveries since the 1970s (among others, see Benzécri, 1992; de Leeuw, 1973; Greenacre, 1984; Lebart, Morineau, & Warwick, 1984; Saporta, 1975; Tenenhaus & Young, 1985). The class of techniques is also known under the names *dual scaling* (Nishisato, 1980, 1994) and *homogeneity analysis* (Gifi, 1981/1990). In the course of its development, the technique has been given many different interpretations. In the original formulation of Guttman (1941), the technique was described as a principal components analysis of qualitative (nominal) variables. There is also an interpretation as a form of generalized canonical correlation analysis (Lebart & Tabard, 1973; Masson, 1974; Saporta, 1975), based on earlier work by Horst (1961a, 1961b), Carroll (1968), and Kettenring (1971).

Another major impetus to optimal scaling was given by work in the area of nonmetric multidimensional scaling (MDS), pioneered by Shepard (1962a, 1962b), Kruskal (1964), and Guttman (1968). In MDS, a set of proximities between objects is approximated by a set of distances in a low-dimensional space, usually Euclidean. Optimal scaling of the proximities was originally performed by monotonic regression; later on, spline transformations were incorporated (Ramsay, 1982). Since the so-called nonmetric breakthrough in MDS in the early 1960s, optimal scaling has subsequently been integrated in multivariate analysis techniques that hitherto were only suited for the analysis of numerical data. Some early contributions include Kruskal (1965), Shepard (1966), and Roskam (1968). In the 1970s and 1980s, psychometric contributions to the area became numerous. Selected highlights from the extensive psychometric literature on the subject include de Leeuw (1973); Kruskal and Shepard (1974); Young, de Leeuw, and Takane (1976); Young, Takane, and de Leeuw (1978); Nishisato (1980); Heiser (1981); Young (1981); Winsberg and Ramsay (1983); Van der Burg and de Leeuw (1983); Van der Burg, de Leeuw, and Verdegaal (1988); and Ramsay (1988). Attempts at systematization resulted in the ALSOS system by Young et al. (1976), Young et al. (1978), and Young (1981) and the system developed by the Leiden "Albert Gifi" group. Albert Gifi's (1990) book, *Nonlinear Multivariate Analysis,* provides a comprehensive system, combining optimal scaling with multivariate analysis, including statistical developments such as the bootstrap. Since the mid-1980s, the principles of optimal scaling have gradually appeared in the mainstream statistical literature (Breiman & Friedman, 1985; Buja, 1990; Gilula & Haberman, 1988; Hastie et al., 1994; Ramsay, 1988). The Gifi system is discussed among traditional statistical techniques in Krzanowski and Marriott (1994).

## 3.2. Graphical Representation

The way we will treat principal components analysis (PCA) is more like a multidimensional scaling (MDS) technique than a technique from the classic multivariate analysis (MVA) domain. The central concept in classical multivariate analysis is the covariance or correlation *among variables.* Consequently, the modeling of the covariance or correlation matrix is the main objective of the analysis; therefore, the persons on which the variables are defined are usually regarded merely as a replication factor. Thus, the role of the persons is confined to acting as intermediaries in obtaining covariance or correlation measures that describe the relationships among the variables. In the multidimensional scaling domain, techniques have been developed for the analysis of a (not necessarily) symmetric square table, with entries representing the degree of dissimilarity *among any kind of objects,* which may be persons. The objective, then, is to map the objects in some low-dimensional space, in which the distances resemble the initial dissimilarities as closely as possible. To make distinctions between MDS and classical MVA more explicit than they would be from a unifying point of view, consider factor analysis, one of the major data-analytic contributions to statistics originating from the behavioral sciences. Unfortunately, from a visualization point of view, the representation of persons became very complicated in the process. The factor-analytic model aggregates observations on persons into an observed covariance matrix for the variables, and the model involved for representing this covariance matrix is focused on the fitting of a matrix incorporating the common covariances among the variables and another (diagonal) matrix that displays the unique variance of each variable. By formulating the data-analytic task through this particular decomposition, the factor scores that would order the persons with respect to the underlying latent variables are undetermined: Although various approaches exist to have the persons reappear, their scores cannot be determined in a unique manner.

In contrast, principal components analysis can be discussed by focusing on the joint representation of persons and variables in a joint low-dimensional space. The variables in the analysis are usually represented as vectors (arrows) in this low-dimensional space. Each variable is associated with a set of component loadings, one for each dimension, and these loadings, which are correlations between the variables and the principal components, give coordinates for the variables to represent them as vectors in the principal component space. The squared length of such a vector corresponds to the percentage of variance accounted for and thus equals the sum of squares of the component loadings across the dimensions. If we sum the squared component loadings in each dimension over the variables, we obtain the eigenvalues. In the CATPCA approach discussed in the sequel of this chapter, a variable can also be viewed as a set of category points. When a variable is visualized as a vector, these category points are located on a line, where the direction is given by the component loadings. There is, however, an alternative to representing the category points on a straight line, which is by displaying them as points in the middle, the *centroid,* of the cloud of associated object points in the low-dimensional representation space. These two ways of representing a variable will be called the *vector* and the *centroid model,* respectively.

### 3.2.1. The Vector Model

A very first description of the vector model can be found in Tucker (1960); Kruskal (1978) used the term *bilinear model,* and Gabriel (1971) invented the name *biplot.* A comprehensive book on biplots is by Gower and Hand (1996). The prefix *bi-* in *bilinear* and *biplot* refers to two sets of entities, the objects and the variables (and not to two dimensions, as is sometimes erroneously assumed). In PCA, the observed values on the $M$ variables are approximated by the inner product of the $P$-dimensional component scores and component loadings for the variables, with $P$ much smaller than $M$. Usually, the classic reference to lower rank approximation is Eckart and Young (1936), but it might be worthwhile to note that this reference is challenged by Stewart (1993), who remarks that the contribution of Schmidt (1907) was much earlier, which is also noted by Gifi (1990). Because the fit is defined on an inner product, one has to make a coherent choice of normalization.[2] Usually, the component scores are normalized to have means of zero and variances equal to 1; the coherent normalization implies that the component loadings are correlations between the variables and the $P$ dimensions of the space fitted to the objects. Component loadings give coordinates for a variable vector in the space, and the angles between the vectors then approximate the correlations between the variables. The inner product of the matrix of component scores and a variable vector approximates a column of the data matrix, and the length of the variable vector in the space equals the correlation between the variable and its approximation.

In the classical PCA biplot, persons are represented as points, and variables are represented as vectors in the same low-dimensional space. In contrast, in the analysis of preference data, in which Tucker's (1960) vector model originated, the persons are represented as vectors and the items are represented as points (for an extended treatment of the vector model in the context of preference analysis, see Carroll, 1968, 1972; Heiser & de Leeuw, 1981). Because we include nonlinear optimal scaling transformations for the variables in principal components analysis, the vector/bilinear model represents not the original categorical variable but the transformed variable, which is given optimal (non)monotonic quantifications for its categories.

### 3.2.2. The Centroid Model

Unlike the vector model that is based on projection, the centroid model is most easily viewed in terms of distances between object points and category points. In the centroid model, each category obtains coordinates that represent the category in the same space as the objects. The centroid model originates from multiple-correspondence analysis (MCA), where a nominal variable is represented as a set of category points, which are in the centroids of the associated objects. The categories of a particular variable partition the cloud of object points into subclouds. When these subclouds overlap considerably, we say that the corresponding variable is a relatively bad discriminator. On the other hand, well-separated subclouds are associated with a good discriminator. When we have chosen the centroid model for two or more variables, and when the solution has a decent fit, the category points that are associated with the same objects will be close together, whereas categories of the same variable will be far apart (each representing a subcloud of object points through its centroid). The weighted mean squared distance of the category points toward the origin gives a measure similar to variance accounted for and has been called the *discrimination measure* (Gifi, 1990).

2. Because the inner product between two vectors $\mathbf{a}$ and $\mathbf{b}$ is defined as $\mathbf{a}'\mathbf{b}$, it remains unchanged if we transform $\mathbf{a}$ into $\tilde{\mathbf{a}} = \mathbf{Ta}$ and $\tilde{\mathbf{b}}$ into $\mathbf{b} = \mathbf{Sb}$, with $\mathbf{S} = (\mathbf{T}')^{-1}$, because $\mathbf{a}'\mathbf{b} = \mathbf{a}'\mathbf{T}'\mathbf{Sb} = \tilde{\mathbf{a}}'\tilde{\mathbf{b}}$. Choosing principal axes and a coherent normalization settles the choice of $\mathbf{T}$ and $\mathbf{S}$ (also see Section 3.2.4).

A special feature of the CATPCA approach is the possibility to fit the vector (bilinear) model and the centroid (distance) model for different variables (or even for the same variable) in a single analysis, a feature not available in other software programs that perform nonlinear principal components analysis.

### 3.2.3. Clustering and Forced Classification

The CATPCA method accommodates differential weights for separate variables. In this way, the centroid model can be used for *forced classification* (a term coined by Nishisato, 1984), which can also be called *supervised learning.* Forced classification is obtained by applying a (very) large weight for the particular variable that we have selected for the classification. Applying this large weight in combination with the centroid model will cause the object points that belong together to cluster into subclouds in the low-dimensional space. The larger the weight that is given, the tighter the clustering will be. This feature is especially attractive when the number of objects is very large and when they can be identified as members of a particular subgroup, such as citizens of different countries (as in the example given below) or members of a particular social group. In these cases, we would not be so much interested in the individual results but in the results for the groups. Because we are dealing with categorical data, it would not make sense to average the data beforehand. The use of a weighted classification variable takes care of this averaging during the analysis, and the size of the weight controls the subsequent clustering of the object points around their centroid.

In this way, we make certain that the classification variable plays a significant role in the first few dimensions of the principal components analysis solution. This property is extremely useful when we would use PCA as a first step in a discriminant analysis to diminish the number of predictors. Such a particular strategy is often used when the number of predictors exceeds the number of objects in the data matrix, as is the case, among others, in genometrics (the analysis of microarray gene expression data), proteometrics, and chemometrics but also in Q-sort data, with judges acting as variables, and with a classification variable available for the objects. In the same manner, CATPCA can be used as a prestep in a multiple regression analysis when the number of predictors exceeds the number of objects. In the latter case, the response variable is included in the analysis, with a much larger weight than the other variables and with the application of the vector model.

### 3.2.4. Different Normalizations

Different normalization options are possible for the display of objects and variables in the low-dimensional Euclidean space. The most commonly used normalization option in principal components analysis is to display the objects in an orthonormal cloud of object points, in which the dimensions themselves have equal variance. Then, the representation of the variables accounts for the differential fit in subsequent dimensions, with the first dimension accounting for most of the variance and subsequent dimensions displaying the variance accounted for (VAF) in a decreasing order. When the object scores are normalized, however, one loses a straightforward distance interpretation with respect to the objects. To attain the latter, one should normalize the component loadings and leave the object scores free (but keeping the inner product fixed). Therefore, an alternative option is provided that should be used if we wish CATPCA to perform a principal coordinates analysis as described in Gower (1966), which is equivalent to the classical MDS method usually attributed to Torgerson (1958). In principal coordinates analysis, the emphasis is on the representation of the objects, and the cloud of object points displays the differential fit in subsequent dimensions (the cloud is not orthonormal but shows a definite shape). The interpretation of nonlinear PCA in terms of distances between objects is given, among others, in Heiser and Meulman (1983) and Meulman (1986, 1992). Whether the object points or the (category points of the) variables are normalized depends algebraically on the allocation of the eigenvalues in the use of the singular-value decomposition to represent both sets of entities in the low-dimensional space. Therefore, in CATPCA, the impact of the eigenvalues (symbolizing the fit) could also be distributed symmetrically over objects and variables (enhancing the joint display, especially when the overall fit is not very large) or handled in a completely customized way to optimize the quality of the joint representation.

### 3.2.5. Different Biplots and a Triplot

For the display of the results, a variety of biplots is available in CATPCA. A biplot can display the objects (as points) and the variables (as vectors), the objects and groups among them (represented by centroids), or the variables with groups of objects (represented by centroids). Combining these three options reveals relationships between objects, groups of objects, and variables, and we call this display a *triplot.* The ultimate summary of the analysis combines the information in the biplots and triplots in one-dimensional displays. These are obtained by taking centroids of the objects, according to a particular (classification) variable, and projecting these centroids on the vectors representing variables of particular interest in the analysis. In this way, the graph identifies particular groups in the data that stand out on the selected variables. The use of the projected centroids representation is demonstrated in Section 3.4.6.

## 3.3. MVA With Different Nonlinear Optimal Scaling Transformations

In the nonlinear transformation process in CATPCA, an appropriate quantification level has to be chosen for each of the variables. The most restricted transformation level is called *numerical;* it applies a linear transformation to the original integer scale values, so that the resulting variables will be standardized. The numerical scaling level fits category points on a straight line through the origin, with equal distances between the points. Instead of a linear transformation, we have the choice between different nonlinear transformations, and these can either be monotonic with the original order of the categories or nonmonotonic.

### 3.3.1. Nominal Transformation and Multiple Nominal Quantifications

When the only fact we will take into account is that a particular subset of the objects is in the same category (whereas others are in different ones), we call the transformation *nominal* (or *nonmonotonic*); the quantifications only maintain the class membership, and the original categories are quantified to give an optimal ordering. The nonlinear transformation can be carried out either by a least squares identity regression (which amounts to averaging over objects in the same category) or by fitting a nonmonotonic regression spline. Geometrically, the nominal scaling level fits category points in an optimal order on a straight line through the origin. The direction of this straight line is given by the corresponding component loadings.

What has been labeled the centroid model above (a categorical variable represented by a set of points located in the centroid of the objects that are in the associated categories) is also called a *multiple* nominal quantification. The quantification is called multiple because there is a separate quantification for each dimension (the average of the coordinates of the objects in the first dimension, the second dimension, etc.) and nominal because there is no prespecified order relationship between the original category numbers and the order in any of the dimensions. An example of the difference between a nominal and a multiple nominal quantification will be given later on. We choose a nominal transformation when we wish the category points to be represented on a vector and a multiple quantification when we wish them to be in the centroids of the associated objects.

### 3.3.2. Monotonic and Nonmonotonic Splines

Within the domain of either monotonic or nonmonotonic transformations, two approaches are available: optimal least squares transformations or optimal spline transformations. As indicated above, the class of monotonic transformations has its origin in the nonmetric multidimensional scaling literature (Kruskal, 1964; Shepard, 1962a, 1962b), in which original dissimilarities were transformed into pseudo-distances to be optimally approximated by distances between object points in low-dimensional space. Free monotonic transformations have been implemented since then to generalize multivariate analysis techniques as well (e.g., see Gifi, 1990; Kruskal, 1965; Kruskal & Shepard, 1974;Young et al., 1978). We call these transformations free monotonic because the number of parameters that is used is free. Because this freedom could lead to overfitting of the MVA model over the transformation of the variables, a more restricted class of transformations was introduced into the psychometric literature. The most important ones form the class of regression splines, and these were introduced in multiple regression analysis and principal components analysis in Winsberg and Ramsay (1980, 1983; for a nice overview, see Ramsay, 1988). For splines, the number of parameters is determined by the degree of the spline that is chosen and the number of interior knots. Because splines use fewer parameters, they usually will be smoother and more robust, albeit at the cost of less goodness of fit with respect to the overall loss function that is minimized.

### 3.3.3. Goodness of Fit: Component Loadings, Variance Accounted For, Eigenvalues, and Cronbach's $\alpha$

Principal components analysis studies the interdependence of the variables. Nonlinear transformations maximize the average interdependence, and this optimality property can be expressed in various forms. When variables obtain an ordinal (monotonic spline) transformation or a nominal (nonmonotonic spline) transformation, the technique maximizes the sum of the $P$ largest eigenvalues of the correlation matrix between the transformed variables (where $P$ indicates the number of dimensions that are chosen in the solution). The sum of the eigenvalues, the overall goodness-of-fit index, is equal to the total variance accounted for (in the transformed variables). The variance accounted for in each dimension for each variable separately is equal to the squared component loading, and the component loading itself is the correlation between the transformed variable and a principal component (given by the object scores) in a particular dimension.

There is a very important relationship between the eigenvalue (the total sum of squared component loadings in each dimension) and probably the most frequently used coefficient for measuring internal consistence in applied psychometrics: Cronbach's $\alpha$ (e.g., see Heiser & Meulman, 1994; Lord, 1958; Nishisato, 1980). The relationship between $\alpha$ and the total variance accounted for, as expressed in the eigenvalue $\lambda$, is

$$\alpha = M(\lambda - 1)/(M - 1)\lambda, \qquad (1)$$

where $M$ denotes the number of variables in the analysis. Because $\lambda$ corresponds to the largest eigenvalue of the correlation matrix, and because CATPCA maximizes the largest eigenvalue of the correlation matrix over transformations of the variables, it follows that CATPCA maximizes Cronbach's $\alpha$. This interpretation is straightforward when the CATPCA solution is one-dimensional. Generalized use of this coefficient in more-dimensional CATPCA is described in Section 3.4.2.

## 3.4. CATPCA in Action, Part 1

Throughout this chapter, the principles behind categorical principal components analysis (CATPCA), or principal components analysis with nonlinear optimal scaling transformations, will be illustrated by using a large-scale multivariate data set from the ISSP (1995) that can be considered exemplary for data collected in the social and behavioral sciences. The ISSP is a continuous annual cross-national data collection project that has been running since 1985. It brings together preexisting social science projects and coordinates research goals, thereby adding a cross-national perspective to the individual national studies. Since 1985, the ISSP grew from 6 to 30 participating countries in 1998. The ISSP Internet pages give access to detailed information about the ISSP data service provided by the Zentral Archiv, Cologne. The homepage of the ISSP-Secretariat provides information on ISSP history, membership, publications, and the ISSP listserver.

The original data concern feelings of national identity from about 28,500 respondents in 23 different countries all over the world. Because the number of respondents in the sample in each of the participating countries is not proportional to the population size, a random sample from the original data was taken so that all countries have the same weight in the analysis, with all being represented by 500 respondents. This selection makes the total number of individuals in our examples equal to 11,500.

For the first application, we have selected three groups of variables from the National Identity Study. The first group of five variables indicates how close the respondents feel toward their neighborhood (CL-1), their town (CL-2), their county (CL-3), their country (CL-4), and their continent (CL-5). (The data were recoded so that a score of 1 indicates *not close at all* and a score of 5 indicates *very close.*) The next five variables indicate whether the respondents are willing to move from their neighborhood to improve their work or living conditions, either to another neighborhood (MO-1), another city (MO-2), another county (MO-3), another country (MO-4), or another continent (MO-5), with the score 1 indicating *very unwilling* and the score of 5 indicating *very willing.* The third set of variables concerns statements about immigrants, asking the respondents on a scale from 1 to 5 whether they *strongly disagree* (1) or *strongly agree* (5) with the following statements: "Foreigners should not be allowed to buy land [in this country]" (I-Land), "Immigrants increase crime rates" (I-Crime), "Immigrants are generally good for the economy" (I-Econ), "Immigrants take jobs away from people who were born [in this country]" (I-Jobs), and " Immigrants make [this] country more open to new ideas and cultures" (I-Ideas). Also, respondents were asked to scale themselves with respect to the statement, "The number of immigrants to [my country] nowadays should be *reduced a lot* (1) . . . *increased a lot* (5)." More than 50% of the respondents have one or more missing values on these 16 variables; therefore, a missing data treatment strategy other than deleting all cases with missing data is required, and it was decided to use the straightforward CATPCA option of imputing the modal category for each of the variables. (See Section 3.6.3 on the treatment of missing data for more sophisticated approaches available in the optimal scaling framework.)

### 3.4.1. VAF and Cronbach's $\alpha$

The results of a two-dimensional solution with monotonic spline transformations will be presented that explains 41% of the variance of the scores of the 11,500 respondents on the 16 variables. The percentage of variance accounted for (PVAF) in the first dimension (26.7%) is almost twice the PVAF in the second dimension (14.4%). The VAF in the first dimension equals $.267 \times 16$ (number of variables) $= 4.275$, and in the second dimension, $.144 \times 16 = 2.305$. As explained above, the VAF is closely related to Cronbach's $\alpha$.

As illustrated in Heiser and Meulman (1994), the relationship between $\alpha$ and the VAF (eigenvalue) is not linear but monotonically increasing, and it is severely nonlinear when $M$, the number of variables, grows. For $M = 16$, as in our example, the VAF in the first dimension corresponds to a value of $\alpha = .817$, and the VAF in the second dimension corresponds to a value of $\alpha = .604$. If we take the total variance accounted for (6.580) as the value of $\lambda$ in equation (1), $\alpha = .905$ (the maximum is 1). This use of equation (1) clearly gives a much more general interpretation of $\alpha$ than was originally intended but provides an indication of the global fit of the CATPCA solution. The VAF per dimension is equal to the sum of squares of the component loadings and equal to the associated eigenvalue of the correlation matrix between the optimally transformed variables. Note that the value of $\alpha$ for a particular dimension becomes negative when the associated eigenvalue is less than 1.0. The largest eigenvalue of the correlation matrix between the original variables is 4.084, so the increase in VAF is $1 - 4.084/4.275 = 4.5\%$, which is not a dramatic overall increase. For most of the individual variables, however, the transformation is clearly nonlinear, as shown in Figure 3.1.

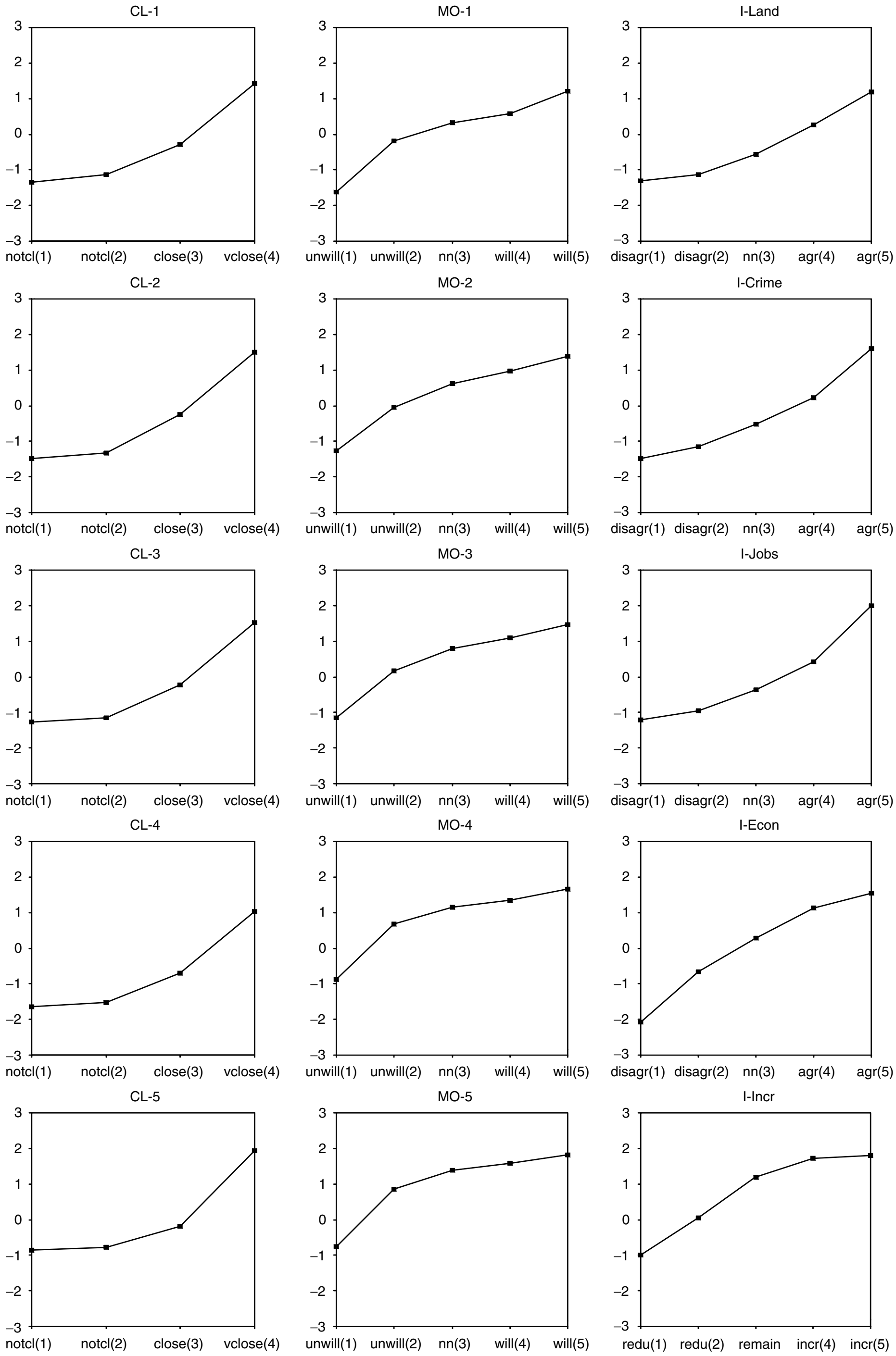
### 3.4.2. Nonlinear Transformations

In Figure 3.1, the transformations for CL-1 unto CL-5, MO-1 unto MO-5, and I-Land unto I-Incr are displayed in its columns; the optimal quantifications are given on the vertical axes versus the original values on the horizontal axes. The nonlinear transformations for CL-1 unto CL-5 show convexity, indicating that there is less distinction between the *not close at all* = ncl(1) and *not close* = ncl(2) categories, which are contrasted to the *very close* = ncl(4) category; the *close* = ncl(3) category is almost always near to the mean of 0. The MO-1 unto MO-5 quantifications show the opposite pattern: The nonlinear transformations approximate a concave function, grouping the *willing, very willing* categories, which are contrasted to the *very unwilling* category. The *unwilling* category has quantifications close to the mean, except for MO-4 and MO-5, which show the most concave functions. When we then inspect the quantifications for I-Land, I-Crime, and I-Jobs (the statements in which a high score expresses a negative attitude toward immigrants), we see that the transformations are convex again, contrasting the flat part for the (*strongly*) *disagree* categories at the lower end from the steep part toward the *strongly agree* category at the upper end. So these transformations resemble those for the CL variables. Looking at the quantifications for I-Econ and I-Incr, which express a positive attitude toward immigrants, we see that their quantifications give concave functions, just as for the MO variables: *strongly disagree* (at the lower end) is contrasted with *agree* and *strongly agree* (at the upper end) for I-Econ, and *reduced a lot* is contrasted with *increase* and *increase a lot* at the upper end for I-Incr ("the number of immigrants should be . . ."). The overall conclusion is that the steep parts of each of the transformations express negative feelings toward immigrants because they occur at the upper end for the negatively stated attitudes and at the lower end for the positively stated attitudes. Simultaneously, this pattern is reflected in the transformations for the CL variables, with the steep part indicating that one feels very close to one's living environment, and the MO variables, with the steep part indicating that one is very unwilling to move.

### 3.4.3. Representing Variables as Vectors

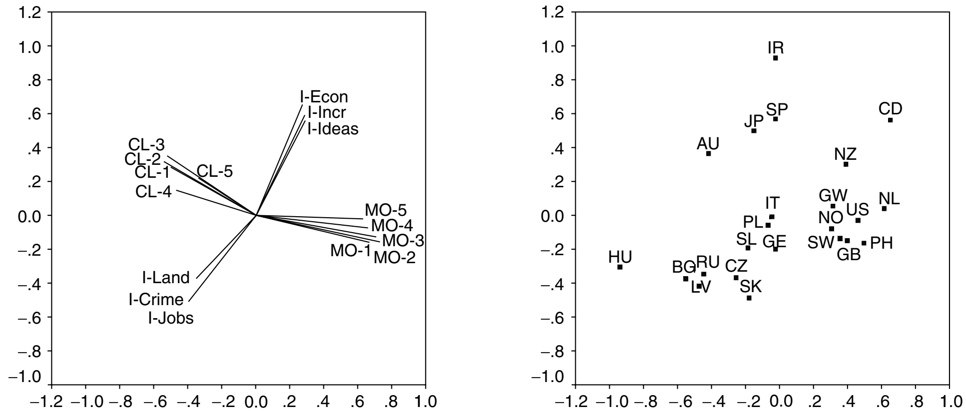
The optimal quantification process turns a qualitative, nominal (or ordinal) variable into a quantitative, numerical variable. The resulting nonlinearly transformed variable can be represented as a vector in the space that is determined for the objects. The coordinates for such a vector are given by the associated component loadings that give the correlation between the transformed variable and the dimensions of the

**Figure 3.1** Spline Transformation of CL Variables (First Column), MO Variables (Second Column), and IM Variables From the CATPCA of the 1995 ISSP National Identity Study




object space. The graph of the component loadings is given in Figure 3.2 (left-hand panel), which shows vectors going in four different directions from the origin (the point 0, 0). Going clockwise, the first group of vectors points in the north-northeast direction, containing I-Econ, I-Incr, and I-Idea; the second

**Figure 3.2** Loadings for MO, CL, and IM Variables (Left-Hand Panel) and Category Points for Country (Right-Hand Panel) From the CATPCA Analysis

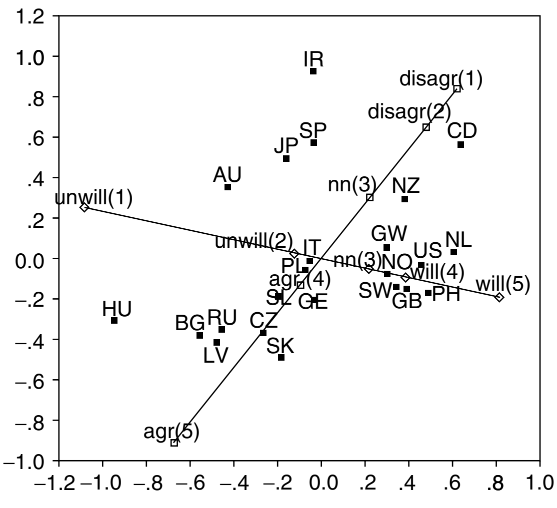



NOTE: Countries are identified as follows: IT = Italy; PL = Poland; SL = Slovenia; GE = East Germany; HU = Hungary; BG = Bulgaria; LV = Latvia; RU = Russia; CZ = Czech Republic; SK = Slovak Republic; AU = Austria; JP = Japan; SP = Spain; IR = Ireland; GW = West Germany; NO = Norway; SW = Sweden; GB = Great Britain; PH = Philippines; US = United States; NL = Netherlands; NZ = New Zealand; CD = Canada.

group points to the east-southeast, comprising the MO variables. The I-Land, I-Crime, and I-Jobs variables point in the south-southwest direction, and the CL variables, finally, point toward the west-northwest. From the transformation plots described above, we know that these directions indicate positive attitudes toward immigrants, willingness to move, very negative attitudes toward immigrants, and feeling very close to one's environment, respectively. It should be noted that each of these four groups of vectors has starting points representing the opposite meaning extending at the opposite side of the origin. So very close to the I-Econ, I-Incr, and I-Idea vectors, we should also envision the starting points of I-Land, I-Crime, and I-Jobs representing positive attitudes, as in the flat parts of the corresponding transformation plots. The reverse, therefore, is also true: The lower, very negative sides of the vectors for I-Econ, I-Incr, and I-Idea are very close to the plotted very negative sides of the vectors for I-Land, I-Crime, and I-Jobs. This whole story can be repeated for the MO and CL vectors that extend either to the right or to the left from the origin (also, see Figure 3.3).

The *very unwilling to move* lower endpoints are close to the *very close* upper endpoints, whereas the *not close* lower endpoints are near the *willing to move* upper endpoints. Now that we have interpreted the extremes of the optimally scaled categories depicted in the transformation plots, we can also interpret the full range of quantifications with respect to their original category labels. Before this will be done in Section 3.4.5, however, we will first inspect a

**Figure 3.3** Joint Category Points for Country, MO-1, and I-Crime




NOTE: Countries are identified as follows: IT = Italy; PL = Poland; SL = Slovenia; GE = East Germany; HU = Hungary; BG = Bulgaria; LV = Latvia; RU = Russia; CZ = Czech Republic; SK = Slovak Republic; AU = Austria; JP = Japan; SP = Spain; IR = Ireland; GW = West Germany; NO = Norway; SW = Sweden; GB = Great Britain; PH = Philippines; US = United States; NL = Netherlands; NZ = New Zealand; CD = Canada.

different type of variables that can be introduced into the analysis described thus far.

### 3.4.4. Supplementary Variables

In the analysis of the CL, MO, and IM variables, we added a supplementary variable labeled *country.* This variable indicates from which of the 23 different countries the respondent originates. A supplementary variable has no influence on the actual analysis, but its quantifications are computed afterwards to establish its relationship with the solution obtained. In the case of the National Identity Study data, the number of respondents is too large to inspect the object scores on an individual level. Having the Country variable as a supplementary variable, however, gives the opportunity to display clusters of respondents from the same country by a single point. When the respondents from a particular country are very heterogeneous, their individual points will be scattered all over the two-dimensional space, and their associated country point, computed as the centroid of the appropriate individual points, will be located close to the origin of the configuration. To obtain these centroids for the 23 different countries in the National Identity Study, we have to specify that the country variable should obtain multiple nominal quantifications. The result is shown in the right-hand panel of Figure 3.2. In this graph, we see various clusters of points in three different directions, starting from the origin, which itself is close to Italy (IT) and Poland (PL) (and Slovenia [SL] and East Germany [GE]). First, a cluster of points contains Hungary (HU), Bulgaria (BG), Latvia (LV), Russia (RU), the Czech Republic (CZ), and the Slovak Republic (SK) in the lower left corner. Going in the upper left direction, we see Austria (AU), Japan (JP), Spain (SP), and Ireland (IR). Finally, going from the origin straight to the right, we have West Germany (GW), Norway (NO), Sweden (SW), Great Britain (GB), the Philippines (PH), the United States (US), and the Netherlands (NL). New Zealand (NZ) and Canada (CD) are almost on a straight line from the origin toward the upper right corner of the graph. Having these coordinates for the 23 countries, we can construct a biplot of the country points and the vectors for the CL, MO, and IM variables.

### 3.4.5. A Biplot of Centroids and Vectors

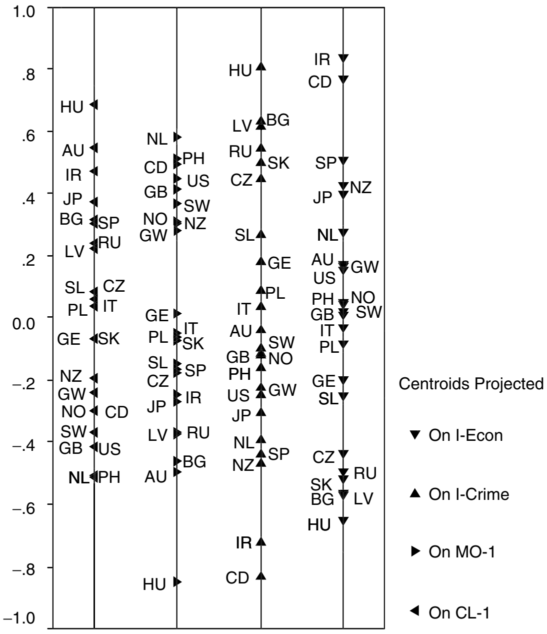
As described above, the CATPCA methodology allows a variety of different biplots. Because the number of objects in the National Identity Study is too large to inspect the relationship between the objects and the variables on the individual level, we represent the individual points by the centroids that are obtained by the supplementary country variable. There are two different ways for joint representation of country points and the vectors for the variables. The most straightforward one is a graph with the centroids from the right-hand panel of Figure 3.2 superimposed on the component loadings depicted in the left-hand panel. Elements of this plot (not shown) can be highlighted by the joint representation of the centroids and category points for selected variables. For illustration in our case, MO-1 and I-Crime were chosen, and the resulting plot is given in Figure 3.3. Here we notice the three most important clusters: Cluster 1 contains HU, BG, RU, LV, SK, and CZ; Cluster 2 contains AU, JP, SP, and IR; and Cluster 3 contains GW, NO, SW, GB, PH, US, and NL, located between the vectors given for MO-1 and I-Crime. In contrast to the component plot in Figure 3.2, a variable is now represented by the full set of category points on a straight line through the origin. For I-Crime, the category points “disagr(1) = *strongly disagree*” and “disagr(2) *disagree*” are both located at the side of the vector that points toward the north, whereas “agr(5) = *strongly agree*” is located at the opposite end, pointing to the south. The category “agr(4) = *agree*” is located close to the origin (compare the quantification close to zero in the transformation plot). The vector for the MO-1 variable contrasts “unwill(1) = *very unwilling*” on the left with “will(4) = *willing*” and “will(5) = *very willing*” on the right; here, the category “unwill(2) = *unwilling*” is close to the origin.

From the location of the country points with respect to the vectors for the variables, we can derive the relative positions by projection; for example, Ireland (IR) and Canada (CD) score high on the *disagree* end of the “Immigrants increase crime” vector. With respect to the cluster structure described above, the first cluster (with Russia [RU] in the center) agrees with the statement that immigrants increase the crime rate, and it is unwilling to move. Cluster 2, containing Japan, is also unwilling to move but (strongly) disagrees with the I-Crime statement. The third cluster, containing the United States, mildly disagrees but is willing to move (from its neighborhood).

### 3.4.6. Projected Centroids

The relative joint position of countries on statements is most clearly represented in the “projected centroids”

**Figure 3.4** Projected Centroids for Country on Selected Variables (From Right to Left)




NOTE: Countries are identified as follows: IT = Italy; PL = Poland; SL = Slovenia; GE = East Germany; HU = Hungary; BG = Bulgaria; LV = Latvia; RU = Russia; CZ = Czech Republic; SK = Slovak Republic; AU = Austria; JP = Japan; SP = Spain; IR = Ireland; GW = West Germany; NO = Norway; SW = Sweden; GB = Great Britain; PH = Philippines; US = United States; NL = Netherlands; NZ = New Zealand; CD = Canada.

plot, shown in Figure 3.4. Here the 23 countries have been projected on the vectors for the statements as in the biplot, but now these projections themselves are shown on straight parallel lines representing the statements. From left to right, the following statements were used: CL-1, MO-1, I-Crime, and I-Econ. As we know from Figure 3.2 (left-hand panel), CL-1 and MO-1 are each other's opposite, and so this is also seen in Figure 3.4, with HU, AU, IR, JP, and BG scoring high on CL-1 (and low on MO-1) and NL, PH, CD, US, and the other countries from Cluster 3 scoring high on MO-1 (and low on CL-1). The two other variables represented show contrasts between Cluster 1 (scoring high on I-Crime and low on I-Econ) and CD, IR, NZ, SP, and NL (scoring low on I-Crime and high on I-Econ).

We should remind ourselves, however, that the data analyzed are from 1995 and that points of view will most probably have changed since then for at least some of the countries in this study.

## 3.5. Technical Background of Nonlinear Principal Components Analysis

### 3.5.1. Indicator Matrices

The nonlinear transformation approach deals with categorical variables in the following way. A categorical variable $\mathbf{h}_m$ defines a binary indicator matrix $\mathbf{G}_m$ with $N$ rows and $C_m$ columns, where $C_m$ denotes the number of categories. Elements $h_{im}$ then define elements $g_{ic(m)}$ as follows:

$$g_{ic(m)} = \begin{cases} 1 \text{ if } h_{im} = c_m \\ 0 \text{ if } h_{im} \neq c_m \end{cases}, \qquad (2)$$

where $c_m = 1, \ldots, C_m$ is the running index indicating a category number in the $m$th variable. If category quantifications are denoted by the vector $\mathbf{y}_m$ (with $C_m$ elements), then a transformed variable $\mathbf{q}_m$ can be written as $\mathbf{G}_m\mathbf{y}_m$. For instance, in a standard linear model, with $M$ predictor variables in $\mathbf{X}$ and $b_m$ denoting the regression weight for the $m$th variable, the linear combination of the predictors that correlates maximally with the response $\mathbf{z}$ can be written as $\hat{\mathbf{z}} = \sum_{m=1}^{M} b_m \mathbf{x}_m$. Incorporating the nonlinear scaling of the predictor variables $\varphi_m(\mathbf{x}_m)$, for $m = 1, \ldots, M$ with $\varphi_m(\mathbf{x}_m)$ an admissible nonlinear function of $\mathbf{x}_m$, the optimal linear combination is now written as $\hat{\mathbf{z}} = \sum_m^M b_m \varphi_m(\mathbf{x}_m) = \sum_m^M b_m \mathbf{G}_m \mathbf{y}_m$. By mapping a categorical variable into an indicator matrix, invariance is ensured under the one-to-one nonlinear transformation of the original variable. The idea to replace a categorical variable by an indicator matrix can already be found in Guttman (1941). The term *indicator matrix* was coined by de Leeuw (1968); other names used are *attribute* or *trait matrix* (Lingoes, 1968), *response-pattern table* (Nishisato, 1980), *incidence matrix,* or *dummy variables* (in experimental design).

### 3.5.2. The Joint Objective Function

In this section, we will describe the objective function that jointly fits the vector model and the centroid model. We suppose that there are $M_V$ variables fitted according to the vector model and $M_B$ variables fitted according to the centroid model; thus, we have $M_V + M_B = M$. We start by defining the following terminology. The $N \times M$ matrix $\mathbf{Q}$ contains the scores for the $N$ objects on $M$ variables. The nature of the individual variables $\mathbf{q}_m$ will be discussed shortly. The $N \times P$ matrix $\mathbf{X}$ contains the coordinates for the $N$ objects in a $P$-dimensional representation space, and the matrix $\mathbf{A}$ (of size $M_V \times P$) gives the coordinates in the same space for the endpoints of the vectors that are fitted to the variables in the bilinear (vector) model. Thus, $\mathbf{a}_m$ contains the coordinates for the representation of the $m$th variable. Consequently, the part of the objective function that minimizes the value of the objective function with respect to the bilinear/vector model can be written as follows:

$$\overline{L}_V(\mathbf{Q};\mathbf{X};\mathbf{A}) = M_V^{-1} \sum_{m \in K_V} \|\mathbf{q}_m - \mathbf{X}\mathbf{a}_m\|^2, \qquad (3)$$

where $K_V$ denotes the index set that contains the indices of the variables that are fitted with the vector model, and $\|\cdot\|^2$ means taking the sum of squares of the elements. Assuming the data in $\mathbf{q}_m$ to have $C_m$ different values, we can also write

$$\overline{L}_V(\mathbf{y}_V;\mathbf{X};\mathbf{A}) = M_V^{-1} \sum_{m \in K_V} \|\mathbf{G}_m\mathbf{y}_m - \mathbf{X}\mathbf{a}_m\|^2, \qquad (4)$$

where $\mathbf{G}_m$ is an indicator matrix that classifies each of the objects in one and only one category. The optimal category quantifications that will be obtained are contained in the $C_m$ vector $\mathbf{y}_m$, where $C_m$ denotes the number of categories for the $m$th variable. The vector $\mathbf{y}_V$ collects the quantifications for the $M_V$ different variables and has length $\sum_{m \in K_V} C_m$.

The projection of the object points $\mathbf{X}$ onto the vector $\mathbf{a}_m$ gives the approximation of the nonlinearly scaled (optimally quantified) variable $\mathbf{q}_m = \mathbf{G}_m\mathbf{y}_m$ in $P$-dimensional Euclidean space. Minimization of the loss function $\overline{L}_V$ for the bilinear/vector model can be shown to be equivalent to the minimization of

$$L_V(\mathbf{y}_V;\mathbf{A};\mathbf{X}) = M_V^{-1} \sum_{m \in K_V} \|\mathbf{G}_m\mathbf{y}_m\mathbf{a}_m' - \mathbf{X}\|^2 \qquad (5)$$

(see Gifi, 1990). Here a $P$-dimensional matrix $\mathbf{X}$ is being approximated by the inner product $\mathbf{G}_m\mathbf{y}_m\mathbf{a}_m'$, which gives the coordinates of the categories of the $m$th variable located on a straight line through the origin in the joint $P$-dimensional space. The major advantage of this reformulation of the objective function is its capacity of capturing the centroid model in the same framework. The latter can simply be written as

$$L_B(\mathbf{Y}_B;\mathbf{X}) = M_B^{-1} \sum_{m \in K_B} \|\mathbf{G}_m\mathbf{Y}_m - \mathbf{X}\|^2, \qquad (6)$$

where $K_B$ denotes the index set of the variables for which a centroid model is chosen. The $C_m \times P$ matrix $\mathbf{Y}_m$ contains the coordinates of the categories in the $P$-dimensional space, and $\mathbf{Y}_B$ collects the quantities for the $M_B$ variables stacked upon each other. The objective function for the centroid model implies that to obtain perfect fit, an object point in $\mathbf{X}$ should coincide with its associated category point in one of the rows of $\mathbf{Y}_m$.

At this point, we can write the joint objective function for CATPCA as a weighted linear combination of the separate losses:

$$L(\mathbf{Y};\mathbf{A};\mathbf{X}) = (M_V + M_B)^{-1}[M_V L_V(\mathbf{y}_V;\mathbf{A};\mathbf{X}) + M_B L_B(\mathbf{Y}_B;\mathbf{X})], \qquad (7)$$

where the first part is minimized for variables indexed by $m$ for which a vector representation is chosen, and the second part is minimized for the representation of categorical variables. The optimal $\hat{\mathbf{X}}$ is found as

$$\hat{\mathbf{X}} = M^{-1}\left[\sum_{m \in K_V} \mathbf{G}_m\mathbf{y}_m\mathbf{a}_m' + \sum_{m \in K_B} \mathbf{G}_m\mathbf{Y}_m\right],$$

after which the object scores are orthonormalized as $\hat{\mathbf{X}}'\hat{\mathbf{X}} = N\mathbf{I}$ (thus, they are uncorrelated).

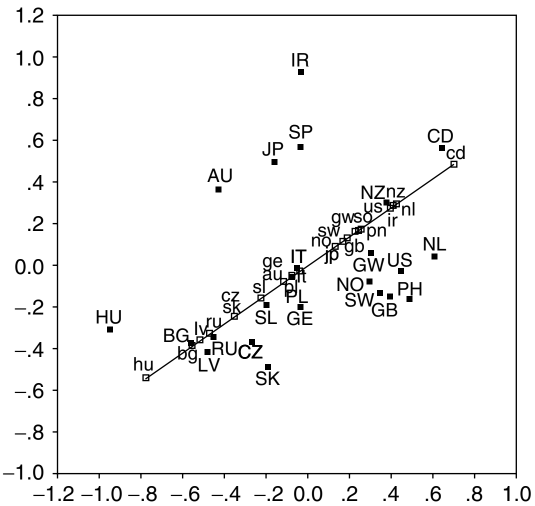
### 3.5.3. Quantifications and Geometry

In this section, we will describe the iterative process that turns multiple quantifications $\mathbf{Y}_k$ into vector coordinates $\mathbf{y}_m\mathbf{a}_m'$, possibly incorporating ordinal and numerical information from the original variables. Recall that in Figure 3.3, a joint representation was given for centroids (for the categories of the country variable) and for vector coordinates (for the categories of the MO-1 and I-Crime variables). The very same representation can also be given for one and the same variable. This idea is illustrated by including a copy of the supplementary Country variable in the analysis as well and giving this supplementary copy not multiple nominal quantifications but a nominal transformation that positions category points on a vector. The result is illustrated in Figure 3.5, in which the uppercase labels are for the centroids from the previous analysis, and the lowercase labels are for the additional vector coordinates. We see that in the cloud of the country points, the dominant direction is from northeast to southwest, from CD to HU and through Clusters 1 and 3. Computationally, the transition from centroids into vector coordinates involves the following steps.

#### *3.5.3.1. From Centroids to Unordered Vector Coordinates*

For each variable, we start with fitting a centroid model according to (6), which gives the minimum over

**Figure 3.5** Centroids (Multiple Nominal Quantification) and Vector Coordinates (Nominal Transformation) for Country in CATPCA




NOTE: Countries are identified as follows: IT = Italy; PL = Poland; SL = Slovenia; GE = East Germany; HU = Hungary; BG = Bulgaria; LV = Latvia; RU = Russia; CZ = Czech Republic; SK = Slovak Republic; AU = Austria; JP = Japan; SP = Spain; IR = Ireland; GW = West Germany; NO = Norway; SW = Sweden; GB = Great Britain; PH = Philippines; US = United States; NL = Netherlands; NZ = New Zealand; CD = Canada.

$\mathbf{Y}_m$ as $\mathbf{Y}_m = \mathbf{D}_m^{-1}\mathbf{G}'_m\mathbf{X}$, where $\mathbf{D}_m = \mathbf{G}'_m\mathbf{G}_m$ contains the marginal frequencies of the categories of the $m$th variable. Next, for the vector model, the centroids $\mathbf{Y}_m$ are projected on a best-fitting line, denoted by $\mathbf{a}_m$, a vector through the origin. The least squares fit that is the minimum of

$$\begin{aligned}&\|\mathbf{G}_m\mathbf{Y}_m - \mathbf{G}_m\mathbf{y}_m\mathbf{a}'_m\|^2 \\ &\quad = \mathrm{tr}(\mathbf{Y}_m - \mathbf{y}_m\mathbf{a}'_m)'\mathbf{D}_m(\mathbf{Y}_m - \mathbf{y}_m\mathbf{a}'_m) \qquad (8)\end{aligned}$$

over both $\mathbf{y}_m$ and $\mathbf{a}_m$ determines the category quantifications $\mathbf{y}_m$ and (the orientation of) the vector $\mathbf{a}_m$. The coordinates $\mathbf{y}_m\mathbf{a}'_m$, the outer product of the category quantifications $\mathbf{y}_m$ and the vector $\mathbf{a}_m$, represent the category points on this line, which represents the $m$th variable in the joint space of objects and variables. The $\mathbf{a}_m$ are also called the component loadings, and they give the correlations between the variables and the dimensions of the principal components space. Setting the partial derivatives in (8) with respect to the component loadings $\mathbf{a}_m$ to zero gives the optimal $\hat{\mathbf{a}}_m$ as

$$\hat{\mathbf{a}}_m = \frac{\mathbf{Y}'_m\mathbf{D}_m\mathbf{y}_m}{(\mathbf{y}'_m\mathbf{D}_m\mathbf{y}_m)}. \qquad (9)$$

Next, setting the partial derivatives in (8) with respect to $\mathbf{y}_m$ to zero shows that the optimal unnormalized $\tilde{\mathbf{y}}_m$ is found as

$$\tilde{\mathbf{y}}_m = \frac{\mathbf{Y}_m\mathbf{a}_m}{\mathbf{a}'_m\mathbf{a}_m}. \qquad (10)$$

To satisfy the normalization conventions $\mathbf{q}'_m\mathbf{q}_m = N$, the standardized variable $\mathbf{q}_m$ should contain quantifications $\hat{\mathbf{y}}_m$ that are rescaled:

$$\hat{\mathbf{y}}_m = N^{1/2}\tilde{\mathbf{y}}_m(\tilde{\mathbf{y}}'_m\mathbf{D}_m\tilde{\mathbf{y}}_m)^{-1/2}. \qquad (11)$$

Note that the length of the vector $\mathbf{a}_m$ has to be diminished to the same extent as the size of the quantifications $\hat{\mathbf{y}}_m$ is increased to keep $\mathbf{y}_m\mathbf{a}'_m$ the same. Equation (10) symbolizes the projection of the centroids $\mathbf{Y}_m$ on the vector $\mathbf{a}_m$ and defines the category coordinates for a nominal transformation. It is very unlikely that the category quantifications in $\mathbf{y}_m$ will be proportional to, or even only in the same order as the original integer scale values $1, \ldots, C_m$. In many cases, however, we would like to maintain the original numeric and/or rank-order information in the transformation, which can be dealt with as follows.

#### *3.5.3.2. From Nominal to Ordinal and Numerical Transformations*

If the ordinal, rank-order information should be maintained, an ordinal, monotonic transformation is chosen for variable $m$, and the quantifications $\mathbf{y}_m$ have to be constrained to be monotonic with the order of the original categories. As described above, this requirement can be satisfied by the use of one of two major classes of monotonic transformations. The first, also historically, is the class of least squares monotonic transformations, obtained by a monotonic regression of the values in $\hat{\mathbf{y}}_m$ upon the original scale values $1, \ldots, C_m$, taking the marginals on the diagonal of $D_m$ into account. The second class is defined by monotonic regression splines. As indicated in Section 3.3.2, transformations by regression splines use fewer parameters than transformations obtained by monotonic regression. For monotonic regression, the number of parameters to be fitted is $C_m - 2$; for regression splines, the number of parameters is determined by the degree of the spline that is chosen and the number of interior knots. If the number of categories is small, monotonic regression and regression splines will basically give the same result. When the number of categories is large, it is usually advised to use regression splines because monotonic regression may result in overfitting: The variance accounted for will increase, but so will the instability. (Note: There is a trade-off between the number of categories and the number of objects in those categories. If the number of objects is large, and all categories are sufficiently filled, monotonic regression will usually not result in overfitting.)

When it is decided to give the $m$th variable a numerical transformation, the implication is that the distances between the category points $\mathbf{y}_m \boldsymbol{a}'_m$ have to be equal, and the category quantifications $\mathbf{y}_m$ will be proportional to the original category numbers. This can be done by linear regression of the $\hat{\mathbf{y}}_m$ on the original scale values and will result in a standardized version of the set of the integer scale values $1, \ldots, C_m$, $\mathbf{G}_m \mathbf{y}_m = \alpha_m \mathbf{h}_m + \beta_m$, where the multiplicative constant and the intercept are fitted taking into account the marginal frequencies. If the distances between the categories have to be stretched very much to obtain unit variance, the VAF (expressed in the squared length of the vector $a_m$) will be very small. It is important to realize that this also applies to ordinary PCA with continuous variables (which can be considered as a CATPCA with $N$ categories, where $N$ is the number of objects, as usual).

## 3.6. Some Additional Options of the Program CATPCA

### 3.6.1. External Fitting of Variables

The CATPCA program not only provides an option for the analysis of supplementary variables, as we saw in Section 3.4.4, but for supplementary objects as well. As was true for supplementary variables, supplementary objects are not active in the analysis but enter into the representation afterwards. Another interesting application for the supplementary variables option is the following. CATPCA offers the possibility of reading a fixed configuration of object points, and thus the CATPCA method may be used for so-called property fitting or external unfolding (Carroll & Chang, 1967; Meulman, Heiser, & Carroll, 1986). In this way, external information on objects (contained in so-called external variables) is fitted into the fixed representational space by the use of the vector model (or the centroid model). The option accommodates the same variety of transformation levels as a standard CATPCA analysis (with nominal, ordinal, and numerical treatment of the variables, including the use of splines).

### 3.6.2. Making Continuous Variables Discrete—Binning

Although the CATPCA algorithm is tuned to the analysis of categorical variables, continuous variables can be introduced into the analysis as well, and this is after they have been made discrete using one of a variety of options provided. This process is comparable to fitting a histogram to a continuous distribution. The grouping options described below can also be used to merge a large initial number of categories into less, which is especially warranted when the distribution of the objects over the original categories is very skew or when some of the categories have very few observations.

#### *3.6.2.1. Grouping in a Specified Number of Categories for a Uniform or Normal Distribution*

In Max (1960), optimal discretization points were computed to transform a continuous variable into a categorical one, in which the number of categories can vary from 2 to 36. These discretization points are optimal with respect to an assumed distribution, particularly a univariate standard normal distribution or a univariate uniform distribution. As an illustration, we use the age variable from the National Identity Study: Respondents varied in age from 14 to 98; the modal age category is 30. When this variable is made discrete with seven categories, assuming the population distribution is normal, the following ranges (with corresponding marginal frequencies in parentheses) are obtained: 14–17 (107), 18–30 (2,596), 31–40 (2,335), 41–49 (2,002), 50–59 (1,794), 60–72 (1,916), and 73–98 (699). If, on the other hand, a uniform distribution would be assumed, the following categories and marginal frequencies result: 14–25 (1,653), 26–33 (1,691), 34–39 (1,444), 40–46 (1,657), 47–55 (1,731), 56–65 (1,639), and 66–98 (1,634).

#### *3.6.2.2. Grouping in Equal Intervals With Specified Size*

When it is preferred to have a continuous variable replaced by a categorical variable in which the original values are grouped into intervals of equal size, this is a feasible option as well. Of course, the choice of a specific range for the interval determines the number of categories (bins in a histogram). For the age variable, choosing intervals of 10 years gives the following: 14–23 (1,216), 24–33 (2,128), 34–43 (2,394), 44–53 (2,066), 54–63 (1,669), 64–73 (1,397), 74–83 (493), 84–93 (79), and 94–98 (7). With this option, the groupings for the higher age ranges have rather low marginal frequencies. Comparing this distribution with the two previous ones, we would prefer the uniform option.

#### *3.6.2.3. Ranking*

This particular form of preprocessing is appropriate for at least two different situations. In the first place, it should be noted again that the optimal scaling framework guarantees that any ordinal transformation of the original data, among which is replacing numeric values by ranks, will leave the analysis results the same when variables are treated ordinally. When there are no ties in the original variable, the number of categories in the new variable will be $N$, the number of objects. However, such an ordinal analysis might involve too many parameters to be fitted. When the number of categories approaches the number of objects, it is often a better choice to fit a monotonic spline of a low degree with a limited number of knots. Another use of ranking is to give the resulting rank-order variables a numerical transformation level. In the latter case, the principal components analysis amounts to the analysis of the Spearman rank correlations. If the ranking operation is applied to a variable that contains a unique identification for the objects in the analysis, then the resulting variable, defined as *supplementary,* can be used to identify individual objects in various plots (e.g., in the projected centroids). Of course, this labeling is only feasible and useful when the number of objects is not too large.

#### *3.6.2.4. Multiplying*

The distributional properties of a continuous variable that contains noninteger values can be maintained as closely as possible by the particular linear transformation that transforms the real-valued variable into a discrete variable containing integers. The result of this process is a variable that could be treated as numerical; when all the variables in the analysis are treated this way, we are back to classical principal components analysis. However, when one assumes monotonic (instead of linear) relationships between such a variable and other variables in the analysis, it is advised to fit a monotonic spline transformation. When relationships are completely nonlinear, nonmonotonic splines should be fitted to allow these relationships to be revealed in the analysis.

### 3.6.3. Missing Data

To handle incomplete data in the analysis, a sophisticated option is available that only takes into account the nonmissing data when the loss function is minimized. The indicator matrix for a variable with incomplete data will, in this case, contain rows with only zeros for an object having a missing observation. The loss function in Section 3.5.2 is extended by the use of (internally generated) object weights, collected in a diagonal matrix in which the diagonal elements indicate the number of nonmissing observations for each of the objects. Although this option is very attractive (missing data are completely ignored), it also has a number of drawbacks that need not be severe, however (see Meulman, 1982). Because objects have a different number of observations, the *weighted* mean of the object scores is now equal to 0, and because the mean itself is not 0, various optimality properties of nonlinear PCA are no longer valid. The maximum/minimum value of the component loadings is no longer equal to 1.0 and –1.0, and therefore a component loading can no longer be interpreted as a correlation. (We can still project a transformed variable in the space of the objects, however.) Also, the property that nonlinear PCA optimizes the sum of the $P$ largest eigenvalues of the correlation matrix between the transformed variables is no longer true. (However, when this correlation matrix is computed, there are various choices available for imputing values for the missing data.) Indications on how many data elements can be missing without too much disturbance are given by Nishisato and Ahn (1994).

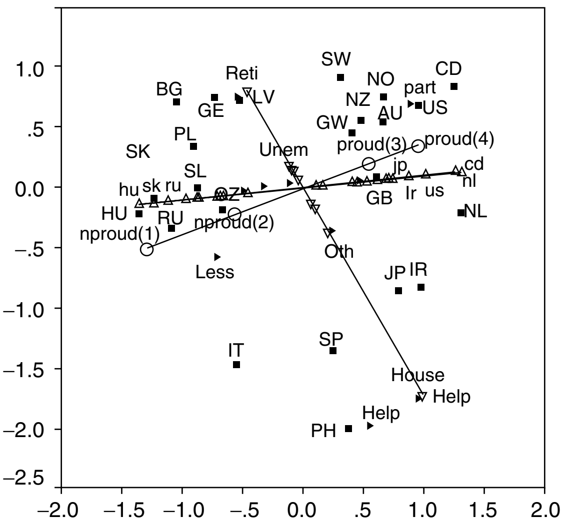
Alternatively, there are other straightforward strategies for treating the missing data in the primary analysis. The first is to exclude objects with missing values; the second provides a very straightforward imputation method, using the value of the modal category. Also, a separate, additional category can be fitted for all objects having a missing value on a particular variable. For all transformation levels, this extra category is positioned optimally with respect to the nonmissing categories. If other, more advanced, missing data strategies are called for (such as the imputation strategy of Van Buuren & Van Rijckevorsel, 1992), these would have to be part of a preprocessing process performed before the actual CATPCA analysis.

## 3.7. CATPCA in Action, Part 2

Having the CATPCA methodology available gives various interesting possibilities compared to a standard correspondence analysis in which two variables are fitted according to the centroid model (Gifi, 1990). First, consider the nominal variables, country and employment status (Emp-Stat), from the National Identity Study. A standard correspondence analysis would display the category points for both variables in a joint low-dimensional space. An extended correspondence analysis may include the same two multiple nominal variables but with a third ordinal variable included as well. This idea will be illustrated by using the Country and Emp-Stat variables, which are now joined with the Democ variable (also from the National Identity Study). The Democ variable indicates, on a scale from 1 to 4, whether the respondent is *very proud* (4), *somewhat proud* (3), *not very proud* (2), or *not proud at all* (1) with respect to his or her country's democracy. The distribution of the original variable shows that the modal respondent is "somewhat proud" ($n$ =4,140); the smallest category is "very proud" ($n$ = 1,361), followed by the category "not proud at all" ($n$ = 1,606), with "not very proud" the second largest category ($n$ = 3,496). Where does this variable fit into the country × Emp-Stat space? The answer is given in Figure 3.6, the joint plot of the categories for Country, Emp-Stat, and Democ. Moreover, we added in this plot the vector representations for Country and Emp-Stat as well, obtained by including copies of these as supplementary variables to be fitted with the vector model.

The centroid representation for Country and Emp-Stat shows their relationship in terms of their category points. The vector representation for Emp-Stat shows that the two extreme categories on a one-dimensional scale would be "Retired" and "Unemployed" at the north-northwest endpoint and "Housewives" at the end pointing south-southeast. From the vector representation, it is easy to see that the category "House = house wives" scores relatively high in the Philippines (PH), Spain (SP), Ireland (IR), Japan (JP), Italy (IT), and the Netherlands (NL). The categories "Retired" and "Unemployed" score high in East Germany (GE), Bulgaria (BG), and Sweden (SW). The one-dimensional projection of the country category points shows that the major direction goes from west to east. The relationship between Country and Emp-Stat in an ordinary correspondence analysis changes when Democ is taken into account. The ordinal transformation of Democ (not shown) turned out to be close to linear but gives emphasis to the modal category "somewhat proud," which is quantified with a higher value

**Figure 3.6** Use of CATPCA: Extended Correspondence Analysis of Country and Em-Stat With Democ (nproud(1) = *not proud at all* to proud(4) = *very proud*) as an Extra Ordinal Variable




NOTE: Countries are identified as follows: IT = Italy; PL = Poland; SL = Slovenia; GE = East Germany; HU = Hungary; BG = Bulgaria; LV = Latvia; RU = Russia; CZ = Czech Republic; SK = Slovak Republic; AU = Austria; JP = Japan; SP = Spain; IR = Ireland; GW = West Germany; NO = Norway; SW = Sweden; GB = Great Britain; PH = Philippines; US = United States; NL = Netherlands; NZ = New Zealand; CD = Canada.

than its numeric counterpart if the original variable containing the scores 1 to 4 had been standardized. The vector for Democ is orthogonal to the direction that connects the categories "Retired" and "Housewives" and is mostly related to the vector representation of Country, contrasting the "very proud of democracy" countries of Canada, the United States, and the Netherlands with the "not proud at all" countries of Italy, Russia, the Slovak Republic, and Hungary.

## 3.8. Discussion

### 3.8.1. Optimal Scaling and (Multiple) Correspondence Analysis

Although we stated earlier that it is beyond the scope of the present chapter to discuss the technique called multiple correspondence analysis (MCA), we need to mention explicitly the relationship between principal components with nonlinear optimal scaling transformations and MCA. When the transformation level is chosen to render multiple nominal quantifications for all variables, the two techniques are completely equivalent. So the current CATPCA program, with all its special options for discretization, missing data, supplementary objects and variables, and a variety of plots, can be used to perform MCA as well. In terms of the loss function in Section 3.5.2, we have $M_V = 0$ and $M_B = M$, and we minimize (6) for all variables.

The classic case of simple correspondence analysis concerns two categorical variables, displayed in a cross table, with the categories of the first variable in the rows and the categories of the second variable in the columns. The cells of the table then contain the frequencies of the joint occurrence of category $C_A$ from variable $A$ and category $C_B$ from variable $B$, and correspondence analysis displays the residuals from independence between the two variables (their interdependence). There are some details that should be taken into account with respect to normalizing the dimensions of the space, but a standard correspondence analysis and a CATPCA are basically equivalent when the two variables are given multiple nominal quantifications. The similarity is largest when the object scores in CATPCA are standardized so that the categories are the average of the object scores, and geometrically the category points will be in the centroid of the object points.

When we have *two* variables, a number of optimal scaling techniques are in fact equivalent. CATPCA with two *nominal* variables, combined with optimization in one dimension, is equivalent to simple regression with optimal scaling, and maximizes the Pearson correlation coefficient over all possible nominal quantifications (Hirschfeld, 1935). When the two variables have a nonlinear relationship, the regression is linearized because categories are allowed to be reordered (permuted), and distances between them are optimally scaled. The term *optimal scaling,* in this context, is due to Bock (1960); also, see Fisher (1940, 1948) for the maximal mutual discrimination principle, as well as the overview in de Leeuw (1990). Applying ordinal (spline) transformations maximizes the correlation coefficient under monotonicity restrictions. When one of the variables is treated as numeric, and the other is given a nominal transformation, the CATPCA technique would be equivalent to linear discriminant analysis but with one single predictor. Obviously, allowing an ordinal transformation instead of the numerical transformation level generalizes the latter technique, maximizing the between to total variation ratio under monotonic transformation of the predictor variable.

### 3.8.2. Special Applications

In the following subsections, we will briefly discuss some special types of applications of CATPCA. For a selection of concrete applications, sometimes using the precursor program PRINCALS, the user is referred to the following: Arsenault, Tremblay, Boulerice, and Saucier (2002); Beishuizen, Van Putten, and Van Mulken (1997); de Haas, Algera, Van Tuijl, and Meulman (2000); de Schipper, Tavecchio, Van IJzendoorn, and Linting (2003); Eurelings-Bontekoe, Duijsens, and Verschuur (1996); Hopman-Rock, Tak, and Staats (2001); Huyse et al. (2000); Theunissen et al. (2003); Vlek and Stallen (1981); Zeijl, te Poel, du Bois-Reymond, Ravesloot, and Meulman (2000); and Van der Ham, Meulman, Van Strien, and Van Engeland (1997).

#### *3.8.2.1. Preferential Choice Data*

In preferential choice data, respondents give a ranking of objects (sometimes called *stimuli*) according to some attribute, giving an explicit comparison. Consumers, for example, can be asked to rank a set of product brands, or psychologists may be asked to rank a number of psychology journals (see Gifi, 1990, pp. 183–187). Such rankings are usually collected in a data matrix, with the stimuli, options, or objects in the rows and the persons (judges) in the columns acting as the variables of the analysis. This situation was actually the very same one in which Tucker's (1960) vector model was applied in Carroll (1972) to preference data. In the latter mentioned application, the analysis was metric because no optimal scaling of the rankings was possible. Because rankings are ordinal by definition, optimal scaling by monotonic (spline) transformations appears most appropriate.

#### *3.8.2.2. Q-Sort and Free-Sort Data*

Another situation for which the persons act as variables is in the so-called analysis of Q-sort data. Here, a number of judges have to group $N$ given objects in a predetermined number of piles (categories), in which the categories have a particular order and the frequencies have to follow a normal distribution as closely as possible. Again, this is a very natural situation for a CATPCA analysis with ordinal transformations. When the $M$ judges are merely given a set of objects and have the liberty to group them in as many categories as they like, without any given order of the categories, we use the term *free-sort* data. Nominal quantification options are called for in this case, either in the form of nominal (nonmonotonic spline) transformations, when the judges seem to group on one (unknown) dimension, or in the form of multiple nominal quantifications, when judges use more than one latent dimension and when different orderings of the categories are allowed for each dimension. (Nominal or nonmonotonic spline transformations will give the same reordering in each dimension.) Examples of multiple nominal quantifications in free-sort data can be found, among others, in Van der Kloot and Van Herk (1991) and Meulman (1996). In the latter paper, groupings were analyzed in the form of a free-sort of statements about the so-called rape myth.

#### *3.8.2.3. The Analysis of Ratings Scales and Test Items*

The application of CATPCA in one dimension is extremely useful because it explores the homogeneity between a set of variables that are assumed to measure the same property (latent characteristic). Optimal scaling minimizes the heterogeneity and maximizes the largest eigenvalue of the correlation matrix. For an extensive treatment of this particular application with its relationship to differential weighting of variables and classical psychometrics, see Heiser and Meulman (1994).

### 3.8.3. CATPCA and the Correlation Matrix Between the Transformed Variables

In ordinary PCA, the results in a two-dimensional solution are identical to those in the first two dimensions of a three-dimensional solution. This property is called *nestedness.* When quantifications have been chosen to be optimal in one dimension, the largest eigenvalue of the correlation matrix is maximized. When they are optimal for $P$ dimensions, the sum of the first $P$ eigenvalues is optimized. The latter does imply that the first eigenvalue, by itself, does not need to be as large as possible, and because this is true by definition for the one-dimensional solution, it implies that CATPCA solutions with different dimensionalities are not necessarily nested. Inspection of the eigenvalues of the transformed correlation matrix shows the distribution of the total sum of the eigenvalues (which is equal to $M$, the number of variables) over the optimized and nonoptimized dimensions. When the CATPCA includes variables with multiple nominal quantifications and a more-dimensional solution is obtained, the situation is somewhat more complicated. The first CATPCA dimension optimizes the largest eigenvalue between the transformed variables, including the first set of the multiple nominal quantifications, whereas the second dimension optimizes the largest eigenvalue of the same correlation matrix, but now including the second set of the multiple nominal quantifications. Therefore, if the primary objective is to maximize the homogeneity, either in one dimension for all variables together or in two dimensions, when the variables seem to form two groups (as in our example in Section 3.4.3), unordered variables should be given a nominal (or nonmonotonic spline) transformation.

### 3.8.4. Prospects

Because unordered or ordered categorical variables are so common in the behavioral sciences, the prospects for nonlinear principal components analysis seem to be good, especially in contexts where a relatively large number of responses have been collected and their mutual relationships have to be sorted out, as in survey research. Another clear application area for CATPCA is instrument development, where it can supplement the usual factor analysis and Cronbach's $\alpha$ calculations for item selection. Because CATPCA directly analyzes the data matrix and not the derived correlation matrix, there need not be the usual concern to have at least 15 times as many observations as the number of variables. In fact, CATPCA is eminently suited for analyses in which there are (many) more variables than objects.

Finally, we would like to mention that there is similar optimal scaling software in the SPSS Categories module for related multivariate analysis techniques. Among these are CATREG for (multiple) regression analysis with optimal scaling, CORRESPONDENCE for correspondence analysis, and OVERALS for nonlinear canonical correlation analysis (Meulman et al., 1999). Like CATPCA, these methods allow one to pursue classic objectives of multivariate analysis when the data do not satisfy the classic quantitative measurement requirements but are qualitative.

## REFERENCES

Arsenault, L., Tremblay, R. E., Boulerice, B., & Saucier, J. F. (2002). Obstetrical complications and violent delinquency: Testing two developmental pathways. *Child Development, 73,* 496–508.

Beishuizen, M., Van Putten, C. M., & Van Mulken, F. (1997). Mental arithmetic and strategy use with indirect number problems up to hundred. *Learning and Instruction, 7,* 87–106.

Benzécri, J.-P. (1973). *L'analyse des Données, Tome II, L'analyse des Correspondances* (Data analysis: Vol. 2. Correspondence analysis). Paris: Dunod.

Benzécri, J.-P. (1992). *Correspondence analysis handbook.* New York: Marcel Dekker.

Bock, R. D. (1960). *Methods and applications of optimal scaling* (Report 25). Chapel Hill: L. L. Thurstone Lab, University of North Carolina.

Breiman, L., & Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association, 80,* 580–598.

Buja, A. (1990). Remarks on functional canonical variates, alternating least squares methods and ACE. *Annals of Statistics, 18,* 1032–1069.

Burt, C. (1950). The factorial analysis of qualitative data. *British Journal of Psychology, 3,* 166–185.

Carroll, J. D. (1968). Generalization of canonical correlation analysis to three or more sets of variables. *Proceedings of the 76th Annual Convention of the American Psychological Association, 3,* 227–228.

Carroll, J. D. (1972). Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences* (Vol. 1, pp. 105–155). New York: Seminar Press.

Carroll, J. D., & Chang, J. J. (1967, April). *Relating preferences data to multidimensional scaling solutions via a generalization of Coomb's unfolding model.* Paper presented at the annual meeting of the Psychometric Society, Madison, WI.

de Haas, M., Algera, J. A., Van Tuijl, H. F. J. M., & Meulman, J. J. (2000). Macro and micro goal setting: In search of coherence. *Applied Psychology, 49,* 579–595.

de Leeuw, J. (1968). *Canonical discriminant analysis of relational data* (Research Report RN-007–68). Leiden, The Netherlands: University of Leiden.

de Leeuw, J. (1973). *Canonical analysis of categorical data.* Unpublished doctoral dissertation, University of Leiden, Leiden, The Netherlands. (Reissued in 1986 by DSWO Press, Leiden, The Netherlands.)

de Leeuw, J. (1990). Multivariate analysis with optimal scaling. In S. Das Gupta & J. Sethuraman (Eds.), *Progress in multivariate analysis.* Calcutta: Indian Statistical Institute.

de Schipper, J. C., Tavecchio, L. W. C., Van IJzendoorn, M. H., & Linting, M. (2003). The relation of flexible child care to quality of center day care and children's socio-emotional functioning: A survey and observational study. *Infant Behavior & Development, 26,* 300–325.

Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika, 1,* 211–218.

Eurelings-Bontekoe, E. H. M., Duijsens, I. J., & Verschuur, M. J. (1996). Prevalence of DSM-III-R and ICD-10 personality disorders among military conscripts suffering from homesickness. *Personality and Individual Differences, 21,* 431–440.

Fisher, R. A. (1940). The precision of discriminant functions. *Annals of Eugenics, 10,* 422–429.

Fisher, R. A. (1948). *Statistical methods for research workers* (10th ed.). Edinburgh, UK: Oliver & Boyd.

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal components analysis. *Biometrika, 58,* 453–467.

Gifi, A. (1990). *Nonlinear multivariate analysis.* Chichester, UK: John Wiley. (Original work published 1981)

Gilula, Z., & Haberman, S. J. (1988). The analysis of multivariate contingency tables by restricted canonical and restricted association models. *Journal of the American Statistical Association, 83,* 760–771.

Gower, J. C. (1966). Some distance properties of latent roots and vector methods used in multivariate analysis. *Biometrika, 53,* 325–338.

Gower, J. C., & Hand, D. J. (1996). *Biplots.* London: Chapman & Hall.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis.* London: Academic Press.

Guttman, L. (1941). The quantification of a class of attributes: A theory and method of scale construction. In P. Horst et al. (Eds.), *The prediction of personal adjustment* (pp. 319–348). New York: Social Science Research Council.

Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika, 33,* 469–506.

Hastie, T., Tibshirani, R., & Buja, A. (1994). Flexible discriminant analysis by optimal scoring. *Journal of the American Statistical Association, 89,* 1255–1270.

Hayashi, C. (1952). On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics, 2,* 93–96.

Heiser, W. J. (1981). *Unfolding analysis of proximity data.* Unpublished doctoral dissertation, University of Leiden, Leiden, The Netherlands.

Heiser, W. J., & de Leeuw, J. (1981). Multidimensional mapping of preference data. *Mathématiques et Sciences Humaines, 19,* 39–96.

Heiser, W. J., & Meulman, J. J. (1983). Analyzing rectangular tables by joint and constrained multidimensional scaling. *Journal of Econometrics, 22,* 139–167.

Heiser, W. J., & Meulman, J. J. (1994). Homogeneity analysis: Exploring the distribution of variables and their nonlinear relationships. In M. Greenacre & J. Blasius (Eds.), *Correspondence analysis in the social sciences: Recent developments and applications* (pp. 179–209). New York: Academic Press.

Hirschfeld, H. O. (1935). A connection between correlation and contingency. *Proceedings of the Cambridge Philosophical Society, 31,* 520–524.

Hopman-Rock, M., Tak, E. C. P. M., & Staats, P. G. M. (2001). Development and validation of the Observation List for early signs of Dementia (OLD). *International Journal of Geriatric Psychiatry, 16,* 406–414.

Horst, P. (1961a). Generalized canonical correlations and their applications to experimental data. *Journal of Clinical Psychology, 17,* 331–347.

Horst, P. (1961b). Relations among *m* sets of variables. *Psychometrika, 26,* 129–149.

Huyse, F. J., Herzog, T., Lobo, A., Malt, U. F., Opmeer, B. C., Stein, B., et al. (2000). European consultation-liaison psychiatric services: The ECLN Collaborative Study. *Acta Psychiatrica Scandinavica, 101,* 360–366.

International Social Survey Programme (ISSP). (1995). *National identity study.* Cologne, Germany: Zentralarchiv für Empirische Sozialforschung.

Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika, 58,* 433–460.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29,* 1–28.

Kruskal, J. B. (1965). Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society, Series B, 27,* 251–263.

Kruskal, J. B. (1978). Factor analysis and principal components analysis: Bilinear methods. In W. H. Kruskal & J. M. Tanur (Eds.), *International encyclopedia of statistics* (pp. 307–330). New York: Free Press.

Kruskal, J. B., & Shepard, R. N. (1974). A nonmetric variety of linear factor analysis. *Psychometrika, 39,* 123–157.

Krzanowski, W. J., & Marriott, F. H. C. (1994). *Multivariate analysis: Part I. Distributions, ordination and inference.* London: Edward Arnold.

Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate descriptive statistical analysis.* New York: John Wiley.

Lebart, L., & Tabard, N. (1973). *Recherche sur la Description Automatique des Données Socio-Economiques* (Research on the automatic description of socioeconomic data). Paris: CORDES-CREDOC.

Lingoes, J. P. (1968). The multivariate analysis of qualitative data. *Multivariate Behavioral Research, 3,* 61–94.

Lord, F. M. (1958). Some relation between Guttman's principal components of scale analysis and other psychometric theory. *Psychometrika, 23,* 291–296.

Masson, M. (1974). Analyse non-linéaire des données (Nonlinear data analysis). *Comptes Rendues de l'Academie des Sciences (Paris), 287,* 803–806.

Max, J. (1960). Quantizing for minimum distortion. *Proceedings IEEE (Information Theory), 6,* 7–12.

Meulman, J. J. (1982). *Homogeneity analysis of incomplete data.* Leiden, The Netherlands: DSWO Press.

Meulman, J. J. (1986). *A distance approach to nonlinear multivariate analysis.* Leiden, The Netherlands: DSWO Press.

Meulman, J. J. (1992). The integration of multidimensional scaling and multivariate analysis with optimal transformations of the variables. *Psychometrika, 57,* 539–565.

Meulman, J. J. (1996). Fitting a distance model to homogeneous subsets of variables: Points of view analysis of categorical data. *Journal of Classification, 13,* 249–266.

Meulman, J. J., Heiser, W. J., & Carroll, J. D. (1986). *PREFMAP-3 users' guide.* Murray Hill, NJ: AT&T Bell Laboratories.

Meulman, J. J., Heiser, W. J., & SPSS. (1999). *SPSS Categories 10.0.* Chicago: SPSS.

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications.* Toronto: University of Toronto Press.

Nishisato, S. (1984). Forced classification: A simple application of a quantification method. *Psychometrika, 49,* 25–36.

Nishisato, S. (1994). *Elements of dual scaling: An introduction to practical data analysis.* Hillsdale, NJ: Lawrence Erlbaum.

Nishisato, S., & Ahn, H. (1994). When not to analyse data: Decision making on missing responses in dual scaling. *Annals of Operations Research, 55,* 361–378.

Ramsay, J. O. (1982). Some statistical approaches to multidimensional scaling data. *Journal of the Royal Statistical Society, Series A, 145,* 285–312.

Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical Science, 3*(4), 425–461.

Roskam, E. E. C. I. (1968). *Metric analysis of ordinal data in psychology.* Voorschoten: VAM.

Saporta, G. (1975). *Liaisons entre Plusieurs Ensembles de Variables et Codage de Données Qualitatives* (Connections between several sets of variables and coding of qualitative data). Unpublished doctoral dissertation, Université Paris VI, Paris.

Schmidt, E. (1907). Zur Theorie der linearen und nichtlinearen Integralgleichungen (On the theory of the linear and nonlinear integral equations). *Mathematische Annalen, 63,* 433–476.

Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function: I. *Psychometrika, 27,* 125–140.

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function: II. *Psychometrika, 27,* 219–246.

Shepard, R. N. (1966). Metric structures in ordinal data. *Journal of Mathematical Psychology, 3,* 287–315.

Stewart, G. W. (1993). On the early history of the singular value decomposition. *SIAM Review, 35,* 551–566.

Tenenhaus, M., & Young, F. W. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis, and other methods for quantifying categorical multivariate data. *Psychometrika, 50,* 91–119.

Theunissen, N. C. M., Meulman, J. J., Den Ouden, A. L., Koopman, H. M., Verrips, G. H., Verloove-Vanhorick, S. P., et al. (2003). Changes can be studied when the measurement instrument is different at different time points. *Health Services and Outcomes Research Methodology, 4* (2).

Torgerson, W. S. (1958). *Theory and methods of scaling.* New York: John Wiley.

Tucker, L. R. (1960). Intra-individual and inter-individual multidimensionality. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: Theory and applications* (pp. 155–167). New York: John Wiley.

Van Buuren, S., & Van Rijckevorsel, L. A. (1992). Imputation of missing categorical data by maximizing internal consistency. *Psychometrika, 57,* 567–580.

Van der Burg, E., & de Leeuw, J. (1983). Non-linear canonical correlation. *British Journal of Mathematical and Statistical Psychology, 36,* 54–80.

Van der Burg, E., de Leeuw, J., & Verdegaal, R. (1988). Homogeneity analysis with *k* sets of variables: An alternating least squares method with optimal scaling features. *Psychometrika, 53,* 177–197.

Van der Ham, T., Meulman, J. J., Van Strien, D. C., & Van Engeland, H. (1997). Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry, 170,* 363–368.

Van der Kloot, W. A., & Van Herk, H. (1991). Multidimensional scaling of sorting data: A comparison of three procedures. *Multivariate Behavioral Research, 26,* 563–581.

Vlek, C., & Stallen, P. J. (1981). Judging risks and benefits in the small and in the large. *Organizational Behavior and Human Performance, 28,* 235–271.

Winsberg, S., & Ramsay, J. O. (1980). Monotonic transformations to additivity using splines. *Biometrika, 67,* 669–674.

Winsberg, S., & Ramsay, J. O. (1983). Monotone spline transformations for dimension reduction. *Psychometrika, 48,* 575–595.

Young, F. W. (1981). Quantitative analysis of qualitative data. *Psychometrika, 46,* 357–387.

Young, F. W., de Leeuw, J., & Takane, Y. (1976). Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. *Psychometrika, 41,* 505–528.

Young, F. W., Takane, Y., & de Leeuw, J. (1978). The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features. *Psychometrika, 43,* 279–281.

Zeijl, E., te Poel, Y., du Bois-Reymond, M., Ravesloot, J., & Meulman, J. J. (2000). The role of parents and peers in the leisure activities of young adolescents. *Journal of Leisure Research, 32,* 281–302.

# Section II

## Testing and Measurement