Political Analysis, 8:2

Nonparametric Unfolding of Binary Choice Data

Keith T. Poole
Carnegie-Mellon University

This paper shows a general nonparametric unfolding technique for maximizing the correct classification of binary choice or two-category data. The motivation for and the primary focus of the unfolding technique are parliamentary roll call voting data. However, the procedures that implement the unfolding also can be applied to the problem of unfolding rank order data as well as analyzing a data set that would normally be the subject of a probit, logit, or linear probability analysis. One aspect of the scaling method greatly improves Manski's "maximum score estimator" technique for estimating limited dependent variable models. To unfold binary choice data two subproblems must be solved. First, given a set of chooser or legislator points, a cutting plane must be found such that it divides the legislators/choosers into two sets that reproduce the actual choices as closely as possible. Second, given a set of cutting planes for the binary choices, a point for each chooser or legislator must be found which reproduces the actual choices as closely as possible. Solutions for these two problems are shown in this paper. Monte Carlo tests of the procedure show it to be highly accurate in the presence of voting error and missing data.

1 Introduction

THE PURPOSE OF this paper is to show a new scaling method for analyzing parliamentary roll call data. The scaling method employs the same spatial model used by Poole and Rosenthal (1997) in their NOMINATE procedure and the scaling method is "NOMINATE-like" in structure. However, rather than maximizing the likelihood of the legislators' choices, the scaling method developed below *maximizes correct classification* of the legislators' choices. The scaling method is *nonparametric* because no assumptions are made about the probability distribution of the legislators' errors in making choices. The only assumptions made are that the choice space is Euclidean and that individuals making choices behave as if they utilize symmetric, single-peaked preferences. Strictly speaking, the scaling method presented here is *not* a statistical model. However, standard errors can be estimated via bootstrapping for the legislator coordinates.

The remainder of this introduction is devoted to the geometric intuition behind the scaling procedure and where the scaling method fits in the existing literature in psychometrics. With

Author's note: I would like to thank Howard Rosenthal, Nolan McCarty, Larry Rothenberg, Tim Groseclose, Tom Romer, and three anonymous referees for their comments on various drafts of this paper. The programs that perform the analyses shown in the paper and detailed examples are available at *http://k7moa.gsia.cmu.edu/dwnl.htm* and the Appendix to the paper can be found at the *Political Analysis* website. This paper is dedicated to the memory of Jerry Salancik, who gave me invaluable comments just prior to his untimely death.

Copyright 1999 by the Society for Political Methodology

respect to econometrics, one aspect of the scaling method—the *cutting plane procedure*—greatly improves Manski's "maximum score estimator" technique (Manski 1975, 1985; Manski and Thompson 1986) for estimating limited dependent variable (LDV) models. This is discussed in Section 3. Readers interested in only this aspect of the paper should read the notational definitions below and then skip to Sections 2 and 3.

The spatial model estimated by the scaling method assumes that legislators have Euclidean preferences defined over some ideological/policy space and that they vote sincerely for the alternative closest to their ideal point. Each roll call vote has two policy points in the space corresponding to the policy consequences of a Yes (Yea) or No (Nay) vote on the roll call.[1] In the perfect case a plane can be placed in the space such that it separates the legislators voting Yea from the legislators voting Nay. Geometrically, this *cutting plane* both is perpendicular to the line joining the Yea and Nay policy points and passes through the midpoint of the Yea and Nay policy points. The *normal vector* to this cutting plane is parallel to the line joining the Yea and Nay policy points.

The scaling method finds estimates of the legislator ideal points and cutting planes for the roll calls that maximize the number of votes classified correctly. The program has two phases. Given the legislator ideal points, for each roll call a cutting plane is found that maximizes correct classification of legislators' votes on that roll call. The *cutting plane procedure* developed in Section 2 shows how this is done. In the second phase, given the cutting planes for the roll calls, for each legislator a point is found in the space vis a vis the cutting planes that maximizes the correct classification of the legislator's votes across the roll calls. The *legislative procedure* developed in Section 4 shows how this is done. The output of the cutting plane procedure is a new set of cutting planes that are passed to the legislative procedure to obtain a new set of legislator ideal points, and so on. The correct classification *can never decrease* going from one phase to the next. These two phases can be repeated in sequence until no further improvement in correct classification occurs.

In one dimension the scaling problem consists of finding a joint rank ordering of the legislators and roll call midpoints (ties are permitted; an example is shown in Table 6) that maximizes correct classification. Given a rank order of legislators, the global maximum in correct classification can be found for every roll call. Similarly, given a rank order of the roll call midpoints, the global maximum in correct classification can be found for every legislator. The two are symmetric in one dimension. The one-dimensional scaling method resembles classical Guttman scaling (see Fig. 2). However, the assumptions underlying the two are fundamentally different. Guttman scaling is based upon a *cumulative* response function, that is, a utility function that is always monotonically increasing over the relevant dimension or space. In contrast, the spatial model employed here is based upon a *proximity* response function (legislators vote for the option closest to their ideal points—symmetric utility). Weisberg (1968) shows that these two models correspond to two very different substantive processes underlying roll call voting. The proximity response function is clearly more realistic in a legislative context.[2]

In two or more dimensions this symmetry disappears—points must be estimated for the legislators and cutting planes must be estimated for roll calls. For example, in two dimensions, suppose that there are q roll calls so that there are q cutting lines. If a va-

[1] This model was first proposed by MacRae (1958) and later developed by Poole and Rosenthal (1997) in their NOMINATE procedure.

[2] See van Schuur (1992) for a discussion of some Guttman-like models. The multidimensional generalization of Guttman scaling is known as Multidimensional Scalogram Analysis (Lingoes, 1963). For a survey see Shye (1978, Chaps. 9–11).

riety of voting coalitions forms among the legislators, then these cutting lines will criss-cross one another in a myriad of directions, creating a very large number of regions in the plane (see Fig. 4). Indeed, Coombs (1964, p. 262) shows that q binary choices (roll calls) create a maximum of $\sum_{k=0}^{s} \binom{\mathbf{q}}{\mathbf{k}}$ regions where s is the number of dimensions. For two dimensions, $1 + q + q(q-1)/2$ regions are possible, with each region corresponding to a voting pattern on the q roll calls, e.g., yynnynynnny.... There are 2^q possible voting patterns and in practical applications this number will greatly exceed the maximum number of regions in the space. Hence, the problem is to find the region that best matches the legislator's observed pattern of roll call votes. This is what the legislative procedure does.

Let s denote the number of policy dimensions, which are indexed by $k = 1, \ldots, s$; let p denote the number of legislators ($i = 1, \ldots, p$) and q denote the number of roll call votes ($j = 1, \ldots, q$). Let $\mathbf{T}$ be the $p \times q$ matrix of observed choices. The choices are simply Yea or Nay and are represented as "y" or "n," respectively. $\mathbf{T}$ can contain missing entries. Let $\mathbf{X}$ be the $p \times s$ matrix of legislator coordinates. The roll call cutting planes are defined in terms of their *normal vectors*. Specifically, let $\mathbf{n_j}$ be the $s \times 1$ normal vector for the jth roll call and let $\mathbf{N}$ be the $q \times s$ matrix of normal vectors for the q cutting planes. A plane is defined as the vector equation, $\mathbf{z'n_j} = \mathbf{v'n_j}$, where $\mathbf{z}$, $\mathbf{n_j}$, and $\mathbf{v}$ are $s \times 1$ vectors and the plane consists of all points $\mathbf{z}$ such that $(\mathbf{z} - \mathbf{v})$ is perpendicular to the normal vector, $\mathbf{n_j}$, and $\mathbf{v}$ is a *specific* point in the plane.[3] Note that if $\mathbf{v_1}$ and $\mathbf{v_2}$ are both points in the plane, then $\mathbf{v'_1 n_j} = \mathbf{v'_2 n_j} = m_j$, where m_j is a scalar constant. Geometrically, every point in the plane projects onto the same point on the line defined by the normal vector. By definition, the normal vector for a roll call cutting plane is parallel to the line joining the Yea and Nay policy points and passes through the midpoint of the Yea and Nay policy points. Because the midpoint of the Yea and Nay policy points is on the cutting plane, it too projects to the point m_j on the line defined by the normal vector, $\mathbf{n_j}$. Let $\mathbf{z_{jy}}$ and $\mathbf{z_{jn}}$ be the Yea and Nay policy points, and let $\mathbf{z_{mj}} = (\mathbf{z_{jy}} + \mathbf{z_{jn}})/2$ be the midpoint. Hence, $\mathbf{z'_{mj} n_j} = m_j$.

Given this geometry, the rule for a correct classification is the following. Let $\mathbf{x_i}$ be the ith legislator's ideal point, let c denote the legislator's chosen outcome, and let the projection of the ith legislator's ideal point onto the projection line be $w_i = \mathbf{x'_i n_j}$. Then the rule for a correct classification is as follows.

If legislator i votes c:

$\delta_{ij} = 1$ if $w_i \geq m_j$ and $\mathbf{z'_{jc} n_j} > m_j$, or $w_i < m_j$ and $\mathbf{z'_{jc} n_j} < m_j$;
$\delta_{ij} = 0$ if $w_i < m_j$ and $\mathbf{z'_{jc} n_j} > m_j$, or $w_i > m_j$ and $\mathbf{z'_{jc} n_j} < m_j$.

In other words, if the legislator votes "Yea"/"Nay" and her ideal point is on the Yea/Nay side of the plane, the legislator's vote is correctly classified. Note that the assumption of *symmetric* single-peaked preferences means that if a legislator votes "Yea" and her ideal point is *anywhere on the Yea side of the plane*, then that counts as a correct classification. If preferences are *not symmetric*, then this might not be true.

[3] In many vector calculus textbooks this is taught in terms of the general expression (in three dimensions) $A(x - x_0) + B(y - y_0) + C(z - z_0) = 0$, where A, B, and C are the components of the normal vector and (x_0, y_0, z_0) is a point in the plane. If the normal vector is $(3, -2, 1)$ and the point in the plane is $(1, 0, 2)$, then this produces the equation $3x - 2y + z = 5$. See Salas and Hille (1974, Chap. 12) and Lang (1979, Chap. 1) for standard treatments of the topic.

The total correct classification is therefore

$$\delta(\mathbf{X}, \mathbf{N}) = \sum_{i=1}^{p} \sum_{j=1}^{q} \delta_{ij} \tag{1}$$

In sum, given the number of dimensions, s, the classification problem consists of finding estimates of $\mathbf{X}$ and $\mathbf{N}$, denoted $\mathbf{X}^*$ and $\mathbf{N}^*$, respectively, which maximize Eq. (1).

The scaling method developed below finds $\mathbf{X}^*$ and $\mathbf{N}^*$ that maximize Eq. (1). It is *not* a statistical model because no error process is incorporated into Eq. (1). However, standard errors for the legislator ideal points and the roll call normal vectors can be estimated via bootstrapping.

Within the psychometrics field, the scaling method developed here is a *nonmetric unfolding* procedure. It is an "unfolding" because the roll calls are treated as preferential choice data and parameters for individuals (legislators) and stimuli (roll calls) are being estimated. It is "nonmetric" because no assumptions are made about the parametric form of the individuals' "true" preference functions other than that they are symmetric and single peaked.

Unfolding was developed for the one-dimensional case by Coombs (1950) and generalized to the multidimensional case by Bennett and Hays (1960). The original unfolding model—later dubbed the *ideal-point model*—represented individuals and stimuli as points and was originally developed to analyze rank orderings of stimuli by individuals. Later, Tucker (1960) developed the *vector model* of unfolding in which the individuals are treated as vectors and the stimuli as points. The vector model is a special case of the unfolding model where the individual's ideal point goes off to infinity (Carroll 1980; Borg and Groenen 1997, Chap. 15). This model is much like Guttman scaling in that the individual utility rises/falls monotonically from the center of the space off to infinity along the individual's vector. The projections of the stimuli onto the individual's vector reproduce the observed rank ordering. The vector model is the basis of the MDPREF program developed by Chang and Carroll (1969).

With respect to the roll call voting problem, the ideal point and vector unfolding models are closely related. If the individuals are treated as roll calls and the roll calls are treated as individuals, then the individual becomes a cutting plane through the space and the point where the cutting plane passes through the normal (individual) vector is the *individual's threshold*. That is, the individual approves/accepts the stimuli on one side of the plane and disapproves/does not accept the stimuli on the other side of the plane. "Pick Any/N" data that are widely used in marketing applications (DeSarbo and Cho 1989) have this form. For example, respondents are given a list of soda pops and asked if they drink/not drink each soda. The soda pops are then displayed as points in a space and the individuals as cutting lines that divide the soda pops into drink/not drink.

Psychometricians have largely abandoned the nonmetric approach in the past fifteen years "because they suspected instability and identification problems" and have focused their efforts on probabilistic and metric alternatives.[4] The most important recent work along these lines is by Londregan (2000). Londregan links the psychometrics testing literature with the spatial theory of legislative voting and derives important statistical results about the

[4]Personal communication to the author from Willem J. Heiser, 27 March 1998. Some recent examples of probabilistic/metric models within the psychometrics tradition are Heiser (1981), DeSarbo and Hoffman (1987), Gifi (1990), Blokland-Vogelesang (1991), Hojo (1994), and Andrich (1995).

parameters of the spatial model. In particular, when the preferential choices are *nominal*, Londregan shows that consistency in its usual statistical sense does not hold. With nominal choices standard maximum-likelihood estimators that attempt to recover *simultaneously* legislators' ideal points and roll call parameters inherit the "granularity" of the choice data and so cannot recapture the underlying *continuous* parameter space. If legislators could report continuous "feeling thermometer" scores instead of just Yea/Nay, then this source of inconsistency would disappear. In addition, note that *as error goes to zero*,[5] legislators are only identified up to *regions* in the space. Londregan's work in part inspired the development of the scaling procedure shown in this paper.

The algorithms that are at the center of the cutting plane procedure and the legislative procedure are unique and are very stable. In particular, when the number of legislators is 100 or greater and the number of roll calls is on the order of 500—typical of national legislatures, for example, the U.S. Senate—then the recovery of the legislators and cutting lines is very precise. With 500 roll calls, there are a maximum 125,251 regions in two dimensions and a maximum of 20,833,751 in three dimensions. Most of these regions are so small that a typical legislator's point is very precisely pinned down. In fact the recovery of the legislator coordinates is virtually identical to those recovered by parametric procedures that must make strong assumptions about the interpersonal comparability of individuals' utility and the functional form of the error distribution (e.g., Heckman and Snyder 1997; Poole and Rosenthal 1997).

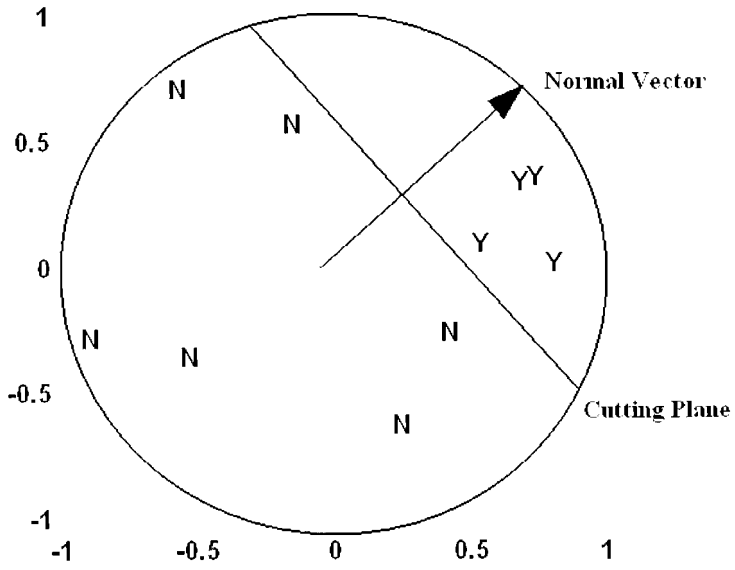
Sections 2–6 develop the nonparametric unfolding procedure. Section 2 develops the *cutting plane procedure*—a method for finding the optimal cutting plane given a configuration of legislators. Section 3 shows the relationship between the cutting plane procedure and Manski's "maximum score estimator" technique for estimating limited dependent variable (LDV) models. Section 4 develops the *legislative procedure* that finds the optimal legislator point given a set of cutting planes. Section 5 shows Monte Carlo tests of the unfolding procedure and empirical applications are shown in Section 6. The Appendix contains additional examples and Monte Carlo tests. It can be found at the *Political Analysis* website.

2 Finding the Optimal Cutting Plane

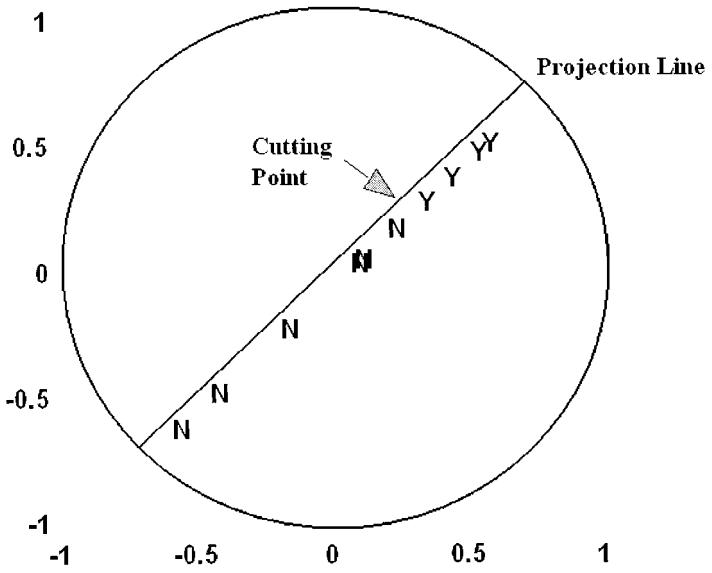
Given the $p \times s$ matrix, $\mathbf{X}$, of legislator coordinates and the $p \times 1$ vector of votes on the jth roll call, $\mathbf{t_j}$, the problem is to find the plane that divides the legislators into two groups such that the number of correct classifications is maximized. Figure 1 shows an example in two dimensions.

Figure 1 illustrates the fact that the cutting plane problem is equivalent to finding a vector—in this case, $\mathbf{n}$—such that when the legislator points are projected onto the vector, a *cutting point* can be found that maximizes the correct classifications. By definition, all points in the cutting plane are projected onto this cutting point. The problem has two distinct parts. First, given an estimated normal vector, the plane perpendicular to the normal vector that maximizes correct classification must be found; and second, given an estimated cutting plane, the orientation of the plane in the space must be changed so that a better estimate of the normal vector is found.

[5] This is not as far-fetched as it sounds. Several European parliaments classify at 95% or above in one or two dimensions. For example, most legislative sessions during the French 4th Republic (personal communication from Howard Rosenthal and Erik Voeten), recent sessions of the Czech parliament (personal communication from Abdul Noury), and the 1841 English parliament (personal communication from Cheryl Schonhardt-Bailey).



(A)



(B)

Fig. 1 (A) Original positions in two dimensions. (B) Points projected onto line.

2.1 *Calculating the Correct Classifications*

Let the legislator coordinates lie within the s-dimensional unit hypersphere and let the origin of the space be placed at the centroid of the legislator coordinates; that is, let

$$\sum_{k=1}^{s} x_{ik}^2 \leq 1, \qquad i = 1, \ldots, p, \qquad \text{and} \qquad \sum_{i=1}^{p} x_{ik} = 0, \qquad k = 1, \ldots, s$$

Actual Voting Pattern

Y Y Y Y Y Y .... Y Y * N Y * N Y * N N ... N N

-1.0 w_1 w_2 w_3 w_4 ..... 0.0 ..... w_{p-1} w_p +1.0

Perfect Voting Patterns

(-1, w_1) produces nnnnnnnnn.....nn	or	yyyyyyyyy.....yy
(w_1, w_2) produces ynnnnnnnn.....nn	or	nyyyyyyyy.....yy
(w_2, w_3) produces yynnnnnnn.....nn	or	nnyyyyyyy.....yy
(w_3, w_4) produces yyynnnnnn.....nn	or	nnnyyyyyy.....yy
etc.		
(w_{p-1}, w_p) produces yyyyyyyyy.....yn	or	nnnnnnnnn.....ny
(w_p, +1) produces yyyyyyyyy.....yy	or	nnnnnnnnn.....nn

Fig. 2 Calculating the correct classification.

In addition, let $\mathbf{n_j}$ be the normal vector for the jth roll call that maximizes correct classifications. Without loss of generality, $\mathbf{n_j}$ can be constrained to be of unit length; i.e., $\mathbf{n_j'n_j} = 1$. The projections (see Fig. 1B) are, therefore,

$$\mathbf{Xn_j} = \mathbf{w_j} \tag{2}$$

Note that the elements in the p-length vector, $\mathbf{w_j}$, range from -1 to $+1$. The elements in $\mathbf{w_j}$ all lie on a line that passes through the origin of the s-dimensional unit hypersphere in the direction of the normal vector with exit points $-\mathbf{n_j}$ and $+\mathbf{n_j}$, respectively. Hereafter, this is referred to as the *projection line*.

Let $\mathbf{n_j^*}$ be an estimate of $\mathbf{n_j}$ and let $\mathbf{w_j^*}$ be the corresponding estimate of $\mathbf{w_j}$. The correct classifications associated with $\mathbf{n_j^*}$ can be calculated quite easily. Figure 2 illustrates the method.

In Fig. 2 the index j has been omitted to reduce clutter. In addition, let the projected legislator coordinates from left to right be denoted in order as w_1 to w_p such that $-1 \leq w_1 \leq w_2 \leq w_3 \leq \cdots \leq w_p \leq +1$ and the Y's and N's above the projection line in the figure indicate how the corresponding legislators voted on the jth roll call. There are $p+1$ possible regions that the cutting point could be in—$(-1, w_1)$, (w_1, w_2), ..., $(w_p, +1)$—and for each region there are exactly two possible perfect voting patterns, for an overall total of $2(p+1)$ possible perfect voting patterns. However, region $(w_p, +1)$ is redundant since it produces the same perfect patterns as the region $(-1, w_1)$ so it may be discarded, leaving $2p$ unique perfect voting patterns to consider.

Since there are only $2p$ perfect patterns, it is a simple matter to compare each perfect pattern with the actual pattern of votes, $\mathbf{t_j}$. This can be done very efficiently by first assuming that the cutting point is in the region $(-1, w_1)$ and calculating the corresponding number of correct classifications. Next assume that the cutting point is in the region (w_1, w_2). Only one calculation has to be made to get the correct classifications for this cutting point since the only change is that the cutting point has been moved from the left of w_2 to the right of w_2. If there is no missing data, the correct classification either increases by 1 or decreases by 1 when the cutting point is moved from the left of w_2 to the right of w_2. Similar reasoning holds for the remaining points. For each possible cutting point the correct classification corresponding to the two possible perfect patterns can be calculated. The estimated cutting

point is set equal to the midpoint of the region for which correct classification is a maximum. For the example shown in Fig. 2, placing the cutting point at the position of any of the three asterisks would produce only two classification errors, for a correct classification of $p - 2$.[6]

Note that this process is equivalent to moving the cutting plane through the unit hypersphere along the line defined by the estimated normal vector, $\mathbf{n_j^*}$, and its reflection, $-\mathbf{n_j^*}$.

2.2 *Calculating the Optimal Normal Vector*

Let m_j^* denote the cutting point that maximizes correct classification on the projection line formed by the elements of $\mathbf{Xn_j^*} = \mathbf{w_j^*}$. The estimated cutting plane consists of all points $\mathbf{v}$ such that

$$\mathbf{v'n_j^*} = m_j^* \tag{3}$$

In order to get a new estimate of $\mathbf{n_j}$, the estimated cutting plane given by Eq. (3) must be moved through the space in a direction that increases correct classification. This is accomplished by moving the cutting plane toward the legislator points that are classification errors.

To do this, a matrix is created by projecting all the *correctly classified* legislator points onto the surface of the cutting plane while leaving the incorrectly classified legislators at their original positions. In two dimensions this produces a line through the space made up of correctly classified legislators (the current cutting plane) around which is a scattering of points corresponding to the incorrectly classified legislators (see Fig. 3). Specifically, let $\mathbf{x_i}$ be the $s \times 1$ vector denoting the ith legislator's point in the space and let w_{ij} be the corresponding point on the projection line from Eq. (2). Construct a $p \times s$ matrix, $\mathbf{\Psi}$, as follows: if legislator i is correctly classified, then her point is projected onto the cutting plane and that point becomes the ith row of $\mathbf{\Psi}$; if legislator i is incorrectly classified, then her point remains at its original position and that point becomes the ith row of $\mathbf{\Psi}$. That is, on the jth roll call,

$$\begin{aligned} \boldsymbol{\psi}_\mathbf{i} &= \mathbf{x_i} + (m_j^* - w_i)\mathbf{n_j^*} && \text{if correctly classified} \\ \boldsymbol{\psi}_\mathbf{i} &= \mathbf{x_i} && \text{if incorrectly classified} \end{aligned} \tag{4}$$

Without loss of generality, the centroid of $\mathbf{\Psi}$ can be placed at the origin. That is, let $\boldsymbol{\mu}$ be the s-length vector of the means of the columns of $\mathbf{\Psi}$, and let $\mathbf{J_p}$ be a $p \times 1$ vector of ones. Define $\mathbf{\Psi}^*$ as

$$\mathbf{\Psi}^* = \mathbf{\Psi} - \mathbf{J_p}\boldsymbol{\mu}' \tag{5}$$

Figure 3A shows a vote in two dimensions that would be perfectly classified by the indicated cutting line. Figure 3B shows the $\mathbf{\Psi}^*$ produced by using an initial estimate of $\mathbf{n_j^{*\prime}} = (0, 1)$—that is, an estimated normal vector perpendicular to the true normal vector. All the "y" and "n" tokens off the plane are classification errors. Clearly, if the plane were moved counterclockwise toward the errors, a better fit would be obtained.

[6] When there are multiple solutions like the example shown in Fig. 2, the cutting point is placed at the midpoint of the region closest to the center of the space.

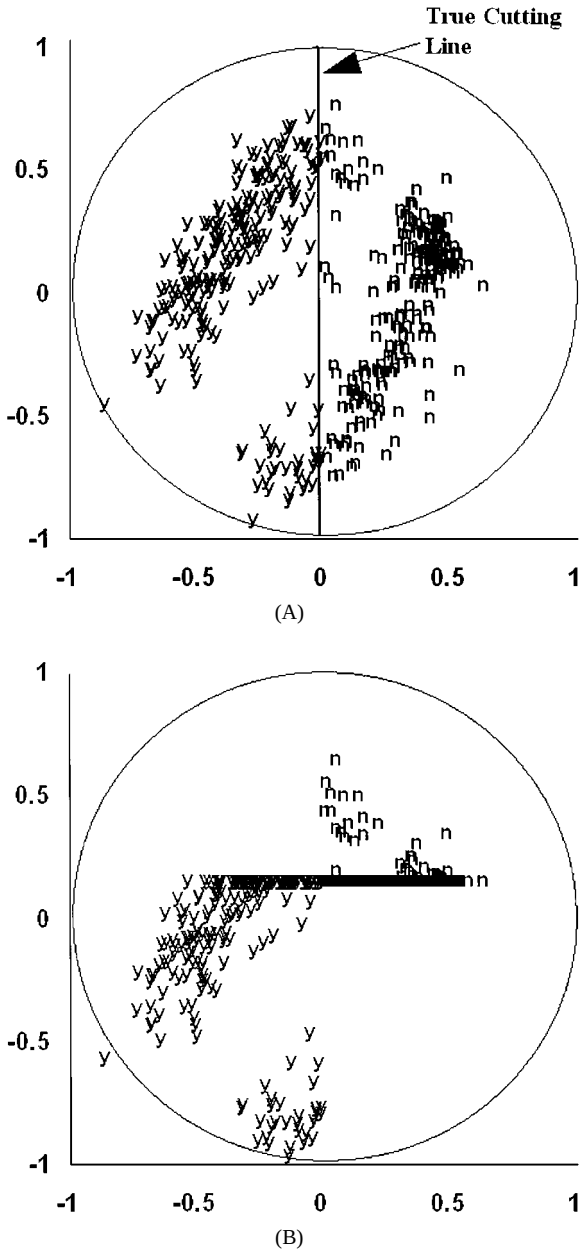
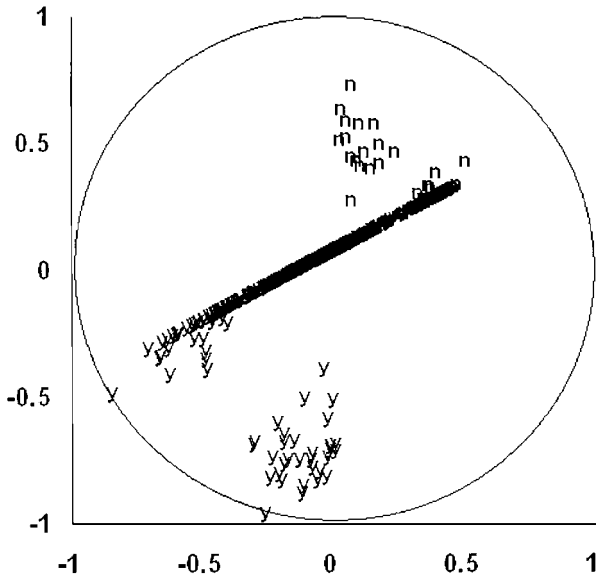


Fig. 3 (A) True cutting line. (B) Initial estimate of cutting plane.

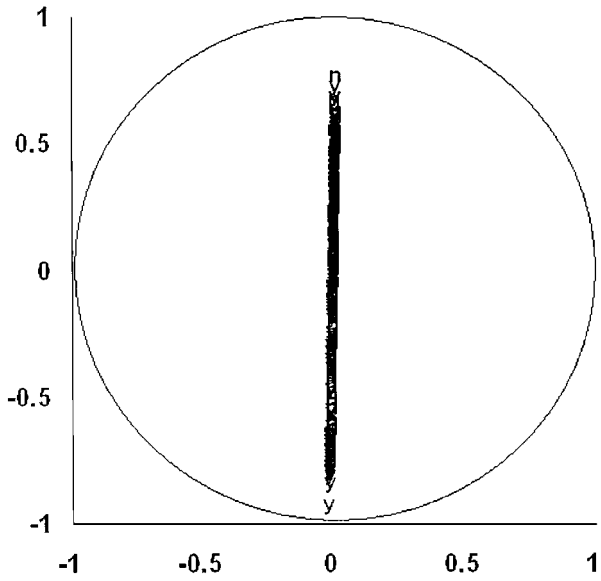
This is accomplished by using the Eckart and Young (1936) lower-rank-matrix approximation theorem. Let the singular value decomposition of $\mathbf{\Psi}^*$ be

$$\mathbf{\Psi}^* = \mathbf{U}\mathbf{\Lambda}\mathbf{\Theta}' \tag{6}$$

where $\mathbf{U}$ is a $p \times s$ orthogonal matrix, $\mathbf{\Theta}$ is an $s \times s$ orthogonal matrix, and $\mathbf{\Lambda}$ is an $s \times s$ diagonal matrix containing the singular values in descending order on the diagonal. By definition, $\mathbf{U}'\mathbf{U} = \mathbf{\Theta}'\mathbf{\Theta} = \mathbf{I}_s$, where $\mathbf{I}_s$ is an $s \times s$ identity matrix.



(C)



(D)

Fig. 3 *(Continued)* (C) Second estimate of cutting plane. (D) Thirty-fifth estimate of cutting plane.

By the Eckart–Young theorem, the best-fitting line through the scatterplot shown in Fig. 3B is found by inserting a zero in place of the second singular value in $\mathbf{\Lambda}$ and remultiplying. That is, let $\mathbf{\Lambda}^{\#}$ be the $s \times s$ diagonal matrix identical to $\mathbf{\Lambda}$ except for the replacement of the sth singular value by zero; then the estimated hyperplane is

$$\mathbf{V} = \mathbf{U}\mathbf{\Lambda}^{\#}\mathbf{\Theta}' \tag{7}$$

where the $p \times s$ matrix $\mathbf{V}$ will have rank $s - 1$ by construction.

Let $\mathbf{n}_j^{\#}$ be the normal vector of the hyperplane defined by $\mathbf{V}$, let θ_s be the sth singular vector of $\mathbf{\Theta}$, and let $\mathbf{0_p}$ be a p-length vector of zeroes. It is straightforward to show that $\mathbf{n}_j^{\#} = \theta_s$. This stems from the simple fact that, by construction,

$$\mathbf{V}\theta_s = \mathbf{0_p}$$

which is identical to Eq. (3) (see Appendix A1 for a simple demonstration).

In sum, calculating the optimal $\mathbf{n_j}$ consists of the following steps:

1. Obtain a starting estimate of $\mathbf{n}_j^*$ using simple OLS (linear probability) or two-group linear discriminant analysis.
2. Calculate the correct classifications associated with $\mathbf{n}_j^*$.
3. Construct $\mathbf{\Psi}^*$ using Eq. (4) and Eq. (5).
4. Perform singular value decomposition of $\mathbf{\Psi}^*$, $\mathbf{U\Lambda\Theta'}$.
5. Use the sth singular vector of $\mathbf{\Theta}$, θ_s, as the new estimate of $\mathbf{n_j}$.
6. Go to step 2.

In an error-free case like that shown in Fig. 3, this cutting plane procedure will almost always quickly iterate into the true cutting plane. With error-ridden data, the rate of convergence is a function of the number of errors. As the number of errors decreases, the mass of the correctly classified choices increases, thereby producing very small changes in the newly estimated normal vectors. The procedure is stopped when the sum of squared differences in $\mathbf{n}_j^*$ divided by s changes by less than .0001 between iterations.

Table 1 shows a Monte Carlo study of the cutting plane procedure using perfect data for 100 legislators and 500 roll calls for 2 through 10 dimensions. (A Monte Carlo study of contaminated data in one to three dimensions with varying error levels and types of error is shown in Appendix A2.) Results for one dimension are not shown in Table 1 since correct classification will always be 100% if error-free data are used. The 100 legislators and 500 pairs of policy points were randomly drawn from a uniform distribution through the unit hypersphere. The policy points were randomly drawn but in such a way so as to produce an average majority margin of about 67% (typical of U.S. congressional roll call data; see Table 6). A maximum of 50 iterations through steps 2–5 above was allowed.

The cutting plane procedure performs very well. The number of dimensions does not appear to play any role in the accuracy of the procedure. For example, for the 10 trials in 10 dimensions, the 5000 total estimated $\mathbf{n}_j^*$'s correctly classified 499,936 of 500,000 choices (99.99%). With 100 legislators the recovery of the true normal vectors is quite good. In two dimensions, the average of the cosines computed between the true normal vectors and the estimated normal vectors is .998 for all roll calls and .999 for roll calls with at least 10% in the minority (90–10 or better). This average cosine falls as the number of dimensions is increased because of the simple fact that there is more space between the legislators so that the cutting planes can be moved slightly without affecting the classification. Note that, in four or fewer dimensions—the dimensionality of most practical applications—the recovery is almost exact.

When error is present the cutting plane procedure is highly accurate and converges very quickly. Table A1 in the Appendix reports the results of the Monte Carlo study in the presence of error.

Table 1 Monte-Carlo tests of cutting plane procedure 100 legislators and 500 votes (perfect data) (each entry average of 10 trials, standard deviations in parentheses)

S	*Average majority margin*	*Average number of errors*	*Average percent correctly classified*	*Average fit with true normal vectors all*[a]	*Average fit with true normal vectors 10% min.*[b]
2	65.6	4.1	99.99	.998	.999
	(0.7)	(3.4)	(0.01)	(.001)	(.000)
3	66.4	4.9	99.99	.996	.997
	(0.5)	(2.6)	(0.01)	(.000)	(.000)
4	66.7	4.8	99.99	.992	.995
	(1.0)	(1.9)	(0.00)	(.001)	(.001)
5	67.3	6.5	99.99	.989	.993
	(0.7)	(2.3)	(0.00)	(.001)	(.000)
6	66.9	7.2	99.99	.984	.990
	(0.7)	(1.7)	(0.00)	(.001)	(.000)
7	67.4	5.6	99.99	.979	.987
	(0.7)	(2.6)	(0.01)	(.002)	(.001)
8	67.4	5.5	99.99	.975	.984
	(0.7)	(2.4)	(0.01)	(.001)	(.001)
9	67.6	6.5	99.99	.970	.980
	(0.5)	(2.3)	(0.00)	(.002)	(.001)
10	67.5	6.4	99.99	.964	.976
	(0.8)	(3.0)	(0.01)	(.001)	(.001)

[a]Average cosine computed between true normal vectors and estimated normal vectors.
[b]Average cosine computed between true normal vectors and estimated normal vectors for those roll calls with at least 10 percent or better in the minority.

3 The Relationship of the Cutting Plane Procedure to Manski's Maximum Score Estimator for Estimating LDV Models

Given a simple two-category dependent variable and a set of fixed independent variables, the cutting plane procedure can be used to estimate a vector of coefficients for the independent variables that maximizes correct classification of the dependent variable. In this instance, with the independent variables scaled so as to lie within a unit hypersphere, the normal vector, $\mathbf{n_j^*}$, produced by the cutting plane procedure, plays the role of the coefficient vector, $\boldsymbol{\beta}$, in a standard Probit, Logit, or linear probability analysis. For example, in a Probit analysis, if the estimated β's for the independent variables, $\beta_1, \beta_2, \ldots, \beta_s$, are normalized so that their sum of squares is equal to one, then they constitute a normal vector to a plane upon which the choice probabilities are exactly .5/.5. That is, in terms of Fig. 1, all the points on the cutting plane have choice probabilities of .5/.5 and are projected onto the projection line at the cutting point, m_j^*, which is determined by the intercept term, β_0. In this context, the normal vector is the direction of maximum increase/decrease in probability and the $1/\sigma$ in the standard Probit expression determines how rapidly the probability rises/falls along the normal vector from the cutting plane to the rim of the hypersphere, that is, from m_j^* to $-\mathbf{n_j}$ and $+\mathbf{n_j}$.

The cutting plane procedure is a very efficient method for implementing Manski's maximum score estimator (MSE) for limited dependent variable models (Manski 1975, 1985; Manski and Thompson 1986). In its simplest form the MSE chooses $\boldsymbol{\beta}$ to maximize correct

classification. Part of the MSE process is very similar to the part of the cutting plane procedure illustrated by Fig. 2. Namely, given a direction, it is easy to find the classification maximum (Manski and Thompson 1986, pp. 89–90). However, the MSE approach has no method other than exhaustive search to find a better direction through the space.

Intuitively, in terms of the notation developed above, the MSE algorithm consists of two phases. First, given $\mathbf{n_j^*}$, the cutpoint m_j^* is found. Second, let $\mathbf{n_j^1}, \mathbf{n_j^2}, \mathbf{n_j^3}, \ldots, \mathbf{n_j^{s-1}}$, be a set of normal vectors orthogonal to $\mathbf{n_j^*}$. The algorithm then searches along these orthogonal vectors for a solution better than the current solution. Unlike the cutting plane procedure, which uses Eq. (3) and Eq. (6) to arrive at a better solution for the normal vector, the MSE algorithm has no systematic criterion for selecting a better search direction (Greene 1993, pp. 658–659).

Greene (p. 659) compared the MSE algorithm with Probit using a data set gathered by Spector and Mazzeo (1980) that studied a new method of teaching economics. Greene found that Probit correctly classified 26 of the 32 observations, while the MSE algorithm classified only 22 of the 32 observations. The cutting plane procedure recovers essentially the same coefficient vector as that shown by Greene (after normalization) for Probit and correctly classifies 28 of 32 observations. This result, along with an additional empirical example and a Monte Carlo study of the cutting plane procedure with classification error, is shown in Appendix A2.

4 Finding the Optimal Legislator Coordinates

Given the $q \times s$ matrix, $\mathbf{N}$, of normal vectors and the $q \times 1$ vector of votes of the ith legislator, $\mathbf{t_i}$, the problem is to find the legislator point, $\mathbf{x_i}$, which maximizes the correct classification. Figure 4 shows an example in two dimensions.

Figure 4 shows five cutting lines indicated by the numbering at the rim of the circle. The "Y" or "N" on either side of each cutting line indicates how a legislator on that side of the

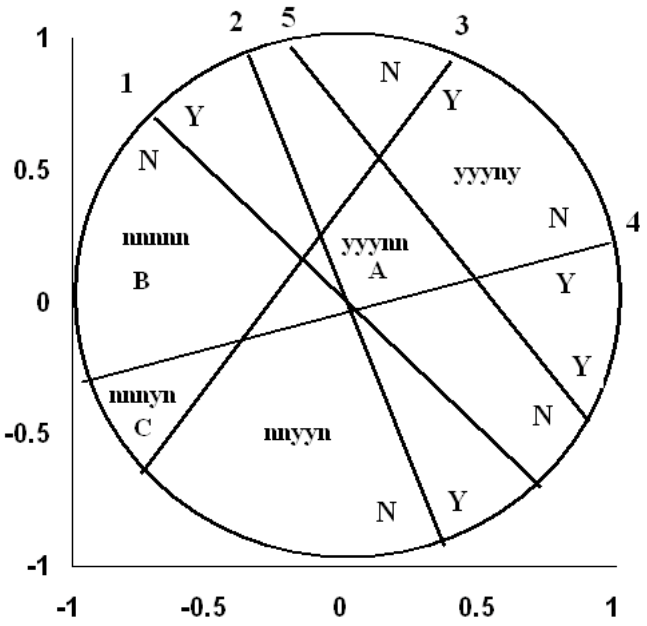


Fig. 4

cutting line should vote—"yea" or "nay," respectively. The maximum number of regions created by five cutting lines in two dimensions is 16 and each of these 16 regions can be characterized by a unique vector of votes. Figure 4 shows only 13 regions to emphasize the practical issue that in real-world data not all of the theoretically possible regions will be present. For example, cutting lines 2 and 5 intersect outside the circle so that a legislator who voted "y" on both cannot be placed inside the circle without a classification error. In practice, the restriction that the legislator points lay within a unit hypersphere does not pose a problem since the legislator points and cutting planes are iteratively adjusted to maximize correct classification. In this case, if the data were perfect (error-free) and two dimensional, the procedure would move the legislator points in such a way that lines 2 and 5 would intersect inside the circle.

Given a legislator's pattern of votes, in this case NNNYN (technically, $\mathbf{t}_\mathbf{i}' = [\text{NNNYN}]$), the problem is to find the region in Fig. 4 that maximizes the correct classification. In this example point C is located in the region corresponding to perfect classification. Suppose the initial estimate of the legislator's coordinates is at the origin, point A in Fig. 4. This initial estimate is very poor, as it correctly classifies only one of the five votes. The problem is to move the point representing the legislator in a direction that increases the number of correct classifications.

Below a method is shown for finding the maximum classification point along any arbitrary line passing through the space. This method is used to move the legislator point through the space in a city-block fashion by searching along a line parallel to the first dimension and then solving for the point along this line that maximizes classification. Then the legislator point is moved along a line through this new point but parallel to the second dimension. This is done for each dimension in turn and can be repeated as many times as desired. This always converges to a point for which the coordinates are at a *local maximum* in terms of classification. That is, the point cannot be moved parallel to any dimension and have the correct classifications increase.

Let $\mathbf{x}_\mathbf{i}^{(h)}$ be the initial estimate for legislator i where h is the iteration number (1, 2, 3, etc.) and let $\mathbf{x}_\mathbf{i}^{(a)}$ be a second point. The problem is to find a new estimate, $\mathbf{x}_\mathbf{i}^{(h+1)}$, on the line passing through $\mathbf{x}_\mathbf{i}^{(h)}$ and $\mathbf{x}_\mathbf{i}^{(a)}$ which increases correct classification. Using Eq. (2), the projection of $\mathbf{x}_\mathbf{i}^{(h)}$ onto the jth normal vector is

$$\mathbf{x}_\mathbf{i}^{(h)\prime}\mathbf{n}_\mathbf{j} = w_{ij}^{(h)} \tag{8}$$

Similarly, the projection of the second point onto the jth normal vector is $w_{ij}^{(a)}$. These projections correspond to a correct classification on roll call j depending upon on which side of the cutpoint, m_j, they fall. There are six possible orderings of $w_{ij}^{(h)}$, $w_{ij}^{(a)}$, and m_j. For each ordering there are two possible classification outcomes, for a total of 12 cases. Table 2 shows each case.

For example, in case 1 both $\mathbf{x}_\mathbf{i}^{(h)}$ and $\mathbf{x}_\mathbf{i}^{(a)}$ project to the right of m_j and are on the correct side of the cutting plane for the jth roll call and are therefore correctly classified. Case 2 is the same geometrically, only now $\mathbf{x}_\mathbf{i}^{(h)}$ and $\mathbf{x}_\mathbf{i}^{(a)}$ are on the wrong side of the cutting plane and are therefore projected as classification errors. Cases 1 to 8 represent no change in classification from moving the legislator point from $\mathbf{x}_\mathbf{i}^{(h)}$ to $\mathbf{x}_\mathbf{i}^{(a)}$. For $\mathbf{x}_\mathbf{i}^{(a)}$ to be an improvement over $\mathbf{x}_\mathbf{i}^{(h)}$, the number of cases 10 and 12 must be greater than the number of cases 9 and 11.

Consider the effect of moving $\mathbf{x}_\mathbf{i}^{(a)}$ farther from $\mathbf{x}_\mathbf{i}^{(h)}$. This has no effect on cases 1, 2, and 7–12. Only those cases where $\mathbf{x}_\mathbf{i}^{(a)}$ is between $\mathbf{x}_\mathbf{i}^{(h)}$ and m_j—cases 3, 4, 5, and 6—are affected. Depending upon how far $\mathbf{x}_\mathbf{i}^{(a)}$ is moved away from $\mathbf{x}_\mathbf{i}^{(h)}$, case 3 could change to case

Table 2

Case	*Ordering*	*Classification*		*Limits of α that correctly project $\mathbf{x}_i^{(h+1)}$*
		h	*a*	
1.	$-1 < m_j < w_{ij}^{(h)} < w_{ij}^{(a)} < +1$	C[1]	C	$\frac{m_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < \alpha_j < \frac{1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
2.	$-1 < m_j < w_{ij}^{(h)} < w_{ij}^{(a)} < +1$	I	I	$\frac{-1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < \alpha_j < \frac{m_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
3.	$-1 < m_j < w_{ij}^{(a)} < w_{ij}^{(h)} < +1$	C	C	$\frac{1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < \alpha_j < \frac{m_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
4.	$-1 < m_j < w_{ij}^{(a)} < w_{ij}^{(h)} < +1$	I	I	$\frac{m_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < \alpha_j < \frac{-1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
5.	$-1 < w_{ij}^{(h)} < w_{ij}^{(a)} < m_j < +1$	C	C	$\frac{-1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < \alpha_j < \frac{m_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
6.	$-1 < w_{ij}^{(h)} < w_{ij}^{(a)} < m_j < +1$	I	I	$\frac{m_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < \alpha_j < \frac{1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
7.	$-1 < w_{ij}^{(a)} < w_{ij}^{(h)} < m_j < +1$	C	C	$\frac{m_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < \alpha_j < \frac{-1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
8.	$-1 < w_{ij}^{(a)} < w_{ij}^{(h)} < m_j < +1$	I	I	$\frac{1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < \alpha_j < \frac{m_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
9.	$-1 < w_{ij}^{(h)} < m_j < w_{ij}^{(a)} < +1$	C	I	$\frac{-1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < \alpha_j < \frac{m_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
10.	$-1 < w_{ij}^{(h)} < m_j < w_{ij}^{(a)} < +1$	I	C	$\frac{m_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < \alpha_j < \frac{1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
11.	$-1 < w_{ij}^{(a)} < m_j < w_{ij}^{(h)} < +1$	C	I	$\frac{1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < \alpha_j < \frac{m_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
12.	$-1 < w_{ij}^{(a)} < m_j < w_{ij}^{(h)} < +1$	I	C	$\frac{m_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < \alpha_j < \frac{-1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$

[1]"C" is correctly classified; "I" is incorrectly classified.

11, increasing the error by one; case 5 could change to case 9, also increasing the error by one; case 4 could change to case 12, decreasing the error by one; and case 6 could change to case 10, also decreasing the error by one. A similar analysis of the effect of moving $\mathbf{x}_i^{(a)}$ toward $\mathbf{x}_i^{(h)}$ can also be done.

More generally, consider the equation of the line

$$\mathbf{x}_i^{(h+1)} = \mathbf{x}_i^{(h)} + \alpha\left(\mathbf{x}_i^{(a)} - \mathbf{x}_i^{(h)}\right) \tag{9}$$

which, when projected onto the jth normal vector, becomes:

$$w_{ij}^{(h+1)} = w_{ij}^{(h)} + \alpha\left(w_{ij}^{(a)} - w_{ij}^{(h)}\right) \tag{10}$$

For a single roll call, it is easy to solve for α; these are shown in Table 2 for all 12 cases. For example, for case 2, α must be chosen so that the projection of $\mathbf{x}_\mathbf{i}^{(h+1)}$, $w_{ij}^{(h+1)}$, is in the region $(-1, m_j)$.

Given $\mathbf{x}_\mathbf{i}^{(h)}$ and $\mathbf{x}_\mathbf{i}^{(a)}$, Table 2 can be used to find the limits of α for each roll call. Let the upper and lower limits for the jth roll call be U_{ij} and L_{ij}, respectively. The correct classification associated with $\mathbf{x}_\mathbf{i}^{(h)}$ can be obtained by setting $\alpha = 0$ and counting the number of roll calls for which $0 \in (L_{ij}, U_{ij})$. Similarly, the correct classification associated with $\mathbf{x}_\mathbf{i}^{(a)}$ is obtained by setting $\alpha = 1$ and counting the number of roll calls for which $1 \in (L_{ij}, U_{ij})$. In general, define

$$\delta_{ij} = 1 \qquad \text{if} \qquad \alpha \in (L_{ij}, U_{ij})$$
$$\delta_{ij} = 0 \qquad \text{if} \qquad \alpha \notin (L_{ij}, U_{ij})$$

and the correct classification is simply

$$\delta(\alpha) = \sum_{j=1}^{q} \delta_{ij} \tag{11}$$

The α that maximizes $\delta(\alpha)$, the number of correct classifications, can be calculated in a simple manner. First, compute the L_{ij} and U_{ij} for each roll call. Second, rank order the L_{ij} and U_{ij} and use the classification algorithm described in Section 2 to calculate the optimal α. Here the L_{ij} play the role of "y" and the U_{ij} play the role of "n." If there exists an α that results in perfect classification, the ordering of L's and U's will look like (dropping the i subscript to reduce clutter and numbering left to right for convenience)

$$L_1 < L_2 < L_3 < \cdots < L_q < U_1 < U_2 < U_3 < \cdots < U_q$$

that is, all the L_j will be less than all the U_j. In this example, perfect classification, $\delta(\alpha) = q$, results from $\alpha \in (L_q, U_1)$.

For example, using the configuration shown in Fig. 4, if the starting estimate ($h = 1$) $\mathbf{x}_\mathbf{i}^{(1)}$, is placed at the origin—point A—and the second point, $\mathbf{x}_\mathbf{i}^{(a)}$, is placed just to the right of $\mathbf{x}_\mathbf{i}^{(1)}$, then the resulting rank order of the upper and lower limits is

$$L_1 < L_5 < L_3 < L_2 < U_3 < U_2 < U_1 < U_5 < L_4 < U_4$$

The rank ordering is almost a perfect pattern in that four of the lower limits are below the five upper limits; only L_4 is wrongly placed producing one classification error. Consequently, the point resulting from using $\alpha \in (L_2, U_3)$, $\mathbf{x}_\mathbf{i}^{(2)}$, point B in Fig. 4, has only one classification error, with four correct classifications. [In practice, α is set equal to the midpoint; in this case, $(L_2 + U_3)/2$.] Note that in Fig. 4 point B is on the wrong side of the cutting line for roll call 4 in the region associated with the pattern NNNNN.

For the second iteration, $h = 2$, the starting estimate is $\mathbf{x}_\mathbf{i}^{(2)}$ and the second point, $\mathbf{x}_\mathbf{i}^{(a)}$, is placed just below $\mathbf{x}_\mathbf{i}^{(2)}$ so that the resulting line is parallel to the second dimension. This produces the rank ordering

$$L_4 < L_2 < L_5 < L_1 < L_3 < U_4 < U_1 < U_3 < U_5 < U_2$$

The rank ordering is now a perfect pattern, with all five lower limits below the five upper limits, so that there are no classification errors. The point resulting from using $\alpha \in (L_3, U_4)$, $\mathbf{x}_\mathbf{i}^{(3)}$, point C in Fig. 4, has five correct classifications and no classification error.

The search for the optimal $\mathbf{x_i}$ is conducted in a city-block manner. If the starting point is placed at the origin, then in the first iteration the search is along a line through the origin with all but the first dimension coordinates in $\mathbf{x_i}^{(1)}$ and $\mathbf{x_i}^{(a)}$ set to zero. In the second iteration, the first dimension coordinates are all set equal to the value corresponding to the optimal first dimension value and the 3rd-, 4th-, ..., sth-dimensional coordinates in $\mathbf{x_i}^{(2)}$ and $\mathbf{x_i}^{(a)}$ remain at their original values of zero. The search is along the corresponding line through $\mathbf{x_i}^{(2)}$ and $\mathbf{x_i}^{(a)}$, which is orthogonal to the first dimension. In the third iteration, the first- and second-dimension coordinates are set equal to the optimal values from the first- and second iterations respectively, and the 4th-, 5th-, ..., sth-dimensional coordinates in $\mathbf{x_i}^{(3)}$ and $\mathbf{x_i}^{(a)}$ remain at their original values of zero. The search is along the corresponding line through $\mathbf{x_i}^{(3)}$ and $\mathbf{x_i}^{(a)}$, which is orthogonal to the second dimension. This process continues in the same fashion through the sth dimension. Since the search for the optimal $\mathbf{x_i}$ is being done city-block-wise, dimensions *1* to *s* can now be searched again.

In sum, calculating the optimal $\mathbf{x_i}$ consists of the following steps.

1. Obtain a realistic starting estimate, $\mathbf{x_i}^{(1)}$ (or set $\mathbf{x_i}^{(1)}$ equal to the origin, that is, $\mathbf{x_i}^{(1)} = \mathbf{0}$).
2. Set $\mathbf{x_i}^{(a)\prime} = (0.01, x_{i2}^{(1)}, x_{i3}^{(1)}, x_{i4}^{(1)}, x_{i5}^{(1)}, \ldots, x_{is}^{(1)})$; find optimal α and $\mathbf{x_i}^{(2)} = \mathbf{x_i}^{(1)} + \alpha(\mathbf{x_i}^{(a)} - \mathbf{x_i}^{(1)})$.
3. Set $\mathbf{x_i}^{(a)\prime} = (x_{i1}^{(2)}, 0.01, x_{i3}^{(1)}, x_{i4}^{(1)}, x_{i5}^{(1)}, \ldots, x_{is}^{(1)})$; find optimal α and $\mathbf{x_i}^{(3)} = \mathbf{x_i}^{(2)} + \alpha(\mathbf{x_i}^{(a)} - \mathbf{x_i}^{(2)})$.
4. Set $\mathbf{x_i}^{(a)\prime} = (x_{i1}^{(2)}, x_{i2}^{(3)}, 0.01, x_{i4}^{(1)}, x_{i5}^{(1)}, \ldots, x_{is}^{(1)})$; find optimal α and $\mathbf{x_i}^{(4)} = \mathbf{x_i}^{(3)} + \alpha(\mathbf{x_i}^{(a)} - \mathbf{x_i}^{(3)})$.
5. Set $\mathbf{x_i}^{(a)\prime} = (x_{i1}^{(2)}, x_{i2}^{(3)}, x_{i3}^{(4)}, 0.01, x_{i5}^{(1)}, \ldots, x_{is}^{(1)})$; find optimal α and $\mathbf{x_i}^{(5)} = \mathbf{x_i}^{(4)} + \alpha(\mathbf{x_i}^{(a)} - \mathbf{x_i}^{(4)})$.

   Etc.

$s+1$. Set $\mathbf{x_i}^{(a)\prime} = \left(x_{i1}^{(2)}, x_{i2}^{(3)}, x_{i3}^{(4)}, x_{i4}^{(5)}, \ldots, x_{is-1}^{(s)}, 0.01\right)$; find optimal α and $\mathbf{x_i}^{(s+1)} = \mathbf{x_i}^{(s)} + \alpha(\mathbf{x_i}^{(a)} - \mathbf{x_i}^{(s)})$.

$s+2$. Go to step 2.

Note that classification error *can never increase from one step to the next*. This is true because setting $\alpha = 0$ preserves the current value of classification. This process converges very quickly (usually fewer than 10 iterations through steps 2 to $s+1$ above) to a vector of coordinates that is a local maximum in terms of classification. That is, it converges to a point such that $\alpha = 0$ for all s dimensions.

In practice, the starting estimate, $\mathbf{x_i}^{(1)}$, and the second point, $\mathbf{x_i}^{(a)}$, could be placed anywhere within the s-dimensional unit hypersphere. In practical applications the starting estimate is not the origin; rather, realistic starting values for the $\mathbf{x_i}^{(1)}$'s are obtained from an eigenvalue/eigenvector decomposition of the double-centered[7] agreement score matrix computed between legislators. The first s eigenvectors normalized to lie in the unit hypersphere are used as the starts.

[7] Technically, given a matrix of squared distances, double-centering is subtracting from each entry in the matrix the mean of the row and the mean of the column and adding the mean of the matrix. This has the effect of removing the squared terms from the matrix leaving just the cross-product matrix. It also reduces the rank of the matrix by one (see Young and Householder 1938; Ross and Cliff 1964).

If the line through $\mathbf{x_i}^{(h)}$ and $\mathbf{x_i}^{(a)}$ is parallel to a cutting line then the corresponding difference between $w_{ij}^{(a)}$ and $w_{ij}^{(h)}$, $w_{ij}^{(a)} - w_{ij}^{(h)}$, which is used in Table 2 to find α_j, may be equal to zero. This presents no problem since if the line through $\mathbf{x_i}^{(h)}$ and $\mathbf{x_i}^{(a)}$ is parallel to a cutting line, then the classification on that roll call is the same no matter where on the line $\mathbf{x_i}^{(h+1)}$ is located. Consequently, the roll call is not used to locate $\mathbf{x_i}^{(h+1)}$. In addition, if the line through $w_{ij}^{(a)}$ and $w_{ij}^{(h)}$ goes through the hypersphere so that it never intersects a cutting plane, this can result in a value of α_j that produces a point that lies outside the unit hypersphere. This is easily handled by computing the upper and lower feasible limits of $\mathbf{x_i}^{(h+1)}$—that is, the values corresponding to the two exit points of the line from the unit hypersphere— and discarding all the L_{ij} and U_{ij} outside the hypersphere. This requires some bookkeeping but it has no effect on the search process. Finally, the search process does not have to be done by moving orthogonally (i.e., city-block-wise) through the hypersphere. However, considerable experimentation shows that it is the most efficient way to proceed.

To guard against bad local maxima ($\alpha = 0$ in s orthogonal directions), multiple starting points for the $\mathbf{x_i}^{(1)}$'s are utilized. If different solutions are found (which are rare and almost always close together), then the lines joining the unique local maxima are searched for the best solution. After considerable experimentation, three starting points were found to work very well in practice. One starting point is from the eigenvalue–eigenvector decomposition of the double-centered agreement score matrix and the other two are randomly generated. After the first iteration, the legislator point from the previous iteration is used as one of the starting points.

Table 3 shows a Monte Carlo study of the legislator procedure using perfect data—the true cutting planes are known—for 100 legislators and 500 roll calls in 2 through 10 dimensions. (Appendix A3 shows results for contaminated data following the same design as that shown in Appendix A2.) To make the test reasonably stringent, only "unreasonable" starting points are used—namely, the origin and two randomly generated points. Results for one dimension are not shown since classification will always be 100% if perfect data are used. The 100 legislators and pairs of policy points were randomly drawn from a uniform distribution through the unit hypersphere. The pairs of policy points were drawn in such a way so as to produce cutting lines with an average majority margin of about 67% (typical of U.S. congressional roll call data; see Table 6). A maximum of 25 iterations through steps 2 to $s + 1$ above were allowed.

The legislator procedure works very well—especially at seven dimensions and below. There is some deterioration in accuracy at 10 dimensions but it still makes an average of only about 43 misclassifications of 50,000 total choices. For five dimensions and below it is practically perfect. Table 3 also shows the average squared Pearson correlations between the true and the reproduced legislator coordinates. The average of the worst and best r^2's for the s dimensions are shown.

These r^2's are very high. Even though the legislator procedure is nonparametric, with 500 roll call cutting planes, the unit hypersphere is chopped up into enough regions that, in effect, metric (i.e., ratio scale) information is being extracted from the roll call matrix. In three dimensions with 500 roll calls, there is a theoretical maximum of 20,833,751 regions created by the 500 cutting planes. Obviously, even if only a fraction of these regions is present, their average volume must be very small.[8]

[8] Actually measuring the volume of these regions is very difficult and is a problem that has not been satisfactorily solved [see Best et al. (1979) for a discussion]. To gain an idea of their size, an exhaustive search was conducted on one solution from each row in Table 3 and Table A5 in the Appendix. For each legislator point, the maximum

Table 3 Monte-Carlo tests of legislator procedure 100 legislators and 500 votes (perfect data) (each entry average of 10 trials, standard deviations in parentheses)

S	*Average majority margin*	*Average number of errors*	*Average percent correctly classified*	*Average worst leg. R-square*[a]	*Average best leg. R-square*[b]
2	65.8	0	100	.941	.998
	(1.1)	(0.0)	(0.000)	(.023)	(.001)
3	66.6	.4	99.999	.985	.998
	(0.8)	(0.7)	(0.001)	(.010)	(.000)
4	66.5	1.6	99.997	.992	.998
	(0.6)	(1.4)	(0.002)	(.003)	(.000)
5	67.4	3.2	99.99	.994	.998
	(0.6)	(2.2)	(0.004)	(.002)	(.000)
6	67.2	5.7	99.99	.995	.998
	(0.4)	(3.2)	(0.006)	(.001)	(.000)
7	67.9	13.7	99.98	.992	.998
	(0.8)	(5.5)	(0.01)	(.004)	(.001)
8	67.4	21.0	99.96	.992	.997
	(0.6)	(6.2)	(0.01)	(.002)	(.000)
9	67.4	30.7	99.94	.991	.996
	(0.7)	(6.8)	(0.01)	(.003)	(.001)
10	67.4	42.6	99.91	.990	.996
	(0.5)	(12.4)	(0.02)	(.002)	(.001)

[a]R-Squares computed between true and reproduced legislator coordinates. The number shown is the average of the worst r-squares across the 10 trials.
[b]R-Squares computed between true and reproduced legislator coordinates. The number shown is the average of the best r-squares across the 10 trials.

5 Nonparametric Unfolding of Binary Choice Matrices

The nonparametric unfolding algorithm is simply the cutting plane procedure and the legislative procedure chained together. Given starting estimates of the legislator coordinates, cutting planes are found that maximize correct classification of legislators' votes. These cutting planes are passed to the legislative procedure to obtain a new set of legislator points. The new set of legislator points is passed to the cutting plane procedure to get a new set of cutting planes, and so on. During this process the correct classification *can never decrease* from one phase to the next. These two phases can be repeated in sequence until no further improvement in correct classification occurs. In sum:

1. Generate starting values for $\mathbf{X}$, $\mathbf{X}^*$, from an eigenvalue/eigenvector decomposition of the legislator by legislator agreement score matrix.

distance to the cutting planes that bound the legislator was found by moving away from the legislator point in 100 random directions until a boundary plane was encountered. The average maximum distance is computed over the 100 legislators. For two dimensions, the average maximum distance of a legislator point to a boundary containing the legislator point was found to be about .028 with a standard deviation of about .019. For 10 dimensions, the numbers are about .030 and .012, respectively. These distances are closer to being "diameters" than "radiuses" because typically the legislator point was very close (less than .004) to a majority of the boundaries.

Table 4 Monte-Carlo tests: Non-parametric unfolding of binary choice matrices 100 legislators and 500 votes (each entry average of 10 trials, standard deviations in parentheses)

S	*Average majority margin*	*Average number of errors*	*Average percent correctly classified*	*Average worst leg. R-square*	*Average best leg. R-square*[a]	*Average fit with true normal vectors all*	*Average fit with true normal vectors 10% min.*
1	68.2	0	100.00	—	1.000	1.000	1.000
	(1.3)	(0.0)	(0.00)		(.000)	(.000)	(.000)
2	65.2	22.5	99.96	.940	.979	.995	.996
	(0.8)	(10.2)	(0.02)	(.014)	(.005)	(.002)	(.002)
3	66.0	10.1	99.98	.964	.981	.991	.993
	(1.0)	(4.0)	(0.01)	(.016)	(.007)	(.002)	(.002)
4	66.7	6.7	99.99	.967	.983	.987	.990
	(1.2)	(3.0)	(0.01)	(.011)	(.004)	(.003)	(.003)
5	66.8	6.5	99.99	.970	.980	.984	.987
	(0.6)	(3.0)	(0.01)	(.005)	(.003)	(.001)	(.001)
6	67.0	5.3	99.99	.959	.978	.978	.987
	(0.7)	(2.5)	(0.00)	(.006)	(.003)	(.001)	(.001)
7	67.3	6.5	99.99	.961	.977	.972	.979
	(0.6)	(3.5)	(0.01)	(.004)	(.003)	(.002)	(.001)
8	67.2	7.5	99.99	.953	.976	.966	.975
	(0.5)	(2.2)	(0.00)	(.012)	(.003)	(.002)	(.002)
9	67.8	6.6	99.99	.952	.971	.960	.970
	(0.7)	(1.3)	(0.00)	(.008)	(.003)	(.002)	(.003)
10	67.8	6.7	99.99	.943	.972	.957	.966
	(0.8)	(1.8)	(0.00)	(.012)	(.003)	(.002)	(.002)

[a]For s = 1, the squared Spearman Rank Correlation is computed between the 100 true and reproduced legislator ranks.

2. Given $\mathbf{X}^*$, find the optimal estimate of $\mathbf{N}$, $\mathbf{N}^*$ using the cutting plane procedure.
3. Given $\mathbf{N}^*$, find the optimal $\mathbf{X}^*$ using the legislative procedure.
4. Go to step 2.

Table 4 shows a Monte Carlo study of the nonparametric unfolding algorithm using perfect data for 100 legislators and 500 roll calls in 1 through 10 dimensions. Only roll calls with margins of 97–3 to 50–50 were used because unanimous and near-unanimous roll calls trivially inflate the number of correct classifications. A maximum of 25 iterations through steps 2 and 3 above was allowed.

The algorithm works well regardless of the number of dimensions. The worst result is for two dimensions, where, on average, about 23 of 50,000 choices were misclassified. The accuracy of the recovery of the true configuration of legislators and the true normal vectors declines after three dimensions but not very substantially. Even at 10 dimensions the average worst Pearson r^2 between the true and the reproduced legislator coordinates is .943. For four and fewer dimensions, the recovery is very precise.

The algorithm also works reasonably well when the dimensions are not equally salient. For example, in two dimensions if 85% of the cutting lines are nearly parallel to the second

dimension, the legislator configuration is recovered with reasonable precision—the average r^2's for the first and second dimensions are .94 and .89, respectively. However, in real-world applications where noise is present, such data will look like they fit a one-dimension model. Consequently, there is no substitute for the researcher's substantive understanding of the data.

Given the history of other multidimensional scaling techniques, most empirical applications of the nonparametric unfolding technique shown here will be to data matrices with missing entries and the estimated configurations will be in three or fewer dimensions. Missing data present no problem for the algorithm. In the cutting plane procedure it simply means that the total number of legislators may vary from vote to vote. In the legislator procedure it simply means that the number of cutting lines may vary from legislator to legislator. Handling missing data requires a little bookkeeping but it has no effect on the algorithm.

Appendix A4 shows a set of experiments with and without error at various levels of missing data. The algorithm performs well with real-world data at realistic levels of missing entries and error levels.

6 Empirical Examples

6.1 *Nonparametric Unfolding of U.S. Senate Roll Call Data*

Roll call voting in the U.S. Congress has been extensively analyzed by researchers using a wide variety of techniques. This will facilitate the interpretation of the nonparametric unfolding results. Two-dimensional senator coordinates from the nonparametric unfoldings are compared with those produced by KYST, a multidimensional scaling program developed by Kruskal et al. (1973), and W-NOMINATE, a maximum-likelihood procedure developed by Poole and Rosenthal (1997).

Table 5 reports the classification results for Senates 80 to 104 in one and two dimensions for the nonparametric unfolding procedure. These percentages are about 3 to 5% points better than W-NOMINATE in both one and two dimensions (Poole and Rosenthal 1997, Chap. 3). This is not surprising given that the W-NOMINATE procedure maximizes a likelihood function and does not attempt to maximize correct classifications.

Table 5 also shows the squared Pearson correlations between the estimated dimensions of the nonparametric unfolding and those produced by KYST and NOMINATE, respectively, in two dimensions. The nonparametric configuration was rotated to best match the NOMINATE and KYST configurations using Schonemann's (1966) technique. These r^2's are for the most part, very high—most of the first dimension r^2's are above .95 and the second-dimension r^2's are mostly above .9. r^2's were also computed for the Heckman–Snyder (1997) configurations. These were nearly the same as those reported for W-NOMINATE because the Heckman–Snyder and NOMINATE configurations are highly correlated (Poole and Rosenthal 1997, Appendix B). For example, the r^2's between the Heckman–Snyder configuration and the nonparametric configuration for the 85th Senate (shown in Fig. 5) are .973 and .944, respectively.

Table 6 shows the estimated rank order from the one-dimensional scaling of the 104th Senate. The ordering is from most liberal (1) to most conservative (103) and it correctly classifies 90.0% of the choices (70,976 of 78,882). Campbell of Colorado switched from Democrat to Republican in April of 1995 so he appears twice (ranks 48 and 55). If two or more senators tied in the ranking, the average of the associated ranks was used. For example, 85 senators were more liberal and 15 more conservative than the threesome Mack (R-FL),

Table 5 U.S. senate: 1947–1996 non-parametric unfolding of roll call data

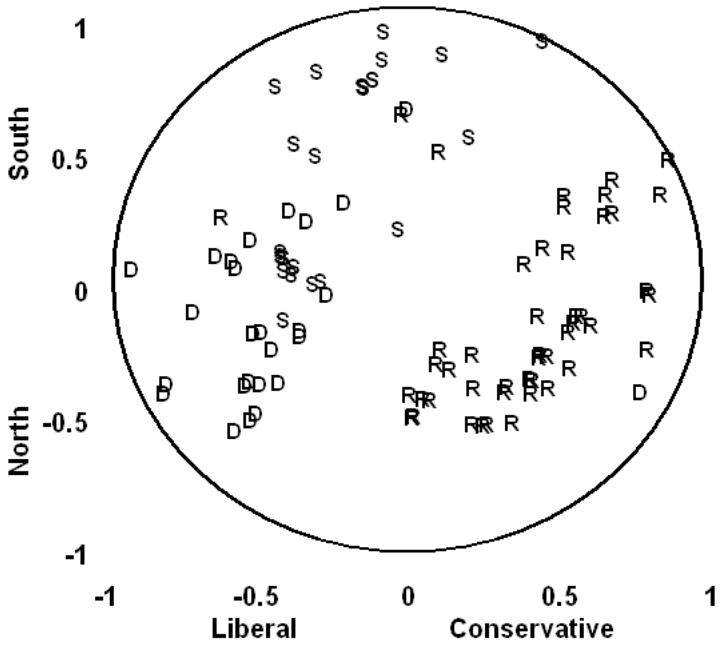
Senate	*Years*	*Senators*	*Roll calls*	*Total choices*	*Average margin*	*Non-P one*	*Non-P two*	*kyst R 1st*	*kyst R 2nd*	*nom R 1st*	*nom R 2nd*
104	1995–96	103[a]	805[b]	78,882[c]	.637	90.0[d]	91.3	.985[e]	.656	.980[f]	.785
103	1993–94	101	647	63,023	.672	89.2	90.4	.984	.744	.983	.818
102	1991–92	102	481	46,208	.685	86.9	88.5	.983	.812	.979	.879
101	1989–90	101	499	48,649	.680	85.4	87.1	.990	.860	.985	.884
100	1987–88	101	635	59,631	.709	87.7	89.5	.981	.722	.981	.925
99	1985–86	101	661	63,104	.688	84.7	86.8	.994	.841	.976	.936
98	1983–84	101	578	53,330	.698	84.8	87.3	.990	.900	.979	.954
97	1981–82	101	818	77,672	.682	85.5	88.1	.995	.897	.987	.956
96	1979–80	101	928	82,937	.683	83.5	85.8	.988	.803	.988	.965
95	1977–78	104	1037	92,868	.691	84.5	86.4	.989	.757	.977	.864
94	1975–76	100	1144	100,328	.691	86.3	88.6	.990	.888	.982	.934
93	1973–74	101	983	87,699	.695	85.1	87.5	.993	.908	.991	.961
92	1971–72	102	783	68,588	.676	85.0	88.6	.991	.944	.981	.971
91	1969–70	102	557	49,219	.681	84.5	88.1	.991	.898	.984	.973
90	1967–68	101	518	46,081	.699	83.6	87.2	.988	.899	.988	.949
89	1965–66	102	441	40.618	.681	85.4	88.4	.988	.901	.975	.912
88	1963–64	102	505	47,797	.686	85.0	90.1	.974	.963	.913	.937
87	1961–62	105	400	38,189	.675	87.3	90.6	.947	.960	.963	.933
86	1959–60	103	360	33,855	.686	84.9	89.6	.976	.962	.963	.956
85	1957–58	98	255	23,097	.669	84.7	89.5	.982	.924	.974	.895
84	1955–56	99	184	16,798	.659	85.5	90.4	.980	.927	.975	.925
83	1953–54	103	242	20,991	.672	86.9	90.3	.949	.731	.950	.906
82	1951–52	96	208	17,368	.659	86.0	89.4	.961	.769	.977	.928
81	1949–50	102	447	38,074	.667	85.0	88.6	.969	.862	.957	.955
80	1947–48	97	237	20,321	.665	88.0	90.8	.970	.885	.961	.917

[a]Number of Senators may exceed two times the number of States because of within Congress replacements.
[b]Number of roll calls with at least 2.5% voting, paired, or announced, on losing side.
[c]Total choices may not equal number of Senators times number of roll calls because of non-voting due to absences, etc.
[d]Classifications from non-parametric unfolding algorithm.
[e]Squared Pearson correlation between Senator coordinates from KYST and Senator coordinates from non-parametric unfolding. Non-parametric unfolding configuration rotated to best match KYST configuration.
[f]Squared Pearson correlation between Senator coordinates from W-NOMINATE and Senator coordinates from non-parametric unfolding. Non-parametric unfolding configuration rotated to best match W-NOMINATE configuration.

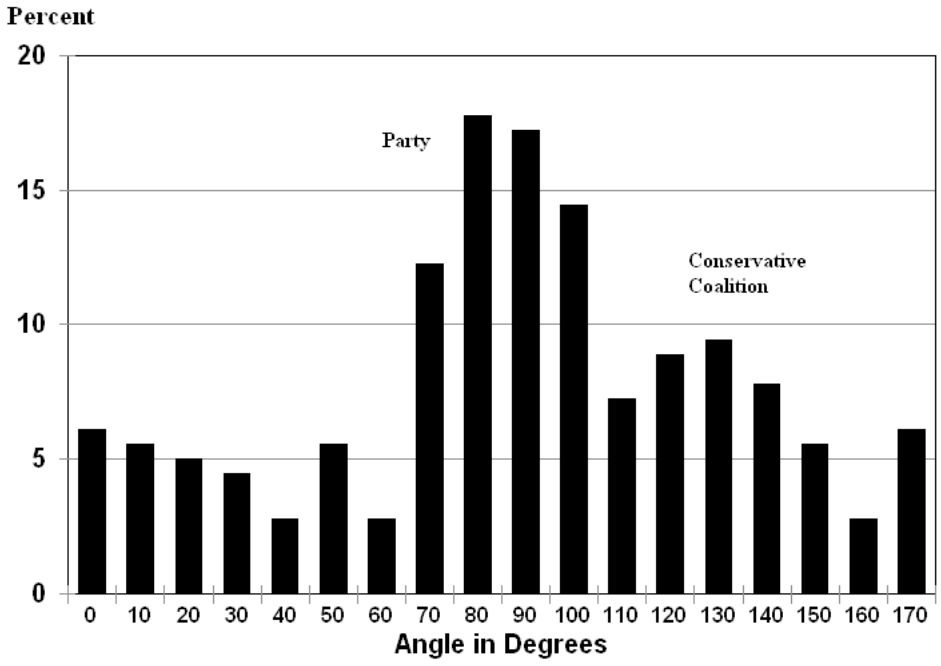
Coverdell (R-GA), and Coats (R-IN), who were tied. Consequently they all were assigned the average rank of 87.

The polarization of American politics (Poole and Rosenthal 1997; King 1998) is evident from an inspection of the table. There is no overlap of the two parties. Campbell's voting record as a Democrat made him the most conservative Democrat in the Senate. His conversion moved him only from 48th to 55th rank—from the right edge of the Democratic Party to the midst of the moderates of the Republican Party.

Figure 5 shows the two-dimensional configuration of senators for the 85th Senate along with a histogram of the roll call cutting line angles. The correct classification was 89.5% (20,679 of 23,097). The two major parties are clearly separated with the Democratic Party being split into its Northern and Southern wings. The 85th Senate occurred during the



(A)



(B)

Fig . 5 (A) Senator locations. D, Northern Democrat; S, Southern Democrat; R, Republican. (B) Eighty-fifth Senate, 1957–1958: cutting line angles.

Table 6 104th (1995–96) U.S. senate

Name	*Rank*	*Name*	*Rank*	*Name*	*Rank*
Simon (D-IL)	1	Hollings (D-SC)	36	Bennett (R-UT)	71
Wellstone (D-MN)	2	Inouye (D-HI)	37	Frist (R-TN)	72
Feingold (D-WI)	3	Ford (D-KY)	38	Santorum (R-PA)	73
Wyden (D-OR)	4	Reid (D-NV)	39	Shelby (R-AL)	74
Boxer (D-CA)	5	Robb (D-VA)	40	Burns (R-MT)	75
Kennedy (D-MA)	6	Exon (D-NE)	41	Hatch (R-UT)	76
Moseley-Braun (D-IL)	7	Lieberman (D-CT)	42	Abraham (R-MI)	77
Levin (D-MI)	8	Breaux (D-LA)	43	Gregg (R-NH)	78
Bradley (D-NJ)	9	Johnston (D-LA)	44	Thompson (R-TN)	79
Lautenberg (D-NJ)	10	Baucus (D-MT)	45	Grassley (R-IA)	80
Leahy (D-VT)	11	Nunn (D-GA)	46	Thurmond (R-SC)	81
Murray (D-WA)	12	Heflin (D-AL)	47	McConnell (R-KY)	82
Bumpers (D-AR)	13.5	Campbell (D-CO)	48	Hutchison (R-TX)	83.5
Harkin (D-IA)	13.5	Jeffords (R-VT)	49	Thomas (R-WY)	83.5
Pell (D-RI)	15	Cohen (R-ME)	50	Frahm (R-KS)	85
Kerry (D-MA)	16	Specter (R-PA)	51	Mack (R-FL)	87
Pryor (D-AR)	17	Snowe (R-ME)	52	Coverdell (R-GA)	87
Moynihan (D-NY)	18	Chafee (R-RI)	53	Coats (R-IN)	87
Kohl (D-WI)	19	Hatfield (R-OR)	54	Dole (R-KS)	89
Sarbanes (D-MD)	20	Campbell (R-CO)	55	Lott (R-MS)	90
Rockefeller (D-WV)	21	Kassebaum (R-KS)	56	Craig (R-ID)	91
Akaka (D-HI)	22	Packwood (R-OR)	57	Kempthorne (R-ID)	92
Daschle (D-SD)	23	Simpson (R-WY)	58	Grams (R-MN)	93
Dodd (D-CT)	24	Roth (R-DE)	59	Nickles (R-OK)	94
Mikulski (D-MD)	25	Stevens (R-AK)	60	Smith (R-NH)	95
Glenn (D-OH)	26	Gorton (R-WA)	61	McCain (R-AZ)	96
Conrad (D-ND)	27	D'Amato (R-NY)	62	Ashcroft (R-MO)	97
Dorgan (D-ND)	28	Dewine (R-OH)	63	Inhofe (R-OK)	98
Bingaman (D-NM)	29	Domenici (R-NM)	64	Gramm (R-TX)	99
Biden (D-DE)	30	Lugar (R-UT)	65	Helms (R-NC)	100
Byrd (D-WV)	31	Bond (R-MO)	66	Faircloth (R-NC)	101
Kerrey (D-NE)	32	Cochran (R-MS)	67	Brown (R-CO)	102
Bryan (D-NV)	33	Murkowski (R-AK)	68	Kyl (R-AZ)	103
Graham (D-FL)	34	Warner (R-VA)	69		
Feinstein (D-CA)	35	Pressler (R-SD)	70		

height of the three-party system that lasted from the late 1930s to the late 1970s (Cox and McCubbins 1993; Poole and Rosenthal 1997). The approximate angle of a party-line vote and the approximate angle of a conservative coalition vote (Northern Democrats versus a coalition of Southern Democrats and Republicans) are indicated in the histogram of the cutting line angles. The second dimension picked up the split in the Democratic Party over race-related issues.

Figure A3 in Appendix A3 shows bootstrapped standard errors for the 98 Senators shown in Fig. 5. Eighty of 98 Senators have standard deviations of less than .10 on the first dimension. The standard deviations for the second dimension are larger reflecting the fact that the bulk of the cutting lines are between 60 and 120° (see Fig. 5). Even so, 72 of 98 Senators have standard deviations of less than .15, which is small relative to the 2-unit diameter of the space. The standard deviations tend to be larger for those Senators near the

rim of the space. These "standard errors" must be taken with a grain of salt, however. They reflect both the volume of the polytopes within which the legislator points are placed and the underlying, unknown, error process. Nevertheless, given the stability of the unfolding procedure, they are useful descriptive statistics.

6.2 *Analyzing Rank-Order Data*

The nonparametric unfolding technique developed above can also be used to analyze the general rank-ordering problem. Suppose we have the rank orders over a set of stimuli from a group of individuals. For example, for six alternatives:

1. A > B > D > F > C > E
2. B > C > D > A > E > F
3. C > A > B > D > F > E
   Etc.
p. D > A > C > F > E > B

The individual rank orders can be converted into roll call data by viewing the stimuli as "voting" between pairs of individuals. Namely, for each pair of individuals, let the stimulus vote for the individual who has the stimulus higher in her ordering. For example, consider individuals 1 and 2 shown above. Let "Yea" be a vote for individual 1 and "Nay" be a vote for individual 2. Stimulus A would vote Yea, B Nay, C Nay, D would abstain, E Nay, and F Yea. This produces a q by $p(p-1)/2$ matrix of "roll calls." The unfolding technique will recover points representing the stimuli and cutting planes between each pair of individuals. The estimates of the stimuli will be very precise using this technique.

7 Conclusion

This paper shows a general nonparametric unfolding technique for maximizing the correct classification of binary preferential choice data. The motivation for and the primary focus of the unfolding technique are parliamentary roll call voting data but the procedures that implement the unfolding can also be applied to a variety of other problems.

Although neither the cutting plane nor the legislative procedure can be formally shown to converge to the global classification maximum, Monte Carlo tests show that both in fact work very well in practice. The Appendix to this paper shows that in the presence of error, the cutting plane procedure almost certainly passes through or very near to the classification maximum and the maximum can be recovered from the iteration record. The legislative procedure is guaranteed to converge to a very strong local maximum. That is, a local maximum for which the point cannot be moved in any orthogonal direction and have the correct classifications increase. When the two procedures are used together in an alternating framework to analyze binary choice matrices, their performance is very good. The Monte Carlo tests in Section 5 and the empirical applications in Section 6 are testimony to this fact.

References

Andrich, David. 1995. "Hyperbolic Cosine Latent Trait Models for Unfolding Direct Responses and Pairwise Preferences." *Applied Psychological Measurement* 19:269–290.

Bennett, Joseph F., and William L. Hays. 1960. "Multidimensional Unfolding: Determining the Dimensionality of Ranked Preference Data." *Psychometrika* 25:27–43.

Best, Alvin M., Forrest W. Young, and Robert G. Hall. 1979. "On the Precision of a Euclidean Structure." *Psychometrika* 44:395–408.

van Blokland-Vogelesang, Rian. 1991. *Unfolding and Group Consensus Ranking for Individual Preferences*. Leiden: DWSO Press.

Borg, Ingwer, and Patrick Groenen. 1997. *Modern Multidimensional Scaling: Theory and Applications*. New York: Springer-Verlag.

Carroll, J. Douglas. 1980. "Models and Methods for Multidimensional Analysis of Preferential Choice (or Other Dominance) Data." In *Similarity and Choice*, eds. E. D. Lantermann and H. Feger. Bern, Switzerland: Huber.

Chang, J. J., and J. Douglas Carroll. 1969. *How to Use MDPREF, a Computer Program for Multidimensional Analysis of Preference Data*, Computer manual. Murray Hill, NJ: Bell Labs.

Coombs, Clyde. 1950. "Psychological Scaling Without a Unit of Measurement." *Psychological Review* 57:148–158.

Coombs, Clyde. 1964. *A Theory of Data*. New York: Wiley.

Cox, Gary, and Mathew D. McCubbins. 1993. *Legislative Leviathan: Party Government in the House*. Berkeley: University of California Press.

DeSarbo, Wayne S., and Jaewun Cho. 1989. "A Stochastic Multidimensional Scaling Vector Threshold Model for the Spatial Representation of 'Pick Any/N' Data." *Psychometrika* 54:105–129.

DeSarbo, Wayne S., and Donna L. Hoffman. 1987. "Constructing MDS Joint Spaces from Binary Choice Data: A Multidimensional Unfolding Threshold Model for Marketing Research." *Journal of Marketing Research* 24:40–54.

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

Gifi, Albert. 1990. *Nonlinear Multivariate Analysis*. Chicester, England: Wiley.

Greene, William H. 1993. *Econometric Analysis*. Englewood Cliffs, NJ: Prentice Hall.

Heckman, James J., and James M. Snyder. 1997. "Linear Probability Models of the Demand for Attributes With an Empirical Application to Estimating the Preferences of Legislators." *Rand Journal of Economics* 28:142–189.

Heiser, Willem J. 1981. *Unfolding Analysis of Proximity Data*. Leiden: University of Leiden.

Hojo, Hiroshi. 1994. "A New Method for Multidimensional Unfolding." *Behaviormetrika* 21:131–147.

King, David C. 1998. "Party Competition and Polarization in American Politics." Paper presented at the 1998 Annual Meeting of the Midwest Political Science Association, Chicago.

Kruskal, Joseph B., Forrest W. Young, and Judith B. Seery. 1978. "How to Use KYST-2, a Very Flexible Program to Do Multidimensional Scaling and Unfolding." Murray Hill, NJ: Bell Laboratories (unpublished).

Lang, Serge. 1979. *Calculus of Several Variables*. Reading, MA: Addison Wesley.

Lingoes, James C. 1963. "Multiple Scalogram Analysis: A Set-Theoretic Model for Analyzing Dichotomous Items." *Education and Psychological Measurement* 23:501–524.

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

MacRae, Duncan, Jr. 1958. *Dimensions of Congressional Voting*. Berkeley: University of California Press.

Manski, Charles F. 1975. "Maximum Score Estimation of the Stochastic Utility Model of Choice." *Journal of Econometrics* 3:205–228.

Manski, Charles F. 1985. "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator." *Journal of Econometrics* 27:313–333.

Manski, Charles F., and T. Scott Thompson. 1986. "Operational Characteristics of Maximum Score Estimation." *Journal of Econometrics* 32:85–108.

McFadden, Daniel. 1976. "Quantal Choice Analysis: A Survey." *Annals of Economic and Social Measurement* 5:363–390.

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

Ross, John, and Norman Cliff. 1964. "A Generalization of the Interpoint Distance Model." *Psychometrika* 29:167–176.

Salas, Saturnino L., and Einar Hille. 1974. *Calculus: One and Several Variables with Analytic Geometry*. New York: John Wiley & Sons.

Schonemann, Peter H. 1966. "A Generalized Solution of the Orthogonal Procrustes Problem." *Psychometrika* 31:1–10.

Shye, Samuel, 1978. *Theory Construction and Data Analysis in the Behavioral Sciences*. San Francisco: Jossey–Bass.

Spector, L., and M. Mazzeo. 1980. "Probit Analysis and Economic Education." *Journal of Economic Education* 11:37–44.

Tucker, L. R. 1960. "Intra-individual and Inter-individual Multidimensionality." In *Psychological Scaling: Theory and Applications*, eds. H. Gulliksen and S. Messick. New York: Wiley.

Van Schuur, Wijbrandt H. 1992. "Nonparametric Unidimensional Unfolding for Multicategory Data." In *Political Analysis*, Vol. 4, ed. John H. Freeman. Ann Arbor: University of Michigan Press.

Weisberg, Herbert F. 1968. *Dimensional Analysis of Legislative Roll Calls*, Doctoral dissertation. Ann Arbor: University of Michigan.

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