Stochastic Matrices

This is a special kind of difference equation in which the state change matrix encodes probabilities.

Red Box Example:

Pretend there are 3 Red Box kiosks in Durham, and that everyone who rents Prognosis Negative today will return it tomorrow. Suppose that someone from kiosk i will return to kiosk j with the following probabilities:

	Renting 1	Renting 2	Renting 3
Returning 1	30%	40%	50%
Returning 2	30%	40%	30%
Returning 3	40%	20%	20%

If $v_k = \begin{pmatrix} x_k \\ y_k \\ z_k \end{pmatrix} = \#$ movies in kiosk $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ on day k then

$$\begin{aligned} x_{k+1} &= .3x_k + .4y_k + .5z_k \\ + \; y_{k+1} &= .3x_k + .4y_k + .3z_k \\ + \; z_{k+1} &= .4x_k + .2y_k + .3z_k \\ \hline x_{k+1}+y_{k+1}+z_{k+1} &= x_k + y_k + z_k \end{aligned}$$

$$\rightsquigarrow \quad v_{k+1} = \underbrace{\begin{pmatrix} .3 & .4 & .5 \\ .3 & .4 & .3 \\ .4 & .2 & .2 \end{pmatrix}}_{A} v_k$$

Note the columns of A sum to 1 because we're assuming every movie has a 100% chance of being returned somewhere.

$\rightarrow$ this means the total # movies won't change.

Def: A square matrix is **stochastic** if its entries are nonnegative & the entries in each column sum to 1. A stochastic matrix is **positive** if all entries are positive (i.e., nonzero)

Eg: positive stochastic

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

stochastic

$$\begin{pmatrix} .6 & .4 & .5 \\ 0 & .4 & .3 \\ .4 & .2 & .2 \end{pmatrix}$$

not stochastic

$$\begin{pmatrix} .6 & .4 & .5 \\ -.1 & .4 & .3 \\ .5 & .2 & .2 \end{pmatrix}$$

not stochastic

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

NB: Columns sum to 1 means $A^T\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix} = \begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}$:

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

sum of col 1

$$A^T\begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix} .3+.3+.4 \\ .4+.4+.2 \\ .5+.3+.2 \end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix}$$

Fact: If A is stochastic then 1 is an eigenvalue.

$\rightarrow$ $\det(A-\lambda I_n) = \det[(A-\lambda I_n)^T] = \det(A^T - \lambda I_n^T)$
$= \det(A^T - \lambda I_n)$

(HW)

So A & A^T have the same eigenvalues, and $\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$ is a 1-eigenvector of A^T.

Fact: If λ is an eigenvalue of a stochastic matrix then $|\lambda| \le 1$.

Why? λ is also an eigenvalue of A^T.

Let v be an eigenvector: $A^T v = \lambda v$

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \rightsquigarrow A^T = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$$

$$v = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \lambda x_3 \end{pmatrix} = A^T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{21}x_2 + a_{31}x_3 \\ a_{12}x_1 + a_{22}x_2 + a_{32}x_3 \\ a_{13}x_1 + a_{23}x_2 + a_{33}x_3 \end{pmatrix}$$

Suppose $|x_1| \ge |x_2|$ and $|x_1| \ge |x_3|$

(choose the coordinate with largest abs. value)

1^{st} coordinate:

$$\lambda x_1 = a_{11}x_1 + a_{21}x_2 + a_{31}x_3$$

$$\Rightarrow |\lambda| \cdot |x_1| = |a_{11}x_1 + a_{21}x_2 + a_{31}x_3|$$

$$\le \overset{\ge 0}{a_{11}}|x_1| + a_{21}\overset{\le |x_1|}{|x_2|} + a_{31}|x_3|$$

$$\le (a_{11} + a_{21} + a_{31})|x_1| = |x_1|$$

$$\Rightarrow |\lambda| \le 1 \quad \checkmark$$

Better Fact: If $\lambda \neq 1$ is an eigenvalue of a positive stochastic matrix then $|\lambda| < 1$.

(so 1 is the dominant eigenvalue)

Eg: The Red Box matrix has characteristic polynomial

$$p(\lambda) = -\lambda^3 + .9\lambda + 0.12\lambda - 0.02$$
$$= -(\lambda-1)(\lambda+0.2)(\lambda-0.1)$$

Eigenvals are $1, -0.2, 0.1$

and $|-0.2|<1$, $|0.1|<1$

In this case, there are 3 (different) eigenvalues, so the matrix is diagonalizable. In fact, the eigenvectors are

$$1: \ w_1 = \begin{pmatrix} 7 \\ 6 \\ 5 \end{pmatrix} \qquad -0.2: \ w_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \qquad 0.1: \ w_3 = \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix}$$

Suppose you start with $v_0 = \begin{pmatrix} 48 \\ 36 \\ 42 \end{pmatrix}$ movies.

Expand in the eigenbasis:

$$v_0 = x_1 w_1 + x_2 w_2 + x_3 w_3 \rightsquigarrow x_1 = 7 \quad x_2 = 3 \quad x_3 = 2$$
$$\Rightarrow v_0 = 7w_1 + 3w_2 + 2w_3$$

Solve the difference equation:

$$v_k = A^k v_0 = (1)^k 7w_1 + (-0.2)^k 3w_2 + (0.1)^k 2w_3$$
$$\xrightarrow[k\to\infty]{} 7w_1 = \begin{pmatrix} 49 \\ 42 \\ 35 \end{pmatrix}$$

Observation 1:

if $v_0 = x_1 w_1 + x_2 w_2 + x_3 w_3$

then $v_k = x_1 w_1 + (-0.2)^k x_2 w_2 + (0.1)^k x_3 w_3$

$$\xrightarrow[k\to\infty]{} x_1 w_1 \quad (\text{if } x_1 \neq 0)$$

So v_k converges to a 1-eigenvector [demo]

Observation 2:

Since the total # movies doesn't change, we even knew which eigenvector: it's the multiple of w_1 whose entries have the same sum as v_0.

In our case, we started with

$$v_0 = \begin{pmatrix} 48 \\ 36 \\ 42 \end{pmatrix} \to 126 \text{ total movies}$$

The sum of the entries of $w_1 = \begin{pmatrix} 7 \\ 6 \\ 5 \end{pmatrix}$ is 18, so the sum of the entries of $\frac{126}{18} w_1 = 7 w_1$ is 126,

$$\text{so } v_k \xrightarrow[k\to\infty]{} 7 w_1 = \begin{pmatrix} 49 \\ 42 \\ 35 \end{pmatrix} \checkmark$$

→ This would've been easier if we'd replaced w_1 by $\frac{1}{18} w_1$ to assume the entries of w_1 sum to 1.

Observation 3:

The coordinates of $w_1 = \begin{pmatrix} 7 \\ 6 \\ 5 \end{pmatrix}$ are positive numbers

It's good they're not negative — that would mean negative movies in some kiosk!

These observations turn out to hold for any positive stochastic matrix, even if it's not diagonalizable.

Perron-Frobenius Theorem: If A is a positive stochastic matrix, then there is a unique 1-eigenvector w with positive coordinates summing to 1.

If v_0 is a vector with coordinates summing to c, then $v_k = A^k v_0 \xrightarrow[k\to\infty]{} c \cdot w$.

Def: The 1-eigenvector of a positive stochastic matrix whose coordinates sum to 1 is the steady state of that matrix.

This is easy to compute!

$\rightarrow$ Find a 1-eigenvector $v \in \mathrm{Nul}(A - I_n)$

$\rightarrow$ $w = \dfrac{v}{\text{sum of coords of } v}$

So the steady state of the Red Box matrix is

$$w = \frac{1}{18}\begin{pmatrix}7\\6\\5\end{pmatrix}.$$

Positive Stochastic Matrices: Summary

If A is positive stochastic, then:

- The 1-eigenspace of A is a line.
- There is a 1-eigenvector with positive coordinates.

  Divide by the sum of the coordinates $\rightsquigarrow$
- There is a unique 1-eigenvector w with positive coordinates summing to 1
- $|\lambda| < 1$ for all other eigenvalues, so 1 is the dominant eigenvalue.
- If v_0 is any vector then

  $$v_k = A^k v_0 \xrightarrow[k\to\infty]{} c \cdot w$$
- The scalar multiple c is the sum of the coordinates of v_0

  (the total # movies doesn't change.)

Google's PageRank

or, how Larry Page & Sergei Brin used linear algebra to make the internet searchable.

Idea: each web page has an "importance", or rank. This is a positive number. If page P links to n other pages $Q_1, \dots, Q_n$, then each Q_i inherits $\frac{1}{n}$ of P's importance.

→ so if an important page links to your page, then your page is important too.

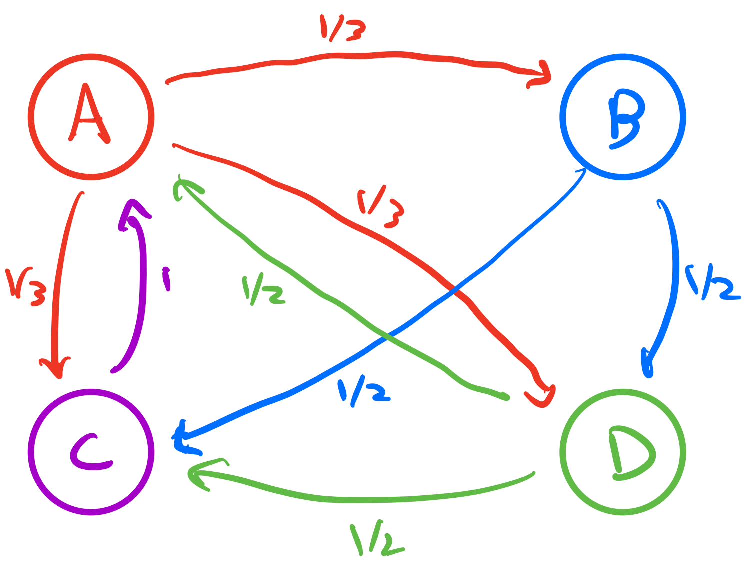
→ or, if a million unimportant pages link to your pages then your page is important.

→ but if only one crappy page links to you, then your page is not important.

Random surfer interpretation:

The random surfer sits at his computer all day clicking links at random. The pages he visits most often are the most important in the above sense, as it turns out.

Eg: Here's an internet with 4 pages. Links are indicated by arrows.



- Page A has 3 links
  ⇝ passes 1/3 of its importance to B C D
- Page B has 2 links
  ⇝ passes 1/2 of its importance to C D
- Page C has 1 link
  ⇝ passes all of its importance to A
- Page D has 2 links
  ⇝ passes 1/2 of its importance to A C

So if the pages have importance a b c d then

$$\begin{aligned} a &= c + \tfrac{1}{2}d \\ b &= \tfrac{1}{3}a \\ c &= \tfrac{1}{3}a + \tfrac{1}{2}b + \tfrac{1}{2}d \\ d &= \tfrac{1}{3}a + \tfrac{1}{2}b \end{aligned} \quad \rightsquigarrow \quad \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

importance matrix

Observation:

- → The importance matrix is stochastic (columns sum to 1: eg. A has 3 links, each with importance $1/3$) (unless there's a page with no links...)
- → The rank vector is an eigenvector with eigenvalue 1 (the $25 billion eigenvector)

In this case, the 1-eigenspace is spanned by

$$w = \frac{1}{31}\begin{pmatrix} 12 \\ 4 \\ 9 \\ 6 \end{pmatrix} \rightsquigarrow \quad a = \frac{12}{31} \quad c = \frac{9}{31} \quad b = \frac{4}{31} \quad d = \frac{6}{31}$$

(normalize so they sum to 1).

$\rightsquigarrow$ A is most important!

Random Surfer Interpretation:

If the random surfer has probabilities (a, b, c, d) of being on pages A B C D, then after the next click he has probabilities

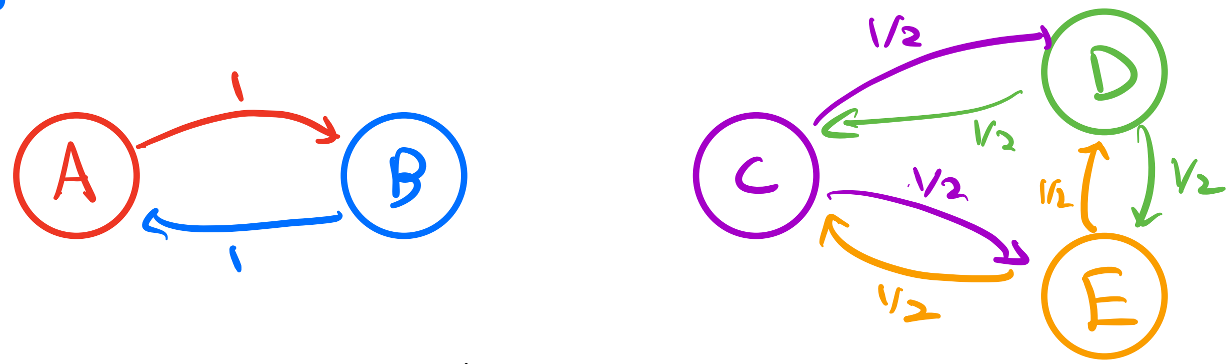
$$\begin{pmatrix} c + \frac{1}{2}d \\ \frac{1}{3}a \\ \frac{1}{3}a + \frac{1}{2}b + \frac{1}{2}d \\ \frac{1}{3}a + \frac{1}{2}b \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

of being on each page.

So the rank vector is the steady state for the random surfer $\rightsquigarrow$ spends more time on important pages.

Observation: this importance matrix is usually stochastic but not positive stochastic, so we can't apply Perron-Frobenius. Does this cause problems? Yes!

Eg (Disconnected Internet): Consider this Internet:



Importance matrix:

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}$$

But both $(1,1,0,0,0)$ and $(0,0,1,1,1)$ are 1-eigenvectors: rank vector is not unique!

The Google Matrix

Page & Brin's solution is as follows.

Fix a damping factor $p \in (0,1)$ (eg. $p = 0.15$).

Let A be the importance matrix and let

$$B = \frac{1}{N}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} \qquad N = \#\text{ pages} \quad (N \times N)$$

The **Google Matrix** is

$$G = (1-p)A + pB$$

Eg: in the disconnected internet example,

$$G = (1-p)\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix} + p\begin{pmatrix} 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \end{pmatrix}$$

Fact: the Google matrix is positive stochastic.

→ **stochastic:** the cols of $(1-p)A$ sum to $1-p$
the cols of pB sum to p

$\Rightarrow$ cols of G sum to 1

→ **positive:** because pB has positive entries.

Random Surfer Interpretation:

With probability p, the random surfer navigates to a random page anywhere on the Internet; he clicks on a random link otherwise.

Def: The $\underbrace{\text{Page}}_{\text{Larry Page}}$Rank vector is the steady state of the Google matrix.

So the importance of a page is the value of its coordinate!