DATA ANALYSIS AND INFORMATICS, III
E. Diday et al. (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

THE GIFI SYSTEM OF NONLINEAR MULTIVARIATE ANALYSIS

Jan de Leeuw

Department of Data Theory FSW/RUL
Leiden University
The Netherlands

There are many different ways in which multivariate data analysis techniques can be organized in a system. This paper discusses the organization chosen by Albert Gifi in a recent series of publications. The Gifi-system uses multiple correspondence analysis as the fundamental multivariate analysis technique.

INTRODUCTION

In Gifi (1981a) a system of nonlinear multivariate analysis methods is presented. It is based on earlier work of Guttman (1941), Burt (1950), Hayashi (1956), De Leeuw (1973), Benzécri (1973), Nishisato (1980). The particular approach chosen by Gifi is new, however. An attempt is made to build the system around *homogeneity analysis*, a technique which is also known as *multiple correspondence analysis*. The system tries to integrate ideas from multivariate analysis with ideas from multidimensional scaling. The main components of the system are a general least squares loss function, the use of additive restrictions to define sets of variables, the use of rank restrictions to define single quantifications, and the alternating least squares principle for algorithm construction. There is some overlap with the ALSOS-system of Young, De Leeuw, and Takane (1980) or Young (1981), but in ALSOS the emphasis is almost completely on the algorithms. In this paper we present the basic components of the Gifi-system, without going into too much technical detail. This means that we do not try to attain maximal generality or coverage. It also means that we do not explain algorithms and computer programs in any detail.

INDICATOR MATRICES AND QUANTIFICATION

Multivariate analysis deals with variables. Variables are functions defined on a given set of objects. In this paper we assume that there is only a finite number of objects, and that the variables assume only a finite number of values. Moreover we only have a finite number of variables. To be more specific: there are n objects, m variables, and each variable can assume at most k different values. At first sight it may seem that these finiteness conditions imply a considerable loss of generality. From a strictly empiristic point of view this is not the case, however. Infinity is always an idealization, it can never be realized in actual measurement.

Each variable defines an *indicator matrix*. The indicator matrix G_j corresponding with variable j is the n x k matrix in which row i indicates the category of variable j that object i is in. Thus row i has all elements, except one, equal to zero. The remaining element is equal to one. It follows that $D_j = G_j'G_j$ is diagonal, while $G_ju = u$ (we use u for a vector with all elements equal to one, its number of elements is clear from the context). The diagonal matrix D_j has on its diagonal the *marginals* of variable j. If d_j is the vector of marginals, then $d_j = G_j'u = D_ju$. For two different variables j and ℓ the matrix $C_{j\ell} = G_j'G_\ell$ contains the *bimarginals* (the cross table) of the two variables. Again $C_{j\ell}u = d_j$, i.e. the row and column sums of the bimarginals are the univariate marginals. The (mk) x (mk) supermatrix C, which has the $C_{j\ell}$ as its submatrices, is called the *tableau de Burt* in the French literature on correspondence analysis. The n x (mk) supermatrix of the form $G = (G_1|\ldots|G_m)$ is the *tableau sous forme disjonctif complète.*

A (p-dimensional) *quantification* of a set is a mapping of the set into $\mathbb{R}^p$. The quantifications of the set of objects (also called *object scores*) are collected in n x p matrices X, quantifications of the categories of variable j in the k x p matrices Y_j. The dimensionality p is chosen by the investigator. The nonlinear multivariate analysis problem is to choose *optimum quantifications*. Thus the first important step in the Gifi-system is to define optimality in a satisfactory and convenient way.

HOMOGENEITY ANALYSIS

In order to measure optimality we first need a definition of perfect fit. We actually give three closely related definitions. Object scores X and category quantifications Y_j are perfectly *consistent* if $X = G_1Y_1 = \ldots = G_mY_m$. Object scores X are perfectly *discriminating* if there exist category quantifications $Y_1,\ldots,Y_m$ such that $X = G_1Y_1 = \ldots = G_mY_m$. And category quantifications $Y_1,\ldots,Y_m$ are perfectly *homogeneous* if there exist object scores X such that $X = G_1Y_1 = \ldots = G_mY_m$. The relationship of these three definitions is clear. If object scores and category quantifications are perfectlyconsistent, then the object scores are perfectly discriminating and the category quantifications are perfectly homogeneous. Conversely if the object scores are perfectly discriminating, then we can find corresponding *induced* category quantifications which are perfectly homogeneous, and together with the object scores perfectly consistent. In the same way perfectly homogeneous category quantifications induce perfectly discriminating object scores, and together they form a perfectly consistent system again.

The algebra corresponding with our definitions is very simple. Let us first start with *direct* object scores X. If they are perfectly discriminating, then the induced category quantifcations $Y_j = D_j^+G_j'X$ are perfectly homogeneous. Superscript + is used for the Moore-Penrose inverse, which we need because some of the categories may be empty. Thus X is perfectly discriminating if $X = P_1X = \ldots = P_mX$, with $P_j = G_jD_j^+G_j'$. Or equivalently if $X'X = X'P_*X$, with P_* the average of the P_j. Observe that the P_j are symmetric idempotents, with rank equal to the number of nonempty categories of variable j. The matrix $X'P_jX$ is the between-category dispersion of variable j, or the sum of squares of all between category distances of the object scores. Of course X'X is the total dispersion, or the sum of squares of all distances. These facts link our definitions directly with the ideas of discriminant analysis, and through the use of distance with multidimensional scaling. We see that object scores are perfectly discriminating if objects in the same category of a variable have the same score, and if this is true for all variables.

Dually we can also start with direct category quantifications. If they are perfectly homogeneous, then any G_jY_j can be used to define perfectly discriminating object scores. Clearly $Y_1,\ldots,Y_m$ is perfectly homogeneous if $G_1Y_1 = \ldots = G_mY_m$. This can also be written as the equation Y'CY = mY'DY, where the Y_j are collected in the (mk) x p supermatrix Y and the D_j in the (mk) x (mk) diagonal supermatrix D. Category quantifications are perfectly homogeneous if categories which contain the same object get the same quantification. Perfect consistency shows that the three definitions of perfect fit amount to the same thing. This is the basic duality of homogeneity analysis, which was already explored by Guttman (1941).

It is not realistic to expect perfect fit in real data. This means that we need a definition of *loss*, i.e. of departure from perfect fit. *Loss of consistency* is defined as

$$\sigma(X;Y_1,\ldots,Y_m) = \frac{1}{m}\sum_{j=1}^{m} \mathrm{tr}\,(X - G_jY_j)'(X - G_jY_j). \tag{1}$$

In conformity with our earlier treatment we also define *loss of discrimination*, which is $\sigma(X;*,\ldots,*)$, the minimum of loss of consistency over category quantifications. Thus, substituting the induced $Y_j = D_j^+G_j'X$,

$$\sigma(X;*,\ldots,*) = \mathrm{tr}\ X'(I - P_*)X. \tag{2}$$

Loss of homogeneity is defined as $\sigma(*;Y_1,\ldots,Y_m)$, the minimum of $\sigma(X;Y_1,\ldots,Y_m)$ over X. By substituting

$$X = \frac{1}{m}\sum_{j=1}^{m} G_jY_j \tag{3}$$

for the induced object scores we find

$$\sigma(*;Y_1,\ldots,Y_m) = m^{-2}\ \mathrm{tr}\ Y'(C - mD)Y. \tag{4}$$

Given the three loss functions it is now easy to define *homogeneity analysis*, which is the technique that minimizes either one of these loss functions. Another familiar duality result, again due in all essentials to Guttman (1941), says that it does not matter which one of the three we minimize, provided we choose appropriate *normalizations*. It is clear that we must impose some sort of normalization, because X = 0 with Y = 0 is always trivially perfectly consistent. We also emphasize that what we call homogeneity analysis is known in the French literature as *analyse (factorielle) des correspondances multiples*. In the Benzécri-system homogeneity analysis is correspondence analysis on the *tableau sous forme disjonctif complète*. In the Gifi-system correspondence analysis is homogeneity analysis with only two variables. Thus the starting points of the two systems are different.

We now proceed with the minimization. Call object scores X *normalized* if X'X = I and u'X = 0, and call category quantifications Y *normalized* if Y'DY = mI and $u'D_jY_j = 0$ for all j. Minimizing loss of consistency $\sigma(X;Y)$ over all X and over all normalized Y is trivially equivalent to minimizing loss of homogeneity $\sigma(*;Y)$ over all normalized Y. Minimizing loss of consistency $\sigma(X;Y)$ over all normalized X and over all Y is equivalent to minimizing loss of discrimination over all normalized X. The first of these two problems (X free, Y normalized) amounts to finding the eigenvectors corresponding with the largest eigenvalues of $Cy = m\lambda Dy$, the second problem (X normalized, Y free) to solving the eigenproblem $P_*x = \mu x$. By using the singular value decomposition of the matrix $m^{-\frac{1}{2}}GD^{-\frac{1}{2}}$ we can show that both problems have the same eigenvalues, and that the eigenvectors we look for are the left and right singular vectors of $m^{-\frac{1}{2}}GD^{-\frac{1}{2}}$. Moreover the singular value decomposition also solves the problem of minimizing loss of consistency $\sigma(X;Y)$ over all normalized X and over all normalized Y. Observe that computing this singular value decomposition effectively performs a correspondence analysis on the complete disjoint table G.

We have shown that is does not matter if we normalize X or Y or both X and Y. More precisely, it only matters up to scale factors. If we normalize X, then the induced quatifications $Y_j = D_j^+G_j'X$ satisfy $Y_j'D_jY_j = X'P_jX$, and thus $Y'DY = mX'P_*X = m\Lambda$. If we normalize the category quantifications Y then the induced X, given by (3), satisfies $X'X = m^{-2}\ Y'CY = m^{-1}\ Y'DY\Lambda = \Lambda$. If we normalize X the induced quantification of a category is the center of gravity of the scores of the objects in the category, if we normalize Y then the induced score of the object is the centroid of the quantifications of the categories the object is in. These are the two *principes barycentriques* of Benzécri (1973), which are of cardinal importance for the interpretation and graphical presentation of the solutions of homogeneity analysis. Details on plots and interpretations can be found in Gifi (1981a, 1981b). We merely remark that in the Gifi-system the convention is used to normalize X, which means that the Y_j are found by the *principe barycentrique*. Because of this convention we refer to this procedure as the *first centroid principle*. Unless otherwise indicated we shall always use this convention, but we must remember that it is completely arbitrary and that we can switch to the *second centroid principle* without changing anything essential.

Some useful auxilary statistics in homogeneity analysis are the *discrimination matrix* and the *loading matrix*. They are defined for each variable separately. The discrimination matrix for variable j is $\Delta_j = Y_j'D_jY_j$. This is the dispersion matrix of the induced scores for variable j. We have already seen that also $\Delta_j = X'P_jX$, and thus Δ_j is also the between category dispersion for variable j. Discrimination matrices add up to $m\Lambda$. The diagonal matrix $\Omega_j = \mathrm{diag}(\Delta_j)$ is the matrix of *discrimination measures*. It is shown in De Leeuw (1983a) that if variable j is independent of the other variables, then the discrimination measures can be scaled in such a way that they have a chi-squared distribution. The *loading matrix* for variable j consists of the correlations between the object scores X and the object scores induced by variable j, which is G_jY_j. Thus the loading matrix is $\Gamma_j = X'G_jY_j\Omega_j^{-\frac{1}{2}} = X'P_jX\Omega_j^{-\frac{1}{2}} = \Delta_j\Omega_j^{-\frac{1}{2}}$.

There are some other important areas which we mention only briefly. The first one is computation. In Gifi's HOMALS program (Gifi 1981a, 1981b) loss of consistency is minimized by *alternating least squares*, in this context also known as *reciprocal averaging*. We start with normalized scores X, compute induced category quantifications by the first centroid principle, compute new object scores by the second centroid principle, and so on. We have to normalize in some way or another along the way. The program HOMALS normalizes X every couple of iterations by modified Gram-Schmidt. Many other choices are possible, however. The second area which we do not discuss is *missing data*. The details are in Gifi (1981a), and even more completely in Meulman (1982). The relationships of homogeneity analysis with multidimensional scaling and unfolding are discussed extensively in Heiser (1981). Statistical stability analyses of homogeneity analysis, using both the delta-method and the bootstrap-method, are again in Gifi (1981a).

RANK RESTRICTIONS ON CATEGORY QUANTIFICATIONS

Homogeneity analysis uses only the purely nominal information provided by the data. We only use the fact that some objects are in the first category of a variable, some are in the second category, and so on. Any prior information we may have on the categories and theit relationships is not used in the analysis. The numbering of the categories is just a labelling, a different labelling leads to the same results. In this section we discuss the method proposed by Gifi (1981a), which does incorporate prior information into homogeneity analysis. This has the effect that the technique becomes more similar to classical linear multivariate analysis, notably to principal component analysis.

The prior information we want to incorporate is that a variable can be *ordinal* or *numerical* in stead of merely *nominal*. In the ordinal case the range of the variable is interpreted as an ordered set, in the numerical case it is interpreted as a subset of the reals. This interpretation can be incorporated by imposing restrictions on the category quantifications. One such system of restrictions has been discussed by Young, De Leeuw, and Takane (1980), compare also Young (1981). Gifi (1981a) uses a different system of restrictions. Before we discuss its main components we emphasize that it is nonsense to say that a variable *is* numerical, ordinal, or nominal, at least if this means that the measurement level is some intrinsic property that belongs to the variable. It is not. The measurement level is a *model*, more concretely a set of restrictions on the quantifications that we may or may not impose.

In the Gifi-system quantifications can be either *multiple* or *single*, and both multiple and single quantifications can be either nominal, ordinal, or numerical. The category quantifications of homogeneity analysis discussed above are *multiple nominal*. They are nominal because no prior ordinal or numerical information is used in the analysis, and they are multiple because each dimension has a separate quantification. There are p dimensions, Y_j is a matrix of order k x p, each column of Y_j defines a single quantification of variable j. The columns of Y_j are not related in any particular way, we only require that $Y'DY = m\Lambda$, i.e. the category quantifications must be orthogonal 'on the average'. It is by now easy to see how *multiple ordinal* variables can be defined. The columns of Y_j must be in the appropriate order, which means that they must be monotonic with the prior order defined over the categories. If we combine this with the first centroid principle, then we require that the projections of the category centroids on the p dimensions must be in the appropriate order. There are obviously a number of possibilities between multiple nominal and multiple ordinal. We can require monotonicity only for selected dimensions, for example only for the first one. Different definitions of partial orders on p-space can also be tried. *Multiple numerical* can be defined in various ways. We can require, for example, that the columns of Y_j are all polynomials of degree less than or equal to q of the prior numerical scores. In stead of polynomials we can also use trigonometric polynomials or splines. This is discussed in Van Rijkevorsel (1982). A very special case is *multiple linear*, in which we require that the columns of Y_j are all linear functions of the prior quantifications. Because of the normalization we use (cf infra) this means that they all must be proportional to a given vector. Thus there is only really a single quantification for each variable, which is moreover known completely. Because of this we decide to identify multiple linear and *single numerical*.

Single numerical requires that all columns of Y_j must be proportional to a given vector, say to z_j. Thus we want $Y_j = z_j a_j'$, the matrix Y_j must be of rank one. If p = 1 this clearly is no restriction, if k = 2 it is no restriction either. Without loss of generality we also require that $u'D_j z_j = 0$ and $z_j' D_j z_j = 1$. Then *single ordinal* is defined by requiring that $Y_j = z_j a_j'$ where z_j is only restricted to be in the appropriate order. And *single nominal* merely requires $Y_j = z_j a_j'$, with no further restrictions on z_j, except for the normalization. Because homogeneity analysis with all variables multiple is very similar to ordinary multiple nominal homogeneity analysis, we shall not discuss the multiple versions in detail any more. For non-nominal multiple variables the first centroid principle becomes more complicated, especially from a geometrical point of view. For ordinal multiple variables the first centroid principle can still be used almost in its original form, but the ordinal restrictions eliminate the rotational indeterminacy of the optimal quantifications and they destroy the Guttman-duality of scores and quantifications. We concentrate below on the case with all variables single. Observe, however, that one of the unique features of the Gifi-system is that programs such as PRINCALS (Gifi, 1981a, 1983) can handle mixed cases, with some variables single and some variables multiple.

If all variables are single the loss of consistency is

$$\sigma(X;z_1,\ldots,z_m;a_1,\ldots,a_m) = \frac{1}{m}\sum_{j=1}^{m} \mathrm{tr}\ (X - G_j z_j a_j')'(X - G_j z_j a_j'). \qquad (5)$$

If we let $q_j = G_j z_j$ then

$$\sigma(X;Q;A) = \mathrm{tr}\ X'X - \frac{2}{m}\,\mathrm{tr}\ AX'Q + \frac{1}{m}\,\mathrm{tr}\ A'A, \qquad (6)$$

where A is the m x p matrix with the a_j as rows and Q is the n x m matrix with the q_j as columns. Remember that $u'q_j = 0$ and $q_j'q_j = 1$ by our normalization conventions.

We still have the basic Guttman-duality between scores X and quantifications Y_j.

There is some additional flexibility, because we effectively deal with three sets of variables. We can minimize (6) over all X. This gives loss of homogeneity

$$\sigma(*;Q;A) = \frac{1}{m}\ \mathrm{tr}\ A'(I - \frac{1}{m} R)A, \tag{7}$$

where R = Q'Q. The minimum in (7) is attained for $X = m^{-1} QA$, which is the same as (3). Now Y is normalized if Y'DY = mI. If all variables are single this becomes A'A = mI. Thus minimizing loss of homogeneity over all normalized Y is the same thing as finding Q such that the sum of the p largest eigenvalues of R is maximized. In fact

$$\sigma(*;Q;*) = p - \underline{\lambda}_p(\frac{1}{m} R), \tag{8}$$

with $\underline{\lambda}_p$ the sum of the p largest eigenvalues. Loss of discrimination can be written as

$$\sigma(X;*;*) = \mathrm{tr}\ X'(I - \frac{1}{m} \hat{Q}\hat{Q}')X, \tag{9}$$

where $\hat{Q}$ is the optimal Q for given X, and where $\hat{A} = \hat{Q}'X$. Maximizing (9) over all normalized X can again be interpreted as maximizing the sum of the p largest eigenvalues of QQ', which is the same thing as maximzing the sum of the p largest eigenvalues of R = Q'Q. Homogeneity with all variables single can be interpreted in terms of the dimensionality of the *induced correlation matrix* R. Observe that if all variables are single numerical then Q is fixed, and the technique 'degenerates' to ordinary principal component analysis. Also observe that (8) implies that $\sigma(*;*;*) \geq p - 1$, with equality if and only if the induced correlation matrix has rank p. To get a loss function with lower bound equal to zero, we must subtract (p - 1)/m for each single variable.

If a variable is single its discrimination matrix is $\Delta_j = Y_j'D_jY_j = a_ja_j'$, and its loading matrix is $\Gamma_j = X'G_jY_j\Omega_j^{-\frac{1}{2}} = a_js_j'$, where $a_j = X'q_j$ are the *single loadings* and where $s_j = \Omega_j^{-\frac{1}{2}}a_j$ contains the signs of the elements of a_j. We can also compute the *category centroids* $D_j^{-1}G_j'X$. If the variable is multiple, then category centroids are category quantifications. If the variable is single then we compute the *single category quantifications* z_j, the single loadings a_j, together with the *rank-one category quantifications* $Y_j = z_ja_j'$. Additional insight can be obtained from a partitioning of the loss function (5). If all variables are single

$$\sigma(X;Y;A) = \frac{1}{m} \sum_{j=1}^{m} \mathrm{tr}(X - G_j\overline{Y}_j)'(X - G_j\overline{Y}_j) + \frac{1}{m} \sum_{j=1}^{m} \mathrm{tr}\ (\overline{Y}_j - z_ja_j')'D_j(\overline{Y}_j - z_ja_j'). \tag{10}$$

In (10) the $\overline{Y}_j$ are the category centroids. The first component in the partitioning is *multiple loss*, the second component is *single loss*. For multiple variables we can write Y_j for $\overline{Y}_j$, and single loss is zero. Multiple loss is equal to $\mathrm{tr}\ X'(I - P_*)X$, and single loss is $\mathrm{tr}\ X'P_*X - m^{-1}\ \mathrm{tr}\ A'A$. *Total loss*, the sum of multiple and single loss, is $p - m^{-1}\ \mathrm{tr}\ A'A$, and at the optimum this is $p - \underline{\lambda}_p(m^{-1}R)$, as before. Multiple loss indicates in how far the object scores deviate from the category quantifications of the categories they fall in. Single loss indicates in how far the category quantifications deviate from the best fitting line through the origin. This also explains why the lower bound for total loss is nonzero. For perfect fit we want object points to coincide with corresponding category points, which also must be on a line through the origin. But taken together this implies that the rank of X must be equal to one, which is contrary to our normalization of X if p > 1.

It is of some interest to observe that in the case of a multiple variable the category quantifications Y_j, which are also category centroids, can be decomposed according to

$$Y_j = \sum_{s=1}^{t} z_{js}a_{js}'. \tag{11}$$

Such a decomposition is possible in many different ways. If t = p, $Z_j = Y_j\Omega_j^{-\frac{1}{2}}$ and $A_j = \Omega_j^{\frac{1}{2}}$, then indeed $Y_j = Z_jA_j'$ and also $\mathrm{diag}(Z_j'D_jZ_j) = I$, which makes the matrix

$Z_j'D_jZ_j$ a correlation matrix. Another possibility is to take $t = k$, $Z_j = D_j^{-\frac{1}{2}}$ and in addition $A_j = Y_j'D_j^{\frac{1}{2}}$. Again $Y_j = Z_jA_j'$, but now $Z_j'D_jZ_j = I$. In general it may be interesting to decompose Y_j with the additional restriction that $Z_j'D_jZ_j = I$, because this implies that $\Delta_j = A_jA_j'$, which is convenient for the interpretation. Two interesting choices for which this additional restriction is true are orthogonal polynomials on prior scores or the singular value decomposition of Y_j, both with the D_j as weights. Remember, however, that decomposition (11) does not influence the computations and can be constructed afterwards. It does show a possible relationship between the multiple and the single approach, because for the single approach decomposition (11) is essentially unique.

In this section we have seen that using rank-one restrictions on the category quantifications makes it possible to bridge the gap between homogeneity analysis and ordinary principal component analysis in a convenient way. Another possibility is to use *dédoublement*, as in Benzécri (1973). The computational details and the graphics of the mixed multiple-single program PRINCALS are in Gifi (1981a, 1983). Relationships between multiple and single quantifications in some very interesting idealized theoretical cases, which also seem to be relevant for a large class of practical problems, are investigated mathematically in De Leeuw (1982) and Bekker (1983).

ADDITIVITY RESTRICTIONS ON THE CATEGORY QUANTIFICATIONS

In ordinary multivariate analysis the partitioning of variables into sets plays an important role. A partitioning into two sets is used in canonical analysis. If one of the two sets consists of a single variable we are dealing with regression or discriminant analysis. Special alternating least squares programs with optimal scaling for two-set analysis were written by De Leeuw, Young, and Takane (1976), Young, De Leeuw, and Takane (1976). A greatly improved version of CANALS has recently been described by Van der Burg (1983), Van der Burg and De Leeuw (1983). Two sets is a rather extreme special case, however, just as principal component analysis is the special case of m sets, each set containing a single variable. In Gifi's categorical data framework an elegant treatment of the concept of sets of variables is possible. An algorithm and a computer program OVERALS, for the general case of any number of sets, are currently being developed. Van der Burg and Verdegaal (1983) present the first results of OVERALS. Related theory has been developed in France by Masson (1974), Dauxois and Pousse (1976), Saporta (1975), and Tenenhaus (1982).

Suppose the k categories of variable j have product structure. With this we mean that there are sets $T_1,\ldots,T_\upsilon$, such that each category corresponds with an element of the Cartesian product $T_1 \times \ldots \times T_\upsilon$. Quantifications under *additivity restrictions* are defined as follows. Suppose $Y_{j\alpha}$ contains quantifications of *factor* T_α, then we require for each category $(t_1,\ldots,t_\upsilon)$ that

$$y_j(t_1,\ldots,t_\upsilon) = y_{j1}(t_1) + \ldots + y_{j\upsilon}(t_\upsilon). \tag{12}$$

The notation is a bit clumsy, but the meaning is probably clear. Categories of a variable can be organized in a υ-dimensional *grid* or *factorial design*, and we require that the category quantifications have no *interaction* or consist of *main effects* only. It is more convenient to write these restrictions using *design matrices*. This means that there is a binary matrix S_j, with structure $S_j = (S_{j1}|\ldots|S_{j\upsilon})$, such that (12) is simply $Y_j = \sum S_{j\alpha}Y_{j\alpha}$. Matrices $S_{j\alpha}$ are indicator matrices. If T_α has k_α elements, then the matrix $S_{j\alpha}$ has k_α columns and $k_1 \times \ldots \times k_\upsilon = k$ rows. Thus $G_jY_j = \sum G_{j\alpha}Y_{j\alpha}$, where $G_{j\alpha} = G_jS_{j\alpha}$. Each $G_{j\alpha}$ is an indicator matrix.

We have used the design matrix to split a variable into υ factors. It is easy to see that the formalism can also be used into group variables into *sets of variables*. If we want to organize υ variables into a set, then we can form the *interactive variable* with $k_1 \times \ldots \times k_\upsilon$ categories, and then impose additivity restrictions. This shows that we can also choose to ignore the fact that a variable is interactive, and work with the interactive indicators without imposing restrictions. As a consequence it becomes a part of the model what we wish to consider as a variable. If we have three variables to start with, for instance, then we can use interactive coding in various ways. We can code them as a single variable with $k_1k_2k_3$ categories, as two variables with k_1k_2 and k_3 categories, as three variables with k_1, k_2, and k_3 categories, and so on. In stead of performing a principal component analysis on m variables, we can also form all $\binom{m}{2}$ pairs of variables, code them interactively, and perform a principal component analysis on all pairs. Using interactive coding and additivity restrictions means that we can leave various higher-order interactions intact, while assuming others away. In a sense this bridges the gap between homogeneity analysis and loglinear analysis, although there remain substantial differences between the two classes of techniques. Observe that in homogeneity analysis the additivity restrictions can be combined with the rank one restrictions. In the program OVERALS we can restrict each $Y_{j\alpha}$ by $Y_{j\alpha} = z_{j\alpha}a_{j\alpha}'$, we can also require this for some factors but not for others.

The least squares loss function used in OVERALS is, again, loss of consistency (1). If we impose additivity restrictions for all variables, and we assume for ease of notation only that all sets contain the same number of variables (or: all variables have the same number of factors), then

$$\sigma(X,Y) = \frac{1}{m}\sum_{j=1}^{m} \mathrm{tr}\,(X - \sum_{\alpha=1}^{\upsilon} G_{j\alpha}Y_{j\alpha})'(X - \sum_{\alpha=1}^{\upsilon} G_{j\alpha}Y_{j\alpha}). \tag{13}$$

In OVERALS we normalize X, but the Guttman-duality linking loss of homogeneity and loss of discrimination is still true, even if there are additivity and rank one restrictions.

The additivity restrictions have some interesting consequences for the centroid principle. This is due to the fact that in most cases the indicator matrices $G_{j\alpha}$ within set j are not orthogonal, which introduces complications similar to those in nonorthogonal analysis of variance. Consider factor α in variable j. Let

$$U_{j\alpha} = \sum_{\beta\neq\alpha}^{\upsilon} G_{j\alpha}Y_{j\alpha}. \tag{14}$$

The optimal $Y_{j\alpha}$, given X and the remaining category quantifications, is given by the centroids $D_{j\alpha}^{+}G_{j\alpha}'(X - U_{j\alpha})$. At the optimum these are the multiple category quantifications, which are unequal in general to the category centroids $D_{j\alpha}^{+}G_{j\alpha}'X$. The discrimination matrix is

$$\Delta_{j\alpha} = Y_{j\alpha}'D_{j\alpha}Y_{j\alpha} = (X - U_{j\alpha})'P_{j\alpha}(X - U_{j\alpha}), \tag{15}$$

and the loading matrix is

$$\Gamma_{j\alpha} = X'G_{j\alpha}Y_{j\alpha}/\Omega_{j\alpha}^{-\frac{1}{2}} = X'P_{j\alpha}(X - U_{j\alpha})\Omega_{j\alpha}^{-\frac{1}{2}}. \tag{16}$$

Expressions (15) and (16) simplify considerably if $P_{j\alpha}U_{j\alpha} = 0$, i.e. if factor α is independent of the other. If there are no additivity restrictions then this condition is vacuously true, because $U_{j\alpha} \equiv 0$.

For single factors we have $\Delta_{j\alpha} = a_{j\alpha}a_{j\alpha}'$ and $\Gamma_{j\alpha} = b_{j\alpha}s_{j\alpha}'$. Here $b_{j\alpha} = X'G_{j\alpha}z_{j\alpha}$, $a_{j\alpha} = (X - U_{j\alpha})'G_{j\alpha}z_{j\alpha}$ and $s_{j\alpha}$ contains the signs of the elements of $a_{j\alpha}$. The vector $b_{j\alpha}$ consists of correlation coefficients (*single loadings*), but in general $a_{j\alpha}$ does not consist of correlation coefficients. If all variables are single it is possible again to interpret the OVERALS solution in terms of the induced correlation matrix. The largest p eigenvalues of the between-set correlation matrix relative to the within-set correlation matrix are maximized. Partitioning the loss is less simple if there are additivity restrictions, again for the same reasons are in nonorthogonal analysis of variance. Something similar to (10) remains true, however. We must adjust (10) because the total loss of a variable can be partitioned in υ ways, one for each factor. This is seen most easily by substituting $X - U_{j\alpha}$ for X in (10). An additional partitioning which can be made quite easily is

$$\sigma(X;Y) = \frac{1}{m}\sum_{j=1}^{m} \mathrm{tr}\,(X - G_j\overline{Y}_j)'(X - G_j\overline{Y}_j) + \\ + \frac{1}{m}\sum_{j=1}^{m} \mathrm{tr}\,(\overline{Y}_j - \sum_{\alpha=1}^{\upsilon} S_{j\alpha}Y_{j\alpha})'D_j(\overline{Y}_j - \sum_{\alpha=1}^{\upsilon} S_{j\alpha}Y_{j\alpha}). \qquad (17)$$

The second component is the *loss due to additivity*.

The Gifi-program OVERALS is the natural end-product of the system. Additional developments in the direction of path analysis and partial canonical correlation are mentioned briefly in Gifi (1981a). They require special procedures, which do not fit naturally into the loss function (13). At the moment we are experimentin with a related, but slightly different system, based on the notion of *copies*. In this system all variables are single, which means that results can be easily formulated in terms of induced correlation matrices. Object scores get less emphasis, and the normalization is on the Y_j. What used to be called multiple variables in the Gifi-system is introduced now by allowing for multiple occurrence of the same variable in a set (this is what is meant by using copies). There is some gain in generality, because now partial canonical correlation fits in naturally, and there may be some gain in computational efficiency. But in the new system, which is explained in more detail in De Leeuw (1983b), we loose a great deal of the geometrical appeal of the Gifi-system. The geometry of the Gifi-system is based on the first centroid principle, applied to the object scores represented as points in low dimensional space. We have seen that even in the more complicated restricted situations the centroid principle can still be used, and it remains a powerful interpretational tool. The *Copy-system* shifts the emphasis from distances to correlations, from object scores to category quantifications, from multidimensional scaling to multivariate analysis, from least squares loss functions to eigenvalue problems. Although the results computed by the two systems are the same, this shift in emphasis does have consequences for plots and interpretations.

REFERENCES

(1) Benzécri, J.P. a.o., L'analyse des données (Dunod, Paris, 1973).

(2) Burt, C. The factorial analysis of qualitative data, British J. Statist. Psychol. 3 (1950) 166-185.

(3) Dauxois, J. and Pousse, A., Les analyses factorielles en calcul des probabilités et en statistique, Thèse d'Etat, Université Paul Sabatier de Toulouse (1976).

(4) De Leeuw, J. Canonical analysis of categorical data, Doctoral dissertation, Leiden University (1973).

(5) De Leeuw, J. A significance test for multiple correspondence analysis, Department of Data Theory, University of Leiden (1983a).

(6) De Leeuw, J. Nonlinear multivariate analysis on bimarginals, Paper presented at the Psychometric Society Meeting, Jouy-en-Josas (July 1983).

(7) De Leeuw, J., Young, F.W., Takane, Y., Additive structure in qualitative data, Psychometrika 41 (1976) 471-504.

(8) Gifi, A. Nonlinear multivariate analysis. Department of Data Theory, Leiden University (1981a).

(9) Gifi, A. HOMALS users guide. Department of Data Theory, Leiden University (1981b).

(10) Gifi, A. PRINCALS users guide. Department of Data Theory, Leiden University, (1983).

(11) Guttman, L. The quantification of a class of attributes: a theory and method of scale construction, in: Horst, P. (ed), The prediction of personal adjustment (Social Science Research Council, New York, 1941).

(12) Hayashi, C., Theory and examples of quantifiaction II, Proc. Inst. Statist. Math. 4 (1956) 19-30.

(13) Heiser, W.J., Unfolding analysis of proximity data, Doctoral dissertation, Leiden University (1981).

(14) Masson, M., Processus linéaires, analyse non linéaire des données. Thèse d'état, Université de Paris VI (1974).

(15) Meulman, J. Homogeneity analysis of incomplete data, (DSWO Press, Leiden, 1982).

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

(17) Saporta, G., Liaisons entre plusieurs ensembles de variables et codage de données qualitatives, Thèse 3ème cycle, Université Paris VI (1975).

(18) Tenenhaus, M. Dissertation proposal, Université Paul Sabatier de Toulouse (1983).

(19) Van der Burg, E., CANALS users guide, Department of Data Theory, Leiden University (1983).

(20) Van der Burg, E., and De Leeuw, J., Nonlinear canonical correlation. British J. Math. Statist. Psychol. 36 (1983) 54-80.

(21) Van der Burg, E., and Verdegaal, R. Canonical analysis of M sets of variables, Paper presented at the Psychometric Society Meeting, Jouy-en-Josas, (july 1983).

(22) Van Rijckevorsel, J. Canonical analysis with B-splines, In: Caussinus, H., Ettinger, P., and Tomassone, R. (eds) COMPSTAT 1982, (Physika Verlag, Wien, 1982).

(23) Young, F.W., Quantitative analysis of qualitative data, Psychometrika 46 (1981) 357-388.

(24) Young, F.W., De Leeuw, J., and Takane, Y. Regression with qualitative and quantitative variables, Psychometrika 41 (1976) 505-530.

(25) Young, F.W., De Leeuw, J., and Takane, Y. Quantifying qualitative data, In: Lantermann, E.D., and Feger, H. (eds): Similarity and choice. (Huber, Bern, 1980).

(26) Bekker, P. Relation between different forms of nonlinear principal component analysis, Department of Data Theory, Leiden Unibersity, (1983).

(27) De Leeuw, J. Nonlinear principal component analysis, In: Caussinus, H., Ettinger, P., and Tomassone, R. (eds) COMPSTAT 1982, (Physika Verlag, Wien, 1982).