Last time: 3) Power method for computing eigenvalues

Goal: Find a simple eigenvalue/eigenvector pair. (This method gives the eigenvalue with biggest absolute value and the corresponding eigenvector.)

Setting: Let $A$ be an $n \times n$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ and corresponding eigenvectors $v_1, \ldots, v_n$ such that:

1. $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n|$.
2. $\{v_1, \ldots, v_n\}$ is an eigenbasis;
3. $|v_j|_2 = 1$ for all $j \in \{1, \ldots, n\}$.
4. $\mathbb{R}_+ = \text{just for simplicity}$.

Idea: Use that $\lambda_1$ is in absolute value bigger than all the other eigenvalues.

Power method: Pick an arbitrary $x_0 \in \mathbb{R}^n$ with $x_0 \neq 0$. Then $x_0 = c_1 v_1 + \cdots + c_n v_n$, for some coefficients $c_1, \ldots, c_n \in \mathbb{R}$. Then

$A^k x_0 = c_1 \lambda_1^k v_1 + c_2 \lambda_2^k v_2 + \cdots + c_n \lambda_n^k v_n$

$= c_1 \lambda_1^k v_1 + c_2 \lambda_2^k v_2 + \cdots + c_n \lambda_n^k v_n$

$= \lambda_1^k (c_1 v_1 + c_2 \frac{\lambda_2}{\lambda_1} v_2 + \cdots + c_n \frac{\lambda_n}{\lambda_1} v_n)$

$= \lambda_1^k (c_1 v_1 + \epsilon_2)$

Since $|\lambda_1| > |\lambda_2|$, for all $k \in \{2, \ldots, n\}$ we have that $\epsilon_2 \to 0$ as $k \to \infty$.

Thus

$\frac{A^k x_0}{||A^k x_0||} \to v_1$ as $k \to \infty$

$\Rightarrow <v_1, Av_1> = \lambda_1 ||v_1||^2 = \lambda_1$
Last time continued: 1) Modification of Power method

Prop. Let $s \in \mathbb{R}$ and $A$ be an $n \times n$ matrix with real eigenvalues as in the setting on the last page. Then one of the following two holds:

(a) $s$ is an eigenvalue of $A$, i.e. $A - sI$ is not invertible

OR

(b) The eigenvalues of $(A - sI)^{-1}$ are exactly $\frac{1}{\lambda_j - s}$ and the corresponding eigenvectors are still $v_j$, $j = 1, \ldots, n$.

Fact. Using the Power method on $(A - sI)^{-1}$, we obtain that the dominate eigenvalue of $(A - sI)^{-1}$ is $\frac{1}{\lambda_j - s}$ if and only if $|\lambda_j - s| < |\lambda_k - s|$ for all $k \neq j$.

$\Rightarrow$ The power method gives us the eigenvector $v_j$ corresponding to this $\lambda_j$ and we can calculate the eigenvalue closest to $s$ by computing $\langle v_j, A v_j \rangle$.

Note that we can choose any $s \in \mathbb{R}$ and let this algorithm run as often as we like.

2) Recursion relations

Setting: Recursive definition of a sequence $a_n$

Goal: Find direct formula for each $a_n$ and understand the large $n$ behavior of $a_n$.

How?

i) Rewrite recursion relation as a matrix equation.

ii) Find eigenvalues/eigenvectors $\lambda$ of the resulting matrix $A$

iii) Diagonalize $A$ and use that $A^n = S \Lambda^n S^{-1}$ if $A$ is diagonalizable.

iv) Calculate the formula for the $n$-th sequence element.

v) To understand the large $n$ behavior let $n \to \infty$. 
Markov chains

Example: Consider the following directed graph

Here $P_{ij}$ are probabilities, more precisely the probabilities to transition from state $j$ to state $i$. They are called transition probabilities and have the following properties:

a) $0 \leq P_{ij} \leq 1$

b) $P_{11} + P_{21} + P_{31} = 1$

$P_{12} + P_{22} + P_{32} = 1$

$P_{13} + P_{23} + P_{33} = 1$

$\Rightarrow \sum_{i=1}^{3} P_{ij} = 1$ for each $j$

Now let $x_{ni}$ be the probability of being at state $i$ at time $n$. Then $0 \leq x_{ni} \leq 1$ and $x_{n1} + x_{n2} + x_{n3} = 1$, i.e. $\sum_{i=1}^{3} x_{ni} = 1$.

$\Rightarrow$ Define $x_n = \begin{bmatrix} x_{n1} \\ x_{n2} \\ x_{n3} \end{bmatrix}$
Definition: A vector $v$ is a state vector if $0 \leq v_j \leq 1$ for all $j$ and $\sum v_j = 1$.

Question: Given a state vector $x_n$ and transition probabilities $P_{ij}$ can we find $x_{n+1}$, i.e. the state of the system at time $n+1$?

Answer: Let's see:

$$x_{n+1,i} = x_{n,1} \cdot P_{1i} + x_{n,2} \cdot P_{2i} + x_{n,3} \cdot P_{3i}$$

$$= \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix} \begin{bmatrix} x_{n,1} \\ x_{n,2} \\ x_{n,3} \end{bmatrix} = x_n$$

So we get $x_{n+1} = P x_n$

$$\Rightarrow \quad x_n = P^n x_0$$

Remark: If there are 12 states (also called nodes) instead of 3, then

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{112} \\ P_{21} & P_{22} & \cdots & P_{212} \\ \vdots & \vdots & \ddots & \vdots \\ P_{11} & P_{21} & \cdots & P_{112} \end{bmatrix}$$

where

a) $0 \leq P_{ij} \leq 1$

b) Each column of $P$ adds up to 1:

$$\sum_{i=1}^{12} P_{ij} = 1 \quad \forall j$$

$\Rightarrow$ Such matrices are called stochastic matrices.
Back to the example from the beginning:

Q: Given an initial choice of let’s say “sightseeing location,” where will the tourist end up eventually?

This is equivalent to finding the steady state (or long-run behavior) of the system.

Example: \[ P = \begin{bmatrix} \frac{3}{14} & 0 & 0 \\ \frac{1}{8} & \frac{3}{14} & \frac{1}{4} \\ \frac{1}{8} & \frac{1}{4} & \frac{3}{14} \end{bmatrix} \] is a stochastic matrix.

Suppose the tourist starts at \( x_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \), i.e. at location 1.

After \( n=10 \) time steps: \( x_{10} = P^{10} x_0 = \begin{bmatrix} 0.06 \\ 0.47 \\ 0.47 \end{bmatrix} \)

After \( n=100 \) time steps: \( x_{100} = P^{100} x_0 = \begin{bmatrix} 0 \\ 0.5 \\ 0.5 \end{bmatrix} \)

Thus \( P^x x_0 \to \begin{bmatrix} 0 \\ 0.5 \\ 0.5 \end{bmatrix} \) as \( k \to \infty \).

Therefore the tourist ends up in either location 2 or location 3 with 50% 50% chance.

Remark: It turns out that \( P^x x_0 \to \begin{bmatrix} 0 \\ 0.5 \\ 0.5 \end{bmatrix} \) regardless what \( x_0 \) we start from.
Let $P^h x_0 \to x$ as $h \to \infty$. This means

$$
\lim_{h \to \infty} P^h x_0 = x \iff \lim_{h \to \infty} P(P^{h-1} x_0) = x
$$

$$
\iff P\left(\lim_{h \to \infty} P^{h-1} x_0\right) = x \iff P x = x
$$

Thus $x$ is an eigenvector corresponding to the eigenvalue $\lambda = 1$ of $P$.

**Def.** Given a Markov chain with transition matrix $P$ (i.e. $y_n = P y_{n-1}$ for a stochastic matrix $P$). The steady state solution of the Markov chain is the eigenvector $x$ corresponding to the eigenvalue $\lambda = 1$ normalized such that $x_1 + x_2 + \ldots + x_n = 1$.

**Facts:** Let $P$ be a stochastic matrix.

1. If $x$ is a state vector, then so is $Px$.
   
   **Proof:** Let $P$ be $n \times n$ and $\sum_j x_j = 1$, then 
   $$
   \sum_i (Px)_i = \sum_i \left( \sum_j P_{ij} x_j \right) = \sum_i \left( \sum_j P_{ij} x_i \right) = \sum_i \left( \sum_j P_{ij} x_i \right) = 1
   $$
   Also $(Px)_i = \sum_j P_{ij} x_j \geq 0$ since $P_{ij} \geq 0$ and $x_i \geq 0$.
   
   Thus $Px$ is a state vector.

2. $P$ has an eigenvalue $\lambda = 1$.
   
   **Proof:** Observe that the eigenvalues of $P^T$ are the same as the eigenvalues of $P$.
   
   **Proof:** 
   $$
   \det(P - \lambda I) = \det((P - \lambda I)^T) = \det(P^T - \lambda I) = \det(P - \lambda I)
   $$
   
   Next observe that since the columns of $P$ add up to one, the rows of $P^T$ add up to one and $P^T[1] = [1]$
   
   $$
   \Rightarrow \lambda = 1 \text{ is an eigenvalue of } P^T \text{ and thus of } P.
   $$

Note that even though $P$ and $P^T$ have the same eigenvalues they need not have the same eigenvectors!
(3) All (other) eigenvalues of $P$ satisfy $|\lambda_j| < 1$.

Proof: First we prove that $\|P x\|_1 \leq \|x\|_1$ for all state vectors $x$.

Proof: $\|P x\|_1 = \sum_{i=1}^n |(P x)_i| = \sum_{i=1}^n \sum_{j=1}^n \hat{a}_{ij} |x_j| \leq \sum_{i=1}^n \sum_{j=1}^n \hat{a}_{ij} |x_j| = \sum_{j=1}^n |x_j| = \|x\|_1$.

Now suppose $P v = \lambda v$, $\lambda \neq 0$. Then $\|P v\|_1 \leq \|v\|_1$, by what we proved above. However $\|P v\|_1 = \|v\|_1 \|v\|_1$, also. Thus $|\lambda| \leq 1$.

(4) The eigenvector for $\lambda_1 = 1$ has non-negative entries.

Idea of proof: Let $x_0 = a_1 v_1 + a_2 v_2 + \ldots + a_n v_n$. Moreover, let $(\lambda_j, v_j)$ be eigenvalue/eigenvector pairs with $\lambda_1 = 1$ and $|\lambda_j| < 1$ for all $j \neq 1$. Then

$$P^n x_0 = a_1 \lambda_1^n v_1 + a_2 \lambda_2^n v_2 + \ldots + a_n \lambda_n^n v_n.$$  

Recall the power method.

That means $P^n x_0 \to a_1 v_1$ as $n \to \infty$. But $P^n x_0 \geq 0$ for all $n$.

Thus the eigenvector $a_1 v_1$ has all non-negative entries.

Remark: The general case is not so easy to prove and we will not present this proof here.

(5) If $P$ (or $P^2$ for some $k \in \mathbb{N}$) has all positive entries (no zero entries!), then $\lambda_1 = 1$, $v_1$ has all positive entries, and $|\lambda_j| < 1$ for $j \neq 1$.

Remark: We do not require the positivity requirement to guarantee that $\lambda_1 = 1$ is the dominant eigenvalue.

Example: $P = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$ is a stochastic product with $\lambda_1 = 1$, $\lambda_2 = -1$. Thus $|\lambda_2| = 1$, too!

Note: $P^2 = I$, $P^3 = P$, ..., $P^{2n} = I$, $P^{2n+1} = P$.

So $\lim_{n \to \infty} P^n x_0$ does not exist unless $x_0 = \begin{bmatrix} a \\ a \end{bmatrix}$ for $a \in \mathbb{R}$. 

Google Page rank

Suppose we have a "world wide web" with 4 websites

![Diagram of a web with links]

# of links out of
node 1: 2
node 2: 0
node 3: 3
node 4: 1

We define

\[
P = \begin{bmatrix}
0 & 0 & \frac{1}{3} & 0 \\
\frac{1}{2} & 0 & \frac{1}{3} & 1 \\
\frac{1}{2} & 0 & 0 & 0 \\
0 & 0 & \frac{1}{3} & 0
\end{bmatrix}
\]

Idea: Start at some node: let's say \( x_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \), i.e. we start at node 4.

The web surfer does a random walk according to \( P \): \( x_1 = P x_0 \), \( x_2 = P^2 x_0 \), \( x_n = P^n x_0 \)

If there is a limiting state (i.e. \( n \to \infty \)) the probability in each node for this limit can be a measure of importance of the node (i.e. website).

Hence we need to answer the following questions:

i) When does such a limit exist?
ii) Does it depend on \( x_0 \)?
iii) Can we find the limit fast?

\{ Hint: stochastic matrix. \}
Remark: To ensure that the limit exists and that $P^n_{x0}$ converges fast to this limit, we need to modify the matrix $P$ such that it is a proper stochastic matrix.

$$ P = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0 \end{bmatrix} \quad \text{modify} \quad S = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0 \end{bmatrix} $$

The problematic column since it does not sum up to 1.

The matrix $S$ is a stochastic matrix. Interpreting the second column in the context of our example with 4 websites means that if the surfer, is on website 2 and finds no link he/she randomly chooses one of all the websites in the "world-wide-web" to go to next.

Now we can answer our questions:

Using MATLAB it is easy to check that $\lambda = 1$ is the dominant eigenvalue of $P$.

This means that $|\lambda_j| < 1$ for all other eigenvalues. Hence we can use the power method and obtain, again using MATLAB, that

$$ \tilde{v}_1 = \begin{bmatrix} 0.32, 0.81, 0.36, 0.32 \end{bmatrix}^T $$

is the eigenvector corresponding to $\lambda = 1$. Hence we have found a limit in finite time that does not depend on $x_0$.

The page rank of our 4 websites is now given by the entries of $\tilde{v}_1$, which is a normalized eigenvector $\tilde{v}_1$.

$$ \tilde{v}_1 = \begin{bmatrix} 0.18, 0.44, 0.20, 0.18 \end{bmatrix} $$

Note: $\tilde{v}_1 = \frac{v_1}{\|v_1\|}$.
Another example:

\[ S = P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \Rightarrow \text{stochastic matrix} \]

But: \( |\lambda_1| = |\lambda_2| = |\lambda_3| = 1 \), so no dominate eigenvalue.

We cannot use the power method!

How to resolve this?: We introduce some randomness \( \Rightarrow \text{DAMPING} \)

1. Step: Create a matrix \( Q \), same size as \( S \), such that all entries of \( Q \) are identical and \( Q \) is stochastic.

How example this would mean \( Q = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \)

2. Step: Pick a "damping factor" \( \alpha \in [0, 1] \)

3. Step: Define the "Google matrix": \( G = \alpha S + (1 - \alpha) Q \) as long as \( 0 < \alpha < 1 \)

Remark: 1) \( G \) is a stochastic matrix and each entry of \( G \) is positive. Thus \( \lambda = 1 \) is the dominate eigenvalue of \( G \).

4. Step: Now we can use the Power method as in the previous example.

The Page-Rank score of a site is again given as the corresponding entry of the eigenvector \( \tilde{v} \), of the eigenvalue \( \lambda = 1 \), normalized such that its entries add up to 1.