Announcements

November 2

- Read the last two slides of 10/31, or the rest of §5.3 in Lay.
  - You'll be responsible for knowing about non-distinct eigenvalues, but we won't cover it in class.
- WeBWorK assignment 5.2 is due on Friday at 6am.
- Your midterms will be returned to you in recitation on Friday.
  - The solutions are posted online.
  - The score breakdown is on Piazza.
- Midterm 3 will take place in recitation on Friday, 11/18.
- Office hours: today 1–2pm, tomorrow 3:30–4:30pm, and by appointment, in Skiles 221.
  - As always, TAs' office hours are posted on the website.
  - Math Lab is also a good place to visit.

Application

Stochastic matrices

Stochastic Matrices

Definition

A square matrix A is **stochastic** if all of its entries are nonnegative, and the sum of the entries of each column is 1. We say A is **positive** if all of its entries are positive.

These arise very commonly in modeling of probabalistic phenomena (Markov chains).

You'll be responsible for knowing basic facts about stochastic matrices and the Perron–Frobenius theorem, but we will not cover them in depth. These slides are the primary reference; also see §4.9 in Lay.

Stochastic Matrices

Example

Red Box has kiosks all over where you can rent movies. You can return them to any other kiosk. Let A be the matrix whose ij entry is the probability that a customer renting a movie from location j returns it to location i. For example, if there are three locations, maybe

$$A = \begin{pmatrix} .3 & .4 & .5 \\ .3 & .4 & .3 \\ .4 & .2 & .2 \end{pmatrix}.$$

30% probability a movie rented from location 3 gets returned to location 2

The columns sum to 1 because there is a 100% chance that the movie will get returned to *some* location. This is a positive stochastic matrix.

Note that, if $v = (x, y, z)$ represents the number of movies at the three locations, then (assuming the number of movies is large), Red Box will have approximately

$$Av = A \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} .3x + .4y + .5z \\ .3x + .4y + .3z \\ .4x + .2y + .2z \end{pmatrix}$$

movies in its three locations the next day. The *total number* of movies doesn't change because the columns sum to 1.

Eigenvalues of Stochastic Matrices

Fact: 1 is an eigenvalue of a stochastic matrix.

Why? If A is stochastic, then 1 is an eigenvalue of A^T:

$$\begin{pmatrix} .3 & .3 & .4 \\ .4 & .4 & .2 \\ .5 & .3 & .2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Lemma

A and A^T have the same eigenvalues.

Proof: $\det(A - \lambda I) = \det\big((A - \lambda I)^T\big) = \det(A^T - \lambda I)$, so they have the same characteristic polynomial.

Note: this doesn't give a new procedure for finding an eigenvector with eigenvalue 1; it only shows one exists.

Eigenvalues of Stochastic Matrices
Continued

Fact: if λ is an eigenvalue of a stochastic matrix, then $|\lambda| \leq 1$. Hence 1 is the *largest* eigenvalue (in absolute value).

Why? If λ is an eigenvalue of A then it is an eigenvalue of A^T.

$$\text{eigenvector } v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \lambda v = A^T v \implies \lambda x_j = \overset{j\text{th entry of } A^T v}{\boxed{\textstyle\sum_{i=1}^n a_{ij} x_i}}.$$

Choose x_j with the largest absolute value, so $|x_i| \leq |x_j|$ for all i.

$$|\lambda| \cdot |x_j| = \left| \sum_{i=1}^n a_{ij} x_i \right| \leq \sum_{i=1}^n \overset{\text{positive}}{a_{ij}} \cdot |x_i| \leq \sum_{i=1}^n a_{ij} \cdot \underset{\geq |x_i|}{|x_j|} = \overset{= \sum_i a_{ij}}{1} \cdot |x_j|,$$

so $|\lambda| \leq 1$.

Better fact: if $\lambda \neq 1$ is an eigenvalue of a positive stochastic matrix, then $|\lambda| < 1$.

Diagonalizable Stochastic Matrices

The Red Box matrix $A = \begin{pmatrix} .3 & .4 & .5 \\ .3 & .4 & .3 \\ .4 & .2 & .2 \end{pmatrix}$ has characteristic polynomial

$$f(\lambda) = -\lambda^3 + 0.12\lambda - 0.02 = -(\lambda - 1)(\lambda + 0.2)(\lambda - 0.1).$$

So 1 is indeed the largest eigenvalue. Since A has 3 distinct eigenvalues, it is diagonalizable:

$$A = P \begin{pmatrix} 1 & 0 & 0 \\ 0 & .1 & 0 \\ 0 & 0 & -.2 \end{pmatrix} P^{-1}.$$

Hence it is easy to compute the powers of A:

$$A^n = P \begin{pmatrix} 1 & 0 & 0 \\ 0 & .1^n & 0 \\ 0 & 0 & (-.2)^n \end{pmatrix} P^{-1}.$$

Let w_1, w_2, w_3 be the columns of P, i.e. the eigenvectors of P with respective eigenvalues $1, .1, -.2$. Let $\mathcal{B} = \{w_1, w_2, w_3\}$. Recall that $P[x]_{\mathcal{B}} = x$ and $[x]_{\mathcal{B}} = P^{-1}x$ for any vector x.

Diagonalizable Stochastic Matrices

Continued

$$[x]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \implies A^n x = P \begin{pmatrix} 1 & 0 & 0 \\ 0 & .1^n & 0 \\ 0 & 0 & (-.2)^n \end{pmatrix} P^{-1} x$$

$$= P \begin{pmatrix} 1 & 0 & 0 \\ 0 & .1^n & 0 \\ 0 & 0 & (-.2)^n \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

$$= P \begin{pmatrix} c_1 \\ .1^n c_2 \\ (-.2)^n c_3 \end{pmatrix} = c_1 w_1 + .1^n c_2 w_2 + (-.2)^n c_3 w_3.$$

As n becomes large, this approaches $c_1 w_1$, which is an *eigenvector with eigenvalue* 1 (assuming $c_1 \neq 0$).

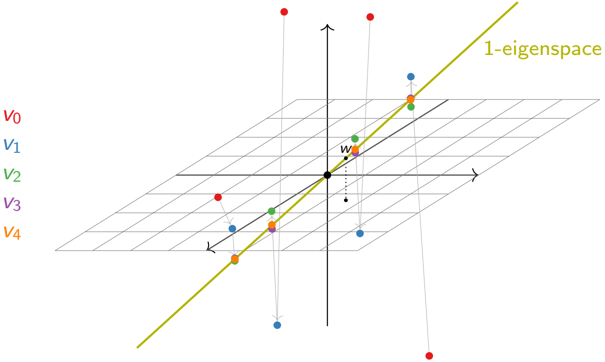
So all vectors get sucked into the 1-eigenspace, which is spanned by

$$w = w_1 = \begin{pmatrix} 0.3889 \\ 0.3333 \\ 0.2778 \end{pmatrix}.$$

Diagonalizable Stochastic Matrices

Picture

Start with a vector v_0 (the number of movies on the first day), let $v_1 = Av_0$ (the number of movies on the second day), let $v_2 = Av_1$, etc.



So v_n approaches an eigenvector with eigenvalue 1 as n gets large.

If A is the Red Box matrix, and v_n is the vector representing the number of movies in the three locations on day n, then

$$v_{n+1} = Av_n.$$

For any starting distribution v_0 of videos in red boxes, after enough days, the distribution v ($= v_n$ for n large) is an eigenvector with eigenvalue 1:

$$Av = v.$$

Moreover, we know *which* eigenvector it is: it is the multiple of $w \sim (0.39, 0.33, 0.28)$ that represents the same number of videos as in v_0. (Remember the total number of videos never changes.)

Presumably, Red Box really does have to do this kind of analysis to determine how many videos to put in each box.

Steady State

Definition

A *steady state* for a stochastic matrix A is an eigenvector w with eigenvalue 1, such that all entries are *positive* and sum to 1.

Perron–Frobenius Theorem

If A is a positive stochastic matrix, then it admits a unique steady state vector w. Moreover, for any vector v_0 with entries summing to some number c, the iterates $v_1 = Av_0$, $v_2 = Av_1$, $\ldots$, approach cw as n gets large.

The fact that A has an eigenvector with eigenvalue 1 and having *positive* entries is very special!

For the Red Box matrix, the steady state was the vector $w \sim (0.39, 0.33, 0.28)$, and if you start with 100 total movies, eventually you'll have $100w = (39, 33, 28)$ movies in the three locations.

The Theorem says that our analysis of the Red Box matrix works for *any* positive stochastic matrix — whether or not it is diagonalizable!

Google's PageRank

Early internet searching was a pain. Yahoo would scan pages for your search text, and just list the results with the most occurrences of those words.

Not surprisingly, the more unsavory websites soon learned that by putting the word "internet" a million times in their pages, they could show up first in every search for the word "internet".

Larry Page and Sergey Brin invented a way to rank pages by *importance*. They founded Google based on their algorithm.

Here's how it works. (roughly)

Reference:

http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/lecture3.html

The Importance of Being Popular

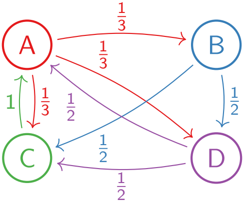
Idea: each webpage has an associated importance, or **rank**. This is a positive number. If page P links to n other pages $Q_1, Q_2, \ldots, Q_n$, then each Q_i should inherit $\frac{1}{n}$ of P's importance.

- So if a very important page links to your webpage, your webpage is considered important.
- And if a ton of unimportant pages link to your webpage, then it's still important.
- But if only one crappy site links to yours, your page isn't important.

Random surfer interpretation: a "random surfer" just sits at his computer all day, randomly clicking on links. The pages he spends the most time on should be the most important. This turns out to be equivalent to the rank.

The Importance Matrix

Consider the following Internet with only four pages. Links are indicated by arrows.



Page A has 3 links, so it passes $\frac{1}{3}$ of its importance to pages B, C, D.
Page B has 2 links, so it passes $\frac{1}{2}$ of its importance to pages C, D.
Page C has one link, so it passes all of its importance to page A.
Page D has 2 links, so it passes $\frac{1}{2}$ of its importance to pages A, C.

In terms of matrices, if $v = (a, b, c, d)$ is the vector containing the ranks a, b, c, d of the pages A, B, C, D, then

importance matrix: ij entry is importance page j passes to page i

$$\begin{pmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} c + \frac{1}{2}d \\ \frac{1}{3}a \\ \frac{1}{3}a + \frac{1}{2}b + \frac{1}{2}d \\ \frac{1}{3}a + \frac{1}{2}b \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

The 25 Billion Dollar Eigenvector

Observations:

- The importance matrix is a stochastic matrix! The columns each contain $1/n$ ($n =$ number of links), n times.
- The rank vector is an eigenvector with eigenvalue 1!

Random surfer interpretation: If a random surfer has probability (a, b, c, d) to be on page A, B, C, D, respectively, then after clicking on a random link, the probability he'll be on each page is

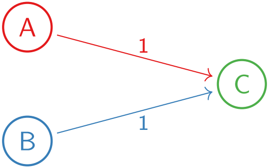
$$\begin{pmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} c + \frac{1}{2}d \\ \frac{1}{3}a \\ \frac{1}{3}a + \frac{1}{2}b + \frac{1}{2}d \\ \frac{1}{3}a + \frac{1}{2}b \end{pmatrix}.$$

The rank vector is a *steady state* for the random surfer: it's the probability vector (a, b, c, d) such that, after clicking on a random link, he'll have the *same probability* of being on each page.

So, the important (high-ranked) pages are those where a random surfer will end up most often.

Observation: the importance matrix is *not* positive: it's only nonnegative. So we can't apply the Perron–Frobenius theorem. Does this cause problems? Yes!

Consider the following Internet:



The importance matrix is $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}$. This has characteristic polynomial

$$f(\lambda) = \det \begin{pmatrix} -\lambda & 0 & 0 \\ 0 & -\lambda & 0 \\ 1 & 1 & -\lambda \end{pmatrix} = -\lambda^3.$$

So 1 is not an eigenvalue at all: there is no rank vector! (It is not stochastic.)

Consider the following Internet:

The importance matrix is $\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$. This has linearly independent eigenvectors $\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$, both with eigenvalue 1. So there is more than one rank vector!

The Google Matrix

Here is Page and Brin's solution. Fix p in $(0,1)$, called the **damping factor**. (A typical value is $p = 0.15$.) The Google matrix is

$$M = (1-p)\cdot A + p\cdot B \quad \text{where} \quad B = \frac{1}{N}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix},$$

N is the total number of pages, and A is the importance matrix.

In the random surfer interpretation, this matrix M says: with probability p, our surfer will surf to a completely random page; otherwise, he'll click a random link.

Lemma

The Google matrix is a positive stochastic matrix.

Hence by the Perron–Frobenius theorem, there is a *unique* eigenvector with eigenvalue 1. It has positive entries. This is the PageRank vector!

The hard part is calculating it: Mathematica doesn't like matrices with dimensions (1 gazillion)$\times$(1 gazillion).