OVERALS

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 ability 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 focuses on the relationships between sets; any particular variable contributes to the results only inasmuch as it provides information that is independent of the other variables in the same set.

Notation

The following notation is used throughout this chapter unless otherwise stated:

n	Number of cases (objects)
m	Total number of variables
p	Number of dimensions
K	Number of sets

For variable j, $j = 1, \ldots, m$

k_j	Number of categories (distinct values) of variable j
$\mathbf{G}_j$	Indicator matrix for variable j, of order $n \times k_j$

The elements of $\mathbf{G}_j$ are defined as $i = 1, \ldots, n;\ r = 1, \ldots, k_j$

$$g_{(j)ir} = \begin{cases} 1 & \text{when the } i\text{th object is in the } r\text{th category of variable } j \\ 0 & \text{when the } i\text{th object is not in the } r\text{th category of variable } j \end{cases}$$

$\mathbf{D}_j$	Diagonal matrix containing the univariate marginals; that is, the column sums of $\mathbf{G}_j$

For set $k,\ k = 1, \ldots, K$

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

m_k	Number of variables in set k (number of elements in $J(k)$)
$\mathbf{M}_k$	Binary, diagonal $n \times n$ matrix, with diagonal elements defined as

$$m_{(k)ii} = \begin{cases} 1 & \text{when the } i\text{th observation is within the range } [1, k_j] \text{ for all } j \in J(k) \\ 0 & \text{when the } i\text{th observation is outside the range } [1, k_j] \text{ for all } j \in J(k) \end{cases}$$

The quantification matrices and parameter vectors are:

$\mathbf{X}$	Object scores, of order $n \times p$
$\mathbf{X}_j$	Auxiliary matrix of order $n \times p$, with corrected object scores when fitting variable j
$\mathbf{Y}_j$	Category quantifications for multiple variables, of order $k_j \times p$: multiple category coordinates for multiple variables
$\mathbf{y}_j$	Category quantifications for single variables, of order k_j
$\mathbf{a}_j$	Variable weights for single variables, of order p
$\mathbf{Q}_k$	Quantified variables of the kth set, of order $n \times m_k$ with columns $\mathbf{q}_j = \mathbf{G}_j \mathbf{y}_j$
$\underline{\mathbf{Y}}$	Collection of multiple and single category quantifications across variables and sets

Note: The matrices $\mathbf{M}_k$, $\mathbf{G}_j$, $\mathbf{M}_j$, and $\mathbf{M}_k$ 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.

Objective Function Optimization

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

$$\sigma(\mathbf{X};\underline{\mathbf{Y}}) = 1/K \sum_k tr\left[\left(\mathbf{X} - \sum_{j \in J(k)} \mathbf{G}_j \underline{\mathbf{Y}}_j\right)' \mathbf{M}_k \left(\mathbf{X} - \sum_{j \in J(k)} \mathbf{G}_j \underline{\mathbf{Y}}_j\right)\right]$$

is minimal, under the normalization restriction $\mathbf{X}'\mathbf{M}_* \mathbf{X} = kn\mathbf{I}$, where the matrix $\mathbf{M}_* = \sum_k \mathbf{M}_k$, and $\mathbf{I}$ is the $p \times p$ identity matrix. The inclusion of $\mathbf{M}_k$ in $\sigma(\mathbf{X};\underline{\mathbf{Y}})$ 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 that 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 $\mathbf{M}_*$ contains the number of "active" sets for each object. The object scores are also centered; that is, they satisfy $\mathbf{u}'\mathbf{M}_* \mathbf{X} = 0$ with u denoting an n-vector with ones.

The following measurement levels are distinguished in OVERALS:

Multiple Nominal

$$\underline{\mathbf{Y}}_j = \mathbf{Y}_j \quad (\text{unrestricted})$$

Single Nominal

$$\underline{\mathbf{Y}}_j = \mathbf{y}_j \mathbf{a}'_j \quad (\text{rank - one and equality restrictions})$$

(Single) Ordinal

$$\underline{\mathbf{Y}}_j = \mathbf{y}_j \mathbf{a}'_j \text{ and } \mathbf{y}_j \in \mathbf{C}_j \quad (\text{rank - one and monotonicity restrictions})$$

(Single) Numerical

$$\underline{\mathbf{Y}}_j = \mathbf{y}_j \mathbf{a}'_j \text{ and } \mathbf{y}_j \in \mathbf{L}_j \quad (\text{rank - one and linearity restrictions})$$

For each variable, these levels can be chosen independently. The general requirement in the “single” options is $\underline{\mathbf{Y}}_j = \mathbf{y}_j \mathbf{a}'_j$; that is, $\underline{\mathbf{Y}}_j$ is of rank one; for identification purposes, $\mathbf{y}_j$ is always normalized so that $\mathbf{y}'_j \mathbf{D}_j \mathbf{y}_j = n$, which implies that the variance of the quantified variable $\mathbf{q}_j = \mathbf{G}_j \mathbf{y}_j$ is 1. In the ordinal case, the additional restriction $\mathbf{y}_j \in \mathbf{C}_j$ means that $\mathbf{y}_j$ must be located in the convex cone of all k_j-vectors with nondecreasing elements; in the numerical case, the additional restriction $\mathbf{y}_j \in L_j$ means that $\mathbf{y}_j$ 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:

1. Initialization I or II
2. Loop across sets and variables
3. Eliminate contributions of other variables
4. Update category quantifications
5. Update object scores
6. Orthonormalization
7. Convergence test: repeat (2)—(6) or continue
8. Rotation

Steps (1) through (8) are explained below.

(1) Initialization

I. Random

The object scores **X** are initialized with random numbers, which are normalized so that $\mathbf{u}'\mathbf{M}_*\mathbf{X} = 0$ and $\mathbf{X}'\mathbf{M}_*\mathbf{X} = Kn\mathbf{I}$, yielding $\tilde{\mathbf{X}}$. For multiple variables, the initial category quantifications are set equal to 0. For single variables, the initial category quantifications $\tilde{\mathbf{y}}_j$ are defined as the first k_j successive integers normalized in such a way that $\mathbf{u}'\mathbf{D}_j\tilde{\mathbf{y}}_j = 0$ and $\tilde{\mathbf{y}}_j\mathbf{D}_j\tilde{\mathbf{y}}_j = n$, and the initial variable weights are set equal to 0.

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 copies from the results of the first one.

(2) Loop across sets and variables

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

(3) Eliminate contributions of other variables

For quantifying variable *j* in set *k*, define the auxiliary matrix

$$\mathbf{V}_{(k)j} = \sum\nolimits_{j \in J(k)} \mathbf{G}_j \underline{\mathbf{Y}}_j - \mathbf{G}_j \underline{\mathbf{Y}}_j$$

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

$$\sigma\left(\mathbf{X}; \underline{\mathbf{Y}}_j\right) = \text{constant} + 1/K \operatorname{tr}\left[\left(\left(\mathbf{X} - \mathbf{V}_{(k)j}\right) - \mathbf{G}_j \underline{\mathbf{Y}}_j\right)' \mathbf{M}_k \left(\left(\mathbf{X} - \mathbf{V}_{(k)j}\right) - \mathbf{G}_j \underline{\mathbf{Y}}_j\right)\right]$$

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

$$\tilde{\mathbf{Y}}_j = \left(\mathbf{G}'_j \mathbf{M}_k \mathbf{G}_j\right)^{-1} \mathbf{G}'_j \mathbf{M}_k \left(\tilde{\mathbf{X}} - \mathbf{V}_{(k)j}\right)$$

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

$$\mathbf{V}_{(k)l} \leftarrow \mathbf{V}_{(k)j} + \mathbf{G}_j \underline{\mathbf{Y}}_j - \mathbf{G}_l \underline{\mathbf{Y}}_l$$

(4) Update category quantifications

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

$$\underline{\mathbf{Y}}_j^{+} = \underline{\tilde{\mathbf{Y}}}_j$$

(b) For single variables one cycle of an ALS algorithm (De Leeuw et al., 1976) is executed for computing the rank-one decomposition of $\tilde{Y}_j$, with restrictions on the left-hand vector.

This cycle starts from the previous category quantification $\tilde{\mathbf{y}}_j$ with

$$\mathbf{a}_j^{+} = \tilde{\mathbf{Y}}'_j \mathbf{D}_j \tilde{\mathbf{y}}_j$$

When the current variable is numerical, we are ready; otherwise we compute

$$\mathbf{y}_j^{*} = \tilde{\mathbf{Y}}_j \mathbf{a}_j^{+} \text{.}$$

Now, when the current variable is single nominal, you can simply obtain $\mathbf{y}_j^{+}$ by normalizing $\mathbf{y}_j^{*}$ in the way indicated below; otherwise the variable must be ordinal, and you have to insert the weighted monotonic regression process

$$\mathbf{y}_j^* \leftarrow WMON\left(\mathbf{y}_j^*\right)$$

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

$$\mathbf{y}_j^+ = n^{1/2}\mathbf{y}_j^*\left(\mathbf{y}_j^{*\prime}\mathbf{D}_j\mathbf{y}_j^*\right)^{-1/2}$$

Finally, we set

$$\underline{\mathbf{Y}}_j^+ = \mathbf{y}_j^+\mathbf{a}_j^{+\prime}$$

(5) Update object scores

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

$$\mathbf{W} \leftarrow \mathbf{W} + \mathbf{M}_k \sum\nolimits_{j\in J(k)} \mathbf{G}_j \underline{\mathbf{Y}}_j^+$$

and centered with respect to $\mathbf{M}_*$:

$$\mathbf{X}^* = \left\{\mathbf{I} - \mathbf{M}_*\mathbf{u}\mathbf{u}'/\mathbf{u}'\mathbf{M}_*\mathbf{u}\right\}\mathbf{W}$$

From these two steps, $\mathbf{M}_*^{-1}\mathbf{X}^*$ would yield the locally best update without orthogonality constraints.

(6) Orthonormalization

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

$$\mathbf{X}^{+} \leftarrow m^{1/2}\mathbf{M}_{*}^{-1/2}\mathrm{PROCRU}\left(\mathbf{M}_{*}^{-1/2}\mathbf{X}^{*}\right)$$

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

(7) 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 values $\sigma(\mathbf{X};\underline{\mathbf{Y}}) = p - \mathrm{tr}\left(\Lambda^4\right)$ is also given. Steps (2) through (6) are repeated as long as the loss difference exceeds ε.

(8) Rotation

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

$$Q_{*} = 1/K\sum_{k}\mathbf{M}_{k}\mathbf{Q}_{k}\left(\mathbf{Q}_{k}'\mathbf{M}_{k}\mathbf{Q}_{k}\right)^{-1}\mathbf{Q}_{k}'\mathbf{M}_{k}$$

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 $\mathbf{M}_{*}^{-1/2}\mathbf{X}^{*}$ diagonal — a matrix identical to the one used in the Procrustes orthonormalization in step (6). In the terminology of that section, we rotate the matrices $\mathbf{X}^{+}, \underline{\mathbf{Y}}^{+}$, and the vectors $\mathbf{a}_j$ with the matrix **L**. The rotation matrix **L** is taken from the last PROCRU operation as described in step (6).

Diagnostics

Maximum Rank

The maximum rank $\rho_{\max}$ 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,

$$\rho_{\max} = \begin{cases} \min\{(n-1), r_1, r_2\} & \text{if } K = 2 \\ \min\{(n-1), \max r_k\} & \text{if } K > 2 \end{cases}$$

where the quantities r_k are defined as

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

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

1. $r_k < n_k - 1$
2. $n_k > 2$
3. $\sum_k r_k \leq \sum_k (n_k - 1) - (n_{\max} - 1)$

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

Marginal Frequencies

The frequencies table gives the univariate marginals and the number of missing values (that is, values that are regarded as out of range for the current analysis) for each variable. These are computed as the column sums of $\mathbf{D}_j$ and the total sum of $\mathbf{M}_k$ for $j \in J(k)$.

Fit and Loss Measures

In the *Summary of Analysis*, loss and fit measures are reported.

Loss Per Set

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

Eigenvalue

The values listed here are 1 minus the means per dimension defined above, forming a partitioning of FIT, which is $\rho-\sigma(\mathbf{X};\underline{\mathbf{Y}})$ when convergence is reached. These quantities are the eigenvalues of $\mathbf{Q}_*$ defined in section (8).

Other fit and loss measures reported are:

Multiple Fit

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

Single Fit

This table gives the squared weights, computed only for variables that are single. The sum of squares of the weights: $\mathbf{a}'_j\mathbf{a}_j$.

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 measurement level restrictions.

Component Loadings and Qualifications

After the *Summary of Analysis*, the weights are reported, then the quantities.

Component Loadings for Single Variables

Loadings are the lengths of the projections of the quantified (single) variables onto the object space: $\mathbf{q}'_j\mathbf{X}$. When there are no missing data, the loadings are equal to the correlations between the quantified variables and the object scores (the principal components).

Category Quantifications (Either Y_j or y_j)

Single Coordinates

For single variables only: $\underline{\mathbf{Y}}_j = \mathbf{y}_j\mathbf{a}'_j$.

Multiple Coordinates

These are $\tilde{\mathbf{Y}}_j$ defined previously; that is, the unconstrained minimizers of the loss function, for multiple variables equal to the category quantifications.

Category Centroids

The centroids of all objects that share the same category, $\mathbf{D}_j^{-1}\mathbf{G}'_j\mathbf{X}$. Note that they are not necessarily equal to the multiple coordinates.

Projected Category Centroids

For single variables only, $\mathbf{y}_j\mathbf{b}'_j$. These are the points on a line in the direction given by the loadings $\mathbf{b}_j$ that result from projection of the category centroids with weights $\mathbf{D}_j$.

References

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. 1972. *Statistical inference under order restrictions*. New York: John Wiley & Sons, Inc.

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, reissued 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*, 31: 33–42.

Gifi, A. 1981. *Nonlinear 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*. Leiden: Internal Report RR-87–13, 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*. Leiden: Internal Report RR-84–12, 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*. Leiden: Internal Report UG-86–01, Department of Data Theory.

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