RR – 86-06

A MULTIDIMENSIONAL SCALING APPROACH TO MULTIVARIATE ANALYSIS

Jan de Leeuw
Department of Data Theory

University of Leiden

Paper presented at the Diana II Conference, Liblice, Czechoslovakia, May 26 through May 30, 1986

Introduction

In *multidimensional scaling* (MDS, from now on) a geometrical representation of the objects in the study is derived from information about the *dissimilarities* of these objects. The objects are represented in a metric space in such a way that dissimilar objects are relatively far apart, while similar objects are relatively close. In most MDS techniques the target space, in which we make the representation, is low-dimensional Euclidean space. But in various forms of *cluster analysis*, which can be interpreted as MDS techniques, the metric space in which we make a picture of the objects is a tree or some other combinatorial structure. For a discussion of the many different MDS techniques we refer to Carroll and Arabee (1980), De Leeuw and Heiser (1980, 1982), and Young (1984). It is clear that MDS has a very strong geometrical orientation, and that the key notion is *distance*.

In these respects MDS is quite different from multivariate analysis (MVA), at least from the usual formulations of MVA. These proceed either by constructing *linear combinations* of variables with optimality properties defined in terms of *correlation coefficients*, or by specifying structural models for correlated variables, which are usually assumed to be *multinormally distributed*. In these formulations the geometrical notions play a relatively minor role, and the emphasis is shifted to linear algebra in the form of matrix calculus. Distances in low-dimensional space, the key concepts for MDS, are replaced by inner products of vectors in high-dimensional space. Nevertheless it is important to realize that the basic mathematical structure used in most forms of MVA and MDS is the same. It is good old Euclidean space, with the familiar inner product defining the angle, and with the corresponding Pythagorean distance measure.

This basic similarity can be exploited in various ways. In Gifi (1981) the more common MVA-techniques are organized into a system which takes *homogeneity analysis* (also known als *multiple correspondence analysis*) as the basic technique, and derives the other techniques as specializations of homogeneity analysis. And homogeneity analysis, as defined and explained by Gifi, is basically an MDS technique which makes low-dimensional pictures of data and focuses on the distance between points in these pictures. Meulman (1986) takes these developments a step further. She defines very

general MDS techniques, which can be specialized to ordinary MDS techniques on one side, and to the Gifi system of techniques on the other side. Thus both classical MDS and classical MVA are special cases of this much more general set-up. In this paper we shall first discuss homogeneity analysis in considerable detail, emphasizing the geometrical properties of the representation. We then show how many of the classical multivariate techniques can be formulated as forms of homogeneity analysis with restrictions on the representation. We extend the results of Gifi (1981) and De Leeuw (1984a), by using the approach of De Leeuw (1984b).

The restrictions that can be used in homogeneity analysis have two major purposes. In the first place they incorporate prior information the investigator may have, in the second place they enhance the stability of the homogeneity analysis solutions. Thus they illustrate one of the general principles of data analysis: incorporating prior information into the technique improves its stability. If this is taken too far, as in much of classical statistics, then the data have not enough room to influence the solution. Prior information (which is often just invented for the purpose of applying a standard technique) dominates the solution, and the data are squeezed into the (possibly ill-fitting) mould provided by the model. On the other hand the opposite extreme also has its problems. Techniques will tend to focus on accidental and non-stable properties in the data, so-called *chance capitalization* or *overfitting*. The representation is highly unstable, and cannot be reproduced by subsequent investigators. There is not enough structure in the results, which can consequently not be related to previous theory, and which cannot be interpreted in a rational way. It is clear that we want to avoid both extremes, but it follows from the nature of our techniques that we shall always tend to be on the empiristic or data-centered side of the continuum. Together with, for example, Tukey (1962), Benzécri (1973), Guttman (1977), Gifi (1981). We refer to these books and papers for more methodological discussion.

Multivariables

We start our formal developments in this paper by providing some definitions. In

MVA we always study a number of *variables* , defined on a set of *objects* . A variable is a function, and we use the familiar notation $\phi: \Omega \to \Gamma$. Here Ω is the *domain* of the variable ϕ, and Γ is its *target* , containing the possible *values* of the variable. Elements of the target are also called the *categories* of a variable. A variable ϕ associates with each $\omega \in \Omega$ a category $\phi(\omega) \in \Gamma$. In practical applications and in actual data analysis the domain Ω will be a finite set $\{\omega_1,...,\omega_n\}$. For theoretical purposes the domain can be infinite. If Ω is a probability space, for instance, and ϕ is measurable, then the variable is a *random variable*. Targets can be finite or infinite. In many cases the target is the set of reals or the integers, i.e. $\Gamma = \mathbb{R} =]-\infty,+\infty[$, or $\Gamma = \mathbb{N} = \{0,1,2,...\}$. But it is also possible that $\Gamma =$ {close, moderate, distant}, or $\Gamma =$ {protestant, catholic, buddhist, other}. In MVA we analyze several variables at the same time. This requires some additional terminology. A *multivariable* is a set of variables with a common domain. We use the notation $\Phi = \{\phi_j \mid j \in J\}$, where $\phi_j : \Omega \to \Gamma$, and where J is the *index set* of the multivariable. Index sets, again, need not be finite, although in practical data analysis they always will be. The variables in Φ have the common domain Ω, but they have possibly different targets Γ_j. *Multivariate analysis* studies the structure of multivariables.

In Table 1 we have presented a small example with 10 objects and three variables. The objects are 10 cars, the variables are price (in $ 1000), gas consumption (litres per 100 km, on the expressway), and weight (in 100 kg). The data are taken from a larger matrix used by Winsberg and Ramsay (1983, p. 587), who took their matrix from the April, 1983 issue of *Consumer Report*. The targets of all three variables are the positive reals $\mathbb{R}^+$ (with just one decimal digit). Table 2 gives another multivariable, derived from the previous one by *discretization* . The targets are now natural numbers, which are used for labelling intervals of the positive real axis.

Indicator functions and matrices

If variable ϕ_j maps Ω into Γ_j, then the *indicator function* η_j of this variable maps $\Omega \times$

	Price	Gas	Weight
Chevrolet Chevette	5.6	6.9	9.7
Dodge Colt	5.7	5.1	8.8
Plymouth Horizon	6.3	5.5	9.9
Fort Mustang	7.6	6.7	12.0
Pontiac Phoenix	8.6	6.9	12.1
Dodge Diplomat	9.4	10.2	15.5
Chevrolet Impala	10.1	7.5	16.9
Buick Regal	10.5	7.8	15.0
AMC Eagle	10.7	11.7	15.7
Oldsmobile 98	13.3	8.7	18.3

Table 1: Car Data, numerical

	Price	Gas	Weight
Chevrolet Chevette	1	1	1
Dodge Colt	1	1	1
Plymouth Horizon	1	1	1
Fort Mustang	2	1	2
Pontiac Phoenix	2	1	2
Dodge Diplomat	2	3	2
Chevrolet Impala	3	2	3
Buick Regal	3	2	2
AMC Eagle	3	3	2
Oldsmobile 98	4	2	3

Table 2: Car Data, categorized

	Price	Gas	Weight
Chevrolet Chevette	1000	100	100
Dodge Colt	1000	100	100
Plymouth Horizon	1000	100	100
Fort Mustang	0100	100	010
Pontiac Phoenix	0100	100	010
Dodge Diplomat	0100	001	010
Chevrolet Impala	0010	010	001
Buick Regal	0010	010	010
AMC Eagle	0010	001	010
Oldsmobile 98	0001	010	001

Table 3: Car Data, indicators

Γ_j into $\{0,1\}$ by the rule $\eta_j(\omega,\gamma) = 1$ if $\phi_j(\omega) = \gamma$, and $\eta_j(\omega,\gamma) = 0$ otherwise. Table 3 shows the indicator functions corresponding with Table 2. Indicator functions for Table 1 would look more complicated, because they are defined on $\{\omega_1,...,\omega_{10}\} \times \mathbb{R}^+$, which does not fit on the page. For variables with a finite target and a finite domain (which are the only ones actually occurring in practice) the indicator function takes the form of an *indicator matrix* . For indicator matrices we use G_j. The number of rows is $\text{card}(\Omega) = n$ and its number of columns is $\text{card}(\Gamma_j) = k_j$.

The indicator matrix (or the *indicator supermatrix* , which consists of all matrices G_j next to each other) can be interpreted as the incidence matrix of a graph. The graph for the car data is drawn in Figure 1. It has n x m = 30 lines. It looks kind of messy, because there are many lines that cross. Now think of the graph as a picture of the multivariable, i.e. as a joint picture of the objects and the targets of the variables, in the Euclidean plane. The graph will be much less messy if the lines are as short as possible, i.e. if objects are close to the categories of the variables that they score in. This is the basic idea of homogeneity analysis, in words. We want to make a picture of the graph of a multivariable in low-dimensional Euclidean space in such a way that the points connected by a line are relatively close together (and the points not connected by lines are relatively far apart). By the triangle inequality this implies that objects with similar *profiles* (i.e. objects that are often in the same categories) will be close, and categories containing roughly the same objects will be close as well. We shall now make these notions quantitative by defining a suitable loss function to be minimized.

Loss of homogeneity

Let us try to find a quantification of objects and categories in p-dimensional space. The quantifications of the n objects can be collected in an n x p matrix X, the quantifications of the k_j categories of variable j in a k_j x p matrix Y_j. The sum of squares of the distances between objects and the categories they score in is given by

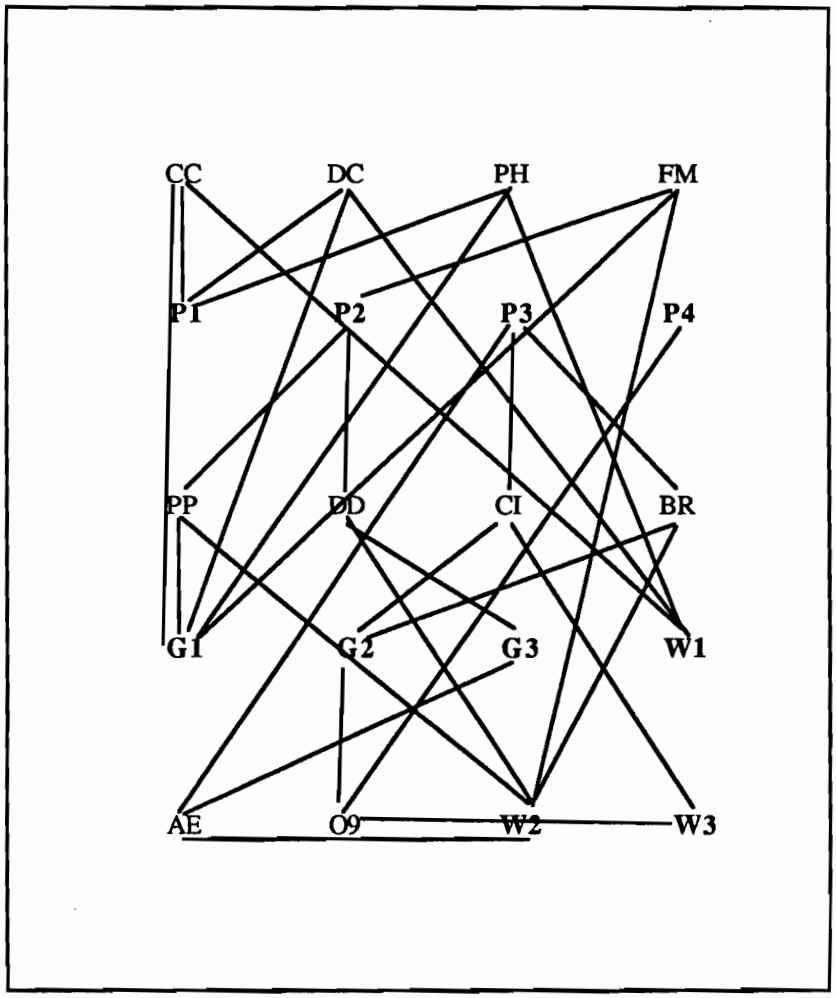


Figure 1: Car Data, arbitrary solution

$$\sigma(X;Y_1,...,Y_m) = \Sigma_j \, SSQ(X - G_jY_j) \qquad (1)$$

where SSQ(.) is convenient shorthand for the sum of squares of the elements of a matrix or vector.

We could now say that homogeneity analysis is a technique that minimizes (1) over X and the Y_j, but this would not be very satisfactory. In the first place we can set X = 0 and $Y_j = 0$ for all j. This gives loss equal to zero, and consequently perfect homogeneity. In fact, more generally, taking all elements of X equal to a constant c, and taking all elements of all Y_j equal to c as well, gives loss zero. We want to exclude these trivial solutions by imposing suitable normalizations. The one we choose, for the moment, is X'u = 0, with u a vector with all elements equal to +1, and X'X = I, the identity matrix. A matrix X satisfying these restrictions is *normalized* . We now define *homogeneity analysis* as minimization of the loss function (1) over all Y_j (j=1,...,m) and over all normalized X.

Define

$$\sigma(X;*,...,*) = \min \{\sigma(X;Y_1,...,Y_m) \mid Y_1,...,Y_m\}. \qquad (2)$$

In order to compute this partial minimum we define $\underline{Y}_j = \{G_j\}^+X$, with + denoting the Moore-Penrose inverse. Write $Y_j = \underline{Y}_j + (Y_j - \underline{Y}_j)$, and substitute in (1). This gives

$$\sigma(X;Y_1,...,Y_m) = \Sigma_j \, tr \, X'(I - P_j)X +$$

$$+ \Sigma_j \, tr \, (Y_j - \underline{Y}_j)'D_j(Y_j - \underline{Y}_j), \qquad (3)$$

where $D_j = G_j'G_j$ and $P_j = G_j\{G_j\}^+$. Thus

$$\sigma(X;*,...,*) = \Sigma_j \text{ tr } X'(I - P_j)X =$$
$$m(p - \text{tr } X'P_*X), \qquad (4)$$

with P_* equal to the average of the P_j.

For the interpretation it is convenient to remember that D_j is a diagonal matrix containing the marginals of the categories of variable j. P_j is an orthogonal projector, which projects on the space $\mathbb{L}_j$ spanned by the columns of G_j. The space $\mathbb{L}_j$ is the subspace of $\mathbb{R}^n$ consisting of all vectors which assign the same real number to objects scoring in the same category of variable j. Or, in somewhat different terminology, if x is any vector in $\mathbb{R}^n$, then P_jx replaces the elements of x by their category means $\underline{y}_j = \{G_j\}^+x$. Thus $x'P_jx = \underline{y}_j'D_j\underline{y}_j$ is the variance *between* categories, and $x'(I - P_j)x$ is the variance *within* categories.

The next step in the derivation of homogeneity analysis is to compute

$$\sigma(*;*,...,*) = \min \{ \sigma(X;*,...,*) \mid X \text{ normalized}\}. \qquad (5)$$

From (4) we obtain directly

$$\sigma(*;*,...,*) = mp \Sigma_s \{1 - \lambda_s(P_*)\}, \qquad (6)$$

where $\lambda_1(P_*) \geq ... \geq \lambda_p(P_*)$ are the p largest nontrivial eigenvalues of the average projector P_*. We use 'nontrivial' because P_* always has a largest trivial eigenvalue $\lambda_0(P_*) = 1$, corresponding with the trivial eigenvector u. All other eigenvectors can consequently be chosen such that $X'u = 0$.

It is of some interest that alternatively we could also minimize the loss function (1)

with the condition that X is free and the Y_j are normalized. This means that

$$\Sigma_j Y_j'D_j u = 0, \tag{7a}$$

$$\Sigma_j Y_j'D_j Y_j = I. \tag{7b}$$

It has been shown in detail by Gifi (1981), also compare De Leeuw (1984a), that this gives essentially the same solution as normalizing X and leaving the Y, the concatenation of the Y_j, free. Alternatively we can also normalize both X and Y, or we could require $mX'X + Y'DY = I$. Again this gives essentially the same solution. This basic result is, in a somewhat less general form, already due to Guttman (1941) in his pioneering paper on (one-dimensional) homogeneity analysis.

Reciprocal averaging

In the previous section we have shown that optimal quantifications for the objects can be found by computing the p dominant eigenvalues, with corresponding eigenvectors, of P_*. The corresponding optimal scores for the categories of variable j are then $Y_j = \{G_j\}^+X = \{D_j\}^+G_j'X$. In words: the optimal quantification of a category is the centroid of the optimal scores of the objects in that category. In many situations in which homogeneity analysis is applied, an iterative technique for computing the optimal scores and quantifications is very convenient. It is usually called *reciprocal averaging* , it has been around since the thirties, and we give the algorithm in Figure 2. It is exceedingly simple, and hardly needs any explanation. The optimal category quantifications are computed in each iteration step as the averages of the relevant current object scores, and the optimal object scores are the averages of the current optimal quantified variables. The only minor complication is that we have to normalize X in each

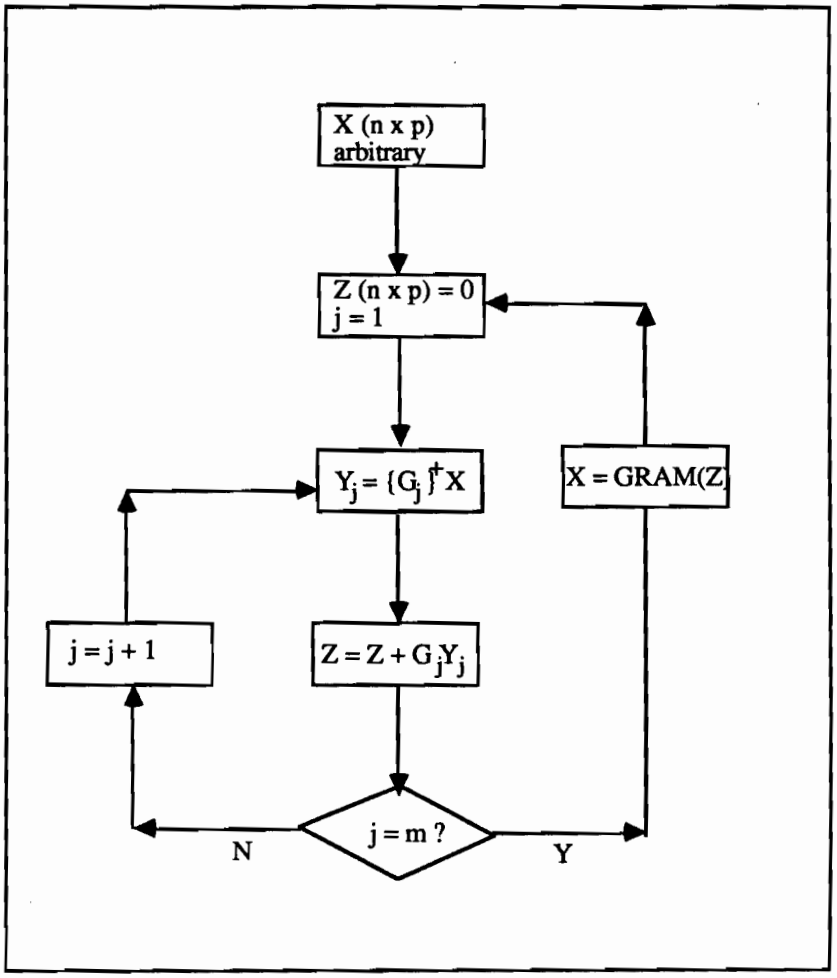


Figure 2: Reciprocal averaging algorithm

iteration, which is done by Gram-Schmidt orthogonalization. In the flow-diagram this reads X = GRAM(Z). Observe that the algorithm in Figure 2 has no stopping criterion, but there are several obvious places where we can test for convergence. The algorithm is implemented in the computer program HOMALS (homogeneity analysis by alternating least squares), compare Van der Geer (1985), and it has been widely applied.

Now let us now apply it to the car data. We find dominant eigenvalues of .81 and .53. The two-dimensional solution is plotted in Figure 3. We see that both cars and categories curve along a convex curve in two-space. This is a common type of representation in homogeneity analysis, which indicates that there is only one really dominant dimension (roughly the size of the car). For more information on 'horseshoes' we refer to Gifi (1981), Schriever (1985), Heiser (1985), Van Rijckevorsel (1986). Comparing Figure 1 and Figure 3 shows clearly how much neater the graph of the indicator supermatrix is now presented in $\mathbb{R}^2$. Of course the example is in no way typical. It is merely an illustration. Usually homogeneity analysis is applied to many more objects and many more variables.

Rank restrictions

After explaining the basic idea of homogeneity analysis we now discuss various types of restrictions that can be imposed. The k_j points corresponding with the categories of a variable can be located anywhere in $\mathbb{R}^p$, there is no restriction on their location relative to one another (except for $Y_j'D_ju = 0$). If the categories of a variable are ordered, however, we would like to see this order represented in the picture. There are various ways of ordering the points in the plane, but the most obvious one is to take a direction and let it define the order. In formulas this means that we want the category points of a variable to be on a straight line through the origin. Thus we require $Y_j = z_ja_j'$, where z_j is a k_j-vector with *single quantifications* and a_j is a p-vector with *weights* or *loadings* . These restrictions are called the *rank-one restrictions* . For identification purposes we

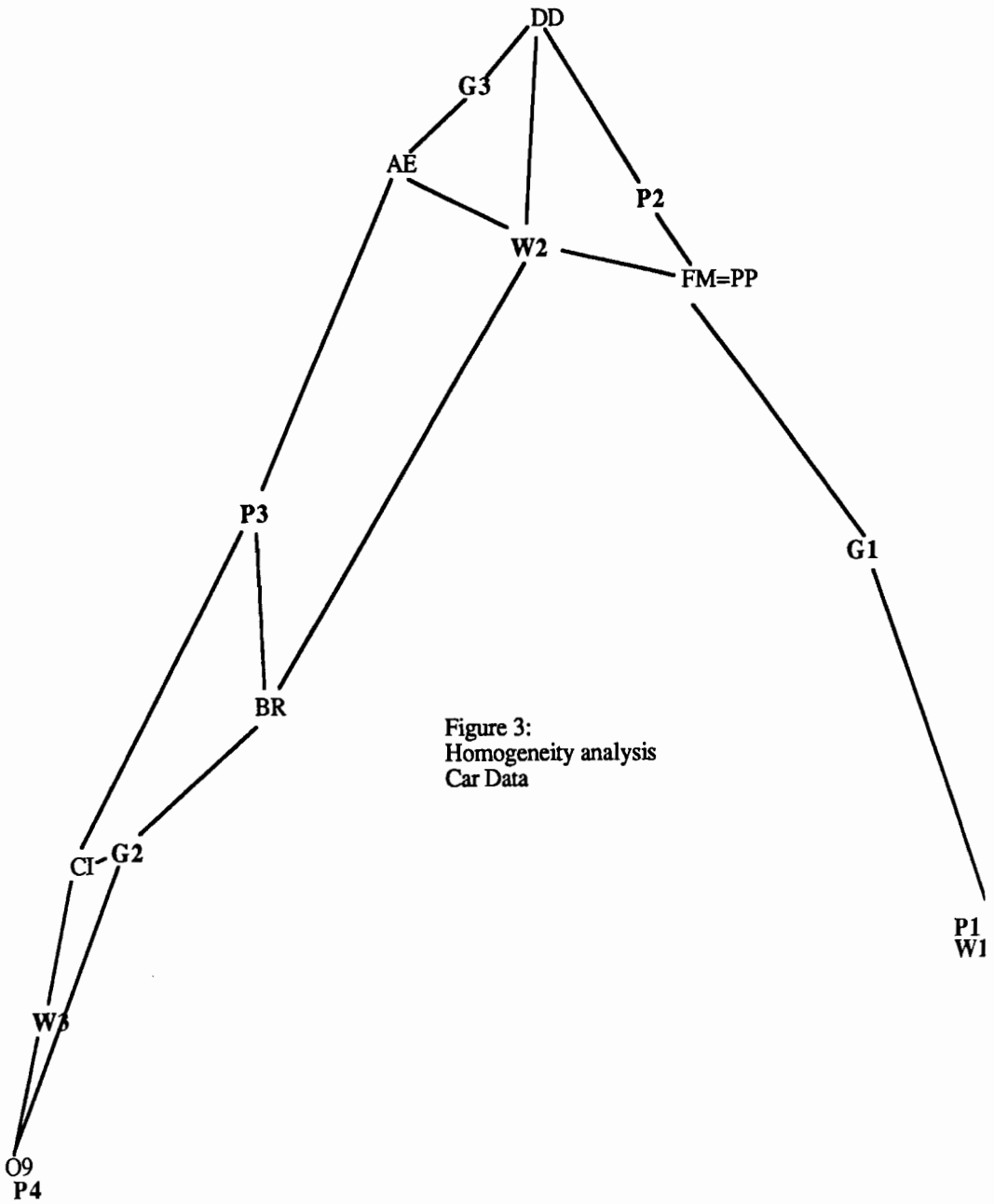


Figure 3:
Homogeneity analysis
Car Data

require $z_j'D_jz_j = 1$ and $u'D_jz_j = 0$ for all j.

The loss function, with restrictions, becomes

$$\sigma(X;Y_1,...,Y_m) = \sigma(X;z_1,...,z_m;a_1,...,a_m) =$$

$$= \Sigma_j \, SSQ(X - G_jz_ja_j') =$$

$$= mp - 2 \, tr \, X'QA + tr \, A'A =$$

$$= m(p - 1) + SSQ(Q - XA'). \qquad (8)$$

Here Q is the n x m matrix with columns $q_j = G_jz_j$, and A is the m x p matrix with the a_j as rows. Letting R = Q'Q we obtain

$$\sigma(*;z_1,...,z_m;*,...,*) =$$

$$= \min \{\sigma(X;z_1,...,z_m;a_1,...,a_m) \mid X;z_1,...,z_m\} =$$

$$= m(p - 1) + \Sigma \{\lambda_s(R) \mid s=p+1,...,m\}. \qquad (9)$$

This shows that homogeneity analysis with rank-one restrictions has a loss which is always at least m(p - 1). It is exactly equal to m(p - 1) if we can choose the z_j in such a way that the rank of the correlation matrix R is less than or equal to p. This means that for all j we must have $G_jz_j = Xa_j$. Geometrically this means that each variable defines a line $\{y \mid y = \tau a_j\}$ through the origin with direction cosines proportional to a_j. If we project object points x_i orthogonally on this line then objects in the same category must project into the same point of the line. This means that categories of a variable are

represented as parallel hyperplanes perpendicular to the line of the variable. We have a perfect solution if all object points are in the appropriate category planes, i.e. if the objects scoring in category one of variable j are in plane one of variable j, and so on.

It also follows from (9) that homogeneity with rank-one restrictions is a form of *nonlinear principal component analysis* (Kruskal and Shepard, 1974, Young, Takane, and De Leeuw, 1978, De Leeuw, 1982, Koyak, 1985). We must find category quantifications z_j in such a way that the sum of the p largest eigenvalues of the correlation matrix of the quantified variables is as large as possible. This becomes more clear if we combine the rank-one restrictions with *cone restrictions* , i.e. if we require in addition that $z_j \in \mathbb{K}_j$, where $\mathbb{K}_j$ is a convex cone in k_j-dimensional space. It follows that $z_j \in \mathbb{K}_j \cap \mathbb{S}_j$, with $\mathbb{S}_j$ the set of vectors that are normalized. If $\mathbb{K}_j$ is the ray of vectors proportional to a known vector, for instance the numerical values in Table 1, then z_j is fixed by the normalization requirements, and homogeneity analysis with single restrictions becomes ordinary principal component analysis. If $\mathbb{K}_j$ is the cone of monotone transformations, then we obtain nonmetric principal component analysis, and so on.

There is little which needs to be said about the algorithm. It has the same basic structure as the algorithm in Figure 2, with one major modification. After finding $\underline{Y}_j = \{G_j\}^+X$ we make steps to solve the minimization problem

$$\text{tr } (\underline{Y}_j - z_j a_j')'D_j(\underline{Y}_j - z_j a_j') \text{ min !} \qquad (10)$$

over a_j and z_j (possibly with restriction $z_j \in \mathbb{K}_j$). We know from (3) that this is the appropriate loss component to minimize. In order to minimize it, or at least decrease it, we apply alternating least squares again. For the current z_j we find the optimal a_j, which is $\underline{Y}_j'D_jz_j$, and then we find the optimal z_j for given a_j. This can be done by defining $\underline{z}_j = \underline{Y}_ja_j$, and by solving

$$(\underline{z}_j - z_j)'D_j(\underline{z}_j - z_j) \text{ min !} \qquad (11)$$

which is, in the most complicated case, a cone-projection problem. Improving z_j and a_j is done one or more times (the *inner iterations*). The final z_j and a_j define a new Y_j, which then is used in the homogeneity analysis. The corresponding computer program for homogeneity analysis, in which there can be rank-one and cone restrictions for some or all variables, is called PRINCALS (Gifi, 1985).

In De Leeuw (1984b) the idea of using rank and cone restrictions was explained with more generality. Any k_j x p matrix Y_j can be written, in many ways, in the form

$$Y_j = \Sigma_t z_{jt}a_{jt}', \qquad (12)$$

provided the number of components in this decomposition is large enough. Various interesting types of restrictions can be analyzed as special cases of (12). The vectors z_t can be given vectors with orthogonal polynomials for instance, or they can be free but restricted to be q in number (rank-q restrictions), or they can be monotone with the data and q in number, and so on. This defines many possible options, of which the more common ones are implemented in the programs HOMALS and PRINCALS.

We say that in ordinary homogeneity analysis (or multiple correspondence analysis, implemented in HOMALS) the variables are treated as *multiple nominal*. They are multiple, because each dimension has its own quantifications, and they are nominal because there are no restrictions on the location of the categories. In PRINCALS, with cone restrictions, we can choose between a multiple nominal treatment, and a *single nominal* , *single ordinal* , or *single numerical* treatment. Single refers to the fact that there is only one quantification for each variable, due to the rank-one restrictions. Single nominal means no further restrictions on z_j, single nominal means that they must be in the 'correct' order, and single numerical means they must be linear with the prior scores. The use of general rank-q restrictions makes it possible to define multiple ordinal and

multiple numerical as well, and shows that there are various possibilities between single with rank(Y_j) = 1 and multiple with rank(Y_j) $\geq$ min(p,k_j - 1).

We give a single numerical analysis of the car example in Figure 4a, and a single nominal analysis in Figure 4b. In both solutions we have plotted the object points x_i and the category quantifications $z_j a_j'$ (in bold type). We have also drawn the three lines through the origin representing the three variables, and for variable 'gas' we have drawn the category lines perpendicular to the direction defined by the variable. The solutions are very different from the multiple nominal solution in Figure 3, the horseshoe has disappeared (of course the horseshoe is incompatible with rank-one restrictions). The single numerical solution, which is an ordinary principal component analysis of Table 2, shows us a very high correlation between price and weight. It also shows that Ford Mustang and Pontiac Phoenix use less gas than expected on the basis of their price/weight, while Dodge Diplomat and AMC Eagle use more. The single nominal solution give a slightly different picture. The correlations between the three variables are now more equal. The program has used the additional freedom to cluster the cars in the four clusters {{FM,PP},{DD,AE},{CC,DC,PH},{CI,O9,BR}}. Such clustering is often seen in small examples, in which the clusters are located close to the category points, which are also clustered. The program clusters categories $\{\mathbf{W}_2,\mathbf{W}_3\}$ and $\{\mathbf{P}_2,\mathbf{P}_3,\mathbf{P}_4\}$ by giving them the same quantification. This creating of ties makes it possible to arrive at a more homogeneous representation than in the single numerical case.

Pseudo-indicators

Indicator matrices are practical if the number of categories is small relative to the number of objects. If the number of categories is equal to the number of objects the indicator matrix will be a square permutation matrix, and any transformation or quantification whatsoever is permissible. This obviously allows for too much freedom.

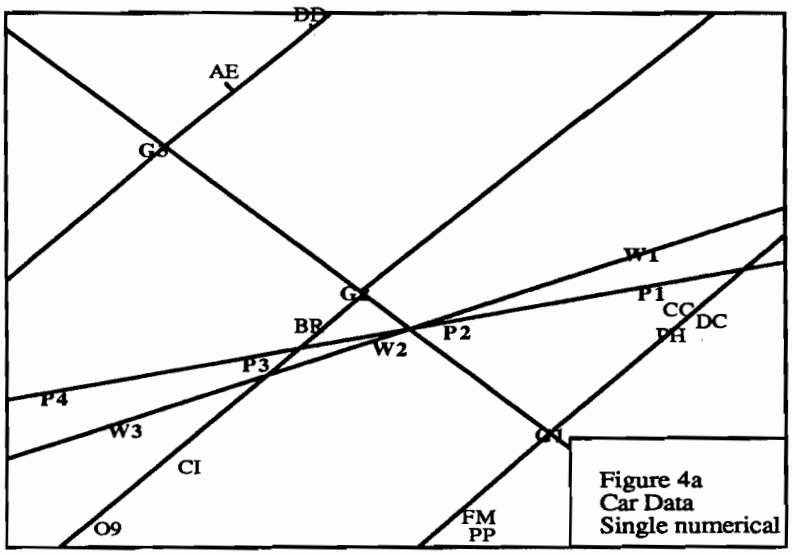


Figure 4a
Car Data
Single numerical

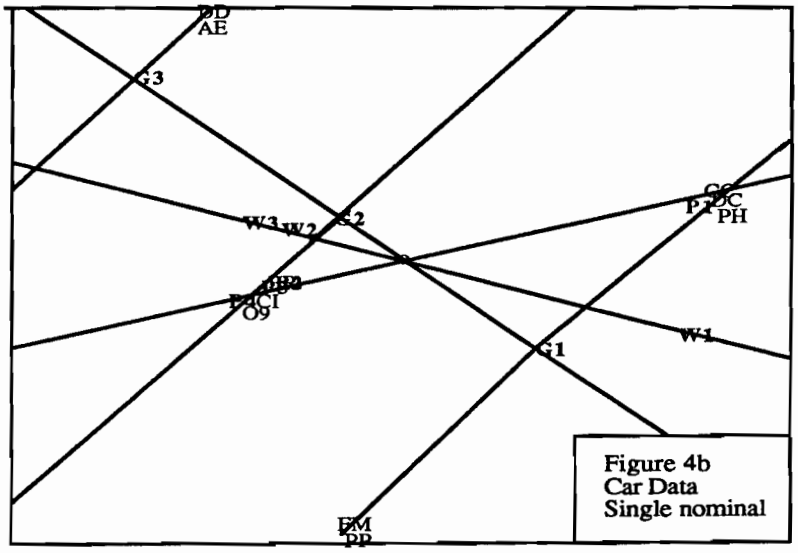


Figure 4b
Car Data
Single nominal

Variables with lots of categories occur in cases in which we are dealing with variables which are 'continuous'. The car data in Table 1 for example have targets with 10 different elements. This implies that $P_j = I$ for j=1,2,3, and thus $P_* = I$, and the homogeneity analysis solution is completely arbitrary.

There are three different ways out of this dilemma. The first one, which we have already discussed above, is to impose rank-one restrictions together with monotonicity restrictions. Rank-one restrictions by themselves are not enough, we do need the monotonicity here. This is the PRINCALS solution. If the number of categories is equal to the number of observations then the restriction that the objects scoring in the same category must be in the same hyperplane perpendicular to the direction defining the variable is no longer a restriction. The restriction is merely that the objects project in the correct order on the direction defining the variable. For the car data these constraints are not very restrictive. The first five cars have lower values on all three variables as the last five cars. Thus if we give the first five cars value -1 and the last five cars value +1 on all three variables, we have a monotonic transformation which gives the transformed data matrix rank exactly equal to one.

The second solution has been used by Breiman and Friedman (1985), and several of their students. It proceeds from a somewhat different, less geometrical, interpretation of homogeneity analysis. The purpose of the technique is formulated as quantifying or transforming the variables in such a way that they are as homogeneous as possible, in the sense that the sum of the p largest eigenvalues of their correlation matrix is maximized. Compare equation (9) and the discussion following it. Other criteria are sometimes also useful (De Leeuw, 1986). The ACE-method of Breiman and Friedman involves computing the optimal quantifications, and then *smoothing* these by running them through a linear smoother. This has the disadvantage that the usual linear smoothers are not projectors, and thus the smoothed optimal transformations are not optimal in terms of projection on a given subspace any more. The usual convergence theory for the alternating least squares methods consequently does not apply. Nevertheless using smoothers is a viable alternative, which has been shown to produce interesting results in practical situations.

In this paper we shall concentrate on a third alternative procedure, however. This is the use of *splines*, or, more generally, of *fuzzy coding* and *pseudo-indicators* . We introduce them in a purely geometrical way again, starting with the basic loss function

$$\sigma(X;Y_1,...,Y_m) = \Sigma_j \, SSQ(X - G_jY_j) \, . \qquad (13)$$

The difference now is that we do not assume that the basic data matrices G_j are indicator matrices, they can be *pseudo-indicators* . A pseudo-indicator matrix is a nonnegative matrix whose row-sums are equal to one. This means that in (13) we compute the squared distances between the object scores x_i and certain convex combinations of the rows of Y_j. Pseudo-indicators may occur because of various reasons. We discuss some of the more common ones.

The first one is that we start out with ordinary indicators G_j, but we require that some of the Y_j must be equal. This could happen, for instance, if some variables have the same categories. We consider the extreme case in which all Y_j are required to be equal to illustrate what will happen. From (13), using G_* for the average of the G_j and D_* for the average of the D_j,

$$\sigma(X;Y_1,...,Y_m) = \Sigma_j \, SSQ(X - G_jY) =$$

$$= m \, \{SSQ(X) - 2 \, tr \, X'G_*Y + tr \, Y'D_*Y\} =$$

$$= m \, SSQ(X - G_*Y) \; + \Sigma_j \, SSQ[(G_j - G_*)Y]. \qquad (14)$$

The matrix G_* is a pseudo-indicator. Homogeneity analysis in this case amounts to eigen-analysis of $G_*(D_*)^{-1}G_*'$. This procedure, and intermediate cases in which some

Y_j must be equal, occurs in the analysis of categorical time series data (Deville and Saporta, 1980; De Leeuw, Van der Heijden, and Kreft, 1985).

A second way in which pseudo-indicators can arise is by *fuzzy coding* . This concept has been studied in great detail by Martin (1980) and Van Rijckevorsel (1986). A simple example is the following. Suppose we have missing data, i.e. for some objects we do not know which category of a given variable they belong to. The corresponding row of the corresponding indicator matrix G_j can then be coded by making all elements equal to $1/k_j$. This is not the only possible way, let alone the best way, to deal with missing data in homogeneity analysis. In fact three other ways of incorporating them are discussed and compared by Meulman (1982). But it does give rise to a pseudo-indicator matrix, which is what interests us here. More generally we can use the rows of the G_j to code other forms of uncertainty. Missing data can also be coded, for instance, by making the rows equal to the marginals of the non-missing objects. Ordinal data can be coded with uncertainty by using .5 for the category the object scores in, and by using .25 for the two neighbouring categories.

A somewhat more systematic procedure, which is very useful for continuous data, is to use B-spline bases for the indicator matrices. We do not intend to discuss B-splines in detail. For this we refer, for example, to De Boor (1978). Applications to homogeneity analysis are discussed in more detail in De Leeuw, Van Rijckevorsel, and Van der Wouden (1981), and in Van Rijckevorsel and Van Kooten (1985). For our purposes here it suffices to say that a B-spline basis for the space of all polynomial splines of a given order and knot-sequence defines a pseudo-indicator. In the simplest case the splines, of order one, are step-functions, and the basis defines a true indicator. In the order two case the splines are continuous piecewise linear functions, and the spline basis are the *hat functions*, which define a pseudo-indicator with at most two consecutive nonzero elements in each row. In Table 4 we have a spline basis of order two for the car data. Each column is a hat function of the corresponding values in Table 1. This means that there are three consecutive knots ($knot_1$,$knot_2$,$knot_3$) such that the function is zero for all values less than or equal to $knot_1$ and larger than or equal to

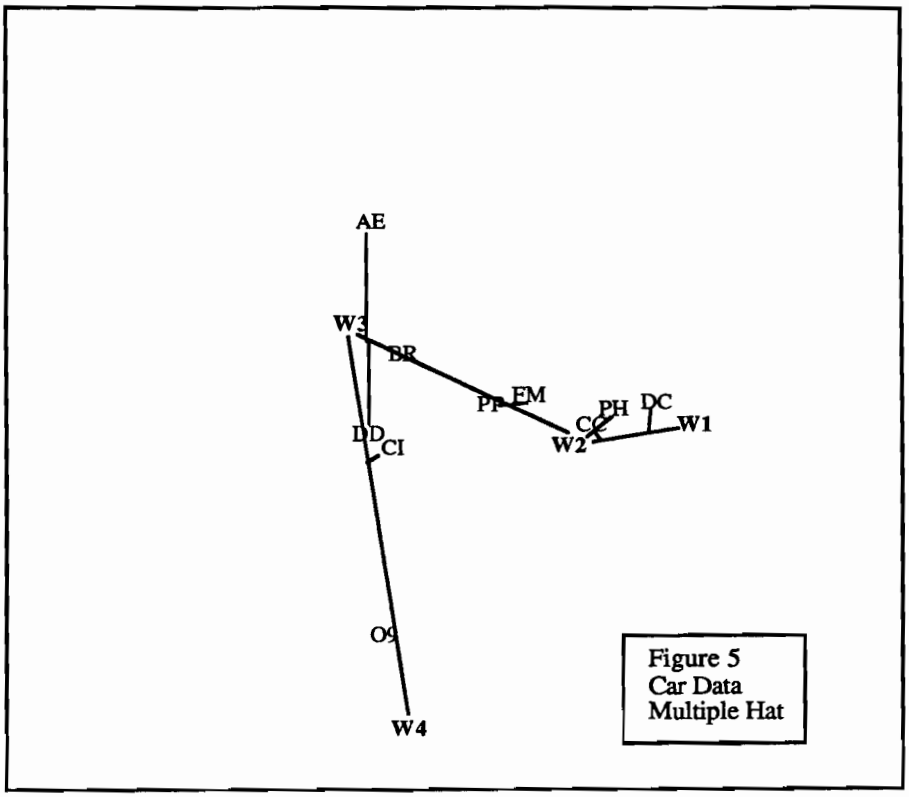
$knot_3$. In the interval ($knot_1$,$knot_2$) it increases from zero to its maximum, in ($knot_2$,$knot_3$) it decreases from the maximum to zero again.

What does the use of splines, or fuzzy coding in general, mean in terms of the geometry of homogeneity analysis ? This is illustrated in Figure 5, in which we have the multiple hat solution for the car data. Only the variable weight is drawn in. We still make a joint picture of object points and category points (or knot points), but the lines are no longer drawn from the object points to the category points, but from the object points to fixed places on the lines connecting two consecutive category points. If the order is three, then objects points are connected with points in the triangle spanned by three consecutive category points, and so on. If we code missing data by assigning $1/k_j$ to all categories of a variable, then 'missing' object points must be connected with the centroid of the category configuration. By incorporating the continuous information we move away from the graph-interpretation of homogeneity analysis, which is of course only natural becaue this is inherently discrete.

B-splines, and fuzzy coding in general, can easily be combined with rank-restrictions on the quantifications. In the case of rank-one restrictions, for example, the category points must be on a line through the origin, and the objects must be connected with points on this line that are the appropriate convex combinations of the category points. In case of order two this means that objects must project into fixed places in disjoint intervals on the line defining the variable, if the order is larger than two the intervals may overlap. This is explained in more detail in De Leeuw (1985b) and Van Rijckevorsel (1986). De Leeuw also discusses the case in which the pseudo-indicators can be adjusted as well. Only the zero-elements are fixed, the nonzero elements are variables over which we optimize under the restriction that they are nonnegative and add up to one. In the case of rank-one restrictions again, this means that the object must project into disjoint intervals of the line defining the variable, but they can project anywhere in these intervals. This generalizes the 'primary approach to ties' of Kruskal and Shepard (1974), and the 'continuous ordinal' scaling of De Leeuw, Young, and Takane (1976) or Young (1981). We shall not illustrate these further developments here, because they obviously should not be applied to small examples such as our car data.

	Price	Gas	Weight
Chevrolet Chevette	.70 .30 .00 .00 .00	.05 .95 .00 .00	.15 .85 .00 .00
Dodge Colt	.65 .35 .00 .00 .00	.95 .05 .00 .00	.60 .40 .00 .00
Plymouth Horizon	.35 .65 .00 .00 .00	.75 .25 .00 .00	.05 .95 .00 .00
Fort Mustang	.00 .80 .20 .00 .00	.15 .85 .00 .00	.00 .67 .33 .00
Pontiac Phoenix	.00 .47 .53 .00 .00	.05 .95 .00 .00	.00 .65 .35 .00
Dodge Diplomat	.00 .20 .80 .00 .00	.00 .00 .90 .10	.00 .08 .92 .00
Chevrolet Impala	.00 .00 .95 .05 .00	.00 .83 .17 .00	.00 .00 .70 .30
Buick Regal	.00 .00 .75 .25 .00	.00 .73 .27 .00	.00 .17 .83 .00
AMC Eagle	.00 .00 .65 .35 .00	.00 .00 .15 .85	.00 .05 .95 .00
Oldsmobile 98	.00 .00 .00 .45 .55	.00 .43 .57 .00	.00 .00 .23 .77

Table 4: Car Data, pseudo-indicators, hat functions.



Additivity restrictions

The title of the paper suggests more than we have offered so far. Of the classical linear multivariate analysis techniques we have only encountered principal component analysis so far as an important special case of homogeneity analysis (all variables single numerical). But what about canonical correlation analysis, discriminant analysis, multiple regression, and multivariate analysis of variance ? In this chapter we show, briefly, how these other techniques can be displayed as versions of homogeneity analysis with restrictions as well. We merely discuss this for canonical correlation analysis, because it is well known from the linear theory that the other techniques we have mentioned are special cases of this (Gittins, 1985). Nonlinear generalizations of canonical correlation analysis are discussed in detail by Van der Burg and De Leeuw (1983), and in even greater generality by Van der Burg, De Leeuw, and Verdegaal (1985).

Let us first specialize homogeneity analysis to the case in which there are just two variables (a bivariable). Homogeneity analysis then becomes identical to *correspondence analysis* (Benzécri et al., 1973; Benzécri et al., 1980; Lebart, Morineaux, and Warwick, 1985; Greenacre, 1985). If both variables are single numerical we merely compute a correlation coefficient, if they are single nominal or ordinal we compute the maximum correlation coefficient (Lancaster, 1969, has many references on the maximum correlation problem). In canonical correlation analysis we do not have two variables, but two sets of variables. But, as the literature on maximal correlation shows, the difference between the two situations is not large.

With our formalism this can be shown quite easily. Suppose $\Phi = \{\phi_j \mid j \in J\}$ and $\Psi = \{\psi_l \mid l \in L\}$ are two multivariables with a common domain. Variable ϕ_j maps Ω into Γ_j, and ψ_j maps Ω into Ξ_j. We can now define two new *product variables* $\phi_J: \Omega \rightarrow \Pi_j \Gamma_j$ and $\psi_L: \Omega \rightarrow \Pi_l \Xi_l$. Variables ϕ_J and ψ_L have $\Pi_j k_j$ and $\Pi_l k_l$ categories. Applying homogeneity analysis to these two product variables is a form of nonlinear canonical analysis, and does not introduce anything new. Of course in practical situations the number of

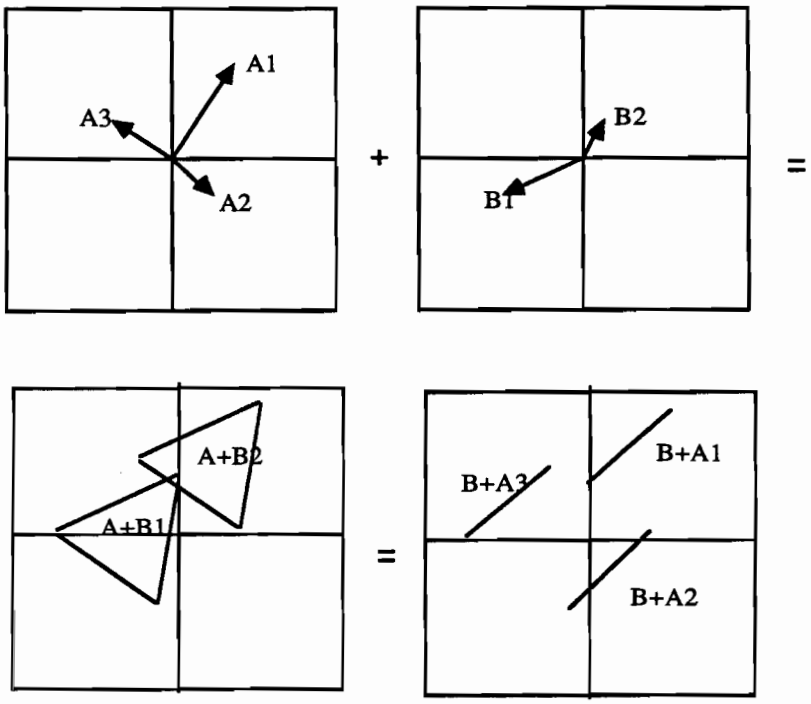


Figure 6: Additivity restrictions, multiple

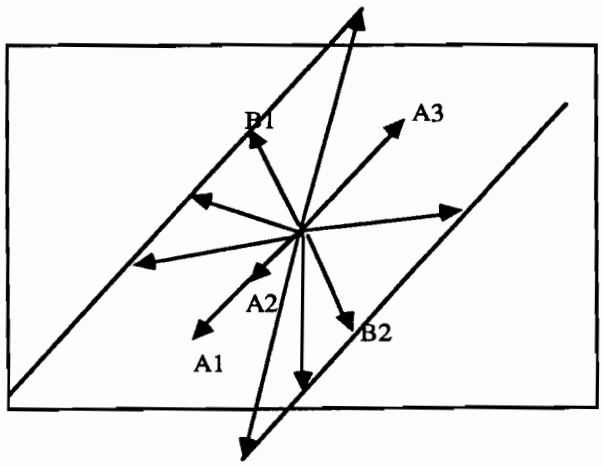


Figure 7: Additivity restrictions, single.

categories of the product variables can easily become much too large. In order to remedy this we can impose *additivity restrictions* . These are illustrated in Figures 6 and 7, for the multiple and single case respectively. In Figure 6 we start with two variables A and B with three and two categories. The product variable A x B has six categories, and the additivity restrictions have as a result that the quantification of (A1,B1) is the vector sum of the quantifications A1 and B1. In stead of 2 x (6 - 1) = 10 parameters we have only 2 x [(2 - 1) + (3 - 1)] = 6 parameters for the quantifications. The category quantifications are on the B1- and B2-translation of the (A1,A2,A3)-triangle, or equivalently on the A1-, A2, or A3-translation of the (B1,B2)-line. This shows the geometrical interpretation of additivity restrictions in the multiple case. Figure 7 shows the effect of additivity restrictions for two variables with rank-one restrictions. In this case the category quantifications are on parallel hyperplanes.

The use of additivity restrictions means that we deal with sets of variables in a somewhat roundabout way. First we code them interactively, by using product variables. Then we apply additivity restrictions to the product variables, with the consequence that the sets are not coded interactively any more, but additively. The loss function for this form of homogeneity analysis is

$$\sigma(X;Y_1,\ldots,Y_m) = \Sigma_t \, SSQ(X - \Sigma_{j\in J(t)} G_j Y_j), \qquad (15)$$

where t now indexes *sets of variables* , and J(t) shows which variables are in set t. Again of course (15) can be combined with rank restrictions and cone restrictions. This produces a very general nonlinear multivariate analysis technique, with corresponding computer program OVERALS (Verdegaal, 1986). Observe that ordinary homogeneity analysis (and consequently also principal component analysis) is the special case in which each set only contains a single variable.

Conclusion

We shall not continue to illustrate the forms of canonical analysis and regression analysis in detail with examples. The basic idea is probably clear. It is possible to develop most of the existing linear MVA techniques within homogeneity analysis by choosing suitable restrictions on the quantifications. In as far as homogeneity analysis is a form of MDS, we have shown that linear MVA and its generalizations can be imbedded in the distance framework, using the concept of homogeneity and using the joined space representation of objects and variables (or categories). For the many details of these developments, the available software, applications in many of the sciences, and for the statistical aspects of the techniques we refer to the book by Gifi (1981), and the other literature we have mentioned. In this paper we could only show some of the more important principles of the MDS approach to MDA.

References and bibliography

Benzécri, J. P. et al. (1973). **L'Analyse des Données.** (2 vols). Paris: Dunod.

Benzécri, J.P. et al. (1980). **Pratique de l'Analyse des Données.** (3 vols). Paris: Dunod.

Breiman, L, Friedman, J.H. (1985). Estimating Optimal Transformations for Multiple Regression and Correlation. **J. Am. Statist. Assoc.**, 80., 580-619.

Carroll, J.D., Arabie, P. (1980). Multidimensional Scaling. **Ann. Rev. Psychol.**, 31, 607-649.

De Boor, C. (1978). **A Practical Guide to Splines.** Berlin: Springer.

De Leeuw, J. (1973). **Canonical Analysis of Categorical Data.** Unpublished dissertation. Reissued DSWO-Press, Leiden, 1984.

De Leeuw, J. (1982). Nonlinear Principal Component Analysis. In H. Caussinus et al. (eds), **COMPSTAT 82.** Wien: Physika Verlag.

De Leeuw, J. (1984a). The Gifi System of Nonlinear Multivariate Analysis. In E. Diday et al. (eds), **Data Analysis and Informatics IV.** Amsterdam: North Holland Publishing Co.

De Leeuw, J. (1984b). **Beyond Homogeneity Analysis.** Report RR-84-08, Department of Data Theory, University of Leiden.

De Leeuw, J. (1986). **Multivariate Analysis with Optimal Scaling.** Report RR-86-01, Department of Data Theory, University of Leiden.

De Leeuw, J., Heiser, W.J. (1980). Multidimensional Scaling with Restrictions on the Configuration. In P.R. Krishnaiah (ed.), **Multivariate Analysis V.** Amsterdam: North Holland Publishing Co.

De Leeuw, J., Heiser, W.J. (1982). Theory of Multidimensional Scaling. In P.R. Krishnaiah and L. Kanal (eds.), **Handbook of Statistics II.** Amsterdam: North Holland Publishing Co.

De Leeuw, J., Van der Heijden, P., Kreft, I. (1984). Homogeneity Analysis of Event-history data. **Methods of Operations Research,50,** 299-316.

De Leeuw, J., Van Rijckevorsel, J, Van der Wouden, H. (1981). Nonlinear Principal Component Analysis using B-splines. **Methods of Operations**

Research, 23, 211-234.

De Leeuw, J., Young, F.W., Takane, Y. (1976). Additive Structure in Qualitative Data. **Psychometrika**, 41, 471-503.

Deville, J.C., Saporta, G. (1980). Analyse Harmonique Qualitative. In E. Diday et al. (eds.), **Data Analysis and Informatics.** Amsterdam: North Holland Publishing Company.

Gifi, A. (1981). **Nonlinear Multivariate Analysis.** Department of Data Theory FSW, University of Leiden. To be reissued by DSWO-Press, 1987.

Gifi, A. (1985). **PRINCALS.** User's Guide UG-85-02. Department of Data Theory FSW, University of Leiden.

Gittins, R. (1985). **Canonical Analysis. A Review with Applications in Ecology.** Berlin: Springer.

Greenacre, M.J. (1984). **Theory and Applications of Correspondence Analysis.** New York: Academic Press.

Guttman, L. (1941). The Quantification of a Class of Attributes: a Theory and Method of Scale Construction. In P. Horst (ed.), **The Prediction of Personal Adjustment.** New York: Social Science Research Council.

Guttman, L. (1977). What is not what in statistics. **The Statistician**, 26, 81-107.

Heiser, W.J. (1981). **Unfolding Analysis of Proximity Data.** Department of Data Theory, University of Leiden.

Heiser, W.J. (1985). Undesired Nonlinearities in Nonlinear Multivariate Analysis. In E. Diday et al. (eds.), **Data Analysis and Informatics IV**. Amsterdam: North Holland Publishing Company.

Koyak, R. (1985). **Nonlinear Dimensionality Reduction.** Unpublished Ph.D. Thesis. Department of Statistics, University of California, Berkeley.

Kruskal, J.B., Shepard, R.N. (1974). A Nonmetric variety of Linear Factor Analysis. **Psychometrika**, 39, 123-157.

Lancaster, H.O. (1969). **The Chi-square Distribution.** New York: Wiley.

Lebart, L., Morineau, A., Warwick, K.M. (1984). **Multivariate Descriptive Statistical Analysis.** New York: Wiley.

Martin, J.F. (1980). **Le Codage Floue et ses Applications.** Doctoral Dissertation,

Department of Statistics, Université de Pau.

Meulman, J. (1982). **Homogeneity Analysis of Incomplete Data**. Leiden: DSWO-Press.

Meulman, J. (1986). **A Distance Approach to Nonlinear Multivariate Analysis.** Doctoral Dissertation, Department of Data Theory, University of Leiden.

Nishisato, S. (1980). **The Analysis of Categorical Data. Dual Scaling and its Application**. Toronto: University of Toronto Press.

Schriever, B.F. (1985). **Order Dependence.** Amsterdam: Mathematical Centre.

Tukey, J. (1962). The Future of Data Analysis. **Ann. Math. Statist.**, 33, 1-67.

Van der Burg, E., De Leeuw, J. (1983). Nonlinear Canonical Correlation. **British Journal of Mathematical and Statistical Psychology**, 36, 54-80.

Van der Burg, E., De Leeuw, J., Verdegaal, R. (1984). **Non-linear Canonical Correlation with M Sets of Variables.** Report RR-84-12, Department of Data Theory, University of Leiden.

Van de Geer, J.P. (1985). **HOMALS.** User's Guide UG-85-01. Department of Data theory FSW, University of Leiden.

Van Rijckevorsel, J. (1986). About Horseshoes in Multiple Correspondence Analysis. In W. Gaul & M. Schader (eds.), **Classication as a Tool of Research.** Amsterdam: North Holland Publishing Company.

Van Rijckevorsel, J., Van Kooten, G. (1985). Smooth PCA of Economic Data. **Computational Statistics Quarterly**, 2, 143-172.

Verdegaal, R. (1986). **OVERALS.** User's Guide UG-86-01. Department of Data Theory FSW, University of Leiden

Winsberg, S., Ramsay, J.O. (1983). Monotone Spline Transformations for Data Reduction. **Psychometrika**, 48, 575-595.

Young, F.W. (1981). Quantitative Analysis of Qualitative Data. **Psychometrika**, 46, 347-388.

Young, F.W. (1984). Scaling. **Ann. Rev. Psychol.**, 35, 55-81.

Young, F.W., De Leeuw, J., Takane, Y. (1976). Regression with Qualitative and Quantitative Variables. **Psychometrika**, 41, 505-529.