Solving the Thermometer Problem

Let **T** be the p by q matrix of thermometer scores where i=1,…,p is the number of respondents and j=1,…,q is the number of political/social stimuli receiving ratings. **T** can be regarded as a matrix of inverse distances between the respondents and the stimuli. Specifically, apply the linear transformation:

$$\mathbf{d}_{\mathbf{ij}}^{*} = \left(\frac{100 - \mathbf{T}}{50}\right) = \left(2 - 0.02\mathbf{T}\right) = \mathbf{d}_{\mathbf{ij}} + \varepsilon_{\mathbf{ij}} \qquad (1)$$

Where the observed data are now *noisy* distances that range from zero to two, that is, $0 \le \mathbf{d}_{\mathbf{ij}}^{*} \le 2$, which are assumed to be equal to some true distance plus a random error term -- $\mathbf{d}_{\mathbf{ij}} + \varepsilon_{\mathbf{ij}}$. This transformation is convenient because it tends to confine the estimated respondent and stimuli points to a unit hypersphere.

Recall that our p by s matrix of individual (respondent) coordinates is:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_{11} & \mathbf{x}_{12} & . & . & . & \mathbf{x}_{1s} \\ \mathbf{x}_{21} & \mathbf{x}_{22} & . & . & . & \mathbf{x}_{2s} \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ \mathbf{x}_{p1} & \mathbf{x}_{p2} & . & . & . & \mathbf{x}_{ps} \end{bmatrix}$$

and our q by s matrix of stimuli coordinates is:

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_{11} & \mathbf{z}_{12} & . & . & . & \mathbf{z}_{1s} \\ \mathbf{z}_{21} & \mathbf{z}_{22} & . & . & . & \mathbf{z}_{2s} \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ \mathbf{z}_{q1} & \mathbf{z}_{q2} & . & . & . & \mathbf{z}_{qs} \end{bmatrix}$$

The p by q matrix of squared distances between **X** and **Z** (individuals [respondents] and stimuli) is:

$$\mathbf{D} = \begin{bmatrix} \sum_{k=1}^{s}(x_{1k} - z_{1k})^2 & \sum_{k=1}^{s}(x_{1k} - z_{2k})^2 & . & . & . & \sum_{k=1}^{s}(x_{1k} - z_{qk})^2 \\ \sum_{k=1}^{s}(x_{2k} - z_{1k})^2 & \sum_{k=1}^{s}(x_{2k} - z_{2k})^2 & . & . & . & \sum_{k=1}^{s}(x_{2k} - z_{qk})^2 \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ \sum_{k=1}^{s}(x_{pk} - z_{1k})^2 & \sum_{k=1}^{s}(x_{pk} - z_{2k})^2 & . & . & . & \sum_{k=1}^{s}(x_{pk} - z_{qk})^2 \end{bmatrix} \qquad \textbf{(2)}$$

This can be written in matrix algebra as:

$$\mathbf{D = diag(XX')J_q' - 2XZ' + J_p diag(ZZ')' =}$$

$$\mathbf{[diag(XX') \mid -2X \mid J_p]} \begin{bmatrix} \mathbf{J_q'} \\ \hline \mathbf{Z'} \\ \hline \mathbf{diag(ZZ')'} \end{bmatrix}$$

Note that the rank of **D**, $\rho(\mathbf{D})$, must be less than or equal to s+2; i.e., $\rho(\mathbf{D}) \leq s + 2$**.**

If there was no error then equation (2) can be solved using the method of Schonemann (1970). Part of Schonemann's solution is to work with the double-centered matrix. Recall:

$$\mathbf{Y} = \mathbf{X}^*\mathbf{Z}^{*\prime} = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & . & . & . & x_{1s} - \bar{x}_s \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & . & . & . & x_{2s} - \bar{x}_s \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ x_{p1} - \bar{x}_1 & x_{p2} - \bar{x}_2 & . & . & . & x_{ps} - \bar{x}_s \end{bmatrix} \begin{bmatrix} z_{11} - \bar{z}_1 & z_{12} - \bar{z}_2 & . & . & . & z_{1s} - \bar{z}_s \\ z_{21} - \bar{z}_1 & z_{22} - \bar{z}_2 & . & . & . & z_{2s} - \bar{z}_s \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ z_{q1} - \bar{z}_1 & z_{q2} - \bar{z}_2 & . & . & . & z_{qs} - \bar{z}_s \end{bmatrix}' \qquad (3)$$

Y is a p by q matrix which is equal to the product of a p by s matrix **X*** and a q by s matrix **Z*.** It is *double-centered* because **X** is defined with respect to the coordinate system *centered* at the origin $\bar{\mathbf{x}}$ and **Z** is defined with respect to the coordinate system *centered* at the origin $\bar{\mathbf{z}}$.

Where the double-centered matrix comes into play in the Thermometer problem is that we can use singular value decomposition to get *starting coordinates* for either **X** or **Z** to use in a gradient-style solution. Specifically, consider the standard squared error loss function:

$$\mu = \sum_{i=1}^{p}\sum_{j=1}^{q}\varepsilon_{ij}^2 = \sum_{i=1}^{p}\sum_{j=1}^{q}\left(\mathbf{d}_{ij}^* - \mathbf{d}_{ij}\right)^2 \qquad (4)$$

Where, from above,

$$\mathbf{d}_{ij} = \sqrt{\sum_{k=1}^{s}(\mathbf{x}_{ik} - \mathbf{z}_{jk})^2}$$

SMACOF Solution

SMACOF is an iterative technique that constructs a quadratic function that *always lies above* the loss function, (4). To see the logic, expand (4):

$$\sum_{i=1}^{p}\sum_{j=1}^{q}\left(\mathbf{d}_{ij}^* - \mathbf{d}_{ij}\right)^2 = \sum_{i=1}^{p}\sum_{j=1}^{q}\mathbf{d}_{ij}^{*2} + \sum_{i=1}^{p}\sum_{j=1}^{q}\mathbf{d}_{ij}^2 - 2\sum_{i=1}^{p}\sum_{j=1}^{q}\mathbf{d}_{ij}^*\mathbf{d}_{ij}$$

Now, let $\mathbf{X}^{(h)}$ **and** $\mathbf{Z}^{(h)}$ be the coordinates at the h^{th} iteration and let $\mathbf{X}^{(h+1)}$ **and** $\mathbf{Z}^{(h+1)}$ be the coordinates at the $h+1^{th}$ iteration. The first term, $\sum_{i=1}^{p}\sum_{j=1}^{q}\mathbf{d}_{ij}^{*2}$ is always a constant, at iteration h+1 the second term is $\sum_{i=1}^{p}\sum_{j=1}^{q}\mathbf{d}_{ij}^{(h+1)^2}$. Because the first and second terms are trivial, bounding the loss function boils down to finding a bound such that $d_{ij} \geq \Delta_{ij}$ for all i,j where Δ_{ij} is easily constructed. De Leeuw's solution is:

$$\Delta_{ij}^{(h+1)} = \frac{\sum_{k=1}^{s}\left(x_{ik}^{(h+1)} - z_{jk}^{(h+1)}\right)\left(x_{ik}^{(h)} - z_{jk}^{(h)}\right)}{d_{ij}^{(h)}}$$

The numerator comes from the Cauchy-Schwarz inequality which in this case is:

$$\sum_{k=1}^{s}\left(x_{ik}^{(h+1)} - z_{jk}^{(h+1)}\right)\left(x_{ik}^{(h)} - z_{jk}^{(h)}\right) \leq \left(\sum_{k=1}^{s}\left(x_{ik}^{(h+1)} - z_{jk}^{(h+1)}\right)^2\right)^{1/2}\left(\sum_{k=1}^{s}\left(x_{ik}^{(h)} - z_{jk}^{(h)}\right)^2\right)^{1/2} = d_{ij}^{(h+1)} d_{ij}^{(h)}$$

This implies that: $\Lambda_{ij}^{(h+1)} \leq d_{ij}^{(h+1)}$

Note that when $\mathbf{X}^{(\mathbf{h})} = \mathbf{X}^{(\mathbf{h+1})}$ and $\mathbf{Z}^{(\mathbf{h})} = \mathbf{Z}^{(\mathbf{h+1})}$ then $\Delta_{ij}^{(h+1)} = d_{ij}^{(h+1)} = d_{ij}^{(h)}$**.** Hence:

$$\sum_{\mathbf{i}=1}^{\mathbf{p}}\sum_{\mathbf{j}=1}^{\mathbf{q}} \mathbf{d}_{\mathbf{ij}}^{*}\, \mathbf{d}_{\mathbf{ij}}^{(\mathbf{h+1})} \geq \sum_{\mathbf{i}=1}^{\mathbf{p}}\sum_{\mathbf{j}=1}^{\mathbf{q}} \mathbf{d}_{\mathbf{ij}}^{*}\Delta_{\mathbf{ij}}^{(\mathbf{h+1})} = \sum_{\mathbf{i}=1}^{\mathbf{p}}\sum_{\mathbf{j}=1}^{\mathbf{q}} \mathbf{d}_{\mathbf{ij}}^{*} \frac{\sum_{\mathbf{k}=1}^{\mathbf{s}}\left(\mathbf{x}_{\mathbf{ik}}^{(\mathbf{h+1})} - \mathbf{z}_{\mathbf{jk}}^{(\mathbf{h+1})}\right)\left(\mathbf{x}_{\mathbf{ik}}^{(\mathbf{h})} - \mathbf{z}_{\mathbf{jk}}^{(\mathbf{h})}\right)}{\mathbf{d}_{\mathbf{ij}}^{(\mathbf{h})}}$$

This gives us the bounding function we seek:

$$\sum_{\mathbf{i}=1}^{\mathbf{p}}\sum_{\mathbf{j}=1}^{\mathbf{q}} \mathbf{d}_{\mathbf{ij}}^{*2} + \sum_{\mathbf{i}=1}^{\mathbf{p}}\sum_{\mathbf{j}=1}^{\mathbf{q}} \mathbf{d}_{\mathbf{ij}}^{(\mathbf{h+1})^2} - 2\sum_{\mathbf{i}=1}^{\mathbf{p}}\sum_{\mathbf{j}=1}^{\mathbf{q}} \mathbf{d}_{\mathbf{ij}}^{*}\, \mathbf{d}_{\mathbf{ij}}^{(\mathbf{h+1})} \leq \sum_{\mathbf{i}=1}^{\mathbf{p}}\sum_{\mathbf{j}=1}^{\mathbf{q}} \mathbf{d}_{\mathbf{ij}}^{*2} + \sum_{\mathbf{i}=1}^{\mathbf{p}}\sum_{\mathbf{j}=1}^{\mathbf{q}} \mathbf{d}_{\mathbf{ij}}^{(\mathbf{h+1})^2} - 2\sum_{\mathbf{i}=1}^{\mathbf{p}}\sum_{\mathbf{j}=1}^{\mathbf{q}} \mathbf{d}_{\mathbf{ij}}^{*}\, \Delta_{\mathbf{ij}}^{(\mathbf{h+1})}$$

Taking derivatives of the right hand side:

$$\frac{\delta rhs}{\delta x_{ik}^{(h+1)}} = 2\sum_{j=1}^{q}\left(x_{ik}^{(h+1)} - z_{jk}^{(h+1)}\right) - 2\sum_{j=1}^{q} d_{ij}^{*} \frac{\left(x_{ik}^{(h)} - z_{jk}^{(h)}\right)}{d_{ij}^{(h)}} = q x_{ik}^{(h+1)} - \sum_{j=1}^{q} z_{jk}^{(h+1)} - \sum_{j=1}^{q} d_{ij}^{*} \frac{\left(x_{ik}^{(h)} - z_{jk}^{(h)}\right)}{d_{ij}^{(h)}}$$

$$\frac{\delta rhs}{\delta z_{jk}^{(h+1)}} = -2\sum_{i=1}^{p}\left(x_{ik}^{(h+1)} - z_{jk}^{(h+1)}\right) + 2\sum_{i=1}^{p} d_{ij}^{*} \frac{\left(x_{ik}^{(h)} - z_{jk}^{(h)}\right)}{d_{ij}^{(h)}} = p z_{jk}^{(h+1)} - \sum_{i=1}^{p} x_{ik}^{(h+1)} + \sum_{i=1}^{p} d_{ij}^{*} \frac{\left(x_{ik}^{(h)} - z_{jk}^{(h)}\right)}{d_{ij}^{(h)}}$$

We have s(p+q) equations with s(p+q) unknowns so we can solve for the minimum of the bounding function. Note that:

$$x_{ik}^{(h+1)} = \bar{z}_k^{(h+1)} + \frac{1}{q}\sum_{j=1}^{q} d_{ij}^* \frac{\left(x_{ik}^{(h)} - z_{jk}^{(h)}\right)}{d_{ij}^{(h)}}$$

$$z_{jk}^{(h+1)} = \bar{x}_k^{(h+1)} - \frac{1}{p}\sum_{i=1}^{p} d_{ij}^* \frac{\left(x_{ik}^{(h)} - z_{jk}^{(h)}\right)}{d_{ij}^{(h)}}$$

Hence, if we assume that $Z^{(h+1)}$ is centered at the origin, $\bar{z}_1^{(h+1)} = \bar{z}_2^{(h+1)} = ... = \bar{z}_s^{(h+1)} = 0$, then we have a solution for all the $x_{ik}^{(h+1)}$. Given the $x_{ik}^{(h+1)}$ we can compute the $\bar{x}_k^{(h+1)}$ and then the $z_{jk}^{(h+1)}$.

Another solution would be:

$$z_{jk}^{(h+1)} = \bar{z}_k^{(h+1)} + \frac{1}{pq}\sum_{i=1}^{p}\sum_{j=1}^{q} d_{ij}^* \frac{\left(x_{ik}^{(h)} - z_{jk}^{(h)}\right)}{d_{ij}^{(h)}} - \frac{1}{p}\sum_{i=1}^{p} d_{ij}^* \frac{\left(x_{ik}^{(h)} - z_{jk}^{(h)}\right)}{d_{ij}^{(h)}}$$

MLSMU6 Solution

The first derivatives are:

$$\frac{\partial \mu}{\partial \mathbf{z}_{jk}} = 2\sum_{i=1}^{p}\left\{\left(\mathbf{d}_{ij}^* - \mathbf{d}_{ij}\right)\left(-\frac{1}{2}\right)\left[\sum_{k=1}^{s}(\mathbf{x}_{ik} - \mathbf{z}_{jk})^2\right]^{-\frac{1}{2}}\left(-2\left[\mathbf{x}_{ik} - \mathbf{z}_{jk}\right]\right)\right\}$$

$$= 2\sum_{i=1}^{p}\left\{\left(\frac{\mathbf{d}_{ij}^*}{\mathbf{d}_{ij}} - 1\right)\left(\mathbf{x}_{ik} - \mathbf{z}_{jk}\right)\right\} \qquad \textbf{(5)}$$

$$\frac{\partial \mu}{\partial \mathbf{x}_{ik}} = 2\sum_{j=1}^{q}\left\{\left(\mathbf{d}_{ij}^* - \mathbf{d}_{ij}\right)\left(-\frac{1}{2}\right)\left[\sum_{k=1}^{s}(\mathbf{x}_{ik} - \mathbf{z}_{jk})^2\right]^{-\frac{1}{2}}\left(2\left[\mathbf{x}_{ik} - \mathbf{z}_{jk}\right]\right)\right\}$$

$$= -2\sum_{j=1}^{q}\left\{\left(\frac{\mathbf{d}_{ij}^*}{\mathbf{d}_{ij}} - 1\right)\left(\mathbf{x}_{ik} - \mathbf{z}_{jk}\right)\right\} \qquad \textbf{(6)}$$

Setting (5) equal to zero and solving for $\mathbf{z}_{jk}$:

$$\sum_{i=1}^{p}\left[\frac{d_{ij}^{*}}{d_{ij}}\left(x_{ik} - z_{jk}\right)\right] - \sum_{i=1}^{p}\left(x_{ik} - z_{jk}\right) = 0$$

Rearranging:

$$pz_{jk} - \sum_{i=1}^{p}\left[x_{ik} + \frac{d_{ij}^{*}}{d_{ij}}\left(z_{jk} - x_{ik}\right)\right] = 0$$

Therefore:

$$\hat{z}_{jk} = \frac{1}{p}\sum_{i=1}^{p}\left[x_{ik} + \frac{d_{ij}^{*}}{d_{ij}}\left(z_{jk} - x_{ik}\right)\right] \qquad \textbf{(7)}$$

Continuing, setting (6) equal to zero and solving for $\mathbf{x_{ik}}$**:**

$$-\sum_{j=1}^{q}\left[\frac{d_{ij}^{*}}{d_{ij}}\left(x_{ik} - z_{jk}\right)\right] + \sum_{j=1}^{q}\left(x_{ik} - z_{jk}\right) = 0$$

Rearranging:

$$qx_{ik} - \sum_{j=1}^{q}\left[z_{jk} + \frac{d_{ij}^{*}}{d_{ij}}\left(x_{ik} - z_{jk}\right)\right] = 0$$

Therefore:

$$\hat{x}_{ik} = \frac{1}{q}\sum_{i=1}^{q}\left[z_{jk} + \frac{d_{ij}^{*}}{d_{ij}}\left(x_{ik} - z_{jk}\right)\right] \qquad \textbf{(8)}$$

Note that the solution [equations (7) and (8)] is in the form:

z=f(x,z) and x=g(x,z)

That is, the solutions for **z** and **x** are values such that when they are plugged into **f(x,z)** and **g(x,z)** they reproduce themselves!

Define:

$$z_{jki} = x_{ik} + \frac{d_{ij}^{*}}{d_{ij}}\left(z_{jk} - x_{ik}\right) \qquad \textbf{(9)}$$

$$\mathbf{x_{ikj}} = \mathbf{z_{jk}} + \frac{\mathbf{d_{ij}^{*}}}{\mathbf{d_{ij}}}\left(\mathbf{x_{ik}} - \mathbf{z_{jk}}\right) \qquad \mathbf{(10)}$$

So that equations (7) and (8) can be re-written as:

$$\hat{\mathbf{z}}_{\mathbf{jk}} = \frac{1}{\mathbf{p}}\sum_{\mathbf{i=1}}^{\mathbf{p}} \mathbf{z_{jki}} \qquad \mathbf{(11)}$$

$$\hat{\mathbf{x}}_{\mathbf{ik}} = \frac{1}{\mathbf{q}}\sum_{\mathbf{j=1}}^{\mathbf{q}} \mathbf{x_{ikj}} \qquad \mathbf{(12)}$$

Using equation (9), note that the *point* $\mathbf{z_{j,m}}$ is:

$$\mathbf{z_{j,i}} = \begin{bmatrix} \mathbf{x_{i1}} + \frac{\mathbf{d_{ij}^{*}}}{\mathbf{d_{ij}}}\left(\mathbf{z_{j1}} - \mathbf{x_{i1}}\right) \\ \mathbf{x_{i2}} + \frac{\mathbf{d_{ij}^{*}}}{\mathbf{d_{ij}}}\left(\mathbf{z_{j2}} - \mathbf{x_{i2}}\right) \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{x_{is}} + \frac{\mathbf{d_{ij}^{*}}}{\mathbf{d_{ij}}}\left(\mathbf{z_{js}} - \mathbf{x_{is}}\right) \end{bmatrix} = \mathbf{x_i} + \frac{\mathbf{d_{ij}^{*}}}{\mathbf{d_{ij}}}\left(\mathbf{z_j} - \mathbf{x_i}\right) \qquad (13)$$

Similarly:

$$\mathbf{x_{i,j}} = \begin{bmatrix} \mathbf{z_{j1}} + \frac{\mathbf{d_{ij}^{*}}}{\mathbf{d_{ij}}}\left(\mathbf{x_{i1}} - \mathbf{z_{j1}}\right) \\ \mathbf{z_{j2}} + \frac{\mathbf{d_{ij}^{*}}}{\mathbf{d_{ij}}}\left(\mathbf{x_{i2}} - \mathbf{z_{j2}}\right) \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{z_{js}} + \frac{\mathbf{d_{ij}^{*}}}{\mathbf{d_{ij}}}\left(\mathbf{x_{is}} - \mathbf{z_{js}}\right) \end{bmatrix} = \mathbf{z_j} + \frac{\mathbf{d_{ij}^{*}}}{\mathbf{d_{ij}}}\left(\mathbf{x_i} - \mathbf{z_j}\right) \qquad (14)$$

Where $\mathbf{z_j} = \begin{bmatrix} \mathbf{z_{j1}} \\ \mathbf{z_{j2}} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{z_{js}} \end{bmatrix}$ and $\mathbf{x_i} = \begin{bmatrix} \mathbf{x_{i1}} \\ \mathbf{x_{i2}} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{x_{is}} \end{bmatrix}$ are points and $\frac{\mathbf{d^*_{ij}}}{\mathbf{d_{ij}}}$ is a scalar.

Equations (13) and (14) are basic equations of a line that passes through $\mathbf{z_j}$ ***and*** $\mathbf{x_i}$**!** The general formula for a line equation is:

$$\mathbf{Y(t)=A+t(B-A)} \qquad \mathbf{(15)}$$

Where **A** and **B** are points and **t** is a scalar. Note that if $\mathbf{0<t<1}$ then equation (15) defines a line that runs between points **A** and **B.**

Once specific values are plugged into equations (11) or (12) then the solution for the point

$\hat{\mathbf{z}}_\mathbf{j} = \begin{bmatrix} \hat{\mathbf{z}}_\mathbf{j1} \\ \hat{\mathbf{z}}_\mathbf{j2} \\ \cdot \\ \cdot \\ \cdot \\ \hat{\mathbf{z}}_\mathbf{js} \end{bmatrix}$ is simply the centroid of the p $\mathbf{z_{j,i}}$ points and $\hat{\mathbf{x}}_\mathbf{i} = \begin{bmatrix} \hat{\mathbf{x}}_\mathbf{i1} \\ \hat{\mathbf{x}}_\mathbf{i2} \\ \cdot \\ \cdot \\ \cdot \\ \hat{\mathbf{x}}_\mathbf{is} \end{bmatrix}$ is simply the centroid of the q $\mathbf{x_{i,j}}$ points**!**

Finally, note that the squared distance between the points $\mathbf{z_j}$ and $\mathbf{z_{j,m}}$ is:

$$\sum_{\mathbf{k}=1}^{\mathbf{s}}\left(\mathbf{z_{jk}}-\mathbf{z_{jki}}\right)^2=\sum_{\mathbf{k}=1}^{\mathbf{s}}\left[\mathbf{z_{jk}}-\left(\mathbf{x_{ik}}+\frac{\mathbf{d^*_{ij}}}{\mathbf{d_{ij}}}\left(\mathbf{z_{jk}}-\mathbf{x_{ik}}\right)\right)\right]^2=$$

$$\sum_{\mathbf{k}=1}^{\mathbf{s}}\left[\left(\mathbf{z_{jk}}-\mathbf{x_{ik}}\right)\left(1-\frac{\mathbf{d^*_{ij}}}{\mathbf{d_{ij}}}\right)\right]^2=\frac{\left(\mathbf{d_{ij}}-\mathbf{d^*_{ij}}\right)^2}{\mathbf{d^2_{ij}}}\left(\sum_{\mathbf{k}=1}^{\mathbf{s}}\left(\mathbf{z_{jk}}-\mathbf{x_{ik}}\right)^2\right)=\boldsymbol{\varepsilon}^2_{\mathbf{ij}} \qquad (16)$$

So that the squared error is represented directly on the s-dimensional hyperplane (see below).

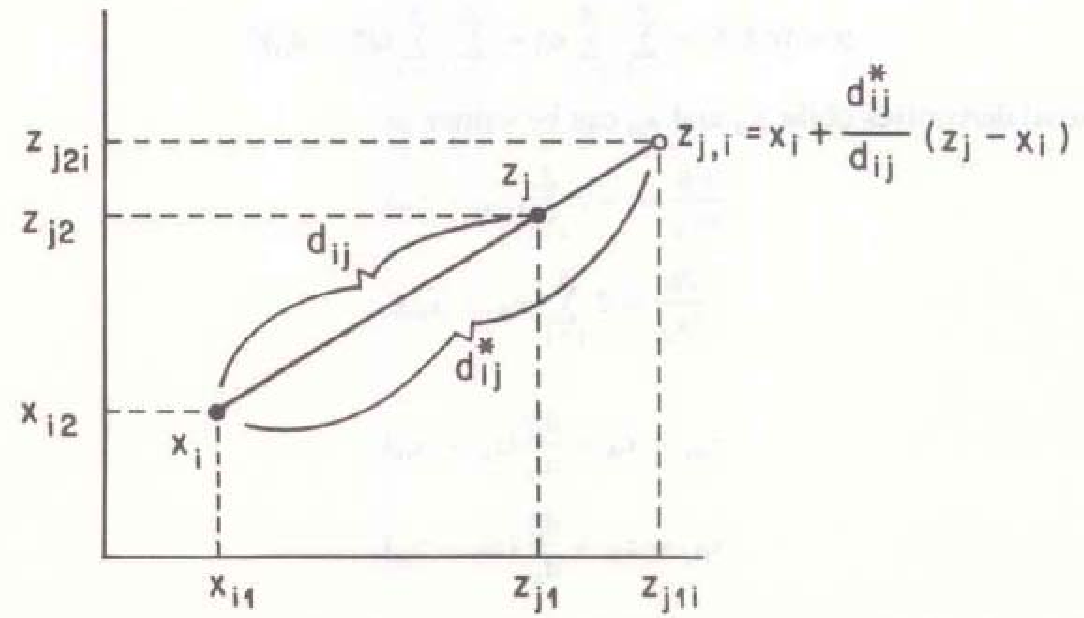

FIGURE 1
Parametric equation of a straight line.

Similarly, the squared distance between the points $\mathbf{x_i}$ and $\mathbf{x_{i,j}}$ is:

$$\sum_{k=1}^{s}\left(x_{ik}-x_{ikj}\right)^2 = \sum_{k=1}^{s}\left[x_{ik}-\left(z_{jk}+\frac{d^*_{ij}}{d_{ij}}\left(x_{ik}-z_{jk}\right)\right)\right]^2 =$$

$$\sum_{k=1}^{s}\left[\left(x_{ik}-z_{jk}\right)\left(1-\frac{d^*_{ij}}{d_{ij}}\right)\right]^2 = \frac{\left(d_{ij}-d^*_{ij}\right)^2}{d^2_{ij}}\left(\sum_{k=1}^{s}\left(x_{ik}-z_{jk}\right)^2\right) = \varepsilon^2_{ij} \quad (17)$$

Another interesting property are the following identities:

$$d^{*2}_{ij} = \sum_{k=1}^{s}\left(x_{ik}-z_{jki}\right)^2 = \sum_{k=1}^{s}\left\{x_{ik}-\left[x_{ik}+\frac{d^*_{ij}}{d_{ij}}\left(z_{jk}-x_{ik}\right)\right]\right\}^2 = \frac{d^{*2}_{ij}}{d^2_{ij}}\sum_{k=1}^{s}\left(z_{jk}-x_{ik}\right)^2 = d^{*2}_{ij}$$

$$d^{*2}_{ij} = \sum_{k=1}^{s}\left(z_{jk}-x_{ikj}\right)^2 = \sum_{k=1}^{s}\left\{z_{jk}-\left[z_{jk}+\frac{d^*_{ij}}{d_{ij}}\left(x_{ik}-z_{jk}\right)\right]\right\}^2 = \frac{d^{*2}_{ij}}{d^2_{ij}}\sum_{k=1}^{s}\left(x_{ik}-z_{jk}\right)^2 = d^{*2}_{ij}$$

Intuitively, the observed distances, the $\mathbf{d}^*_{ij}$ are the lengths of the vectors *attached* to the $\mathbf{x_i}$ that produce the $\mathbf{z_{j,i}}$ points (see figure above). Similarly, the $\mathbf{d}^*_{ij}$ are the lengths of the vectors *attached* to the z_j that produce the $\mathbf{x_{i,j}}$ points.

It is a relatively simple process to iterate back and forth between equations (7) and (8) [equivalently, equations (11) and (12)] until the $\hat{\mathbf{x}}$'s and $\hat{\mathbf{z}}$'s reproduce each other.

Note that this process is *strictly descending*; that is:

$$\sum_{i=1}^{p} \varepsilon_{ij}^{(h+1)^2} = \sum_{i=1}^{p}\sum_{k=1}^{s}\left(\mathbf{z}_{jk}^{(h+1)} - \mathbf{z}_{jki}^{(h+1)}\right)^2 \le \sum_{i=1}^{p}\sum_{k=1}^{s}\left(\mathbf{z}_{jk}^{(h+1)} - \mathbf{z}_{jki}^{(h)}\right)^2 \le \sum_{i=1}^{p}\sum_{k=1}^{s}\left(\mathbf{z}_{jk}^{(h)} - \mathbf{z}_{jki}^{(h)}\right)^2 = \sum_{i=1}^{p} \varepsilon_{ij}^{(h)^2} \quad (18)$$

where h is the iteration number. The third and fourth terms of the inequality are true because $\mathbf{z}_{jk}^{(h+1)}$, by equation (11), is the centroid of the $\mathbf{z}_{jki}^{(h)}$. The second and third terms of the inequality are true because the $\mathbf{z}_{jki}^{(h+1)}$ are computed using $\mathbf{z}_{jk}^{(h+1)}$ as shown in equations (11) and (13). That is, the point $\mathbf{z}_{j,i}^{(h+1)}$, *by construction,* lies on the line that passes through $\mathbf{x_i}$ and $\mathbf{z}_j^{(h+1)}$ hence *it must be closer to* $\mathbf{z}_j^{(h+1)}$ than the point $\mathbf{z}_{j,i}^{(h)}$ because $\mathbf{z}_{j,i}^{(h)}$ lies on the straight line through $\mathbf{x_i}$ and $\mathbf{z}_j^{(h)}$. Furthermore, by the identities shown above, the distance from $\mathbf{x_i}$ to $\mathbf{z}_{j,i}^{(h)}$ is $\mathbf{d}^*_{ij}$ and the distance from $\mathbf{x_i}$ to $\mathbf{z}_{j,i}^{(h+1)}$ is $\mathbf{d}^*_{ij}$. Hence, by the triangle inequality the distance from $\mathbf{z}_j^{(h+1)}$ to $\mathbf{z}_{j,i}^{(h+1)}$ must be less than the distance from $\mathbf{z}_j^{(h+1)}$ to $\mathbf{z}_{j,i}^{(h)}$.

Finally, the problem of a zero distance is easily handled in this framework. Suppose at some point that $\mathbf{d}_{ij}^{(h)} = 0$. Then the corresponding vector of length $\mathbf{d}^*_{ij}$ could not be "aimed" at a point because $\mathbf{x_i} = \mathbf{z_j}$ so that no $\mathbf{z}_{j,i}^{(h)}$ [or $\mathbf{x}_{i,j}^{(h)}$] could be computed. In this instance *any* $\mathbf{z}_{j,i}^{(h)}$ [or $\mathbf{x}_{i,j}^{(h)}$] point with distance $\mathbf{d}_{ij}^*$ from $\mathbf{x}_i$ [or $\mathbf{z}_j$] will still satisfy the inequality above.

Although this algorithm works very well and converges quickly the fly in the ointment is that the statistical properties of the $\boldsymbol{\varepsilon}_{ij}$ are not clear. For example, suppose $\boldsymbol{\varepsilon}_{ij} \sim \mathbf{N}(0, \boldsymbol{\sigma}^2)$ then $\mathbf{d}_{ij}^* \sim \mathbf{N}(\mathbf{d}_{ij}, \boldsymbol{\sigma}^2)$. But this will not work because $\mathbf{d}_{ij}^* \geq 0$.

The Cahoon-Hinich Solution

Cahoon and Hinich assume that the observed data are noisy squared distances and transform the data by picking one of the stimuli to be the origin and subtracting the column corresponding to that stimulus from the other columns in the data matrix**. (**I will denote that matrix as **Δ**.) Let **D*** be the noisy matrix of squared distances. Cahoon and Hinich assume:

$$\mathbf{D}^* = \mathbf{D} + \mathbf{E} \qquad \textbf{(19)}$$

where **D** is the p by q matrix of true squared distances given in equation (2) and **E** is a p by q matrix of error. When one of the columns is subtracted from the others this has the effect of canceling out the $\sum_{k=1}^{s} \mathbf{x}_{ik}^2$ terms from the other columns in **D**.

To simplify notation assume that there are q+1 stimuli and we subtract column q+1 from the first q columns of **D**. Let the transformed matrix be denoted as $\tilde{\mathbf{D}}$. The entries in $\tilde{\mathbf{D}}$ become:

$$\sum_{k=1}^{s}(\mathbf{x}_{ik} - \mathbf{z}_{jk})^2 - \sum_{k=1}^{s}(\mathbf{x}_{ik} - \mathbf{z}_{q+1,k})^2 = -2\sum_{k=1}^{s}\mathbf{x}_{ik}\left(\mathbf{z}_{jk} - \mathbf{z}_{q+1,k}\right) + \sum_{k=1}^{s}\mathbf{z}_{jk}^2 - \sum_{k=1}^{s}\mathbf{z}_{q+1,k}^2$$

Now setting stimulus q+1 to the origin, that is, $\mathbf{z}_{q+1} = \begin{bmatrix} \mathbf{z}_{q+1,1} \\ \mathbf{z}_{q+1,2} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{z}_{q+1,s} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$, the entries in $\tilde{\mathbf{D}}$ simplify to $-2\sum_{k=1}^{s} \mathbf{x}_{ik}\mathbf{z}_{jk} + \sum_{k=1}^{s} \mathbf{z}_{jk}^{2}$. When error is taken into account the Cahoon-Hinich model becomes:

$$\Delta_{ij} = -2\sum_{k=1}^{s} \mathbf{x}_{ik}\mathbf{z}_{jk} + \sum_{k=1}^{s} \mathbf{z}_{jk}^{2} + \boldsymbol{\varepsilon}_{ij} - \boldsymbol{\varepsilon}_{i,q+1} \quad (20)$$

Now consider the column means of the p by q matrix **Δ:**

$$\bar{\Delta}_{j} = -2\sum_{k=1}^{s} \bar{\mathbf{x}}_{k}\mathbf{z}_{jk} + \sum_{k=1}^{s} \mathbf{z}_{jk}^{2}$$

The sum of the error terms across the respondents can be assumed to be zero.

Subtracting the column means from each entry in the corresponding column cancels the $\sum_{k=1}^{s} \mathbf{z}_{jk}^{2}$ terms; that is,

$$\Delta_{j}^{*} = \Delta_{j} - \mathbf{J}_{p}\bar{\Delta}_{j} = \begin{bmatrix} -2\sum_{k=1}^{s} \mathbf{z}_{jk}\left(\mathbf{x}_{1k} - \bar{\mathbf{x}}_{k}\right) \\ -2\sum_{k=1}^{s} \mathbf{z}_{jk}\left(\mathbf{x}_{2k} - \bar{\mathbf{x}}_{k}\right) \\ \cdot \\ \cdot \\ \cdot \\ -2\sum_{k=1}^{s} \mathbf{z}_{jk}\left(\mathbf{x}_{pk} - \bar{\mathbf{x}}_{k}\right) \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{1j} - \boldsymbol{\varepsilon}_{1,q+1} \\ \boldsymbol{\varepsilon}_{2j} - \boldsymbol{\varepsilon}_{2,q+1} \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{\varepsilon}_{pj} - \boldsymbol{\varepsilon}_{p,q+1} \end{bmatrix} \quad (21)$$

Using equation (20) and the regularity assumptions on the errors, the q by q covariance matrix of the $\mathbf{\Delta_j - J_p \bar{\Delta}_j}$ is:

$$\mathbf{\Theta = 4Z'\Sigma_x Z + \Psi + \sigma^2_{q+1} J_q J_q'} \qquad (22)$$

Where $\mathbf{\Sigma_x}$ is the s by s variance-covariance matrix for **X,** $\mathbf{\Psi}$ is a q by q diagonal matrix of the variances for the stimuli, namely, $\mathbf{\sigma^2_j = E(\varepsilon^2_{ij})}$, and $\mathbf{\sigma^2_{q+1}}$ is the corresponding variance for stimulus q+1, $\mathbf{\sigma^2_{q+1} = E(\varepsilon^2_{i,q+1})}$

Let

$$\mathbf{\tilde{\Delta}^* = \Delta^* - J_p \bar{\delta}^*} \qquad (23)$$

where $\mathbf{\Delta^*}$ is the p by q matrix from equation (21) and $\mathbf{\bar{\delta}^*}$ is the vector of the means of the q columns of $\mathbf{\Delta^*}$. The q by q sample covariance matrix is:

$$\mathbf{S} = \frac{1}{\mathbf{p}}\left(\tilde{\Delta}^{*\prime}\tilde{\Delta}^{*}\right) \qquad (24)$$

Cahoon and Hinich then recover estimates of the parameters of the model by analyzing **S** with maximum likelihood factor analysis and the vector of means $\mathbf{\bar{\delta}^*}$ with OLS and bootstrapping (errors in variables problems).

The statistical properties of the method are not completely clear as no parametric assumptions are made about the error.

Transforming the Thermometer Scores into Roll Call Votes

In 1968 respondents were asked to give feeling thermometer ratings to 12 political figures: George Wallace, Hubert Humphrey, Richard Nixon, Eugene McCarthy, Ronald

Reagan, Nelson Rockefeller, Lyndon Johnson, George Romney, Robert Kennedy, Edmund Muskie, Spiro Agnew, and Curtis LeMay. The NES survey was conducted after Robert Kennedy's assassination in June, 1968. This obviously affects the ratings Kennedy received.

Suppose a respondent gave ratings of 30, 80, and 55 to Wallace (W), Humphrey (H), and Nixon (N) respectively. With respect to these three candidates, the rank order is H > N > W. Now suppose a second respondent gave ratings of 45, 65, and 95, respectively, for a rank order of N > H > W. These rank orders can be converted to binary choice data by treating each pair of candidates as a roll call vote. For example, consider the pair of Wallace and Humphrey. If a respondent rates Wallace higher than Humphrey make that Yea, and if Humphrey is rated higher than Wallace, make that Nay. Doing this consistently across respondents creates a roll call vote where the outcomes are Wallace and Humphrey, respectively.

With the actual 1968 data, I used the order of the 12 political figures listed above (which is their actual order in the NES data set) to create the roll calls. That is, given a pair of politicians, the one earlier in the NES ordering was treated as a Yea and the later one a Nay. So if the pair was Ronald Reagan and Curtis LeMay, then if a respondent rated Reagan higher than LeMay that is a Yea vote. If a respondent gave a pair of politicians the same rating, for example, 55 and 55, then I treated it as missing data (that is, as if the respondent abstained on the roll call).

Example:

50 97 85 60 40 40 75 30 80 60 40 30

WALLACE	HUMPHREY	6
WALLACE	NIXON	6
WALLACE	MCCARTHY	6
WALLACE	REAGAN	1
WALLACE	ROCKEFELLER	1
WALLACE	LBJ	6
WALLACE	ROMNEY	1
WALLACE	R.KENNEDY	6
WALLACE	MUSKIE	6
WALLACE	AGNEW	1
WALLACE	LEMAY	1
HUMPHREY	NIXON	1
HUMPHREY	MCCARTHY	1
HUMPHREY	REAGAN	1
HUMPHREY	ROCKEFELLER	1
HUMPHREY	LBJ	1
HUMPHREY	ROMNEY	1
HUMPHREY	R.KENNEDY	1
HUMPHREY	MUSKIE	1
HUMPHREY	AGNEW	1
HUMPHREY	LEMAY	1
NIXON	MCCARTHY	1
NIXON	REAGAN	1
NIXON	ROCKEFELLER	1
NIXON	LBJ	1
NIXON	ROMNEY	1
NIXON	R.KENNEDY	1
NIXON	MUSKIE	1
NIXON	AGNEW	1
NIXON	LEMAY	1
MCCARTHY	REAGAN	1
MCCARTHY	ROCKEFELLER	1
MCCARTHY	LBJ	6
MCCARTHY	ROMNEY	1
MCCARTHY	R.KENNEDY	6
MCCARTHY	MUSKIE	0
MCCARTHY	AGNEW	1
MCCARTHY	LEMAY	1
REAGAN	ROCKEFELLER	0
REAGAN	LBJ	6
REAGAN	ROMNEY	1
REAGAN	R.KENNEDY	6
REAGAN	MUSKIE	6
REAGAN	AGNEW	0
REAGAN	LEMAY	1
ROCKEFELLER	LBJ	6
ROCKEFELLER	ROMNEY	1
ROCKEFELLER	R.KENNEDY	6
ROCKEFELLER	MUSKIE	6
ROCKEFELLER	AGNEW	0
ROCKEFELLER	LEMAY	1
LBJ	ROMNEY	1
LBJ	R.KENNEDY	6
LBJ	MUSKIE	1
LBJ	AGNEW	1
LBJ	LEMAY	1
ROMNEY	R.KENNEDY	6
ROMNEY	MUSKIE	6
ROMNEY	AGNEW	6
ROMNEY	LEMAY	0
R.KENNEDY	MUSKIE	1
R.KENNEDY	AGNEW	1
R.KENNEDY	LEMAY	1
MUSKIE	AGNEW	1
MUSKIE	LEMAY	1
AGNEW	LEMAY	1

Figure 9C. 1968 Humphrey Voters

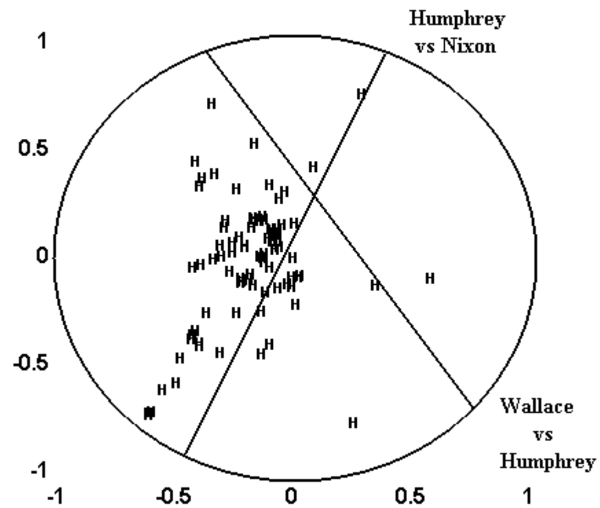

Figure 9D. 1968 Nixon Voters

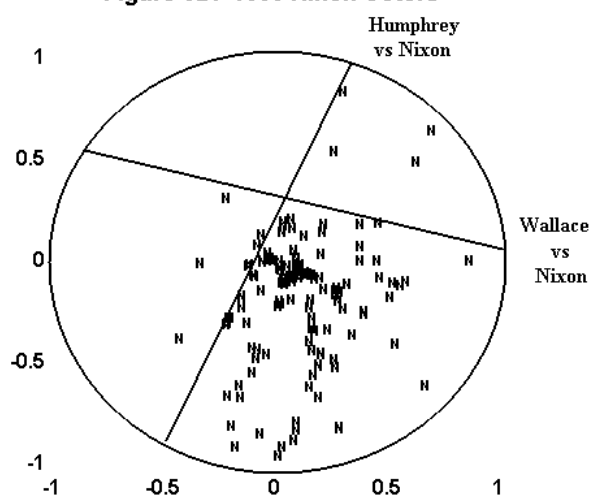

Figure 9E. 1968 Wallace Voters