

**Proof that if Voting is Perfect in One Dimension, then the First
Eigenvector Extracted from the Double-Centered Transformed
Agreement Score Matrix has the Same Rank Ordering as the True Data**

Keith T. Poole

University of Houston

24 February 2001

Notation and Definitions

Let the true ideal points of the \mathbf{p} legislators be denoted as $\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \dots, \tilde{\mathbf{X}}_p$.

Without loss of generality let the ordering of the true ideal points of the legislators on the dimension from left to right be:

$$\tilde{\mathbf{X}}_1 \leq \tilde{\mathbf{X}}_2 \leq \tilde{\mathbf{X}}_3 \leq \dots \leq \tilde{\mathbf{X}}_p$$

Let \mathbf{q} be the number of *non-unanimous* roll call votes with $\mathbf{q} > \mathbf{0}$ and let the cutpoint for the j th roll call be \mathbf{Z}_j . Voting is *perfect*. That is, all legislators are sincere voters and all legislators to the left of a cutpoint vote for the same alternative and all legislators to right of a cutpoint vote for the opposite alternative. For example, if all legislators to the left of \mathbf{Z}_j vote “Nay”, then all legislators to the right of \mathbf{Z}_j vote “Yea”. Without loss of generality we can assume that every legislator to the left of \mathbf{Z}_j votes “Yea” and every legislator to the right of \mathbf{Z}_j votes “Nay”. That is, the “polarity” of the roll call does not affect the analysis below.

Let k_1 be the number of cutpoints between legislators 1 and 2, k_2 be the number of cutpoints between legislators 2 and 3, and so on, with k_{p-1} being the number of cutpoints between legislators $p-1$ and p . Hence

$$q = \sum_{i=1}^{p-1} k_i > 0 \quad (1)$$

The *agreement score* between two legislators is the simple proportion of roll calls that they vote for the same outcome. Hence, the agreement score between legislators 1

and 2 is simply $\frac{q - k_1}{q}$ because 1 and 2 agree on all roll calls except for those with

cutpoints between them. Similarly, the agreement score between legislators 1 and 3 is

$\frac{q - k_1 - k_2}{q}$ and the agreement score between legislators 2 and 3 is $\frac{q - k_2}{q}$. In general,

for two legislators X_a and X_b , where $a \neq b$, the agreement score is:

$$A_{ab} = \frac{q - \sum_{i=a}^{b-1} k_i}{q} \quad (2)$$

These definitions allow me to state the following theorem:

***Theorem:* If Voting is Perfect in One Dimension, then the First Eigenvector Extracted From the Double Centered p by p Matrix of Squared Distances from Equation (3) has at Least the Same Weak Monotone Rank Ordering as the Legislators.**

Proof: The agreement scores can be treated as Euclidean distances by simply subtracting them from 1. That is:

$$\mathbf{d}_{ab} = 1 - \mathbf{A}_{ab} = 1 - \frac{\mathbf{q} - \sum_{i=a}^{b-1} \mathbf{k}_i}{\mathbf{q}} = \frac{\sum_{i=a}^{b-1} \mathbf{k}_i}{\mathbf{q}} \quad (3)$$

The \mathbf{d} 's computed from equation (3) satisfy the three axioms of distance: they are non-negative because by (2) $0 \leq \mathbf{A}_{ab} \leq 1$ so that $0 \leq \mathbf{d}_{ab} \leq 1$; they are symmetric, $\mathbf{d}_{ab} = \mathbf{d}_{ba}$; and they satisfy the triangle inequality. To see this, consider any triple of points $\mathbf{X}_a < \mathbf{X}_b < \mathbf{X}_c$. The distances are:

$$\mathbf{d}_{ab} = \frac{\sum_{i=a}^{b-1} \mathbf{k}_i}{\mathbf{q}} \quad \text{and} \quad \mathbf{d}_{bc} = \frac{\sum_{i=b}^{c-1} \mathbf{k}_i}{\mathbf{q}} \quad \text{and} \quad \mathbf{d}_{ac} = \frac{\sum_{i=a}^{c-1} \mathbf{k}_i}{\mathbf{q}}$$

Hence

$$\mathbf{d}_{ac} = \mathbf{d}_{ab} + \mathbf{d}_{bc} \quad (4)$$

Because all the triangle inequalities are *equalities*, in Euclidean geometry this implies that $\mathbf{X}_a, \mathbf{X}_b$, and \mathbf{X}_c *all lie on a straight line* (Borg and Groenen, 1997, ch. 18).

Because all the triangle inequalities are equalities and all triples of points lie on a straight line, the distances computed from (2) can be directly written as distances between points:

$$\mathbf{d}_{ab} = \frac{\sum_{i=a}^{b-1} \mathbf{k}_i}{\mathbf{q}} = |\mathbf{X}_a - \mathbf{X}_b| \quad (5)$$

where $\mathbf{d}_{aa} = 0$. The p by p matrix of squared distances is:

$$\mathbf{D} = \begin{bmatrix} \mathbf{0} & (\mathbf{X}_2 - \mathbf{X}_1)^2 & \cdot & \cdot & \cdot & (\mathbf{X}_p - \mathbf{X}_1)^2 \\ (\mathbf{X}_1 - \mathbf{X}_2)^2 & \mathbf{0} & \cdot & \cdot & \cdot & (\mathbf{X}_p - \mathbf{X}_2)^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ (\mathbf{X}_1 - \mathbf{X}_p)^2 & (\mathbf{X}_2 - \mathbf{X}_p)^2 & \cdot & \cdot & \cdot & \mathbf{0} \end{bmatrix} \quad (6)$$

To recover the \mathbf{X} 's, simply double-center \mathbf{D} and perform an eigenvalue-eigenvector decomposition. The first eigenvector is the solution. To see this:

$$\text{Let the mean of the } j\text{th column of } \mathbf{D} \text{ be } \mathbf{d}_{\cdot j}^2 = \frac{\sum_{i=1}^p \mathbf{d}_{ij}^2}{p} = \mathbf{X}_j^2 - 2\mathbf{X}_j\bar{\mathbf{X}} + \frac{\sum_{i=1}^p \mathbf{X}_i^2}{p}.$$

$$\text{Let the mean of the } i\text{th row of } \mathbf{D} \text{ be } \mathbf{d}_{i\cdot}^2 = \frac{\sum_{j=1}^p \mathbf{d}_{ij}^2}{p} = \mathbf{X}_i^2 - 2\mathbf{X}_i\bar{\mathbf{X}} + \frac{\sum_{j=1}^p \mathbf{X}_j^2}{p}.$$

$$\text{Let the mean of the matrix } \mathbf{D} \text{ be } \mathbf{d}_{\cdot\cdot}^2 = \frac{\sum_{i=1}^p \sum_{j=1}^p \mathbf{d}_{ij}^2}{p^2} = \frac{\sum_{j=1}^p \mathbf{X}_j^2}{p} - 2\bar{\mathbf{X}}^2 + \frac{\sum_{i=1}^p \mathbf{X}_i^2}{p}.$$

Where $\bar{\mathbf{X}} = \frac{\sum_{i=1}^p \mathbf{X}_i}{p}$ is the mean of the \mathbf{X}_i .

The matrix \mathbf{D} is double-centered as follows: from each element subtract the row mean, subtract the column mean, add the matrix mean, and divide by -2 ; that is,

$$y_{ij} = \frac{(\mathbf{d}_{ij}^2 - \mathbf{d}_{\cdot j}^2 - \mathbf{d}_{i\cdot}^2 + \mathbf{d}_{\cdot\cdot}^2)}{-2} = (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})$$

This produces the p by p symmetric positive semidefinite matrix \mathbf{Y} :

$$\mathbf{Y} = \begin{bmatrix} \mathbf{X}_1 - \bar{\mathbf{X}} \\ \mathbf{X}_2 - \bar{\mathbf{X}} \\ \vdots \\ \vdots \\ \mathbf{X}_p - \bar{\mathbf{X}} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 - \bar{\mathbf{X}} & \mathbf{X}_2 - \bar{\mathbf{X}} & \dots & \mathbf{X}_p - \bar{\mathbf{X}} \end{bmatrix} \quad (7)$$

Because \mathbf{Y} is symmetric with a rank of one, its eigenvalue-eigenvector decomposition is simply:

$$Y = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{u}_p \end{bmatrix} \lambda_1 [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdot \quad \cdot \quad \cdot \quad \mathbf{u}_p] \quad (8)$$

Hence, the solution is

$$\begin{bmatrix} \mathbf{X}_1 - \bar{\mathbf{X}} \\ \mathbf{X}_2 - \bar{\mathbf{X}} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{X}_p - \bar{\mathbf{X}} \end{bmatrix} = \sqrt{\lambda_1} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{u}_p \end{bmatrix}$$

Because, without loss of generality, the origin can be placed at zero, that is, $\bar{\mathbf{X}} = \mathbf{0}$, the solution can also be written as:

$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{X}_p \end{bmatrix} = \sqrt{\lambda_1} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{u}_p \end{bmatrix} \quad (9)$$

The points from (9) *exactly reproduce* the distances in (4), the agreement scores in (2), and the original roll call votes. In addition, note that the *mirror image* of the points in (9) (a multiplication by minus one) also exactly reproduces the original roll call votes.

Furthermore, for any pair of true legislator ideal points $\tilde{\mathbf{X}}_a$ and $\tilde{\mathbf{X}}_b$ with one or more midpoints between them, $\tilde{\mathbf{X}}_a < \mathbf{Z}_j < \tilde{\mathbf{X}}_b$, the recovered legislator ideal points *must have the same ordering*, $\mathbf{X}_a < \mathbf{X}_b$. If there are no midpoints between $\tilde{\mathbf{X}}_a$ and $\tilde{\mathbf{X}}_b$ -- that is,

their roll call voting pattern is *identical* -- then the recovered legislator ideal points are identical; $\mathbf{X}_a = \mathbf{X}_b$. Hence, if there are cutting points between every pair of adjacent legislators, that is, $\mathbf{k}_i \geq 1$ for $i=1, \dots, p-1$, then the rank ordering of the recovered ideal points is the same as the true rank ordering. If some of the $\mathbf{k}_i = \mathbf{0}$, then the recovered ideal points have a weak monotone transformation of the true rank ordering (in other words there are ties, some legislators have the same recovered ideal points).

This completes the proof. **QED.**

Discussion

Note that an interval level set of points is recovered. However, *this is an artifact of the distribution of cutting points*. For example, if $\mathbf{k}_1 > \mathbf{k}_2$, this has the effect of making $\mathbf{d}_{12} > \mathbf{d}_{23}$ even if the true coordinates $\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \tilde{\mathbf{X}}_3$ were *evenly spaced*. With perfect one dimensional voting, the legislator configuration is only identified up to a *weak monotone transformation of the true rank ordering*.

The rank ordering can also be recovered from the matrix \mathbf{Y} given in (7) without performing an eigenvalue-eigenvector decomposition. Note that, with the origin at zero, the diagonal elements of \mathbf{Y} are simply the legislator coordinates squared. The rank ordering can be recovered by taking the square root of the first diagonal element and then dividing through the first row of the matrix. Note that this sets $\mathbf{X}_1 > \mathbf{0}$ and the remaining points are identified vis a vis \mathbf{X}_1 .