Math 2270-004 Week 11 notes
We will not necessarily finish the material from a given day's notes on that day. We may also add or subtract some material as the week progresses, but these notes represent an in-depth outline of what we plan to cover. These notes cover material in 5.4-5.6

Mon Mar 26

- 5.4 matrices for linear transformations as a general framework to understand change of bases, diagonalization, and similar matrices.

Announcements:

Warm-up Exercise:

Monday Review and look ahead:

We've been studying *linear transformations* $T : V \rightarrow W$ between vector spaces, which include matrix transformations $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ given as $T(\underline{\boldsymbol{x}}) = A\,\underline{\boldsymbol{x}}$.

We've been studying how coordinates change when we change bases in $\mathbb{R}^n$.

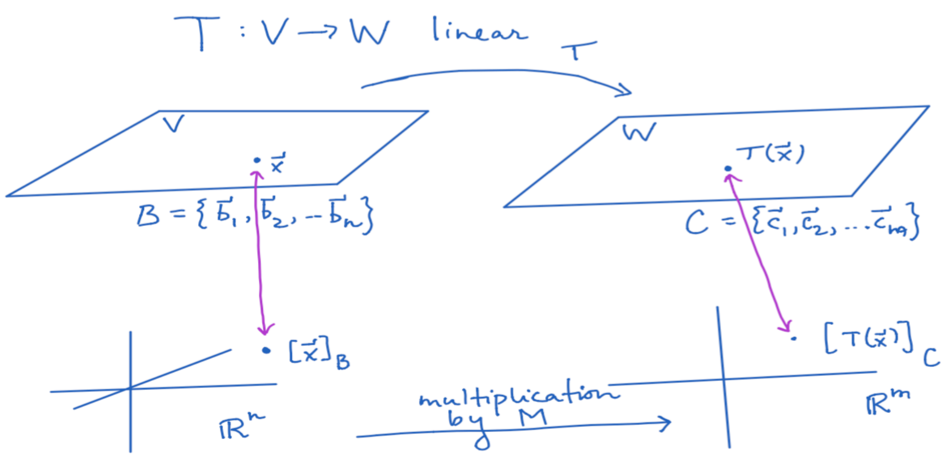
The last thing we studied in depth before the midterm was eigenvectors and eigenvalues for square matrices A, and the notion of diagonalizability, which we understood in an algebraic sense.

On the Wednesday before the midterm we introduced section 5.4, about how linear transformations $T : V \rightarrow W$ are associated with matrix transformations from $\mathbb{R}^n \rightarrow \mathbb{R}^m$, once we choose bases for V and W. We didn't have time to explain how this general framework is connected to all of our previous change of coordinates discussion, to matrix diagonalizability, and to the more general notion of similar matrices. That's what we'll do today.

Tomorrow we'll study section 5.5 on complex eigenvalues and eigenvectors. To understand the geometry of matrix transformations with complex eigendata we'll use "similar matrices" notions from today, to see that (in the 2×2 case), such matrices are similar to "rotation-dilations". You saw a hint of this on a food for thought problem before break, if you dared.

Wednesday we'll start section 5.6 on discrete dynamical systems, and we'll continue that discussion into Friday with google page rank. These section 5.6 topics are more expository than comprehensive, and for fun. There will be some follow-up homework problems.

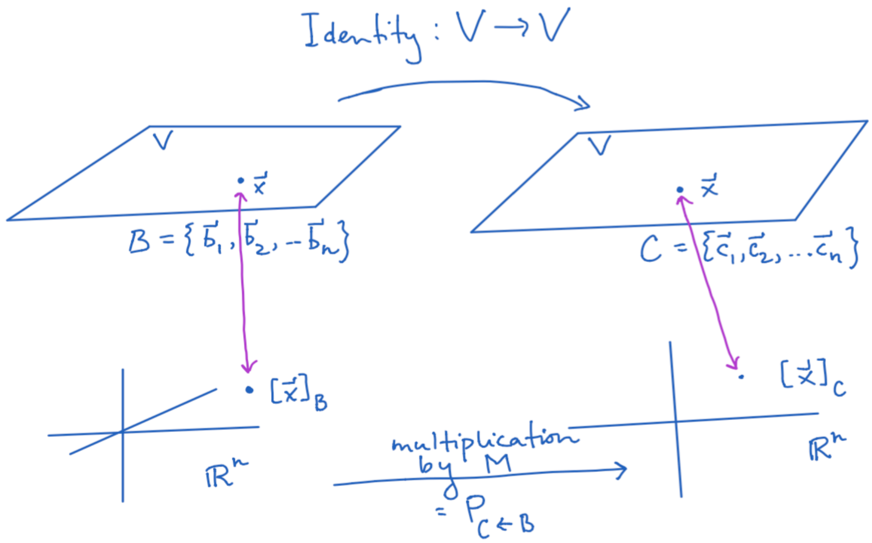
Recall, if we have a linear transformation $T : V \rightarrow W$ and bases $B = \{\underline{\boldsymbol{b}}_1, \underline{\boldsymbol{b}}_2, \ldots \underline{\boldsymbol{b}}_n\}$ in V, $C = \{\underline{\boldsymbol{c}}_1, \underline{\boldsymbol{c}}_2, \ldots \underline{\boldsymbol{c}}_m\}$ in W, then the matrix of T with respect to these two bases transforms the B coordinates of vectors $\underline{\boldsymbol{x}} \in V$ to the C coordinates of $T(\underline{\boldsymbol{x}})$ in a straightforward way:



$$M = \left[[T(\vec{b}_1)]_C \quad [T(\vec{b}_2)]_C \quad \cdots \quad [T(\vec{b}_n)]_C \right]$$

<u>Exercise 1)</u> Explain why the columns of the matrix M *have to be* the C coordinate vectors of T applied the B basis vectors. Do this two ways: (1) using the chart. AND (2) seeing what must happen when you multiply M by the standard basis vectors. This should help you remember in case you get confused.

Exercise 2) Fill in the matrix M for changing coordinates in a general vector space. We focused on changing coordinates in $\mathbb{R}^n$ in section 4.7, which is a special case of this when $V = \mathbb{R}^n$ itself. The text also discussed the more general context below, in that section.



Example: On the midterm: Let $B = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ and $C = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right\}$ be two bases for $V = \mathbb{R}^2$.

Find the change of coordinates matrix $P_{C \leftarrow B}$:

Solution:

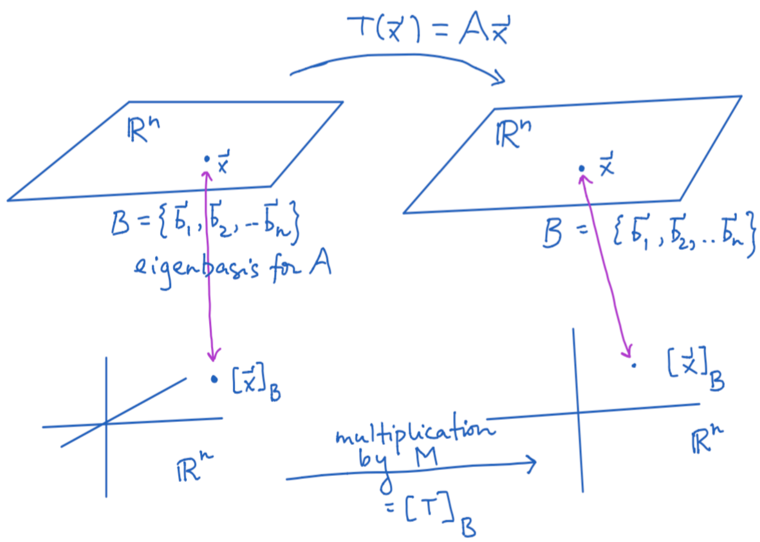
$$P_{C \leftarrow B} = \left[[\underline{\boldsymbol{b}}_1]_C \ [\underline{\boldsymbol{b}}_2]_C \right]$$

Since $\begin{bmatrix} 1 & 2 & 1 & 0 \\ 1 & 3 & 2 & 1 \end{bmatrix}$ reduces to $\begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 1 & 1 \end{bmatrix}$ the answer was $P_{C \leftarrow B} = \begin{bmatrix} -1 & -2 \\ 1 & 1 \end{bmatrix}$.

(Which, as the diagram indicates, could also have been computed as

$$P_{C \leftarrow B} = P_{C \leftarrow E} P_{E \leftarrow B} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} .)$$

Exercise 4) What if a matrix A is diagonalizable? What is the matrix of $T(\underline{\boldsymbol{x}}) = A\,\underline{\boldsymbol{x}}$ with respect to the eigenbasis? How does this connect to our matrix identities for diagonalization? Fill in the matrix M below, and then compute another way to express it, as a triple product using the diagram.

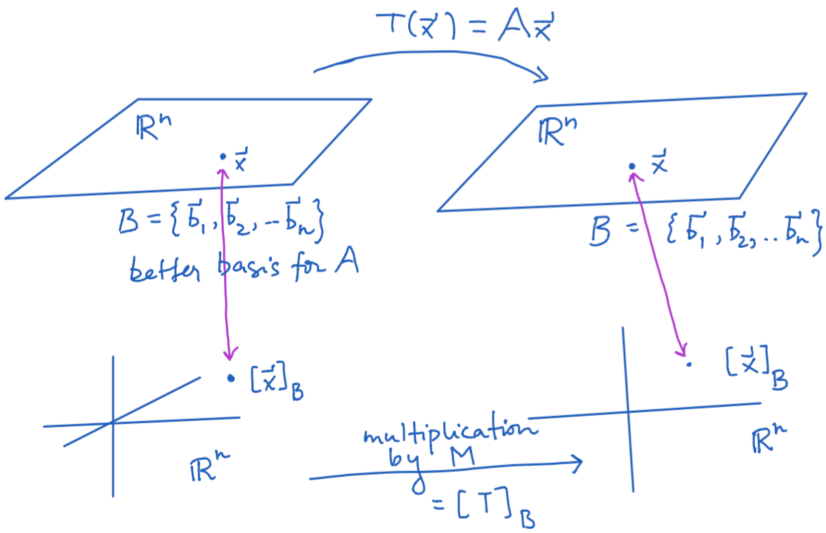


Example, from the week before break:

$$A = \begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix} \qquad E_{\lambda=4} = span\left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right\} \qquad E_{\lambda=1} = span\left\{\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right\} \qquad B = \left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right\}$$

Write the various matrices corresponding to the diagram above.

Even if the matrix A is not diagonalizable, there may be a better basis to help understand the transformation $T(\underline{\boldsymbol{x}}) = A\,\underline{\boldsymbol{x}}$. The diagram on the previous page didn't require that B be a basis of eigenvectors....maybe it was just a "better" basis than the standard basis, to understand T.



Exercise 5 (If we have time - this one is not essential.) Try to pick a better basis to understand the matrix transformation $T(\underline{\boldsymbol{x}}) = C\,\underline{\boldsymbol{x}}$, even though the matrix C is not diagonalizable. Compute $M = P^{-1} A\, P$ or compute M directly, to see if it really is a "better" matrix.

$$C = \begin{bmatrix} 4 & 4 \\ -1 & 0 \end{bmatrix}$$

Tues Mar 27

- 5.5 Complex eigenvalues and eigenvectors

Announcements:

Warm-up Exercise:

We'll focus on 2 × 2 matrices, for simplicity. In this case it will turn out that a matrix with real entries and complex eigenvalues is always similar to a rotation-dilation matrix.

<u>Definition</u> A matrix of the form $A = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ is called a *rotation-dilation* matrix, because for $r = \sqrt{a^2 + b^2}$ we can rewrite A as

$$A = r \begin{bmatrix} \frac{a}{r} & -\frac{b}{r} \\ \frac{b}{r} & \frac{a}{r} \end{bmatrix} = r \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

So the transformation $T(\underline{\mathbf{x}}) = A\,\underline{\mathbf{x}}$ rotates vectors by an angle θ and then scales them by a factor of r. (So A^2 rotates by an angle $2\,\theta$ and scales by r^2; A^3 rotates by an angle $3\,\theta$ and scales by r^3, etc.

<u>Exercise 1</u> Draw the transformation picture for

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

and interpret this transformation as a rotation-dilation.

Exercise 2) What are the eigenvalues of a rotation-dilation matrix $A = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$?

It is possible for a matrix A with real entries to be diagonalizable if one allows complex scalars and vectors, even if it's not diagonalizable with real eigenvalues and eigenvectors. You saw an example of that on a food for thought problem, if you weren't afraid. We'll use a matrix today that we'll use later as well, in section 5.6, to study an interesting discrete dynamical system. This matrix is not a rotation-dilation matrix, but it is *similar* to one, and that fact will help us understand the discrete dynamical system.

Exercise 3) Let

$$B = \begin{bmatrix} .9 & -.4 \\ .1 & .9 \end{bmatrix}$$

Find the (complex) eigenvalues and eigenvectors for B.

General facts we saw illustrated in the example, about complex eigenvalues and eigenvectors: Let A be a matrix with real entries, and let

$$A\,\underline{\mathbf{v}} = \lambda\,\underline{\mathbf{v}}$$

with $\lambda = a + b\,i$, $\underline{\mathbf{v}} = \underline{\mathbf{u}} + i\,\underline{\mathbf{w}}$ complex, $(a, b \in \mathbb{R}, \underline{\mathbf{u}}, \underline{\mathbf{w}} \in \mathbb{R}^n)$. Then we write

$$\operatorname{Re} \lambda = a, \quad \operatorname{Im} \lambda = b$$

$$\operatorname{Re} \underline{\mathbf{v}} = \underline{\mathbf{u}}, \qquad \operatorname{Im} \underline{\mathbf{v}} = \underline{\mathbf{w}}.$$

So, the equation $A\,\underline{\mathbf{v}} = \lambda\,\underline{\mathbf{v}}$ expands as

$$A\,(\underline{\mathbf{u}} + i\,\underline{\mathbf{w}}) = (a + b\,i)\,(\underline{\mathbf{u}} + i\,\underline{\mathbf{w}}).$$

It will always be true then that the conjugate $\lambda = a - b\,i$ is also an eigenvalue, and the conjugate vector $\underline{\mathbf{v}} = \underline{\mathbf{u}} - i\,\underline{\mathbf{w}}$ will be a corresponding eigenvector, because it will satisfy

$$A\,(\underline{\mathbf{u}} - i\,\underline{\mathbf{w}}) = (a - b\,i)\,(\underline{\mathbf{u}} - i\,\underline{\mathbf{w}})$$

<u>Exercise 4</u> Verify that if the first eigenvector equation holds, then

$$A\,\underline{\mathbf{u}} = a\,\underline{\mathbf{u}} - b\,\underline{\mathbf{w}}$$
$$A\,\underline{\mathbf{w}} = b\,\underline{\mathbf{u}} + a\,\underline{\mathbf{w}}$$

Then check that these equalities automatically make the second conjugate eigenvector equation true as well.

<u>Theorem</u> Let A be a real 2×2 matrix with complex eigenvalues. Then A is similar to a rotation-dilation matrix.

proof: Let a complex eigenvalue and eigenvector be given by $\lambda = a + b\,i$, $\underline{\boldsymbol{v}} = \underline{\boldsymbol{u}} + i\,\underline{\boldsymbol{w}}$ complex, ($a, b \in \mathbb{R}$, $\underline{\boldsymbol{u}}, \underline{\boldsymbol{w}} \in \mathbb{R}^n$) Choose

$$P = [\ \mathrm{Re}\,\underline{\boldsymbol{v}} \quad \mathrm{Im}\,\underline{\boldsymbol{v}}\,] = [\,\underline{\boldsymbol{u}}\ \ \underline{\boldsymbol{w}}\,]$$

(One can check that $\{\underline{\boldsymbol{u}}, \underline{\boldsymbol{w}}\}$ is automatically independent.) Then, using the equations of Exercise 4, we mimic what we did for diagonalizable matrices...

$$A\,[\,\underline{\boldsymbol{u}}\ \ \underline{\boldsymbol{w}}\,] = [\,a\,\underline{\boldsymbol{u}} - b\,\underline{\boldsymbol{w}},\ b\,\underline{\boldsymbol{u}} + a\,\underline{\boldsymbol{w}}\,]$$

$$= [\ \underline{\boldsymbol{u}}\ \ \underline{\boldsymbol{w}}\ ] \begin{bmatrix} a & b \\ -b & a \end{bmatrix}.$$

$$A\,P = P \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$$

$$P^{-1}\,A\,P = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}.$$

(The matrix on the right is a rotation-dilation matrix ... nobody ever said what the sign of b was. :-))

It's a mess, but we can carry out the procedure of the theorem, for the matrix B in exercise 3,

$$B = \begin{bmatrix} .9 & -.4 \\ .1 & .9 \end{bmatrix}$$

using $\lambda = .9 - .2\,i$, $\underline{\boldsymbol{v}} = \begin{bmatrix} -2\,i \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + i \begin{bmatrix} -2 \\ 0 \end{bmatrix}$, one gets

$$P = [\ \mathrm{Re}\,\underline{\boldsymbol{v}} \quad \mathrm{Im}\,\underline{\boldsymbol{v}}] = \begin{bmatrix} 0 & -2 \\ 1 & 0 \end{bmatrix}, \qquad P^{-1} = \frac{1}{2}\begin{bmatrix} 0 & 2 \\ -1 & 0 \end{bmatrix}$$

$$P^{-1} B\, P = \frac{1}{2}\begin{bmatrix} 0 & 2 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} .9 & -.4 \\ .1 & .9 \end{bmatrix}\begin{bmatrix} 0 & -2 \\ 1 & 0 \end{bmatrix}$$

$$P^{-1} B\, P = \begin{bmatrix} .9 & -.2 \\ .2 & .9 \end{bmatrix} = \sqrt{.85}\begin{bmatrix} \dfrac{.9}{\sqrt{.85}} & -\dfrac{.2}{\sqrt{.85}} \\ \dfrac{.2}{\sqrt{.85}} & \dfrac{.9}{\sqrt{.85}} \end{bmatrix}$$

$$P^{-1} B\, P = r\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

for $r = \sqrt{.85} \approx .92$, $\theta = \arctan\left(\dfrac{2}{9}\right) \approx .22$ radians.

Wed Mar 28

- 5.6 Discrete dynamical systems

Announcements:

Warm-up Exercise:

what is a discrete dynamical system, with constant transition matrix?

Example: (See text, page 304). A predator-prey system: "Deep in the redwood forests of California, dusky-footed wood rats povide up to 80 % of the diet for the spotted owl, the main predator of the wood rat..." This model is a simplified version of how the owls and wood rats interact: Denote the owl and wood rat populations at time k months by

$$x_k = \begin{bmatrix} \mathrm{O}_k \\ R_k \end{bmatrix},$$

where O_k is the number of owls in the region studied, and R_k is the number of rats (measured in the thousands). Suppose

$$\begin{aligned} \mathrm{O}_{k+1} &= .5\,\mathrm{O}_k + .4\,R_k \\ R_{k+1} &= -p\,\mathrm{O}_k + 1.1\,R_k \end{aligned}$$

where p is a positive parameter (predation constant) to be specified. The $(.5)\,\mathrm{O}_k$ in the first equation says that with no wood rats for food, only half the owls will survive each month, while the $1.1\,R_k$ says that with no owls as predators, the rat population will grow by 10 % each month. If the rats are plentiful, the $.4\,R_k$ will tend to make the owl population rise, while the negative term $-p\,\mathrm{O}_k$ measures the deaths of rats due to predation by owls. (In fact, $1000\,p$ is the average number of rats eaten by one owl in one month.) Determine the evolution of this system when the predation parameter p is .104.

solution We see that

$$\begin{bmatrix} \mathrm{O}_{k+1} \\ R_{k+1} \end{bmatrix} = \begin{bmatrix} .5 & .4 \\ -.104 & 1.1 \end{bmatrix} \begin{bmatrix} \mathrm{O}_k \\ R_k \end{bmatrix}.$$

Writing

$$\underline{\boldsymbol{x}}_k = \begin{bmatrix} \mathrm{O}_k \\ R_k \end{bmatrix},\quad A = \begin{bmatrix} .5 & .4 \\ -.104 & 1.1 \end{bmatrix}$$

we see

$$\underline{\boldsymbol{x}}_k = A^k\,\underline{\boldsymbol{x}}_0$$

where $\underline{\boldsymbol{x}}_0 = \begin{bmatrix} \mathrm{O}_0 \\ R_0 \end{bmatrix}$ are the owl and rat populations at the start. This is a *dynamical system* because it involves quantities that are changing over time. It is a *discrete* dynamical system because we are letting time change by discrete amounts (of one month). If we allowed time to vary continuously we would get statements about derivatives and would be studying differential equations instead. (See Math 2280 or 2250.) A is the constant transition matrix.

The way to understand this problem is to use the fact that A is diagonalizable:

Input:

eigenvalues $\begin{pmatrix} 0.5 & 0.4 \\ -0.104 & 1.1 \end{pmatrix}$

Results:

$\lambda_1 \approx 1.02$

$\lambda_2 \approx 0.58$

Corresponding eigenvectors:

$v_1 \approx (0.769231, 1)$

$v_2 = (5, 1)$

(An exact eigenvector for $\lambda_1 = 1.02$ is actually $\underline{\boldsymbol{v}}_1 = \begin{bmatrix} 10 \\ 13 \end{bmatrix}$.)

Exercise 1 Describe the long term behavior of the solutions $\underline{\boldsymbol{x}}_k$ to this problem. Begin by writing $\underline{\boldsymbol{x}}_0$ in terms of the eigenbasis. Then apply A repeatedly. (We could do this in terms of diagonalization and the matrix of A with respect to the eigenbasis, but that would be unneccesarily confusing.)

$$\underline{\boldsymbol{x}}_0 = c_1 \underline{\boldsymbol{v}}_1 + c_2 \underline{\boldsymbol{v}}_2$$

$$= c_1 \begin{bmatrix} 10 \\ 13 \end{bmatrix} + c_2 \begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

Exercise 2 Suppose we have a general discrete dynamical system

$$\underline{\boldsymbol{x}}_k = A^k \, \underline{\boldsymbol{x}}_0$$

and that the matrix A is diagonalizable (over the real numbers, or even over the complex numbers). What can you say about the long term behavior of solutions, depending on the absolute value of the eigenvalues of A?

72. Use the method outlined in Exercise 70 to check for which values of the constants a, b, and c the matrix

$$A = \begin{bmatrix} 1 & a & b \\ 0 & 0 & c \\ 0 & 0 & 1 \end{bmatrix} \text{ is diagonalizable.}$$

73. Prove the Cayley-Hamilton theorem, $f_A(A) = 0$, for diagonalizable matrices A. See Exercise 7.3.54.

74. In both parts of this problem, consider the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix},$$

with eigenvalues $\lambda_1 = 5$ and $\lambda_2 = -1$ (see Example 1).

a. Are the column vectors of the matrices $A - \lambda_1 I_2$ and $A - \lambda_2 I_2$ eigenvectors of A? Explain. Does this work for other 2×2 matrices? What about diagonalizable $n \times n$ matrices with two distinct eigenvalues, such as projections or reflections? (*Hint:* Exercise 70 is helpful.)

b. Are the column vectors of

$$A - \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

eigenvectors of A? Explain.

"Linear Algebra with Applications"
by Otto Bretscher

7.5 Complex Eigenvalues

Imagine that you are diabetic and have to pay close attention to how your body metabolizes glucose. After you eat a heavy meal, the glucose concentration will reach a peak, and then it will slowly return to the fasting level. Certain hormones help regulate the glucose metabolism, especially the hormone insulin. (Compare with Exercise 7.1.52.) Let $g(t)$ be the excess glucose concentration in your blood, usually measured in milligrams of glucose per 100 milliliters of blood. (*Excess* means that we measure how much the glucose concentration deviates from the fasting level.) A negative value of $g(t)$ indicates that the glucose concentration is below fasting level at time t. Let $h(t)$ be the excess insulin concentration in your blood. Researchers have developed mathematical models for the glucose regulatory system. The following is one such model, in slightly simplified (linearized) form.

$$\begin{aligned} g(t+1) &= ag(t) - bh(t) \\ h(t+1) &= cg(t) + dh(t) \end{aligned}$$

(These formulas apply between meals; obviously, the system is disturbed during and right after a meal.)

In these formulas, a, b, c, and d are positive constants; constants a and d will be less than 1. The term $-bh(t)$ expresses the fact that insulin helps your body absorb glucose, and the term $cg(t)$ represents the fact that the glucose in your blood stimulates the pancreas to secrete insulin.

For your system, the equations might be

$$\begin{aligned} g(t+1) &= 0.9g(t) - 0.4h(t) \\ h(t+1) &= 0.1g(t) + 0.9h(t), \end{aligned}$$

with initial values $g(0) = 100$ and $h(0) = 0$, after a heavy meal. Here, time t is measured in hours.

After one hour, the values will be $g(1) = 90$ and $h(1) = 10$. Some of the glucose has been absorbed, and the excess glucose has stimulated the pancreas to produce 10 extra units of insulin.

The rounded values of $g(t)$ and $h(t)$ in the following table give you some sense for the evolution of this dynamical system.

t	0	1	2	3	4	5	6	7	8	15	22	29
$g(t)$	100	90	77	62.1	46.3	30.6	15.7	2.3	−9.3	−29	1.6	9.5
$h(t)$	0	10	18	23.9	27.7	29.6	29.7	28.3	25.7	−2	−8.3	0.3

We can "connect the dots" to sketch a rough trajectory, visualizing the long-term behavior. See Figure 1.

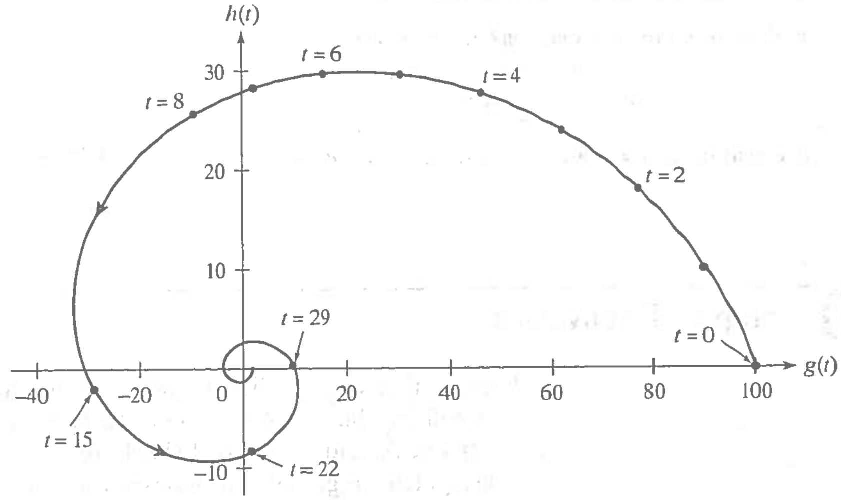


Figure 1

We see that after 7 hours the excess glucose is almost gone, but now there are about 30 units of excess insulin in the system. Since this excess insulin helps to reduce glucose further, the glucose concentration will now fall below fasting level, reaching about −30 after 15 hours. (You will feel awfully hungry by now.) Under normal circumstances, you would have taken another meal in the meantime, of course, but let's consider the case of (voluntary or involuntary) fasting.

We leave it to the reader to explain the concentrations after 22 and 29 hours, in terms of how glucose and insulin concentrations influence each other, according to our model. The *spiraling trajectory* indicates an *oscillatory behavior* of the system: Both glucose and insulin levels will swing back and forth around the fasting level, like a damped pendulum. Both concentrations will approach the fasting level (thus the name).

Another way to visualize this oscillatory behavior is to graph the functions $g(t)$ and $h(t)$ against time, using the values from our table. See Figure 2.

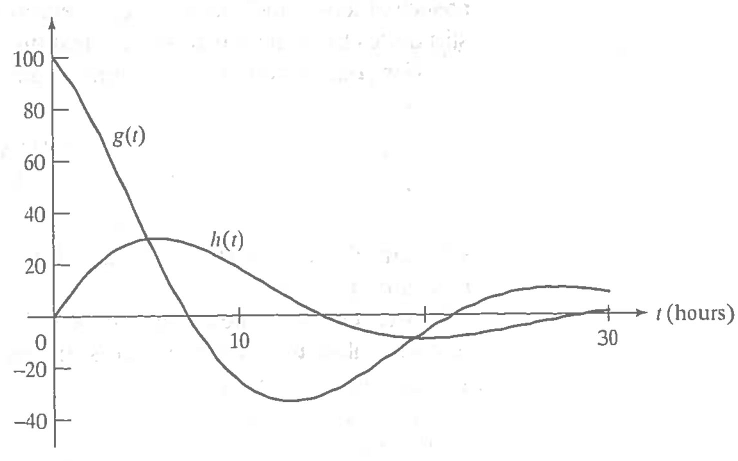


Figure 2

Example: From the algebra yesterday, and after a fair amount of work, For the dynamical system

$$\begin{bmatrix} g_{k+1} \\ h_{k+1} \end{bmatrix} = \begin{bmatrix} .9 & -.4 \\ .1 & .9 \end{bmatrix} \begin{bmatrix} g_k \\ h_k \end{bmatrix}$$

and with $\begin{bmatrix} g_0 \\ h_0 \end{bmatrix} = \begin{bmatrix} 100 \\ 0 \end{bmatrix}$, one can calculate and understand the spiral picture...

$$\begin{bmatrix} g_k \\ h_k \end{bmatrix} = \begin{bmatrix} .9 & -.4 \\ .1 & .9 \end{bmatrix}^k \begin{bmatrix} 100 \\ 0 \end{bmatrix} = = .92^k \begin{bmatrix} 100 \cos(k\,\theta) \\ 50 \sin(k\,\theta) \end{bmatrix}$$

$$\theta \approx .22 \text{ radians.}$$

yipes!

For $r = \sqrt{.85} \approx .92$, $\theta = \arctan\left(\frac{2}{9}\right) \approx .22$ radians.

$$B = r\,P\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}P^{-1}$$

$$B^2 = r^2\,P\begin{bmatrix} \cos(2\,\theta) & -\sin(2\,\theta) \\ \sin(2\,\theta) & \cos(2\,\theta) \end{bmatrix}P^{-1}$$

$$B^n = r^n\,P\begin{bmatrix} \cos(n\,\theta) & -\sin(n\,\theta) \\ \sin(n\,\theta) & \cos(n\,\theta) \end{bmatrix}P^{-1}$$

$$B^n\begin{bmatrix} 100 \\ 0 \end{bmatrix} = .92^n\begin{bmatrix} 0 & -2 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \cos(n\,\theta) & -\sin(n\,\theta) \\ \sin(n\,\theta) & \cos(n\,\theta) \end{bmatrix}\frac{1}{2}\begin{bmatrix} 0 & 2 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 100 \\ 0 \end{bmatrix}$$

$$.92^n\begin{bmatrix} 0 & -2 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \cos(n\,\theta) & -\sin(n\,\theta) \\ \sin(n\,\theta) & \cos(n\,\theta) \end{bmatrix}\begin{bmatrix} 0 \\ -50 \end{bmatrix}$$

$$= .92^n\begin{bmatrix} 0 & -2 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 50\sin(n\theta) \\ -50\cos(n\theta) \end{bmatrix}$$

$$= .92^n\begin{bmatrix} 100\cos(n\theta) \\ 50\sin(n\theta) \end{bmatrix}$$

Fri Mar 30

- 5.6 Discrete dynamical systems: Google page rank, using some notes I wrote a while ago. I'll probably also bring a handout to class. There are other documents on the internet about this subject. One that a lot of people like is called "THE $25,000,000,000* EIGENVECTOR THE LINEAR ALGEBRA BEHIND GOOGLE", which you can google. This topic is also related to section 4.9 of our text.

Announcements:

Warm-up Exercise:

The Giving Game: Google Page Rank

University of Utah Teachers' Math Circle

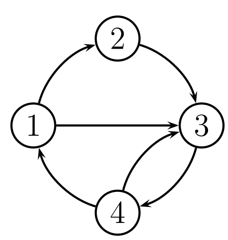
Nick Korevaar

March 24, 2009

Stage 1: The Game

Imagine a game in which you repeatedly distribute something desirable to your friends, according to a fixed template. For example, maybe you're giving away "play–doh" or pennies! (Or it could be you're a web site, and you're voting for the sites you link to. Or maybe, you're a football team, and you're voting for yourself, along with any teams that have beaten you.)

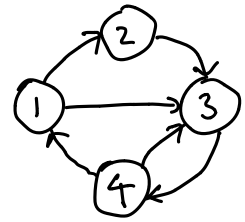
Let's play a small–sized game. Maybe there are four friends in your group, and at each stage you split your material into equal sized lumps, and pass it along to your friends, according to this template:



The question at the heart of the basic Google page rank algorithm is: in a voting game like this, with billions of linked web sites and some initial vote distribution, does the way the votes are distributed settle down in the limit? If so, sites with more limiting votes must ultimately be receiving a lot of votes, so must be considered important by a lot of sites, or at least by sites which themselves are receiving a lot of votes. Let's play!

1. Decide on your initial material allocations. I recommend giving it all to one person at the start, even though that doesn't seem fair. If you're using pennies, 33 is a nice number for this template. At each stage, split your current amount into equal portions and distribute it to your friends, according to the template above. If you have remainder pennies, distribute them randomly. Play the game many (20?) times, and see what ultimately happens to the amounts of material each person controls. Compare results from different groups, with different initial allocations.

2. While you're playing the giving game, figure out a way to model and explain this process algebraically!



Stage 2: Modeling the game algebraically

The game we just played is an example of a *discrete dynamical system*, with constant *transition matrix*. Let the initial fraction of play dough distributed to the four players be given by

$$\mathbf{x}_0 = \begin{bmatrix} x_{0,1} \\ x_{0,2} \\ x_{0,3} \\ x_{0,4} \end{bmatrix}, \quad \sum_{i=1}^{4} x_{0,i} = 1$$

Then for our game template on page 1, we get the fractions at later stages by

$$\begin{bmatrix} x_{k+1,1} \\ x_{k+1,2} \\ x_{k+1,3} \\ x_{k+1,4} \end{bmatrix} = x_{k,1} \begin{bmatrix} 0 \\ 0.5 \\ 0.5 \\ 0 \end{bmatrix} + x_{k,2} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_{k,3} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} + x_{k,4} \begin{bmatrix} 0.5 \\ 0 \\ 0.5 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} x_{k+1,1} \\ x_{k+1,2} \\ x_{k+1,3} \\ x_{k+1,4} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0 \\ 0.5 & 1 & 0 & 0.5 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_{k,1} \\ x_{k,2} \\ x_{k,3} \\ x_{k,4} \end{bmatrix}$$

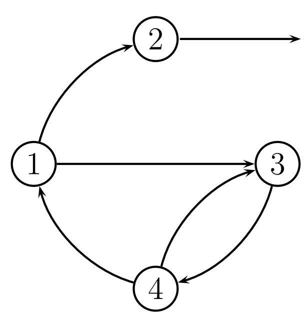
So in matrix form, $\mathbf{x}_k = A^k \mathbf{x}_0$ for the transition matrix A given above.

3. Compute a large power of A. What do you notice, and how is this related to the page 1 experiment?

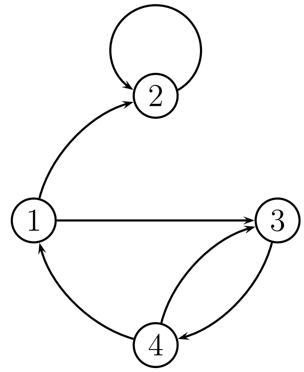
4. The limiting "fractions" in this problem really are fractions (and not irrational numbers). What are they? Is there a matrix equation you could solve to find them, for this small problem? Hint: the limiting fractions should remain fixed when you play the game.

5. Not all giving games have happy endings. What happens for the following templates?

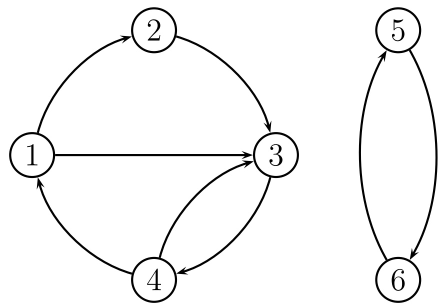
   (a)



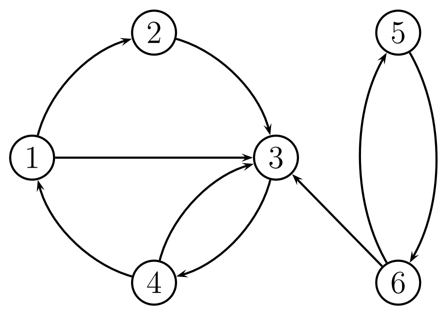
(b)



(c)



(d)



Here's what separates good giving–game templates, like the page 1 example, from the bad examples 5a,b,c,d.

Definition: A square matrix S is called *stochastic* if all its entries are positive, and the entries in each column add up to exactly one.

Definition: A square matrix A is *almost stochastic* if all its entries are non–negative, the entries in each column add up to one, and if there is a positive power k so that A^k is stochastic.

6. What do these definitions mean *vis-à-vis* play–doh distribution? Hint: if it all starts at position j, then the inital fraction vector $\mathbf{x}_0 = \mathbf{e_j}$, i.e. has a 1 in position j and zeroes elsewhere. After k steps, the material is distributed according to $A^k\mathbf{e_j}$, which is the j^{th} column of A^k.

Stage 3: Theoretical basis for Google page rank

Theorem. *(Perron–Frobenius) Let A be almost stochastic. Let $\mathbf{x}_0$ be any "fraction vector" i.e. all its entries are non–negative and their sum is one. Then the discrete dynamical system*

$$\mathbf{x}_k = A^k \mathbf{x}_0$$

has a unique limiting fraction vector $\mathbf{z}$, and each entry of $\mathbf{z}$ is positive. Furthermore, the matrix powers A^k converge to a limit matrix, each of whose columns are equal to $\mathbf{z}$.

proof: Let $A = [a_{ij}]$ be almost stochastic. We know, by "conservation of play–doh", that if $\mathbf{v}$ is a fraction vector, then so is $A\mathbf{v}$. As a warm–up for the full proof of the P.F. theorem, let's check this fact algebraically:

$$\begin{aligned}\sum_{i=1}^{n}(A\mathbf{v})_i &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}v_j = \sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij}v_j \\ &= \sum_{j=1}^{n} v_j \left(\sum_{i=1}^{n} a_{ij}\right) = \sum_{j=1}^{n} v_j = 1\end{aligned}$$

Thus as long as $\mathbf{x}_0$ is a fraction vector, so is each iterate $A^N\mathbf{x}_0$.

Since A is almost stochastic, there is a power l so that $S = A^l$ is stochastic. For any (large) N, write $N = kl + r$, where $N/l = k$ with remainder r, $0 \le r < l$. Then

$$A^N\mathbf{x}_0 = A^{kl+r}\mathbf{x}_0 = \left(A^l\right)^k A^r\mathbf{x}_0 = S^k A^r \mathbf{x}_0$$

As $N \to \infty$ so does k, and there are only l choices for $A^r\mathbf{x}_0$, $0 \le r \le l-1$. Thus if we prove the P.F. theorem for stochastic matrices S, i.e. $S^k\mathbf{y}_0$ has a unique limit independent of $\mathbf{y}_0$, then the more general result for almost stochastic A follows.

So let $S = [s_{ij}]$ be an $n \times n$ stochastic matrix, with each $s_{ij} \ge \varepsilon > 0$. Let 1 be the matrix for which each entry is 1. Then we may write:

$$B = S - \varepsilon 1; \quad S = B + \varepsilon 1. \tag{1}$$

Here $B = [b_{ij}]$ has non–negative entries, and each column of B sums to

$$1 - n\varepsilon := \mu < 1. \tag{2}$$

We prove the P.F. theorem in a way which reflects your page 1 experiment: we'll show that whenever $\mathbf{v}$ and $\mathbf{w}$ are fraction vectors, then $S\mathbf{v}$ and $S\mathbf{w}$ are geometrically closer to each other than were $\mathbf{v}$ and $\mathbf{w}$. Precisely, our "metric" for measuring the distance "d" between two fraction vectors is

$$\mathrm{d}(\mathbf{v}, \mathbf{w}) := \sum_{i=1}^{n} |v_i - w_i|. \tag{3}$$

Here's the magic: if $\mathbf{v}$ is any fraction vector, then for the matrix 1, of ones,

$$(1\mathbf{v})_i = \sum_{j=1}^{n} 1 v_j = 1.$$

So if $\mathbf{v}, \mathbf{w}$ are both fraction vectors, then $1\mathbf{v} = 1\mathbf{w}$. Using matrix and vector algebra, we compute using equations (1), (2):

$$\begin{aligned} S\mathbf{v} - S\mathbf{w} &= (B + \varepsilon 1)\mathbf{v} - (B + \varepsilon 1)\mathbf{w} \\ &= B(\mathbf{v} - \mathbf{w}) \end{aligned} \tag{4}$$

So by equation (3),

$$\begin{aligned} \mathrm{d}(S\mathbf{v}, S\mathbf{w}) &= \sum_{i=1}^{n} \left| \sum_{j=1}^{n} b_{ij}(v_j - w_j) \right| \\ &\leq \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} |v_j - w_j| \\ &= \sum_{j=1}^{n} |v_j - w_j| \sum_{i=1}^{n} b_{ij} \\ &= \mu \sum_{j=1}^{n} |v_j - w_j| \\ &= \mu \mathrm{d}(\mathbf{v}, \mathbf{w}) \end{aligned} \tag{5}$$

Iterating inequality (5) yields

$$\mathrm{d}(S^k\mathbf{v}, S^k\mathbf{w}) \leq \mu^k \mathrm{d}(\mathbf{v}, \mathbf{w}). \tag{6}$$

Since fraction vectors have non–negative entries which sum to 1, the greatest distance between any two fraction vectors is 2:

$$\mathrm{d}(\mathbf{v}, \mathbf{w}) = \sum_{i=1}^{n} |v_i - w_i] \leq \sum_{i=1}^{n} v_i + w_i = 2$$

So, no matter what different initial fraction vectors experimenters begin with, after k iterations the resulting fraction vectors are within $2\mu^k$ of each other, and by choosing k large enough, we can deduce the existence of, and estimate the common limit $\mathbf{z}$ with as much precision as desired. Furthermore, if all initial material is allotted to node j, then the initial fraction vector $\mathbf{e_j}$ has a 1 in position j and zeroes elsewhere. $S^k\mathbf{e_j}$, (or $A^N\mathbf{e_j}$) is on one hand the j^{th} column of S^k (or A^N), but on the other hand is converging to $\mathbf{z}$. So each column of the limit matrix for S^k and A^N equals $\mathbf{z}$. Finally, if $\mathbf{x}_0$ is any initial fraction vector, then $S(S^k\mathbf{x}_0) = S^{k+1}(\mathbf{x}_0)$ is converging to $S(\mathbf{z})$ and also to $\mathbf{z}$, so $S(\mathbf{z}) = \mathbf{z}$ (and $A\mathbf{z} = \mathbf{z}$). Since the entries of $\mathbf{z}$ are non–negative (and sum to 1) and the entries of S are all positive, the entries of $S\mathbf{z}$ $(= \mathbf{z})$ are all positive. ∎

Stage 4: The Google fudge factor

Sergey Brin and Larry Page realized that the world wide web is not almost stochastic. However, in addition to realizing that the Perron–Frobenius theorem was potentially useful for ranking URLs, they figured out a simple way to guarantee stochasticity—the "Google fudge factor."

Rather than using the voting matrix A described in the previous stages, they take a combination of A with the matrix of 1s we called 1. For (Brin an Pages' choice of) $\varepsilon = .15$ and n equal the number of nodes, consider the Google matrix

$$G = (1-\varepsilon)A + \frac{\varepsilon}{n}1.$$

(See [Austin, 2008]).

If A is almost stochastic, then each column of G also sums to 1 and each entry is at least ε/n. This G is stochastic! In other words, if you use this transition matrix everyone gets a piece of your play–doh, but you still get to give more to your friends.

7. Consider the giving game from 5c. Its transition matrix

$$A = \left[\begin{array}{cccc|cc} 0 & 0 & 0 & .5 & 0 & 0 \\ .5 & 0 & 0 & 0 & 0 & 0 \\ .5 & 1 & 0 & .5 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right]$$

is not almost stochastic. For $\varepsilon = .3$ and $\varepsilon/n = .05$, work out the Google matrix G, along with the limit rankings for the six sites. If you were upset that site 4 was ranked as equal to site 3 in the game you played for stage 1, you may be happier now.

Historical notes

The Perron–Frobenius theorem had historical applications to input–output economic modeling. The idea of using it for ranking seems to have originated with Joseph B. Keller, a Stanford University emeritus mathematics professor. According to a December 2008 article in the Stanford Math Newsletter [Keller, 2008], Professor Keller originally explained his team ranking algorithm in the 1978 Courant Institute Christmas Lecture, and later submitted an article to Sports Illustrated in which he used his algorithm to deduce unbiased rankings for the National League baseball teams at the end of the 1984 season. His article was rejected. Utah professor James Keener visited Stanford in the early 1990s, learned of Joe Keller's idea, and wrote a SIAM article in which he ranked football teams [Keener, 1993].

Keener's ideas seem to have found their way into some of the current BCS college football ranking schemes which often cause boosters a certain amount of heartburn. I know of no claim that there is any direct path from Keller's original insights, through Keener's paper, to Brin and Pages' amazing Google success story. Still it is interesting to look back and notice that the seminal idea had been floating "in the air" for a number of years before it occurred to anyone to apply it to Internet searches.

Acknowledgement: Thanks to Jason Underdown for creating the graph diagrams and for typesetting this document in LaTeX.

References

David D. Austin. How Google Finds Your Needle in the Web's Haystack. 2008. URL http://www.ams.org/featurecolumn/archive/pagerank.html.

Sergey Brin and Lawrence Page. The anatomy of a large–scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 33:107–117, 1998. URL http://infolab.stanford.edu/pub/papers/google.pdf.

James Keener. The Perron–Frobenius Theorem and the ranking of football teams. *SIAM Rev.*, 35:80–93, 1993.

Joseph B. Keller. Stanford University Mathematics Department newsletter, 2008.