

THE DIMENSION OF THE SUPREME COURT

*Paul H. Edelman**

It is a rare occurrence when the *New York Times*,¹ *Washington Post*,² NPR,³ and even Jack Kilpatrick⁴ discuss a political science paper. Nonetheless, that is what happened after *A Pattern of Analysis of the Second Rehnquist US Supreme Court*, by Lawrence Sirovich,⁵ was published in the *Proceedings of the National Academy of Science* in June 2003. Sirovich's paper applies two unusual mathematical techniques to the decisions of the Court with the aim of "extracting key patterns and latent information."⁶ Using information theory, and in particular the idea of entropy, Sirovich claims that the "Court acts as if composed of 4.68 ideal Justices."⁷ After applying a singular value decomposition to the decision data he concludes that the Court's decisions can be accurately approximated by a suitably chosen two-dimensional space.

While some commentary has questioned whether Sirovich's conclusions are novel, at least one of the methods of analysis is new (in the context of political science) and might also prove useful in other circumstances. Moreover the methods themselves raise interesting questions about the Court. It is therefore worthwhile to consider the methods more carefully.

* Professor of Mathematics and Law, Vanderbilt University. I thank Erin O'Hara, Bob Rasmussen, Suzanna Sherry, Stephen Siegel, and an editor of this fine journal for their assistance in making this commentary accessible to a legal audience. To the extent that their efforts were in vain the blame falls solely on yours truly.

1. Nicholas Wade, *A Mathematician Crunches the Supreme Court's Numbers*, N.Y. TIMES, June 24, 2003, at F3.

2. Guy Gugliotta, *Supreme Court Independence, by the Numbers*, WASH. POST, July 28, 2003, at A7.

3. Weekend Edition Saturday, NPR, June 28, 2003 (transcript from Lexis).

4. Jack Kilpatrick, *Justices Don't Fit into Predictable Ideological Boxes*, AUGUSTA CHRON. (Ga.), Aug. 3, 2003, at A4.

5. Lawrence Sirovich, *A Pattern Analysis of the Second Rehnquist US Supreme Court*, 100 PROC. NAT'L ACAD. SCI. 7432 (2003).

6. *Id.* at 7432.

7. *Id.*

Before discussing the methods themselves, we need to explore how Sirovich encodes data from the Court. He starts by ordering the Justices in alphabetical order (although any order would work) and then encodes each decision by a vector with nine entries in which a 1 signifies a Justice who was in the majority and a -1 signifies a Justice in the minority. For example, a case decided unanimously is coded (1,1,1,1,1,1,1,1,1) and a case decided by the classic 5-4 conservative-liberal split (say *Garrett*⁸) is coded (-1,-1,1,1,1,1,-1,-1,1) where the first -1 indicates that Breyer (the alphabetically first Justice) was in the minority and the last 1 indicates that Thomas (the alphabetically last Justice) was in the majority.⁹ Thus, Sirovich reduces each case to a string of 1 and -1's of length 9. I will refer to these codings as vote-patterns.

There are two things worth noting about Sirovich's data set. First, he records the *decisions* of the Court and not the opinions. For instance, *Lawrence v. Texas*¹⁰ is recorded as (1,1,1,1,-1,-1,1,-1) with O'Connor listed in the majority even though she did not join the majority opinion. The second fact worth noting is that Sirovich discarded 30% of the cases because "the vote was incomplete or ambiguous (*per curiam* . . . decisions furnished no details of the vote and were deemed inadmissible, as were cases in which a Justice was absent or voted differently on the parts of a case)."¹¹ Later I will reexamine his decision to exclude these cases.

ENTROPY

The most original part of Sirovich's paper is his use of information theory, and in particular the idea of entropy, to analyze the Supreme Court. Sirovich uses information theory to measure the variability of the set of vote patterns of the Court. While others have discussed the distribution of decisions from the Court,¹² and the correlation of votes among the Justices¹³ no

8. Board of Trustees of the Univ. of Alabama v. Garrett, 531 U.S. 356 (2001).

9. The alphabetical order of the Justices is Breyer, Ginsburg, Kennedy, O'Connor, Rehnquist, Scalia, Souter, Stevens, and Thomas.

10. 123 S. Ct. 2472 (2003).

11. Sirovich, *supra* note 5, at 7432.

12. See, e.g., Paul H. Edelman & Jim Chen, *The Most Dangerous Justice Rides Again: Revisiting the Power Pageant of the Justices*, 86 MINN. L. REV. 131 (2001); Paul H. Edelman & Suzanna Sherry, *All or Nothing: Explaining the Size of Supreme Court Majorities*, 78 N.C. L. REV. 1225 (2000).

13. See, e.g., *The Supreme Court, 1999 Term—Leading Cases*, 114 HARV. L. REV. 179, 402 tbl.I (2000) ("Harvard Table").

one has proposed an overall measure of the variability of decisions until now. This fact alone makes Sirovich's paper worth reading.

Entropy is a measure of the total amount of variability in a situation. Suppose there are n different possible outcomes which we list as $1, 2, \dots, n$ and that outcome j occurs with probability p_j . The entropy of this set of outcomes is defined to be

$$I = -\sum_{j=1}^n p_j \log p_j$$

where the logarithm is taken to be base 2.¹⁴ Entropy measured in these terms can be interpreted as the smallest average code word needed to convey the outcomes.¹⁵ It is infeasible to provide a complete explanation here, but a few examples should suffice to explain how it works.

First, a small example to help clarify the ideas. Suppose that when you talk to your stockbroker he recommends Buy with probability $\frac{1}{2}$, Hold with probability $\frac{1}{4}$ and Sell with probability $\frac{1}{4}$. What is the entropy of his recommendations? Applying the formula above yields

$$I = -\frac{1}{2} \times \log \frac{1}{2} - \frac{1}{4} \times \log \frac{1}{4} - \frac{1}{4} \times \log \frac{1}{4} = -\frac{1}{2} \times (-1) - \frac{1}{4} \times (-2) - \frac{1}{4} \times (-2) = \frac{3}{2}.$$

Remember that the log function here is base 2, so $\log(2^n) = n$. What is the significance of the value $3/2$? Suppose your stockbroker had to communicate by Morse Code with you and he wanted to use, on average, as short a set of symbols as possible. If he encoded Buy with a dot, Sell with a dash-dot and Hold with a dash-dash, the expected length of his recommendation signal would be $3/2$. This is because Buy is coded with a single symbol and occurs half the time, and Hold and Sell each are coded with two symbols and occur a quarter of the time. The expected

14. The log base 2 of a number x is the number y with the property that 2 raised to the y^{th} power is x . In other words, $y = \log_2 x$ means that $2^y = x$. So $\log_2 8 = 3$ since $2^3 = 8$. Information theory commonly uses logarithms base 2 because information is typically encoded into binary signals. Most calculators do not provide a key for computing the log base 2 of a number, but it can be computed using the fact that the log base 2 of a number x is $\ln x / \ln 2$ where $\ln x$ is the natural log of x . Most calculators provide a key for computing the natural log.

15. There is a huge amount of mathematics stuffed in this sentence. Entropy in this context is a part of information theory and concerns itself with the inherent amount of information contained in a set of outcomes. One measure of that amount of information is how efficiently one can code it and transmit it to a third party. For a more detailed technical exposition, see STEVEN ROMAN, *CODING AND INFORMATION THEORY*, 39-68 (1992).

length therefore is $\frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{4} \times 2 = 3/2$. One can show that this is the best possible way to encode this information.¹⁶

Let us examine Sirovich's examples. Sirovich calls an "Omniscient Court" one for which every decision is unanimous. For such a court, there is only one outcome, which occurs with probability 1. By the above formula, $I=0$. This accords with our intuition since we assume that all the opinions will come down the same way and hence we get no new information from seeing one.

It is worth noting, though, that unanimity has nothing to do with this analysis. What matters is that every opinion is the same. If every case were to be decided by the canonical conservative versus liberal 5-4 margin, the entropy would still be 0. Omniscience, in and of itself, has nothing to do with the amount of entropy.

The other extreme case, as proposed by Sirovich, is the "platonic" Court. In this instance, he assumes that each Justice is equally likely to vote for one side as the other, which is to say "the vote of a platonic justice is as predictable as the toss of a fair coin."¹⁷ In this situation there are $2^8=256$ different possible outcomes and all of them are equally likely. To see this, note that the total number of strings of length nine where each entry is a 1 or -1 is $2^9=512$, because there are two possibilities for each coordinate. By the way we have encoded the decisions, we only consider those with more 1's than -1's (since a 1 indicates that Justice is in the majority) and exactly half of the 512 strings have that property. Applying the entropy formula leads to

$$I = -\sum_{j=1}^{256} \frac{1}{256} \log \frac{1}{256} = -\log \frac{1}{256} = -\log 2^{-8} = 8$$

for the entropy calculation. The significance of the value 8 is that, if we were clever, we could convey this information using only strings whose average length is 8 instead of 9.¹⁸

How, then, does Sirovich compute the entropy of the Court? Using data from October Term 1994 through October Term 2001, he computes the probability of a given majority coa-

16. You might think one could do better by encoding Buy as dot, Hold as dash, and Sell as dash-dot lowering the average length. This is not acceptable (by information theoretic standards, at least) because if, thereby you were to see dash-dot you would not know whether that was a single Sell recommendation or two recommendations, one a Buy and one a Hold. I told you there was a lot buried in here.

17. Sirovich, *supra* note 5, at 7433.

18. The intuition behind this is that there is one less degree of freedom in the data because, by the nature of the encoding, the number of 1's will always be larger than the number of -1's.

lition by counting the number of times it occurred and dividing by the total number of cases. Then he computes entropy using the above formula to get a value of 3.68. He concludes that the Court behaves like 4.68 "platonic" Justices. The equivalent number of "platonic" Justices is one more than the entropy because a Court with no entropy still has a judge to cast the unique decision. Alternatively, as noted earlier,¹⁹ since the judgments are encoded so that there is always a larger number of 1's than -1's, there is always an extra degree of freedom in the number of judges over the entropy.

This method of computing a global measure of uncertainty on the Court is really interesting. It captures in one number a lot of the different ideas of covariance among the Justices. Most other measures of variance consider correlation only between pairs of Justices.²⁰ I want to raise two questions about this analysis, however. The first is methodological and the second is more philosophical.

If one is truly interested in entropy on the Court, then it is hard to justify ignoring cases in which some Justice voted differently in different parts of the case. Entropy is a measure of variance in the output of the Court, and surely fractured opinions should count as contributing both to output and to variance. One might argue that such cases are not very common and can be ignored, but that is not obvious at all from the mathematics. If it were really true, then it would become evident in the entropy calculation. It makes sense to include fractured opinions and see what happens.

The other difficulty with the way Sirovich calculates entropy is that he works with the set of Justices who concur in the judgment of the Court and not with the set of Justices who join a specific opinion. As noted earlier, Sirovich codes *Lawrence v. Texas* as having been decided 6-3 with O'Connor in the majority, even though she did not join the majority opinion and concurred in the judgment on other grounds. The five in the majority struck down all sodomy laws on due process grounds; O'Connor would have struck down, on equal protection grounds, only those laws that applied solely to homosexual conduct. If one is concerned with variance in output, it is hard to justify lumping O'Connor with the rest of the majority since she disagreed both on the

19. Sirovich, *supra* note 18.

20. See Keith Poole, *The Unidimensional Supreme Court*, at http://voteview.uh.edu/the_unidimensional_supreme_court.html; Harvard Table, *supra* note 13.

source of the constitutional violation and the scope of the decision.

How then should one compute entropy? It would be better to examine all of the coalitions of Justices that formed. By coalitions I mean those collections of Justices who signed on to a common opinion. This is the method of analysis I have used earlier in analyzing voting power²¹ and I think it makes sense in this context as well. Based on data from the October Term 1994 to October Term 2000,²² I compute entropy as 5.87, which is equivalent to 6.87 "platonic" Justices. This way of computing might tend to overstate variance, since different coalitions do not always significantly differ on the law. If I were forced to choose a specific number, I would probably average this value with that derived by Sirovich.

Suppose, though, that we all agree on the appropriate way to measure entropy. This still leaves the most interesting question unresolved: What level of entropy is appropriate for the Court? Does the level of entropy make any difference to the Court?

In thinking about this question one must be careful to remember that entropy is dependent on the likelihood of distinct coalitions and not on the alignments of those coalitions as such. A court that was unanimous in every case would have the same entropy as a court that decided every case 5-4 with the same five Justices always in the majority. One would suspect that the latter arrangement would undermine political support for the Court while the former would be much better in that regard. Entropy, therefore, says little about the political stability of the Court.

Moreover, having a lot of variability implies nothing about the quality of the judgments. As Sirovich notes, nine monkeys trained to flip coins and register votes accordingly would produce decisions with maximum entropy. And it is not clear that one should really expect a court to have that much variability. Some of the Justices will almost necessarily share some legal views. Their votes will inevitably be correlated, leading to a lowering of entropy. Docket selection would also likely reduce the variability from that of completely "platonic" court.

Perhaps the most useful thing to do would be to see whether this measure of entropy correlates in some way with what is

21. Edelman & Chen, *supra* note 12.

22. Data available from me, but reported in Edelman & Chen, *supra* note 12.

known historically about the Court. Is the entropy on Courts in which the Justices were thought to be on good personal terms? Is it really just a measure of unanimity? If the unanimous decisions are factored out, do all the Courts have the same entropy? Do Courts with stable membership have more or less entropy than Courts that have had a more rapid turnover?²³ I think these questions are potentially interesting ones, and ones that would be difficult to investigate empirically without thinking about them in the context of entropy.

GEOMETRY AND THE COURT

In the second part of his paper, Sirovich uses an entirely different technique to assess the variability of the decisions of the Court. In this section he takes a geometric point of view in an attempt to measure how many degrees of freedom there are in the observed decisions. Before commenting on Sirovich's results, I want to give some elementary geometric background to make the following discussion more accessible.

Imagine that we have a court with five judges, creatively named A, B, C, D, and E. In principle, the number of different vote patterns that we might observe (using Sirovich's convention that 1 indicates the judge is in the majority and a -1 indicates a vote in the minority) is $16=2^{n-1}$ where n is 5, the number of judges on the court.²⁴ Suppose, though, that on this court we know that A and B always vote the same way and that C, D, and E always vote the same way. The only vote patterns we observe are (1,1,1,1,1) and (-1,-1,1,1,1). In this situation, all of the vote patterns lie on the straight line that goes between the two points (1,1,1,1,1) and (-1,-1,1,1,1). Even though in principle we need all five coordinates to describe the vote patterns, in practice we

23. Thomas Merrill has hypothesized that Courts with stable membership are more likely to be innovative than one with more rapid turnover. See Thomas Merrill, *The Making of the Second Rehnquist Court: A Preliminary Analysis*, 47 ST. LOUIS U. L.J. 569, 653 (2003) ("Given the closely divided nature of the Rehnquist Court and the divided government and society in which it operates, I think it is very unlikely that the Court would have generated the important changes in the law of constitutional federalism we have seen in the last eight years without the added boost from subnormal turnover."). *But cf.* Jim Chen, *Judicial Epochs in Supreme Court History: Sifting through the Record for Stitches in Time and Switches in Nine*, 47 ST. LOUIS U. L.J. 677, 734 (2003) ("[W]e cannot discern a clear causal connection, if any, between the stability of the Supreme Court's membership and the Justices' propensity to craft aggressive new doctrines.")

24. They are: (1,1,1,1,1), (1,1,1,1,-1), (1,1,1,-1,1), (1,1,-1,1,1), (1,-1,1,1,1), (-1,1,1,1,1), (1,1,1,-1,-1), (1,1,-1,1,-1), (1,1,-1,-1,1), (1,-1,1,1,-1), (1,-1,1,-1,1), (1,-1,-1,1,1), (-1,1,1,1,-1), (-1,1,1,-1,1), (-1,1,-1,1,1), (-1,-1,1,1,1) where the first coordinate corresponds to A's vote and the last to E's.

really only need to know whether the judges A and B are in the minority or not to decide what the vote pattern was. Every vote pattern observed can be described algebraically by the equation $(0,0,1,1,1) + t(1,1,0,0,0)$. If t is 1 then we have $(1,1,1,1,1)$, and if t is -1 we get $(-1,-1,1,1,1)$. In this sense the vector $(1,1,0,0,0)$ is a coordinate vector that allows us to describe the decisions, and the value of t is the distance along that coordinate.

Suppose that in 100 observations of our small court the pattern of A and B voting together and C, D, and E voting together holds, except that in one case A joins with B and C to form a majority, and in a complete surprise, in one other case B, C, and D form a majority. The first of these cases is encoded as $(1,1,1,-1,-1)$; the second as $(-1,1,1,1,-1)$.

Now there are vote patterns that do not fall on the line we described before. However, it should be intuitively clear that the line that we found is the best line that could be found—that is, among all lines it best “fits” the observed vote patterns. The measure for how well the line “fits” the vote patterns is calculated by computing for each vote pattern the square of its distance from the line, taking the average of all of these square distances, and then taking the square root.²⁵ Moreover, there would seem to be little reason to give up on our assumption that the line is still the best description of the vote patterns of the Court just because we have two outliers.

This second assertion is subject to a number of caveats. The two outliers might have been cases of particular salience, in which case we might not want to ignore them. In addition these cases might represent a new trend that foreshadows a breakdown in our observed pattern. These are important concerns. I will address the former in this section. The latter I leave for another time, except to say that assessing changes in patterns of voting is an interesting and difficult empirical issue.²⁶

Consider a somewhat different hypothetical five-judge court. Suppose that there are only three observed decisions: $(1,1,1,1,1)$, $(-1,-1,1,1,1)$ and $(1,1,1, -1,-1)$. These three decisions do not lie on any line, but they do line on a plane. By a plane I mean a set of points that can be described using two coordinate

25. This is usually referred to as the root mean square and is the standard measure of deviation in statistics. See DAVID FREEDMAN, ROBERT PISANI & ROGER PURVES, *STATISTICS* 66 (3d ed. 1998).

26. For one sophisticated approach see Andrew D. Martin & Kevin M. Quinn, *The Dimensions of Supreme Court Decision Making: Again Revisiting The Judicial Mind* (2001).

vectors, which in this case would be $(1,1,1,1,1) + t(0,0,0,1,1) + s(1,1,0,0,0)$ where if $t = s = 0$ we get the unanimity vector, if $t = -2$ and $s = 0$ we get $(1,1,1,-1,-1)$, and so forth.²⁷ As before, if a few decisions deviate from that form, this plane may still be the best approximation for the data even if it is not a perfect fit. We refer to $(0,0,0,1,1)$ and $(1,1,0,0,0)$ as coordinate vectors as before.

Let us now return to the data that Sirovich collected from voting patterns of the Supreme Court. As we just saw, it is possible that the data, although nominally requiring nine coordinates to describe, really only require a smaller number of coordinates to get a good "fit." There a number of different methods one might use to fit the data. Sirovich employs the technique of singular value decomposition. Singular value decomposition simultaneously approximates a set of data by the best line, plane, three-dimensional space, etc. It will also provide the best set of coordinates for describing the data. It is closely related to factor analysis in statistics and is now in common use in many fields including computer recognition and mathematical biology.²⁸

Sirovich's principal conclusion is that the vote patterns of the Court can be best approximated by a certain plane whose coordinate vectors are called v_1 ,²⁹ which is close to the unanimity vector $(1,1,1,1,1,1,1,1,1)$, and a vector v_2 ,³⁰ which is close to the conservative-liberal (c-l) 5-4 vector $(-1,-1,1,1,1,1,-1,-1,1)$.³¹ The unanimity and c-l vectors play a key role because they are the two most common voting patterns on the Court, accounting for almost 57% of all decisions; as a result they are weighted quite heavily in any approximation scheme.

As Keith Poole has noted,³² Sirovich's result is consistent with current work on the Court by many political scientists. This

27. What I've really described here is an affine plane since it does not go through the origin $(0,0,0,0,0)$. In Sirovich's case the approximating plane does go through the origin.

28. GILBERT STRANG, *LINEAR ALGEBRA AND ITS APPLICATIONS* 137 (3d ed. 1988).

29. The coordinates of v_1 are (.341, .333, .353, .363, .347, .313, .238, .264, .316) (to three significant digits).

30. The coordinates of v_2 are (-.327, -.368, .174, .104, .304, .403, -.313, -.446, .406) (to three significant digits).

31. The axis that is close to the unanimity vector gives a smaller value to the Stevens coordinate than to the others, presumably because Stevens is the most likely Justice to be a lone dissenter. The axis near the conservative-liberal 5-4 vector is noticeably smaller in the Kennedy and O'Connor coordinates, presumably because they are the conservatives most likely to abandon this coalition. Keith Poole has noted that the coordinates in v_2 mirror a plot of the Justices on a liberal-conservative axis. See Poole, *supra* note 20.

32. *Id.*

work employs spatial voting models, which attempt to identify each Justice with an ideal point in a policy space and then model issues as cutting planes in such a way that the set of Justices on one side of the plane vote one way and those on the other vote the opposite. The techniques used to approximate these ideal points are closely related to those employed by Sirovich. They lead to the same conclusion, that the decisions of the Court are best approximated by two dimensions. Poole avers that the better way to analyze the data is to ignore the unanimous decisions. He concludes that the Court is in fact one-dimensional.³³

What is novel about Sirovich's analysis is that he is interested in approximating the decisions rather than using the decisions to approximate the ideal points of the Justices.³⁴ The more common way to deal with spatial voting analysis is to simultaneously approximate preference points for the voters and equations for the cutting planes that define a vote. Sirovich takes the outcome of the vote as fixed and tries to approximate that data. As we will see, this particular choice of coordinates for the vote is problematic. It is surprising, actually, that his results are so close to those obtained by other methods.

Before examining his results more closely, we should recall what the criterion is for the measure of "best" approximation. The measure used to define the "best" approximation is the sum of the squares of the distances from the vote patterns to the approximating object. The best line is the line that minimizes the sum of the squares of the distances from each data point to the line. This distance is computed for each outcome of the Court so that, for example, the distance from the unanimous vote pattern to the line is added each time the Court rendered a unanimous decision. As a consequence it is not surprising the two coordinate vectors of the approximating plane are close to the two most common vote patterns.

Sirovich shows that the plane spanned by v_1 and v_2 accounts, on average, for roughly 80 percent of the data.³⁵ While this is a substantial amount, I am not sure that it can support the claim that "[t]he implication is that the decision space of the Rehnquist Court requires only two dimensions for its description."³⁶ I say

33. *Id.* Poole gives a sketch of the details of this argument on his web site.

34. For a sophisticated algorithm see KEITH T. POOLE & HOWARD ROSENTHAL, CONGRESS: A POLITICAL-ECONOMIC HISTORY OF ROLL CALL VOTING, 233-51 App. A (1997).

35. Sirovich, *supra* note 5, at 7434 tbl.3.

36. *Id.* at 7434.

this for two reasons. The first is that vote patterns of some of the most salient decisions are not well-approximated by this plane. Second, the geometry that Sirovich uses is not consistent with any legal analysis.

First, let us consider how well this plane really approximates the data. Take, for example, the decision in *Lawrence v. Texas*, in which Justice Kennedy joined the liberals to form a 5-4 majority opinion of the Court. Sirovich would code this as a 6-3 since O'Connor joined the other five as to the outcome, even though she did not join the majority opinion. This 5-4 opinion I would encode as (1,1,1,-1,-1,-1,1,1,-1). Because this vote pattern is not on the plane of approximation, we might want to ask which point on the approximating plane is closest to it, that is, how well this plane approximates this vote pattern. A short calculation³⁷ shows that the point on the approximating plane closest to (1,1,1,-1,-1,-1,1,1,-1) is (to three decimal places)

(0.924, 1.022, -0.322, -0.147, -0.652, -0.909, 0.890, 1.196, -0.915).

One curious thing about this vector, which is supposed to be a good approximation to (1,1,1,-1,-1,-1,1,1,-1), is that in Kennedy's coordinate (the third one) the approximating point is negative. Because Kennedy was in the majority the value should be 1 in that coordinate. In fact, Kennedy's coordinate is closer to being -1 than is O'Connor's (the fourth coordinate), even though she was in the minority and should therefore display a value of -1.

Indeed, the distance from the *Lawrence v. Texas* vote pattern to the approximating plane is 1.6, over half the length of the original vote pattern.³⁸ A similar calculation shows that the approximation of the vote pattern encoding a 5-4 decision consisting of O'Connor and the liberals in the majority (say, as in *Grutter*³⁹) is similarly inaccurate. Sirovich's approximation assigns the wrong sign to O'Connor.

On what, then, does Sirovich base his claim that his plane is a good approximation of the data? The criterion he uses is the ability of the approximation to estimate the margin by which the majority is carried. In a 5-4 decision the majority is carried by a margin of one; in a unanimous decision the margin is nine. It is

37. It is a standard problem in elementary linear algebra to find the point on a plane closest to some point that is not on the plane. See Strang, *supra* note 28, at 103.

38. The length of the encoding vector is 3 (as are all coding vectors since each entry is either a 1 or -1). Recall that the length of a vector is the square root of the sum of the squares of its entries. Since all the entries are either 1 or -1, and there are nine entries in total, the sum of the squares is 9 and hence the length is 3.

39. *Grutter v. Bollinger*, 123 S. Ct. 2325 (2003).

easy to see that the margin of a decision is just the sum of the coordinates of the vote pattern. For example, the c-1 vote pattern, with coordinates $(-1, -1, 1, 1, 1, 1, -1, -1, 1)$, has five 1's and four -1's, and therefore the sum of the entries is 1, the margin of victory. Sirovich tests his approximation by how the sum of the coordinates of the approximating point approximates the sum of the coordinates of the vote pattern.

As an example, consider the vote pattern for *Lawrence*. The margin of the decision is 1. The sum of the coordinates of the approximating vector is 1.087. If this is rounded to the nearest integer we get 1, the same number as the actual margin. This, for Sirovich, shows that the approximation is a good one. If one were only interested in the margin by which the majority prevails, this is a reasonable measure. But if one is interested in a more detailed assessment, in particular how an individual Justice might vote, this measure is much less informative.

Moreover, a peculiar mathematical feature of this criterion affects the accuracy of the approximation. For some vote patterns the approximating plane that Sirovich uses gives a worse approximation for the voting margin than the best approximating line! That is, this measure of fit does not improve with more dimensions in which to work. An example of this is *Lawrence*. As noted above, the sum of the coordinates of the point on the approximating plane closest to this vote pattern is 1.087. The sum of the coordinates of the approximating vector on the line of best fit is .933, which is closer to the actual margin.⁴⁰ In spite of Sirovich's comments to the contrary,⁴¹ it is difficult to see how his two-dimensional approximation optimizes the distance from the true margin of the decision.

Sirovich has quite an intriguing figure in his paper that graphically depicts how well each vote pattern of the Court is approximated by his plane.⁴² Not surprisingly the twelve most common vote patterns are well approximated (and these account for roughly 80% of all outcomes) but many vote patterns do not seem very close to their approximation at all. As noted by Sirovich in his final comments,⁴³ this geometric model is not very good at distinguishing those vote patterns that do occur from those that don't. That is, there are a number of vote patterns that

40. The line of best fit is the one obtained by taking multiples of the vector v_1 . This follows from the nature of the singular value decomposition.

41. See, e.g., Sirovich, *supra* note 5, at 7436.

42. *Id.* (fig. 1).

43. *Id.* at 7437.

are well approximated by the plane but which for some reason have never been realized, as well many vote patterns that have appeared but are poorly approximated by the plane.

Even if Sirovich's planar approximation should be considered a good fit, it is far from clear that it is an appropriate way to think about the decisions of the Court. To see this, consider the following question: Which decision is closer to the c-l split, a unanimous decision or a 5-4 decision with O'Connor and the four liberals in the majority? I think that most people concerned with the behavior of the Court would say the latter, since the only difference in the alignment is the movement of O'Connor from one side to the other, whereas a unanimous decision would require four Justices to alter their votes. But that is not the right answer according to Sirovich.

For Sirovich, the c-l split is coded $(-1,-1,1,1,1,1,-1,-1,1)$ and so the distance between it and the unanimous decision (which is coded as $(1,1,1,1,1,1,1,1,1)$) is 4.⁴⁴ The distance between the c-l split and the other 5-4 split (coded as $(1,1,-1,1,-1,-1,1,1,-1)$) is approximately 5.657.⁴⁵ What happens in this encoding is that the c-l split differs from the unanimous decision in only 4 coordinates but it differs from the other 5-4 split in eight coordinates (every one other than O'Connor).

When Sirovich produces a plane that provides a "good" approximation for the vote patterns of the Court, this approximation is with respect to the geometry of his coordinatization, and not necessarily with respect to any legal intuition as to what is close. And in fact it is on the 5-4 decisions other than the c-l split that this approximation performs at its worst, which is not that surprising given that these are the decisions for which no one has a good theoretical explanation.

CONCLUSIONS

Sirovich's paper should not be dismissed lightly as using "a more complicated method to belabor the obvious."⁴⁶ The idea of using entropy as a way of quantifying variance in the decisions of the Court is novel and interesting, although I suggest an alterna-

44. The difference of the two vectors is $(1,1,1,1,1,1,1,1,1)-(-1,-1,1,1,1,1,-1,-1,1)=(2,2,0,0,0,2,2,0)$ whose length is the square root of the sum of the square of the entries—that is, the square root of $2^2+2^2+2^2+2^2=16$ which is 4.

45. $(1,1,-1,1,-1,-1,1,1,-1)-(-1,-1,1,1,1,1,-1,-1,1)=(2,2,-2,0,-2,-2,2,2,2)$ whose length is the square root of 32 which is approximately 5.657.

46. Gugliotta, *supra* note 2.

tive method to doing this calculation. His use of singular value decomposition amounts to a variation on standard spatial voting models and as such is less novel.⁴⁷ It does highlight the incommensurability between certain geometric methods of analysis and legal methods of analysis in a way that might be helpful in the future.

47. The fact that his results are so close to more sophisticated models actually raises interesting mathematical issues, but those are not likely to be interesting to the readers of this journal.