SPECTRAL ANALYSIS OF THE SUPREME COURT

BRIAN L. LAWSON, MICHAEL E. ORRISON AND DAVID T. UMINSKY

1. Introduction

Imagine a survey in which 110 people were given a set $\{A, B, C, D, E\}$ of five items, and were asked to vote for their two favorite items from the set. The data has been presented to you as vector $f \in \mathbb{R}^{10}$ where

$$f = \begin{bmatrix} 2 \\ 12 \\ 11 \\ 6 \\ 17 \\ 8 \\ 4 \\ 24 \\ 20 \\ 6 \end{bmatrix} \qquad \begin{matrix} AB \\ AC \\ AD \\ AE \\ BC \\ BD \\ BE \\ CD \\ CE \\ DE \end{matrix}$$

with the numbers corresponding to the list of unordered pairs on the right. For example, two people chose items A and B, while 24 people chose items C and D.

If your task was to "figure out what's going on with this data," then where would you start? If your first thought is to "check out a book from the library," then you will be pleased to know that there are a handful of resources available (see, for example, [1, 5]). A relatively new answer, however, starts with some simple counting statistics, and then goes on to involve an intriguing mixture of ideas and techniques from introductory courses in linear algebra, abstract algebra, numerical analysis, and graph theory. The end result, which extends some of the work in [1, 2, 3] and can be found in [6, 7], is an efficient procedure for doing exploratory data analysis for data similar to that above. For example, as we will show, it can be used to pinpoint voting coalitions in small voting bodies like the United States Supreme Court.

Before we say anything else, though, it is important to point out that the overall approach to analyzing voting data that is presented in this paper goes by the name of *generalized spectral analysis*. It was initially pioneered by Diaconis in [1, 2]. The interested reader is strongly encouraged to delve into these sources, both of which are teeming with tantalizing open questions and deep ideas.

The focus of this paper is the linear algebraic framework in which this analysis is carried out. In particular, our goal is to show how simple ideas from linear algebra can come together to say something interesting about voting. And what could be more simple than where our story begins—with counting.

2. From Counting to Orthogonal Subspaces

There are 110 people involved in the survey above. The sum of the entries in the vector f is therefore 110, so the average number of votes given to each of the ten pairs of items is simply $110/10 = 11$. In the long run, this information may or may not be useful, but it seems like a reasonable place to start. After all, if each of the entries in f had been near the average, we could summarize the data by saying that each pair seems just as likely to have been chosen.

After computing the average, our next step might be to compute the number of times an individual item, such as A or C, was chosen. This would help us to see if there was an item that was particularly popular or unpopular, regardless of with which item it was paired. Since we would want to do this for each item, we could use a matrix for this calculation:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \begin{array}{c} \begin{bmatrix} 2 \\ 12 \\ 11 \\ 6 \\ 17 \\ 8 \\ 4 \\ 24 \\ 20 \\ 6 \end{bmatrix} \end{array} \begin{array}{c} AB \\ AC \\ AD \\ AE \\ BC \\ BD \\ BE \\ CD \\ CE \\ DE \end{array} = \begin{bmatrix} 31 \\ 31 \\ 73 \\ 49 \\ 36 \end{bmatrix} \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \tag{1}$$

This matrix-vector product shows, for example, that A was chosen $2+12+11+6 = 31$ times, while C was chosen $12+17+24+20 = 73$ times. For reference, we will refer to the 5×10 matrix used in (1) as T_1. In other words,

$$T_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

But why stop with one matrix? For example, the average could have been computed using the matrix

$$T_0 = \begin{bmatrix} \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \end{bmatrix}$$

since the product of T_0 and our data vector is precisely the sum of the entries in the vector divided by 10. In fact, we could even construct the matrix T_2 that computes the number of times that each pair was chosen. Of course, this turns out to just be the 10×10 identity matrix since the data was originally defined in terms of pairs of items. Nonetheless, the matrices T_0, T_1, and T_2 seem to be just the ticket when it comes to counting. But wait, there's more!

It turns out that if we define N_i to be the nullspace of T_i, then we get the chain

$$N_0 \supset N_1 \supset N_2.$$

Of course, all of the useful counting information is actually contained in the orthogonal complements of the N_i, which leads to the "complementary" chain

$$N_0^\perp \subset N_1^\perp \subset N_2^\perp.$$

This chain of subspaces seems to makes intuitive sense. If we know how many times every pair was chosen, we can figure out the number times each individual item was chosen. This information can then be used easily to compute the average.

We will refer to the 1-dimensional subspace $N_0^\perp$ as the *mean effects* space; it contains all of the information needed to compute the average. The 5-dimensional subspace $N_1^\perp$ is the *first order effects* space; it contains all of the information needed to compute the number of times a particular item was chosen. Lastly, the 10-dimensional subspace $N_2^\perp$ is the *second order effects* space; it is simply the original vector space containing all of the information associated with pairs of items.

Given the chain $N_0^\perp \subset N_1^\perp \subset N_2^\perp$ of subspaces, we could next ask about the effect that each subspace has on our data vector f. For example, we could compute the parts of f that are contained in $N_0^\perp$, $N_1^\perp$, and $N_2^\perp$. Because these subspaces form a chain, however, it is more instructive to compute the parts of f that are introduced as we move up the chain, i.e., to compute what we need to build the vector f as we move from $N_0^\perp$ to $N_1^\perp$ to $N_2^\perp$. This gives rise to an orthogonal decomposition

$$M_0 \oplus M_1 \oplus M_2$$

of the original vector space, where $N_0^\perp = M_0$, $N_1^\perp = M_0 \oplus M_1$, and $N_2^\perp = M_0 \oplus M_1 \oplus M_2$. The vector f can therefore be written uniquely as a sum $f = f_0 + f_1 + f_2$ where $f_i \in M_i$. In our case, we have

$$f = \begin{bmatrix} 2 \\ 12 \\ 11 \\ 6 \\ 17 \\ 8 \\ 4 \\ 24 \\ 20 \\ 6 \end{bmatrix} \quad f_0 = \begin{bmatrix} 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \end{bmatrix} \quad f_1 = \begin{bmatrix} -26/3 \\ 16/3 \\ -8/3 \\ -7 \\ 16/3 \\ -8/3 \\ -7 \\ 34/3 \\ 7 \\ -1 \end{bmatrix} \quad f_2 = \begin{bmatrix} -1/3 \\ -13/3 \\ 8/3 \\ 2 \\ 2/3 \\ -1/3 \\ 0 \\ 5/3 \\ 2 \\ -4 \end{bmatrix} \quad \begin{matrix} AB \\ AC \\ AD \\ AE \\ BC \\ BD \\ BE \\ CD \\ CE \\ DE \end{matrix}$$

The 1-dimensional space M_0 is still just the mean effects space. The 4-dimensional subspace M_1, however, can be thought of as the space of *pure* first order effects, since we have removed the mean effects contained in $N_0^\perp$. Likewise, the 5-dimensional subspace M_2 can be thought of as the space of *pure* second order effects, since we have removed the mean and first order effects from $N_2^\perp$.

Now that we have isolated f_0, f_1, and f_2, the next step might be to explore more deeply the way in which these vectors contribute to the data vector f. We could begin by comparing the squared norms of the f_i. One reason for doing this lies in the fact that, since the M_i are orthogonal to each other, we know that

$$||f||^2 = ||f_0||^2 + ||f_1||^2 + ||f_2||^2.$$

By comparing the squared norms, we can therefore get a sense for where the data is concentrated. In our case, $||f_0||^2 = 1210$, $||f_1||^2 \approx 422.67$, and $||f_2||^2 \approx 53.33$. Now the norm of f_0 captures nothing more than the number of people voting. The relatively large size of f_1, however, suggests that the first order effects are contributing heavily to this data.

Before we attempt to pinpoint which item or items might actually be contributing to f_1, notice that even though $\dim M_1 = 4$, there are five natural effects to consider, namely the individual effects of each of the five items. In other words, there are too many items to just find a basis vector in M_1 for each and to then write f_1 in terms of that basis. As noted in [1], however, there is a straightforward way around this that makes use of inner products.

For each item x, consider the function g_x which is defined on the pairs of items, and whose value at a pair is 1 if x is in the pair, and 0 otherwise. Project each of these functions into M_1, normalize the projection, then compute their inner products with a normalized version of the projection f_1. The resulting numbers, all of which are between -1 and 1, then allow us to try to interpret the projection f_1. (See [1, 2] for examples and more details.)

For our data vector f, this approach leads to the numbers

A	B	C	D	E
-0.41	-0.41	0.91	0.16	-0.25

which suggest that the respondents in the survey really liked item C but were slightly adverse to choosing items A and B. Indeed, a quick glance back at the original data confirms this.

We could also compute similar inner products for the pure second order effects. Here the natural functions to consider correspond to the original pairs, with a 1 in just one position and zeros elsewhere. The resulting numbers

AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
-0.06	-0.84	0.52	0.39	0.13	-0.06	0.00	0.32	0.39	-0.77

suggest, for example, that the 12 votes for the pair $\{A, C\}$ are due mostly to C's popularity, not the popularity of the pair.

Now for such a small data set, it may seem as though we went to a lot of trouble to end up only saying that "people seem to really like item C." In fact, you may have already come to that conclusion when you first saw the data, or after you saw how many people chose a pair containing C. The point, of course, is that this approach applies to any survey in which people are asked to choose their top k items from a list of size n. In fact, if we assume that $0 \leq k \leq n/2$ (we will ask them to choose their least favorites if we must), then we get an orthogonal decomposition

$$M = M_0 \oplus M_1 \oplus \cdots \oplus M_{k-1} \oplus M_k$$

where M is the underlying $\binom{n}{k}$-dimensional vector space of real-valued functions defined on the k-sets of an n-set. The subspace M_i captures the *pure i-th order effects* of the voting data. The projections of a data vector $f \in M$ into each of the M_i can be computed with the hope of uncovering hidden large-scale structure. Subsequent inner product calculations can then lead to the uncovering of hidden small-scale structure.

As we will see in the next section, this simple approach to untangling survey data can also be applied to voting data that arises when committees vote "yea" or "nay" on several issues. The trick, perhaps to the delight of committee members everywhere, is to let the issues do all of the voting!

3. From Surveys to the Supreme Court

When the members of a committee are asked to vote "yea" or "nay" on an issue, and none of the members abstain from the vote, the result is a splitting of the committee into two groups—the winners (or majority) and the losers (or minority). Of course, once we know either group, we automatically know the other, so for convenience, we will focus on the minority members for each issue.

Now although the committee members are really doing all of the voting, we can turn the tables by pretending as though the issues are actually voting on the subset of members that it wants to make up the minority when it comes before the committee. In this way, we can use the issues and their "votes" to try to pinpoint coalitions in the committee. Moreover, to analyze the resulting data, we simply separate it into different functions, one for each of the possible number of members in the minority. We then analyze each of the functions using the techniques described in the earlier sections. (See [4] for more details.)

As a proof of concept, consider the well-studied nine member "committee" of justices on the United States Supreme Court, say from 1994 to 1998. For each case (issue) in which there are no abstentions, there can be zero, one, two, three, or four justices that form the minority. We limit our analysis to non-unanimous cases, and for the Supreme Court buffs out there, we have also limited our analysis to the cases in which a signed opinion was issued. Our data comes from the database maintained by Spaeth [9], and the results of our analysis are summarized in Table 1.

Table 1. Rehnquist Court 1994-1998, 192 non-unanimous cases

split	subspace	norm^2	four largest (using absolute value) inner products							
8-1	M_1	703	Sv	.996	By	-.165	Gi	-.165	Sc	-.125
7-2	M_2	183	ThSc	.732	SvGi	.476	SvBy	.354	SvSc	-.330
7-2	M_1	59	Sv	.695	O'	-.452	So	-.348	Ke	-.348
6-3	M_3	72	ReThSc	.647	O'ThSc	.345	ByThSc	-.321	ReO'Sc	-.290
6-3	M_2	105	ThSc	.626	SvGi	.345	GiTh	-.309	ReTh	.237
6-3	M_1	16	Th	.656	Sc	.540	Ke	-.501	By	-.270
5-4	M_4	316	SvGiBySo	.954	KeReThSc	.344	O'ReThSc	.341	SvBySoSc	-.269
5-4	M_3	199	SvBySo	.379	SvGiSo	.368	GiBySo	.293	SvGiBy	.293
5-4	M_2	360	SvSo	.315	SvBy	.301	ThSc	.282	SvGi	.265
5-4	M_1	22	Ke	-.646	So	.418	O'	-.380	Gi	.380

Sv Stevens Gi Ginsburg By Breyer So Souter Ke Kennedy
O' O'Connor Re Rehnquist Th Thomas Sc Scalia

As we will see, after looking at the results in the table, it is easy to go back and find the information that supports it. We want to stress, however, that finding the most important coalitions in a committee would be labor intensive and unsystematic without something like generalized spectral analysis. This would be especially true if you were starting with raw data and knew essentially nothing about the committee members. Moreover, as described in [6, 7], by using a combination of ideas and techniques from introductory courses in linear algebra, abstract algebra, numerical analysis, and graph theory, the results presented in Table 1 can actually be computed within seconds!

Cases with 8-1 splits. There were 37 cases which split 8-1. The lone dissenter was Stevens 29 times, Thomas wrote 3 such dissents, neither Ginsburg nor Breyer ever dissented on their own, and the remaining five justices each wrote one lone dissent. Not surprisingly, the first line of the table show that the 8-1 data points strongly in the direction of Stevens dissenting.

Cases with 7-2 splits. There are first and second order effects for the cases which split 7-2. In this case, the squared norm of the projection onto M_2 is 183, while the squared norm of the projection onto M_1 is 59. This portion of the data is therefore dominated by the pure second order effect.

The pure second order effect points in the direction of Thomas-Scalia dissenting, but also has noticeable components in the Stevens-Ginsburg and Stevens-Breyer directions. The pure first order effect, although not as strong as the pure second order effect, points in the direction of Stevens dissenting.

In this portion of the data, there are 48 cases with 7-2 splits, where 11 of them are Thomas-Scalia, 9 are Stevens-Ginsburg and 8 are Stevens-Breyer dissenting. Of the 48 cases with pairs dissenting, Stevens is a dissenter in 24 of them. In fact, Steven dissents with everyone except Rehnquist, Scalia, and O'Connor.

Cases with 6-3 splits. For the 6-3 cases, the second order effect is the largest with a squared norm of 105. The largest inner product for the pure second order effects corresponds to Thomas-Scalia dissenting. Moreover, the other projections for the 6-3 cases also point in this general direction.

The largest pure first order effect is Thomas dissenting, and the next largest is Scalia dissenting. The two largest third order effects are Thomas-Scalia dissenting joined first by Rehnquist, and then by O'Connor. The negative sign on the Bryer-Thomas-Scalia triple suggests that these three justices seldom dissent together in a 6-3 split.

Again, when we examine the data we find that this is a good summary. Of the 44 cases with a 6-3 split, Thomas and Scalia join together in dissent in 21 of them. They are joined ten times by Rehnquist, seven times by O'Connor, and at least once by each remaining justice with the exception of Ginsburg and Breyer.

Cases with 5-4 splits. The pure second order effect is the largest for the 5-4 cases with a squared norm of 360. The five largest inner products for the pure second order effect are relatively close to each other. They correspond to "liberal" dissents (Stevens-Souter, Stevens-Breyer, and Stevens-Ginsburg) and "conservative" dissents (Thomas-Scalia and Rehnquist-Scalia, the latter of which is not in the table).

The pure fourth order effect is the next largest with a squared norm of 316. The pure fourth order effect is in the "liberal" dissenting direction (Stevens-Ginsberg-Breyer-Souter) with smaller components in the "conservative" dissenting directions (Rehnquist-Thomas-Scalia-Kennedy and Rehnquist-Thomas-Scalia-O'Connor). This seems to fit the data well. For the 63 cases with 5-4 splits, 28 cases are dissents by Stevens-Ginsberg-Breyer-Souter, eight are Rehnquist-Thomas-Scalia-O'Connor, and seven are Rehnquist-Thomas-Scalia-Kennedy.

Now although the pure first order effects are the smallest, they have an interesting interpretation. The "swing" voter Kennedy has a large significant negative value (-.646) suggesting that he rarely ends up in the minority in 5-4 splits, which is what would be expected of swing voters. The case for O'Connor, while smaller, is similar.

4. Conclusion

We hope we have convinced you that simple ideas from linear algebra can come together to say something worthwhile about voting data. Generalized spectral analysis is a powerful tool for doing exploratory data analysis, and given that efficient algorithms for doing this type of analysis exist (see, for example, [6, 7, 8]), political scientists, economists, and market research analysts seem to have every incentive to include it in their arsenal.

5. Acknowledgments

We gratefully acknowledge a Harvey Mudd College Beckman Research Award which supports research by undergraduates with faculty. Special thanks also to Francis Su for comments on a preliminary version of this paper.

References

1. P. Diaconis, *Group representations in probability and statistics*, Institute of Mathematical Statistics, Hayward, CA, 1988.
2. ———, *A generalization of spectral analysis with application to ranked data*, Ann. Statist. **17** (1989), no. 3, 949–979.
3. P. Diaconis and D. Rockmore, *Efficient computation of isotypic projections for the symmetric group*, Groups and computation (New Brunswick, NJ, 1991), Amer. Math. Soc., Providence, RI, 1993, pp. 87–104.
4. M. Iwasaki, *Spectral analysis of multivariate binary data*, J. Japan Statist. Soc. **22** (1992), no. 1, 45–65.
5. J. Marden, *Analyzing and modeling rank data*, Monographs on Statistics and Applied Probability, vol. 64, Chapman & Hall, London, 1995.
6. D. Maslen, M. Orrison, and D. Rockmore, *Computing isotypic projections with the Lanczos iteration*, SIAM Journal on Matrix Analysis and Applications **25** (2004), no. 3, 784–803.
7. M. Orrison, *An eigenspace approach to decomposing representations of finite groups*, Ph.D. thesis, Dartmouth College, 2001.
8. D. Rockmore, *Some applications of generalized FFTs*, Groups and computation, II (New Brunswick, NJ, 1995), Amer. Math. Soc., Providence, RI, 1997, pp. 329–369.
9. H. Spaeth, UNITED STATES SUPREME COURT JUDICIAL DATABASE, 1953-1997 TERMS [Computer file]. 9th ICPSR version. East Lansing, MI: Michigan State University, Dept. of Political Science [producer], 1998. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1999.

Department of Political Science, University of Cincinnati, Cincinnati, OH 45221
E-mail address: `brian.lawson@uc.edu`

Department of Mathematics, Harvey Mudd College, Claremont, CA 91711
E-mail address: `orrison@hmc.edu`

Department of Mathematics and Statistics, Boston University, Boston, MA 02215
E-mail address: `duminsky@math.bu.edu`