RR 90-05

PRINCIPAL COMPONENTS MODELS USING DISTANCES, INCLUDING WEIGHTS FOR VARIABLES AND DIMENSIONS

Jacqueline J. Meulman

Department of Data Theory

University of Leiden

PRINCIPAL COMPONENTS MODELS USING DISTANCES, INCLUDING WEIGHTS FOR VARIABLES AND DIMENSIONS

Abstract

In this paper various approaches to analyze the same set of multivariate data will be discussed and compared. In the first place, conventional principal components analysis will be contrasted with an approach that fits distances between the objects in multivariate data. In addition, two different ways of including weights in the analysis will be proposed. Although the possibilities pertain to multivariate analysis in general, the applications will be confined to principal component analysis models. Another important feature is the incorporation of optimal transformation of the variables while fitting the models to multivariate data.

Key words: multivariate analysis, principal components analysis, multi–dimensional scaling, individual differences scaling, nonlinear transformations, distance approximation.

INTRODUCTION

The starting point of this paper is the distance approach to (non)linear multivariate analysis (MVA), which takes the fitting of distances between the objects as a major objective (Meulman, 1986). In this approach MVA techniques are viewed as particular multidimensional scaling (MDS) problems; these concern the approximation of distances between the objects in an observation space (the space that is described by the variables in multivariate data) by distances in a low-dimensional representation space. With respect to approximation, conventional MVA redefines the MDS task by changing the quantities to be approximated from distances to scalar products. The distance approach to MVA defines the task directly on the distances, so a better fit of the distances in representation space to the distances in observation space may be obtained.

The class of basic MVA models, which include principal components analysis, canonical correlation analysis, redundancy analysis, and multiple correspondence analysis, is extended in this paper by including various differential weighting schemes. To be more specific, these include differential weighting of the variables, and differential weighting of the dimensions of the representation space (the INDSCAL model, proposed by Carroll & Chang, 1970). The number of different options that are available within these models are numerous. Therefore the basic ideas and motivation behind the distance approach to multivariate analysis will be discussed by taking one of the most widely used MVA techniques, principal components analysis (PCA), as a reference point. As an extension to both conventional MVA and MVA using distances, nonlinear generalizations will be discussed. These concern optimal transformation (scaling) of the variables, while fitting a particular MVA model.

We assume data are available for n objects or individuals and m variables. The columns of the nxm data matrix $\mathbf{Z}$ are defined by nx1 vectors $\mathbf{z}_j, j=1,...,m$, that contain observations on the variables and are assumed to have a mean of zero and a sum of squares equal to 1. The measurements on the objects for the m variables define the rows in $\mathbf{Z}$, and give the coordinates for each object in an m-dimensional *observation* space. The rows of the matrix $\mathbf{Z}$ will bé denoted by $\mathbf{z}_1',\ldots,\mathbf{z}_i',\ldots,\mathbf{z}_k',\ldots,\mathbf{z}_n'$. More generally, if the columns representing the m variables in $\mathbf{Z}$ are transformed, the set of transformed variables will be denoted by $\mathbf{Q}$. For each object we wish to find coordinates in a *representation* space; the dimensionality of this representation space is assumed to be p, and the coordinates for the n objects in this (unknown) space are contained in the rows of the nxp matrix X. The rows of X will be denoted by $\mathbf{x}_1',\ldots,\mathbf{x}_i',\ldots,\mathbf{x}_k',\ldots,\mathbf{x}_n'$.

Given this notation, a squared distance between a pair of objects $\{i,k\}$ in the m-dimensional observation space $\mathbf{Z}$ is defined by

$$d_{ik}^2(\mathbf{Z}) = (\mathbf{z}_i - \mathbf{z}_k)'(\mathbf{z}_i - \mathbf{z}_k) = (\mathbf{e}_i - \mathbf{e}_k)'\mathbf{ZZ}'(\mathbf{e}_i - \mathbf{e}_k),$$

where $\mathbf{e}_i$ is the i^{th} column of the nxn identity matrix $\mathbf{I}$. Applying the squared distance function $D^2(\cdot)$ that maps coordinates, $\mathbf{Z}$, into squared distances gives the matrix formulation:

$$D^2(\mathbf{Z}) = \boldsymbol{\alpha}\mathbf{1}' + \mathbf{1}\boldsymbol{\alpha}' - 2\mathbf{ZZ}',$$

with $\mathbf{1}$ an nx1 vector of all 1's and $\boldsymbol{\alpha}$ an nx1 vector containing the diagonal elements of $\mathbf{ZZ}'$; this will be denoted as $\boldsymbol{\alpha}$ = vecdiag($\mathbf{ZZ}'$). For the same pair of objects $\{i,k\}$, low-dimensional distances can be defined. The distance in the representation space X is written as

$$d_{ik}^2(\mathbf{X}) = (\mathbf{x}_i - \mathbf{x}_k)'(\mathbf{x}_i - \mathbf{x}_k) = (\mathbf{e}_i - \mathbf{e}_k)'\mathbf{XX}'(\mathbf{e}_i - \mathbf{e}_k),$$

and in matrix notation,

$$D^2(\mathbf{X}) = \boldsymbol{\alpha}\mathbf{1}' + \mathbf{1}\boldsymbol{\alpha}' - 2\mathbf{XX}',$$

with $\boldsymbol{\alpha}$ = vecdiag($\mathbf{XX}'$). Given these preliminaries, classical principal components analysis can be described as follows.

PRINCIPAL COMPONENTS ANALYSIS

From the geometrical point of view, principal components analysis is closely related to classical multidimensional scaling (Torgerson, 1952; called principal coordinates analysis by Gower, 1966). Both PCA and classical MDS are based on the approximation of a matrix of scalar products by another one that is of lower rank. A basic ingredient is the Young & Householder (1938) process that transforms a squared distance matrix $D^2(\mathbf{Z})$ into an nxn scalar product matrix $\mathbf{ZZ}'$, locating the origin in the centroid of points by using the nxn centering operator $\mathbf{J} = \mathbf{I} - (\mathbf{11}'/\mathbf{1}'\mathbf{1})$, so that $\mathbf{Z}$ satisfies $\mathbf{Z} = \mathbf{JZ}$:

$$-1/2\mathbf{J}D^2(\mathbf{Z})\mathbf{J} = -1/2\mathbf{J}(\boldsymbol{\alpha}\mathbf{1}' + \mathbf{1}\boldsymbol{\alpha}' - 2\mathbf{ZZ}')\mathbf{J} = \mathbf{ZZ}'. \tag{1}$$

The scalar product matrix $\mathbf{ZZ}'$ must be approximated by another scalar product matrix of lower rank, which objective can be written in the form of a loss function that will be called STRAIN:

$$\text{STRAIN}(\mathbf{X}) = \|\mathbf{ZZ}' - \mathbf{XX}'\|^2,$$

where $\|\cdot\|^2$ denotes a least squares discrepancy measure such that

$$\|\mathbf{ZZ}' - \mathbf{XX}'\|^2 = \text{tr}\,(\mathbf{ZZ}' - \mathbf{XX}')'(\mathbf{ZZ}' - \mathbf{XX}').$$

In the classical scaling approach to PCA, first an eigenanalysis of $\mathbf{ZZ}'$ is performed:

$$\mathbf{ZZ}' = \mathbf{K\Lambda K}',$$

where $\mathbf{K}$ is an nxt matrix containing the eigenvectors as columns, $\mathbf{\Lambda}$ is a txt diagonal matrix containing the ordered positive eigenvalues, $\lambda_1 \geq ... \geq \lambda_t$, and t denotes the rank of $\mathbf{Z}$ (for $t \leq m$). The optimal solution for obtaining a p-dimensional $\mathbf{X}$ (for $p \leq t$), is given by $\mathbf{X} = \mathbf{K}_p\mathbf{\Lambda}_p^{1/2}$, where the subscript p indicates the use of the first p columns in $\mathbf{K}$ (and the first p rows and columns of $\mathbf{\Lambda}$).

In contrast with the distance approach, PCA is usually written as a bilinear model (Kruskal, 1978). Loss functions for linear and bilinear models will be given the name STRIFE, and PCA minimizes

$$\text{STRIFE}(\mathbf{X};\mathbf{A}) = \|\mathbf{Z} - \mathbf{XA}'\|^2, \tag{2}$$

by performing a singular value decomposition of the matrix $\mathbf{Z}$:

$\mathbf{Z} = \mathbf{V\Psi W}'$.

A representation for the objects is obtained by choosing $\mathbf{X}$ as $\mathbf{V}_p\mathbf{\Psi}_p$; when we wish to represent the variables at the same time, $\mathbf{A}$ must be chosen as $\mathbf{A} = \mathbf{W}_p$. $\mathbf{Z}$ has t left singular vectors in $\mathbf{V}$ that are equal to the first t eigenvectors $\mathbf{K}_t$ of the matrix $\mathbf{ZZ}'$, and $\mathbf{Z}$ has t singular values in $\mathbf{\Psi}_t$ that are equal to the square root of the first t eigenvalues Λ_t of $\mathbf{ZZ}'$, i.e. $\mathbf{\Psi}_t = \Lambda_t^{1/2}$. So the optimal solution in classical scaling solution $\mathbf{X} = \mathbf{K}_p\Lambda_t^{1/2}$ is identical to the optimal PCA solution $\mathbf{X} = \mathbf{V}_p\mathbf{\Psi}_p$. This result establishes the distance properties of conventional PCA.

LEAST SQUARES DISTANCE APPROXIMATION

As an alternative to classical MDS, an approach was developed (Kruskal, 1964; Guttman, 1968) that is defined directly on the distances, and not on the scalar products. This form of MDS can be introduced in the distance framework by writing the loss function (called STRESS) for PCA as

$$\text{STRESS}(\mathbf{X}) = \|D(\mathbf{Z}) - D(\mathbf{X})\|^2. \tag{3}$$

To emphasize the difference between classical scaling and (3), we first note that classical scaling involves squared distances, due to (1). Secondly, the approximation is *from below*, which means that for each pair of observation and representation distances, $d_{ik}^2(\mathbf{X}) \leq d_{ik}^2(\mathbf{Z})$. Explicitly, when a p-dimensional approximation is obtained,

$$\begin{aligned} d_{ik}^2(\mathbf{X}) &= (\mathbf{e}_i\text{-}\mathbf{e}_k)'\mathbf{XX}'(\mathbf{e}_i\text{-}\mathbf{e}_k) = (\mathbf{e}_i\text{-}\mathbf{e}_k)'\mathbf{K}_p\Lambda_p\mathbf{K}_p'(\mathbf{e}_i\text{-}\mathbf{e}_k) \leq \\ &\leq (\mathbf{e}_i\text{-}\mathbf{e}_k)'\mathbf{K}\Lambda\mathbf{K}'(\mathbf{e}_i\text{-}\mathbf{e}_k) = d_{ik}^2(\mathbf{Z}), \end{aligned}$$

because the residual matrix $\mathbf{K}\Lambda\mathbf{K}' - \mathbf{K}_p\Lambda_p\mathbf{K}_p'$ is positive semidefinite. Using the fact that the eigenvalues are ordered, we obtain the sequence

$$d_{ik}^2(\mathbf{X}_1) \leq d_{ik}^2(\mathbf{X}_2) \leq \ldots \leq d_{ik}^2(\mathbf{X}_t) = d_{ik}^2(\mathbf{Z}),$$

where the subscript on $\mathbf{X}$ indicates the dimensionality of the representation space, and $\mathbf{X}_t$, with t equal to the rank of $\mathbf{Z}$, denotes the solution with maximum dimensionality.

Geometrically, approximation from below results from the fact that the points in the observation space are projected onto a low-dimensional subspace. When PCA is used as a multidimensional scaling technique, one should realize that conclusions based on small distances can be suspect. For instance, two object points that are close together could result from the representation of a small distance between the objects in observation space, or alternatively, of a badly fitted large distance, giving a false impression of the object pair's initial similarity.

Least squares distance approximation in principal components analysis based on STRESS cannot be described in terms of a projection. In contrast with approximation just from below, distances in observation space can also be approximated from above, and may actually be larger in representation space. By approximating some distances from above, the overall fit is improved. Low-dimensional solutions will generally give more reliable information with respect to small distances: small distances in low-dimensional space usually represent small distances in observation space, and large distances tend to be approximated from above (Meulman, 1986). The latter is typically less distorting in the interpretation of the structure given by the analysis than when large distances are represented by small ones. The need for approximation either from below or above is reflected in badness-of-fit. But compared to conventional PCA, an approximation that does allow both can reduce the number of dimensions necessary to describe the distances in observation space with an acceptable badness-of-fit.

NONLINEAR PCA

The incorporation of nonlinear transformations in conventional PCA has been discussed by various authors, including Kruskal and Shepard (1974), Young, Takane & De Leeuw (1978), Gifi (1981; 1990), and Winsberg & Ramsay (1983). For each variable in $\mathbf{Z}$ we wish to find an optimal transformation; an important possibility is the monotonic transformation. Monotonic transformations only take the order of the original values into account, and they maximize the fit of the monotonically transformed data to the PCA model. Other transformations include nominal transformations (these refer to categorical variables, with unordered categories), and smooth spline transformations (see Ramsay, 1989, for a review).

The optimality of nonlinear conventional PCA can be described as follows. In classical PCA the minimum loss is a function of the eigenvalues of the nxn scalar product matrix $\mathbf{ZZ}'$, which matrix has its first t eigenvalues in common with the eigenvalues of the mxm correlation matrix R($\mathbf{Z}$) of rank t; when m dimensions in observation space are replaced by p dimensions in representation space, the loss is equal to the sum of the $t - p$ smallest eigenvalues. If $\mathbf{Q}$ denotes the set of optimally transformed variables, and R($\mathbf{Q}$) is the correlation matrix, then the sum of the $t-p$ smallest eigenvalues of R($\mathbf{Q}$) will be as small as possible.

Nonlinear transformation of the variables can also be combined with least squares distance fitting. Here we minimize the loss function

$$\text{STRESS}(\mathbf{Q};\mathbf{X}) = \|\mathbf{D}(\mathbf{Q}) - \mathbf{D}(\mathbf{X})\|^2, \tag{4}$$

both over **Q** and **X**, which can be done by using an algorithm based on majorization (Meulman, 1986). The algorithm consists of three substeps: in the first step **X** is improved for fixed **Q** by using the Guttman transform (see De Leeuw & Heiser, 1980), in the second step an unconditional improvement **Y** of **Q** is found for fixed **X** by using the reversed Guttman transform (see Meulman, 1986), and in the third step **Q** is updated by ensuring that its columns are transformations of **Z**, while staying as closely as possible to **Y**. This is achieved by minimizing

$$\text{PROJECT}(\mathbf{Q}) = \|\mathbf{Y} - \mathbf{Q}\|^2,$$

over columns of **Q** satisfying both the transformation constraints and the normalization constraints $\mathbf{q}_j'\mathbf{q}_j = 1$. It has been shown that this three-step algorithm guarantees a convergent series of updates for both **Q** and **X**. In the next section the two approaches, i.e., nonlinear conventional PCA and nonlinear PCA using distances, will be compared.

APPLICATION 1

For all applications to follow, a set of variables have been analyzed that consist of 7 social indicator statistics taken from the statistical abstracts of the U.S.A. (1977). The data were collected by Wainer & Thissen (1981) in order to re-examine the search for "The Worst American State", and the variables are described in Table 1.

TABLE 1.

Social Indicator Statistics for the United States

POPULATION	1975 population
INCOME	Per capita income in dollars
ILLITERACY	Illiteracy rate in percent of population
LIFE	Life expectancy in years
HOMICIDE	1976 homicide and non-negligent manslaughter rate (per 1000)
SCHOOL	Percent of the population over age 25 who are high school graduates
FREEZE	Average number of days of the year in which temperature falls below zero

Previous analyses of these data, with transformations of the variables chosen as either third degree (De Leeuw & Meulman, 1986) or second degree polynomials (Meulman, 1986) pointed to the special role for the states Alaska (AK), Nevada (NV), and Hawaii (HI). The transformations of several variables displayed nonlinear functions that curved back at the extremes, although this phenomenon did not occur for the same variables in a conventional PCA compared to a PCA using distances. When transformations are so different, it is difficult to distinguish the transformation effect from the effect of approximation. To study here both effects in the same analysis, the transformations were chosen from the more limited class of monotonic second degree polynomials, and fitted using the approach to integrated splines discussed in Winsberg & Ramsay (1983), and Ramsay (1989). Effectively, only two parameters, the M-spline coefficients, are estimated for each transformation, and because the transformed variables are normalized, transformations in conventional PCA and PCA using distances become identical when for both either the first or the second coefficient becomes zero to ensure monotonicity.

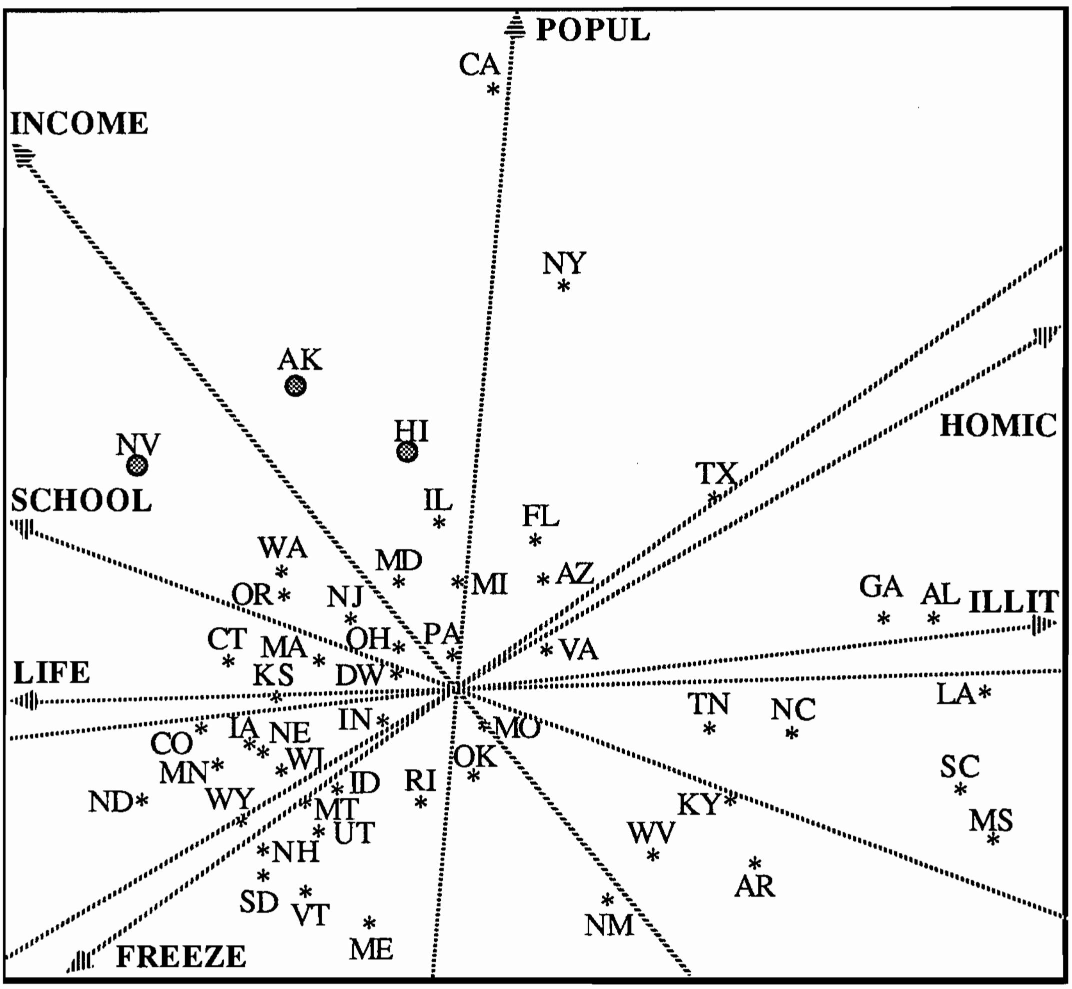

Figure 1. Two-dimensional solution for United States: PCA Using distances

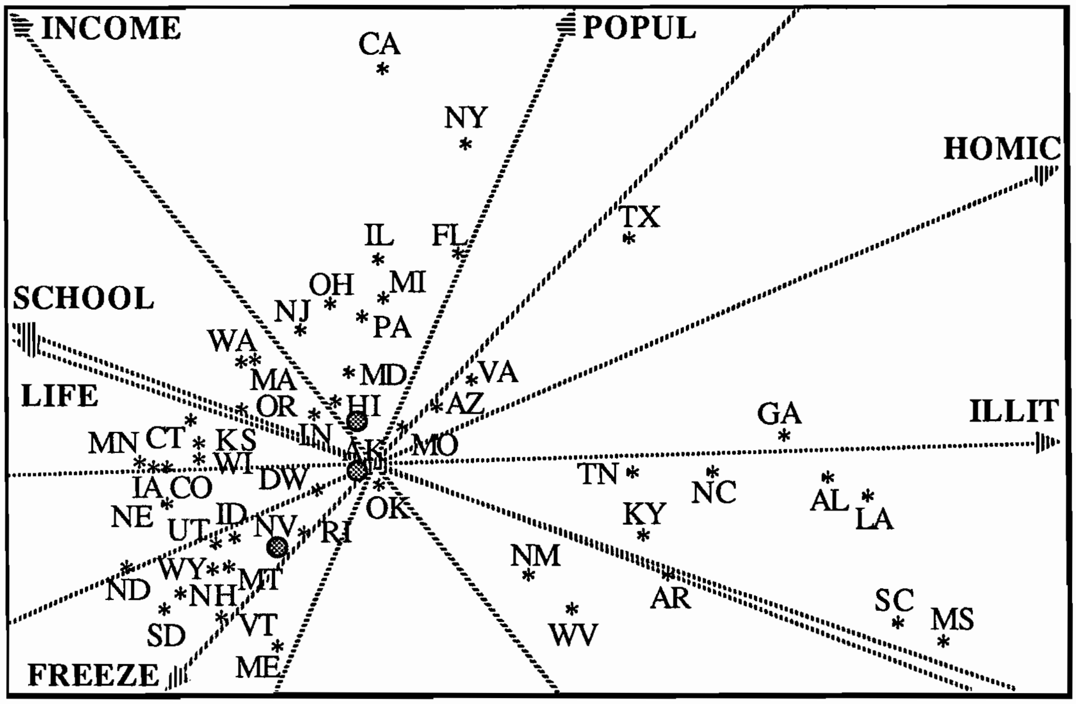

Figure 2. Two-dimensional configuration for United States: Conventional PCA

The points for the 50 states are shown in Figure 1 for the distance approach to PCA, and in Figure 2 for conventional PCA, both using spline transformations. The two Figures are on the same scale, showing that the overall configuration for conventional PCA is much smaller. With respect to the distance approximation, as measured by STRESS in (4), PCA using distances fits the transformed data 2.5 times better than the conventional PCA (see Table 2); this is reflected in the size of the configurations.

TABLE 2

Fit Measures for two approaches to Principal Components Analysis

	PCA Using Distances	Conventional PCA
STRESS	.159	.404
Correlations between transformed variables and fitted variables		
POPULATION	.72	.91
INCOME	.92	.86
ILLITERACY	.88	.90
LIFE	.75	.84
HOMICIDE	.87	.88
SCHOOL	.89	.88
FREEZE	.81	.81

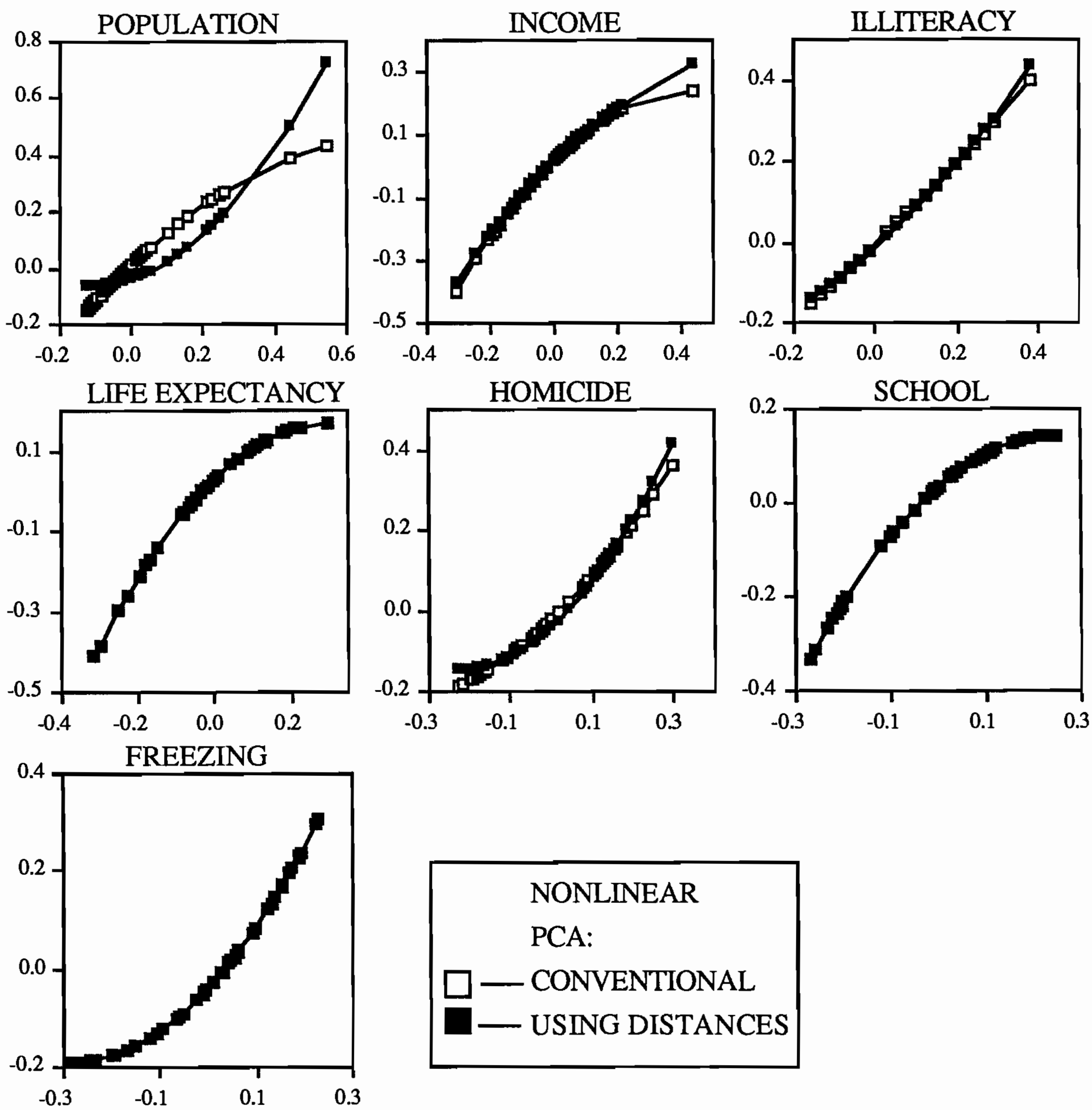

Figure 3. Second degree spline transformations for United States data

Except for this scale factor, the two configurations have many local resemblances. In both spaces, directions for the transformed variables are fitted. These are obtained by a projection, defined by $\mathbf{Q}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$. For conventional PCA $\mathbf{Q}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{A}$ in (2), using $\mathbf{Q}$ in the role of $\mathbf{Z}$. The fit of the transformed variables in the two configurations, the correlation between the $\mathbf{q}_j$ and $\mathbf{Xa}_j'$, which is the criterion in conventional PCA, is given in Table 2. The fit measures for ILLITERACY, HOMICIDE, SCHOOL, and FREEZE are much alike; LIFE and especially POPULATION are much better fitted in the conventional PCA solution, while INCOME fits better in PCA using distances. In both configurations, the overall structure of the USA data is fitted quite well; the first dimension can be characterized by a "quality of life" factor; the second gives the relative positions of the states with respect to POPULATION and FREEZE. The first dimension discriminates the Southern states at the right hand side of the Figures 1 and 2, scoring poorly on the quality of life variables, from states in the Northeast and the Midwest. The second dimension shows that especially California and New York score high on POPULATION and INCOME, and low on FREEZE. Although the two representations of the structure are much alike, there are a few notable exceptions. Alaska, Nevada, and Hawaii do not fit very well in the structure. In the conventional PCA solution, they are quite near the origin of the configuration, which follows from the projection of the points into the subspace; in the PCA using distances, these points are located in the upper left corner. Consider, for example, Alaska. It has a small population, many homicides, a high level of education, it's cold, and has the highest income. In the conventional PCA, none of these characteristics is apparent; in the PCA using distances, Alaska is located in a correct position according to INCOME, SCHOOL, and HOMICIDE, but in a wrong one with respect to FREEZE and especially POPULATION.

When we look at the transformation plots in Figure 3 we see that the transformations for the two techniques are quite similar, except for INCOME and POPULATION. The conventional PCA analysis has given Alaska a lower value for INCOME, whereas the distance PCA has maintained the high value. This has been reversed for POPULATION; the small values have been transformed linearly in conventional PCA, but in the distance PCA, this part of the transformation is flat, giving Alaska the same value as many other states. As a result, Alaska has a high position for INCOME and POPULATION in Figure 1, whereas the fit for INCOME is large and for POPULATION rather small. In the conventional PCA solution, POPULATION fits better than INCOME.

COMPARING DIFFERENT SETS OF VARIABLES

Another important class of techniques in multivariate analysis is called generalized canonical analysis. Here two or more sets of variables are analyzed; important special cases are multiple regression, discriminant analysis, redundancy analysis, and multiple correspondence analysis. To investigate the relationship between M groups of variables, $\mathbf{Z}_1,...,\mathbf{Z}_J,...,\mathbf{Z}_M$, a set of weight matrices $\mathbf{A} = \{\mathbf{A}_1,...,\mathbf{A}_J,...,\mathbf{A}_M\}$, of order $m_1 \text{x} p,...,m_J \text{x} p,...,m_M \text{x} p$, must be found so the linear combinations $\mathbf{Z}_1\mathbf{A}_1,...,\mathbf{Z}_J\mathbf{A}_J,...,\mathbf{Z}_M\mathbf{A}_M$ resemble each other as closely as possible. One way of characterizing maximal resemblance is to minimize the least squares discrepancy between each separate $\mathbf{Z}_J\mathbf{A}_J$ and an unknown $n\text{x}p$ comparison matrix $\mathbf{X}$. Optimal transformations have been incorporated also in this class of techniques, so that the general loss function can be written as

$$\text{STRIFE}(\mathbf{Q};\mathbf{X};\mathbf{A}) = M^{-1}\sum_{J=1}^{M} \|\mathbf{Q}_J\mathbf{A}_J - \mathbf{X}\|^2, \tag{5}$$

to be minimized over $\mathbf{X}$, $\mathbf{A}_J$, and $\mathbf{Q}_J$, $J = 1,\ldots,M$, under some suitable chosen normalization constraint. Details about the minimization of this loss function and its applications can be found in Van der Burg (1988), Van der Burg, De Leeuw, & Verdegaal (1988), and Gifi (1990).

In the latter mentioned book a system of MVA techniques is developed that has the notion of homogeneity as starting point. In this system, nonlinear PCA can be viewed as a special case of (5), and is written in the form of the loss function:

$$\text{STRIFE}(\mathbf{Q};\mathbf{x};\mathbf{a}) = m^{-1}\sum_{j=1}^{m} \|\text{a}_j\mathbf{q}_j - \mathbf{x}\|^2. \tag{6}$$

Here each set contains only one optimally transformed variable $\mathbf{q}_j$ that is given a weight a_j so that the weighted variables resemble the unknown $\mathbf{x}$ as closely as possible. Differential weighting of variables can also be incorporated in the distance approach to PCA; the basic line of thought is extended by combining features from conventional MVA, which seeks weighted sums of variables, and the alternative approach, which seeks to fit distances.

DIFFERENTIAL WEIGHTING OF VARIABLES

The general procedure derives distances from the M different sets $\mathbf{Q}_J$; to stay close to PCA as in (6), each variable defines a set, so the data consist of m matrices $\text{D}(\mathbf{q}_j)$ of order nxn, j=1,...,m. From here, there a two ways to go. First of all, we can minimize the loss function

$$\text{STRESS}(\mathbf{Q};\mathbf{X};\mathbf{a}) = m^{-1}\sum_{j=1}^{m} \|(\text{D}(\text{a}_j\mathbf{q}_j) - \text{D}(\mathbf{X})\|^2, \tag{7}$$

which is a straightforward extension of (6) that does not approximate the weighted transformed variables themselves, but the distances they generate. On the other hand we can minimize

$$\text{STRESS}(\mathbf{Q};\mathbf{X};\mathbf{d}_1,\ldots,\mathbf{d}_j,\ldots,\mathbf{d}_m) = m^{-1}\sum_{j}^{m}\|(\text{a}_j\mathbf{d}_j - \mathbf{d}_{(\text{X})}\|^2. \tag{8}$$

Here $\mathbf{d}_j$ is an $1/2n(n-1)$x1 vector that contains the lower diagonal elements of the matrix $\mathbf{D}_j = \{\text{d}_{ik}\}$ for variable j. The $\mathbf{d}_j$ could be used as variables in a classical PCA as is done by Tucker & Messick (1963), but a linear combination of these variables does not necessarily give a principal component that consists of Euclidean distances between the objects. In contrast, the component scores $\mathbf{d}_{(X)}$ in (8) do give Euclidean distances between objects in a p-dimensional configuration $\mathbf{X}$; minimization over $\mathbf{X}$ ensures optimality with respect to the MDS task. In (7) the optimal transformation is defined on the $\mathbf{q}_j$; (8) can be generalized by defining the transformation on the $\mathbf{d}_j$, giving $\mathbf{d}_j^*$ which are optimal transformations of the given vectors of dissimilarities $\mathbf{d}_j$. The latter can be viewed as a form of (6), with $\mathbf{d}_j$ in the role of $\mathbf{q}_j$, with restrictions on the component scores; it is also a form of nonmetric multidimensional scaling (Kruskal, 1964).

When $\mathbf{q}_j$ is (7) is chosen as $\mathbf{z}_j$, and $\mathbf{d}_j$ in (8) as the vector with lower diagonal elements of $D(\mathbf{z}_j)$, then due to the homogeneity of the Euclidean distance function $D(a_j\mathbf{z}_j) = a_jD(\mathbf{z}_j)$, and (7) and (8) will give the same solution for $\mathbf{X}$. When we have at the same time both multivariate data and given dissimilarity measures (that are not necessarily Euclidean distances) defined on the same objects, and transformations are desired for both, the two transformation approaches can be combined. In that case we obtain, in the same analysis, transformations for the given dissimilarity measures and for the variables in the multivariate data. The algorithm that attains this objective is given in Meulman & Verboon (1989).

APPLICATION 2

In this application the United States data are analyzed again, and because we wish to transform the variables, we choose the form of PCA with differential weighting of variables as given in (7). The weights that are obtained are given in Table 3, and presented in a diagram in Figure 4. The average weight equals .859, so FREEZE but especially POPULATION obtain weights that are below average, and INCOME is only slightly above.

TABLE 3

DIFFERENTIAL WEIGHTS FOR VARIABLES IN PCA USING DISTANCES

1	POPULATION	.711	5	HOMIC	.903
2	INCOME	.867	6	SCHOOL	.896
3	ILLIT	.912	7	FREEZE	.849
4	LIFE	.878			

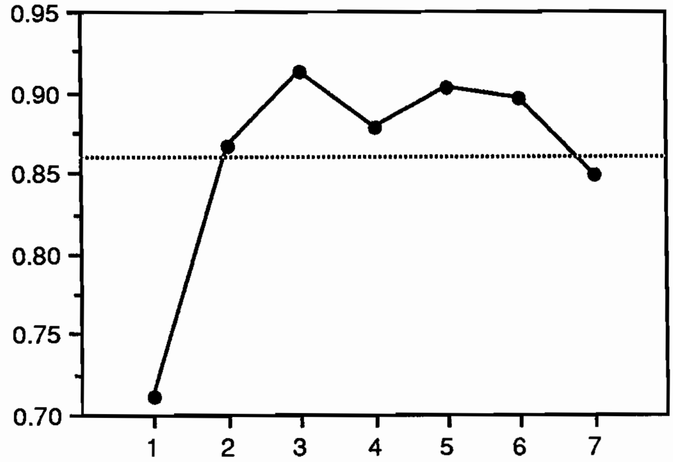

Figure 4. Differential weights for the 7 variables in PCA

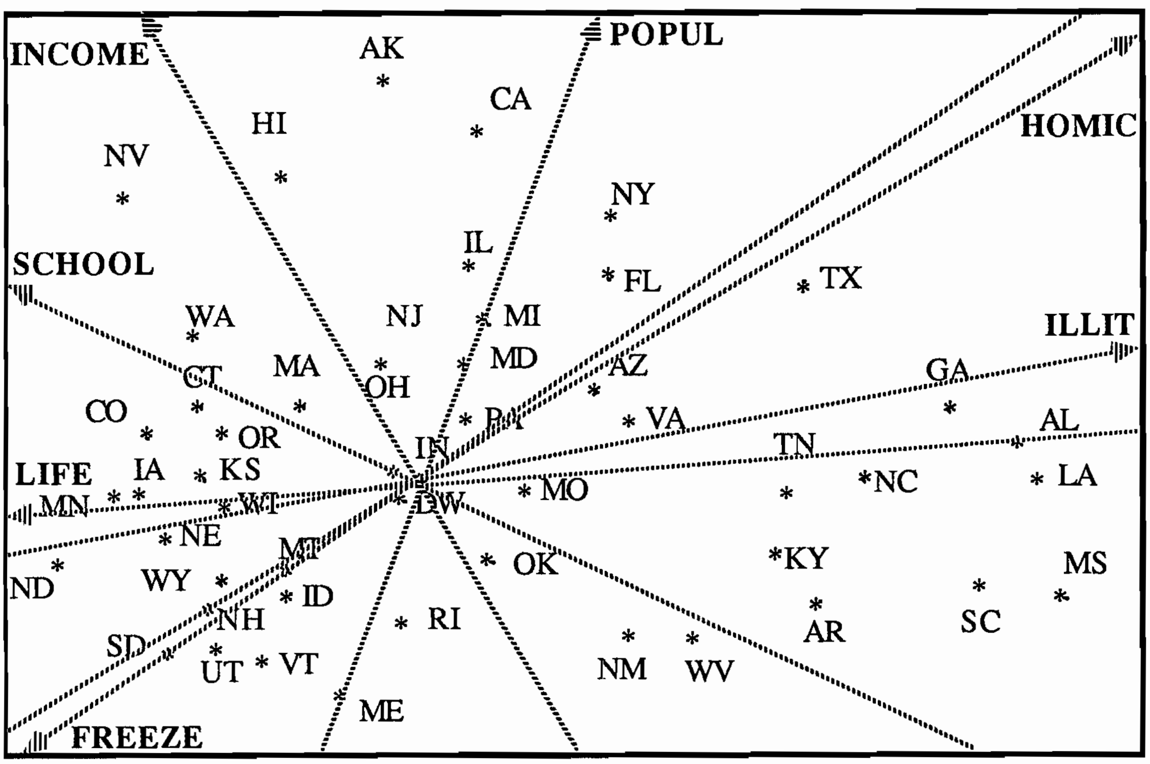

Figure 5. Two-dimensional configuration for United States: differential weights for variables

Looking at the configuration for the model with weighted variables in Figure 5, we see that the position of most of the states is similar to the unweighted PCA using distances in Figure 1. Their position of Alaska, Hawaii and Nevada, however, is much more dominant in the second dimension; this due to the fact that POPULATION and FREEZE are less important than the other variables in the analysis. Again, to facilitate the interpretation, the transformed variables are projected into the space to obtain a joint plot.

DIFFERENTIAL WEIGHTS FOR DIMENSIONS

In Meulman (1986) a loss function was developed that compares sets of multivariate data through the distances between the objects as an alternative to (5). In this section, however, it is shown how an analysis with several sets of variables can be integrated with the individual differences scaling tradition. Again individual distance matrices are created on the basis of each variable $\mathbf{q}_j$. In the previous sections there was always one single representation space; according to individual differences scaling the objective is to find m representation spaces $\mathbf{X}_j$ in which the distances approximate the distances in $\mathbf{q}_j$, where the separate spaces are restricted by requiring that the $\mathbf{X}_j$ are linear combinations of a common space $\mathbf{X}$. This restriction is written as $\mathbf{X}_j = \mathbf{X}\mathbf{A}_j$, where $\mathbf{A}=\{\mathbf{A}_1,...,\mathbf{A}_j,...,\mathbf{A}_m\}$ denotes the set of individual differences parameters. In formula the model can be written as

$$\mathrm{STRESS}(\mathbf{X};\mathbf{Q};\mathbf{A}) = m^{-1}\sum_j^m \|\mathrm{D}(\mathbf{q}_j) - \mathrm{D}(\mathbf{X}\mathbf{A}_j)\|^2.$$

There are a variety of choices for $\mathbf{A}_j$ (cf. the class of models in Heiser & Stoop, 1986). The Carroll & Chang (1970) INDSCAL model requires the $\mathbf{A}_j$ to be diagonal weight matrices, which give saliences for the dimensions of the common space $\mathbf{X}$. Another attractive model requires the $\mathbf{A}_j$ to be of reduced rank; the common space $\mathbf{X}$ may be high-dimensional, while the individual spaces $\mathbf{X}\mathbf{A}_j$ are low-dimensional. The full and reduced rank model have been discussed under the name General Euclidean Model in Bloxom (1978) and Young (1984). In the most general case $\mathbf{A}_j$ may be a matrix of full rank: this is the Carroll & Chang (1972) IDIOSCAL model. Here $\mathbf{A}_j$ can be decomposed, giving a rotation matrix $\mathbf{T}_j$, which rotates $\mathbf{X}$ into an optimal position and a diagonal weight matrix $\mathbf{W}_j$, which gives saliences for the dimensions in the rotated space. The algorithm that fits the various options for $\mathbf{A}_j$ together with an optimal $\mathbf{X}$ and transformations in $\mathbf{Q}$ is described in Meulman & Heiser (1989).

APPLICATION 3

In the third application the INDSCAL model is applied to the United States data, so $\mathbf{A}_j$ is diagonal and we obtain differential weights for the two dimensions in $\mathbf{X}$. The weights obtained in the analysis are given in Table 4. The weights can be depicted in a two-dimensional diagram

(Figure 6) as vectors extending from the origin; it clearly shows the influence of variables on the two dimensions (the diagonal represents an equal influence). SCHOOL, LIFE, ILLIT are dominant in the first dimension, HOMIC and INCOME have somewhat more influence on the second dimension, although primarily on the first, the latter pattern is reversed for FREEZE, and POPULATION dominates the second dimension and has almost no influence on the first. When we recall the variable-weights in the previous application, we see that the variables that had weights below average now have dominant weights in the second dimension.

TABLE 4

DIFFERENTIAL WEIGHTS FOR DIMENSIONS
IN PCA USING DISTANCES

	Weights	
Dimension	1	2
POPULATION	.014	.756
INCOME	.388	.261
ILLIT	.440	.185
LIFE	.449	.160
HOMIC	.414	.283
SCHOOL	.442	.154
FREEZE	.295	.442

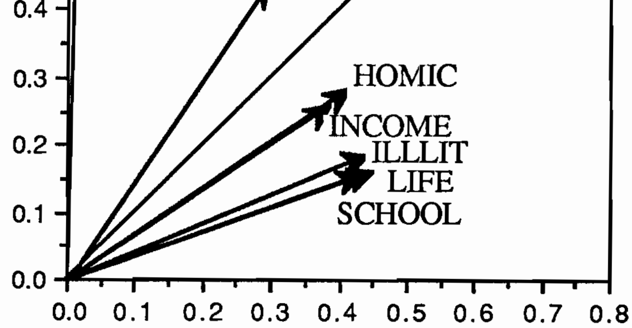

Figure 6. Differential weights for dimensions
in the INDSCAL model

When we inspect the configuration in Figure 7 we see that we have obtained quite a different placement for Alaska, Nevada and Hawaii compared to Figure 1 and 4. Especially Alaska and Nevada are placed in a position that is determined by POPULATION and FREEZE: they have 'moved' to the bottom of the configuration.

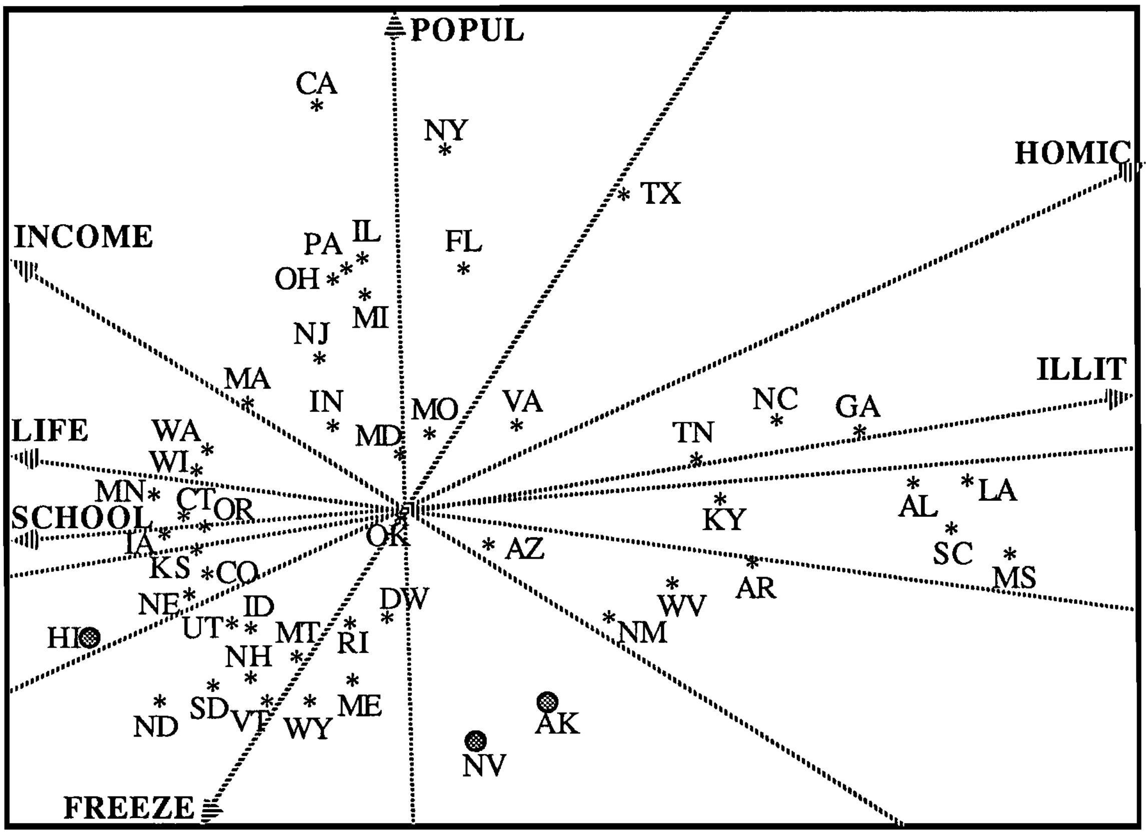

Figure 7. Two-dimensional configuration for United States: differential weights for dimensions

The transformations for differential weighting of variables and dimensions are given in Figure 8. The transformations in the two models are very similar; when we compare them with Figure 3 we see that the transformations of POPULATION and INCOME are closer to the transformations in conventional PCA than in PCA using distances. Thus the fact that we have included features from conventional MVA in the distance approach comes up in the transformations. The two configurations in Figure 4 and Figure 7, however, are quite different from Figure 2, as far as Alaska, Nevada and Hawaii are concerned. This difference must be due to the approximation from above and below, since the transformations are very similar.

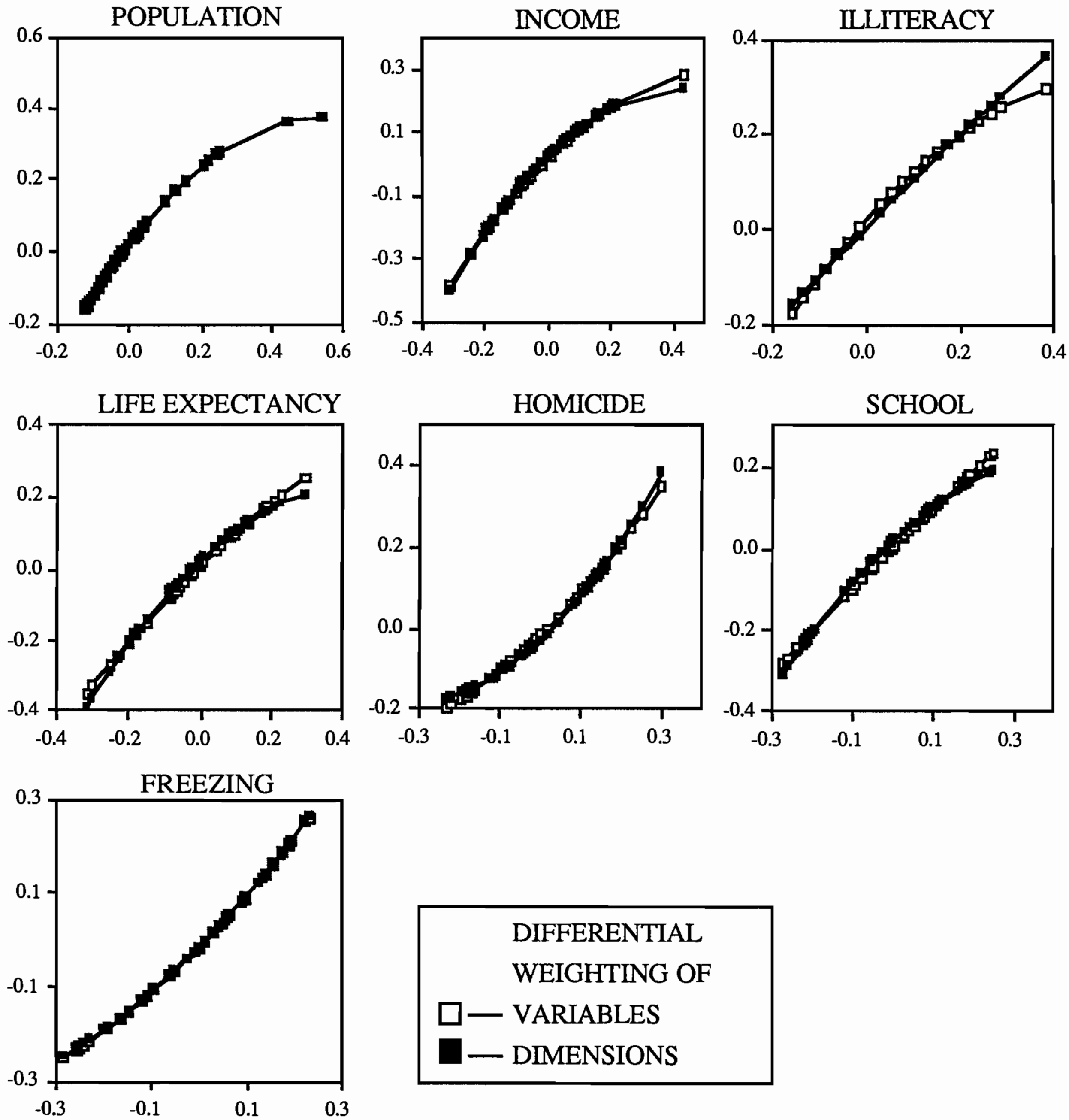

Figure 8. Second degree spline transformations for United States data

CONCLUDING REMARKS

By combining differential weighting of either variables or dimensions with least squares distance fitting in MVA, we have extended the basic class of techniques. Throughout this paper ordinary Euclidean distances have been used in the observation space(s), because Euclidean distances are related to PCA. When we have several sets of variables $\mathbf{Q}_1, \ldots \mathbf{Q}_J, \ldots \mathbf{Q}_M$, we may also use other distance functions, for instance the Mahalanobis distance that corrects for the covariances between the variables in a set. When the Mahalanobis distance is applied to each of the sets, we obtain an alternative for (5). When the $\mathbf{Q}_J$ consist of binary indicator variables, we may apply the chi-squared distance to obtain an alternative for multiple correspondence analysis. We may also apply a mixture of Euclidean and Mahalanobis distances to get alternatives for redundancy analysis and multiple regression, or a mixture of Mahalanobis and chi-squared distances for discriminant analysis. This generalizes the approach discussed to most other MVA techniques apart from PCA.

The advantage of the distance approach over conventional MVA techniques is a least squares approximation of the distances. We have also the most interesting possibility to analyze given dissimilarity data and multivariate data in the same analysis, while both may be transformed. The variables, however, play an inferior role: whether they are transformed or not, they are merely used to derive the distances between the objects. By including differential weighting schemes the variables are given more influence in the analysis compared to straightforward distance fitting.

REFERENCES

Bloxom, B. (1978). Constrained multidimensional scaling in N spaces. *Psychometrika, 43*, 397-408.

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

Carroll, J.D., & Chang, J.J. (1972). IDIOSCAL(Individual differences in orientation scaling): *A generalization of INDSCAL allowing IDIOsyncratic reference systems as well as an analytic approximation to INDSCAL.* Paper presented at the Psychometric Society Meeting, Princeton, NJ.

De Leeuw, J., & Heiser, W.J. (1980). Multidimensional scaling with restrictions on the configuration. In P.R. Krisnaiah (Ed.), *Multivariate analysis, Vol. V* (pp. 501-522). Amsterdam: North-Holland.

De Leeuw, J., & Meulman, J. (1986). Principal component analysis and restricted multidimensional scaling. In W. Gaul & M. Schader (Eds.), *Classification as a tool of research* (pp. 83-96). Amsterdam: North-Holland.

Gifi, A. (1990). *Nonlinear multivariate analysis* [first edition 1981, Leiden: Dept. of Data Theory]. Chichester: Wiley.

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

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

Heiser, W.J., & Stoop, I. (1986). *Explicit SMACOF algorithms for individual differences scaling*. Research Report RR-86-14. Leiden: Dept. of Data Theory.

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

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: The Free Press.

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

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

Meulman, J., & Heiser, W.J. (1989). *Nonlinear multivariate analysis through generalized individual differences models*. Paper presented at the second Conference of the International Federation of Classification Societies, Charlottesville, Virginia, USA.

Meulman, J., & Verboon, P. (1989). *Nonlinear principal components analysis with distance restrictions on the component scores*. Paper presented at the Sixth European Meeting of the Psychometric Society, University of Leuven, Belgium.

Ramsay, J.O. (1989). Monotone regression splines in action. *Statistical Science, 4,* 425-441.

Tucker, L.R., & Messick, S. (1963). An individual differences model for multidimensional scaling. *Psychometrika, 28,* 333-367.

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

Van der Burg, E. (1988). *Nonlinear canonical correlation and some related techniques*. Leiden: DSWO Press.

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.

Wainer, H., & Thissen, D. (1981). Graphical data analysis. *Annual Review of Psychology, 32,* 191-241.

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

Young, F.W. (1984). The general Euclidean model. In H.G. Law et al. (Eds.), *Research methods for multimode data analysis* (pp. 440-469). New York: Praeger.

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.

Young, G., & Householder, A.S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika, 3,* 19-22.