RR 89-01

ALGORITHM DESCRIPTIONS FOR ANACOR, HOMALS, PRINCALS, AND OVERALS

Albert Gifi

Department of Data Theory
University of Leiden

Preface

This report has four parts, each of which offers self-contained algorithm descriptions of the optimal scaling programs ANACOR, HOMALS, PRINCALS, and OVERALS. I apologize to those readers who find the overlap in explications and references boring. The descriptions are valid for those versions of the programs that are supported by the Department of Data Theory at the time of writing, which will also be available from the software company SPSS Inc. in the very near future.

Leiden, February 1989 Albert Gifi

ANACOR

GENERAL

The ANACOR algorithm consists of three major parts: (1) a singular value decomposition (SVD), (2) centering and rescaling of the data, and various rescalings of the results, and (3) variance estimation by the delta method. Other names for SVD are "Eckart-Young decomposition", after Eckart and Young (1936) who introduced the technique in psychometrics, and "basic structure" (Horst, 1963). The rescalings and centering, including their rationale, are well explained in Benzécri (1969), Nishisato (1980), Gifi (1981), and Greenacre (1984). Those who are interested in the general framework of matrix approximation and reduction of dimensionality with positive definite row and column metrics are referred to Rao (1980). The delta method is a method that can be used for the derivation of asymptotic distributions, and is particularly useful for the approximation of the variance of complex statistics. There are many versions of the delta method, differing in the assumptions made and in the strength of the approximation (Rao, 1973, ch.6; Bishop et al., 1975, ch. 14; Wolter, 1985, ch. 6).

(1) NOTATION

(a) k_1 number of rows (row objects)

k_2 number of columns (column objects)

p number of dimensions.

(b) Data related quantities

f_{ij} nonnegative data value for row i and column j; collected in table F

f_{i+} marginal total of row i, $i = 1,...,k_1$

f_{+j} marginal total of column j, $j = 1,...,k_2$

N grand total of F

(c) Scores and statistics

r_{is} score of row object i on dimension s

c_{js} score of column object j on dimension s

I total inertia.

(2) BASIC CALCULATIONS

One way to phrase the ANACOR objective (cf. Heiser, 1981) is to say that we wish to find row scores $\{r_{is}\}$ and column scores $\{c_{js}\}$ so that the function

$$\sigma(\{r_{is}\};\{c_{js}\}) = \sum_i \sum_j f_{ij} \sum_s (r_{is} - c_{js})^2$$

is minimal, under the standardization restriction either that $\sum_i f_{i+} r_{is} r_{it} = \delta^{st}$ or $\sum_j f_{+j} c_{js} c_{jt} = \delta^{st}$, where δ^{st} is Kronecker's delta, and t is an alternative index for dimensions. The trivial set of scores ({1},{1}) is excluded.

The ANACOR algorithm can be subdivided into five steps, as explained below.

(i) Data scaling and centering

The first step is to form the auxiliary matrix Z with general element

$$z_{ij} = f_{ij} / (\sqrt{f_{i+}}\sqrt{f_{+j}}) - (\sqrt{f_{i+}}\sqrt{f_{+j}}) / N.$$

(ii) Singular value decomposition

Let the singular value decomposition of Z be denoted by

$$Z = K\Lambda L', \quad \text{with } K'K = I,\ L'L = I, \text{ and } \Lambda \text{ diagonal.}$$

This decomposition is calculated by a routine based on Golub and Reinsch (1971). It involves Householder reduction to bidiagonal form and diagonalization by a QR procedure with shifts. The routine requires an array with more rows than columns, so when $k_1 < k_2$ the original table is transposed and the parameter transfer is permuted accordingly.

(iii) Adjustment to the row and column metric

The arrays of both the left-hand singular vectors and the right-hand singular vectors are adjusted row-wise to form scores that are standardized in the row and in the column marginal proportions, respectively:

$$r_{is} = k_{is} / \sqrt{f_{i+}}/N,$$
$$c_{js} = l_{js} / \sqrt{f_{+j}}/N.$$

This way both sets of scores satisfy the standardization restrictions simultaneously.

(iv) Determination of variances and covariances

For the application of the delta method to the results of generalized eigenvalue methods under multinomial sampling the reader is referred to Gifi (1981, chapter 12) and Israëls (1987, Appendix B). It is shown there that N times the variance-covariance matrix of a function ϕ of the observed cell proportions $p = \{p_{ij} = f_{ij}/N\}$ asymptotically reaches the form

$$N\,COV(\phi(p)) \rightarrow \quad \Sigma_i\Sigma_j \pi_{ij} (\partial\phi/\partial p_{ij}) (\partial\phi/\partial p_{ij})' - (\Sigma_i\Sigma_j \pi_{ij} \partial\phi/\partial p_{ij}) (\Sigma_i\Sigma_j \pi_{ij} \partial\phi/\partial p_{ij})'.$$

Here the quantities π_{ij} are the cell probabilities of the multinomial distribution, and $\partial\phi/\partial p_{ij}$ are the partial derivatives of ϕ (which is either a generalized eigenvalue or a generalized eigenvector) with respect to the observed cell proportion. Expressions for these partial derivatives can also be found in the above-mentioned references.

(v) Normalization of row and column scores

Depending on the normalization option chosen the scores are normalized, which implies a compensatory rescaling of the coordinate axes of the row scores and the column scores. The general formula for the weighted sum of squares that results from this rescaling is

$$\text{row scores: } \Sigma_i f_{i+} r_{is}^2 = N\lambda_s^{(1+q)} \qquad \text{column scores: } \Sigma_j f_{+j} c_{js}^2 = N\lambda_s^{(1-q)}.$$

The parameter q can be chosen freely, or it can be specified according to the following designations: canonical: q = 0; row principal: q = 1; column principal: q = -1. There is a fifth possibility, choosing the designation "principal", that does not fit into this scheme. It implies that the weighted sum of squares of both sets of scores becomes equal to $N\lambda_s^2$. The estimated variances and covariances are adjusted according to the type of normalization chosen.

(3) DIAGNOSTICS

After printing the data ANACOR optionally also prints a table of row profiles and column profiles, which are $\{f_{ij} / f_{i+}\}$ and $\{f_{ij} / f_{+j}\}$, respectively.

(a) Singular values, maximum rank, and inertia

All singular values λ_s defined in 2(ii) are printed up to a maximum of $\min[(k_1-1),(k_2-1)]$.

Small singular values and corresponding dimensions are suppressed when they don't exceed the quantity $(k_1k_2)^{1/2}\ 10^{-7}$; in this case a warning message is issued. Dimension-wise inertia and total inertia are given by the relationships

$$I = \sum_s \lambda_s^2 = \sum_s (\sum_i f_{i+} r_{is}^2 / N),$$

where the right-hand part of this equality is only true if the normalization is row principal (but for the other normalizations similar relationships are easily derived from 2(v)). The quantities "proportion explained" are equal to inertia divided by total inertia: λ_s^2 / I.

(b) Scores and contributions

The next output is given first for rows, then for columns, and always preceeded by a column of marginal proportions (f_{i+}/N and f_{+j}/N, respectively). The table of scores is printed in p dimensions. The contribution to the inertia of each dimension is given by

$$\tau_{is} = (f_{i+} / N)\ r_{is}^2 / \lambda_s^{(1+q)}$$
$$\tau_{js} = (f_{+j} / N)\ c_{js}^2 / \lambda_s^{(1-q)},$$

where q, as before, denotes the parameter which controls the normalization. The contribution of dimensions to the inertia of each point is given by, for s,t = 1,...p,

$$\sigma_{is} = r_{is}^2\lambda_s^{(1+q)} / \sum_t r_{it}^2\lambda_t^{(1+q)}$$
$$\sigma_{js} = c_{js}^2\lambda_s^{(1-q)} / \sum_t c_{jt}^2\lambda_t^{(1-q)}\ .$$

(c) Variances and correlation matrix of singular values and scores

The computation of variances and covariances is explained in section 2(iv). Since the row and column scores are linear functions of the singular vectors, an adjustment is necessary

depending on the normalization option chosen. From these adjusted variances and covariances the correlations are derived in the standard way.

(d) Permutations of the input table

For each dimension s let $\rho(i|s)$ be the permutation of the first k_1 integers that would sort the sth column of $\{r_{is}\}$ in ascending order; similarly, let $\rho(j|s)$ be the permutation of the first k_2 integers that would sort the sth column of $\{c_{js}\}$ in ascending order. Then the permuted data matrix is given by $\{f_{\rho(i|s),\rho(j|s)}\}$.

(4) PRECISION

All calculations are in single precision.

(5) REFERENCES

Benzécri, J.-P. (1969). Statistical analysis as a tool to make patterns emerge from data. In Methodologies of Pattern Recognition (Watanabe, S., Ed.). New York: Academic Press.

Bishop, Y.M., Fienberg, S.E., and Holland, P.W. (1975). Discrete Multivariate Analysis. Cambridge, Massachusetts: M.I.T. Press.

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

Gifi, A. (1981). Nonlinear Multivariate Analysis. Leiden: Department of Data Theory.

Golub, G.H., and Reinsch, C. (1971), Chapter I.10 in Linear Algebra, Vol. II of Handbook for Automatic Computation (Wilkinson, J.H. and C. Reinsch, Eds.). New York: Springer.

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

Heiser (1981). Unfolding Analysis of Proximity Data. Doctoral dissertation. Department of Data Theory, University of Leiden.

Horst, P. (1963). Matrix Algebra for Social Scientists. New York: Holt, Rinehart and Winston.

Israëls, A. (1987). Eigenvalue Techniques for Qualitative Data. Leiden: DSWO Press.

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

Rao, C.R. (1973). Linear Statistical Inference and its Applications. New York: Wiley.

Rao, C.R. (1980). Matrix approximations and reduction of dimensionality in multivariate statistical analysis. In: Multivariate Analysis (Krishnaiah, P.R., Ed.) Vol 5. Amsterdam: North-Holland.

Wolter, K.M. (1985). Introduction to Variance Estimation. Berlin: Springer Verlag.

HOMALS

GENERAL

The iterative HOMALS algorithm is a modernized version of Guttman (1941). The treatment of missing values - to be described below - is based on setting weights in the loss function equal to zero, and was first described in De Leeuw and Van Rijckevorsel (1980). Other possibilities do exist, and can be accomplished by recodings of the data (Gifi, 1981; Meulman, 1982).

(1) NOTATION

(a) n number of cases (objects)

m number of variables

p number of dimensions.

(b) For variable j, j = 1,...,m

h_j n-vector with categorical observations

k_j number of categories (distinct values) of variable j

G_j indicator matrix for variable j, of order n x k_j

the elements of G_j are defined as (i = 1,...,n; r = 1,...,k_j)

$g_{(j)ir} = 1$ when the ith object is in the rth category of variable j

$g_{(j)ir} = 0$ when the ith object is not in the rth category of variable j

M_j binary, diagonal n x n matrix, with diagonal elements defined as

$m_{(j)ii} = 1$ when the ith observation is within the range [1,k_j]

$m_{(j)ii} = 0$ when the ith observation is outside the range [1,k_j]

D_j diagonal matrix containing the univariate marginals, i.e. the column sums of G_j.

(c) The quantification matrices are

X object scores, of order n x p

Y_j category quantifications, of order k_j x p

Y concatenated category quantification matrices, of order $\sum_j k_j$ x p.

Note: the matrices G_j, M_j, and D_j are exclusively notational devices; they are stored in reduced form, and the program fully profits from their sparseness by replacing matrix multiplications with selective accumulation.

(2) OBJECTIVE FUNCTION OPTIMIZATION

The HOMALS objective is to find object scores X and a set of Y_j (for j = 1,...m) so that the function

$$\sigma(X;Y) = 1/m \sum_j \text{tr } (X - G_jY_j)'M_j(X - G_jY_j)$$

is minimal, under the normalization restriction $X'M_*X = mn\ I$, where the matrix $M_* = \sum_j M_j$, and I is the p x p identity matrix. The inclusion of M_j in $\sigma(X;Y)$ ensures that there is no influence of data values outside the range $[1,k_j]$, which may be really missing or merely regarded as such; M_* contains the number of "active" data values for each object. The object scores are also centered, i.e. they satisfy $u'M_*X = 0$ with u denoting an n-vector with ones.

Optimization is achieved through the following iteration scheme:

(i) Initialization

(ii) Update object scores

(iii) Orthonormalization

(iv) Update category quantifications

(v) Convergence test: repeat (ii) - (iv) or continue

(vi) Rotation

These steps are explained below.

(i) Initialization

The object scores X are initialized with random numbers, which are normalized so that $u'M_*X = 0$ and $X'M_*X = mn\ I$, yielding $X^{\sim}$. Then the first category quantifications are obtained as $Y_j^{\sim} = D_j^{-1}G_j'X^{\sim}$.

(ii) Update object scores

First the auxiliary score matrix Z is computed as

$$Z \leftarrow \sum_j M_j G_j Y_j^{\sim}$$

and centered with respect to M_*:

$$Z^{\sim} \leftarrow \{M_* - (M_* uu' M_* / u' M_* u)\}\, Z.$$

These two steps yield locally the best updates when there would be no orthogonality constraints.

(iii) Orthonormalization

The orthonormalization problem is to find an M_*-orthonormal X^+ that is closest to $Z^{\sim}$ in the least squares sense. In HOMALS this is done by setting

$$X^+ \leftarrow m^{1/2} M_*^{-1/2} \text{GRAM}\,(M_*^{-1/2} Z^{\sim}),$$

which is equal to the genuine least squares estimate up to a rotation - see (vi). The notation GRAM( ) is used to denote the Gram-Schmidt transformation (Björk and Golub, 1973).

(iv) Update category quantifications

For j = 1,...,m the new category quantifications are computed as:

$$Y_j^+ = D_j^{-1} G_j' X^{\sim}$$

(v) Convergence test

The difference between consecutive loss function values $\sigma(X^{\sim};Y^{\sim}) - \sigma(X^+;Y^+)$ is compared with the user specified convergence criterion ε - a small positive number. Steps (ii) to (iv) are repeated as long as the loss difference exceeds ε.

(vi) Rotation

As remarked in (iii), during iteration the orientation of X and Y with respect to the coordinate system is not necessarily correct; this also reflects the fact that $\sigma(X;Y)$ is invariant under simultaneous rotations of X and Y. From theory it is known that solutions in different dimensionality should be nested, i.e. the p-dimensional solution should be equal to the first p columns of the (p+1)-dimensional solution. Nestedness is achieved by computing the eigenvectors of the matrix $1/m \sum_j Y_j' D_j Y_j$. The corresponding eigenvalues

are printed after the convergence message of the program. The calculation involves tridiagonalization with Householder transformations followed by the implicit QL algorithm (Wilkinson, 1965).

(3) DIAGNOSTICS

(a) Maximum rank (may be issued as a warning when exceeded)

The maximum rank p_{max} indicates the maximum number of dimensions that can be computed for any data set. In general we have:

$$p_{max} = \min \{(n - 1), ((\sum_j k_j) - \max (m_1, 1))\},$$

where m_1 is the number of variables with no missing values. Although the number of non-trivial dimensions may be less than p_{max} when m = 2, HOMALS does allow dimensionalities all the way up to p_{max}.

(b) Marginal frequencies

The frequencies table gives the univariate marginals and the number of missing values (i.e., values that are regarded as out of range for the current analysis) values for each variable. These are computed as the column sums of D_j and the total sum of M_j.

(c) The discrimination measures

These are the dimensionwise variances of the quantified variables. For variable j and dimension s we have:

$$\eta_{js}^2 = y_{(j)s}' D_j y_{(j)s} / n,$$

where $y_{(j)s}$ is the sth column of Y_j, corresponding to the sth quantified variable $G_j y_{(j)s}$.

(d) Eigenvalues

The computation of the eigenvalues that are reported after convergence is discussed in (vi). When the HISTORY option is in effect the sum of the eigenvalues is reported during iteration, under the heading "total fit". Due to the fact that the sum of the eigenvalues is equal to the trace of the original matrix the sum can be computed as $1/m \sum_j \sum_s \eta_{js}^2$. The value of $\sigma(X;Y)$ is equal to $p - 1/m \sum_j \sum_s \eta_{js}^2$.

(4) PRECISION

All basic calculations are in double precision, while the data and some of the simple statistics are kept in single word integer arrays.

(5) REFERENCES

Björk, A. and Golub, G.H. (1973). Numerical methods for computing angles between linear subspaces. Mathematics of Computation, 27, 579-594.

De Leeuw, J. and Van Rijckevorsel, J. (1980). HOMALS and PRINCALS - Some generalizations of principal components analysis. In E. Diday et al. (Eds.), Data Analysis and Informatics. Amsterdam: North-Holland.

Gifi, A. (1981). Non-linear Multivariate Analysis. Leiden: Department of Data Theory.

Guttman. L. (1941). The quantification of a class of attributes: A theory and method of scale construction. In P. Horst et al. (Eds.), The Prediction of Personal Adjustment. New York: Social Science Research Council.

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

Wilkinson, J.H. (1965). The algebraic eigenvalue problem. Oxford: Clarendon Press.

PRINCALS

GENERAL

The PRINCALS algorithm was first described in Van Rijckevorsel and De Leeuw (1979) and De Leeuw and Van Rijckevorsel (1980); also see Gifi (1981, 1985). Characteristic features of PRINCALS are the possibility to specify any of a number of measurement levels for each variable separately, and the treatment of missing values by setting weights in the loss function equal to zero.

(1) NOTATION

(a) n number of cases (objects)

m number of variables

p number of dimensions.

(b) For variable j, j = 1,...,m

h_j n-vector with categorical observations

k_j number of categories (distinct values) of variable j

G_j indicator matrix for variable j, of order n x k_j

the elements of G_j are defined as (i = 1,...,n; r = 1,...,k_j)

$g_{(j)ir} = 1$ when the ith object is in the rth category of variable j

$g_{(j)ir} = 0$ when the ith object is not in the rth category of variable j

M_j binary, diagonal n x n matrix, with diagonal elements defined as

$m_{(j)ii} = 1$ when the ith observation is within the range [1,k_j]

$m_{(j)ii} = 0$ when the ith observation is outside the range [1,k_j]

D_j diagonal matrix containing the univariate marginals, i.e. the column sums of G_j.

(c) The quantification matrices and parameter vectors are

X object scores, of order n x p

Y_j multiple category quantifications, of order k_j x p

y_j single category quantifications, of order k_j

a_j variable weights (equal to component loadings), of order p

Q transformed data matrix of order n x m, with columns $q_j = G_j y_j$

$\underline{Y}$ collection of multiple and single category quantifications.

Note: the matrices G_j, M_j, and D_j are exclusively notational devices; they are stored in reduced form, and the program fully profits from their sparseness by replacing matrix multiplications with selective accumulation.

(2) OBJECTIVE FUNCTION OPTIMIZATION

The PRINCALS objective is to find object scores X and a set of $\underline{Y}_j$ (for j = 1,...,m) - the underlining indicates that they are possibly restricted - so that the function

$$\sigma(X;\underline{Y}) = 1/m \sum_j \text{tr } (X - G_j\underline{Y}_j)'M_j(X - G_j\underline{Y}_j)$$

is minimal, under the normalization restriction $X'M_*X = mn\ I$, where the matrix $M_* = \sum_j M_j$, and I is the p x p identity matrix. The inclusion of M_j in $\sigma(X;\underline{Y})$ ensures that there is no influence of data values outside the range $[1,k_j]$, which may be really missing or merely regarded as such; M_* contains the number of "active" data values for each object. The object scores are also centered, i.e. they satisfy $u'M_*X = 0$ with u denoting an n-vector with ones.

The following measurement levels are distinguished in PRINCALS:

(a) multiple nominal: $\underline{Y}_j = Y_j$ / unrestricted

(b) single nominal: $\underline{Y}_j = y_j a_j'$ / rank-one restrictions only

(c) (single) ordinal: $\underline{Y}_j = y_j a_j'$ and $y_j \in C_j$ / rank-one and monotonicity restrictions

(d) (single) numerical: $\underline{Y}_j = y_j a_j'$ and $y_j \in L_j$ / rank-one and linearity restrictions.

For each variable these levels can be chosen independently. The general requirement in the "single" options is $\underline{Y}_j = y_j a_j'$, i.e. $\underline{Y}_j$ is of rank one; for identification purposes y_j is always normalized so that $y_j'D_j y_j = n$, which implies that the variance of the quantified variable $q_j = G_j y_j$ is 1. In the ordinal case the additional restriction $y_j \in C_j$ means that y_j must be located in the convex cone of all k_j-vectors with nondecreasing elements; in the numerical case the additional

restriction $y_j \in L_j$ means that y_j must be located in the subspace of all k_j-vectors that are a linear transformation of the vector consisting of k_j successive integers.

Optimization is achieved by executing the following iteration scheme twice:

(i) Initialization I or II

(ii) Update object scores

(iii) Orthonormalization

(iv) Update category quantifications

(v) Convergence test: repeat (ii) - (iv) or continue

(vi) Rotation

The first time (for the initial configuration) initialization I is used and all single variables are temporarily treated as single numerical, the second time (for the final configuration) initialization II is used.Steps (i) to (vi) are explained below.

(i) Initialization

I. The object scores X are initialized with random numbers, which are normalized so that $u'M_*X = 0$ and $X'M_*X = mn\ I$, yielding $X^{\sim}$. For multiple variables the initial category quantifications are obtained as $Y_j^{\sim} = D_j^{-1}G_j'X^{\sim}$. For single variables the initial category quantifications are defined as the first k_j successive integers, normalized so that $u'D_jy_j^{\sim} = 0$ and $y_j^{\sim\prime}D_jy_j^{\sim} = n$, and the initial variable weights are calculated as the vector $a_j^{\sim} = X^{\sim\prime}G_jy_j^{\sim}$, rescaled to unit length.

II. All relevant quantities are copied from the results of the first cycle.

(ii) Update object scores

First the auxiliary score matrix Z is computed as

$$Z \leftarrow \Sigma_j M_jG_jY_j^{\sim}$$

and centered with respect to M_*:

$$Z^{\sim} \leftarrow \{M_* - (M_*uu'M_* / u'M_*u)\}\ Z.$$

These two steps yield locally the best updates when there would be no orthogonality constraints.

(iii) Orthonormalization

The orthonormalization problem is to find an M_*-orthonormal X^+ that is closest to $Z^\sim$ in the least squares sense. In PRINCALS this is done by setting

$$X^+ \leftarrow m^{1/2} M_*^{-1/2} \text{GRAM} (M_*^{-1/2} Z^\sim),$$

which is equal to the genuine least squares estimate up to a rotation - see (vi). The notation GRAM( ) is used to denote the Gram-Schmidt transformation (Björk and Golub, 1973).

(iv) Update category quantifications; loop across variables j = 1,...,m:

(a) For multiple nominal variables the new category quantifications are computed as:

$$Y_j^+ = D_j^{-1} G_j' X^+ .$$

(b) For single variables first an unconstrained update is computed in the same way:

$$Y_j^\sim = D_j^{-1} G_j' X^+ .$$

Next one cycle of an ALS algorithm (De Leeuw et al., 1976) is executed for computing a rank-one decomposition of $Y_j^\sim$, with restrictions on the left-hand vector. This cycle starts from the previous single quantifications $y_j^\sim$ with

$$a_j^+ = Y_j^{\sim\prime} D_j y_j^\sim .$$

When the current variable is numerical we are ready; else we compute

$$y_j^* = Y_j^\sim a_j^+ .$$

Now, when the current variable is single nominal we can simply obtain y_j^+ by normalizing y_j^* in the way indicated below; else the variable must be ordinal, and we have to insert the weighted monotonic regression process

$$y_j^* \leftarrow \text{WMON} (y_j^*),$$

which makes y_j^* monotonically increasing. The weights used are the diagonal elements of D_j, and the subalgorithm used is the up-and-down-blocks minimum violators algorithm (Kruskal, 1964; Barlow et al., 1972). The result is normalized:

$$y_j^+ = n^{1/2}\, y_j^*\, (y_j^{*\prime} D_j y_j^*)^{-1/2}\,.$$

Finally, we set $\underline{Y}_j^+ = y_j^+ a_j^{+\prime}$.

(v) Convergence test

The difference between consecutive values of the quantity

$$\mathrm{TFIT} = 1/m \sum_s [\, \sum_{j \in J} y_{(j)s}' D_j y_{(j)s} + \sum_{j \notin J} a_j' a_j \,],$$

where $y_{(j)s}$ denotes the sth column of Y_j and where J is an index set recording which variables are multiple, is compared with the user specified convergence criterion ε - a small positive number. It can be shown that TFIT = p - $\sigma(X;\underline{Y})$. Steps (ii) to (iv) are repeated as long as the loss difference exceeds ε.

(vi) Rotation

As remarked in (iii), during iteration the orientation of X and Y with respect to the coordinate system is not necessarily correct; this also reflects the fact that $\sigma(X;\underline{Y})$ is invariant under simultaneous rotations of X and Y. From the theory of principal components it is known that if all variables would be single the matrix A - which can be formed by stacking the row vectors a_j' - has the property that A'A is diagonal. Therefore we may rotate so that the matrix

$$1/m\; A'A = 1/m \sum_j a_j a_j' = 1/m \sum_j Y_j' D_j Y_j$$

becomes diagonal. The corresponding eigenvalues are printed after the convergence message of the program. The calculation involves tridiagonalization with Householder transformations followed by the implicit QL algorithm (Wilkinson, 1965).

(3) DIAGNOSTICS

(a) Maximum rank (may be issued as a warning when exceeded)

The maximum rank p_{max} indicates the maximum number of dimensions that can be computed for any data set. In general we have:

$$p_{max} = \min \{(n - 1), ((\Sigma_{j \in J} k_j + m_2) - \max (m_1, \max(0, 1 - m_2)))\},$$

where m_1 is the number of multiple variables with no missing values, m_2 the number of single variables, and J an index set recording which variables are multiple. Although the number of non-trivial dimensions may be less than p_{max} when m = 2, PRINCALS does allow dimensionalities all the way up to p_{max}. When due to empty categories in the actual data the rank deteriorates below the specified dimensionality the programs stops.

(b) Marginal frequencies

The frequencies table gives the univariate marginals and the number of missing values (i.e., values that are regarded as out of range for the current analysis) values for each variable. These are computed as the column sums of D_j and the total sum of M_j.

(c) Fit and loss measures

When the HISTORY option is in effect the following fit and loss measures are reported:

(i) Total fit. This is the quantity TFIT defined in (v).

(ii) Total loss. This is $\sigma(X;\underline{Y})$, computed as the sum of multiple loss and single loss defined below.

(iii) Multiple loss. This measure is computed as

$$TMLOSS = p - 1/m \Sigma_j \text{ tr } Y_j'D_jY_j .$$

(iv) Single loss. This measure is computed only when some of the variables are single:

$$SLOSS = 1/m \Sigma_{j \notin J} \text{ tr } Y_j'D_jY_j + \Sigma_{j \notin J} a_j'a_j .$$

(d) Eigenvalues and correlations between optimally scaled variables

If there are no missing data the eigenvalues printed by PRINCALS are those of 1/m R(Q), where R(Q) denotes the matrix of correlations between the optimally scaled variables in the

columns of Q. For multiple variables q_j is defined here as $G_j y_{(j)1}$. When all variables are single, or when p = 1, R(Q) itself is also printed. If there are missing data then the eigenvalues are those of the matrix with elements $q_j' M_*^{-1} q_l$, which is not necessarily a correlation matrix, although it is positive semi-definite.

(4) PRECISION

All basic calculations are in double precision, while the data and some of the simple statistics are kept in single word integer arrays.

(5) REFERENCES

Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972). Statistical Inference under order restrictions. New York: Wiley.

Björk, A. and Golub, G.H. (1973). Numerical methods for computing angles between linear subspaces. Mathematics of Computation, 27, 579-594.

De Leeuw, J. and Van Rijckevorsel, J. (1980). HOMALS and PRINCALS - Some generalizations of principal components analysis. In E. Diday et al. (Eds.), Data Analysis and Informatics. Amsterdam: North-Holland.

De Leeuw, J., Young, F.W., and Takane, Y. (1976). Additive structure in qualitative data: an alternating least squares method with optimal scaling features. Psychometrika, 41, 471-503.

Gifi, A. (1981). Non-linear Multivariate Analysis. Leiden: Department of Data Theory.

Gifi , A. (1985). PRINCALS. Leiden: Department of Data Theory, Internal Report UG-85-02.

Kruskal, J.B. (1964). Nonmetric multidimensional scaling: a numerical method. Psychometrika, 29, 115-129.

Van Rijckevorsel, J. and De Leeuw, J. (1979). An Outline of PRINCALS. Leiden: Department of Data Theory, Internal Report RB 002-'79.

Wilkinson, J.H. (1965). The algebraic eigenvalue problem. Oxford: Clarendon Press.

OVERALS

GENERAL

The OVERALS algorithm was first described in Gifi (1981) and Van der Burg, De Leeuw and Verdegaal (1984); also see Verdegaal (1986), Van de Geer (1987), Van der Burg, De Leeuw and Verdegaal (1988), and Van der Burg (1988). Characteristic features of OVERALS, conceived by De Leeuw (1973), are the partitioning of the variables into K sets, and the possibility to specify any of a number of measurement levels for each variable separately. Analogously to the situation in multiple regression and canonical correlation analysis OVERALS focusses on the relationships between sets; any particular variable only contributes to the results in as much as it provides information that is independent of the other variables in the same set.

(1) NOTATION

(a) n number of cases (objects)

m total number of variables

p number of dimensions

K number of sets.

(c) For variable j, j = 1,...,m

k_j number of categories (distinct values) of variable j

G_j indicator matrix for variable j, of order n x k_j

the elements of G_j are defined as (i = 1,...,n; r = 1,...,k_j)

$g_{(j)ir} = 1$ when the ith object is in the rth category of variable j

$g_{(j)ir} = 0$ when the ith object is not in the rth category of variable j

D_j diagonal matrix containing the univariate marginals, i.e. the column sums of G_j.

(b) For set k, k = 1,...,K

J(k) index set of the variables that belong to set k (so that we may write $j \in J(k)$)

m_k number of variables in set k (number of elements in J(k))

M_k binary, diagonal n x n matrix, with diagonal elements defined as

$m_{(k)ii} = 1$ when the ith observation is within the range $[1,k_j]$ for all $j \in J(k)$

$m_{(k)ii} = 0$ when the ith observation is outside $[1,k_j]$ for any $j \in J(k)$.

(d) The quantification matrices and parameter vectors are

X object scores, of order n x p

X_j auxiliary matrix of order n x p, with corrected object scores when fitting variable j

Y_j category quantifications for multiple variables, of order k_j x p

y_j category quantifications for single variables, of order k_j

a_j variable weights for single variables, of order p

Q_k quantified variables of the kth set, of order n x m_k, with columns $q_j = G_j y_j$

$\underline{Y}$ collection of multiple and single category quantifications across variables and sets.

Note: the matrices G_j, D_j, and M_k are exclusively notational devices; they are stored in reduced form, and the program fully profits from their sparseness by replacing matrix multiplications with selective accumulation.

(2) OBJECTIVE FUNCTION OPTIMIZATION

The OVERALS objective is to find object scores X and a set of $\underline{Y}_j$ (for j = 1,...,m) - the underlining indicates that they maybe restricted in various ways - so that the function

$$\sigma(X;\underline{Y}) = 1/K \sum_k \mathrm{tr}\,(X - \sum_{j \in J(k)} G_j\underline{Y}_j)'M_k(X - \sum_{j \in J(k)} G_j\underline{Y}_j)$$

is minimal, under the normalization restriction $X'M_*X = Kn\, I$, where the matrix $M_* = \sum_k M_k$, and I is the p x p identity matrix. The inclusion of M_k in $\sigma(X;\underline{Y})$ provides the following mechanism for weighting the loss: whenever any of the data values for object i in set k falls outside its particular range $[1,k_j]$, a circumstance which may indicate either genuine missing values or simulated missing values for the sake of analysis, all other data values for object i in set k are disregarded (listwise deletion per set). The diagonal of M_* contains the number of "active" sets for each object. The object scores are also centered, i.e. they satisfy $u'M_*X = 0$ with u denoting an n-vector with ones.

The following measurement levels are distinguished in OVERALS:

(a) multiple nominal: $\underline{Y}_j = Y_j$ / unrestricted

(b) single nominal: $\underline{Y}_j = y_j a_j'$ / rank-one and equality restrictions

(c) (single) ordinal: $\underline{Y}_j = y_j a_j'$ and $y_j \in C_j$ / rank-one and monotonicity restrictions

(d) (single) numerical: $\underline{Y}_j = y_j a_j'$ and $y_j \in L_j$ / rank-one and linearity restrictions.

For each variable these levels can be chosen independently. The general requirement in the "single" options is $\underline{Y}_j = y_j a_j'$, i.e. $\underline{Y}_j$ is of rank one; for identification purposes y_j is always normalized so that $y_j' D_j y_j = n$, which implies that the variance of the quantified variable $q_j = G_j y_j$ is 1. In the ordinal case the additional restriction $y_j \in C_j$ means that y_j must be located in the convex cone of all k_j-vectors with nondecreasing elements; in the numerical case the additional restriction $y_j \in L_j$ means that y_j must be located in the subspace of all k_j-vectors that are a linear transformation of the vector consisting of k_j successive integers (= normalized data vector).

Optimization is achieved by executing the following iteration scheme:

(i) Initialization I or II

(ii) Loop across sets and variables:

(iii) Eliminate contributions of other variables

(iv) Update category quantifications

(v) Update object scores

(vi) Orthonormalization

(vii) Convergence test: repeat (ii) - (vi) or continue

(viii) Rotation

Steps (i) to (viii) are explained below.

(i) Initialization

I. Random. The object scores X are initialized with random numbers, which are normalized so that $u'M_*X = 0$ and $X'M_*X = Kn\ I$, yielding $X^{\sim}$. For multiple variables the initial category quantifications are set equal to zero. For single variables the initial category

quantifications $y_j^{\sim}$ are defined as the first k_j successive integers normalized in such a way that $u'D_jy_j^{\sim} = 0$ and $y_j^{\sim\prime}D_jy_j^{\sim} = n$, and the initial variable weights are set equal to zero.

II. Nested. In this case the above iteration scheme is executed twice. In the first cycle (initialized with initialization I) all single variables are temporarily treated as single numerical, so that for the second, proper cycle all relevant quantities can be copied from the results of the first one.

(ii) Loop across sets and variables

The next two steps are repeated for k = 1,...,K and all $j \in J(k)$. During the updating of variable j all parameters of the remaining variables are fixed at their current values.

(iii) Eliminate contributions of other variables

For quantifying variable j in set k we define the auxiliary matrix

$$V_{(k)j} = \sum_{l \in J(k)} G_l\underline{Y}_l - G_j\underline{Y}_j,$$

which accumulates the contributions of the other variables in set k; then in $(X - V_{(k)j})$ the contributions of the other variables are eliminated from the object scores. This device enables us to write the loss $\sigma(X;\underline{Y})$ as a function of X and $\underline{Y}_j$ only:

$$\sigma(X;\underline{Y}_j) = \text{constant} + 1/K \text{ tr } ((X - V_{(k)j}) - G_j\underline{Y}_j)'M_k((X - V_{(k)j}) - G_j\underline{Y}_j).$$

With fixed current values $X^{\sim}$ the unconstrained minimum over $\underline{Y}_j$ is attained for the matrix

$$Y_j^{\sim} = (G_j'M_kG_j)^{-1}G_j'M_k(X^{\sim} - V_{(k)j}),$$

which forms the basis of the further computations. When switching to another variable *l* in the same set the matrix $V_{(k)l}$ is not computed from scratch, but updated:

$$V_{(k)l} \leftarrow V_{(k)j} + G_jY_j - G_lY_l.$$

(iv) Update category quantifications

(a) For multiple nominal variables the new category quantifications are simply

$$\underline{Y}_j^{+} = Y_j^{\sim}.$$

(b) For single variables one cycle of an ALS algorithm (De Leeuw et al., 1976) is executed for computing the rank-one decomposition of $Y_j^{\sim}$, with restrictions on the left-hand vector. This cycle starts from the previous category quantifications $y_j^{\sim}$ with

$$a_j^{+} = Y_j^{\sim\prime} D_j y_j^{\sim} .$$

When the current variable is numerical we are ready; else we compute

$$y_j^{*} = Y_j^{\sim} a_j^{+} .$$

Now, when the current variable is single nominal we can simply obtain y_j^{+} by normalizing y_j^{*} in the way indicated below; else the variable must be ordinal, and we have to insert the weighted monotonic regression process

$$y_j^{*} \leftarrow \mathrm{WMON}\,(y_j^{*}),$$

which makes y_j^{*} monotonically increasing. The weights used are the diagonal elements of D_j, and the subalgorithm used is the up-and-down-blocks minimum violators algorithm (Kruskal, 1964; Barlow et al., 1972). The result is normalized:

$$y_j^{+} = n^{1/2}\, y_j^{*}\, (y_j^{*\prime} D_j y_j^{*})^{-1/2} .$$

Finally, we set $\underline{Y}_j^{+} = y_j^{+} a_j^{+\prime}$.

(v) Update object scores

During the loop across sets the auxiliary score matrix W is accumulated as

$$W \leftarrow W + M_k \sum_{j \in J(k)} G_j \underline{Y}_j^{+}$$

and centered with respect to M_*:

$$X^{*} = \{I - (M_* uu' / u'M_* u)\}\, W.$$

From these two steps $M_*^{-1} X^{*}$ would yield the locally best update when there would be no orthogonality constraints.

(vi) Orthonormalization

The orthonormalization problem is to find an M_*-orthonormal X^+ that is closest to $M_*^{-1}X^*$ in the M_*-weighted least squares sense. In OVERALS this is done by setting

$$X^+ \leftarrow m^{1/2} M_*^{-1/2} \text{ PROCRU } (M_*^{-1/2} X^*),$$

The notation PROCRU( ) is used to denote the Procrustes orthonormalization process. If the singular value decomposition of the input matrix $M_*^{-1/2} X^*$ is denoted by $K\Lambda L'$, with $K'K = I$, $L'L = I$, and Λ diagonal, then the output matrix $KL' = M_*^{-1/2} X^* L\Lambda^{-1}L'$ satisfies orthonormality in the unit metric (Cliff, 1966). The matrix X^+ defined above satisfies orthonormality in the metric M_*. The calculation of L and Λ is based on tridiagonalization with Householder transformations followed by the implicit QL algorithm (Wilkinson, 1965).

(vii) Convergence test

The difference between consecutive values of tr Λ^4 is compared with the user specified convergence criterion ε - a small positive number. After convergence the badness-of-fit value $\sigma(X;\underline{Y}) = p - \text{tr } \Lambda^4$ is also given. Steps (ii) to (vi) are repeated as long as the loss difference exceeds ε.

(viii) Rotation

The OVERALS loss function $\sigma(X;\underline{Y})$ is invariant under simultaneous rotations of X and $\underline{Y}$. It can be shown that the solution is related to the principal axes of the average projection operator

$$Q_* = 1/K \sum_k M_k Q_k (Q_k' M_k Q_k)^{-1} Q_k' M_k .$$

In order to achieve principal axes orientation, which is useful for purposes of interpretation and comparison, it is sufficient to find a rotation matrix that makes the cross products of the matrix $M_*^{-1/2}X^*$ diagonal - a matrix identical to the one used in the Procrustes orthonormalization in step (vi). In the terminology of that section we rotate the matrices X^+, $\underline{Y}^+$, and the vectors a_j with the matrix L. The rotation matrix L is taken from the last PROCRU operation as described in (vi).

(3) DIAGNOSTICS

(a) Maximum rank

The maximum rank p_{max} indicates the maximum number of dimensions that can be computed for any data set (if exceeded, OVERALS adjusts the number of dimensions if possible and issues a message). In general we have:

$$p_{max} = \min \{(n - 1), r_1, r_2\} \quad \text{if} \quad K = 2$$
$$p_{max} = \min \{(n - 1), \Sigma_k r_k\} \quad \text{if} \quad K > 2 ,$$

where the quantities r_k are defined as

$$r_k = \Sigma_{j \in JM(k)} k_j + m_{k1} - m_{k2}.$$

Here m_{k1} is the number of multiple variables with no missing values in set k, m_{k2} the number of single variables in set k, and JM(k) an index set recording which variables are multiple in set k. Furthermore, OVERALS stops when either one of the following conditions is not satisfied:

(i) $r_k < n_k - 1$

(ii) $n_k > 2$

(iii) $\Sigma_k r_k \leq \Sigma_k (n_k - 1) - (n_{max} - 1)$.

Here n_k denotes the number of nonmissing objects in set k, and n_{max} the maximum across all of n_k.

(b) Marginal frequencies

The frequencies table gives the univariate marginals and the number of missing values (i.e., values that are regarded as out of range for the current analysis) for each variable. These are computed as the column sums of D_j and the total sum of M_k for $j \in J(k)$.

(c) Fit and loss measures

In the SUMMARY OF ANALYSIS loss and fit measures (i) and (ii) are reported:

(i) Loss per set. This is K times $\sigma(X;\underline{Y})$, partitioned with respect to sets and dimensions; the means per dimension are also given.

(ii) Eigenvalue. The values listed here are 1 minus the means per dimension defined above, forming a partitioning of FIT, which is p - $\sigma(X;\underline{Y})$ when convergence is reached. These quantities are the eigenvalues of Q_* defined in section (viii).

Other fir and loss measure reported are:

(iii) Multiple fit. This measure is computed as the diagonal of the matrix $\underline{Y}_j'D_j\underline{Y}_j$, computed for all variables (rows) with dimensions given in the columns.

(iv) Single fit. This table gives the squared weights, computed only for variables that are single. The sum reported is the sum of squares of the weights: $a_j'a_j$.

(v) Single loss. Single loss is equal to multiple fit minus single fit for single variables only. It is the loss incurred by the imposition of the rank-one and measurement level restrictions.

(d) Component loadings and quantifications

After the SUMMARY OF ANALYSIS the weights are reported; next the quantities:

(i) Component loadings for single variables. Loadings are the lengths of the projections of the quantified (single) variables onto the object space: $q_j'X$. When there are no missing data the loadings are equal to the correlations between the quantified variables and the object scores (the principal components).

(ii) Component loadings for multiple variables. In this case the loadings are computed for each of the multiple quantifications, i.e. each column of Y_j. For dimension s we have $(n\, y_{(j)s}'D_jy_{(j)s})^{-1/2}\, y_{(j)s}'G_j'X$.

Next the object scores are reported. The following results are arranged by variable:

(iiia) Category quantifications (either Y_j or y_j).

(iiib) Single coordinates. For single variables only: $\underline{Y}_j = y_ja_j'$.

(iv) Multiple coordinates. These are $Y_j^{\sim}$ defined in section 2(iii), i.e. the unconstrained minimizers of the loss function, for multiple variables equal to the category quantifications.

(v) Category centroids. The centroids of all objects that share the same category: $D_j^{-1}G_j'X$; note that they are not necessarily equal to the multiple coordinates.

(vi) Projected category centroids. For single variables only: $y_j b_j'$. These are the points on a line in the direction given by the loadings b_j that result from projection of the category centroids with weights D_j.

(4) PRECISION

All arrays are in single precision; the data and some of the simple statistics are kept in single word integer arrays.

(5) REFERENCES

Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972). Statistical Inference under order restrictions. New York: Wiley.

Cliff, N. (1966). Orthogonal rotation to congruence. Psychometrika, 31, 33-42.

De Leeuw, J.(1973). Canonical Analysis of Categorical Data. Doctoral dissertation, University of Leiden. Re-issued by DSWO Press (Leiden, 1984).

De Leeuw, J., Young, F.W., and Takane, Y. (1976). Additive structure in qualitative data: an alternating least squares method with optimal scaling features. Psychometrika, 41, 471-503.

Gifi, A. (1981). Non-linear Multivariate Analysis. Leiden: Department of Data Theory.

Kruskal, J.B. (1964). Nonmetric multidimensional scaling: a numerical method. Psychometrika, 29, 115-129.

Van de Geer, J.P. (1987). Algebra and geometry of OVERALS. Internal Report RR-87-13, Leiden, Department of Data Theory.

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

Van der Burg, E., De Leeuw, J. and Verdegaal, R. (1984). Non-linear canonical correlation analysis. Internal Report RR-84-12, Leiden, Department of Data Theory.

Van der Burg, E., De Leeuw, J. and Verdegaal, R. (1988). Homogeneity analysis with k sets of variables: an alternating least squares method with optimal scaling features. Psychometrika, 53, 177-197.

Verdegaal, R. (1986). OVERALS. Internal Report UG-86-01, Leiden, Department of Data Theory.

Wilkinson, J.H. (1965). The algebraic eigenvalue problem. Oxford: Clarendon Press.