STEP-BY-STEP PROCEDURE FOR ASSN. 3: TORGERSON'S METRIC MDS ALGORITHM

PART A -- Transform the similarity data to obtain "pseudodistances".

The following matrix show the confusion probabilities among pairs of five Morse code signals (for E,H,N,S,W) obtained by Rothkopf (1957).

	E	H	N	S	W
E	97	04	04	07	02
H	09	87	08	37	09
N	08	16	93	12	12
S	11	59	17	96	12
W	09	15	26	12	86

Starting with these probabilities, transform the confusions into dissimilarities that satisfy the axioms of a metric space (="pseudodistances"). This process involves the following steps 1-4:

1. symmetrize the matrix (by averaging the corresponding entries in the upper & lower halves of the matrix), and write it out as a lowerhalf matrix.

2. transform the similarities into dissimilarities (by subtracting each entry from the largest similarity).

3. for every triple of objects, check if the triangle inequality is satisfied (theoretically, it must be satisfied for all possible permutations of the three points: d(x,y)+d(y,z) ≥ d(x,z), d(x,z) + d(z,y) ≥ d(x,y), d(y,x)+d(x,z)≥d(y,z), but a little thought can save you a lot of checking here). If the TI is not satisfied, then find the largest violation, C = d(x,z) - (d(x,y)+d(y,z)), and add this constant C to each of the dissimilarities. Verify that the TI is now EXACTLY satisfied for the triple that gave you the largest violation.

4. write out the resulting matrix as a full matrix (putting 0's in the diagonal). These numbers now satisfy the metric axioms; i.e., they are "pseudodistances".

PART B. transform the "pseudodistances" into "pseudo scalar products" (as follows)

5. square each entry in this matrix to get d_{ij}^2.

6. "double-center" this symmetric matrix using the formula:

$$b_{ij}^* = -\frac{1}{2}\left[d_{ij}^2 - \frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 - \frac{1}{n}\sum_{j=1}^{n} d_{ij}^2 + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2 \right]$$

(include the diagonal entries). The resulting matrix may be thought of as "pseudo scalar products".

PART C. Factor the pseudo-scalar products using PCA.

7. Run a principal components analysis (PCA) of this matrix, treating it as covariances, and requesting no rotation of the factor solution. Do a "scree plot" of the size of each eigenvalue for the five components. How many dimensions appear to approximately characterize the data?

8. Plot the five points in a 2-dimensional space using the "component loadings" for the first two dimensions (label each point appropriately). In a sentence or two, compare your configuration to that obtained by Shepard (and reported in Kruskal & Wish).