Algebraic Properties of Weighted Distance Models (Young & Hamer, Chapter 6)

The General Euclidean Model is:

$$d_{ijk}^2 = (y_i - x_j)V_i W_k (y_i - x_j)'$$

where:

d_{ijk}^2 is the squared GEM distance between points i and j as modified by weights V_i and W_j.

x_i, y_j are r-element row vectors of coordinates of points in an r-dimensional Euclidean space. The y_i are restricted to equal the x_i when the data are square.

V_i is a square, symmetric, order r, rank $s_i \leq r$, positive semi-definite matrix of weights associated with data row object i.

W_k is a square, symmetric, order r, rank $t_k \leq r$, positive semi-definite matrix of weights associated with data matrix k.

Assumptions Leading to Special Cases

GEM is too general to be useful in MDS without certain simplifying assumptions.

It is, however, very useful as a meta-model for MDS, since it includes many specific MDS models as special cases.

For certain kinds of data GEM must be simplified.

1. For two way data k=1 and it must be the case that $W_k = I$.
2. For square data it must be the case that Y=X.
3. For symmetric data it must be the case that all $V_i = I$.

There are simplifying assumptions that can be made about the nature of the weights V_i and W_k. Each of these matrices may be:

1. An identity matrix
2. A diagonal, non-identity matrix
3. A symmetric, rank one matrix
4. A symmetric, reduced rank matrix
5. A symmetric, full rank matrix

Each of these special cases yields interesting special models that have been proposed separately by various investigators.

Analyses using Two-Way GEMs

Since the data are two-way, we must necessarily assume

1 $W_1 = I$

The standard (i.e., symmetric, unweighted) Euclidean Model for analyzing square data results from also assuming

2 $Y = X$

3 $V_i = I$ for all i

The asymmetric (unweighted) Euclidean Model for analyzing asymmetric data results from assuming:

2 $Y = X$

3 V_i= diagonal for all i

Rectangular Euclidean Models for analyzing rectangular data :

2 $Y \neq X$.

3a $V_i = I$ for all i – Classical unfolding

3b V_i= diagonal or reduced rank for all i – Carroll's family of unfolding models

Analyses using Three-Way GEMs

There are 40 different models in the family of three-way GEMs. These result from the factorial combination of (a) the 2 assumptions about Y; (b) the 5 assumptions about V_i; (c) 4 of the 5 assumptions about W_k.

Of these, 20 are for square data, and 20 are for rectangular data.

Four models using Symmetric Three-way GEMs:

There are four models which differ in their assumptions about W_i.
All four make the assumption that
1 Y=X
2 V_i=I
The four models differ as to whether
3 W_k is diagonal, rank-one, reduced rank or full rank.

There are 8 analyses that can be performed with these 4 models, 4 for symmetric data and 4 for asymmetric data.

Sixteen models using Asymmetric Three-way GEMs:

There are sixteen models which differ in their assumptions about W_i.
All sixteen make the assumption that
1 Y=X
There are then four families of models that differ as to whether
2 V_i= is diagonal, rank-one, reduced rank or full rank
which are factorially combined with four assumption as to whether
3 W_k is diagonal, rank-one, reduced rank or full rank.

The 20 rectangular models parallel the 20 symmetric models, but all assume that $Y \neq X$.

Algebra of Diagonal Weight Matrices

Recall that the General Euclidean Model is:

$$d_{ijk}^2 = (y_i - x_j)V_i W_k (y_i - x_j)' \quad .$$

In this section we vertically adjoin (concatenate) the two stimulus matrices X and Y into a single stimulus supermatrix Z:

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix} \; .$$

We define an (r x r) column orthonormal rotation matrix T, where T'T=I.

We show that if we orthogonally rotate the stimulus space Z to obtain the rotated stimulus space Z*=ZT we have violated the basic diagonality condition of the model, unless at least one product $V_i W_k$ is diagonal.

First we see that

$$d_{ijk}^2 = (z_i^* - z_j^*)V_i W_k (z_i^* - z_j^*)' \quad ,$$

$$d_{ijk}^2 = (z_i T - z_j T)V_i W_k (z_i T - z_j T)' \quad ,$$

$$d_{ijk}^2 = (z_i - z_j)T V_i W_k T'(z_i - z_j)' \quad .$$

Then we can define $U_{ik} = V_i W_k$ so that

$$d_{ijk}^2 = (z_i - z_j)T U_{ik} T'(z_i - z_j)' \quad \text{becomes}$$

$$d_{ijk}^2 = (z_i - x_j)U^*_{ik}(z_i - z_j)' \quad .$$

We note that U^*_{ik} is not diagonal unless $V_i W_k$ is diagonal. So rotation is not possible without violating a basic tenent of the model.

Algebra of Non-Diagonal Weight Matrices

Once again, recall that the General Euclidean Model is:

$$d_{ijk}^2 = (y_i - x_j)V_iW_k(y_i - x_j)' \quad .$$

We now consider the case where the weight matrices V_i and W_k are (r x r) positive semi-definite (p.s.d.) matrices whose rank, as set by the data analyst, are some known values t_i (for V_i) and t_k (for W_k).

The rank for each matrix is said to be the “principal directions” for that data source.

We simplify the discussion by once again considering the matrix of combined weights $U_{ik} = V_iW_k$, which must be p.s.d, given the conditions that V_i and W_k are p.s.d.

Orthogonal Decomposition: Since U_{ik} is p.s.d., it is the case that it may be orthogonally decomposed into

$U_{ik} = P_{ik}O_{ik}P'_{ik}$, where

P_{ik} is a (r x t_{ik}) column orthonormal matrix of eigenvectors of U_{ik} ($U_{ik}U_{ik}$=I) and O_{ik} is a (t_{ik} x t_{ik}) diagonal matrix of eigenvalues of U_{ik}.

Then

$$d_{ijk}^2 = (y_i - x_j)P_{ik}O_{ik}P'_{ik}(y_i - x_j)' \quad .$$

Algebra of Non-Diagonal Weight Matrices (continued)

Then we define the coefficients of the principal directions as

$C_{ik} = P_{ik} O_{ik}^{1/2}$, so that

$$d_{ijk}^2 = (y_i - x_j) C_{ik} C'_{ik} (y_i - x_j)' \quad .$$

We can now define the transformed (personal) stimulus space for the combined data sources i and k as the (n+m by t_{ik}) matrix of coordinates of the principal directions Z_{ik}, where

$$Z_{ik} = Z C_{ik} = \begin{bmatrix} X \\ Y \end{bmatrix} C_{ik} \ .$$

We can, alternatively, express the General Euclidean Model as

$$d_{ijk}^2 = (y_i^{ik} - x_j^{ik})(y_i^{ik} - x_j^{ik})' \ ,$$

showing that the stimulus space, as transformed by the combined weights for sources i and k, is a (unweighted) Euclidean space.

Note that Z_{ik}, which is (n+m by t_{ik}), may be lower dimensional than Z, which is (n+m by r).

Note also that, unlike diagonal models, the dimensions of the stimulus space Z *can* be rotated without violating the p.s.d. nature of the model.

Oblique Representation of GEM

We begin by defining a matrix S_{ik} which is the diagonal of the combined weight matrix for data sources i and k:

$$S_{ik} = \text{diag}(U_{ik}) = \text{diag}(V_i W_k) \quad .$$

This permits defining the stimulus space as

$$Z_{ik} = Z S_{ik} \ .$$

Since S_{ik} is the diagonal of the combined weights, Z_{ik} can be interpreted as the *obliquely* transformed space for the combined data sources that results from a simple stretching and shrinking of the stimulus space dimensions. No rotations or translations, etc. are permitted. All Z_{ik} have dimensions that correspond to the dimensions of Z, and they should be interpretable.

We can also define

$$R_{ik} = S_{ik}^{-1/2} U_{ik} S_{ik}^{-1/2} \quad ,$$

which is a matrix of correlations (cosines of angles) between the dimensions of the obliquely transformed stimulus space Z_{ik}.

We have decomposed U_{ik} so that

$U_{ik} = S_{ik}^{1/2} R_{ik} S_{ik}^{1/2}$, which is an oblique decomposition. This is an alternative to the orthogonal decomposition presented above.