Math 2270-002 Week 9 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 4.5, 4.6, 4.9, 5.1-5.2.

Mon Oct 22

- 4.5, 4.6 Finish general theorems about finite dimensional vector spaces, bases, spanning sets, linearly independent sets and subspaces from 4.5; and complete the discussion of the four fundamental subspaces, from 4.6.

Announcements:

Warm-up Exercise:

Monday Review!

We've been studying *vector spaces*, which are a generalization of $\mathbb{R}^n$. They occur as *subspaces* of $\mathbb{R}^n$; also as vector spaces and subspaces of matrices, and of function spaces, for example. There are general theorems for vector spaces having to do with questions of *linear independence, span, basis, dimension* that we already understand well for $\mathbb{R}^n$. We ended Friday in the midst of a discussion of these theorems, and we'll complete that discussion in Part 1 of today's notes.

We've also been studying and using *linear transformations* $T : V \rightarrow W$ between vector spaces, which are generalizations of matrix transformations $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ given as $T(\underline{\boldsymbol{x}}) = A\,\underline{\boldsymbol{x}}$. A particularly useful linear transformation once if we have a basis $\beta = \{\underline{\boldsymbol{b}}_1, \underline{\boldsymbol{b}}_2, \ldots \underline{\boldsymbol{b}}_n\}$ for any vector space V is the coordinate transformation isomorphism:

$$T(\underline{\boldsymbol{v}}) = [\underline{\boldsymbol{v}}]_\beta$$

$$T : V \rightarrow \mathbb{R}^n.$$

The coordinate transformation and its inverse function are helpful because they allow us to translate questions about linear independence and span in V into equivalent questions in $\mathbb{R}^n$, where we already have the tools to answer those questions.

For an $m \times n$ matrix A we've studied the subspaces $Nul\,A \subseteq \mathbb{R}^n$ and $Col\,A \subseteq \mathbb{R}^m$, which are the kernel and range of the associated linear transformation $T(\underline{\boldsymbol{x}}) = A\,\underline{\boldsymbol{x}}$, $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$. On Friday we introduced two more subspaces connected to the geometry of the matrix transformations $T(\underline{\boldsymbol{x}}) = A\,\underline{\boldsymbol{x}}$. There are $Row\,A$ and $Nul\,A^T$. We'll complete the discussion of the four fundamental subspaces associated to matrix transformations in Part 2 of today's notes; we'll see how $Row\,A$ and $Nul\,A$ are related to a decomposition of the domain $\mathbb{R}^n$ of T, which is analogous to how $Col\,A = row\left(A^T\right)$ and $Nul\,A^T$ decompose the codomain $\mathbb{R}^m$.

on Friday :

Part I Monday
vector space theorems

There is a circle of ideas related to linear independence, span, and basis for vector spaces, which it is good to try and understand carefully. That's what we'll do today. These ideas generalize (and use) ideas we've already explored more concretely, and facts we already know to be true for the vector spaces $\mathbb{R}^n$. (A vector space that does not have a basis with a finite number of elements is said to be *infinite dimensional*. For example the space of all polynomials of arbitrarily high degree is an infinite dimensional vector space. We often study finite dimensional subspaces of infinite dimensional vector spaces.)

Theorem 1 (constructing a basis from a spanning set): Let V be a vector space of dimension at least one, and let $span\{\underline{\boldsymbol{v}}_1, \underline{\boldsymbol{v}}_2, \ldots \underline{\boldsymbol{v}}_p\} = V$.
Then a subset of the spanning set is a basis for V. (We followed a procedure like this to extract bases for *Col A*.)

here's how: If $\{\vec{v}_1, \vec{v}_2, \ldots \vec{v}_p\}$ is already independent, it's a basis and Theorem is proved.

otherwise, one of these vectors in a linear combo of the rest
by renumbering, assume

$$\vec{v}_p = d_1\vec{v}_1 + d_2\vec{v}_2 + \cdots + d_{p-1}\vec{v}_{p-1} \quad (\text{so } d_j \text{ may} = 0)$$

then any $\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_p\vec{v}_p$ is actually in $span\{\vec{v}_1, \vec{v}_2, \ldots \vec{v}_{p-1}\}$.
because ↑

continue until the remaining set has same span as original set, but is now independent, so it's a basis

Theorem 2 Let V be a vector space, with basis $\beta = \{\underline{\boldsymbol{b}}_1, \underline{\boldsymbol{b}}_2, \ldots \underline{\boldsymbol{b}}_n\}$. Then any set in V containing more than n elements must be linearly dependent. (We used reduced row echelon form to understand this in $\mathbb{R}^n$.)

Let $A = \{\vec{a}_1, \vec{a}_2, \ldots \vec{a}_N\}$ $N > n$ be a set in V.

Consider

$$c_1\vec{a}_1 + c_2\vec{a}_2 + \cdots + c_N\vec{a}_N = \vec{0}$$

and take coord transformation to $\mathbb{R}^n$, $[\ \ ]_\beta$

$$[\quad " \quad]_\beta = [\vec{0}]_\beta = \vec{0} \in \mathbb{R}^n$$

vector eqtn in $\mathbb{R}^n$

$$c_1[\vec{a}_1]_\beta + c_2[\vec{a}_2]_\beta + \cdots + c_N[\vec{a}_N]_\beta = \vec{0}$$

$$n\left\{\underbrace{\begin{bmatrix}[\vec{a}_1]_\beta & [\vec{a}_2]_\beta & \cdots & [\vec{a}_N]_\beta\end{bmatrix}}_{N}\right. \begin{bmatrix}c_1\\c_2\\\vdots\\c_N\end{bmatrix} = \vec{0}$$

at least $N-n$ free parameters so lots of dependencies to be contined ...

Theorem 3 Let V be a vector space, with basis $\beta = \{\underline{\boldsymbol{b}}_1, \underline{\boldsymbol{b}}_2, \ ... \ \underline{\boldsymbol{b}}_n\}$. Then no set $\alpha = \{\underline{\boldsymbol{a}}_1, \underline{\boldsymbol{a}}_2, \ ... \ \underline{\boldsymbol{a}}_p\}$ with $p < n$ vectors can span V. (We know this for $\mathbb{R}^n$.)

Theorem 4 Let V be a vector space, with basis $\beta = \{\underline{\boldsymbol{b}}_1, \underline{\boldsymbol{b}}_2, \ ... \ \underline{\boldsymbol{b}}_n\}$. Let $\alpha = \{\underline{\boldsymbol{a}}_1, \underline{\boldsymbol{a}}_2, \ ... \ \underline{\boldsymbol{a}}_p\}$ be a set of independent vectors that don't span V. Then $p < n$, and additional vectors can be added to the set α to create a basis $\{\underline{\boldsymbol{a}}_1, \underline{\boldsymbol{a}}_2, \ ... \ \underline{\boldsymbol{a}}_p, \ ... \ \underline{\boldsymbol{a}}_n\}$ (We followed a procedure like this when we figured out all the subspaces of $\mathbb{R}^3$.)

Theorem 5 Let Let V be a vector space, with basis $\beta = \{\underline{\boldsymbol{b}}_1, \underline{\boldsymbol{b}}_2, \ \dots \ \underline{\boldsymbol{b}}_n\}$. Then every basis for V has exactly n vectors. (We know this for $\mathbb{R}^n$.)

Theorem 6 Let Let V be a vector space, with basis $\beta = \{\underline{\boldsymbol{b}}_1, \underline{\boldsymbol{b}}_2, \ \dots \ \underline{\boldsymbol{b}}_n\}$. If $\alpha = \{\underline{\boldsymbol{a}}_1, \underline{\boldsymbol{a}}_2, \ \dots \ \underline{\boldsymbol{a}}_n\}$ is another collection of exactly n vectors in V, and if $span\{\underline{\boldsymbol{a}}_1, \underline{\boldsymbol{a}}_2, \ \dots \ \underline{\boldsymbol{a}}_n\} = V$, then the set α is automatically linearly independent and a basis. Conversely, if the set $\{\underline{\boldsymbol{a}}_1, \underline{\boldsymbol{a}}_2, \ \dots \ \underline{\boldsymbol{a}}_n\}$ is linearly independent, then $span\{\underline{\boldsymbol{a}}_1, \underline{\boldsymbol{a}}_2, \ \dots \ \underline{\boldsymbol{a}}_n\} = V$ is guaranteed, and α is a basis. (We know all these facts for $\mathbb{R}^n$ from reduced row echelon form considerations.)

Corollary Let Let V be a vector space of dimension n. Then the subspaces of V have dimensions $0, 1, 2, \ldots n-1, n$. (We know this for $\mathbb{R}^n$.)

Remark We used the coordinate transformation isomorphism between a vector space V with basis $\beta = \{\underline{\boldsymbol{b}}_1, \underline{\boldsymbol{b}}_2, \ldots \underline{\boldsymbol{b}}_n\}$ for Theorem 2, but argued more abstractly for the other theorems. An alternate (quicker) approach is to just note that because the coordinate transformation is an isomorphism it preserves sets of independent vectors, and maps spans of vectors to spans of the image vectors, so maps subspaces to subspaces. Then every one of the theorems above follows from their special cases in $\mathbb{R}^n$, which we've already proven. But this shortcut shortchanges the conceptual ideas to some extent, which is why we've discussed the proofs more abstractly.

Part 2:
For an $m \times n$ matrix A there are actually four interesting subspaces. We've studied the first two of these quite a bit:

$$Nul\,A = \{\underline{\mathbf{x}} \in \mathbb{R}^n : A\,\underline{\mathbf{x}} = \underline{\mathbf{0}}\} \subseteq \mathbb{R}^n$$

and

$$Col\,A = \{\underline{\boldsymbol{b}} \in \mathbb{R}^m : \underline{\boldsymbol{b}} = A\,\underline{\mathbf{x}}, \underline{\mathbf{x}} \in \mathbb{R}^n\} = span\{\underline{\boldsymbol{a}}_1, \underline{\boldsymbol{a}}_2, \ldots \underline{\boldsymbol{a}}_n\} \subseteq \mathbb{R}^m.$$

(Here we expressed $A = [\underline{\boldsymbol{a}}_1\ \underline{\boldsymbol{a}}_2\ \ldots\ \underline{\boldsymbol{a}}_n]$ in terms of its columns.) Through homework and class discussions we've understood the *rank+nullity Theorem*, that $dim\ Col\,A + dim\ Nul\,A = n$. This theorem follows from considerations of the reduced row echelon form of A and is connected to the number of pivot columns and the number of non-pivot columns in A. On Friday we introduced the other two of the four interesting subspaces connected to A. These are

$$Row\,A := span\{\underline{\boldsymbol{R}}_1, \underline{\boldsymbol{R}}_2, \ldots \underline{\boldsymbol{R}}_m\} \subseteq \mathbb{R}^n, \text{ where we write } A \text{ in terms of its rows, } A = \begin{bmatrix} \underline{\boldsymbol{R}}_1 \\ \underline{\boldsymbol{R}}_2 \\ \vdots \\ \underline{\boldsymbol{R}}_m \end{bmatrix}.$$

Note that $Row\,A = Col\,A^T$. The final subspace is

$$Nul\left(A^T\right) = \{\underline{\boldsymbol{y}} \in \mathbb{R}^m : A^T\,\underline{\boldsymbol{y}} = \underline{\mathbf{0}}\} \subseteq \mathbb{R}^m$$

<u>Exercise 1</u>: Using the reduced row echelon form of A we have realized the following facts. Let's review our reasoning, some of which you understood in last week's homework and all of which we've discussed in class. I've pasted our warm-up discussion from Friday into the following page, where we studied a large example in this context:

$$dim\,(Col\,A) = \#\ pivot\ columns\ in\ A\ (=\#pivots)$$

$$dim\,(Nul\,A) = \#\ non\text{-}pivot\ columns\ in\ A$$

$$dim\,(Row\,A) = \#\ pivot\ rows\ in\ A\ \ (=\#pivots)$$

$$dim\left(Nul\,A^T\right) = \#\ non\text{-}pivot\ columns\ in\ A^T.$$

The $dim\,(Col\,A) = dim\,(Row\,A)$ is called the *rank* of the matrix A and is the number of pivots in both A and A^T. If we call this number "r", then

$$dim\,(Nul\,A) = n - r$$
$$dim\left(Nul\,A^T\right) = m - r.$$

Example from Friday

Math 2270-002
Friday October 19

Big example of four fundamental subspaces associated to each matrix.
Here is a matrix and it's reduced row echelon form, from quiz 7, and related matrices

$$A := \begin{bmatrix} 1 & -1 & 0 & -1 & 5 \\ 2 & -3 & -1 & -4 & 8 \\ 2 & -2 & 0 & -2 & 2 \\ 1 & 2 & 3 & 5 & 1 \end{bmatrix} \quad \text{row reduces to} \quad \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$\vec{a}_1 \ \vec{a}_2 \ \vec{a}_3 \ \vec{a}_4 \ \vec{a}_5$

$$A^T = \begin{bmatrix} 1 & 2 & 2 & 1 \\ -1 & -3 & -2 & 2 \\ 0 & -1 & 0 & 3 \\ -1 & -4 & -2 & 5 \\ 5 & 8 & 2 & 1 \end{bmatrix} \text{ row reduces to } \begin{bmatrix} 1 & 0 & 0 & \frac{9}{2} \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & \frac{5}{4} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \left(A \text{ column reduces to } \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \frac{9}{2} & -3 & \frac{5}{4} & 0 & 0 \end{bmatrix}. \right)$$

What are $dim(Col\,A)$, $dim(Nul\,A)$, $dim(Row\,A)$, $dim(Nul\,A^T)$? Can you find bases for each subspace? How could you find these dimensions in general ?

dim (col A) = 3 = # of pivot cols in rref(A) — because corresponding cols in A $(\vec{a}_1, \vec{a}_2, \vec{a}_5)$ are a pretty good basis for col A

dim (Nul A) = 2 = # of non-pivot cols
= # of free variables in solns $\vec{x}$ to $A\vec{x} = \vec{0}$
= # of basis vectors in Nul A

dim (row A) = 3 = # of pivot rows in rref(A) (same as # pivot cols, i.e # pivots)

dim (Nul A^T) = 1 = # of non-pivot cols in rref(A^T)

$T(x) = A\vec{x}$

$T: \mathbb{R}^5 \to \mathbb{R}^4$

Row A, Nul A $\subset \mathbb{R}^5$; Col A, Nul A^T $\subset \mathbb{R}^4$

Row A = col A^T

$S(\vec{y}) = A^T\vec{y}$

$S: \mathbb{R}^4 \to \mathbb{R}^5$

A is $m \times n$

$T: \mathbb{R}^n \to \mathbb{R}^m$ $\quad T(x) = Ax$

$S: \mathbb{R}^m \to \mathbb{R}^n$ $\quad S(\vec{y}) = A^T\vec{y}$

We figured out that in general A has m rows & n colums

dim Col A = dim Row A — this is called the <u>rank</u> of the matrix

dim Col A + dim Nul A = n (# of columns) (dim of domain $\mathbb{R}^n$)

dim Row A + dim Nul A^T = m (# of rows) (dim of codomain)

Geometry connected to the four fundamental subspaces:

- First, recall the geometry fact that the dot product of two vectors in $\mathbb{R}^n$ is zero if and only if the vectors are perpendicular, i.e.

$$\underline{\boldsymbol{u}} \cdot \underline{\boldsymbol{v}} = 0 \quad \text{if and only if } \underline{\boldsymbol{u}} \perp \underline{\boldsymbol{v}}.$$

(Well, we really only know this in $\mathbb{R}^2$ or $\mathbb{R}^3$ so far, from multivariable Calculus class. But it's true for all $\mathbb{R}^n$, as we'll see in Chapter 6.) So for a vector $\underline{\boldsymbol{x}} \in Nul\, A$ we can interpret the equation

$$A\, \underline{\mathbf{x}} = \underline{\mathbf{0}}$$

as saying that $\underline{\boldsymbol{x}}$ is perpendicular to every row of A. Because the dot product distributes over addition, we see that each $\underline{\boldsymbol{x}} \in Nul\, A$ is perpendicular to every linear combination of the rows of A, i.e. to all of $Row\, A$:

$$Row\, A \perp Nul\, A.$$

And analogously,

$$Col\, A = Row\, A^T \perp Nul\, A^T$$

small example.

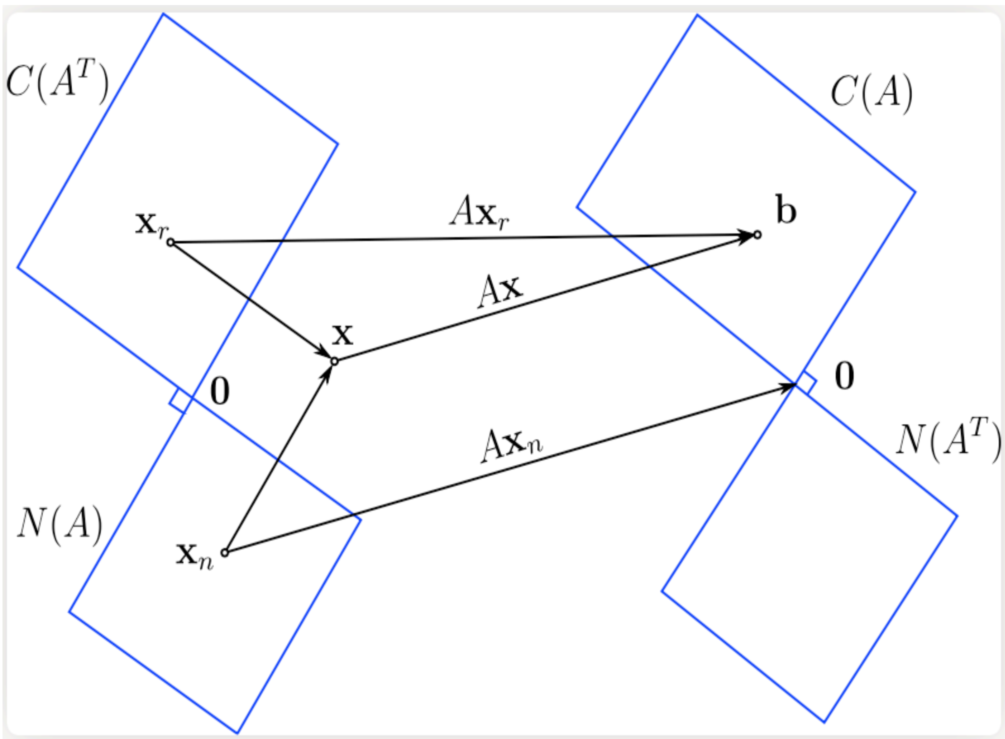
$$T : \mathbb{R}^2 \to \mathbb{R}^2$$

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$S\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

Here's a general schematic of what's going on, stolen from the internet. The web site I stole it from looks pretty good....

http://www.itshared.org/2015/06/the-four-fundamental-subspaces.html



More details on the decompositions .... we'll cover this in more detail in Chapter 6, but here's what true: In the domain $\mathbb{R}^n$, the two subspaces associated to A are *Row A* and *Nul A*. Notice that the only vector in their intersection is the zero vector, since

$$\underline{x} \in Row\ A \cap Nul\ A \quad \Rightarrow \underline{x} \cdot \underline{x} = 0 \quad \Rightarrow \underline{x} = \underline{0}.$$

So, let

$$\{\underline{u}_1, \underline{u}_2, \ \dots \underline{u}_r\} \quad \text{be a basis for} \quad Row\ A$$

$$\{\underline{v}_1, \underline{v}_2, \ \dots \underline{v}_{n-r}\} \quad \text{be a basis for } Nul\ A.$$

Then we can check that set of n vectors obtained by taking the union of the two sets,

$$\{\underline{u}_1, \underline{u}_2, \ \dots \underline{u}_r, \underline{v}_1, \underline{v}_2, \ \dots \underline{v}_{n-r}\}$$

is actually a basis for $\mathbb{R}^n$. This is because we can show that the n vectors in the set are linearly independent, so they automatically span $\mathbb{R}^n$ and are a basis: To check independence, let

$$c_1\underline{u}_1 + c_2\underline{u}_2 + \ \dots + c_r\underline{u}_r + d_1\underline{v}_1 + d_2\underline{v}_2 + \ \dots + d_{n-r}\underline{v}_{n-r} = \underline{0}.$$

then

$$c_1\underline{u}_1 + c_2\underline{u}_2 + \ \dots + c_r\underline{u}_r = -d_1\underline{v}_1 - d_2\underline{v}_2 - \ \dots - d_{n-r}\underline{v}_{n-r}.$$

Since the vector on the left is in *Row A* and the one that it equals on the right is in *Nul A*, this vector is the zero vector:

$$c_1\underline{u}_1 + c_2\underline{u}_2 + \ \dots + c_r\underline{u}_r = \underline{0} = -d_1\underline{v}_1 - d_2\underline{v}_2 - \ \dots - d_{n-r}\underline{v}_{n-r}.$$

Since $\{\underline{u}_1, \underline{u}_2, \ \dots \underline{u}_r\}$ and $\{\underline{v}_1, \underline{v}_2, \ \dots \underline{v}_{n-r}\}$ are each linearly independent sets, we deduce from these two equations that

$$c_1 = c_2 = \dots = c_r = 0, \qquad d_1 = d_2 = \dots = d_{n-r} = 0\ .$$

Q.E.D.

So the picture on the previous page is completely general (also for the decomposition of the codomain). One can check that the transformation $T(\underline{x}) = A\,\underline{x}$ restricts to an isomorphism from *Row A* to *Col A*, because it is $1 - 1$ on these subspaces of equal dimension, so must also be onto. So, T squashes *Nul A*, and maps every translation of *Nul A* to a point in *Col A*. More precisely, Each

$$\underline{x} \in \mathbb{R}^n$$

can be written uniquely as

$$\underline{x} = \underline{u} + \underline{v} \quad \text{with } \underline{u} \in Row\ A,\ \underline{v} \in\ Nul\ A.$$

and

$$T(\underline{u} + \underline{v}) = T(\underline{u}) + T(\underline{v}) = T(\underline{u}) \in Col\ (A).$$

As sets,

$$T(\{\underline{u} + Nul\ A\}) = T(\underline{u}).$$

Tues Oct 23

- 4.9 Applications to Markov Chains

Announcements:

Warm-up Exercise:

Example of a Markov Chain:

Example 1: Consider a model of population movement between a city and its suburbs, described by the following schematic. Every year, 95% of the city dwellers remain in the city, but 5% move to the suburbs; whereas every year 3% of the suburb dwellers move to the city and 97% stay in the suburbs. Our goal is to study what happens to initial populations distributed between the city and the suburbs, over the course of many years.

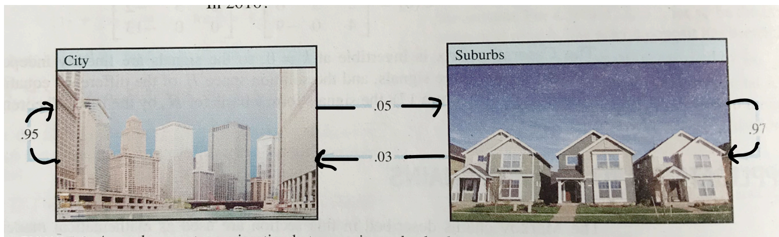


FIGURE 1 Annual percentage migration between city and suburbs.

The way the poulations evolve can be computed using the transition matrix below. The first column encodes the movement from the city to the city and suburbs; the second column encodes the movement from the suburbs to the city and suburbs.

$$M = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix}$$

Exercise 1a) If there are initially 600,000 people in the city and 400,000 people in the suburbs, how are people distributed one year later? How about 2 years later? 10 years later?

Exercise 1b) If we don't know the total population, but we know that .6 of it is initially in the city, and .4 of it is initially in the suburbs, how do we compute the fractions in each location in later years?

The previous example is encompassed in the following framework:

Definitions

a) A vector in $\mathbb{R}^n$ with non-negative entries which add up to 1 is called a *probability vector*.

b) A square matrix $P = [\underline{\boldsymbol{p}}_1\ \underline{\boldsymbol{p}}_2\ \ \dots\ \underline{\boldsymbol{p}}_n]$ (in column form) is called a *stochastic matrix* if each of its columns $\underline{\boldsymbol{p}}_1,\ \underline{\boldsymbol{p}}_2,\ \dots\ \underline{\boldsymbol{p}}_n$ is a probability vector.

c) A *Markov chain* is a sequence of probability vectors $\{\underline{\boldsymbol{x}}_0, \underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, \ \dots.\ \underline{\boldsymbol{x}}_k, \ \dots.\}$ such that

$$\underline{\boldsymbol{x}}_1 = P\,\underline{\boldsymbol{x}}_0,\quad \underline{\boldsymbol{x}}_2 = P\,\underline{x}_1,\ \dots.,\ \underline{\boldsymbol{x}}_{k+1} = P\,\underline{x}_k\ ,\ \dots.$$

$$(\text{so } \underline{\boldsymbol{x}}_k = P^k\,\underline{\boldsymbol{x}}_0 \ \text{ for } k \in \mathbb{N}).$$

Markov chains arise in lots of natural situations in science, engineering, economics. We'll see tomorrow that they are connected to google page rank. Here's another example which isn't actually a realistic model, but it's the second example in the text and it is at least sort of topical.

Exercise 2 Suppose the voting results of a congressional election at a certain voting precinct are represented by a probability vector $\underline{\boldsymbol{x}} \in \mathbb{R}^3$:

$$\underline{\boldsymbol{x}} = \begin{bmatrix} \textit{fraction Democratic } (D) \\ \textit{fraction Republican } (R) \\ \textit{fraction Libertarian } (L) \end{bmatrix}.$$

Suppose we record the outcome of the election every two years using vectors of these type. Further (and this isn't very realisitic), suppose that each election only depends on the results of the preceding one, via the following diagram. Construct the stochastic matrix for the resulting Markov Chain.

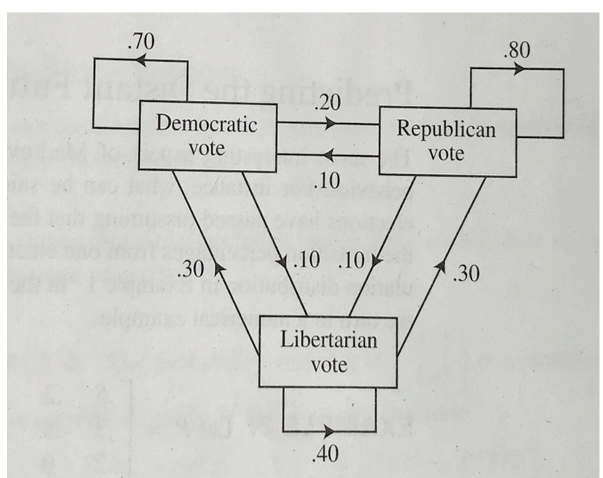


FIGURE 2 Voting changes from one election to the next.

Definition: A stochastic matrix P is called *regular* if some power of P has all positive entries (as opposed to just non-negative).

Definition: A probability vector $\underline{\boldsymbol{q}}$ is called a *steady state* vector for a Markov Chain with transition matrix P if

$$P\,\underline{\boldsymbol{q}} = \underline{\boldsymbol{q}}.$$

(Notice that in this case, if $\underline{\boldsymbol{x}}_0 = \underline{\boldsymbol{q}}$ then each $\underline{\boldsymbol{x}}_k = \underline{\boldsymbol{q}}$ as well.)

Long-time behavior of Markov chains:

Theorem (Perron-Frobenius Theorem) If P is an $n \times n$ regular stochastic matrix, then P has a unique steady state vector $\underline{\boldsymbol{q}}$. Furthermore, if $\underline{\boldsymbol{x}}_0$ is any initial state (probability vector) for the Markov chain

$$\underline{\boldsymbol{x}}_{k+1} = P\,\underline{\boldsymbol{x}}_k \qquad k = 0, 1, 2, \ldots.$$

then the Markov chain $\{\underline{\boldsymbol{x}}_k\}$ converges to the steady state $\underline{\boldsymbol{q}}$ as $k \to \infty$. In particular, since the j^{th} column of P^k is $P^k\,\underline{\boldsymbol{e}}_j$ and $\underline{\boldsymbol{e}}_j$ is an admissable initial state probability vector, each column of P^k converges to $\underline{\boldsymbol{q}}$.

Example 3 For the transition matrix from our first example of a city and its suburbs,

$$M = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix}.$$

Computations:

Input:

$$\begin{pmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{pmatrix}^{100}$$

Open code

Result:

$$\begin{pmatrix} 0.37515 & 0.37491 \\ 0.62485 & 0.62509 \end{pmatrix}$$

So approximately at least, $\underline{\boldsymbol{q}} \approx \begin{bmatrix} .375 \\ .625 \end{bmatrix}$.

Notice that in general the steady state vector $\underline{\boldsymbol{q}}$ for a regular stochastic matrix P satisfies

$$P\underline{\boldsymbol{q}} = \underline{\boldsymbol{q}}$$
$$P\underline{\boldsymbol{q}} - \underline{\boldsymbol{q}} = \underline{\mathbf{0}}$$

$$P\underline{\boldsymbol{q}} - I\underline{\boldsymbol{q}} = \underline{\mathbf{0}}$$

$$(P - I)\,\underline{\boldsymbol{q}} = \underline{\mathbf{0}}.$$

So $\underline{\boldsymbol{q}}$ is a basis vector for what must be a one-dimensional subspace given by $Nul(P - I)$. (It can only be one-dimensional since the steady state vector is unique.) Ahah! a connection to the rest of Chapter 4. :-)

Exercise 3 Find the steady state vector for the matrix

$$M = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix}$$

by finding the nullspace of

$$M - I = \begin{bmatrix} -.05 & .03 \\ .05 & -.03 \end{bmatrix}.$$

Exercise 4 What do these computations have to do with the long-time behavior of the D-R-L voting model in Exercise 2, with transition matrix

$$P = \begin{bmatrix} .7 & .1 & .3 \\ .2 & .8 & .3 \\ .1 & .1 & .4 \end{bmatrix}?$$

Input:

$$\begin{pmatrix} 0.7 & 0.1 & 0.3 \\ 0.2 & 0.8 & 0.3 \\ 0.1 & 0.1 & 0.4 \end{pmatrix}^{20}$$

Open code

Result:

$$\begin{pmatrix} 0.32145 & 0.321413 & 0.321438 \\ 0.535693 & 0.53573 & 0.535705 \\ 0.142857 & 0.142857 & 0.142857 \end{pmatrix}$$

row reduce {{-.3, .1, .3}, {.2, -.2, .3}, {.1, .1, -.6}}

Input interpretation:

row reduce $\begin{pmatrix} -0.3 & 0.1 & 0.3 \\ 0.2 & -0.2 & 0.3 \\ 0.1 & 0.1 & -0.6 \end{pmatrix}$

Result:

$$\begin{pmatrix} 1 & 0 & -2.25000000000000000000000000000 \\ 0 & 1 & -3.75000000000000000000000000000 \\ 0 & 0 & 0 \end{pmatrix}$$

Wed Oct 24

- 4.9 supplement: Google page rank

Announcements:

Warm-up Exercise:

The Giving Game: Google Page Rank

University of Utah Teachers' Math Circle

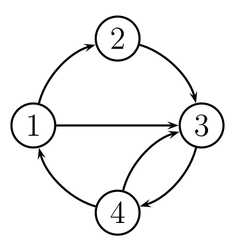
Nick Korevaar

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:

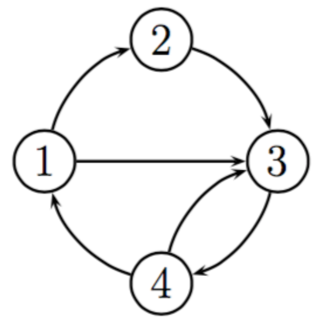


The question at the heart of the basic Google page rank algorithm is: in a voting game like this, with billions of linked web sites and some initial vote distribution, does the way the votes are distributed settle down in the limit? If so, sites with more limiting votes must ultimately be receiving a 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!

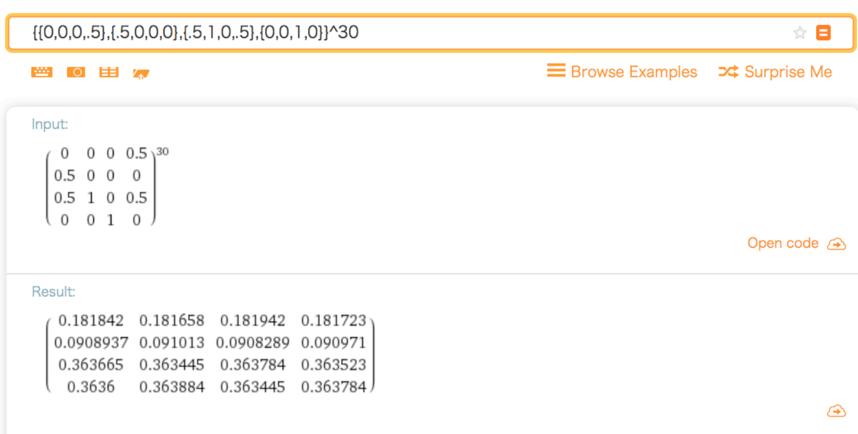
1. Decide on your initial material allocations. I recommend giving it all to one person at the start, even though that doesn't seem fair. If you're using pennies, 33 is a nice number for this template. At each stage, split your current amount into equal portions and distribute it to your friends, according 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!

Play the google game!



Transition matrix for problem **1**, to a large power:



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

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

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

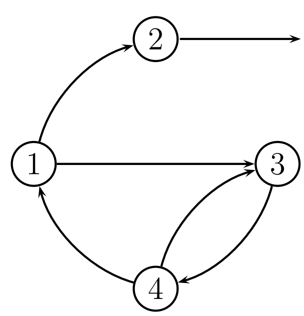
So in matrix form, $\mathbf{x}_k = A^k \mathbf{x}_0$ for the transition matrix A given above.

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

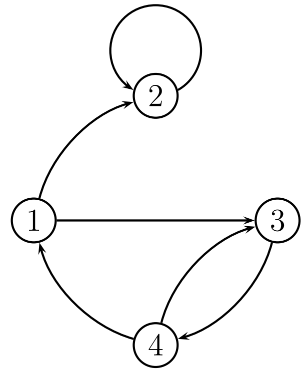
4. The limiting "fractions" in this problem really are fractions (and not irrational numbers). What are they? Is there a matrix equation you could solve to find them, for this small problem? Hint: the limiting fractions should remain fixed when you play the game.

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

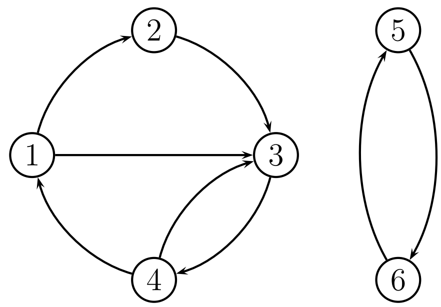
   (a)



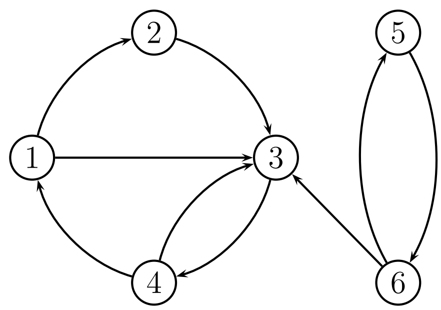
(b)



(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 $\mathbb{1}$. For (Brin an Pages' choice of) $\varepsilon = .15$ and n equal the number of nodes, consider the Google matrix

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

(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

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

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.

Fri Oct 26

- 5.1-5.2 Eigenvectors and eigenvalues for square matrices

Announcements:

Warm-up Exercise:

Eigenvalues and eigenvectors for square matrices.

The steady state vectors for stochastic matrices in section 4.9, i.e. the vectors $\underline{\boldsymbol{x}}$ with $P(\underline{\boldsymbol{x}}) = \underline{\boldsymbol{x}}$ when P is stochastic, are a special case of the concept of eigenvectors and eigenvalues for general square matrices, as we'll see below.

To introduce the idea of eigenvalues and eigenvectors we'll first think geometrically.

<u>Example</u> Consider the matrix transformation $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ with formula

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1\begin{bmatrix} 3 \\ 0 \end{bmatrix} + x_2\begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

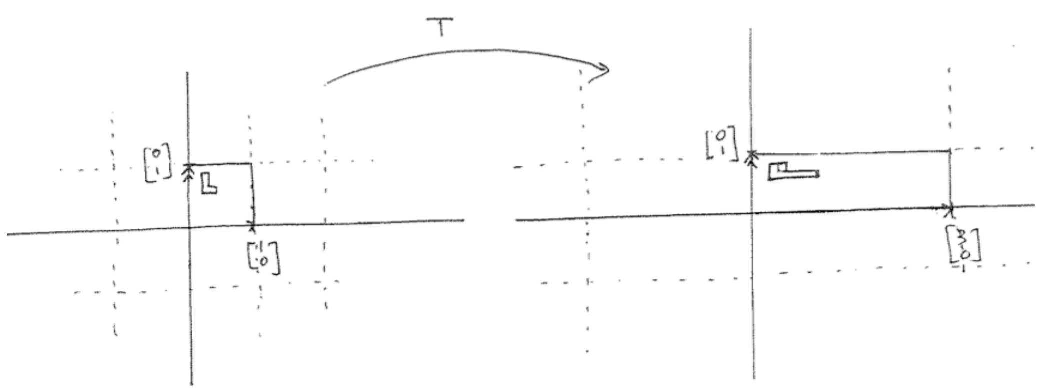
Notice that for the standard basis vectors $\underline{\boldsymbol{e}}_1 = [1, 0]^T, \underline{\boldsymbol{e}}_2 = [0, 1]^T$

$$T(\underline{\boldsymbol{e}}_1) = 3\underline{\boldsymbol{e}}_1$$
$$T(\underline{\boldsymbol{e}}_2) = \underline{\boldsymbol{e}}_2 \ .$$

The facts that T is linear and that it transforms $\underline{\boldsymbol{e}}_1, \underline{\boldsymbol{e}}_2$ by scalar multiplying them, lets us understand the geometry of this transformation completely:

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = T(x_1\underline{\boldsymbol{e}}_1 + x_2\underline{\boldsymbol{e}}_2) = x_1 T(\underline{\boldsymbol{e}}_1) + x_2 T(\underline{\boldsymbol{e}}_2)$$
$$= x_1(3\underline{\boldsymbol{e}}_1) + x_2(1\underline{\boldsymbol{e}}_2) \ .$$

In other words, T stretches by a factor of 3 in the $\underline{\boldsymbol{e}}_1$ direction, and by a factor of 1 in the $\underline{\boldsymbol{e}}_2$ direction, transforming a square grid in the domain into a parallel rectangular grid in the image:



Definition: If $A_{n \times n}$ and if $A\,\underline{\boldsymbol{v}} = \lambda\,\underline{\boldsymbol{v}}$ for a scalar λ and a vector $\underline{\boldsymbol{v}} \neq \underline{\mathbf{0}}$ then $\underline{\boldsymbol{v}}$ is called an eigenvector of A, and λ is called the eigenvalue of $\underline{\boldsymbol{v}}$. (In some texts the words characteristic vector and characteristic value are used as synonyms for these words.)

• In the example above, the standard basis vectors (or multiples of them) are eigenvectors, and the corresponding eigenvalues are the diagonal matrix entries. A non-diagonal matrix may still have eigenvectors and eigenvalues, and this geometric information can still be important to find. (For example, a stochastic matrix P always has eigenvectors with eigenvalue 1, namely the steady-state vector and its multiples are the eigenvectors. But how do you find eigenvectors and eigenvalues for general non-diagonal matrices? ...

Exercise 2) Try to find eigenvectors and eigenvalues for the non-diagonal matrix, by just trying random input vectors $\underline{\boldsymbol{x}}$ and computing $A\,\underline{\boldsymbol{x}}$.

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

How to find eigenvalues and eigenvectors (including eigenspaces) systematically:

If

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

$$\Leftrightarrow A\,\underline{\boldsymbol{v}} - \boldsymbol{\lambda}\,\underline{\boldsymbol{v}} = \underline{\mathbf{0}}$$

$$\Leftrightarrow A\,\underline{\boldsymbol{v}} - \lambda\,I\,\underline{\boldsymbol{v}} = \underline{\mathbf{0}}$$

where I is the identity matrix.

$$\Leftrightarrow (A - \lambda\,I)\underline{\boldsymbol{v}} = \underline{\mathbf{0}}\,.$$

As we know, this last equation can have non-zero solutions $\underline{\boldsymbol{v}}$ if and only if the matrix $(A - \lambda\,I)$ is not invertible, i.e.

$$\Leftrightarrow det(A - \lambda\,I) = 0\,.$$

So, to find the eigenvalues and eigenvectors of matrix you can proceed as follows:

- Compute the polynomial in λ

$$p(\lambda) = det(A - \lambda\,I)\,.$$

If $A_{n \times n}$ then $p(\lambda)$ will be degree n. This polynomial is called the characteristic polynomial of the matrix A.

- λ_j can be an eigenvalue for some non-zero eigenvector $\underline{\boldsymbol{v}}$ if and only if it's a root of the characteristic polynomial, i.e. $p(\lambda_j) = 0$. For each such root, the homogeneous solution space of vectors $\underline{\boldsymbol{v}}$ solving

$$(A - \lambda_j\,I)\underline{\boldsymbol{v}} = \underline{\mathbf{0}}$$

will be eigenvectors with eigenvalue λ_j. This subspace of eigenvectors will be at least one dimensional, since $(A - \lambda_j\,I)$ does not reduce to the identity and so the explicit homogeneous solutions will have free parameters. Find a basis of eigenvectors for this subspace. Follow this procedure for each eigenvalue, i.e. for each root of the characteristic polynomial.

Notation: The subspace of eigenvectors for eigenvalue λ_j is called the λ_j eigenspace, and we'll denote it by $E_{\lambda = \lambda_j}$. The basis of eigenvectors is called an eigenbasis for E_{λ_j}.

Exercise 3) a) Use the systematic algorithm to find the eigenvalues and eigenbases for the non-diagonal matrix of Exercise 2.

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

b) Use your work to describe the geometry of the linear transformation in terms of directions that get scaled:

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Exercise 4) Find the eigenvalues and eigenspace bases for

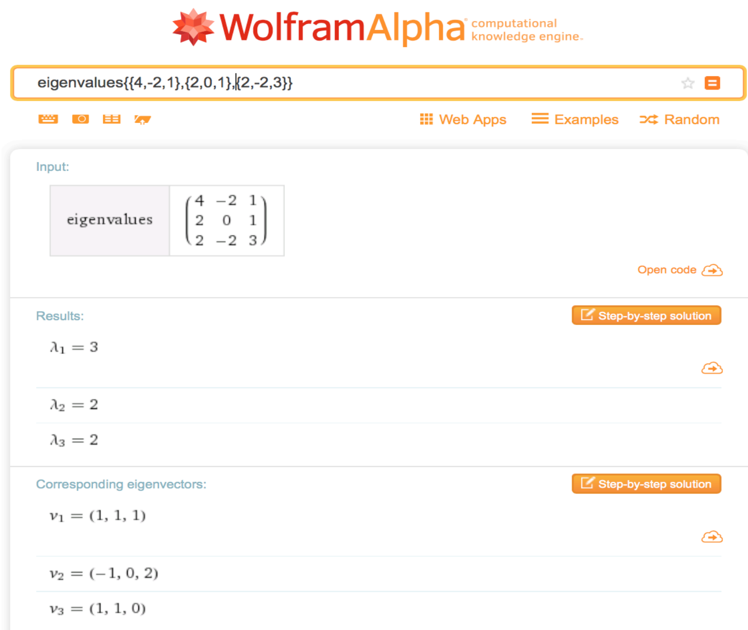
$$B := \begin{bmatrix} 4 & -2 & 1 \\ 2 & 0 & 1 \\ 2 & -2 & 3 \end{bmatrix}.$$

(i) Find the characteristic polynomial and factor it to find the eigenvalues.

(ii) for each eigenvalue, find bases for the corresponding eigenspaces.

(iii) Can you describe the transformation $T(\underline{\mathbf{x}}) = B\underline{\mathbf{x}}$ geometrically using the eigenbases? Does $det(B)$ have anything to do with the geometry of this transformation?

Your solution will be related to the output below:



In all of our new examples so far, it turns out that by collecting bases from each eigenspace for the matrix $A_{n \times n}$, and putting them together, we get a basis for $\mathbb{R}^n$. This lets us understand the <u>geometry</u> of the transformation

$$T(\underline{\mathbf{x}}) = A\,\underline{\mathbf{x}}$$

almost as well as if A is a diagonal matrix. This is actually something that does not always happen for a matrix A. When it does happen, we say that A is <u>diagonalizable.</u> Here's an example of a matrix which is NOT diagonalizable:

<u>Exercise 5</u>: Find matrix eigenvalues and eigenspace basis for each eigenvalue, for

$$A = \begin{bmatrix} 3 & 2 \\ 0 & 3 \end{bmatrix}.$$

Explain why there is no basis of $\mathbb{R}^2$ consisting of eigenvectors of A.