BAYESIAN METRIC MULTIDIMENSIONAL SCALING

by

Ryan Bakker

Keith T. Poole

Department of Political Science

School of Public and International Affairs

University of Georgia

14 August 2011

(PRELIMINARY)

ABSTRACT

In this paper we show how to apply Bayesian methods to noisy ratio scale distances for both the classical similarities problem as well as the unfolding problem. We show that Bayesian methods produce essentially the same point estimates as the classical methods developed by the Psychometricians in the 1950s and 1960s but the Bayesian methods are superior in that they provide more accurate measures of uncertainty in the data.

Identification is non-trivial for this class of problems because only distances are observed so that a configuration of points that reproduces the distances is only identified up to a choice of origin and a rotation. This can be solved by fixing some of the coordinates of the points but if too many points are held fixed then the correct posterior distribution will not be searched with MCMC methods. The approach we take is to find the optimal solution using "hill climbing" methods such as Nedler-Mead and Powell. The configuration of points from the optimizers is then used to identify the Bayesian posterior. Consequently, we are then able to get a complete picture of the parameters of interest using standard MCMC methods.

1. Introduction

In this paper we take a fresh look at the classical similarities and unfolding problems from the Psychometrics literature using Bayesian methods. Because these problems have been studied for 50 years or more, the solutions are known and various data sets have been used to calibrate a succession of statistical methods. Both problems can easily be handled with frequentist or Bayesian models but Markov chain Monte Carlo (MCMC) methods must be properly constrained to yield meaningful results.

The analysis of ratio scale similarities data by psychometricians in the 1930s through the 1960s led to the development of multidimensional scaling methods (MDS). The psychometricians solved the general problem of representing relational or distance data in a spatial or geometric map where the points represented the stimuli and the distances between the points in the geometric map reproduced the observed distance/relational data. The ratio scale similarities problem was solved by Torgerson (1952, 1958) which in turn built upon work done by pschometricians in the 1930s [Eckart and Young, (1936); Young and Householder (1938)].

In the unfolding problem there are two sets of points – one representing individuals and one representing stimuli. The observed distance/relational data are regarded as expressing the

preferences of individuals; namely, the closer a stimulus point is to an individual point the more the individual prefers that stimulus. The unfolding problem for ratio scale data (the "metric unfolding problem") was first solved by Schönemann (1970).

We first discuss the similarities problem and then we turn to the unfolding problem. Because our Bayesian framework is essentially the same for both problems, we spend more time detailing our solution for the similarities problem because the mathematical exposition is simpler. However, the unfolding problem is of greater interest because most public opinion survey data sets include a set of relational data questions in some form ("where would you place George Bush"; "On a scale of zero to 10, how would you rate John Kerry?"; etc.).

We begin with a discussion of Torgerson's solution to the similarities problem and then we discuss the nature of the constraints necessary for identification. Namely, given only distances, a configuration of points is defined up to a choice of origin and a rotation of the configuration as a whole. The unfolding problem has the same set of constraints.

There are two levels of constraints. In two dimensions only three constraints are necessary to identify the global minimum (if least squares) or global maximum (if Log-Likelihood) for both similarities and unfolding. We show this using analytical

and numerical Hessians. However, there are *reflections* of the global minimum/maximum that are identical in every respect. In two dimensions there are always four solutions that have full rank Hessians. In one dimension there are always two solutions.

We then proceed to a discussion of a Bayesian approach to the similarities and unfolding problems. Given the above result, additional constraints are needed to isolate the correct posterior distribution. We show a simple solution to the problem using a combination of classical hill climbing methods (Nedler and Mead, 1965; Powell, 1973) with a constrained form of MCMC.

2. Classical Metric Scaling

Torgerson's solution to the similarities problem is quite elegant. First, transform the observed similarities/dissimilarities into squared distances. (For example, if the matrix is a Pearson correlation matrix subtract all the entries from 1 and square the result.) Next, double-center the matrix of squared distances by subtracting from each entry in the matrix the mean of the row, the mean of the column, adding the mean of the matrix, and then dividing by -2. This has the effect of removing the squared terms from the matrix leaving just the cross-product matrix (Gower, 1966). Finally,

perform an eigenvalue-eigenvector decomposition to solve for the coordinates.

The statistical properties of Torgerson's solution are unclear. This is due to the fact that similarities/dissimilarities data cannot be negative and the derivatives are not everywhere continuous. For example, denote the observed dissimilarity (distance) as d^*_{jm} where

$$d^*_{jm} = d_{jm} + \varepsilon_{jm} \quad (1)$$

Where j and m are both indices for the stimuli; i.e., j=1,…,q; m=1,…,q. Let Z_{jk} be the j^{th} stimulus coordinate on the k^{th} dimension, k=1,…,s, where s is the number of dimensions. Let d_{jm} be the Euclidean distance between stimulus j and stimulus m in the s-dimensional space:

$$d_{jm} = \sqrt{\sum_{k=1}^{s} (Z_{jk} - Z_{mk})^2} \quad (2)$$

and

$$\varepsilon_{jm} = d^*_{jm} - d_{jm} \sim N(0, \sigma^2) \quad (3)$$

or

$$\varepsilon_{jm} \sim \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2}\left(d^*_{jm} - \sqrt{\sum_{k=1}^{s}(Z_{jk} - Z_{mk})}\right)^2}$$

The joint probability distribution of the sample is:

$$\mathbf{f(D^* \mid Z_{jk})} = \prod_{j=1}^{q-1} \prod_{m=j+1}^{q} \frac{1}{\left(2\pi\sigma^2\right)^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2}\left(d^*_{jm} - \sqrt{\sum_{k=1}^{s}\left(Z_{jk} - Z_{mk}\right)}\right)^2} \qquad (4)$$

Where $\mathbf{D^*}$ is the q by q matrix of observed dissimilarities. Following DeGroot (1986, p. 317), if we regard $\mathbf{f(D^* \mid Z_{jk})}$ as a function of the parameters for given values of the d^*_{jm} then it is a likelihood function; that is

$$\mathbf{L^*(Z_{jk} \mid D^*)} = \frac{1}{\left(2\pi\sigma^2\right)^{\frac{q(q-1)/2}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(d^*_{jm} - \sqrt{\sum_{k=1}^{s}\left(Z_{jk} - Z_{mk}\right)^2}\right)^2} \qquad \mathbf{(5)}$$

Taking the log of the right hand side and dropping the unnecessary constants yields a standard squared error loss function:

$$\ell n\xi = -\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(d^*_{jm} - \sqrt{\sum_{k=1}^{s}\left(Z_{jk} - Z_{mk}\right)^2}\right)^2 = -\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(d^*_{jm} - d_{jm}\right)^2 \qquad (6)$$

The first derivatives of (6) are quite unusual as they are sums of line equations. Namely, they can be written as:

$$\frac{\partial \ell n\xi}{\partial Z_{jk}} = 2\sum_{m \neq j}^{q}\left\{\left(\frac{d^*_{jm}}{d_{jm}} - 1\right)\left(Z_{jk} - Z_{mk}\right)\right\} \qquad \mathbf{(7)}$$

Gleason (1967) points out that all the multidimensional scaling methods then in use for similarities problems (Shepard, 1962a,b; Kruskal, 1964a,b; Lingoes, 1965; Guttman, 1968)

employed some variant of equation (7). The problem with the use of equation (7) is the ratio $\frac{d^*_{jm}}{d_{jm}}$ which is undefined when $Z_j = Z_m$ so that $d_{jm} = 0$. In practice this is not a problem but it and the fact that distances cannot be negative means that the statistical properties are not clear and that the assumption about the error, equation (3), is dicey at best. Nevertheless, finding Z's that minimize (or maximize as in equation (6)) the squared error loss function is relatively easy.

We now turn to a more realistic model of the data. We assume that the observed distances, d^*_{jm}, are drawn from the log-normal distribution because distances are inherently positive:

$$\ln(d^*_{jm}) \sim N(\ln(d_{jm}), \sigma^2) \tag{8}$$

That is

$$f(d^*_{jm}) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}} d^*_{jm}} e^{\left(-\frac{1}{2\sigma^2}\left(\ln(d^*_{jm})-\ln(d_{jm})\right)^2\right)}$$

Hence our likelihood function is:

$$\mathbf{L^*}(Z_{jk}|\mathbf{D^*}) = \frac{1}{(2\pi\sigma^2)^{\frac{q(q-1)/2}{2}}} \left(\prod_{j=1}^{q-1}\prod_{m=j+1}^{q}\frac{1}{d^*_{jm}}\right) e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(\ln\left(d^*_{jm}\right)-\ln\left(\sqrt{\sum_{k=1}^{s}\left(Z_{jk}-Z_{mk}\right)^2}\right)\right)^2} \tag{9}$$

To implement our Bayesian model we use simple normal prior distributions for the stimuli coordinates:

$$\xi(Z_{jk}) = \frac{1}{(2\pi\kappa^2)^{\frac{1}{2}}} e^{-\frac{Z_{jk}^2}{2\kappa^2}} \quad (10)$$

and a simple uniform prior for the variance term:

$$\xi(\sigma^2) = \frac{1}{c},\ 0 < c < b \quad (11)$$

where, empirically, b is no greater than 2.

Hence, our posterior distribution is:

$$\xi(Z_{jk} \mid \mathbf{D}^*) \propto \prod_{j=1}^{q-1} \prod_{m=j+1}^{q} \left\{ f_{jm}(Z_{jm} \mid d_{jm}^*) \right\} \xi(Z_{11})\xi(Z_{12})\ldots\xi(Z_{1s})\xi(Z_{21})\ldots\xi(Z_{qs})\xi(\sigma^2) \quad \mathbf{(12)}$$

Taking the log of the right hand side and dropping the unnecessary constants:

$$\ell n\xi \propto -\frac{q(q-1)/2}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(\ln\left(d_{jm}^*\right) - \ln\left(\sqrt{\sum_{k=1}^{s}\left(Z_{jk} - Z_{mk}\right)^2}\right)\right)^2$$
$$-\frac{1}{2\kappa^2}\left(\sum_{j=1}^{q}\sum_{k=1}^{s} Z_{jk}^2\right) - \ln(c) =$$

$$-\frac{q(q-1)/2}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(\ln\left(d_{jm}^*\right) - \ln\left(d_{jm}\right)\right)^2 - \frac{1}{2\kappa^2}\left(\sum_{j=1}^{q}\sum_{k=1}^{s} Z_{jk}^2\right) - \ln(c) \quad (13)$$

We follow standard practice of using vague priors and set κ=100. In the Appendix we show the first and second derivatives

for (13). In our estimation work we check the solutions with both numerical and analytical first and second derivatives.

Before turning to an example of similarities scaling we first discuss the identification problem inherent in similarities and unfolding data. We then resume our analysis with the appropriate set of constraints.

3. The Problem of Constraints

Assume that our dissimilarities data are squared distances between pairs of stimuli. Our q by q symmetric matrix of data has q(q-1)/2 unique entries (we ignore the diagonal of zeroes). Suppose there is an exact solution; that is, a set of q points in s dimensions that exactly reproduces the squared distances. Clearly, given that we only observe the distances, it does not matter what origin or rotation around that origin we select as long as the configuration of points vis a vis one another is not altered.

With q points in s dimensions we have to solve for q*s parameters. However, we can set any point to the origin – (0,0,…,0) – so this leaves us with q*s – s= (q-1)*s parameters. To pin down the configuration we need to set the rotation. In general a rigid rotation of a configuration is determined by s-1 angles from the origin. For example, in two dimensions the general form of the rotation matrix is:

$$\Gamma = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \; 0 \le \theta \le 2\pi$$

However, note that given a *specific* θ we have *four* rotation matrices:

$$\Gamma_1 = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \quad \Gamma_2 = \begin{bmatrix} -\cos\theta & \sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \Gamma_3 = \begin{bmatrix} \cos\theta & -\sin\theta \\ -\sin\theta & -\cos\theta \end{bmatrix} \quad \Gamma_4 = \begin{bmatrix} -\cos\theta & -\sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

Or

$$\Gamma^* = \Delta\Gamma \text{ where } \Delta = \begin{bmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{bmatrix} \qquad (14)$$

That is, given a specific θ, there are 2^s sign flips corresponding to the s columns of the rotation matrix. With *s*=2, suppose that we have a solution $\tilde{Z}$ such that it reproduces our matrix of squared distances, **D**. Then there are three more solutions corresponding to the above rotation matrices that also exactly reproduce **D**. In general, in *s* dimensions, if we have a solution $\tilde{Z}$ that exactly reproduces the matrix of squared distances then there an additional 2^s-1 solutions that exactly reproduce **D**.

This identification problem is very similar to that discussed by Rivers (2003). He discusses the identification of the classical maximum likelihood factor analysis problem and

shows the number of restrictions necessary to get identification (these include fixing the origin and sign flips). However, his main concern is the identification of the multidimensional IRT model where the data are indicators and he shows that fixing s+1 points (or s(s+1) parameters) fully identifies the model. Our result is different because we assume that we observe (noisy) ratio scale data. Identification is simpler in this setting.

4. A Bayesian MDS Model

To illustrate our approach to similarities scaling, we use agreement scores computed between members of the U.S. 90th (1967-68) Senate. We chose the 90th Senate because it is well known that voting was two dimensional during this period (Poole and Rosenthal, 1997). Given q roll call votes, the agreement score is the number of times a pair of senators vote the same way (Yea, Yea or Nay, Nay) divided by the number of times that they both voted on the same roll calls and multiplied by 100. The agreement scores range from 0 to 100 with 100 indicating identical voting records. Table 1 shows a few Senators and their agreement scores.

Table 1: Agreement Scores for 90th Senate (Partial)

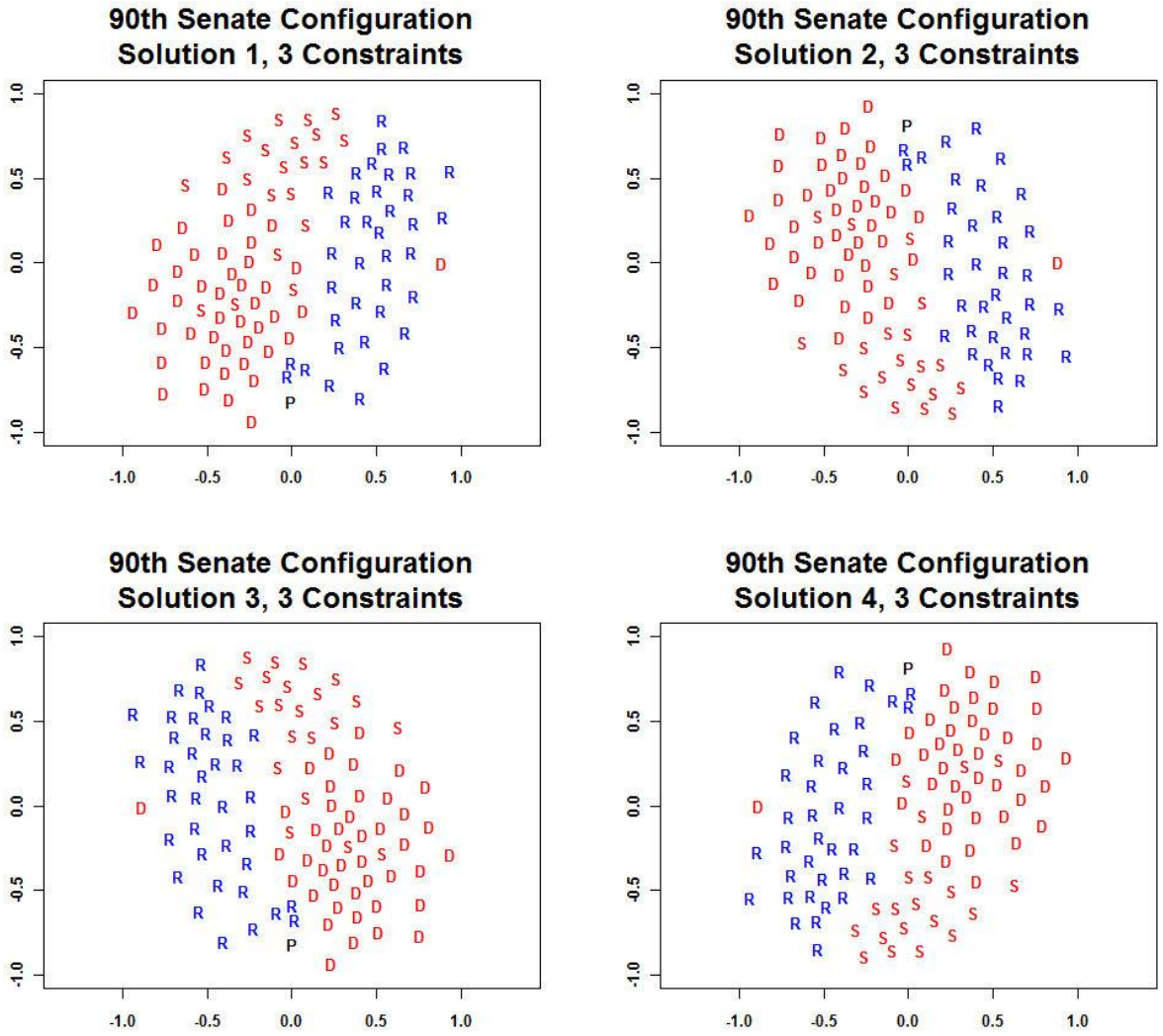
JOHNSON (D-Pres)	100	61	50	52	65	70	37	...
SPARKMAN (D-AL)	61	100	89	50	65	85	65	...
HILL (D-AL)	50	89	100	53	62	78	69	...
GRUENING (D-AK)	52	50	53	100	76	58	43	...
BARTLETT (D-AK)	65	65	62	76	100	70	47	...
HAYDEN (D-AZ)	70	85	78	58	70	100	57	...
FANNIN (R-AZ)	37	65	69	43	47	57	100	...

We convert the agreement scores to distances by subtracting them from 100 and dividing by 50. This is a convenient normalization because the estimated coordinates are usually in the unit hypersphere. Note that we include President Lyndon Johnson in the matrix by using Congressional Quarterly's presidential support roll calls. That is, CQ indicates on a fair number of roll calls whether a Yea/Nay is a vote in favor of the President's position. Hence, the President can be treated as a Senator. He just does not vote as often.

To fix the origin we set Senator Hill (D-AL) at the origin and we fix President Johnson's second dimension coordinate at

zero. We use the Nedler-Mead (1965) amoeba method and the Powell (1973) method to obtain 1001 solutions from random starts. The best solution and its reflections are shown in Figure 1. The tokens in the plots indicate the political party of the member -- "D" for northern Democrat, "S" for southern Democrat, and "R" for Republican. We computed both numerical and analytical first and second derivatives for the optimal solution to show that the Hessian was full rank (i.e., negative-definite; see Appendix A2).

Figure 1: Best 90th Senate Configuration and its Reflections



Of the 1001 solutions for the Bayes posterior, only 3 were the solution (and its reflections) shown in Figure 1. The log-likelihood was about -3100.0. The value for σ^2 was 0.1104. The extreme non-linearity of the log-normal likelihood function meant that a large number of modes were found by the optimizers.

Many of these were quite close together in terms of log-likelihood.

5. Applying MCMC to Similarities Data

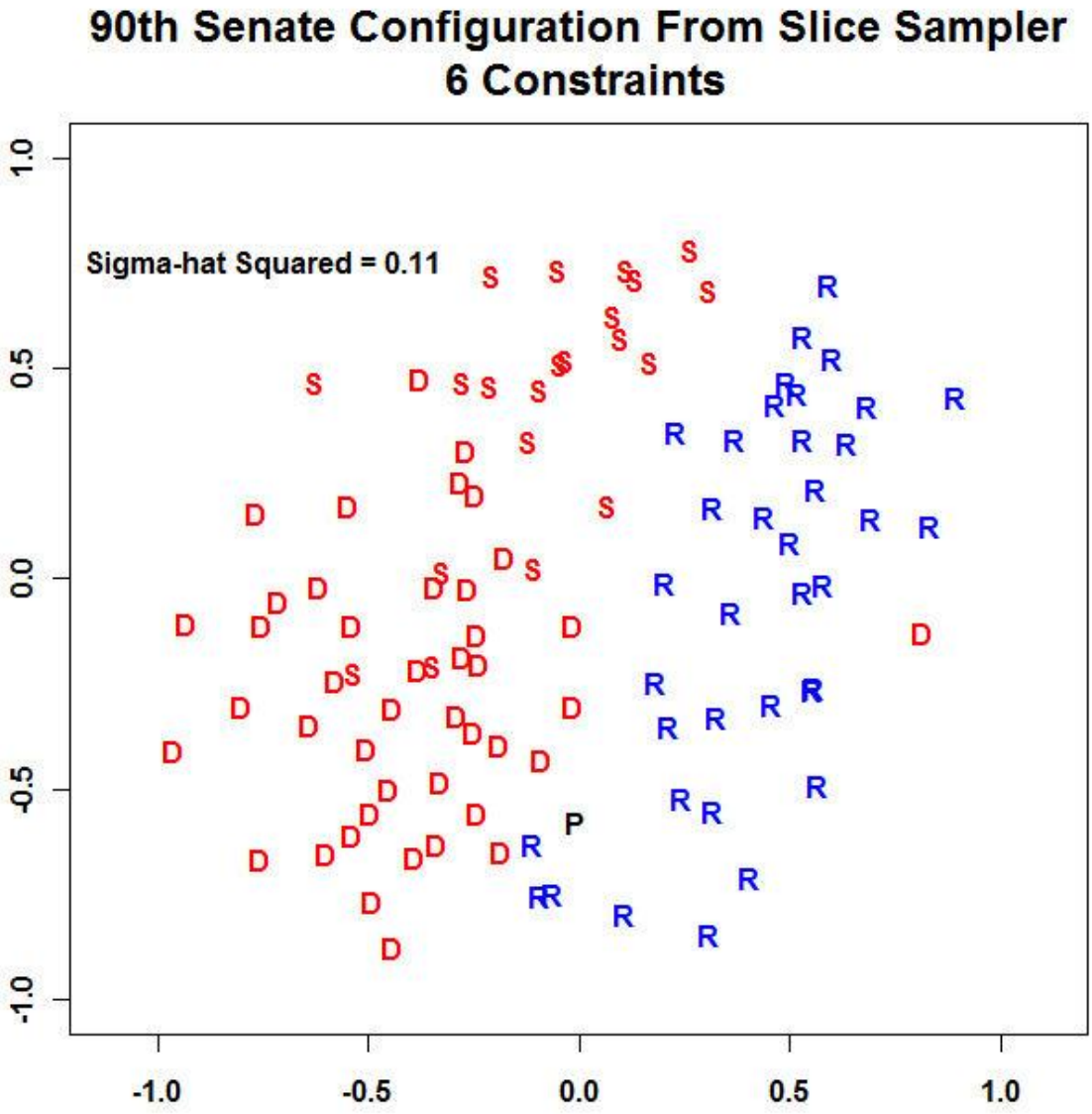
In two dimensions three constraints are enough to pin down four identical posteriors corresponding to the sign flips. This is enough so that an optimizer can find modes. However, three constraints are not enough for the use of MCMC methods because the reflections will cause the chains to settle in around 0.0 for all the parameters. Clearly, if MCMC methods are to be applied to this problem enough constraints must be imposed to remove the reflections. Four loss functions are layered over the hyperplane of the parameters. One must be isolated so that its properties can be analyzed.

For small similarities problems we found that in addition to the origin and one fixed coordinate simply adding three sign constraints to the three fixed coordinates isolated a single posterior. That is, keep the three constraints used to find the modes and then restrict three coordinates to be positive/negative. This works well and it is easy to implement in WinBUGS by using the I(,0) or I(0,) operators.

For larger problems like the 90th Senate agreement scores we retain the origin and one fixed coordinate and then solve for the sign flips by computing simple correlations dimension by dimension between coordinates from each draw in the chain (a configuration of points) and the coordinates from the optimizer solution. This adds four additional constraints corresponding to the sign flips.

Figure 2 shows the results for the 90th Senate (we adjusted the coordinates to -1 to +1 for presentation purposes). We ran our chain out to 110,000 and treated the first 10,000 draws as burn-in. The configuration is the mean of draws 10,001 to 110,000. The configuration is very similar to that shown in Figure 1. The variance term is very precisely estimated with a standard deviation of 0.0026. The standard deviations around the points range from about 0.08 to 0.18 with the largest being 0.25. Additionally, we assessed convergence using the Geweke, Heidelberger-Welch and Raftery and Lewis diagnostics. According to these diagnostics, the posteriors for all parameters meet all criteria for convergence. Note that fixing three coordinates has the effect of "transmitting" the uncertainty associated with those coordinates to other points. There is no solution for this. It is just inherent in the problem.

Figure 2: 90th Senate Using 6 Constraints and Vague Priors



Our approach has the advantage of isolating one posterior distribution and then analyzing it with standard MCMC methods. However, we could simply fix the origin and let the chain wander through the (q-1)*s dimensional hyperplane and post-process the results by rotating each configuration in the chain back to a target configuration. This approach is very similar to that

advocated by Oh and Raftery (2001) and Hoff, Raftery, and Handcock (2002). We prefer our approach because it is computationally simpler and can be implemented in publicly available software such as WinBUGS and JAGS. However, for the unfolding problem we found that a variant of the rotation method to work the best. In Appendix A1 we show our WinBUGS script for the 90th Senate. We used informed priors derived from the Nedler-Mead configuration to stabilize the sampler in WinBUGS.

We now turn to a discussion of how to apply our approach to the unfolding problem.

6. A Bayesian Multidimensional Unfolding Model

In the unfolding problem we have two sets of points – one for individuals and one for stimuli We are given only the noisy ratio scale distances between the two sets and not the distances within each set. Specifically, denote the observed distance as d_{ij}^{*} where

$$d_{ij}^{*} = d_{ij} + \varepsilon_{ij} \qquad (15)$$

Where n is the number of individuals, i=1,…,n, and X_{ik} is the ith individual coordinate on the kth dimension. As before let Z_{jk} be the jth stimulus coordinate on the kth dimension, k=1,…,s, where s

is the number of dimensions. Let d_{ij} be the Euclidean distance between individual i and stimulus j in the s-dimensional space:

$$d_{ij} = \sqrt{\sum_{k=1}^{s}(X_{ik} - Z_{jk})^2} \qquad (16)$$

As before, we assume that the observed distances, d_{ij}^*, are drawn from the log-normal distribution:

$$\ln(d_{ij}^*) \sim N(\ln(d_{ij}), \sigma^2)$$

Which produces the likelihood function:

$$L^*(X_{ik}, Z_{jk} | D^*) = \frac{1}{(2\pi\sigma^2)^{\frac{nq}{2}}} \left(\prod_{i=1}^{n} \prod_{j=1}^{q} \frac{1}{d_{ij}^*} \right) e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \sum_{j=1}^{q} \left(\ln\left(d_{ij}^*\right) - \ln\left(\sqrt{\sum_{k=1}^{s} \left(X_{ik} - Z_{jk}\right)^2} \right) \right)^2} \qquad (17)$$

We use simple normal prior distributions for the individual and stimuli coordinates:

$$\xi(X_{ik}) = \frac{1}{(2\pi\zeta^2)^{\frac{1}{2}}} e^{-\frac{X_{ik}^2}{2\zeta^2}}$$

$$\xi(Z_{jk}) = \frac{1}{(2\pi\kappa^2)^{\frac{1}{2}}} e^{-\frac{Z_{jk}^2}{2\kappa^2}}$$

and a simple uniform prior for the variance term:

$$\xi(\sigma^2) = \frac{1}{c}, \; 0 < c < b$$

where b, empirically, is no greater than 2.

Hence, our posterior distribution is:

$$\xi(X_{ik}, Z_{jk} \mid D^*) \propto \prod_{i=1}^{n}\prod_{j=1}^{q}\left\{ f_{ij}(X_{ik}, Z_{jk} \mid d_{ij}^*)\right\}\xi(X_{11})...\xi(X_{ns})\xi(Z_{11})...\xi(Z_{qs})\xi(\sigma^2) \qquad (18)$$

Taking the log of the right hand side and dropping the unnecessary constants:

$$\ell n\xi \propto -\frac{nq}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{q}\left(\ln\left(d_{ij}^*\right) - \ln\left(\sqrt{\sum_{k=1}^{s}\left(X_{ik} - Z_{jk}\right)^2}\right)\right)^2$$
$$-\frac{1}{2\zeta^2}\left(\sum_{i=1}^{n}\sum_{k=1}^{s}X_{ik}^2\right) - \frac{1}{2\kappa^2}\left(\sum_{j=1}^{q}\sum_{k=1}^{s}Z_{jk}^2\right) - \ln(c) =$$

$$-\frac{nq}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{q}\left(\ln\left(d_{ij}^*\right) - \ln\left(d_{ij}\right)\right)^2 - \frac{1}{2\zeta^2}\left(\sum_{i=1}^{n}\sum_{k=1}^{s}X_{ik}^2\right) - \frac{1}{2\kappa^2}\left(\sum_{j=1}^{q}\sum_{k=1}^{s}Z_{jk}^2\right) - \ln(c) \quad (19)$$

We follow standard practice of using vague priors and set ζ=100 and κ=100. In the Appendix we show the first and second derivatives for (19).

Our unfolding example is the classic 1968 National Election Study feeling thermometers. A feeling thermometer asks individuals to respond to a set of stimuli (political figures in this case) based on their subjective views of warmth towards them. The thermometer ranges from 0 to 100 degrees with 100 indicating warm and very favorable feeling, 50 indicating neutrality towards the political figure, and 0 indicating that the respondent feels cold and very unfavorable towards the

political figure. The 1968 feeling thermometers have been analyzed by Weisberg and Rusk (1970), Wang, et al. (1975), Rabinowitz (1976), Cahoon, et al. (1978), Poole and Rosenthal (1984), and Brady (1990) with the main focus on modeling the latent dimensions underlying the thermometers as well as testing theories of spatial voting.

In the NES 1968 survey twelve political figures were included in the thermometer questions: George Wallace, Hubert Humphrey, Richard Nixon, Eugene McCarthy, Ronald Reagan, Nelson Rockefeller, President Johnson, George Romney, Robert Kennedy, Edmund Muskie, Spiro Agnew, and Curtis LeMay. There were 1,673 respondents and we included in our analysis the 1,392 respondents who rated at least five of the twelve political figures.

We perform our analysis in two dimensions because previous analyses using optimization methods almost all find two dimensions in the data. We think this is due to the idiosyncratic noise in the thermometers (see Abrajano and Poole, 2011, for a discussion) and valence effects (Londregan, 2000; Merrill and Grofman, 1999; Adams, Merrill, and Grofman, 2005). A second dimension is picking up some of these effects and "smoothing" out the first dimension. Modeling valence effects is difficult so we leave that for future work. In any event, our aim here is to show the advantages of our Bayesian approach.

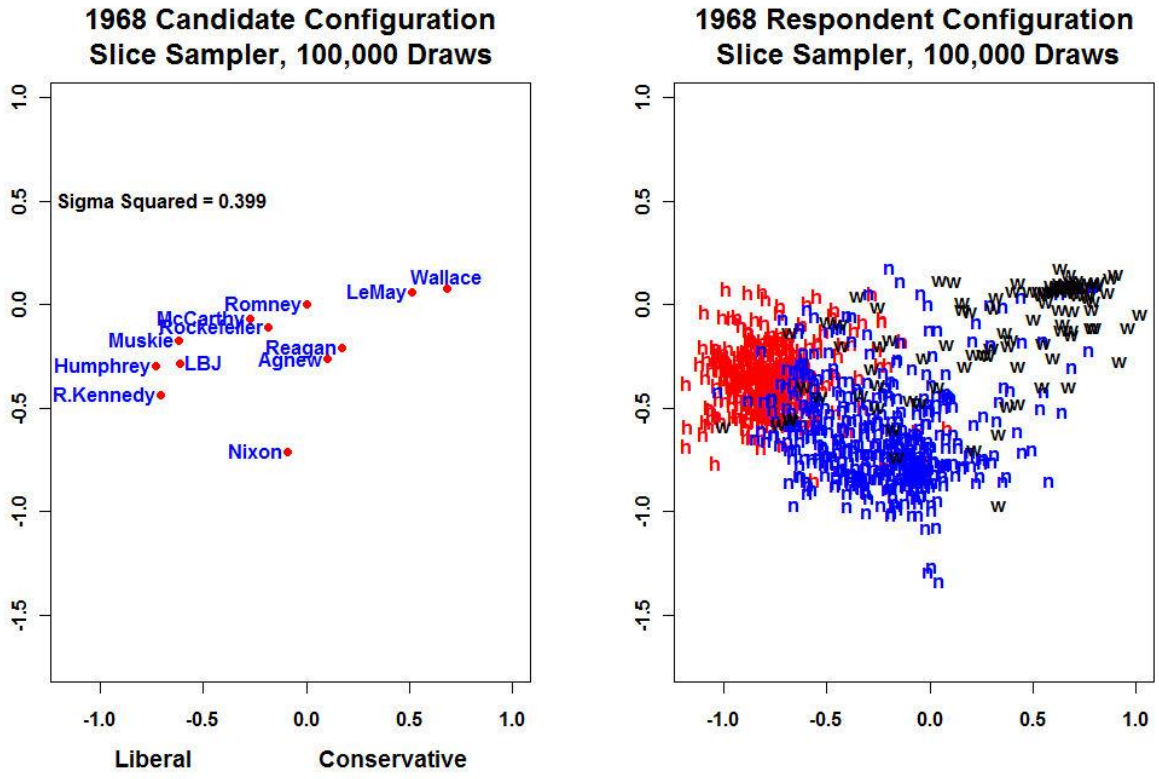
Namely, a properly designed Markov chain reveals much more information than simply the modes of a loss function.

Finding the optimal solution for the unfolding problem is difficult because of the number of parameters. In two dimensions this would require a search over a 2,805 dimensional hyperplane. This was impractical. Instead, what we found to work well was to simply fix one candidate at the origin and another candidate at zero on one of the dimensions. Holding the candidate points fixed, it was easy to find the optimal point for each respondent. Then, holding the respondent points fixed, it was again easy to find the optimal points for the candidates. We continued this until there was no further improvement. At each step we used Nedler-Mead (1965) to find the points. We checked the first derivatives (see Appendix) for the starting configuration to be sure that our points were located on modes of the loss function.

In practice we set George Romney at the origin and Eugene McCarthy's second dimension coordinate at zero. Using the respondent and candidate coordinates as targets we were able to run a slice sampler on the 1968 data. Because the ratio of respondents to the candidates is so large, we found that the method that worked the best was to first draw the respondent coordinates and then the candidate coordinates. We kept Romney at the origin but we did not constrain any other points because

we found that simply rotating the drawn configuration to the optimal configuration with Romney at the origin was simple and easy to implement. We ran our chain to 110,000 draws with the first 10,000 as burn-in. Figure 3 shows the results.

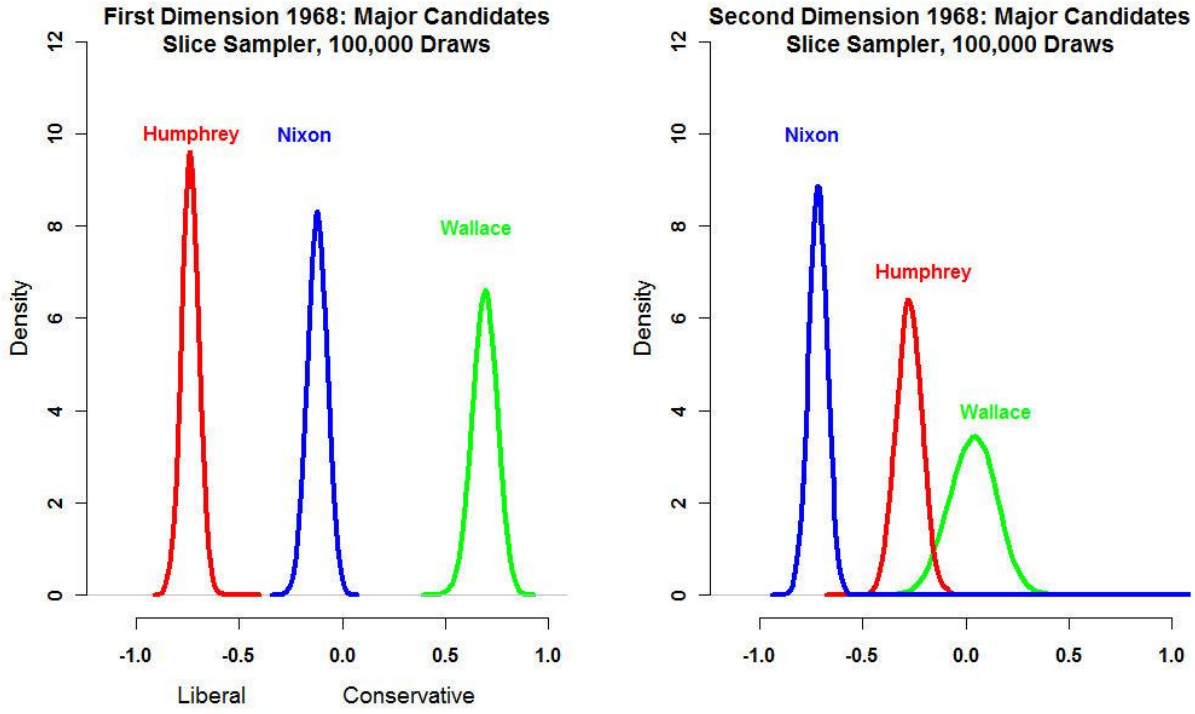
Figure 3: 1968 Thermometer Unfolding Example



The left panel of Figure 3 shows the candidate configuration and the right panel shows the respondents. We display those respondents who indicated that they voted for Humphrey, Nixon, or Wallace, as the tokens "h", "n", or "w", respectively. Humphrey, Nixon, and Wallace are located near where their voters are concentrated.

The candidates are very precisely estimated. The largest standard deviation was for George Wallace's second dimension coordinate at 0.11. Figure 4 shows the 100,000 draws (after burn-in) for the three major Presidential candidates. All are unimodal and appear to be symmetric.

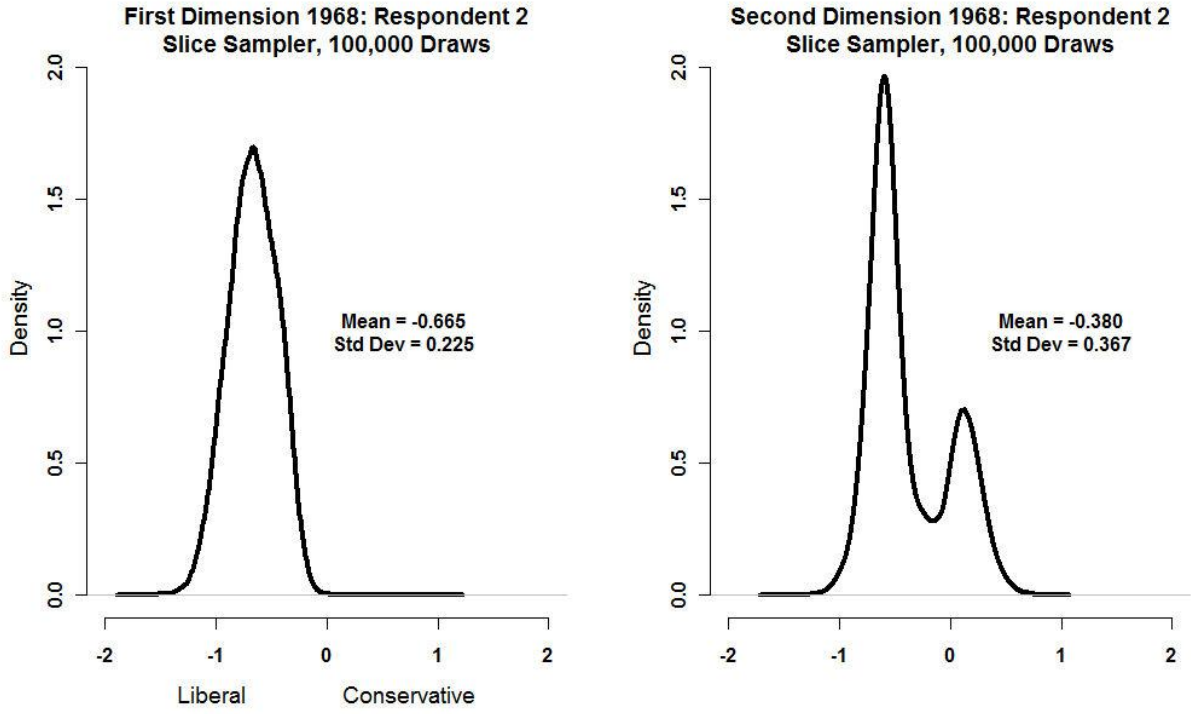
Figure 4: Major Candidates 1968 Presidential Election



The respondents were less precisely estimated. For example, Figure 5 shows the distribution of the 100,000 draws for the coordinates of the respondent number 2. The 2nd respondent was a young white male Democrat. He did not like Wallace, LeMay, Agnew, and Reagan (15, 30, 30, 30) but he was a

little warmer towards Nixon and Romney (40, 40). He was less than enthused with President Johnson and Hubert Humphrey (50 and 60) but he really liked Robert Kennedy, Nelson Rockefeller, and Eugene McCarthy (97, 97, 85). His preferences roughly line up left to right but not entirely. This is reflected in the distribution of the draws. The draws on the first dimension are unimodal with a reasonable standard deviation but the draws on the second dimension have two modes with a large standard deviation. A mode finder (optimization method) will land on one of the two modes whereas a Markov chain "illuminates" the entire distribution and recovers the *means*.

Figure 5: 1968 NES Respondent 2



7. Conclusion

In this paper we have shown how to apply Bayesian methods to noisy ratio scale distances for both the classical similarities problem as well as the unfolding problem. Our approach combines the advantages of traditional mode finders and Bayesian MCMC. We use the mode finders to give us a target that identifies (“freezes”) the posterior for the Markov chain generator.

Our unfolding example using the 1968 candidate thermometers shows the power of MCMC (made possible by the speed of modern computers) to illuminate complex distributions. Instead of modes with their associated standard deviations from the inverse Hessian, "painting" the entire posterior distribution allows us to show means and the complete distribution of the parameters.

Our results are preliminary. We deliberately kept our models simple because our aim was to revisit older problems using modern methods. We think the thermometers are an underutilized resource that potentially can reveal important information about individuals' utilities for political figures. Our aim here was simply to show a basic method that can be used as a springboard to more complex analyses.

Appendix

A1 WINBUGS SIMILARITIES MODEL

```
#
#  MDS Model for 90th Senate--over constrained
#
model{

#  Fix one point
#
        x[8,1] <- -0.626000480
……….   x[8,2] <-  0.46524749
#
# llh and sumllh monitor the log-likelihood
#
for (i in 1:101){
    llh[i,i] <- 0.0
    for (j in i+1:102){
#
#  Read in Distances rather than the similarities (makes handling missing data easier)
#
         dstar[i,j] ~ dlnorm(mu[i,j],tau)
         mu[i,j] <- log(sqrt((x[i,1]-x[j,1])*(x[i,1]-x[j,1])+(x[i,2]-x[j,2])*(x[i,2]-x[j,2])))
         llh[i,j] <- (log(dstar[i,j])-mu[i,j])*(log(dstar[i,j])-mu[i,j])
         llh[j,i] <- (log(dstar[i,j])-mu[i,j])*(log(dstar[i,j])-mu[i,j])
    }
}

   llh[102,102] <- 0.0
   sumllh <- sum(llh[,])
#
  ## priors
  tau ~ dgamma(1,1)

#
# Informed priors placed below (not all shown)
#
  x[1,1] ~ dnorm(0,.1) I(0,)
  x[1,2] ~ dnorm(0,.1) I(,0)
  x[2,1] ~ dnorm(0,.1) I(,0)
  x[2,2] ~ dnorm(0,.1) I(0,)

...etc. etc.

  x[98,1] ~ dnorm(0,.1) I(,0)
  x[98,2] ~ dnorm(0,.1) I(0,)
  x[99,1] ~ dnorm(0,.1) I(,0)
  x[99,2] ~ dnorm(0,.1) I(, -0.5)
  x[100,1] ~ dnorm(0,.1) I(,0)
  x[100,2] ~ dnorm(0,.1) I(,-0.5)
  x[101,1] ~ dnorm(0,.1) I(0.5,)
  x[101,2] ~ dnorm(0,.1) I(0.2,)
  x[102,1] ~ dnorm(0,.1) I(,-0.2)
  x[102,2] ~ dnorm(0,.1) I(,0)

}
```

A2 The Derivatives for the Log-Normal Bayesian Model

Similarities: The first derivatives for the similarities problem are:

$$\frac{\partial \ell n\xi}{\partial Z_{jk}} = -2\frac{1}{2\sigma^2}\sum_{j\neq m}^{q}\left\{\left(\ln\left(d_{jm}^{*}\right)-\ln\left(d_{jm}\right)\right)\left(-\frac{1}{d_{jm}}\right)\left(\frac{1}{2}\right)\left[\sum_{k=1}^{s}(Z_{jk}-Z_{mk})^2\right]^{-\frac{1}{2}}\left(2\left[Z_{jk}-Z_{mk}\right]\right)\right\}-\frac{Z_{jk}}{\kappa^2}$$

which simplifies to

$$\frac{\partial \ell n\xi}{\partial Z_{jk}} = \frac{1}{\sigma^2}\sum_{j\neq m}^{q}\left\{\frac{\left(\ln\left(d_{jm}^{*}\right)-\ln\left(d_{jm}\right)\right)}{d_{jm}^{2}}\left(Z_{jk}-Z_{mk}\right)\right\}-\frac{Z_{jk}}{\kappa^2} \qquad \textbf{(A1)}$$

and

$$\frac{\partial \ell n\xi}{\partial \sigma^2} = -\frac{q(q-1)}{4\sigma^2}+\frac{1}{2\sigma^4}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(\ln\left(d_{jm}^{*}\right)-\ln\left(d_{jm}\right)\right)^2 \qquad \textbf{(A2)}$$

Hence, we get the usual result for the variance term:

$$\hat{\sigma}^2 = \frac{2}{q(q-1)}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(ln\left(d_{jm}^{*}\right)-ln\left(d_{jm}\right)\right)^2 \qquad \textbf{(A3)}$$

Note that since κ^2 is a vague prior, the practical effect is that at an inflection point we have $\frac{\partial^2 \ell n\xi}{\partial Z_{jk}\sigma^2} \approx \frac{\partial \ell n\xi}{\partial Z_{jk}} = 0$. Numerically, this is a handy result because it makes computing the inverse Hessian much easier to accomplish.

The second derivative for the variance is:

$$\frac{\partial^2 \ell n\xi}{\partial\sigma^2\partial\sigma^2} = \frac{q(q-1)}{4\sigma^4} - \frac{1}{\sigma^6}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(\ln\left(d_{jm}^*\right) - \ln\left(d_{jm}\right)\right)^2 \qquad \textbf{(A4)}$$

Substituting (A3) into (A4) it is easy to show that $\frac{\partial^2 \ell n\xi}{\partial\sigma^2\partial\sigma^2} < 0$ so that when the global maximum for the Z_{jk} is found σ^2 will be a maximum as well.

The second derivatives for the coordinates are:

$$\frac{\partial^2 \ell n\xi}{\partial Z_{jk}\partial Z_{jk}} = -4\sum_{j\neq m}^{q}\frac{\left(\ln\left(d_{jm}^*\right) - \ln\left(d_{jm}\right)\right)}{d_{jm}^4}\left(Z_{jk} - Z_{mk}\right)^2 - 2\sum_{j\neq m}^{q}\left[\frac{\left(Z_{jk} - Z_{mk}\right)^2}{d_{jm}^4}\right] + 2\sum_{j\neq m}^{q}\left[\frac{\left(\ln\left(d_{jm}^*\right) - \ln\left(d_{jm}\right)\right)}{d_{jm}^2}\right] - \frac{1}{\kappa^2}$$

(A5)

$$\frac{\partial^2 \ell n\xi}{\partial Z_{jk}\partial Z_{mk}} = 4\frac{\left(\ln\left(d_{jm}^*\right) - \ln\left(d_{jm}\right)\right)}{d_{jm}^4}\left(Z_{jk} - Z_{mk}\right)^2 + 2\frac{\left(Z_{jk} - Z_{mk}\right)^2}{d_{jm}^4} - 2\frac{\left(\ln\left(d_{jm}^*\right) - \ln\left(d_{jm}\right)\right)}{d_{jm}^2} \qquad \textbf{(A6)}$$

In more than one dimension

$$\frac{\partial^2 \ell n\xi}{\partial Z_{jk}\partial Z_{j\ell}} = -2\sum_{j\neq m}^{q}\left\{\left[\frac{\left(Z_{jk} - Z_{mk}\right)\left(Z_{j\ell} - Z_{m\ell}\right)}{d_{jm}^4}\right]\left[2\left(\ln\left(d_{jm}^*\right) - \ln\left(d_{jm}\right)\right) - 1\right]\right\} \qquad \textbf{(A7)}$$

$$\frac{\partial^2 \ell n\xi}{\partial Z_{jk}\partial Z_{m\ell}} = 2\frac{\left(Z_{jk} - Z_{mk}\right)\left(Z_{j\ell} - Z_{m\ell}\right)}{d_{jm}^4}\left[2\left(\ln\left(d_{jm}^*\right) - \ln\left(d_{jm}\right)\right) + 1\right] \qquad \textbf{(A8)}$$

where $\ell = 1,\ldots,s$ and $\ell \neq k$.

Unfolding: The first derivatives for the unfolding problem are:

$$\frac{\partial \ell n\xi}{\partial X_{ik}} = \frac{1}{\sigma^2}\sum_{j=1}^{q}\left\{\frac{\left(\ln\left(d_{ij}^*\right) - \ln\left(d_{ij}\right)\right)}{d_{ij}^2}\left(X_{ik} - Z_{jk}\right)\right\} - \frac{X_{ik}}{\zeta^2} \qquad \textbf{(A9)}$$

$$\frac{\partial \ell n\xi}{\partial Z_{jk}} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}\left\{\frac{\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)}{d_{ij}^{2}}\left(X_{ik} - Z_{jk}\right)\right\} - \frac{Z_{jk}}{\kappa^2} \qquad \textbf{(A10)}$$

and

$$\frac{\partial \ell n\xi}{\partial \sigma^2} = -\frac{nq}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}\sum_{j=1}^{q}\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)^2 \qquad \textbf{(A11)}$$

Hence, we get the usual result for the variance term for the unfolding model:

$$\hat{\sigma}^2 = \frac{1}{nq}\sum_{i=1}^{n}\sum_{j=1}^{q}\left(ln\left(d_{ij}^{*}\right)-ln\left(d_{ij}\right)\right)^2 \qquad \textbf{(A12)}$$

Note that since ζ^2 and κ^2 are vague priors, the practical effect is that at an inflection point we have $\frac{\partial^2 \ell n\xi}{\partial X_{ik}\sigma^2} \approx \frac{\partial \ell n\xi}{\partial X_{ik}} = 0$.

The second derivative for the variance is:

$$\frac{\partial^2 \ell n\xi}{\partial \sigma^2 \partial \sigma^2} = \frac{nq}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}\sum_{j=1}^{q}\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)^2 \qquad \textbf{(A13)}$$

Substituting (A12) into (A13) it is easy to show that $\frac{\partial^2 \ell n\xi}{\partial \sigma^2 \partial \sigma^2} < 0$ so that when the global maximum for the X_{ik} and Z_{jk} is found σ^2 will be a maximum as well.

The second derivatives for the coordinates are:

$$\frac{\partial^2 \ell n\xi}{\partial X_{ik}\partial X_{ik}} = -2\sum_{j=1}^{q}\frac{\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)}{d_{ij}^{4}}\left(X_{ik}-Z_{jk}\right)^{2}-\sum_{j=1}^{q}\left[\frac{\left(X_{ik}-Z_{jk}\right)^{2}}{d_{ij}^{4}}\right]+\sum_{j=1}^{q}\left[\frac{\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)}{d_{ij}^{2}}\right]-\frac{1}{\zeta^{2}} \quad \textbf{(A14)}$$

$$\frac{\partial^2 \ell n\xi}{\partial Z_{jk}\partial Z_{jk}} = 2\sum_{i=1}^{n}\frac{\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)}{d_{ij}^{4}}\left(X_{ik}-Z_{jk}\right)^{2}+\sum_{i=1}^{n}\left[\frac{\left(X_{ik}-Z_{jk}\right)^{2}}{d_{ij}^{4}}\right]-\sum_{i=1}^{n}\left[\frac{\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)}{d_{ij}^{2}}\right]-\frac{1}{\kappa^{2}} \quad \textbf{(A15)}$$

$$\frac{\partial^2 \ell n\xi}{\partial X_{ik}\partial Z_{jk}} = \left[\frac{\left(X_{ik}-Z_{jk}\right)^{2}}{d_{ij}^{4}}\right]\left[2\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)+1\right]-\frac{\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)}{d_{ij}^{2}} \quad \textbf{(A16)}$$

$$\frac{\partial^2 \ell n\xi}{\partial X_{ik}\partial X_{hk}} = \frac{\partial^2 \ell n\xi}{\partial Z_{jk}\partial Z_{mk}} = 0 \quad \textbf{(A17)}$$

Where h=1,...,n and h≠i. In more than one dimension

$$\frac{\partial^2 \ell n\xi}{\partial X_{ik}\partial X_{i\ell}} = -\sum_{j=1}^{q}\left\{\left[\frac{\left(X_{ik}-Z_{jk}\right)\left(X_{i\ell}-Z_{j\ell}\right)}{d_{ij}^{4}}\right]\left[2\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)+1\right]\right\} \quad \textbf{(A18)}$$

$$\frac{\partial^2 \ell n\xi}{\partial Z_{jk}\partial Z_{j\ell}} = -\sum_{i=1}^{n}\left\{\left[\frac{\left(X_{ik}-Z_{jk}\right)\left(X_{i\ell}-Z_{j\ell}\right)}{d_{ij}^{4}}\right]\left[2\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)+1\right]\right\} \quad \textbf{(A19)}$$

$$\frac{\partial^2 \ell n\xi}{\partial X_{ik}\partial Z_{j\ell}} = \left[\frac{\left(X_{ik}-Z_{jk}\right)\left(X_{i\ell}-Z_{j\ell}\right)}{d_{ij}^{4}}\right]\left[2\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)+1\right] \quad \textbf{(A20)}$$

$$\frac{\partial^2 \ell n\xi}{\partial X_{ik}\partial X_{h\ell}} = \frac{\partial^2 \ell n\xi}{\partial Z_{jk}\partial Z_{m\ell}} = 0 \quad \textbf{(A21)}$$

References

Abrajano, Marisa and Keith T. Poole. 2011. "A Method of Linking Surveys Using Affective Signatures with an Application to Racial/Ethnic Groups in the U.S." In *Who Gets Represented?* Edited by Peter Enns and Christopher Wlezien. New York: Russell Sage Foundation.

Adams, James, Samuel Merrill and Bernard Grofman. 2005. *A Unified Theory of Party Competition: A Cross-National Analysis Integrating Spatial and Behavioral Factors*. New York: Cambridge University Press.

Brady, Henry. 1990. "Traits Versus Issues: Factor Versus Ideal-Point Analysis of Candidate Thermometer Ratings." *Political Analysis*, 2:97-129.

Cahoon, Lawrence S., Melvin J. Hinich, and Peter C. Ordeshook. 1978. "A Statistical Multidimensional Scaling Method based on the Spatial Theory of Voting." In *Graphical Representation of Multivariate Data*, ed. by P.C. Wang. New York, Academic Press.

DeGroot, Morris H. 1986. *Probability and Statistics* (2nd edition). New York: Addison Wesley.

Eckart, Carl H. and Gale Young. 1936. "The Approximation of One Matrix by Another of Lower Rank." *Psychometrika*, 1:211-218.

Gleason, Terry C. 1967. "A General Model for Nonmetric Multidimensional Scaling." Working Paper, Michigan Mathematical Psychology Program, MMPP 67-3.

Gower, John C. 1966. "Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis." *Biometrika*, 53:325-338.

Guttman, Louis. 1968. "A General Nonmetric Technique for Finding the Smallest Coordinate Space for a Configuration of Points." *Psychometrika,* 33:469-506.

Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock. 2002. "Latent Space Approaches to Social Network Analysis." *Journal of the American Statistical Association*, 97 (No. 460):1090-1098.

Kruskal, Joseph B. 1964a. "Multidimensional Scaling by Optimizing a Goodness of Fit to a Nonmetric Hypothesis." *Psychometrika*, 29:1-27.

Kruskal, Joseph B. 1964b. "Nonmetric Multidimensional Scaling: A Numerical Method." *Psychometrika*, 29:115-129.

Lingoes, J. C. 1965. "An IBM-7090 Program for Guttman-Lingoes Smallest Space Analysis—I." *Behavioral Science*, 10:183-184.

Londregan, John B. 2000. "Estimating Legislators' Preferred Points." *Political Analysis*, 8(1):35-56.

Merrill, Samuel III and Bernard Grofman. 1999. *A Unified Theory of Voting: Directional and Proximity Spatial Models*. New York: Cambridge University Press.

Nedler, John A. and Roger Mead. 1965. "A simplex method for function minimization." *Computer Journal* **7**: 308–313.

Oh, Man-Suk and Adrian E. Raftery. 2001. "Bayesian Multidimensional Scaling and Choice of Dimension." *Journal of the American Statistical Association*, 96 (No. 455): 1031-1044.

Poole, Keith T. and Howard Rosenthal. 1984. "U.S. Presidential Elections 1968-1980: A Spatial Analysis." *American Journal of Political Science*, 28:282-312.

Poole, Keith T. and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting.* New York: Oxford University Press.

Powell, Michael J. D. 1973. "On Search Directions for Minimization Algorithms." *Mathematical Programming* 4: 193–201.

Rabinowitz, George. 1976. "A Procedure for Ordering Object Pairs Consistent with the Multidimensional Unfolding Model." *Psychometrika*, 45:349-373.

Rivers, Douglas. 2003. "Identification of Multidimensional Spatial Voting Models." Working Paper, Stanford University (http://polmeth.wustl.edu/media/Paper/river03.pdf).

Schönemann, Peter H. 1970. "On Metric Multidimensional Unfolding." *Psychometrika*, 35:349-366.

Shepard, Roger N. 1962a. "The Analysis of Proximities: Multidimensional Scaling With an Unknown Distance Function. I." *Psychometrika*, 27:125-139.

Shepard, Roger N. 1962b. "The Analysis of Proximities: Multidimensional Scaling With an Unknown Distance Function. II." *Psychometrika*, 27: 219-246.

Torgerson, Warren S. 1952. "Multidimensional Scaling: I. Theory and Method. *Psychometrika*, 17:401-419.

Torgerson, Warren S. 1958. *Theory and Methods of Scaling*. New York: Wiley.

Wang, Ming-Mei, Peter H. Schonemann, and Jerrold G. Rusk. 1975. "A Conjugate Gradient Algorithm for the Multidimensional Analysis of Preference Data." *Multivariate Behavioral Research*, 10:45-80.

Weisberg, Herbert F. and Jerrold G. Rusk. 1970. "Dimensions of Candidate Evaluation." *American Political Science Review*, 64:1167-1185.

Young, Gale and Alston S. Householder. 1938. "Discussion of a Set of Points in Terms of their Mutual Distances." *Psychometrika*, 3:19-22.