Submitted to the Annals of Applied Statistics

HORSESHOES IN MULTIDIMENSIONAL SCALING AND KERNEL METHODS*

By Persi Diaconis†, Sharad Goel† and Susan Holmes‡

Stanford University and USC

Classical multidimensional scaling (MDS) is a method for visualizing high-dimensional point clouds by mapping to low-dimensional Euclidean space. This mapping is defined in terms of eigenfunctions of a matrix of interpoint proximities. In this paper we analyze in detail multidimensional scaling applied to a specific dataset: the 2005 United States House of Representatives roll call votes. MDS and kernel projections output 'horseshoes' that are characteristic of dimensionality reduction techniques. We show that in general, a latent ordering of the data gives rise to these patterns when one only has *local* information. That is, when only the interpoint distances for nearby points are known accurately. Our results provide a rigorous set of results and insight into manifold learning in the special case where the manifold is a curve.

1. Introduction. Classical multidimensional scaling is a widely used technique for dimensionality reduction in complex data sets, a central problem in pattern recognition and machine learning. In this paper we carefully analyze the output of MDS applied to the 2005 United States House of Representatives roll call votes. The resultant 3-dimensional mapping of legislators shows 'horseshoes' that are characteristic of a number of dimensionality reduction techniques, including principal components analysis and correspondence analysis. These patterns are heuristically attributed to a latent ordering of the data, e.g. the ranking of politicians within a left-right spectrum. Our work lends insight into this heuristic, and we present a rigorous analysis of the 'horseshoe phenomenon'. In particular, we find that horseshoes result from ordered data in which only local interpoint distances can be estimated accurately.

Ordination techniques are part of the ecologists' standard toolbox ter Braak [25], and methods for accounting for gradients (i.e. detrending the

*This work was part of a project funded by the French ANR under a Chaire d'Excellence at the University of Nice Sophia-Antipolis.

†Supported by a DARPA grant HR 0011-04-1-0025.

‡SH acknowledges funding by NSF-DMS grant 02-41246.

AMS 2000 subject classifications: Primary 60K35, 60K35; secondary 60K35

Keywords and phrases: horseshoes, multidimensional scaling, dimensionality reduction, principal components analysis, kernel methods

axes) Hill and Gauch [14] are standard in the analysis of MDS with chisquare distances, known as correspondence analysis. Some mathematical insight into the horseshoe phenomenon has been proposed in Podani and Miklos [19] and Iwatsubo [15].

The paper is structured as follows: In Section 1.1 we describe our data set and briefly discuss the output of MDS applied to these data. Section 1.2 describes the MDS method in detail. Section 2 states our main assumption–that legislators can be isometrically mapped into an interval–and presents a simple model for voting that is consistent with this metric requirement. In Section 3 we analyze the model and present the main results of the paper. Section 4 connects the model back to the data. The proofs of the theoretical results from Section 3 are presented in the Appendix.

1.1. *The Voting Data.* We apply multidimensional scaling to data generated by members of the 2005 United States House of Representatives with proximity between legislators defined via roll call votes Office of the Clerk - U.S. House of Representatives [17]. A full House consists of 435 members, and in 2005 there were 671 roll calls. The first two roll calls were a call of the House by States and the election of the Speaker, and so were excluded from our analysis. Hence, the data can be ordered into a 435×669 matrix $D = (d_{ij})$ with $d_{ij} \in \{1/2, -1/2, 0\}$ indicating, respectively, a vote of 'yea', 'nay', or 'not voting' by Representative i on roll call j. (Technically, a representative can vote 'present', but for purposes of our analysis this was treated as equivalent to 'not voting'). We further restricted our analysis to the 401 Representatives that voted on at least 90% of the roll calls (220 Republicans, 180 Democrats and 1 Independent) leading to a 401×669 matrix V of voting data. This step removed, for example, the Speaker of House Dennis Hastert (R-IL) who by custom votes only when his vote would be decisive, and Robert T. Matsui (D-CA) who passed away at the start the term.

As a first step, we define an empirical distance between legislators as

$$\hat{d}(l_i, l_j) = \frac{1}{669} \sum_{k=1}^{669} |v_{ik} - v_{jk}|. \tag{1.1}$$

Roughly, $\hat{d}(l_i, l_j)$ is the percentage of roll calls on which legislators l_i and l_j disagreed. This interpretation would be exact if not for the possibility of 'not voting'. In Section 2 we give some theoretical justification for this choice of distance, but it is nonetheless a natural metric on these data.

Now, it is reasonable that the empirical distance above captures the similarity of nearby legislators. To reflect the fact that $\hat{d}$ is most meaningful at

small scales, we define the proximity

$$P(i,j) = 1 - \exp\left(-\hat{d}(l_i, l_j)\right).$$

Then $P(i,j) \approx \hat{d}(l_i, l_j)$ for $d(l_i, l_j) \ll 1$ and $P(i,j)$ is not as sensitive to noise around relatively large values of $\hat{d}(l_i, l_j)$. This localization is a common feature of dimensionality reduction algorithms, for example eigenmap Niyogi [16], isomap Tenenbaum, de Silva, and Langford [24], local linear embedding Roweis and Saul [20] and kernel PCA Schölkopf, Smola, and Muller [22].

We apply MDS by double centering the squared distances built from the proximity matrix P and plotting the first three eigenfunctions weighted by their eigenvalues (see Section 1.2 for details). Figure 1 shows the results of the 3-dimensional MDS mapping. The most striking feature of the mapping is that the data separate into 'twin horseshoes'. In Figure 2 we have added color to indicate the political party affiliation of each Representative (blue for Democrat, red for Republican, and green for the lone independent – Rep. Bernie Sanders of Vermont). The output from MDS is qualitatively similar to that obtained from other dimensionality reduction techniques, such as principal components analysis applied directly to the voting matrix V.

In Sections 2 and 3 we build and analyze a model for the data in an effort to understand and interpret these pictures. Roughly our theory predicts that the Democrats, for example, are ordered along the blue curve in correspondence to their political ideology, i.e. how far they lean to the left. In Section 4 we discuss connections between the theory and the data. In particular, we explain why in the data, legislators at the political extremes are not quite at the tips of the projected curves, but rather are positioned slightly toward the center.

1.2. *Multidimensional Scaling.* Multidimensional Scaling (MDS) is a widely used technique for approximating the interpoint distances, or proximities, of points in a high-dimensional space by actual distances between points in a low-dimensional Euclidean space. See Torgerson [26], Young and Householder [30] for early, clear references, Shepard [23] for extensions from distances to ranked proximities, Cox and Cox [10] and Borg and Groenen [5] for useful textbook accounts. In our setting, applying the usual centering operations of MDS to the proximities we use as data lead to surprising numerical coincidences: the eigenfunctions of the centered matrices are remarkably close to the eigenfunctions of the original proximity matrix. The development below unravels this finding, and describes the multidimensional scaling procedure in detail.

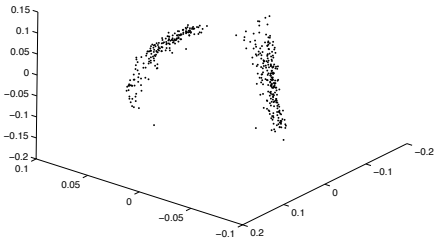

FIG 1. *3-Dimensional MDS output of legislators based on the 2005 U.S. House of Representatives roll call votes.*

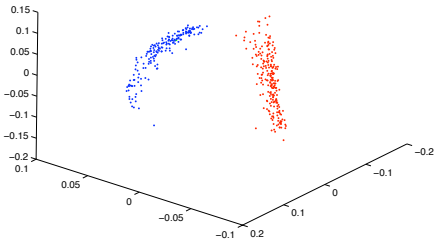

FIG 2. *3-Dimensional MDS output of legislators based on the 2005 U.S. House of Representatives roll call votes. Color has been added to indicate the party affiliation of each Representative.*

Euclidean Points: If $x_1, x_2, \ldots x_n \in \mathbb{R}^p$, let

$$d_{i,j} = \sqrt{(x_i^1 - x_j^1)^2 + \cdots + (x_i^p - x_j^p)^2}.$$

be the interpoint distance matrix. Schoenberg [21] characterized distance matrices as positive definite and gave an algorithmic solution for finding the points given the distances. Albouy [1] discusses the history of this problem, tracing it back to Borchardt [4] in 1866. Of course the points can only be reconstructed up to translation and rotation, thus we assume $\sum_{i=1}^n x_i = 0$.

To describe Schoenberg's procedure, first organize the unknown points into a $n \times p$ matrix X and consider the matrix of dot products $S = XX^T$, i.e. $S_{ij} = x_i x_j^T$. Then the spectral theorem for symmetric matrices yields $S = U\Lambda U^T$ for orthogonal U and diagonal Λ. Thus a set of n vectors which yield S is given by $X = U\Lambda^{1/2}$. This reduces the problem to finding the dot product matrix S from the interpoint distances. For this, observe,

$$d_{i,j}^2 = (x_i - x_j)(x_i - x_j)^T = x_i x_i^T + x_j x_j^T - 2x_i x_j^T$$

or

$$D_2 = s\mathbf{1}^T + \mathbf{1}s^T - 2S \tag{1.2}$$

where D_2 is the $n \times n$ matrix of squared distances, s is the $n \times 1$ vector of the diagonal entries of S, and $\mathbf{1}$ is the $n \times 1$ vector of ones. The matrix S can be obtained by *double centering* D_2:

$$S = -\frac{1}{2} H D_2 H \qquad H = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T. \tag{1.3}$$

To see this, first note that for any matrix A, HAH centers the rows and columns to have mean 0. Consequently, $Hs\mathbf{1}^T H = H\mathbf{1}s^T H = 0$ since the rows of $s\mathbf{1}^T$ and the columns of $\mathbf{1}s^T$ are constant. Pre- and post-multiplying (1.2) by H, we have

$$HD_2H = -2HSH.$$

Since the x's were chosen as centered the row sums of S satisfy

$$\sum_j x_i x_j^T = x_i(\sum_j x_j)^T = 0$$

and so $S = -\frac{1}{2} H D_2 H$ as claimed.

In summary, given an $n \times n$ matrix of interpoint distances, one can solve for the points by:

1. Double centering the interpoint distance squared matrix: $S = -\frac{1}{2} H D_2 H$.
2. Diagonalizing S: $S = U \Lambda U^T$.
3. Extracting X: $X = U \Lambda^{1/2}$.

Approximate Distance Matrices: The analysis above assumes that one starts with points $x_1, x_2, \ldots, x_n$ in a p-dimensional Euclidean space. We may want to find an embedding $x_i \Longrightarrow y_i$ in a space of dimension $k < p$ that preserves the interpoint distances as closely as possible. Assume that $S = U \Lambda U^T$ is such that the diagonal entries of Λ are decreasing. Set Y_k to be the matrix obtained by taking the first k columns of the U and scaling them so that their squared norms are equal to the eigenvalues Λ_k. In particular, this provides the first k columns of X above and solves the minimization problem

$$\min_{y_i \in \mathbb{R}^k} \sum_{i=1}^{n} \left(\|x_i - x_j\|_2 - \|y_i - y_j\|_2 \right)^2 . \tag{1.4}$$

Householder and Young (1938) [30] showed that this minimization can be realized as an eigenvalue problem. In applications, an observed matrix D is often not based on Euclidean distances but may represent 'proximity' or 'dissimilarity', or just the difference of ranks. Then, the MDS solution is a heuristic for finding points in a Euclidean space whose interpoint distances approximate the orders of the proximities D. This is called non-metric MDS [23].

Kernel methods: MDS converts similarities into inner products, whereas modern kernel methods [22] start with a given matrix of inner products. Williams [29] pointed out that Kernel PCA [22] is equivalent to metric MDS in feature space when the kernel function is chosen isotropic, i.e. the kernel $K(x, y)$ only depends on the norm $||x - y||$. The kernels we focus on in this paper have that property.

Relating the eigenfunctions of S to those of D_2: In practice it is easier to think about the eigenfunctions of the squared distances matrix D_2 rather than the recentered matrix $S = -\frac{1}{2} H D_2 H$.

Observe that if v is any vector such that $\mathbf{1}^T v = 0$ (i.e. the entries of v sum to 0) then

$$Hv = \left(I - \frac{1}{n} \mathbf{1}\mathbf{1}^T \right) v = v.$$

Now, suppose w is an eigenfunction of D_2 with eigenvalue λ, and let

$$\bar{w} = \left(\frac{1}{n} \sum_{i=1}^{n} w_i \right) \mathbf{1}$$

be the constant vector whose entries are the mean of w. Then $\mathbf{1}^T(w-\bar{w}) = 0$ and

$$\begin{aligned} S(w-\bar{w}) &= -\frac{1}{2}HD_2H(w-\bar{w}) \\ &= -\frac{1}{2}HD_2(w-\bar{w}) \\ &= -\frac{1}{2}H(\lambda w - \lambda\bar{w} + \lambda\bar{w} - D_2\bar{w}) \\ &= -\frac{\lambda}{2}(w-\bar{w}) + \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n} w_i\right)\begin{bmatrix} r_1 - \bar{r} \\ \vdots \\ r_n - \bar{r} \end{bmatrix} \end{aligned}$$

where $r_i = \sum_{j=1}^{n}(D_2)_{ij}$ and $\bar{r} = (1/n)\sum_{i=1}^{n} r_i$. In short, if w is an eigenfunction of D_2 and $\bar{w} = 0$ then w is also an eigenfunction of S. By continuity, if $\bar{w} \approx 0$ or $r_i \approx \bar{r}$, then $w - \bar{w}$ is an *approximate* eigenfunction of S. In our setting, it turns out that the matrix D_2 has approximately constant row sums (so $r_i \approx \bar{r}$), and its eigenfunctions satisfy $\bar{w} \approx 0$ (in fact, some satisfy $\bar{w} = 0$). Consequently, the eigenfunctions of the centered and uncentered matrix are approximately the same in our case.

2. A Model for the Data. In spatial models of roll call voting, legislators and policies are represented by points in a low-dimensional Euclidean space [11], with votes decided by maximizing a deterministic or stochastic utility function. For a precise description of the mathematics behind unfolding techniques see Jan de Leeuw's article [11], where he treats the particular case of unfolding roll call data such as ours.

Since Coombs [9] it has been understood that there is usually a natural left-right (i.e. unidimensional) model for political data. Recent comparisons [6] between the available left-right indices have shown that there is little difference, and that indices based on multidimensional scaling [13] perform well. They conclude 'standard roll call measures are good proxies of personal ideology and are still among the best measures available'.

In empirical work, it is often convenient to specify a parametric family of utility functions. In that context, the central problem is then to estimate those parameters and to find 'ideal points' for both the legislators and the policies. A robust Bayesian procedure for parameter estimation in spatial models of roll call data was introduced in [8], and provides a statistical framework for testing models of legislative behavior.

Although the empirical distance (1.1) is arguably a natural one to use on our data, we further motivate this choice by considering a theoretical

model in which legislators lie on a regular grid in a unidimensional policy space. In this idealized model it is natural to identify legislators l_i $1 \leq i \leq n$ with points in the interval $I = [0, 1]$ in correspondence with their political ideologies. We define the distance between legislators to be

$$d(l_i, l_j) = |l_i - l_j|.$$

This assumption that legislators can be isometrically mapped into an interval is key to our analysis. In the 'cut-point model' for voting, each bill $1 \leq k \leq m$ on which the legislators vote is represented as a pair

$$(C_k, P_k) \in [0, 1] \times \{0, 1\}.$$

We can think of P_k as indicating whether the bill is liberal ($P_k = 0$) or conservative ($P_k = 1$), and we can take C_k to be the cut-point between legislators that vote 'yea' or 'nay'. Let $V_{ik} \in \{1/2, -1/2\}$ indicate how legislator l_i votes on bill k. Then, in this model,

$$V_{ik} = \begin{cases} 1/2 - P_k & l_i \leq C_k \\ P_k - 1/2 & l_i > C_k \end{cases}.$$

As described, the model has $n + 2m$ parameters, one for each legislator and two for each bill. We reduce the number of parameters by assuming that the cut-points are independent random variables uniform on I. Then,

$$\mathbb{P}(V_{ik} \neq V_{jk}) = d(l_i, l_j) \tag{2.1}$$

since legislators l_i and l_j take opposites sides on a given bill if and only if the cut-point C_k divides them. Observe that the parameters P_k do not affect the probability above.

The empirical distance (1.1) between legislators l_i and l_j generalizes to

$$\hat{d}_m(l_i, l_j) = \frac{1}{m} \sum_{k=1}^{m} |V_{ik} - V_{jk}| = \frac{1}{m} \sum_{k=1}^{m} 1_{V_{ik} \neq V_{jk}}.$$

By (2.1), we can estimate the latent distance d between legislators by the empirical distance $\hat{d}$ which is computable from the voting record. In particular,

$$\lim_{m \to \infty} \hat{d}_m(l_i, l_j) = d(l_i, l_j) \quad \text{a.s.}$$

since we assumed the cut-points are independent. More precisely, we have the following result:

LEMMA 2.1. *For* $m \geq \log(n/\sqrt{\epsilon})/\epsilon^2$

$$\mathbb{P}\left(\left|\hat{d}_m(l_i, l_j) - d(l_i, l_j)\right| \leq \epsilon \ \ \forall\ 1 \leq i, j \leq n\right) \geq 1 - \epsilon.$$

PROOF. By the Hoeffding inequality, for fixed l_i and l_j

$$\mathbb{P}\left(\left|\hat{d}_m(l_i, l_j) - d(l_i, l_j)\right| > \epsilon\right) \leq 2e^{-2m\epsilon^2}.$$

Consequently,

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{1\leq i<j\leq n}\left|\hat{d}_m(l_i, l_j) - d(l_i, l_j)\right| > \epsilon\right) &\leq \sum_{1\leq i<j\leq n}\mathbb{P}\left(\left|\hat{d}_m(l_i, l_j) - d(l_i, l_j)\right| > \epsilon\right) \\
&\leq \binom{n}{2}2e^{-2m\epsilon^2} \\
&\leq \epsilon
\end{aligned}$$

for $m \geq \log(n/\sqrt{\epsilon})/\epsilon^2$ and the result follows. □

We identify legislators with points in the interval $I = [0, 1]$ and define the distances between them to be $d(l_i, l_j) = |l_i - l_j|$. This general description seems to be reasonable not only for applications in political science, but also in a number of other settings. The points and the exact distance d are usually unknown, however one can often estimate d from the data. For our work, we assume that one has access to an empirical distance that is *locally* accurate, i.e. we assume one can estimate the distance between nearby points.

3. Analysis of the Model.

3.1. *Eigenfunctions and Horseshoes.* In this section we analyze multidimensional scaling applied to metric models satisfying

$$d(x_i, x_j) = |i/n - j/n|.$$

This corresponds to the case in which legislators are uniformly spaced in I: $l_i = i/n$. Now, if all the interpoint distances were known precisely, classical scaling would reconstruct the points exactly (up to a reversal of direction). In applications, it is often not possible to have globally accurate information. Rather, one can only reasonably approximate the interpoint distances for nearby points. To reflect this limited knowledge, we work with the proximity

$$P(i, j) = 1 - \exp(-d(x_i, x_j)).$$

In fact, if we apply a linear proximity transformation, so that the proximities are given by

$$P(i,j) = |l_i - l_j|,$$

and we do not apply the double centering and just analyze the distances much of what we develop here stays true. This has been developed by the physicists as what they call the *crystal configuration* of a one-dimensional Anderson model and whose spectral decomposition is analyzed in [3].

As a matrix,

$$P = \begin{pmatrix} 0 & 1-e^{-1/n} & \dots & 1-e^{-(n-1)/n} \\ 1-e^{-1/n} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1-e^{-1/n} \\ 1-e^{-(n-1)/n} & \dots & 1-e^{-1/n} & 0 \end{pmatrix}.$$

We are interested in finding eigenfunctions for the doubly centered matrix

$$S = -\frac{1}{2}HPH = -\frac{1}{2}\left(P - JP - PJ + JPJ\right)$$

where $J = (1/n)\mathbf{1}\mathbf{1}^T$. To prove limiting results, we work with the scaled matrices $S_n = (1/n)S$. Approximate eigenfunctions for S_n are found by considering a limit K of the matrices S_n, and then solving the corresponding integral equation

$$\int_0^1 K(x,y)f(y)dy = \lambda f(x).$$

Standard matrix perturbation theory is then applied to recover approximate eigenfunctions for the original, discrete matrix.

When we continuize the scaled matrices S_n, we get the kernel defined for $(x,y) \in [0,1] \times [0,1]$

$$\begin{aligned} K(x,y) &= \frac{1}{2}\left(e^{-|x-y|} - \int_0^1 e^{-|x-y|}dx - \int_0^1 e^{-|x-y|}dy + \int_0^1\int_0^1 e^{-|x-y|}dxdy\right) \\ &= \frac{1}{2}\left(e^{-|x-y|} + e^{-y} + e^{-(1-y)} + e^{-x} + e^{-(1-x)}\right) + e^{-1} - 2. \end{aligned}$$

Recognizing this as a kernel similar to those in Fredholm equations of the second type suggests that there are trigonometric solutions, as we show in Theorem 3.2 below. The eigenfunctions we derive are in agreement with those arising from the voting data, lending considerable insight into our data analysis problem and, more importantly, the horseshoe phenomenon.

We state a classical perturbation result that relates two different notions of an *approximate eigenfunction*. A proof is included here to aid the reader. For more refined estimates, see [18] (Chapter 4, page 69).

THEOREM 3.1. *Consider an $n \times n$ symmetric matrix A with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$. If for $\epsilon > 0$*

$$\|Af - \lambda f\|_2 \leq \epsilon$$

for some f, λ with $\|f\|_2 = 1$, then A has an eigenvalue λ_k such that $|\lambda_k - \lambda| \leq \epsilon$.

If we further assume that

$$s = \min_{i:\lambda_i \neq \lambda_k} |\lambda_i - \lambda_k| > \epsilon$$

then A has an eigenfunction f_k such that $Af_k = \lambda_k f_k$ and $\|f - f_k\|_2 \leq \epsilon/(s - \epsilon)$.

PROOF. First we show that $\min_i |\lambda_i - \lambda| \leq \epsilon$. If $\min_i |\lambda_i - \lambda| = 0$ we are done; otherwise $A - \lambda I$ is invertible. Then,

$$\begin{aligned} \|f\|_2 &\leq \|(A - \lambda I)^{-1}\| \cdot \|(A - \lambda)f\|_2 \\ &\leq \epsilon \|(A - \lambda I)^{-1}\|. \end{aligned}$$

Since the eigenvalues of $(A - \lambda I)^{-1}$ are $1/(\lambda_1 - \lambda), \ldots, 1/(\lambda_n - \lambda)$, by symmetry

$$\|(A - \lambda I)^{-1}\| = \frac{1}{\min_i |\lambda_i - \lambda|}.$$

The result now follows since $\|f\|_2 = 1$.

Set $\lambda_k = \text{argmin}|\lambda_i - \lambda|$, and consider an orthonormal basis $g_1, \ldots, g_m$ of the associated eigenspace E_{λ_k}. Define f_k to be the projection of f onto E_{λ_k}:

$$f_k = \langle f, g_1 \rangle g_1 + \cdots + \langle f, g_m \rangle g_m.$$

Then f_k is an eigenfunction with eigenvalue λ_k. Writing $f = f_k + (f - f_k)$ we have

$$\begin{aligned} (A - \lambda I)f &= (A - \lambda I)f_k + (A - \lambda I)(f - f_k) \\ &= (\lambda_k - \lambda)f_k + (A - \lambda I)(f - f_k). \end{aligned}$$

Since $f - f_k \in E_{\lambda_k}^{\perp}$, by symmetry we have

$$\langle f_k, A(f - f_k) \rangle = \langle Af_k, f - f_k \rangle = \langle \lambda_k f_k, f - f_k \rangle = 0.$$

Consequently, $\langle f_k, (A - \lambda I)(f - f_k) \rangle = 0$ and by Pythagoras

$$\|Af - \lambda f\|_2^2 = (\lambda_k - \lambda)^2 \|f_k\|_2 + \|(A - \lambda I)(f - f_k)\|_2^2.$$

In particular,

$$\epsilon \geq \|Af - \lambda f\|_2 \geq \|(A - \lambda I)(f - f_k)\|_2.$$

For $\lambda_i \neq \lambda_k$, $|\lambda_i - \lambda| \geq s - \epsilon$. The result now follows since for $h \in E_{\lambda_k}^{\perp}$

$$\|(A - \lambda I)h\|_2 \geq (s - \epsilon)\|h\|_2.$$

□

REMARK 3.1. *The second statement of the theorem allows non-simple eigenvalues, but requires that the eigenvalues corresponding to distinct eigenspaces be well-separated.*

REMARK 3.2. *The eigenfunction bound of the theorem is asymptotically tight in ϵ as the following example illustrates: Consider the matrix*

$$A = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda + s \end{bmatrix}$$

with $s > 0$. For $\epsilon < s$ define the function

$$f = \begin{bmatrix} \sqrt{1 - \epsilon^2/s^2} \\ \epsilon/s \end{bmatrix}.$$

Then $\|f\|_2 = 1$ and $\|Af - \lambda f\|_2 = \epsilon$. The theorem guarantees that there is an eigenfunction f_k with eigenvalue λ_k such that $|\lambda - \lambda_k| \leq \epsilon$. Since the eigenvalues of A are λ and $\lambda + s$, and since $s > \epsilon$, we must have $\lambda_k = \lambda$. Let $V_k = \{f_k : Af_k = \lambda_k f_k\} = \{ce_1 : c \in \mathbb{R}\}$ where e_1 is the first standard basis vector. Then

$$\min_{f_k \in V_k} \|f - f_k\|_2 = \|f - (f \cdot e_1)e_1\| = \epsilon/s.$$

The bound of the theorem, $\epsilon/(s - \epsilon)$, is only slightly larger.

We establish an integral identity in order to find trigonometric solutions to $Kf = \lambda f$ where K is the continuized kernel of the centered exponential proximity matrix.

LEMMA 3.1. *(see proof in the appendix) For constants $a \in \mathbb{R}$ and $c \in [0, 1]$*

$$\int_0^1 e^{-|x-c|} \cos[a(x - 1/2)]dx$$
$$= \frac{2\cos[a(c - 1/2)]}{1 + a^2} + \frac{\left(e^{-c} + e^{c-1}\right)\left(a\sin(a/2) - \cos(a/2)\right)}{1 + a^2}$$

and

$$\int_0^1 e^{-|x-c|}\sin[a(x-1/2)]dx \\ = \frac{2\sin[a(c-1/2)]}{1+a^2} + \frac{\left(e^{-c}-e^{c-1}\right)\left(a\cos(a/2)+\sin(a/2)\right)}{1+a^2}.$$

We now derive eigenfunctions for the continuous kernel.

THEOREM 3.2. *(see proof in the appendix) For the kernel*

$$K(x,y) = \frac{1}{2}\left(e^{-|x-y|}+e^{-y}+e^{-(1-y)}+e^{-x}+e^{-(1-x)}\right)+e^{-1}-2$$

defined on $[0,1]\times[0,1]$ *the corresponding integral equation*

$$\int_0^1 K(x,y)f(y)dy = \lambda f(x)$$

has solutions

$$f(x) = \sin(a(x-1/2)) \qquad a\cot(a/2) = -1$$

and

$$f(x) = \cos(a(x-1/2)) - \frac{2}{a}\sin(a/2) \qquad \tan(a/2) = \frac{a}{2+3a^2}.$$

In both cases, $\lambda = 1/(1+a^2)$.

Theorem 3.2 states specific solutions to our integral equation. Now we show that in fact these are all the solutions with positive eigenvalues. To start, observe that for $0 \le x,y \le 1$, $e^{-1} \le e^{-|x-y|} \le 1$ and $e^{-1}+1 \le e^{-x}+e^{-(1-x)} \le 2e^{-1/2}$. Consequently,

$$-1 < \frac{3}{2}e^{-1}+1+e^{-1}-2 \le K(x,y) \le \frac{1}{2}+2e^{-1/2}+e^{-1}-2 < 1$$

and so $\|K\|_\infty < 1$. In particular, if λ is an eigenvalue of K, then $|\lambda| < 1$. Now suppose f is an eigenfunction of K, i.e.

$$\lambda f(x) = \int_0^1 \left[\frac{1}{2}\left(e^{-|x-y|}+e^{-x}+e^{-(1-x)}+e^{-y}+e^{-(1-y)}\right)+e^{-1}-2\right]f(y)dy.$$

Taking the derivative with respect to x, we see that f satisfies

$$\lambda f'(x) = \frac{1}{2}\int_0^1 \left(-e^{-|x-y|}H_y(x) - e^{-x}+e^{-(1-x)}\right)f(y)dy \tag{3.1}$$

where $H_y(x)$ is the Heaviside function, i.e. $H_y(x) = 1$ for $x \geq y$ and $H_y(x) = -1$ for $x < y$. Taking the derivative again, we get

$$\lambda f''(x) = -f(x) + \frac{1}{2}\int_0^1 \left(e^{-|x-y|} + e^{-x} + e^{-(1-x)}\right) f(y)dy. \tag{3.2}$$

Now, substituting back into the integral equation we see

$$\lambda f(x) = \lambda f''(x) + f(x) + \int_0^1 \left[\frac{1}{2}\left(e^{-y} + e^{-(1-y)}\right) + e^{-1} - 2\right] f(y)dy.$$

Taking one final derivative with respect to x, and setting $g(x) = f'(x)$ we see

$$g''(x) = \frac{\lambda - 1}{\lambda} g(x). \tag{3.3}$$

For $0 < \lambda < 1$, all the solutions to (3.3) can be written in the form

$$g(x) = A\sin(a(x - 1/2)) + B\cos(a(x - 1/2))$$

with $\lambda = 1/(1 + a^2)$. Consequently, $f(x)$ takes the form

$$f(x) = A\sin(a(x - 1/2)) + B\cos(a(x - 1/2)) + C.$$

Note that since $\int_0^1 K(x, y)dy = 0$, the constant function $c(x) \equiv 1$ is an eigenfunction of K with eigenvalue 0. Since K is symmetric, for any eigenfunction f with non-zero eigenvalue, f is orthogonal to c in $L^2(dx)$, i.e. $\int_0^1 f(x)dx = 0$. In particular, for $0 < \lambda < 1$, without loss we assume

$$f(x) = A\sin(a(x - 1/2)) + B\left[\cos(a(x - 1/2)) - \frac{2}{a}\sin(a/2)\right].$$

We solve for a, A and B. First assume $B \neq 0$, and divide f through by B. Then $f(1/2) = 1 - (2/a)\sin(a/2)$. Since $K(x, \cdot)$ is symmetric about $1/2$ and

$\sin(a(x-1/2))$ is skew-symmetric about $1/2$ we have

$$\begin{aligned}
\lambda f(1/2) &= \frac{1-(2/a)\sin(a/2)}{1+a^2} \\
&= \int_0^1 \left[\frac{1}{2}\left(e^{|y-1/2|}+e^{-y}+e^{-(1-y)}\right)+e^{-1/2}+e^{-1}-2\right] f(y)dy \\
&= \frac{1}{2}\int_0^1 \left(e^{|y-1/2|}+e^{-y}+e^{-(1-y)}\right)\cos(a(y-1/2))dy \\
&\quad + \frac{2}{a}\sin(a/2)\left(e^{-1/2}+e^{-1}-2\right) \\
&= \frac{1}{1+a^2}+\frac{e^{-1/2}\left(a\sin(a/2)-\cos(a/2)\right)}{1+a^2} \\
&\quad + \frac{a\sin(a/2)(1+e^{-1})}{1+a^2}+\frac{\cos(a/2)(1-e^{-1})}{1+a^2} \\
&\quad + \frac{2}{a}\sin(a/2)\left(e^{-1/2}+e^{-1}-2\right).
\end{aligned}$$

The last equality follows from Lemma 3.1. Equating the sides, a satisfies

$$\begin{aligned}
0 &= 2\sin(a/2)+e^{-1/2}a\left(a\sin(a/2)-\cos(a/2)\right)+a^2\sin(a/2)(1+e^{-1}) \\
&\quad + a\cos(a/2)(1-e^{-1})+2(1+a^2)\sin(a/2)\left(e^{-1/2}+e^{-1}-2\right) \\
&= \left(1-e^{-1/2}-e^{-1}\right)\left(a\cos(a/2)-2\sin(a/2)-3a^2\sin(a/2)\right).
\end{aligned}$$

From this it is immediate that $\tan(a/2) = a/(2+3a^2)$. Now we suppose $A \neq 0$ and divide f through by A. Then $f'(1/2) = a$ and from (3.1)

$$\begin{aligned}
\lambda f'(1/2) &= \frac{a}{1+a^2} \\
&= -\frac{1}{2}\int_0^1 e^{-|y-1/2|}H_y(1/2)f(y)dy \\
&= -\frac{1}{2}\int_0^1 e^{-|y-1/2|}H_y(1/2)\sin(a(y-1/2))dy \\
&= -\frac{e^{-1/2}}{1+a^2}\left(a\cos(a/2)+\sin(a/2)\right)+\frac{a}{1+a^2}.
\end{aligned}$$

In particular, $a\cot(a/2) = -1$.

The solutions of $\tan(a/2) = a/(2+3a^2)$ are approximately $2k\pi$ for integers k and the solutions of $a\cot(a/2) = -1$ are approximately $(2k+1)\pi$. Lemma 3.2 makes this precise. Since they do not have any common solutions, $A = 0$ if and only if $B \neq 0$. This completes the argument that Theorem 3.2 lists all the eigenfunctions of K with positive eigenvalues.

LEMMA 3.2. *(see proof in the appendix)*

1. *The positive solutions of* $\tan(a/2) = a/(2+3a^2)$ *lie in the set*

$$\bigcup_{k=1}^{\infty} (2k\pi, 2k\pi + 1/3k\pi)$$

with exactly one solution per interval. Furthermore, a is a solution if and only if $-a$ is a solution.

2. *The positive solutions of* $a\cot(a/2) = -1$ *lie in the set*

$$\bigcup_{k=0}^{\infty} ((2k+1)\pi, (2k+1)\pi + 1/(k\pi + \pi/2))$$

with exactly one solution per interval. Furthermore, a is a solution if and only if $-a$ is a solution.

The exact eigenfunctions for the continuous kernel yield approximate eigenfunctions and eigenvalues for the discrete case.

THEOREM 3.3. *Consider the centered and scaled proximity matrix defined by*

$$S_n(x_i, x_j) = \frac{1}{2n}\left(e^{-|i-j|/n} + e^{-i/n} + e^{-(1-i/n)} + e^{-j/n} + e^{-(1-j/n)} + 2e^{-1} - 4\right)$$

for $1 \le i, j \le n$.

1. *Set* $f_{n,a}(x_i) = \cos(a(i/n - 1/2)) - (2/a)\sin(a/2)$ *where a is a positive solution to* $\tan(a/2) = a/(2+3a^2)$. *Then*

$$S_n f_{n,a}(x_i) = \frac{1}{1+a^2} f_{n,a}(x_i) + R_{f,n} \quad \text{where } |R_{f,n}| \le \frac{a+4}{2n}.$$

2. *Set* $g_{n,a}(x_i) = \sin(a(i/n - 1/2))$ *where a is a positive solution to* $a\cot(a/2) = -1$. *Then*

$$S_n g_{n,a}(x_i) = \frac{1}{1+a^2} g_{n,a}(x_i) + R_{g,n} \quad \text{where } |R_{g,n}| \le \frac{a+2}{2n}.$$

That is, $f_{n,a}$ and $g_{n,a}$ are approximate eigenfunctions of S_n.

PROOF. That f and g are approximate eigenfunctions for the discrete matrix follows directly from Theorem 3.2. Suppose K is the continuous

kernel. Then,

$$\begin{aligned} S_n f_{n,a}(x_i) &= \sum_{j=1}^{n} S_n(x_i, x_j)[\cos(a(j/n - 1/2)) - (2/a)\sin(a/2)] \\ &= \int_0^1 K(x_i, y)[\cos(a(y - 1/2)) - (2/a)\sin(a/2)]dy + R_{f,n} \\ &= \frac{1}{1+a^2} f_{n,a}(x_i) + R_{f,n} \end{aligned}$$

where the error term satisfies

$$|R_{f,n}| \leq \frac{M}{2n} \quad \text{for } M \geq \sup_{0\leq x\leq 1} \left| \frac{d}{dx} K(x_i, y)[\cos(a(y - 1/2)) - (2/a)\sin(a/2)] \right|$$

by the standard right-hand rule error bound. In particular, we can take $M = a + 4$ independent of j, from which the result for $f_{n,a}$ follows. The case of $g_{n,k}$ is analogous. □

THEOREM 3.4. *Consider the setting of Theorem 3.3 and let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of S_n.*

1. *For positive solutions to* $\tan(a/2) = a/(2 + 3a^2)$

$$\min_{1\leq i\leq n} \left| \lambda_i - \frac{1}{1+a^2} \right| \leq \frac{a+4}{\sqrt{n}}.$$

2. *For positive solutions to* $a\cot(a/2) = -1$

$$\min_{1\leq i\leq n} \left| \lambda_i - \frac{1}{1+a^2} \right| \leq \frac{a+2}{\sqrt{n}}.$$

3.1.1. *Horseshoes and Twin Horseshoes.* The 2-dimensional MDS mapping is built out of the first and second eigenfunctions of the centered proximity matrix. As shown above, we have the approximate eigenfunctions

- $f_{n,a_1}(x_i) = \sin(3.67(i/n - 1/2))$ with eigenvalue $\lambda_1 \approx 0.07$
- $f_{n,a_2}(x_i) = \cos(6.39(i/n - 1/2))$ with eigenvalue $\lambda_2 \approx 0.02$

where the eigenvalues are for the scaled matrix. Figure 3 shows a graph of these eigenfunctions. Moreover, Figure 4 shows the horseshoe that results from plotting $\Lambda : x_i \mapsto (\sqrt{\lambda_1} f_1(x_i), \sqrt{\lambda_2} f_2(x_i))$. From Λ it is possible to deduce the relative order of the representatives in the interval I. Since $-f_1$ is also an eigenfunction, it is not in general possible to determine the absolute order knowing only that Λ comes from the eigenfunctions. However, as can be seen in Figure 4 the relationship between the two eigenfunctions is a

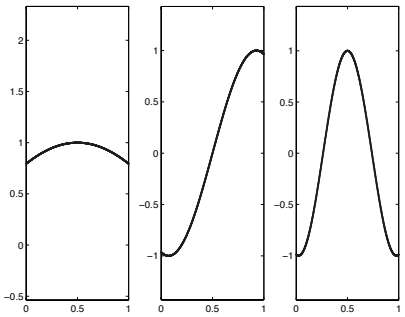

FIG 3. *Approximate eigenfunctions f_1, f_2 and f_3.*

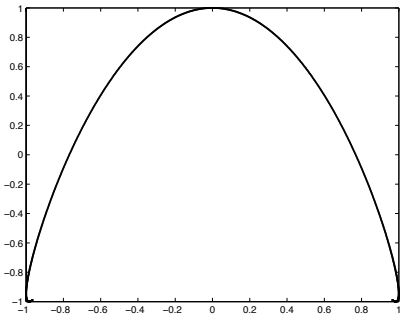

FIG 4. *A horseshoe that results from plotting $\Lambda : x_i \mapsto (\sqrt{\lambda_1} f_1(x_i), \sqrt{\lambda_2} f_2(x_i))$.*

curve for which we have the parametrization given above, but cannot be written in functional form.

With the voting data, we see not one, but two horseshoes. To see how this can happen, consider the two population state space $\mathcal{X} = \{x_1, \ldots, x_n, y_1, \ldots, y_n\}$ with proximity $d(x_i, x_j) = 1 - e^{-|i/n - j/n|}$, $d(y_i, y_j) = 1 - e^{-|i/n - j/n|}$ and $d(x_i, y_j) = 1$. This leads to the partitioned proximity matrix

$$\tilde{P}_{2n} = \left[\begin{array}{c|c} P_n & 1 \\ \hline 1 & P_n \end{array}\right]$$

where $P_n(i,j) = 1 - e^{-|i/n - j/n|}$. Since this proximity is ultimately centered, we can equivalently consider the distance

$$\tilde{P}_{2n} - \mathbf{1}\mathbf{1}^T = -\left[\begin{array}{c|c} A & 0 \\ \hline 0 & A \end{array}\right]$$

where $A(i,j) = e^{-|i/n - j/n|}$. The first eigenvector provides the separation into two groups, this is a well known method for separating clusters known today as spectral clustering, for a nice survey and consistency results see von Luxburg, Belkin, and Bousquet [27].

THEOREM 3.5. *For $1 \le i, j \le n$, consider the matrices defined by*

$$A_n(i,j) = \frac{1}{2n} e^{-|i-j|/n} \text{ and } S_n(i,j) = A_n - \frac{1}{2n}\mathbf{1}\mathbf{1}^T$$

1. *Set $f_{n,a}(x_i) = \cos(a(i/n - 1/2))$ where a is a positive solution to $a\tan(a/2) = 1$*
   Then
   $$A_n f_{n,a}(x_i) = \frac{1}{1+a^2} f_{n,a}(x_i) + R_{f,n} \quad \text{where } |R_{f,n}| \le \frac{a+1}{2n}.$$
2. *Set $g_{n,a}(x_i) = \sin(a(i/n - 1/2))$ where a is a positive solution to $a\cot(a/2) = -1$*
   Then
   $$S_n g_{n,a}(x_i) = \frac{1}{1+a^2} g_{n,a}(x_i) + R_{g,n} \quad \text{where } |R_{g,n}| \le \frac{a+1}{2n}.$$

That is, $f_{n,a}$ and $g_{n,a}$ are approximate eigenfunctions of A_n and S_n.

PROOF. The proof is analogous to Theorem 3.3 by way of Lemma 3.1 and so is omitted here. □

COROLLARY 3.1. *From Theorem 3.5 we have the following approximate eigenfunctions and eigenvalues for* $-(1/2n)\tilde{P}_{2n}$*:*

- $f_1(i) = \cos(a_1(i/n - 1/2))$*, for* $1 \le i \le n$ $f_1(j) = -\cos(a_1((j-n)/n - 1/2))$ *for* $(n+1) \le j \le 2n$*, where* $a_1 \approx 1.3$ *and* $\lambda_1 \approx 0.37$
- $f_2(i) = \sin(a_2(i/n - 1/2))$*, for* $1 \le i \le n$ $f_2(j) = 0$ *for* $(n+1) \le j \le 2n$*, where* $a_2 \approx 3.67$ *and* $\lambda_2 \approx 0.069$
- $f_3(i) = 0$*, for* $1 \le i \le n$*,* $f_3(j) = \sin(a_2((j-n)/n - 1/2))$ *for* $(n+1) \le j \le 2n$*, where* $a_2 \approx 3.67$ *and* $\lambda_3 \approx 0.069$

PROOF.

$$-\frac{1}{2n}\tilde{P}_{2n} = \left[\begin{array}{c|c} A_n & 0 \\ \hline 0 & A_n \end{array}\right] - \frac{1}{2n}\mathbf{1}\mathbf{1}^T$$

If we call u_1 the first eigenvector of A_n then the vector of length $2n$ $(u_1, -u_1)$ will be the eigenvector of $-\frac{1}{2n}\tilde{P}_{2n}$ since

$$\left[\begin{array}{c|c} A_n & 0 \\ \hline 0 & A_n \end{array}\right] - \frac{1}{2n}\mathbf{1}\mathbf{1}^T \begin{pmatrix} u_1 \\ -u_1 \end{pmatrix} = \lambda_1 \begin{pmatrix} u_1 \\ -u_1 \end{pmatrix} + 0$$

□

Moreover, since f_1, f_2 and f_3 are all orthogonal to constant functions, they are also approximate eigenfunctions for the centered, scaled matrix $(-1/2n)HPH$.

The functions are graphed in Figure 5, and the twin horseshoes that result from the 3-dimensional mapping $\Lambda : z \mapsto (\sqrt{\lambda_1} f_1(z), \sqrt{\lambda_2} f_2(z), \sqrt{\lambda_3} f_3(z))$ are shown in Figure 6.

Remark:

In fact the matrices A_n and $\tilde{P}_{2n}$ above are centrosymmetric [28], that is symmetrical around the center of the matrix. Formally, if K is the matrix with 1's in the counter (or secondary) diagonal,

$$K = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & & \vdots & & \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix}$$

then a matrix B is centrosymmetric iff $BK = KB$. A very useful review by Weaver [28], quotes I.J. Good [12] on the connection between centrosymmetric matrices and kernels of integral equations: *"Toeplitz matrices (which are*

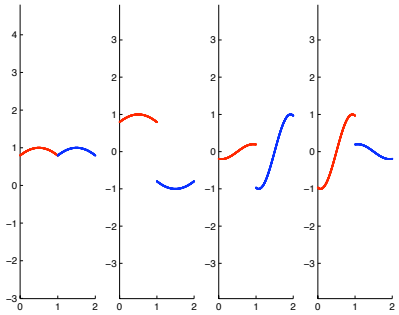

FIG 5. *Approximate eigenfunctions f_0, f_1, f_2, and f_3 for the centered proximity matrix arising from the two population model.*

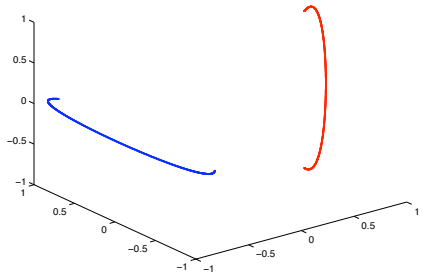

FIG 6. *Twin horseshoes in the two population model that result from plotting $\Lambda : z \mapsto (\sqrt{\lambda_1} f_1(z), \sqrt{\lambda_2} f_2(z), \sqrt{\lambda_3} f_3(z))$.*

examples of matrices which are both symmetric and centrosymmetric) arise as discrete approximations to kernels $k(x,t)$ of integral equations when these kernels are functions of $|x-t|$." (Today we would call these isotropic kernels). "Similarly if a kernel is an even function of its vector argument (x, t), that is, if $k(x,t) = k(-x,-t)$, then it can be discretely approximated by a centrosymmetric matrix."

Centrosymmetric matrices have very neat eigenvector formulas [7]. In particular if the order of the matrix, n, is even, then the first eigenvector is skew symmetric and thus of the form $(u_1, -u_1)$ and orthogonal to the constant vector. This explains the miracle that seems to occur in the simplification of the eigenvectors in the above formulæ.

4. Connecting the Model to the Data. When we apply MDS to the voting data, the first three eigenvalues are

- 0.13192
- 0.00764
- 0.00634

Observe that as our two population model suggests, the second and third eigenvalues are about equal and significantly smaller than the first.

Figure 7 shows the first, second and third eigenfunctions f_1, f_2 and f_3 from the voting data. The 3-dimensional MDS plot in Figure 1 is the graph of $\Lambda : x_i \mapsto (\sqrt{\lambda_1} f_1(x_i), \sqrt{\lambda_2} f_2(x_i), \sqrt{\lambda_3} f_3(x_i))$. Since legislators are not a priori ordered, the eigenfunctions are difficult to interpret. However, our model suggests the following ordering: Split the legislators into two groups G_1 and G_2 based on the sign of $f_1(x_i)$; then the norm of f_2 is larger on one group, say G_1, so we sort G_1 based on increasing values of f_2, and similarly, sort G_2 via f_3. Figure 8 shows the same data as does Figure 7, but with this judicious ordering of the legislators. Figure 9 shows the ordered eigenfunctions obtained from MDS applied to the 2004 roll call data. The results appear to be in agreement with the theoretically derived functions in Figure 5.

The theoretical second and third eigenfunctions are part of a two-dimensional eigenspace. In the voting data, it is reasonable to assume that noise eliminates symmetry and collapses the eigenspaces down to one dimension. Nonetheless, we would guess that the second and third eigenfunctions in the voting data are in the two-dimensional predicted eigenspace, as is seen to be the case in Figures 8 and 9.

Our analysis in Section 3 suggests that if legislators are in fact isometrically embedded in the interval I (relative to the roll call distance), then their MDS derived rank will be consistent with the order of legislators in the

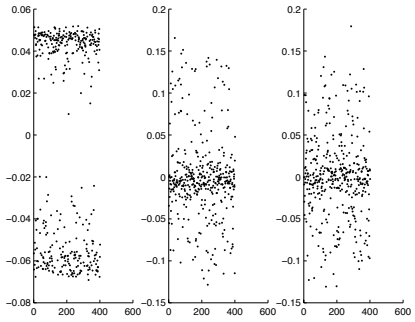

FIG 7. *The first, second and third eigenfunctions output from MDS applied to the 2005 U.S. House of Representatives roll call votes.*

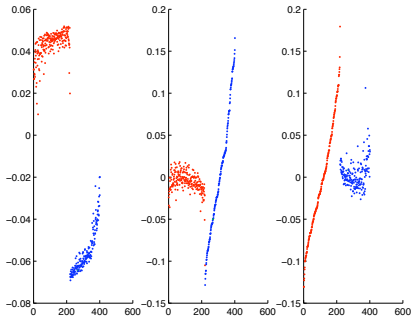

FIG 8. *The re-indexed first, second and third eigenfunctions output from MDS applied to the 2005 U.S. House of Representatives roll call votes. Colors indicate political parties.*

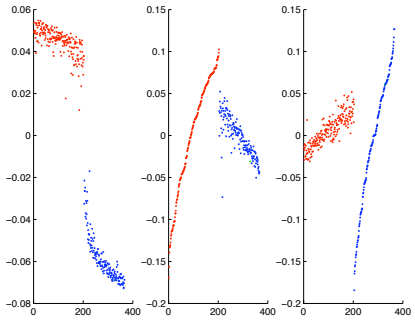

FIG 9. *The re-indexed first, second and third eigenfunctions output from MDS applied to the 2004 U.S. House of Representatives roll call votes. Colors indicate political parties.*

interval. This appears to be the case in the data, as seen in Figure 10, which shows a graph of $\hat{d}(l_i, \cdot)$ for selected legislators l_i. For example, as we would predict, $\hat{d}(l_1, \cdot)$ is an increasing function and $\hat{d}(l_n, \cdot)$ is decreasing. Moreover, the data seem to be in rough agreement with the metric assumption of our two population model, namely that the two groups are well-separated and that the within group distance is given by $d(l_i, l_j) = |i/n - j/n|$.

Our voting model suggests that the MDS ordering of legislators should correspond to political ideology. To test this, we compared the MDS results to the assessment of legislators by Americans for Democratic Action [2]. Each year, ADA selects 20 votes it considers the most important during that session, for example, the Patriot Act reauthorization. Legislators are assigned a Liberal Quotient: the percentage of those 20 votes on which the Representative voted in accordance with what ADA considered to be the liberal position. For example, a representative who voted the liberal position on all 20 votes would receive an LQ of 100%. Figure 11 below shows a plot of LQ vs. MDS rank.

For the most part, the two measures are consistent. However, MDS separates two groups of relatively liberal Republicans. To see why this is the case, consider the two legislators Mary Bono (R-CA) with MDS rank 248 and Gil Gutknecht (R-MN) with rank 373. Both Representatives received an ADA rating of 15%, yet had considerably different voting records. On the 20 ADA bills, both Bono and Gutknecht supported the liberal position 3 times – but never simultaneously. Consequently, the empirical roll call distance between them is relatively large considering that they are both Republicans. Since MDS attempts to preserve local distances, Bono and Gutknecht are

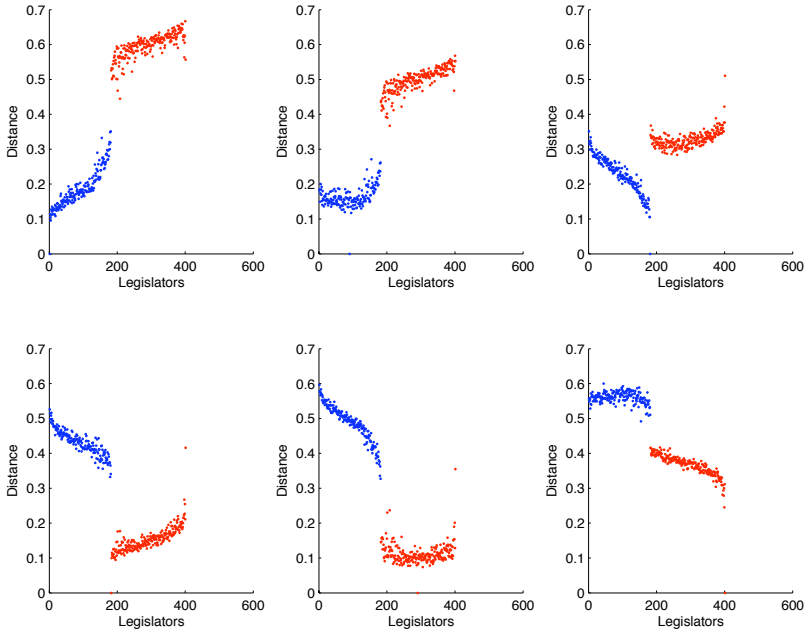

FIG 10. *The empirical roll call derived distance function* $\hat{d}(l_i, \cdot)$ *for selected legislators* $l_i = 1, 90, 181, 182, 290, 401$. *The x-axis orders legislators according to their MDS rank.*

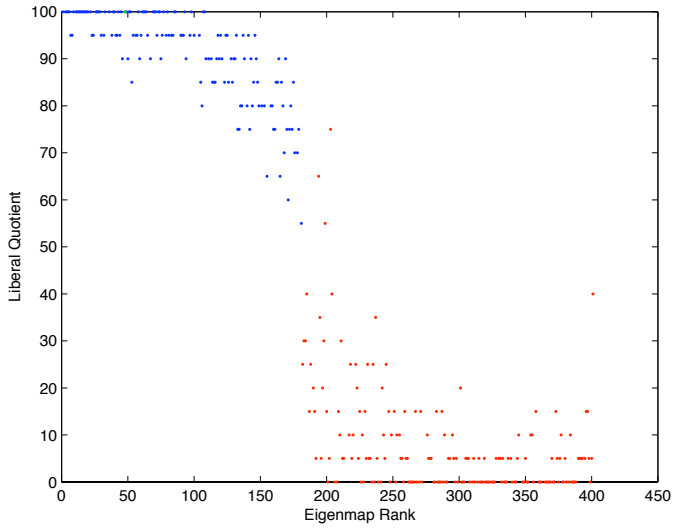

FIG 11. *Comparison of the MDS derived rank for Representatives with the Liberal Quotient as defined by Americans for Democratic Action.*

consequently separated by the algorithm. In this case, distance is directly related to the propensity of legislators to vote the same on any given bill. Figure 11 results because this notion of proximity, although related, does not correspond directly to political ideology. The MDS and ADA rankings complement one another in the sense that together they facilitate identification of two distinct, yet relatively liberal groups of Republicans. That is, although these two groups are relatively liberal, they do not share the same political positions.

Like ADA, the National Journal ranks Representatives each year based on their voting record. In 2005, The Journal chose 41 votes on economic issues, 42 on social issues and 24 dealing with foreign policy. Based on these 107 votes, legislators were assigned a rating between 0 and 100 – lower numbers indicate a more liberal political ideology. Figure 12 is a plot of the National Journal vs. MDS rankings, and shows results similar to the ADA comparison. As in the ADA case, we see that relatively liberal Republicans receive quite different MDS ranks. Interestingly, this phenomenon does not appear for Democrats under either the ADA or the National Journal ranking system.

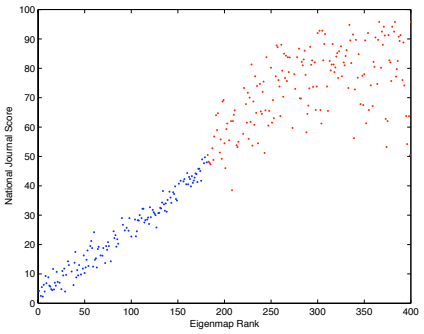

FIG 12. *Comparison of the eigendecomposition derived rank for Representatives with the National Journal's liberal score*

Acknowledgments. We thank Harold Widom, Richard Montgomery, Beresford Parlett and Doug Rivers for bibliographical pointers and helpful conversations. Cajo Ter Braak did a wonderful job educating us as well as pointing out typos and mistakes in an earlier draft.

References.

[1] A. Albouy. Mutual distances in celestial mechanics. Lectures at Nankai Institute, Tianjin, China, 2004.

[2] Americans for Democratic Action. ADA Congressional voting record - U.S. House of Representatives 2005. http://www.adaction.org.

[3] E. Bogomolny, O. Bohigas, and C. Schmit. Spectral properties of distance matrices. *J. Phys. A: Math. Gen.*, 36:3595–3616, 2003. URL http://www.citebase.org/abstract?id=oai:arXiv.org:nlin/0301044.

[4] C.W. Borchardt. Ueber die aufgabe des maximum, welche der bestimmung des tetraeders von grösstem volumen bei gegebenem flächeninhalt der seitenflächen für mehr als drei dimensionen entspricht. *Mathematische Abhandlungen der Akademie der Wissenschaften zu Berlin*, pages 121–155, 1866.

[5] I. Borg and P. Groenen. *Modern multidimensional scaling : theory and applications*. Springer-Verlag, New York, Berlin, 1997.

[6] B. C. Burden, G. A. Caldeira, and T. Groseclose. Measuring the ideologies of U. S. senators: The song remains the same. *Legislative Studies Quarterly*, 25(2):237–258, 2000.

[7] A. Cantoni and P. Butler. Eigenvalues and eigenvectors of symmetric centrosymmetrlc matrices. *Linear Algebra and its Applications*, 13:275–288, 1976.

[8] J. Clinton, S. Jackman, and D. Rivers. The statistical analysis of roll call data. *American Political Science Review*, pages 355–370, 2004.

[9] C. H. Coombs. *A Theory of Data*. Wiley, NY., 1964.

[10] T. F. Cox and M. A. A. Cox. *Multidimensional Scaling*. Chapman & Hall, 2000.

[11] Jan de Leeuw. Multidimensional unfolding. In *Encyclopedia of Statistics in Behavioral Science*. Wiley, 2005.

[12] I. J. Good. The inverse of a centrosymmetric matrix. *Technometrics*, 12:925–928, 1970.

[13] J. J. Heckman and J. M. Snyder. Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators. *RAND Journal of Economics*, 28:S142–89., 1997.

[14] M.O. Hill and H.G. Gauch. Detrended correspondence analysis, an improved ordination technique. *Vegetatio*, 42:47–58, 1980.

[15] S. Iwatsubo. *Data analysis and informatics.*, chapter The analytical solutions of an eigenvalue problem in the case of applying optimal scoring method to some types of data., pages 31–40. North Holland, Amsterdam, The Netherlands., 1984.

[16] P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003. . URL http://www.mitpressjournals.org/doi/abs/10.1162/089976603321780317.

[17] Office of the Clerk - U.S. House of Representatives. U.S. House of Representatives roll call votes 109th Congress - 1st session (2005). http://clerk.house.gov.

[18] B. N. Parlett. *The Symmetric Eigenvalue Problem*. Prentice Hall, 1980.

[19] J. Podani and I. Miklos. Resemblance coefficients and the horseshoe effect in principal coordinates analysis. *Ecology*, pages 3331–3343, 2002.

[20] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, pages 2323 – 2326, 2000.

[21] I.J. Schoenberg. Remarks to Maurice Frechet's article "Sur la definition axiomatique d'une classe d'espace distances vectoriellement applicable sur l'espace de Hilbert. *The Annals of Mathematics*, 36(3):724–732, July 1935.

[22] B. Schölkopf, A. Smola, and Klaus-Robert Muller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319, 1998. . URL http://www.mitpressjournals.org/doi/abs/10.1162/089976698300017467.

[23] R.N. Shepard. The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika*, 27(2):125–140, 1962.

[24] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, pages 2319 – 2323, 2000.

[25] C. J. F. ter Braak. *Data analysis in community and landscape ecology.*, chapter Ordination, pages 81–173. Center for Agricultural Publishing and Documentation, Wageningen, The Netherlands., 1987.

[26] W. S. Torgerson. Multidimensional scaling: I. theory and method. *Psychometrika*, 17:401–419, 1952.

[27] U. von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering. *Annals of Statistics*, page to appear, 2007. To appear in Annals of Statistics.

[28] J. R. Weaver. Centrosymmetric (cross-symmetric) matrices, their basic properties, eigenvalues, and eigenvectors. *The American Mathematical Monthly*, 92:711–717, 1985.

[29] C. K. Williams. On a connection between kernel PCA and metric multidimensional scaling. In *"NIPS"*, pages 675–681, 2000.

[30] G. Young and A.S. Householder. Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 3:19–22, 1938.

Appendix: Proofs.

Proof of Lemma 3.1.

PROOF. The lemma follows from a straightforward integration. First split the integral into two pieces:

$$\int_0^1 e^{-|x-c|}\cos[a(x-1/2)]dx$$
$$= \int_0^c e^{x-c}\cos[a(x-1/2)]dx + \int_c^1 e^{c-x}\cos[a(x-1/2)]dx.$$

By integration by parts applied twice,

$$\int e^{x-c}\cos[a(x-1/2)]dx = \frac{ae^{x-c}\sin(a(x-1/2)) + e^{x-c}\cos(a(x-1/2))}{1+a^2}$$

and

$$\int e^{c-x}\cos[a(x-1/2)]dx = \frac{ae^{c-x}\sin(a(x-1/2)) - e^{c-x}\cos(a(x-1/2))}{1+a^2}.$$

Evaluating these expressions at the appropriate limits of integration gives the first statement of the lemma. The computation of $\int_0^1 e^{-|x-c|}\sin[a(x-1/2)]dx$ is analogous, and so is omitted here. □

Proof of Theorem 3.2.

PROOF. First note that both classes of functions in the statement of the theorem satisfy $\int_0^1 f(x)dx = 0$. Consequently, the integral simplifies to

$$\int_0^1 K(x,y)f(y)dy = \frac{1}{2}\int_0^1 \left(e^{-|x-y|} + e^{-y} + e^{-(1-y)}\right) f(y)dy.$$

Furthermore, since $e^{-y}+e^{-(1-y)}$ is symmetric about $1/2$ and $\sin(a(y-1/2))$ is skew-symmetric about $1/2$, Lemma 3.1 shows that

$$\begin{aligned}\int_0^1 K(x,y)\sin(a(y-1/2))dy &= \frac{1}{2}\int_0^1 e^{-|x-y|}\sin(a(y-1/2))dy \\ &= \frac{\sin[a(c-1/2)]}{1+a^2} + \frac{(e^{-c}-e^{c-1})\,(a\cos(a/2)+\sin(a/2))}{2(1+a^2)}.\end{aligned}$$

This establishes the first statement of the theorem. We examine the second. Since $\int_0^1 K(x,y)dy = 0$

$$\int_0^1 \left(e^{-|x-y|} + e^{-y} + e^{-(1-y)}\right) dy = \left(4 - 2e^{-1} - e^{-x} - e^{-(1-x)}\right)$$

and also, by straightforward integration by parts

$$\begin{aligned}\int_0^1 e^{-y}\cos(a(y-1/2))dy &= \int_0^1 e^{-(1-y)}\cos(a(y-1/2))dy \\ &= \frac{a\sin(a/2)(1+e^{-1})}{1+a^2} + \frac{\cos(a/2)(1-e^{-1})}{1+a^2}.\end{aligned}$$

Using the result of Lemma 3.1 we have

$$\begin{aligned}&\frac{1}{2}\int_0^1 \left[e^{-|x-y|}+e^{-y}+e^{-(1-y)}\right]\left[\cos(a(y-1/2)) - \frac{2}{a}\sin(a/2)\right]dy \\ &\qquad = \frac{\cos[a(x-1/2)]}{1+a^2} + \frac{\left(e^{-x}+e^{x-1}\right)\left(a\sin(a/2)-\cos(a/2)\right)}{2(1+a^2)} \\ &\qquad\quad + \frac{a\sin(a/2)(1+e^{-1})}{1+a^2} + \frac{\cos(a/2)(1-e^{-1})}{1+a^2} \\ &\qquad\quad - \frac{1}{a}\sin(a/2)\left(4-2e^{-1}-e^{-x}-e^{-(1-x)}\right) \\ &\qquad = \frac{\cos[a(x-1/2)]}{1+a^2} - \frac{2\sin(a/2)}{a(1+a^2)} + \frac{\phi(x)}{a(1+a^2)}\end{aligned}$$

where

$$\begin{aligned}\phi(x) &= 2\sin(a/2) + a\left(e^{-x}+e^{x-1}\right)\left(a\sin(a/2)-\cos(a/2)\right)/2 \\ &+ a^2\sin(a/2)(1+e^{-1}) + a\cos(a/2)(1-e^{-1}) \\ &- (1+a^2)\sin(a/2)\left(4-2e^{-1}-e^{-x}-e^{-(1-x)}\right).\end{aligned}$$

The result follows by grouping the terms of $\phi(x)$ so that we see

$$\begin{aligned}\phi(x) &= \left[2-4+2e^{-1}+e^{-x}+e^{-(1-x)}\right]\sin(a/2) \\ &+ \left[e^{-x}/2+e^{x-1}/2+1+e^{-1}-4+2e^{-1}+e^{-x}+e^{-(1-x)}\right]a^2\sin(a/2) \\ &+ \left[-e^{-x}/2-e^{x-1}/2+1-e^{-1}\right]a\cos(a/2) \\ &= \left[-e^{-x}/2-e^{x-1}/2+1-e^{-1}\right]\left[a\cos(a/2)-2\sin(a/2)-3a^2\sin(a/2)\right].\end{aligned}$$

□

Proof of Lemma 3.2.

Proof. Let $f(\theta) = \tan(\theta/2) - \theta/(2+3\theta^2)$. Then f is an odd function, so a is a solution to $f(\theta)=0$ if and only if $-a$ is a solution. Now,

$$f'(\theta) = \frac{1}{2}\sec^2(\theta/2) + \frac{3\theta^2-2}{(3\theta^2+2)^2}$$

and so $f(\theta)$ is increasing for $\theta \geq \sqrt{2/3}$. Recall the power series expansion of $\tan\theta$ for $|\theta| < \pi/2$ is

$$\tan\theta = \theta + \theta^3/3 + 2\theta^5/15 + 17\theta^7/315 + \dots.$$

In particular, for $0 \leq \theta < \pi/2$, $\tan\theta \geq \theta$. Consequently, for $\theta \in (0, \pi/2)$,

$$f(\theta) \geq \frac{\theta}{2} - \frac{\theta}{2+3\theta^2} > 0.$$

So f has no roots in $(0, \pi/2)$, and is increasing in the domain in which we are interested. Furthermore, for $k \geq 1$

$$f(2k\pi) < 0 < +\infty = \lim_{\theta \to (2k+1)\pi^-} f(\theta).$$

The third and fourth quadrants have no solutions since $f(x) < 0$ in those regions. This shows that the solutions to $f(\theta) = 0$ lie in the intervals

$$\bigcup_{k=1}^{\infty} (2k\pi, 2k\pi + \pi)$$

with exactly one solution per interval. Finally, for $k \in \mathbb{Z}_{\geq 1}$

$$\begin{aligned} f(2k\pi + 1/(3k\pi)) &\geq \tan(k\pi + 1/(6k\pi)) - \frac{1}{6k\pi} \\ &= \tan(1/(6k\pi)) - \frac{1}{6k\pi} \\ &\geq 0 \end{aligned}$$

which gives the result.

To prove the second statement of the lemma, set $g(\theta) = \theta\cot(\theta/2)$. Then g is even, so $g(a) = -1$ if and only if $g(-a) = -1$. Since $g'(\theta) = \cot(\theta/2) - (\theta/2)\csc^2(\theta/2)$, $g(\theta)$ is negative and decreasing in third and fourth quadrants (assuming $\theta \geq 0$) and furthermore,

$$g((2k+1)\pi) = 0 > -1 > -\infty = \lim_{\theta \to 2(k+1)\pi^-} g(\theta).$$

The first and second quadrants have no solutions since $g(x) \geq 0$ in those regions. This shows that the solutions to $g(x) = -1$ lie in the intervals

$$\bigcup_{k=0}^{\infty} ((2k+1)\pi, (2k+1)\pi + \pi)$$

with exactly one solution per interval. Finally, for $k \in \mathbb{Z}_{\geq 0}$

$$\begin{aligned} & g((2k+1)\pi + 1/(k\pi+\pi/2)) \\ & \quad = ((2k+1)\pi + 1/(k\pi+\pi/2))\cot(k\pi+\pi/2+1/(2k\pi+\pi)) \\ & \quad = ((2k+1)\pi + 1/(k\pi+\pi/2))\cot(k\pi+\pi/2+1/(2k\pi+\pi)) \\ & \quad = ((2k+1)\pi + 1/(k\pi+\pi/2))\cot(\pi/2+1/(2k\pi+\pi)) \\ & \quad = -((2k+1)\pi + 1/(k\pi+\pi/2))\tan(1/(2k\pi+\pi)) \\ & \quad < -1 \end{aligned}$$

which completes the proof. □

Proof of Theorem 3.4.

Proof. Let $\tilde{f}_{n,a} = f_{n,a}/\|f_{n,a}\|_2$. Then by Theorem 3.3

$$\left| K_n \tilde{f}_{n,a}(x_i) - \frac{1}{1+a^2}\tilde{f}_{n,a}(x_i) \right| \leq \frac{a+4}{2n\|f_{n,a}\|_2}$$

and consequently

$$\left\| K_n \tilde{f}_{n,a}(x_i) - \frac{1}{1+a^2}\tilde{f}_{n,a}(x_i) \right\|_2 \leq \frac{a+4}{2\sqrt{n}\|f_{n,a}\|_2}.$$

By Lemma 3.2, a lies in one of the intervals $(2k\pi, 2k\pi + 1/3k\pi)$ for $k \geq 1$. Then

$$\begin{aligned} |f_{n,a}(x_n)| &= |\cos(a/2) - (2/a)\sin(a/2)| \\ &\geq \cos(1/6\pi) - 1/\pi \\ &\geq 1/2. \end{aligned}$$

Consequently,

$$\|f_{n,a}\|_2 \geq |f_{n,a}(x_n)| \geq 1/2$$

and so the first statement of the result follows from Theorem 3.1. The second statement is analogous. □

Susan Holmes and Persi Diaconis
Department of Statistics
Sequoia Hall
CA 94305 Stanford, USA.
URL: http://www-stat.stanford.edu/ susan/
susan@stat.stanford.edu

Sharad Goel
Department of Mathematics
University of Southern California Denney Research Building (DRB)
Los Angeles, CA 90089, USA.
E-mail: 5harad.6oel@gmail.com
URL: http://www-rcf.usc.edu/ sharadg/