Graphs, Random Walks, and Matrices

Hans R. Schneebeli

Abstract

A link is established between graphs and matrices or between random walks on graphs and stochastic matrices. It is then exploited computationally in order to solve some prototypical problems about random walks on graphs and their applications.

The toolbox of a graphing calculator or a CAS enables us to operate rather easily on matrices. The workshop presents a selection of sample problems which are in the reach of students equipped with handheld technology.

Our principal aim is to shed some light on social rankings and one of the basic ideas behind Google's PageRank. The centrepiece of this kind of application of mathematics is the Theorem of Perron-Frobenius on non-negative matrices.

Requirements Basics of linear algebra, linear maps, the fundamental theorem on linear maps, eigenvalues, eigenvectors. All computations can be done with the help of a CAS, calculator or some numerical software like Octave, Matlab, TI-Nspire.

Aims Explore the power of linear algebra applied to random walks on graphs, applications of Markov processes, stochastic matrices in various situations from climate simulations to Leontjev models for closed economies to social rankings in a group of monkeys or to Google's PageRank.

1 Random walks on graphs, and matrices

Any directed and weighted graph may be interpreted in terms of flows along directed edges between the vertices of the graph. Assume that the graph is finite and has n vertices. Then n^2 directed and weighted connections are possible between these vertices, if a connection from any vertex to itself is allowed. Now label any directed edge with a number indicating the strength of a hypothetical flow along this edge. Then a directed and weighted graph results. It may be conveniently interpreted as an $n \times n$-matrix $G = [g_{ij}]$ whose entry g_{ij} describes the strenght of the flow from vertex number j to vertex number i. This convention depends on the manner in which linear selfmaps of vector spaces are coded as matrices. Specifically, we assume a linear selfmap $g : V \to V$ to be coded in matrixform with respect to some basis $\{\vec{b}_1, \ldots \vec{b}_n\} \subset V$ in such a way that column number r of the associated matrix G contains the coordinates with respect to this basis of the image $g(\vec{b}_r)$ of vector $\vec{b}_r$. The only other notational convention for matrices as coordinate descriptions of linear maps would result in an interchange of rows and columns, i.e. replace any product $G \cdot \vec{x}$ by the transpose notation $(G \cdot \vec{x})^\intercal = \vec{x}^\intercal \cdot G^\intercal$. Both variants may be found in the literature.

A vector is called stochastic if all its entries are non-negative and sum up to 1. The set of stochastic vectors in $\mathbb{R}^n$ is the convex hull of its standard basis. An $n \times n$-matrix is stochastic if its columns are stochastic vectors. Equivalently, it describes a linear transformation mapping the set of stochastic vectors to itself. Here, we assume the standard basis to be chosen for the coordinatization of the selfmap. Any stochastic matrix $S = [s_{ij}]$ is a description of a random walk between n points or states, where s_{ij} is the constant probability for a transition from state j to state i.

Example: Winter weather in Tel Aviv During winter, the weather in Tel Aviv is either rainy or sunny. Climatologists found out that a stochastical model describes the pattern of transitions between the two states S or R at least for the sake of climatology.

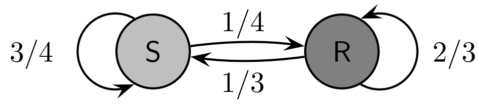

Figure 1: Markov chain model for a two state climate in winter at Tel Aviv

Problems

1. Given, on November 30th the weather at Tel Aviv is sunny and we use the climatological model for a prediction. What is the probability for
   (a) sunny weather on the following days:
   December 2nd, 4th, 8th, 16th, January 1st, January 2nd?
   (b) the same events if the weather on November 30th is rainy?
2. Draw a schematic sketch showing the stochastic vectors in $\mathbb{R}^2$ and in $\mathbb{R}^3$.
3. Why is the set of stochastic vectors in $\mathbb{R}^2$ and in $\mathbb{R}^3$ the convex hull of the canonical basis?
4. Assume S is a stochastic 3×3-matrix and $S^\intercal$ its transpose.
   (a) Let $\vec{e} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and compute $\quad S^\intercal \cdot \vec{e}$.
   (b) Why is 1 an eigenvalue of S?
   (c) Is $\vec{e}$ an eigenvector of S?
   (d) Does S admit a fixed point in the set of stochastic vectors?
   (e) Which of your findings generalize to stochastic matrices acting on the stochastic vectors of $\mathbb{R}^n$?
5. Assume we cast a fair coin and observe the results.
   (a) Which directed graph represents this experiment?
   (b) Which stochastic matrix describes this experiment?
   (c) What about rolling a fair dice?
   (d) What conclusions generalize to arbitrary Laplacian experiments?
   (e) Consider a Bernoulli experiment with states [0] and [1]. The probability to reach [1] is given bei p irrespective of the state the system is in. Then [0] is reached with probability $q := 1 - p$ from both states. Which is the system matrix and what equilibrium distribution does the system reach asymptotically?

2 Linear models for closed economies according to Leontjev

Leontjev, the first Nobel laureate in economics (1973), invented two simplified models for economies. One for closed economies which will be dealt with in the sequel in a didactically simplified form. The only difference between an example in teaching linear algebra and the work of a pioneering Nobel laureate lies in the data, including data processing by one of the early electronic computers, the Harvard Mark I in 1943.

The data for the following example are pure inventions while the ones in Leontjev's model were based on an analysis of empirical data from the US market after 1935.

We are going to explain the working of Leontjev's model using a toy market based on five sectors only. In contrast, Leontjev considered a subdivision of the US economy into at least 70 sectors. The main idea of the so called closed model is to simulate the flow of money during the economic activities of the various branches exchanging goods, services, energy or information for money. The key question to be solved empirically was to find out what happens to a monetary unit in anyone of the various branches in one time step. There is a simplifying assumption: The total amount of money as well as prices are kept constant inside the market.

This results in a graph describing a lossless flow of money between the different sectors of economy. After normalising the currency we find ourselves in the situation of a Markov process and the Leontjev matrix is seen to be a stochastic one. Typically it is ergodic and after some time its dynamics settles down in an equilibrium described mathematically by a fixed point.

For the purpose of illustration our toy economy will consist of five sectors only:

1. industry
2. agriculture
3. transportation
4. energy production and waste treatment
5. information treatment and administration

Then

$$L = \begin{bmatrix} 0.25 & 0.4 & 0.25 & 0.15 & 0.2 \\ 0.1 & 0.2 & 0.1 & 0.2 & 0.25 \\ 0.2 & 0.2 & 0.2 & 0.15 & 0.15 \\ 0.20 & 0.1 & 0.15 & 0.25 & 0.1 \\ 0.25 & 0.1 & 0.3 & 0.25 & 0.3 \end{bmatrix}$$

is a Leontjev matrix for a closed economy.

Questions concerning the Leontjev toy model

1. What is the interpretation of the number 0.4 in the Leontjev matrix of the toy model?
2. How can we check that the total amount of the money circulating in the model remains constant?
3. What happens to the money after a long time in this model?

4. Make a prediction for the diffusion of a currency unit concentrated initially in agriculture into the various sectors after 8, 64, 1024 time steps.
5. Why did the assumptions of a closed economy fit the US market around 1940?
6. Why could Leontjev be awarded a Nobel price in Economics despite the fact that the mathematical model is due to Markov?

3 Monkey business – a prelude to PageRank

Look at this picture from Mount Bukit, the top spot of Singapore. It shows monkey business in action. While father and son are chasing insects, three females interact socially. A measurable expression of social interaction in primates is the intensity of grooming. Grooming is important for social bonding by cultivating sympathy and curing one another's skin and hair. To make a link to the Leontjev model we might say that grooming is an exchange of sympathy capital or respect. From this point of view it is tempting to ask what will happen to a unit of sympathy capital in the monkey society after some time.

For an answer we need a sociogram based on the observation of the group and the mutual grooming activity levels.

Figure 2: A group of macaques, social interaction

We now introduce a measure for the intensity of grooming into the sociogram and derive the 'grooming matrix' in two steps. The intensity levels are coarsly discretized on a scale from 0 to 3. There results a matrix with non-negative entries. Now normalize the columns such that a stochastic matrix G arises

$$G = \begin{bmatrix} 0 & 0 & 0 & 1/5 & 0 & 0 & 1/3 & 1/7 & 0 \\ 1/4 & 0 & 3/5 & 1/5 & 1/6 & 0 & 0 & 0 & 0 \\ 0 & 3/5 & 0 & 2/5 & 0 & 1/5 & 0 & 2/7 & 0 \\ 1/2 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 \\ 1/4 & 1/5 & 0 & 0 & 0 & 0 & 1/6 & 1/7 & 0 \\ 0 & 1/5 & 0 & 0 & 1/6 & 0 & 1/6 & 1/7 & 3/4 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/7 & 0 \\ 0 & 0 & 2/5 & 1/5 & 0 & 2/5 & 1/3 & 0 & 1/4 \\ 0 & 0 & 0 & 0 & 0 & 2/5 & 0 & 1/7 & 0 \end{bmatrix}$$

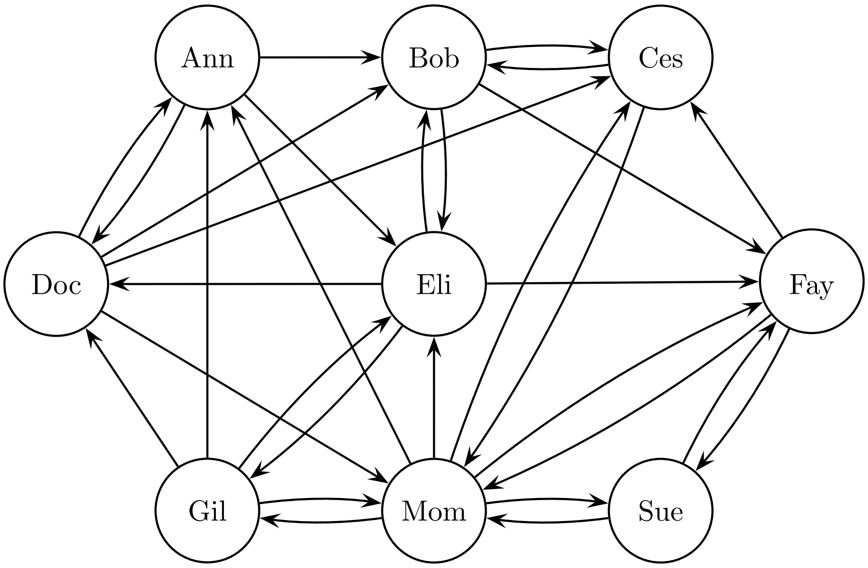

Figure 3: Sociogram of a group of nine macaques

Activities: Social ranking to PageRank

1. Check that the vector

$$\vec{v} := \begin{bmatrix} A \\ B \\ C \\ D \\ E \\ F \\ G \\ M \\ S \end{bmatrix} \approx \begin{bmatrix} 0.054678.. \\ 0.155003.. \\ 0.195428.. \\ 0.053923.. \\ 0.079753.. \\ 0.141545.. \\ 0.052858.. \\ 0.183916.. \\ 0.082892.. \end{bmatrix}$$

   is stochastic and an eigenvector of G corresponding to a fixed point.

2. Argue why the maximal entry identifies the most worshiped individual of the group of macaques.

3. In what way can a social ranking among the macaques be derived from the eigenvector?

4. Imagine a sociogram for web sites on the internet. What corresponds to grooming?

5. What is an important difference between the group of macaques and the ranking of web sites on the Internet? Try to find out the additional mathematical or numerical problems that had to be overcome for a ranking such as the PageRank used in Google's search engine.

4 Some Remarks

For those interested in the mathematical background, we mention the Theorem of Perron - Frobenius. It applies in particular to ergodic stochastic matrices and guarantees the existence of a fixed point.

More generally, the theorem deals with non-negative matrices or linear maps that map the positive orthant of $\mathbb{R}^n$ into itself. Then a dominant positive eigenvalue exists and the corresponding eigenvector may be chosen such that all its entries are non-negative.

References

Anton and Rorres, Elementary Linear Algebra, Applications Version, 7th Ed, John Wiley & Sons Inc., New York 1994, ISBN 0-471-30570-7

Anton and Busby, Contemporary Linear Algebra, John Wiley & Sons Inc., New York 2003, ISBN 0-471-16362-7 [Search Engines, Page Rank, cf. pp 249-263]

S. Brin and L. Page. The anatomy of a large-scale hypertextual (Web) search engine. In The Seventh International World Wide Web Conference, 1998.

D. Carlson et al., Linear Algebra Gems, MAA Notes # 59, 2002, ISBN 0-88385-170-9

P. G. Doyle, J.L. Snell, Random Walks and Electric Networks, Carus Monograph Vol. 22, MAA, 1984, ISBN 0-88385-024-9

Arthur Engel, Wahrscheinlichkeitsrechnung und Statistik, Klett Studienbücher, Band 2, 1978, ISBN 3-12-983170-3

Gabriel, K. R. and Neumann J. (1962); A Markov chain model for daily rainfall occurrence at Tel Aviv, Quart. Jour. Roy. Meteorol. Soc. Vol. 88, pp 90-95.

Dr. E. Garcia www.miislita.com, [10.3.2012]
Matrix Tutorial 3: Eigenvalues and Eigenvectors. A tutorial on eigenvalues, eigenvectors and their properties. Includes step by step how-to calculations. An introduction to vector iteration, the Power Method and the Deflation Method is provided.

Andrew Y. Ng, Alice X. Zheng and Michael Jordan. Link Analysis, Eigenvectors, and Stability. In Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-01), 2001. [ps, pdf]

A. N. Langville und C. D. Meyer, Who is #1? The science of Rating and Ranking, Princeton University Press, ISBN 978-00691-15422-0

A. N. Langville und C. D. Meyer, Google's PageRank and Beyond: The Science of Search Engine Rankings, Princeton University Press, 2006, ISBN 978-0-691-15266-0

Matoušek, Jiří, Thirty-three Miniatures, Mathematical and Algorithmic Applications of Linear Algebra, MAA Student Mathematical Library Vol 53, ISBN 978-8218-4977-4

C. D. Meyer, Matrix Analysis and Applied Linear Algebra, SIAM, Philadelphia, 2000

Cleve B. Moler, Numerical Computing with MATLAB, section 2.11 PageRank and Markov Chains, SIAM, Philadelphia, 2004, ISBN 0-89871-560-1 (pbk)