Optimal scaling of continuous numerical data

Jan de Leeuw

Dep of Data Theory
University of Leiden
Centraal Rekeninstituut
Wassenaarseweg 80

Leiden
The Netherlands

SUMMARY:

In this paper we define continuous numerical data, and discuss several algorithms to find optimal least squares quantifications of data of this type. The various algorithms are combined into a single algorithm that seems to be very efficient.

Jan de Leeuw
Recomputed Version

1: Introduction

An important problem that must be solved repeatedly in many scaling programs is the least squares regression problem

$$\min_{y \in C} (x - y)'(x - y).$$

The vector x in this problem is known, the set C is also known, and determined by the measurement characteristics of the data. If the data are ordinal, for example, we use monotone regression. The set C is a polyhedral convex cone, defined by

$$C = \{y \mid y_1 \leq y_2 \leq \ldots \leq y_n\}.$$

If the data are numerical we use linear regression. In linear regression C is defined by using a known n x p matrix U, and by specifying

$$C = \{y \mid y = Uz\},$$

for some p-element vector z. Clearly C is a linear subspace with dimension rank(U). Important special cases are ratio scale data (p = 1), and interval scale data with

$$C = \{y \mid y_i = z_1 u_i + z_2\},$$

for some pair z_1, z_2.

In the case of numerical data it follows from the definition of C that equal rows in U imply equal values of y. In the interval scale case, for example, we have $y_i = y_j$ if $u_i = u_j$. If the numerical data are categorical this restriction may be undesirable. In survey data, for example, many numerical variables occur, whose possible values are grouped into a small number of intervals. Income is a familiar example, and there are many others. In cases such as these it is usually too restrictive to represent all observations in a category by the same number. We want to represent categories by disjoint intervals of numbers, and the quantifications of individual observations can be chosen freely from the interval the observation is in.

In their discussion of optimal quantification DeLeeuw, Young, and Takane (1976) distinguish between the discrete and the continuous approach to quantifying categorical data. Representing all observations in a category by the same number is the discrete approach, representing categories by disjoint intervals of real numbers is the continuous approach. In the case of ordinal data the distinction reduces to that between the primary and secondary approach to ties discussed by Kruskal (1964). Primary is continuous, secondary is discrete. In the case of continuous numerical data DeLeeuw, Young, and Takane propose to define C by

$$C = \{y \mid u_i^- \leq y_i \leq u_i^+\},$$

where the u_i^- and u_i^+ are known numbers. If the data are categorical then the u_i^- and u_i^+ for observations in the same category are all equal, and we can write

$$C = \{y \mid u^-_{g(i)} \leq y_i \leq u^+_{g(i)}\},$$

where g(i) indicates which category observation i is in. Thus a category g is quantified as the known interval $[u^-_g, u^+_g]$. The regression problem for these continuous numerical restrictions is trivial, we simply set

$$y_i = \begin{cases} u_i^- & \text{if } x_i < u_i^-, \\ x_i & \text{if } u_i^- \leq x_i \leq u_i^+, \\ u_i^+ & \text{if } x_i > u_i^+. \end{cases}$$

The definition of continuous numerical data proposed by DeLeeuw, Young, and Takane is not very useful in generalizations of principal component analysis such as HOMALS (DeLeeuw, 1976). This is because the definition is the continuous version of the discrete restrictions

$$C = \{y \mid y_i = u_i\},$$

which is the type of restriction we use if the data are on an <u>identity scale</u>. In HOMALS we deal with variables which are defined on (at most) interval scales. In DeLeeuw (1976) a possible definition of a continuous interval scale was proposed. We set

$$C = \{y \mid z_1 u_i^- + z_2 \leq y_i \leq z_1 u_i^+ + z_2\},$$

with u_i^- and u_i^+ again known, $u_i^- \leq u_i^+$ for all i, but z_1 and z_2 unknown. The definition merely asserts that there exist z_1 and z_2, such that y is between the resulting bounds. Observe that in this case C is a polyhedral convex cone, while the DeLeeuw, Young, and Takane definition makes C a rectangle. For obvious reasons we have to restrict z_1 and z_2 in the definition by $z_1 u_i^- + z_2 \leq z_1 u_i^+ + z_2$ for all i, which is equivalent to $z_1 \geq 0$ if there is at least one i such that $u_i^- < u_i^+$. If $u_i^- = u_i^+$ for all i, then C becomes the same as in the discrete interval case, and the requirement $z_1 \geq 0$ is not needed any more.

2: <u>Decomposition</u>

The general regression problem for continuous numerical data is of the following form. We want to minimize

$$\sigma(y) = \tfrac{1}{2} \sum_{i=1}^{n} (x_i - y_i)^2$$

over all y in R^n and all z in Ω such that $Uz \leq y \leq Vz$. Here Ω is the subset of R^p defined by

$$\Omega = \{z \mid Uz \leq Vz\}.$$

We assume that Ω is nonempty. It is clear that the regression problem is a quadratic programming problem, but directly applying one of the standard quadratic programming methods seems undesirable, because the number of variables is $n + p$ and the number of constraints is $2n$, and this can both be very large. In HOMALS, for example, applications with n around 1000 are quite common. On the other hand if we consider z as fixed the problem becomes very easy to solve, even for large n, and the number of components of z is usually very small (ratio: $p=1$; interval: $p=2$; degree q polynomial: $p = q+1$). This makes it interesting to reformulate the problem as follows. We define

$$\phi(z) = \inf \{\sigma(y) \mid Uz \leq y \leq Vz\}.$$

By the usual conventions this means that $\phi(z) = +\infty$ if $z \notin \Omega$. If $z \in \Omega$ then we know how to compute $\phi(z)$ from our discussion in the previous section.

Clearly the regression problem we are studying is equivalent to the problem of minimizing $\phi(z)$ over R^p, which is also equivalent to the problem of minimizing $\phi(z)$ over Ω. In the next sections we shall prove some interesting properties of $\phi(z)$, that will lead directly to efficient algorithms for solving the regression problem. Some of these properties could easily be derived from the general theory of convex analysis, but we have chosen to give proofs that are completely elementary and constructive.

We first give some useful definitions and some notations, and then derive a simple preliminary result. Suppose $z \in \Omega$. Define the following index sets.

$$I_1(z) = \{i \mid x_i < u_i(z)\},$$
$$I_2(z) = \{i \mid x_i = u_i(z)\},$$
$$I_3(z) = \{i \mid u_i(z) < x_i < v_i(z)\},$$

$I_4(z) = \{i \mid x_i = v_i(z)\}$,

$I_5(z) = \{i \mid x_i > v_i(z)\}$.

Here $u_i(z)$ is element i of Uz, and $v_i(z)$ is element i of Vz. ~~Observe that if $z \notin \Omega$ these index sets need no longer be disjoint~~. For computational purposes we represent subsets of $\{1,\ldots,n\}$ as n x n diagonal matrices, with diagonal element i equal to unity if i is in the subset, and equal to zero otherwise. This defines the diagonal matrices $M_1(z),\ldots,M_5(z)$ corresponding with the index sets $I_1(z),\ldots,I_5(z)$. We also define the function

$\psi(z) = \frac{1}{2}(x - Uz)'M_1(z)(x - Uz) + \frac{1}{2}(x - Vz)'M_5(z)(x - Vz)$.

By solving the regression problem for fixed z, and by substituting the solution $\hat{y}(z)$ in $\sigma(y)$ we find the following result.

Theorem 1: If $z \in \Omega$ then $\psi(z) = \phi(z)$.

Observe that if $z \notin \Omega$ then $\psi(z)$ is finite, while $\phi(z) = +\infty$. This property sometimes makes it more convenient to study the function $\psi(z)$ in stead of $\phi(z)$.

3: A local result

Suppose we replace z by $z + \delta$, where δ is supposed to be small. How does such a perturbation affect $\psi(z)$ or $\phi(z)$? To answer this question we define more index sets.

$J_1(\delta) = \{i \mid u_i(\delta) < 0\}$,

$J_2(\delta) = \{i \mid u_i(\delta) = 0\}$,

$J_3(\delta) = \{i \mid u_i(\delta) > 0\}$,

$J_4(\delta) = \{i \mid v_i(\delta) < 0\}$,

$J_5(\delta) = \{i \mid v_i(\delta) = 0\}$,

$J_6(\delta) = \{i \mid v_i(\delta) > 0\}$.

The corresponding diagonal matrices indicating these subsets are $N_1(\delta),\ldots,N_6(\delta)$.

Lemma 1: If δ is "sufficiently small" then

$I_1(z + \delta) = I_1(z) \cup \{I_2(z) \cap J_3(\delta)\}$,

$I_5(z + \delta) = I_5(z) \cup \{I_4(z) \cap J_6(\delta)\}$.

The proof is simple, and is omitted. The expression "sufficiently small" can of course be defined rigorously in several different ways.

Theorem 2: If δ is "sufficiently small", then

$$\psi(z + \delta) = \psi(z) - \delta'\{U'M_1(z)(x - Uz) + V'M_5(z)(x - Vz)\} +$$
$$+ \tfrac{1}{2}\,\delta'\{U'M_1(z + \delta)U + V'M_5(z + \delta)V\}\delta.$$

Proof: Substitute $z + \delta$ in the definition of $\psi(z)$. Use the results of lemma 1 in the form

$$M_1(z + \delta) = M_1(z) + \{M_2(z).N_3(\delta)\},$$
$$M_5(z + \delta) = M_5(z) + \{M_4(z).N_4(\delta)\},$$

where the dot indicates elementwise multiplication. Then simplify by using

$$M_2(z)(x - Uz) = 0,$$
$$M_4(z)(x - Vz) = 0.$$

This gives the required formula.//

Corollary 3:1: The function ψ is continuously differentiable, the function ϕ is continuously differentiable in $\text{int}(\Omega)$.

3:2: The gradient of both ψ and ϕ at z, if it exists, is $g(z) = U'M_1(z)(Uz - x) + V'M_5(z)(Vz - x)$.

3:3: The function ψ is almost everywhere twice continuously differentiable, for ϕ the same thing is true in $\text{int}(\Omega)$.

3:4 The Hessian of both ψ and ϕ at z, if it exists, is equal to $H(z) = U'M_1(z)U + V'M_5(z)V$.

Proof: Almost everything follows directly from theorem 2. Continuity of the gradient follows from continuity of $M_1(z)(Uz - x)$ and $M_5(z)(Vz - x)$. The set where both $M_2(z) \neq \emptyset$ and $M_4(z) \neq \emptyset$ is of measure zero. If $M_2(z)$ and $M_4(z)$ are both empty, then $H(z)$ is continuous at z. //

4: A global result

For perturbations which are not assumed to be small we can derive useful and interesting inequality estimates. Define

$$R(z,\delta) = \psi(z + \delta) - \psi(z) - \delta'g(z),$$

and also define the new index sets

$L_1(z,\delta) = I_1(z+\delta) \cap I_1(z)$,

$L_2(z,\delta) = I_1(z+\delta) \cup I_1(z)$,

$L_3(z,\delta) = I_5(z+\delta) \cap I_5(z)$,

$L_4(z,\delta) = I_5(z+\delta) \cup I_5(z)$.

The diagonal indicator matrices corresponding with these index sets are

$O_1(z,\delta),\ldots,O_4(z,\delta)$.

Theorem 1: For all z and δ in R^p it is true that

$$\tfrac{1}{2}\,\delta'U'O_1(z,\delta)U\delta + \tfrac{1}{2}\,\delta'V'O_3(z,\delta)V\delta \leq R(z,\delta) \leq \tfrac{1}{2}\,\delta'U'O_2(z,\delta)U\delta + \tfrac{1}{2}\,\delta'V'O_4(z,\delta)V\delta.$$

Proof: For the proof we need four more index sets.

$K_1(z,\delta) = I_1(z+\delta) - I_1(z)$,

$K_2(z,\delta) = I_1(z) - I_1(z+\delta)$,

$K_3(z,\delta) = I_5(z+\delta) - I_5(z)$,

$K_4(z,\delta) = I_5(z) - I_5(z+\delta)$.

By using the definitions we find that $R(z,\delta)$ is the sum of the following six terms.

Term 1: $\tfrac{1}{2}\sum\{ (x_i - u_i(z) - u_i(\delta))^2 \mid i \in K_1(z,\delta) \}$.

Term 2: $\tfrac{1}{2}\sum\{ (x_i - v_i(z) - v_i(\delta))^2 \mid i \in K_3(z,\delta) \}$.

Term 3: $\tfrac{1}{2}\sum\{-(x_i - u_i(z) - u_i(\delta))^2 \mid i \in K_2(z,\delta) \}$.

Term 4: $\tfrac{1}{2}\sum\{-(x_i - v_i(z) - v_i(\delta))^2 \mid i \in K_4(z,\delta) \}$.

Term 5: $\tfrac{1}{2}\sum\{ u_i(\delta)^2 \mid i \in I_1(z) \}$.

Term 6: $\tfrac{1}{2}\sum\{ v_i(\delta)^2 \mid i \in I_5(z) \}$.

We can bound the first four terms by using the implications:

$i \in K_1(z,\delta)$ implies $-u_i(\delta) \leq x_i - u_i(z) - u_i(\delta) \leq 0$,

$i \in K_2(z,\delta)$ implies $0 \leq x_i - u_i(z) - u_i(\delta) \leq -u_i(\delta)$,

$i \in K_3(z,\delta)$ implies $0 \leq x_i - v_i(z) - v_i(\delta) \leq -v_i(\delta)$,

$i \in K_4(z,\delta)$ implies $-v_i(\delta) \leq x_i - v_i(z) - v_i(\delta) \leq 0$.

This implies

$$0 \leq \text{term 1} \leq \tfrac{1}{2}\sum\{ u_i(\delta)^2 \mid i \in K_1(z,\delta)\},$$

$$0 \leq \text{term 2} \leq \tfrac{1}{2}\sum\{ v_i(\delta)^2 \mid i \in K_3(z,\delta)\},$$

$$-\tfrac{1}{2}\sum\{ u_i(\delta)^2 \mid i \in K_2(z,\delta)\} \leq \text{term 3} \leq 0,$$

$$-\tfrac{1}{2}\sum\{ v_i(\delta)^2 \mid i \in K_4(z,\delta)\} \leq \text{term 4} \leq 0.$$

If we add these inequalities, and also add term 5 and term 6, we find the required result. //

Corollary 2: Both ϕ and ψ are convex.

Proof: This follows directly from $R(z,\delta) \geq 0$. //

It is of some interest to find out when the inequalities in theorem 1 become equalities. By checking the proof of theorem 1 we find the following result.

Theorem 3: Necessary and sufficient for equality in theorem 1 (both upper and lower) is that the indicated cells in the following intersection table are empty:

	$I_1(z+\delta)$	$I_2(z+\delta)$	$I_3(z+\delta)$	$I_4(z+\delta)$	$I_5(z+\delta)$
$I_1(z)$	†	†	$\emptyset$	$\emptyset$	$\emptyset$
$I_2(z)$	†	†	†	†	$\emptyset$
$I_3(z)$	$\emptyset$	†	†	†	$\emptyset$
$I_4(z)$	$\emptyset$	†	†	†	†
$I_5(z)$	$\emptyset$	$\emptyset$	$\emptyset$	†	†

By using theorem 3 in conjunction with lemma 1 from the previous section we find the interesting result that there is double equality in theorem 1 if and only if δ is "sufficiently small", in the sense defined in the previous section.

5: An optimistic algorithm

The definition of $\psi(z)$ in section 2 shows that $\psi(z)$ is piecewise quadratic, where the pieces are defined by the index sets $I_1(z)$ and $I_5(z)$. This suggests the following simple minded algorithm: in each iteration we find the successor of z by minimizing the quadratic function $\psi(z)$ is equal to in a neighborhood of z. The formula is

$$z^+ = H^{-1}(z)t(z),$$

with

$$t(z) = U'M_1(z)x + V'M_5(z)x.$$

The results in section 3 show that the algorithm also has a more reasonable

interpretation: the quadratic pieces of $\psi(z)$ are joined in a smooth way, and the algorithm takes Newton-Raphson steps "almost everywhere". Nevertheless there are several ways in which the algorithm can go wrong. We illustrate this with a small example. Let U and V be defined as the 5 x 2 matrices

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 12 & 1 \end{bmatrix}$$

Let x = 1 2 3 5 4, and let z = 1 0. Then

$I_1(z) = \{5\}$,

$I_2(z) = \{1,2,3\}$,

$I_3(z) = \{4\}$,

$I_4(z) = \{1,2\}$,

$I_5(z) = \emptyset$.

Thus t(z) = 1 1, and H(z) is

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Thus H(z) is singular, the iteration cannot be carried out. Observe that in this case $g(z) = H(z)z - t(z)$ is equal to zero, so we would have stopped anyway. In practical applications singularity of H(z) will be very rare indeed, but a reliable algorithm must have provisions for the possibility.

In the second place the algorithm is not stable, in the sense that it does not necessarily decrease the loss function $\psi(z)$. If we use the same U and V, let x = 3 2 4 5 1 and z = 0 3, then $\psi(z) = 5$. One iteration of our algorithm gives z = .55 .39, with $\psi(z) = 5.20$. The loss has increased. A second iteration gives z = .29 1.83, with $\psi(z) = 3.98$ and g(z) = 0 0. Although stability is not a necessary property of a good algorithm, it certainly is desirable.

In the third place the iterations may move z out of Ω, temporarily or permanently. With the same U and V, with x = 2 4 1 3 5, with z = .28 1.81, we find $\psi(z) = 2.69$. One iteration gives z = -.5 3.33, and $\psi(z) = 12.15$.

The next iteration has z = .23 1.51 and $\psi(z) = 3.07$, and the third iteration gives convergence at z = .25 1.89 with $\psi(z) = 2.66$. Observe that the final solution is very close to the initial one, but the intermediate iterations have taken us very far away from the solution. Observe also that, although something strange and undesirable happens in all three examples, we still have rapid convergence to the correct solution.

A fourth property of the algorithm is much more serious. Again U and V are as before, x = 1 2 5 3 4, z = 1 0, and $\psi(z) = 1$. One iteration gives z = 1 -1, and $\psi(z) = 1.5$. The next iteration gives z = 1 0, and $\psi(z) = 1$. Consequently the algorithm cycles, and can never get to the correct solution.

On the other hand the algorithm works extremely well if it works at all. Convergence usually takes only three or four iterations. And moreover the exceptional cases we have discussed turn out to be extremely rare with real data sets. Our algorithm is too optimistic, in the sense that the quadratic approximation is extended to regions where it may lead us astray, but it is also very practical and very economical. We do not want to throw it away, we only have to build in safeguards that prevent the undesirable possibilities.

6: A pessimistic algorithm

The algorithm in this section is based on the following simple result, which we give without proof.

Lemma 1: For all z,δ in R^p, and for all $\lambda \geq 0$,

$L_2(z,\lambda\delta) \subseteq I_1(z) \cup J_3(\delta)$,

$L_4(z,\lambda\delta) \subseteq I_5(z) \cup J_4(\delta)$.

The diagonal indicators of the two sets on the right are $P_1(z,\delta)$ and $P_2(z,\delta)$. From lemma 1 and the results of section 4 we obtain a useful inequality.

Lemma 2: For all z,δ in R^p, and for all $\lambda \geq 0$,

$$\psi(z + \lambda\delta) \leq \psi(z) + \lambda\delta' g(z) + \tfrac{1}{2}\lambda^2 \delta' U' P_1(z,\delta) U\delta + \tfrac{1}{2}\lambda^2 \delta' V' P_2(z,\delta) V\delta.$$

Observe that we can assume without loss of generality that not both $P_1(z,\delta)$ and $P_2(z,\delta)$ are zero. If they were, then $I_1(z)$ and $I_5(z)$ would be empty, and $\psi(z)$ would be zero. It follows that if $g(z) \neq 0$ we can always choose δ such that $\delta' g(z) < 0$, and such that the number λ_* that minimizes the quadratic in lemma 2 also satisfies $\lambda_* > 0$. In fact it follows from lemma 2 that if $\delta' g(z) < 0$, then $\delta'\{U' P_1(z,\delta) U + V' P_2(z,\delta) V\}\delta$ must be positive. If it were zero, then we could make $\psi(z + \lambda\delta) < 0$ by choosing $\lambda > \psi(z)/|\delta' g(z)|$, which is absurd. A possible choice for δ is, of course, $\delta = -g(z)$.

Theorem 3: The algorithm

$$z^+ = z - \lambda_* g(z)$$

converges to a point $\hat{z}$, with $g(\hat{z}) = 0$.

Proof: It follows from lemma 2 and the definition of λ_* that $\psi(z^+) \leq \psi(z)$, with equality if and only if $g(z) = 0$. The continuity of the gradient guarantees convergence to a stationary point, which is of course a minimum by convexity. //

We now have an optimistic algorithm, which is extremely fast in almost all applications, but which may fail in some. We also have a pessimistic algorithm which always works, but turns out to be very slow. The estimates in lemmaas 1 and 2 are too crude, and convergence often takes more than 100 iterations.

7: A first compromise

An obvious next step is to combine the two procedures into a single algorithm. We try an optimistic move in each iteration, if this does not decrease the loss (or if something else goes wrong) we make a pessimistic step. In the end the pessimistic steps will bring us into a region where optimism works. Because the total number of optimistic steps is always finite it follows that the combined algorithm converges monotonically. Tests with the combined algorithm show that its performance is satisfactory, but not perfect. If we are forced to use the pessimistic approach it often takes around 10 to

20 iterations before optimism works again. Although the pessimistic iterations are simple this does take time.

8: A local algorithm

The optimistic algorithm was based on the fact that the loss function is locally quadratic. It is optimistic because it acts as if the local quadratic approximation (which is exact in a neighborhood of the current point) is exact through the whole space. The pessimistic algorithm uses a majorization approach. It constructs a quadratic function which is an overestimate of the loss function along a line. An alternative approach is to use a local quadratic approximation, and to compute the region in which this approximation is exact. The necessary theory has been explained in section 3. The results there can be used especially nicely if we construct the local approximation along the line.

We consequently choose a direction δ, and define λ_o as the maximum value of λ which satisfies

$$I_1(z + \lambda\delta) = I_1(z) \cup \{I_2(z) \cap J_3(\delta)\},$$
$$I_5(z + \lambda\delta) = I_5(z) \cup \{I_4(z) \cap J_4(\delta)\}.$$

The procedure for computing λ_o is perhaps best described as a sort of computer program.

set i = 1 and $\lambda_o = \infty$;

s0: if $x_i \geq u_i(z)$ then goto s1;
else if $u_i(\delta) \geq 0$ then goto s5;
else $\lambda_o = \min\{\lambda_o, (x_i - u_i(z))/u_i(\delta)\}$;

s1: if $x_i > u_i(z)$ then goto s2;
else if $v_i(\delta) \geq 0$ then goto s5;
else $\lambda_o = \min\{\lambda_o, (x_i - v_i(z))/v_i(\delta)\}$;

s2: if $x_i \geq v_i(z)$ then goto s3;
else if $v_i(\delta) \geq 0$ then goto s2a;
else $\lambda_o = \min\{\lambda_o, (x_i - v_i(z))/v_i(\delta)\}$;

s2a: if $u_i(\delta) \leq 0$ then goto s5;
else $\lambda_o = \min\{\lambda_o, (x_i - u_i(z))/u_i(\delta)\}$;
goto s5;

s3: if $x_i > v_i(z)$ then goto s4;
    else if $u_i(\delta) \leq 0$ then goto s5;
    else $\lambda_o = \min \{\lambda_o, (x_i - u_i(z))/u_i(\delta)\}$;
s4: if $v_i(\delta) \leq 0$ then goto s5;
    else $\lambda_o = \min \{\lambda_o, (x_i - v_i(z))/v_i(\delta)\}$;
s5: if i = n then STOP;
    else i = i + 1;
    goto s0;

After determining λ_o we also determine λ_+, which is the value of λ that minimizes the quadratic approximation in theorem 2 of section 3 along the line $z + \lambda\delta$. We then set $\lambda_* = \min(\lambda_o, \lambda_+)$, and $z^+ = z + \lambda_*\delta$. Although this local algorithm is slightly more complicated than the pessimistic algorithm, it turns out to be considerably faster. In our implementations we have used the direction $H^+(z)g(z)$, with $H^+(z)$ the Moore-Penrose inverse of $H(z)$. We have only implemented the interval scale case, with $p = 2$, and with obvious modifications in the algorithm to guarantee that $z_1 \geq 0$. For a more precise description of this algorithm we refer to Walter (1976).

Notes:

De Leeuw, J. HOMALS
Paper presented at the spring meeting of the Psychometric Society, Murray Hill, N.J., april 1976.

Walter, J. CONNUM: a FORTRAN subroutine for quantifying continuous numerical data.
Unpublished Masters Thesis, Dept. of Methodology, Univ. of Leiden.

References:

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

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