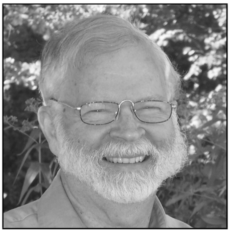
Singular Vectors' Subtle Secrets

David James, Michael Lachance, and Joan Remski



David James, a former Chair of the Michigan Section of the MAA, received his Ph.D. from the University of Wisconsin. In addition to writing on automorphic forms, he has authored interdisciplinary papers with computer scientists, math educators, an anthropologist, an architect, and, currently, even a political scientist. He's devised a course relating mathematics to music, art, architecture, and nature, and another that relates time to literature, philosophy, anthropology, science, and religion. Off time is spent in family travel and collecting musical instruments from various cultures, which he plays when no one else is present.

Michael Lachance is presently Chair of the Department of Mathematics and Statistics at the University of Michigan–Dearborn and completed his Ph.D. in approximation theory at the University of South Florida. He turned early in his career from approximation theory to its practice, consulting for the automotive industry. He has brought this experience to the fore in three symbiotic undergraduate courses: multivariate calculus, matrix algebra, and numerical analysis. The introduction of the SVD to undergraduate audiences has been a passion for some time; distance running, landscape gardening, international travel, and a wonderful family leave no room for other activities.

Joan Remski (remski@umd.umich.edu) received her Ph.D. from Michigan State University and has taught at the University of Michigan-Dearborn since 1997. Her research interests include numerical analysis, partial differential equations, and more recently, matrix algebra and digital image processing. She enjoys running, bicycling, cooking, and assembling Star Wars Lego sets with her two sons.

What unites cryptography, the Supreme Court, and the monks of Sampson's New England Monastery?

A new Dan Brown thriller?

We offer a different answer: the singular value decomposition (SVD). Whenever there is a hierarchy of relationships among individuals that can be captured by an adjacency matrix, e.g., letters follow or precede one another, justices agree or disagree,

doi:10.4169/college.math.j.42.2.086

monks admire or disdain each other, the SVD aids in identifying subsets that exhibit their own internal dynamics as well as external dynamics, shedding light respectively on relationships between and among individual letters, justices, and monks.

In this article, we develop a tool based upon the singular vectors of the SVD, to aid social scientists and simultaneously to shed light on the SVD in general and singular vectors in particular. While the singular value decomposition continues to gain exposure, it remains in many ways an enigma. We believe these examples will help to dispel some of its mystery.

The singular value decomposition

At the heart of the singular value decomposition is a geometric, and not immediately obvious, idea: If $\boldsymbol{A}$ is an $n \times n$ matrix with real values, then $\boldsymbol{A}$ maps the unit sphere in $\mathbb{R}^n$ onto an ellipsoid in $\mathbb{R}^n$ (see Figure 1). (We ask the reader to accept this as fact; its justification and more information can be found in textbooks [**5**], [**10**], reference books [**1**], [**4**], and articles [**2**].)

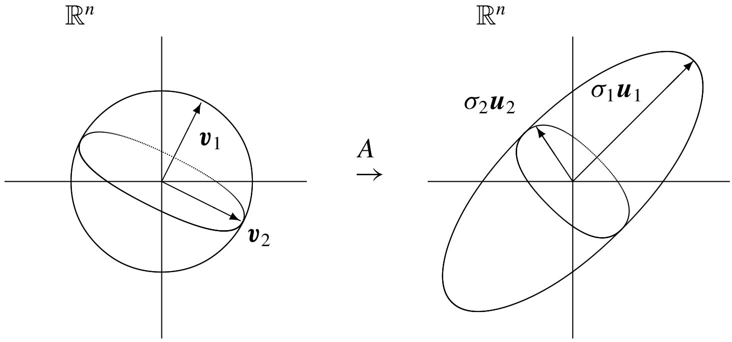


Figure 1. The unit sphere in $\mathbb{R}^n$ and its image under the mapping $\boldsymbol{A}$

Keeping this geometry in mind, let the singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0$, of $\boldsymbol{A}$ be the lengths of the principal semiaxes of the ellipsoid generated by mapping the unit sphere under $\boldsymbol{A}$. In the image space, $\mathbb{R}^n$, let $\boldsymbol{u}_i$ be the *unit* vector in the direction of the semiaxis with length σ_i. In the pre-image space, set $\boldsymbol{v}_i \in \mathbb{R}^n$ to be the vector on the unit sphere such that

$$\boldsymbol{A}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i, \quad i = 1, 2, \ldots, k.$$

The unit vectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ are called the *right singular vectors* and, surprisingly, they are mutually orthogonal in $\mathbb{R}^n$. Similarly, the unit vectors $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ are called the *left singular vectors* and are also pairwise orthogonal. While left and right may seem to be mislabeled for now, in the full matrix factorization this naming scheme makes sense.

If necessary, we extend the orthonormal set $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ to an orthonormal basis for $\mathbb{R}^n$. Let $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ be this basis and notice that $\boldsymbol{A}\boldsymbol{v}_i = 0$ for $i = k + 1, \ldots, n$, since our ellipsoid in the image space has only k principal semiaxes. We also extend the left singular vectors to a basis for $\mathbb{R}^n$, producing the orthonormal set $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$.

Let $\boldsymbol{V}$ be the matrix $[\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n]$ and similarly, let $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$. Now

$$\begin{aligned}\boldsymbol{AV} &= [\boldsymbol{Av}_1, \boldsymbol{Av}_2, \ldots, \boldsymbol{Av}_n] = [\sigma_1\boldsymbol{u}_1, \sigma_2\boldsymbol{u}_2, \ldots, \sigma_k\boldsymbol{u}_k, 0\boldsymbol{u}_{k+1}, \ldots, 0\boldsymbol{u}_n]\\ &= \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots \end{bmatrix}\begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots \end{bmatrix} = \boldsymbol{US},\end{aligned}$$

where $\boldsymbol{S}$ is the $n \times n$ diagonal matrix with diagonal entries $\sigma_1, \sigma_2, \ldots, \sigma_k, 0, \ldots, 0$.

Since the columns of $\boldsymbol{V}$ are orthonormal, $\boldsymbol{V}\boldsymbol{V}^T = \boldsymbol{I}$, which gives the matrix factorization

$$\boldsymbol{A} = \boldsymbol{USV}^T.$$

Now the *left* singular vectors appear on the *left* as the columns of $\boldsymbol{U}$ and the *right* singular vectors appear on the *right*, as the rows of $\boldsymbol{V}^T$. This is the *singular value decomposition* of $\boldsymbol{A}$.

Approximation

We expand $\boldsymbol{A}$ as follows.

$$\boldsymbol{A} = \boldsymbol{USV}^T = \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^T + \sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^T + \cdots + \sigma_k\boldsymbol{u}_k\boldsymbol{v}_k^T. \tag{1}$$

Note that if $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors, then $\boldsymbol{uv}^T$ is an $n \times n$ matrix with the property that each row of $\boldsymbol{uv}^T$ is a multiple of $\boldsymbol{v}^T$. Such a matrix is called *rank one*, since its reduced row echelon form contains only one nonzero row or alternatively since its image space is one-dimensional. Thus in the SVD expansion (1), $\boldsymbol{A}$ is a sum of rank one matrices.

There are lots of expansions of a matrix $\boldsymbol{A}$ in terms of rank one matrices, but this one is special. It can be shown that it is best possible in the following sense: The distance between $\boldsymbol{A}$ and $\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^T$ is the smallest among all the possible rank one approximates. Similarly, if we look at the matrix $\boldsymbol{A} - \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^T$, then its best rank one approximation is $\sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^T$. In this way the SVD parallels other expansions in mathematics (Taylor series, for example) where the error caused by truncating the sum has the same order of magnitude as the first omitted term.

Methodology

In many applications, data is generated in the form of a square matrix such that the row labels match the column labels. For example, the (i, j)th entry of a matrix might count the number of flights from city i to city j or the number of times politician i agrees with politician j. We will refer to such tables as *adjacency matrices* and write them as shown below.

$$\begin{array}{c|ccc} & \ell_1 & \cdots & \ell_n \\ \hline \ell_1 & & & \\ \vdots & & \boldsymbol{A} & \\ \ell_n & & & \end{array}$$

Note that these matrices are not necessarily symmetric, although the labels are.

Our object is to use the rank one expansion $\boldsymbol{A} = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^T + \cdots$ to establish relationships among the labels $\{\ell_i\}$. One important observation follows when the entries of $\boldsymbol{A}$ are nonnegative.

Proposition. *If $\boldsymbol{A}$ is a nonnegative adjacency matrix, then the first pair of singular vectors are of one sign, and we can take $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ to be nonnegative. The second pair's entries cannot be of one sign; both $\boldsymbol{u}_2$ and $\boldsymbol{v}_2$ must have both positive and negative entries.*

Proof. Recall our geometry: σ_1 is the length of the longest semiaxis in the ellipsoid generated by mapping the unit sphere under $\boldsymbol{A}$, $\boldsymbol{u}_1$ is the unit vector in the direction of the longest semiaxis and $\boldsymbol{v}_1$ is the preimage of $\sigma_1 \boldsymbol{u}_1$, so that $\boldsymbol{A}\boldsymbol{v}_1 = \sigma_1 \boldsymbol{u}_1$. So $\boldsymbol{v}_1$ is the unit vector $\boldsymbol{x}$ for which $\|\boldsymbol{A}\boldsymbol{x}\|$ is maximized [**1**]. Since $a_{ij} \geq 0$ it is easy to see that

$$\|\boldsymbol{A}\boldsymbol{x}\|^2 = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_{ij} x_j \right)^2 = \sum_{i=1}^{n} \left(\left(\sum_{j=1}^{n} a_{ij}^2 x_j^2 + 2 \sum_{1 \leq j < k \leq n} a_{ij} a_{ik} x_j x_k \right) \right),$$

is maximized when x_j and x_k have the same sign. Thus we can choose the entries of $\boldsymbol{v}_1$ all to be nonnegative. Since $\boldsymbol{A}\boldsymbol{v}_1 = \sigma_1 \boldsymbol{u}_1$ and $\sigma_1 > 0$, $\boldsymbol{u}_1$ will also have nonnegative components. The second part follows since $\boldsymbol{u}_1$ is orthogonal to $\boldsymbol{u}_2$, $\boldsymbol{u}_1 \neq \boldsymbol{0}$, and all the entries of $\boldsymbol{u}_1$ are nonnegative. A similar argument holds for $\boldsymbol{v}_2$. ■

The first singular vectors Given a nonnegative adjacency matrix $\boldsymbol{A}$, we compute its SVD and first approximate $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T$. Because each row of $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T$ is a multiple of $\boldsymbol{v}_1^T$, a large entry in $\boldsymbol{u}_1$ creates a correspondingly large row in $\boldsymbol{u}_1 \boldsymbol{v}_1^T$. If we think of this matrix as a surface, with the (i, j)th entry representing a height, then a large entry in $\boldsymbol{u}_1$ results in a visually prominent row ridge on this surface. Similarly, a large entry in $\boldsymbol{v}_1^T$ results in a prominent column ridge on this surface. By the same reasoning, small entries in $\boldsymbol{u}_1$ (or $\boldsymbol{v}_1$) create row (or column) valleys.

This gives a way to single out labels that are particularly weak or strong. A large component $\boldsymbol{u}_{1k}$ (or $\boldsymbol{v}_{1k}$) leads to a large row (or column) ridge in the outer product $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T$, and in $\boldsymbol{A}$ too since this outer product is the best rank one matrix approximation of $\boldsymbol{A}$. But a large kth row (or column) in $\boldsymbol{A}$ suggests that ℓ_k is a more prominent or a more frequently tabulated label in our data matrix. Thus, to rank the labels $\{\ell_1, \ell_2, \ldots, \ell_n\}$, we could look for large components in $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$. One way to visualize these entries is to follow Moler and Morrison's lead [**6**] and plot the label ℓ_k at the Cartesian point $(\boldsymbol{u}_{1k}, \boldsymbol{v}_{1k})$. Labels that are farther from the origin represent more dominant or higher ranking individuals, while labels closer to the origin indicate weaker or less prominent labels.

The second singular vectors We gain more interesting information by subtracting off average behavior and seeing how the labels deviate from it. Using the SVD, we think of $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T$ as the center of the data. Subtracting it from $\boldsymbol{A}$ and considering the residual, $\boldsymbol{A} - \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T$, we see the advantage of the SVD: $\sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^T$ is the best rank one approximation to $\boldsymbol{A} - \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T$. In other words,

$$\boldsymbol{A} - \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T \approx \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^T.$$

Now we look at where the residual, $\boldsymbol{A} - \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T$, is positive and where it is negative. Recalling that $\boldsymbol{u}_2$ and $\boldsymbol{v}_2$ take on both positive and negative values, we seek a way to identify groups of labels that exhibit similar behavior. If we rearrange the labels, a clear picture emerges.

First, we sort the entries of $\boldsymbol{u}_2$ from high to low, so the signs of the entries are $(+, +, \ldots, +, -, -, \ldots, -)$, and simultaneously change the order of the entries of $\boldsymbol{v}_2$, so that the row and column labels remain identical for $\boldsymbol{u}_2$ and $\boldsymbol{v}_2$. If the first k entries of $\boldsymbol{u}_2$ now have positive signs, we sort the corresponding first k entries in $\boldsymbol{v}_2$ high to low. Similarly, we rearrange the last $(n-k)$ entries in $\boldsymbol{v}_2$ from high to low. As before, we change the corresponding entries of $\boldsymbol{u}_2$.

For example, after the above sorting process has been completed, we might have $\operatorname{sgn}(\boldsymbol{u}_2) = (+++++----)^T$ and $\operatorname{sgn}(\boldsymbol{v}_2) = (++---++--)^T$. Consequently,

$$\operatorname{sgn}\left(\boldsymbol{u}_2\boldsymbol{v}_2^T\right) = \begin{bmatrix} \boldsymbol{+} & \boldsymbol{+} & - & - & - & + & + & - & - \\ \boldsymbol{+} & \boldsymbol{+} & - & - & - & + & + & - & - \\ + & + & \boldsymbol{-} & \boldsymbol{-} & \boldsymbol{-} & + & + & - & - \\ + & + & \boldsymbol{-} & \boldsymbol{-} & \boldsymbol{-} & + & + & - & - \\ + & + & \boldsymbol{-} & \boldsymbol{-} & \boldsymbol{-} & + & + & - & - \\ - & - & + & + & + & \boldsymbol{-} & \boldsymbol{-} & + & + \\ - & - & + & + & + & \boldsymbol{-} & \boldsymbol{-} & + & + \\ - & - & + & + & + & - & - & \boldsymbol{+} & \boldsymbol{+} \\ - & - & + & + & + & - & - & \boldsymbol{+} & \boldsymbol{+} \end{bmatrix}.$$

This sorting leads to two complementary types of rows, and two complementary types of columns, and consequently to at most four square blocks down the main diagonal, each having uniform signs. The first block has positive entries as a result of positive entries in $\boldsymbol{u}_2$ and $\boldsymbol{v}_2$; the next block has negative entries since $\boldsymbol{u}_2$ has positive entries and $\boldsymbol{v}_2$ has negative entries; the third block also has negative entries, but for the opposite reason, and the fourth block has positive entries since both $\boldsymbol{u}_2$ and $\boldsymbol{v}_2$ are negative.

In this way, each label ℓ_k falls into one of four signed blocks, and a plot of ℓ_k at the position given by the ordered pair $(\boldsymbol{u}_{2k}, \boldsymbol{v}_{2k})$ taken from the diagonal serves to identify potential subsets based upon the quadrants in which each resides. This eliminates the need to execute the sorting process mentioned above, as the ordered pairs $(\boldsymbol{u}_{2k}, \boldsymbol{v}_{2k})$ remain unchanged by rearrangement. Finally since $\sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^T$ approximates $A - \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^T$, the labels that share a particular quadrant all have the same type of deviation from the center matrix $\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^T$.

Case study: Textual analysis and cryptography

Taking Martin Luther King's "I Have a Dream" speech, for example, ignoring spaces and punctuation, we generate an adjacency matrix that counts how often the jth letter of the alphabet follows the ith letter. The first part of this 26×26 matrix is shown below.

	A	B	C	D	E	$\cdots$
A	0	17	25	24	0	$\cdots$
B	14	0	0	0	42	$\cdots$
C	30	0	0	0	29	$\cdots$
D	30	11	2	2	35	$\cdots$
E	86	8	36	80	46	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

The entry a_{ij} counts how many times the ith letter *precedes* the jth letter. In this speech, A precedes B 17 times and follows it 14 times. The full 26-by-26 matrix can be downloaded as a *Mathematica* file at [**7**]. The rank two singular value approximate $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^T$ now gives information on the distribution of letters in our chosen text. On different graphs in Figure 2 we plot the ordered pairs $(\boldsymbol{u}_{1k}, \boldsymbol{v}_{1k})$ and $(\boldsymbol{u}_{2k}, \boldsymbol{v}_{2k})$ as suggested in [**6**].

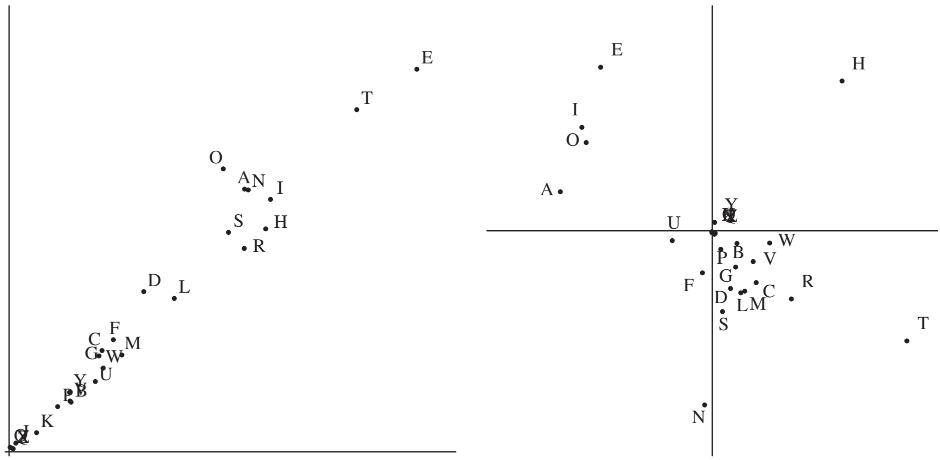


Figure 2. The first and second singular vector plots for the text example.

Looking at the first singular vector plot, we see the letters plotted farthest from the origin in order are E, T, I, N, A, O, H, S, R, L, D, . . . , a list that is strikingly similar to the classical letter frequency in English: E, T, A, O, I, N, S, H, R, D, L, . . . This order also closely matches the letter frequency in this particular speech. In other words, the sizes of the entries of both $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ are a good estimate of the letter-frequency ranking. The second singular vector plot is amazing: the more frequently occurring vowels all appear in the second quadrant, and most of consonants appear in the fourth. How did this happen?

In English, consonants are more likely to precede vowels rather than other consonants. Similarly, a vowel preceding another vowel is less likely than a vowel preceding a consonant. The frequency of these pairs in the "I Have a Dream" speech are in the table below.

preceded by	vowel	consonant
vowel	.052	.349
consonant	.358	.241

Since vowels prefer to precede consonants, along a vowel row in $\boldsymbol{A}$ we expect a positive deviation from the center $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T$ at columns corresponding to consonants, and a negative deviation at a vowel's column. Since $\sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^T$ approximates the residual $\boldsymbol{A} - \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T$, this same behavior is expected in $\sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^T$. In this case study, the signs of $\boldsymbol{u}_2 \boldsymbol{v}_2^T$ are of the form:

	a	e	b	c
a	$-$	$-$	$+$	$+$
e	$-$	$-$	$+$	$+$
b	$+$	$+$	$-$	$-$
c	$+$	$+$	$-$	$-$

where a and e represent typical vowels and b and c represent typical consonants. The only way $\operatorname{sgn}(\boldsymbol{u}_2\boldsymbol{v}_2^T)$ can have this form is when $\boldsymbol{u}_2 = (++--)^T$ and $\boldsymbol{v}_2 = (--++)^T$ or vice versa. This results in labels in the second and fourth quadrants in the second singular vector plot.

A second interesting relationship is the detection of a few cross-vowelers: consonants that partially act like vowels. The strings THE, THEN, THAT, and WHEN are quite common in English and this leaves H in an odd position in that it prefers to follow consonants and precede vowels. The reverse is true for the letter N, due to words containing the strings -ING and the two-letter words IN, AN, and ON. These letters find homes in the first and third quadrants respectively.

One application of this analysis is to single substitution ciphers. As an example, using a random permutation of the alphabet Martin Luther King's speech might be encrypted as "YGMW LPAFW TWNFL NQA N QFWND NIWFMPNB. . . ." Even if spaces and punctuation are omitted, the second singular vector plot applied to this cipher text immediately gives the cryptanalyst a good indication of which letters are vowels and which are consonants.

Much of this has been known since Moler and Morrison's paper [**6**]. It is such a compelling example of singular vector analysis, however, that we couldn't resist including it here. It deserves broader treatment and a wider audience.

Case study: The U.S. Supreme Court

In 2008 and 2009, the U.S. Supreme Court consisted of Justices Souter, Roberts, Ginsburg, Thomas, Alito, Kennedy, Scalia, Stevens, and Breyer. The court delivered over 135 opinions during these terms, including the 5–4 decision that Guantanamo Bay detainees have the right to challenge their detention in federal court. By tabulating each opinion using data from [**11**], we constructed Table 1, which shows the percentage of cases where each justice agrees with each of his or her colleagues. The labels along the left of the chart refer to the author of an opinion, while the labels across the top to whether that judge agreed with the author. For example, Justice Alito agreed with

Table 1. Percentage of agreement among U.S. Supreme Court justices.

	So	Ro	Gi	Th	Al	Ke	Sc	St	Br
So	100	97	83	80	86	87	87	80	82
Ro	53	100	56	91	94	100	100	63	66
Gi	87	83	100	77	77	100	80	73	86
Th	67	93	73	100	93	80	83	60	73
Al	82	93	82	82	100	100	82	75	82
Ke	61	68	64	46	68	100	57	57	73
Sc	50	81	53	95	87	82	100	45	58
St	81	59	75	59	56	75	56	100	69
Br	78	88	78	66	81	88	66	75	100

an opinion authored by Justice Thomas 93% of the time, while Justice Thomas agreed with an opinion authored by Justice Alito 82% of the time. A large jth component in $\boldsymbol{v}_1$ indicates Justice j frequently agrees with others. On the other hand, row j shows how often the other Justices agree with Justice j's opinion, a subtly different measure. A large component in $\boldsymbol{u}_1$ corresponds to a large row, so if Justice j tends to be agreed with, we expect a relatively large value in the jth component of $\boldsymbol{u}_1$.

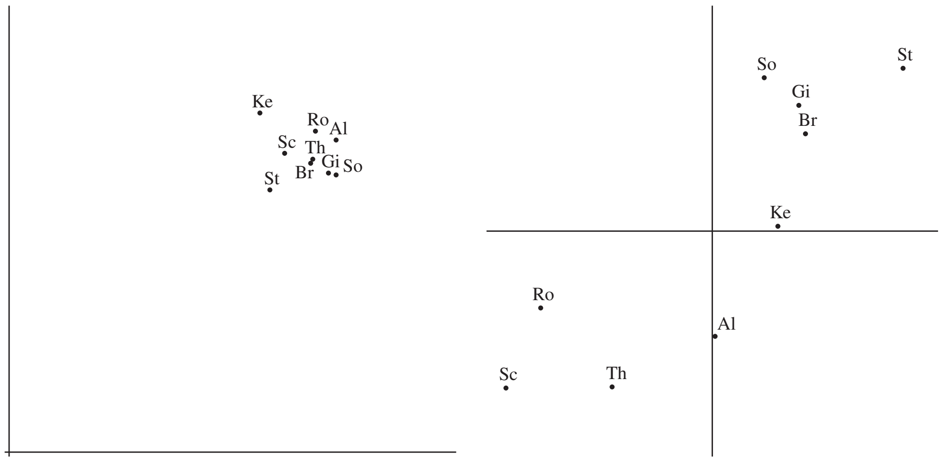


Figure 3. The first and second singular vector plots for the supreme court data.

The first singular vector plot shows Justices Kennedy, Alito, and Roberts as those who agree with others the most—they are the uppermost points in the first singular vector plot, having the largest values in $\boldsymbol{v}_1$, while Stevens agrees the least. The validity of these conclusions can be verified by comparing the sums of the entries in the columns. However in the agreed-with measure, Alito, Souter, and Ginsberg rank the highest—they are the rightmost points, having the largest values of $\boldsymbol{u}_1$, while Kennedy and Stevens are the least agreed with.

The second singular vector plot shows something more interesting. After subtracting off $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T$, we see two distinct ways to deviate from average. The two groups might be generally labeled as progressives and conservatives. That some justices are labeled as progressive or conservative is not surprising, but note that two Clinton appointees, Ginsberg and Breyer, are grouped together, associated with Justice Souter who was appointed by a conservative president (Reagan), but who became more progressive since his appointment. All the other justices appointed by Reagan or by one of the Bushes are situated in the opposite direction. The Ford appointee, Justice Stevens, has two highly individualistic characteristics: he is the most likely to disagree, and a quick glance through the court opinions verifies that the justices he disagrees with most are Alito, Roberts, Thomas, Kennedy, and Scalia.

One difference between this example and the cryptography case study is that when Justice i and Justice j are similar, we anticipate a positive entry where row i meets column j to indicate more agreeability between these two Justices than on average would be expected. Here we expect to end up with a matrix $\text{sgn}(\boldsymbol{u}_2 \boldsymbol{v}_2^T)$ with exactly the opposite signs as in the previous case study, and that is exactly what happens. The methodology section explains why this results in labels in the first and third quadrants.

This example was inspired by work by Lawrence Sirovich [**9**]. Sirovich built a 468 by 9 matrix of $+1/-1$'s for the Rehnquist court (1995–2003) and used SVD rank reduction techniques to conclude that by virtue of strong alliances this court acted as if composed of only 4.68 ideal independent justices.

Case study: Sampson's monastery

In the social network literature, data collected by Sampson [**8**] is a widely used benchmark to test algorithms that partition data into groups or cliques. In 1969 Sampson surveyed eighteen novice monks at a New England monastery and asked them to rank their peers in four different areas: like/dislike, esteem, personal influence, and consistency with the monastic creed. Sampson used direct observation to discern several distinct social groups and his hypotheses were later confirmed when the monks in one group were either expelled or resigned because of religious differences. The rankings generate a matrix with entry (i, j) of $\boldsymbol{A}$ representing the number of times the ith monk ranked the jth monk in his top three in all four relations. We use the same data matrix as [**3**] and apply an SVD analysis. Figure 4 gives our two singular vector plots. The four cliques identified by Sampson are labeled distinctly: ∘ for the 'Loyal Opposition' (Sampson's term); ▴ for the 'Waverers'; ▪ for the 'Young Turks'; and • for the 'Outcasts'.

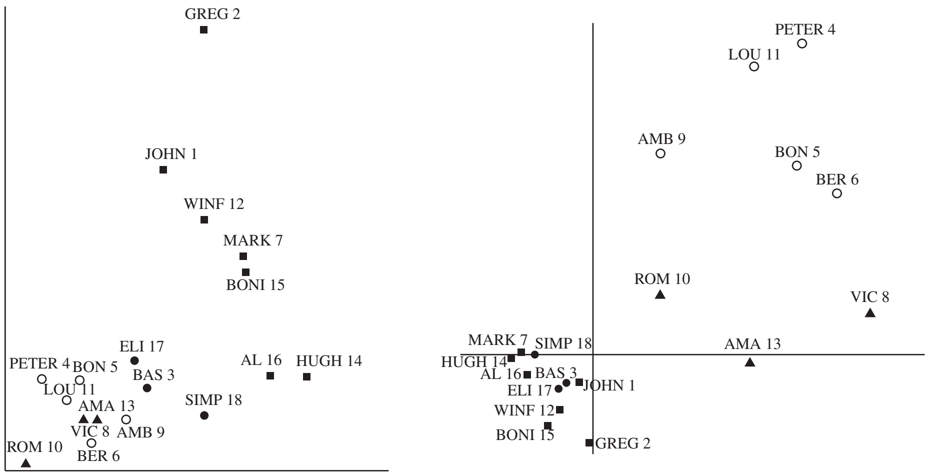


Figure 4. The first and second singular vector plots for Sampson's monastery data.

In our first singular vector plot, novices Greg, John, Winfred, Mark, Boni, and Hugh are the ones farthest from the origin, indicating they have strong positive opinions of others (or a high $\boldsymbol{u}_{1k}$ value) or others have a strong positive opinion of them (high $\boldsymbol{v}_{1k}$ value), or both. All these monks (plus Al) are the Young Turks. Sampson considered them the more liberal newcomers to the monastery. Monks given the highest ranking by their peers (Greg, John, Winfred, Peter and Bon) lie above the line $y = x$ since $\boldsymbol{v}_{1k} > \boldsymbol{u}_{1k}$. They are the leaders, that is, Sampson identified Greg, John, and Winfred as leaders of the Young Turks, while Peter and Bon are considered leaders of the Loyal Opposition. The Outcasts, who were later dismissed for being too immature, are near or below the line $y = x$ indicating that they were more likely to rank their peers highly while their peers were less likely to rank them highly, that is, $\boldsymbol{v}_{1k} < \boldsymbol{u}_{1k}$.

In our second singular vector plot, the first quadrant contains the Loyal Opposition and almost all members of the Waverers. The two remaining groups, the Young Turks and the Outcasts, are in the third quadrant. Sampson describes the Waverers as being in intense conflict and moving between the Loyal Oppsition and the Young Turks. Our analysis places Vic and Rom in quadrant one, weakly allied with members of the Loyal Opposition, identical to the results in [**3**] and [**12**].

Conclusion

The purpose of this paper has been to demonstrate the usefulness of the SVD. In particular, we've shown how it can be used to analyze adjacency matrices, not only to find groups and cliques but also to identify whether individuals are strong or weak members of them, a problem of obvious importance to social scientists. The three applications presented here have strong appeal for students in any linear algebra class and provide motivation for the inclusion of the SVD as a course topic.

Our analysis is of square matrices, but the SVD applies more generally to any matrix ([**5**], [**10**], [**1**]). Using square matrices has some advantages both theoretical and pedagogical. Theoretically, because the labels of rows and columns are the same, we have been able to determine how cliques are formed by examining the second singular vectors. Pedagogically, students will find the algebra and geometry of the SVD easier to understand because all of the singular vectors are elements of $\mathbb{R}^n$.

Acknowledgment. We would like to thank the editor for his valuable suggestions.

Summary. Social scientists use adjacency tables to discover influence networks within and among groups. Building on work by Moler and Morrison, we use ordered pairs from the components of the first and second singular vectors of adjacency matrices as tools to distinguish these groups and to identify particularly strong or weak individuals.

References

1. D. Bau III and L. N. Trefethen, *Numerical Linear Algebra*, SIAM, Philadelphia PA, 1997.
2. J. Clifford, D. James, M. Lachance, and J. Remski, A constructive approach to the singular value decomposition and symmetric Schur factorization, *Amer. Math. Monthly* **112** (2005) 358–363. `doi:10.2307/30037473`
3. G. B. Davis and K. Carley, Clearing the FOG: Fuzzy, overlapping groups for social networks, *Social Networks* **30** (2008) 201–212. `doi:10.1016/j.socnet.2008.03.001`
4. G. Golub and van Loan, Matrix Computations, *SIAM: Series B, Numerical Analysis* **2** (1965) 205–224. `doi:10.1137/0702016`
5. D. C. Lay, *Linear Algebra and Its Applications*, 3rd ed., Pearson/Addison Wesley, Boston MA, 2006.
6. C. Moler and D. Morrison, Singular value analysis of cryptograms, *Amer. Math. Monthly* **90** (1983) 78–87. `doi:10.2307/2975804`
7. J. Remski, textual analysis notebook, available at `http://www.umd.umich.edu/~remski/IHaveADream.nb`, 2010.
8. S. F. Sampson, *A Novitiate in a period of change: An experimental and case study of social relationships*, unpublished doctoral dissertation, 1968.
9. L. Sirovich, A pattern analysis of the second Rehnquist U.S. Supreme Court, *Proc. Natl. Acad. Sci.* **100** (2003) 7432–7437.
10. G. Strang, *Linear Algebra and Its Applications*, Thompson Brooks/Cole, Belmont CA, 2006.
11. U.S. Supreme Court Slip Opinions, available at `http://www.supremecourtus.gov/`, 2008–2009.
12. H. C. White, S. A. Boorman, and R. L. Breiger, Social structure from multiple networks. I. Blockmodels of roles and positions, *American Journal of Sociololgy* **81** (1976) 730–780. `doi:10.1086/226141`