Lecture 34: Perron Frobenius theorem

This is a second lecture on Markov processes. We want to see why the following result is true:

> If all entries of a Markov matrix A are positive then A has a unique equilibrium: there is only one eigenvalue 1. All other eigenvalues are smaller than 1.

To illustrate the importance of the result, we look how it is used in chaos theory and how it can be used for search engines to rank pages.

1 The matrix

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

is a Markov matrix for which all entries are positive. The eigenvalue 1 is unique because the sum of the eigenvalues is $1/2 + 2/3 < 2$.

2 We have already proven Perron-Frobenius for 2×2 Markov matrices: such a matrix is of the form

$$A = \begin{bmatrix} a & b \\ 1-a & 1-b \end{bmatrix}$$

and has an eigenvalue 1 and a second eigenvalue smaller than 1 because $\mathrm{tr}(A)$ the sum of the eigenvalues is smaller than 2.

3 Lets give a brute force proof of the Perron-Frobenius theorem in the case of 3×3 matrices: such a matrix is of the form

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ 1-a-d & 1-b-e & 1-c-f \end{bmatrix} .$$

and has an eigenvalue 1. The determinant is $D = c(d-e) + a(e-f) + b(-d+f)$ and is the product of the two remaining eigenvalues. The trace is $1 + (a-c) + (e-f)$ so that $T = (a-c)+(e-f)$ is the sum of the two remaining eigenvalues. An ugly verification shows that these eigenvalues are in absolute value smaller than 1.

The Markov assumption is actually not needed. Here is a more general statement which is useful in other parts mathematics. It is also one the theorems with the most applications like **Leontief's models** in economics, **chaos theory** in dynamical systems or **page rank** for search engines.

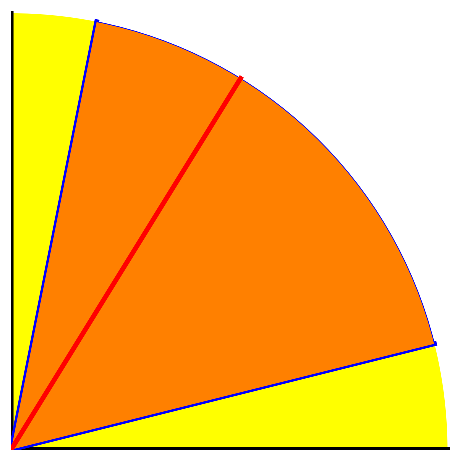
> **Perron Frobenius theorem**: If all entries of a $n \times n$ matrix A are positive, then it has a unique maximal eigenvalue. Its eigenvector has positive entries.

Proof. The proof is quite geometric and intuitive. Look at the sphere $x_1^2+\ldots+x_n^2 = 1$ and intersect it with the space $\{x_1 \geq 0, \ldots, x_n \geq 0\}$ which is a quadrant for $n = 2$ and octant for $n = 3$. This gives a closed, bounded set X. The matrix A defines a map $T(v) = Av/|Av|$ on X because the entries of the matrix are nonnegative. Because they are positive, TX is contained in the interior of X. This map is a **contraction**, there exists $0 < k < 1$ such that $d(Tx, Ty) \leq kd(x, y)$ where d

is the geodesic sphere distance. Such a map has a unique fixed point v by Banach's fixed point theorem. This is the eigenvector $Av = \lambda v$ we were looking for. We have seen now that on X, there is only one eigenvector. Every other eigenvector $Aw = \mu w$ must have a coordinate entry which is negative. Write $|w|$ for the vector with coordinates $|w_j|$. The computation

$$|\mu||w|_i = |\mu w_i| = |\sum_j A_{ij} w_j| \leq \sum_j |A_{ij}||w_j| = \sum_j A_{ij}|w_j| = (A|w|)_i$$

shows that $|\mu|L \leq \lambda L$ because $(A|w|)$ is a vector with length smaller than λL, where L is the length of w. From $|\mu|L \leq \lambda L$ with nonzero L we get $|\mu| \leq \lambda$. The first "$\leq$" which appears in the displayed formula is however an inequality for some i if one of the coordinate entries is negative. Having established $|\mu| < \lambda$ the proof is finished.



Remark. The theorem generalizes to situations considered in **chaos theory**, where **products of random matrices** are considered which all have the same distribution but which do not need to be independent. Given such a sequence of random matrices A_k, define $S_n = A_n \cdot A_{n-1} \cdots A_1$. This is a non commutative analogue of the random walk $S_n = X_1 + ... + X_n$ for usual random variables. But it is much more intricate because matrices do not commute. Laws of large numbers are now more subtle.

Application: Chaos

The **Lyapunov exponent** of a random sequence of matrices is defined as

$$\lim_{n\to\infty} \frac{1}{2n} \log \lambda(S_n^T S_n) ,$$

where $\lambda(B)$ is the maximal eigenvalue of the symmetric matrix $S_n^T S_n$.

Here is a prototype result in Chaos theory due to Anosov for which the proof of Perron-Frobenius can be modified using different contractions. It can be seen as an example of a noncommutative law of large numbers:

If A_k is a sequence of identically distributed random positive matrices of determinant 1, then the Lyapunov exponent is positive.

4 Let A_k be either $\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ or $\begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix}$ with probability 1/2. Since the matrices do not commute, we can not determine the long term behavior of S_n so easily and laws of large numbers do not apply. The Perron-Frobenius generalization above however shows that still, S_n grows exponentially fast.

Positive Lyapunov exponent is also called **sensitive dependence on initial conditions** for the system or simply dubbed **"chaos"**. Nearby trajectories will deviate exponentially. Edward Lorentz, who studied about 50 years ago models of the complicated equations which govern our weather stated this in a poetic way in 1972:

> The flap of a butterfly's wing in Brazil can set off a tornado in Texas.

Unfortunately, Mathematics is quite weak still to mathematically prove positive Lyapunov exponents if the system does not a priori feature positive matrices. There are cases which can be settled quite easily. For example, if the matrices A_k are IID random matrices of determinant 1 and eigenvalues 1 have not full probability, then the Lyapunov exponent is positive due to work of Fuerstenberg and others. In real systems, like for the motion of our solar system or particles in a box, positive Lyapunov exponents is measured but can not be proven yet. Even for simple toy systems like $S_n = dT^n$, where dT is the Jacobean of a map T like $T(x, y) = (2x - c\sin(x), y)$ and T^n is the n'th iterate, things are unsettled. One measures $\lambda \geq \log(c/2)$ but is unable to prove it yet. For our real weather system, where the Navier stokes equations apply, one is even more helpless. One does not even know whether trajectories exist for all times. This existence problem looks like an esoteric ontological question if it were not for the fact that a one million dollar bounty is offered for its solution.

Application: Pagerank

A set of nodes with connections is a **graph**. Any network can be described by a graph. The link structure of the web forms a graph, where the individual websites are the nodes and if there is an arrow from site a_i to site a_j if a_i links to a_j. The adjacency matrix A of this graph is called the **web graph**. If there are n sites, then the adjacency matrix is a $n \times n$ matrix with entries $A_{ij} = 1$ if there exists a link from a_j to a_i. If we divide each column by the number of 1 in that column, we obtain a Markov matrix A which is called the **normalized web matrix**. Define the matrix E which satisfies $E_{ij} = 1/n$ for all i, j. The graduate students and later entrepreneurs **Sergey Brin** and **Lawrence Page** had in 1996 the following one billion dollar idea:

> The **Google matrix** is the matrix $G = dA + (1 - d)E$, where $0 < d < 1$ is a parameter called **damping factor** and A is the Markov matrix obtained from the adjacency matrix by scaling the rows to become stochastic matrices. This is a $n \times n$ Markov matrix with eigenvalue 1.

> Its Perron-Frobenius eigenvector v scaled so that the largest value is 10 is called **page rank** of the damping factor d.

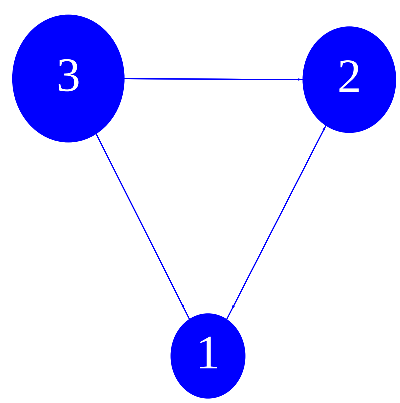
> The **page rank equation** is
>
> $$[dA + (1 - d)E]v = v$$

The damping factor can look a bit misterious. Brin and Page write:

"PageRank can be thought of as a model of user behavior. We assume there is a "random surfer" who is given a web page at random and keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page. The probability that the random surfer visits a page is its PageRank. And, the d damping factor is the probability at each page the "random surfer" will get bored and request another random page. One important variation is to only add the damping factor d to a single page, or a group of pages. This allows for personalization and can make it nearly impossible to deliberately mislead the system in order to get a higher ranking. We have several other extensions to PageRank." [1]

5 Consider 3 sites A, B, C, where A is connected to B, C and B is connected to C and C is connected to A. Find the page rank to $d = 0.1$. **Solution.** The adjacency matrix of the graph is $A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$. The Google matrix is

$$G = (1-d)\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}/3 + d\begin{bmatrix} 0 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1 & 0 \end{bmatrix}.$$



It is said now that page rank is the world's largest matrix computation. The $n \times n$ matrix is huge. It was 8.1 billion 5 years ago. [2]

Homework due April 27, 2011

1 Verify that if a Markov matrix A has the property that A^2 has only positive entries, then A has a unique eigenvalue 1.

2 Take 4 cites A, B, C, D where A links to B, C, D, and B links to C, D and C links to D and D links to A. Find the Google matrix with the damping factor 1/2.

3 Determine the Page rank of the previous system, possibly using technology like Mathematica.

[1] http://infolab.stanford.edu/ backrub/google.html

[2] Amy Langville and Carl Meyer, Googles PageRank and Beyond, Princeton University Press, 2006.