HM 802

February 8, 2013

1 Preamble

The notes below will serve as the basis for the module lectures to be given on the topics of Jordan form, nonnegative matrices, and Markov chains. As you'll see, most of the proofs have been omitted in the notes. The intention is to sketch the proofs during the lectures, but also to have the students work through some of the details (using references such as those listed at the end of this document) as part of the self–study component of module. If you're having trouble getting ahold of some of the reference texts, I have copies of all of them in my office which you're welcome to borrow from time to time. I have also included some sample exercises. These are not intended to be terribly difficult, but if you want to discuss the problems with me once the module is over, feel free to do so.

2 Eigenvalues, eigenvectors, and Jordan canonical form

Suppose that A is a complex matrix of order n. A scalar $\lambda \in \mathbb{C}$ is an *eigenvalue* for A if there is a nonzero vector $v \in \mathbb{C}^n$ such that $Av = \lambda v$. In that case, the vector v is an *eigenvector* for A. Occasionally we refer to such a v as a *right eigenvector* for A. Note that there is a corresponding notion of a *left eigenvector* as well: a nonzero vector $w \in \mathbb{C}^n$ is a left eigenvector for A corresponding to the eigenvalue λ provided that $w^*A = \lambda w^*$.

The *spectral radius* of a square matrix A, denoted $\rho(A)$ is given by

$$\rho(A) = \{|\lambda||\lambda \text{ is an eigenvalue of } A\}.$$

Suppose that the square matrix A has λ as an eigenvalue. A *Jordan chain* for λ is a finite list of nonzero vectors $v_1, \ldots, v_k$ such that $Av_1 = \lambda v_1$, and for $j =$

$2, \ldots, k, Av_j = \lambda v_j + v_{j-1}$. Evidently v_1 is an eigenvector of A corresponding to the eigenvalue λ. The vectors $v_2, \ldots, v_k$ are known as *generalised eigenvectors* of A corresponding to the eigenvalue λ. The *algebraic multiplicity* of an eigenvalue λ of A is its multiplicity as a root of the characteristic polynomial of A, while the *geometric multiplicity* of λ is the dimension of the null space of $A - \lambda I$. For any eigenvalue, the algebraic multiplicity is greater than or equal to the geometric multiplicity.

Suppose that $\lambda \in \mathbb{C}$ and that $k \in \mathbb{N}$. The *Jordan block* $J_k(\lambda)$ is the $k \times k$ upper triangular matrix

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \ldots & & 0 \\ 0 & \lambda & 1 & 0 & \ldots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ & & & & & \\ 0 & 0 & & \ldots & \lambda & 1 \\ 0 & 0 & & \ldots & 0 & \lambda \end{bmatrix}.$$

Note that in the special case that $k = 1, J_1(\lambda)$ is just the 1×1 'matrix' $\begin{bmatrix} \lambda \end{bmatrix}$.

Theorem 2.1 *Suppose that A is a complex matrix of order $n \geq 2$. Then there is a nonsingular $n \times n$ matrix S such that*

$$S^{-1}AS = \begin{bmatrix} J_{k_1}(\lambda_1) & & & \\ & J_{k_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{k_m}(\lambda_m) \end{bmatrix}. \tag{1}$$

Remarks:
• The matrix on the right hand side of (1) is the *Jordan canonical form* for A. It is unique, up to a reordering of the diagonal Jordan blocks.
• The numbers $\lambda_1, \ldots, \lambda_m$ are the eigenvalues of A, and they are not necessarily distinct.
• Evidently $k_1 + \ldots k_m = n$.
• In the case that $m = n$ and all k_js are equal to 1, the Jordan canonical form for A is a diagonal matrix, and A is said to be *diagonable*.
• The columns of the matrix S consist of eigenvectors and generalised eigenvectors for the matrix A.
• Suppose that λ is an eigenvalue of the matrix A, and that the Jordan canonical form for A is given by (1). Let $I = \{j | \lambda_j = \lambda\}$. Then the algebraic multiplicity of λ is given by $\sum_{j \in I} k_j$, while the geometric multiplicity of λ is given by $|I|$.
• Suppose that $k, p \in \mathbb{N}, \lambda \in \mathbb{C}$ and consider the matrix $J_k(\lambda)^p$. An argument by

induction (on p) shows that $J_k(\lambda)^p$ is an upper triangular matrix such that for each pair of indices i, j with $1 \le i < j \le k$, the (i, j) entry of $J_k(\lambda)^p$ is given by:

$$(J_k(\lambda)^p)_{i,j} = \begin{cases} \binom{p}{j-i}\lambda^{p-j+i}, & \text{if } j-i \le p-1, \\ 1, & \text{if } j-i = p, \\ 0, & \text{if } j-i > p. \end{cases}$$

Observe that the search for the Jordan canonical form of a given square matrix A is essentially the same as the search for Jordan chains associated with the eigenvalues of A. For concreteness, suppose that λ is an eigenvalue of A. We can then find a basis for the null space of $A - \lambda I$, say $u_1, \dots, u_p$. Each u_j is an eigenvector of A corresponding to λ, and the geometric multiplicity of λ is p, so we know that there will be p Jordan blocks corresponding to the eigenvalue λ in the Jordan form for A. Further, each u_j can be used to construct a Jordan chain of eigenvectors and generalised eigenvectors, as follows.
i) Select u_1, and set $v_1 \equiv u_1$.
For each $l = 1, 2, \dots$, we iterate the steps below.
ii) If the null space of $(A - \lambda I)^l$ is a proper subspace of the null space of $(A - \lambda I)^{l+1}$, we find v_{l+1} as a solution to the linear system $(A - \lambda I)v_{l+1} = v_l$.
iii) If the null space of $(A - \lambda I)^l$ is the same as the null space of $(A - \lambda I)^{l+1}$, then we stop.
Observe that this procedure produces a Jordan chain of length r (say), $v_1, \dots, v_r$, whose initial vector is u_1. The corresponding Jordan block in the Jordan form for A will be $r \times r$. We then repeat the procedure to produce Jordan chains for λ that start with each of $u_2, \dots, u_p$. A similar process applies to the remaining eigenvalues of A.

3 Nonnegative matrices

3.1 Primitive matrices

A square matrix A with entries in $\mathbb{R}$ is *nonnegative* (respectively, *positive*) if each of its entries is nonnegative (respectively, *positive*). We use the notation $A \ge 0$ (respectively, $A > 0$) to denote this, and a similar terminology and notation applies to vectors in $\mathbb{R}^n$. A nonnegative matrix A is *primitive* if there is a $k \in \mathbb{N}$ such that $A^k > 0$.

Theorem 3.1 (The Perron–Frobenius theorem for primitive matrices) *Suppose that A is a primitive nonnegative matrix of order n. Then:*
a) the spectral radius, $\rho(A)$, is an eigenvalue of A;
b) $\rho(A)$ is an algebraically simple eigenvalue of A;

c) A has positive right and left eigenvectors corresponding to the eigenvalue $\rho(A)$;
d) the left and right eigenspaces of A corresponding to $\rho(A)$ both have dimension 1;
e) if $\lambda \neq \rho(A)$ is an eigenvalue of A, then $|\lambda| < \rho(A)$;
f) if B is an $n \times n$ matrix such that $0 \leq B \leq A$, then $\rho(B) \leq \rho(A)$, and if $\rho(B) = \rho(A)$, then in fact $B = A$;
g) if x is a positive right eigenvector of A, then necessarily $Ax = \rho(A)x$, with an analogous statement holding for positive left eigenvectors.

For a primitive nonnegative matrix A, the eigenvalue $\rho(A)$ is referred to as the *Perron value* for A, while a positive right (respectively, left) eigenvector of A corresponding to the Perron value is called a right (respectively, left) *Perron vector* for A.

Example 3.2 Consider the matrix

$$A = \begin{bmatrix} 0 & \frac{3}{4} & \frac{1}{4} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

It can be verified that $A^5 > 0$, so that A is primitive. The characteristic polynomial of A is given by $\det(xI - A) = x^3 - \frac{3}{4}x - \frac{1}{4}$, and we deduce that the eigenvalues of A are 1 and $-\frac{1}{2}$, the latter with algebraic multiplicity two. In particular, $\rho(A) = 1$ is an eigenvalue of A, and it dominates the modulus of the other eigenvalue(s) of A. The vectors

$$x = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, y = \begin{bmatrix} 4 \\ 4 \\ 1 \end{bmatrix}$$

are readily seen to be right and left eigenvectors (respectively) of A corresponding to the eigenvalue 1. It turns out that the Jordan canonical form for A is given by

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & 1 \\ 0 & 0 & -\frac{1}{2} \end{bmatrix}.$$

The next result follows from Theorems 2.1 and 3.1. Here we use J to denote an all–ones matrix.

Theorem 3.3 *Let A be a primitive nonnegative matrix of order n with Perron value ρ. Let x and y denote right and left Perron vectors of A respectively, normalised so* *that $y^T x = 1$. Label the eigenvalues of A as $\rho, \lambda_2, \ldots, \lambda_n$, where $|\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n|$. If $\lambda_2 \neq 0$, there is a constant C such that*

$$\left|\frac{1}{\rho^k} A^k - xy^T\right| \leq Ck^{n-1} \left(\frac{|\lambda_2|}{\rho}\right)^k J,$$

where the inequality holds entrywise. On the other hand, if $\lambda_2 = 0$, then $\frac{1}{\rho^k} A^k = xy^T$ for all $k \geq n-1$. In particular, in either case we have

$$\lim_{k \to \infty} \frac{1}{\rho^k} A^k = xy^T.$$

Corollary 3.4 *Suppose that A, x, y and ρ are as in Theorem 3.3. Let $z \in \mathbb{R}^n$ be a nonnegative vector having at least one positive entry. Then*

$$\lim_{k \to \infty} \frac{1}{\rho^k} A^k z = (y^T z)x.$$

3.2 Primitive matrix models in mathematical ecology

Suppose that we have a species whose members can be classified into n disjoint classes. Examples of this sort of classification might be: egg, larva, pupa, adult; or age 0–5 year, age 5–10 years, …, age 90–100 years; or according to the size of a particular attribute (such as the carapace, in a population of tortoises). We select a time unit, and want to model the species in discrete time, keeping track of the number of individuals in each sub–class at each time. For each pair of indices i, j between 1 and n, we let $a_{i,j}$ denote the average number of members of class i that, in one time unit, are contributed by a member of class j. This 'contribution' usually takes one of two forms: survival, i.e. survival into class i from class j; and reproduction, i.e. members of class j can give birth to new individuals in class i. (For practical purposes, it is often the case that the model only considers the female members of the population, the assumption being that the growth and structure of the male subpopulation is compatible with that of the female subpopulation.)

Now form the *population projection matrix* $A = \left[\ a_{i,j}\ \right]_{i,j=1,\ldots,n}$. Let $p(0) \in \mathbb{R}^n$ be a nonnegative, nonzero vector that serves as the initial population vector at time zero. That is, for each $i = 1, \ldots, n, p(0)_i$ is the number of individuals in class i at time zero. Then for each $k \in \mathbb{N}$, we have $p(k) = Ap(k-1)$, or equivalently, $p(k) = A^k p(0)$.

How does the structure of the population behave as $k \to \infty$? Let $\mathbf{1}$ denote the all ones vector in $\mathbb{R}^n$, and observe that $\frac{1}{\mathbf{1}^T p(k)} p(k)$ is the vector giving the proportion of members of the population in the various stage classes at time k. If the matrix A

happens to be primitive with Perron vector x, then from Corollary 3.4 we find that as $k \to \infty$,

$$\frac{1}{\mathbf{1}^T p(k)} p(k) \to \frac{1}{\mathbf{1}^T x} x.$$

For this reason, the vector $\frac{1}{\mathbf{1}^T x} x$ is sometimes known as the *stable distribution vector* for the population. Note also that as $k \to \infty$, the total size of the population is $\mathbf{1}^T p(k)$. Since

$$\lim_{k\to\infty} \frac{\mathbf{1}^T p(k+1)}{\mathbf{1}^T p(k)} = \rho(A),$$

the Perron value of A is interpreted as the asymptotic growth rate of the population.

Example 3.5 (North American right whale) The female subpopulation is under consideration here, the time unit is one year, and the population is subdivided into five categories: calf, immature female, mature but non-reproductive female, mother, and resting mother (right whales do not reproduce in the year after giving birth). The corresponding population projection matrix (based on estimated data) is

$$A = \begin{bmatrix} 0 & 0 & .13 & 0 & 0 \\ .9 & .85 & 0 & 0 & 0 \\ 0 & .12 & .71 & 0 & 1 \\ 0 & 0 & .29 & 0 & 0 \\ 0 & 0 & 0 & .85 & 0 \end{bmatrix}. \tag{2}$$

The Perron value is 1.0239 and the stable distribution vector is

$$\begin{bmatrix} 0.0551 \\ 0.2853 \\ 0.4344 \\ 0.1230 \\ 0.1021 \end{bmatrix}.$$

Example 3.6 (Desert tortoise) The population projection matrix for this species is

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1.300 & 1.980 & 2.570 \\ 0.716 & 0.567 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.149 & 0.567 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.149 & 0.604 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.235 & 0.560 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.225 & 0.678 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.249 & 0.851 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.016 & 0.860 \end{bmatrix}.$$

The corresponding Perron value is 0.9581, and the stable distribution vector is

$$\begin{bmatrix} 0.2217 \\ 0.4058 \\ 0.1546 \\ 0.0651 \\ 0.0384 \\ 0.0309 \\ 0.0718 \\ 0.0117 \end{bmatrix}.$$

One question of interest in the mathematical ecology literature is the sensitivity of the asymptotic growth rate to changes in demographic parameters, or, in mathematical language, the sensitivity of the Perron value of A to changes in the entries of A. To frame the issue more precisely, suppose that we're given a primitive matrix A of order n, and we consider its Perron value $\rho(A)$ as a function of the n^2 entries in A. One way of measuring the sensitivity of the Perron value is as follows: fix a pair of indices i, j between 1 and n. Now consider

$$\left.\frac{\partial \rho}{\partial a_{i,j}}\right|_A = \lim_{\epsilon \to 0} \frac{1}{\epsilon}(\rho(A + \epsilon E_{i,j}) - \rho(A)).$$

Here $E_{i,j}$ denotes the $(0,1)$ matrix with a 1 in the (i, j) position, and 0s everywhere else.

Theorem 3.7 *Let A be a primitive $n \times n$ matrix with Perron value $\rho(A)$. Suppose that x and y are right and let Perron vectors for A, respectively, normalised so that $y^T x = 1$. Fix a pair of indices i, j between 1 and n. Then*

$$\left.\frac{\partial \rho}{\partial a_{i,j}}\right|_A = x_j y_i.$$

Example 3.8 We revisit the matrix of Example 3.6. We saw in that example that the stable distribution vector is

$$x = \begin{bmatrix} 0.2217 \\ 0.4058 \\ 0.1546 \\ 0.0651 \\ 0.0384 \\ 0.0309 \\ 0.0718 \\ 0.0117 \end{bmatrix},$$

and it turns out that the left Perron vector y, normalised so that $y^T x = 1$, is given by

$$y = \begin{bmatrix} 0.1955 \\ 0.2616 \\ 0.6866 \\ 1.8019 \\ 2.7148 \\ 4.8029 \\ 4.3813 \\ 5.1237 \end{bmatrix}.$$

From Theorem 3.7 we find that the matrix of derivatives for the Perron value is given by

$$yx^T = \begin{bmatrix} 0.0433 & 0.0793 & 0.0302 & 0.0127 & 0.0075 & \mathbf{0.0060} & \mathbf{0.0140} & \mathbf{0.0023} \\ \mathbf{0.0580} & \mathbf{0.1062} & 0.0405 & 0.0170 & 0.0100 & 0.0081 & 0.0188 & 0.0031 \\ 0.1522 & \mathbf{0.2786} & \mathbf{0.1062} & 0.0447 & 0.0264 & 0.0212 & 0.0493 & 0.0080 \\ 0.3994 & 0.7313 & \mathbf{0.2786} & \mathbf{0.1173} & 0.0692 & 0.0556 & 0.1294 & 0.0211 \\ 0.6018 & 1.1018 & 0.4198 & \mathbf{0.1767} & \mathbf{0.1043} & 0.0838 & 0.1949 & 0.0318 \\ 1.0646 & 1.9492 & 0.7427 & 0.3126 & \mathbf{0.1845} & \mathbf{0.1482} & 0.3448 & 0.0563 \\ 0.9712 & 1.7782 & 0.6775 & 0.2851 & 0.1683 & \mathbf{0.1352} & \mathbf{0.3145} & 0.0513 \\ 1.1357 & 2.0794 & 0.7923 & 0.3334 & 0.1968 & 0.1581 & \mathbf{0.3678} & \mathbf{0.0600} \end{bmatrix}.$$

Here the bolded entries correspond to the demographically relevant entries, where the original population projection matrix has positive entries.

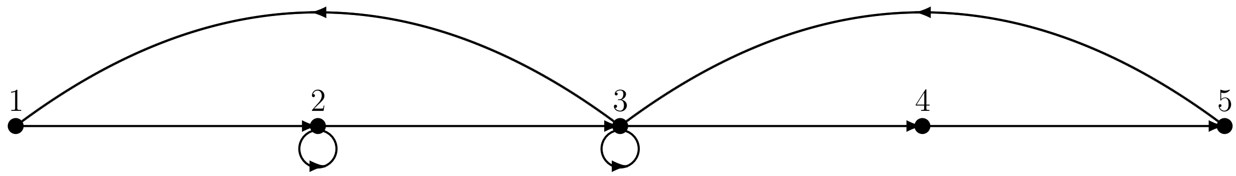
3.3 Nonnegative matrices and directed graphs

Suppose that we have a nonnegative matrix A of order n. We have seen that there are some nice things to be said if A is primitive, so, how can we determine whether or not a given A is primitive? To answer this question, we introduce the useful notion of the directed graph associated with a matrix.

Given an $n \times n$ matrix $A \geq 0$, the *directed graph of* A, denoted $D(A)$ is the defined as follows: the vertices of $D(A)$ are labelled with the numbers $1, \ldots, n$; for each $1 \leq i, j \leq n$, there is an arc from vertex i to vertex j in $D(A)$ if $a_{i,j} > 0$, and no arc between those vertices if $a_{i,j} = 0$.

Example 3.9 Consider the matrix A of (2). Then $D(A)$ is as follows.

Suppose that D is a directed graph on vertices $1, \ldots, n$, and fix a pair of vertices i, j. A *walk* in D from vertex i to vertex j is a sequence of arcs in D of the form $i \equiv i_0 \to i_1 \to i_2 \to \ldots \to i_k \equiv j$. The *length* of a walk is the number of arcs it



contains. If the vertices on a walk are all distinct, then the walk is called a *path*, while if we have a walk from vertex i to vertex i of the form $i \equiv i_0 \to i_1 \to i_2 \to \ldots \to i_k \equiv i$ and the vertices $i_0, i_1, \ldots, i_{k-1}$ are distinct, then the walk is called a *cycle*. A cycle of length 1, that is, an arc of the form $i \to i$, is called a *loop*. Finally, a directed graph D is *strongly connected* if it has the property that for any pair of vertices i, j of D, there is a walk from i to j in D.

A square matrix A is said to be *reducible* if there is a permutation matrix P such that PAP^T is in block triangular form – i.e.

$$PAP^T = \left[\begin{array}{c|c} A_{1,1} & A_{1,2} \\ \hline 0 & A_{2,2} \end{array}\right],$$

where the blocks $A_{1,1}$ and $A_{2,2}$ are square. A matrix A is *irreducible* if no such permutation exists. The follow results describes the relationship between irreducibility and directed graphs.

Theorem 3.10 *A square nonnegative matrix A is irreducible if and only if its directed graph $D(A)$ is strongly connected.*

The connection between primitivity and directed graphs becomes more apparent with the following result.

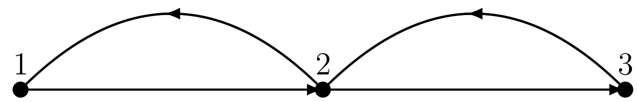
Theorem 3.11 *Let A be a square nonnegative matrix. Then A^k has a positive entry in position (i, j) if and only if $D(A)$ contains a walk from vertex i to vertex j of length k.*

From Theorem 3.11 we deduce that a square nonnegative matrix A is primitive if and only if there is a $k \in \mathbb{N}$ such that for any two vertices in $D(A)$, there is a walk of length exactly k. From this we deduce that any primitive matrix must have a strongly connected directed graph. However, the converse fails, as the following example shows.

Example 3.12 Consider the following 3×3 matrix:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}.$$

The directed graph for A is below.



The following number–theoretic result, sometimes called the 'postage stamp lemma', is helpful.

Lemma 3.13 *Suppose that $m_1, m_2, \ldots, m_k \in \mathbb{N}$, and that $gcd\{m_1, m_2, \ldots, m_k\} = 1$. Then there is a number $L \in \mathbb{N}$ such that for any j in$\mathbb{N}$ with $j \geq L$, we can find nonnegative integers $a_1, \ldots, a_k$ such that $a_1 m_1 + a_2 m_2 + \ldots + a_k m_k = j$.*

Here is a characterisation of primitivity based on directed graphs.

Theorem 3.14 *Suppose that A is an $n \times n$ nonnegative matrix. Then A is primitive if and only if:*
a) $D(A)$ is strongly connected; and
b) the greatest common divisor of the lengths of the cycles in $D(A)$ is equal to 1.

Suppose that A is a primitive matrix of order n. So, there's some $k \in \mathbb{N}$ such that $A^k > 0$, but how big might such a k be? To address that question, we have the following definition. The *exponent* of a primitive matrix A, denoted by $exp(A)$, is defined as

$$exp(A) = \min\{k \in \mathbb{N} | A^k > 0\}.$$

Observe that for any $l \geq exp(A)$, we have $A^l > 0$ as well. Since the exponent depends only on positioning of the positive entries in A and its powers, and not on the sizes of those positive entries, the directed graph $D(A)$ can be used to discuss the exponent. The following result provides an attainable upper bound on the exponent of a primitive matrix.

Theorem 3.15 *Let A be a primitive matrix of order n. Consider $D(A)$, and let s denote the length of a shortest cycle in $D(A)$. Then we have*

$$exp(A) \leq n + s(n-2)$$

and

$$exp(A) \leq (n-1)^2 + 1. \tag{3}$$

Further, equality holds in (3) *if and only if there is a permutation matrix P and a collection of positive numbers $a_1, \dots a_{n+1}$ such that either PAP^T or PA^TP^T has the following form:*

$$\begin{bmatrix} 0 & a_1 & 0 & \dots & 0 \\ 0 & 0 & a_2 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ a_n & 0 & 0 & \dots & a_{n-1} \\ a_{n+1} & 0 & 0 & \dots & 0 \end{bmatrix}. \tag{4}$$

3.4 An application to web search: the *HITS* algorithm

We can view the world wide web as a directed graph. The vertices represent individual web pages, with an arc from vertex i to vertex j whenever the i–th web page has a link out to the j–th web page. If we have an arc from i to j, we can think of page i as 'endorsing' page j, and that philosophy informs the hypertext induced topic search, or *HITS*, algorithm. Here's the idea.

Suppose that we have identified a collection of web pages that deal with a common topic. We consider two interrelated notions of the quality of a particular page. If there are lots of pages that point to page i, then we might think of i as an 'authority', while if there are a lot of links issuing from page j, we can think of page j as a 'hub'. *HITS* proceeds by developing two measures associated with a page, its hub score, and its authority score. These scores are informed by the following philosophy: good authorities are pointed to by good hubs, and good hubs point to good authorities.

Here is how we can take that philosophy and produce something quantitative from it. Suppose that we have a directed graph D representing the link structure of a given collection of n web pages. Label the pages with the numbers $1, \dots, n$, and construct the $n \times n$ *adjacency matrix* A for the directed graph D as follows: for each $i, j = 1, \dots, n$, set

$$a_{i,j} = \begin{cases} 1, & \text{if } i \to j \text{ in } D, \\ 0, & \text{otherwise.} \end{cases}$$

With the adjacency matrix in hand, we now construct two sequences of vectors in $\mathbb{R}^n$, one to approximate the authority scores of the vertices, and the other to approximate the hub scores. Start with two initial vectors $x(0), y(0) \geq 0$, which are our starting approximations for the authority score vectors and hub score vectors, respectively.

For each $k \in \mathbb{N}$, we iterate as follows:

$$x_i(k) = \sum_{j \ni j \to i} y_j(k-1) \tag{5}$$

$$y_i(k) = \sum_{j \ni i \to j} x_j(k). \tag{6}$$

Equation (5) can be interpreted as saying that the k–step authority score for vertex i is given by the sum of the $k-1$–step hub scores of the vertices having outarcs to vertex i. From (6), we see that the k–step hub score for vertex i is the sum of the k–step authority scores of the vertices to which i has an outarc. We can recast (5) and (6) in matrix–vector terms as

$$x(k) = A^T y(k-1), y(k) = Ax(k),$$

which in turn yields

$$x(k) = A^T Ax(k-1), y(k) = AA^T y(k-1).$$

It now follows that $x(k) = (A^T A)^k x(0), y(k) = (AA^T)^k y(0), k \in \mathbb{N}$. Since we're interested in the relative sizes of the entries in $x(k)$ (respectively, $y(k)$), we consider the related sequences $\tilde{x}(k) = \frac{1}{\mathbf{1}^T x(k)} x(k), \tilde{y}(k) = \frac{1}{\mathbf{1}^T y(k)} y(k), k \in \mathbb{N}$.

Suppose now that both AA^T and $A^T A$ are primitive, and let u and v denote the Perron vectors of AA^T and $A^T A$, respectively, normalised so that $\mathbf{1}^T u = \mathbf{1}^T v = 1$. Then we have

$$\lim_{k\to\infty} \tilde{x}(k) = u, \lim_{k\to\infty} \tilde{y}(k) = v.$$

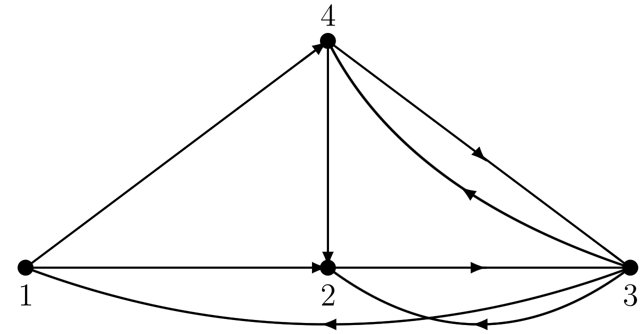
Hence we take u and v as the authority and hub vectors for D, respectively.

Example 3.16 Consider the directed graph D given above. The corresponding adjacency matrix is

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix},$$

so that

$$A^T A = \begin{bmatrix} 2 & 0 & 2 & 1 \\ 0 & 2 & 1 & 0 \\ 2 & 1 & 3 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}, AA^T = \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 3 & 0 & 2 \\ 1 & 0 & 1 & 0 \\ 1 & 2 & 0 & 2 \end{bmatrix}.$$



Computing the corresponding Perron vectors now yields

$$\tilde{x} = \begin{bmatrix} 0.3028 \\ 0.1254 \\ 0.4043 \\ 0.1675 \end{bmatrix}, \tilde{y} = \begin{bmatrix} 0.2368 \\ 0.3910 \\ 0.0561 \\ 0.3161 \end{bmatrix}.$$

The implementation of *HITS* is query–dependent. That is, once the user enters the term(s) on which to be searched, a search of web pages containing those terms (and possibly some semantically–related terms) is performed. That yields the collection of web pages that form the vertex set of the directed graph D; we then include in D all arcs of the form $i \to j$ where i and j are among the pages containing the search terms, and where page i has a link out to page j. Apparently a version of *HITS* is used by the search engine *Teoma*, which now belongs to *Ask.com*.

Remarks:

• The matrix AA^T is primitive if and only if the corresponding directed graph is strongly connected. A similar comment applies to A^TA.

• If A contains neither a zero row nor a zero column, then AA^T is primitive if and only if A^TA is primitive.

• If the authority score vector $\tilde{x}$ has been computed, we may find $\tilde{y}$ easily via the formula $\tilde{y} = \frac{1}{\mathbf{1}^T A\tilde{x}} A\tilde{x}$.

• In practice the vector $\tilde{x}$ can often be estimated by the *power method* – that is, by selecting some nonnegative initial vector $x(0)$, and then computing $x(k) = A^TAx(k-1)$ for several iterations (say, 10–15).

3.5 Irreducible imprimitive matrices

We have seen in the sections above that much can be said about primitive nonnegative matrices. Suppose that $A \geq 0$ and that A is irreducible but not primitive. It turns out that in this setting, a version of the Perron–Frobenius theorem still holds, as we will see below. If our matrix is A is irreducible but not primitive, then we find from Theorem 3.14 that the greatest common divisor of the lengths of the cycles in $D(A)$, sat d, must exceed 1. In this case, we say that our irreducible nonnegative matrix A is *periodic with period d*.

For an irreducible nonnegative matrix A that is periodic with period d, we can simultaneously permute the rows and columns of A to put it in *periodic normal form*. That is, there is a permutation matrix P so that PAP^T has the form

$$PAP^T = \left[\begin{array}{c|c|c|c|c} 0 & A_1 & 0 & \dots & 0 \\ \hline 0 & 0 & A_2 & \dots & 0 \\ \hline & & \ddots & \ddots & \\ \hline 0 & 0 & \dots & 0 & A_{d-1} \\ \hline A_d & 0 & \dots & 0 & 0 \end{array}\right]. \tag{7}$$

Further, each of the cyclic products $A_1 \dots A_d, A_2A_2 \dots A_dA_1, \dots A_dA_1 \dots A_{d-1}$ is primitive, with Perron value equal to $\rho(A)^d$. Observe that if x is a right Perron vector for the matrix $A_dA_1 \dots A_{d-1}$, then for any complex number λ of the form $\rho(A)e^{\frac{2\pi j}{d}}, j = 0, 1, \dots, d-1$, the vector

$$\begin{bmatrix} A_1A_2 \dots A_{d-1}x \\ \lambda A_2 \dots A_{d-1}x \\ \lambda^2 A_3 \dots A_{d-1}x \\ \vdots \\ \lambda^{d-2}A_{d-1}x \\ \lambda^{d-1}x \end{bmatrix}$$

is an eigenvector of the matrix in (7) corresponding to the eigenvalue λ.

Theorem 3.17 (The Perron–Frobenius theorem for irredubile periodic matrices) *Suppose that A is an irreducible nonnegative matrix of order n, and that A is periodic with period d. Then:*
a) the spectral radius, $\rho(A)$, is an eigenvalue of A;
b) $\rho(A)$ is an algebraically simple eigenvalue of A;
c) A has positive right and left eigenvectors corresponding to the eigenvalue $\rho(A)$;
d) the left and right eigenspaces of A corresponding to $\rho(A)$ both have dimension 1;
e) there are precisely d eigenvalues of A having modulus $\rho(A)$, namely $\rho(A)e^{\frac{2\pi i j}{d}}, j = 0, 1, \dots, d-1$, each of which is an algebraically simple eigenvalue of A;
f) if B is an $n \times n$ matrix such that $0 \le B \le A$, then $\rho(B) \le \rho(A)$, and if $\rho(B) = \rho(A)$, then in fact $B = A$;
g) if x is a positive right eigenvector of A, then necessarily $Ax = \rho(A)x$, with an analogous statement holding for positive left eigenvectors.

The following weaker version of Theorem 3.3 holds for irreducible periodic matrices.

Theorem 3.18 *Suppose that A is an irreducible nonnegative matrix that is periodic with period d. Let x and y denote right and left Perron vectors for A, respectively, normalised so that $y^T x = 1$. Then*

$$\lim_{k\to\infty} \frac{1}{d} \sum_{j=k}^{k+d-1} \frac{1}{\rho(A)^j} A^j = xy^T.$$

Example 3.19 Condider the adjacency matrix A of the directed graph in Figure 3.3. Then

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

The eigenvalues of A are $\sqrt{2}, -\sqrt{2}, 0$, with corresponding (right) eigenvectors given by

$$\begin{bmatrix} \frac{1}{2} \\ \frac{1}{\sqrt{2}} \\ \frac{1}{2} \end{bmatrix}, \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{2} \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{bmatrix},$$

respectively. Observe that for the permutation matrix

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

we have

$$PAP^T = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

which is in periodic normal form.

Given an irreducible nonnegative matrix A of order n, how can we determine a) whether it's primitive, and b) if it's not primitive, what its period is. Here is straightforward techniques for answering those questions; not surprisingly, it relies on the directed graph $D(A)$.

1. Pick an index i between 1 and n, and set $S_0 \equiv \{i\}$.
2. For each $j = 1, 2, \dots,$ define S_j by $S_j = \{k | p \to k \text{ for some } p \in S_{j-1}\}$.

If we initialise with step 1, and then iterate with step 2, one of two things will happen:
a) for some $m \in \mathbb{N}, S_m = \{1, 2, \dots, n\}$; in this case, A is primitive;
b) there is a $d \geq 2$, an $m \in \mathbb{N}$ and sets $T_1, \dots T_d$ such that for all $k \geq m, S_k = T_{k \bmod d}$; in this case, A is periodic with period d.

Example 3.20 Consider the matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Starting with $S_0 = \{1\}$, we iterate the procedure as above to find that $S_2 = \{2\}, S_3 = \{3\}, S_4 = \{4\}, S_5 = \{5\}, S_6 = \{6\}, S_7 = \{1,4\}, S_8 = \{2,5\}, S_9 = \{3,6\}$. Further, for all $k \geq 7$, we find that

$$S_k = \begin{cases} S_7, & if\ k \equiv 1 \mod 3, \\ S_8, & if\ k \equiv 2 \mod 3, \\ S_9, & if\ k \equiv 0 \mod 3. \end{cases}$$

Hence A is periodic with period 3.

We close the section with the following result on irreducible nonnegative matrices.

Theorem 3.21 *Let A be an irreducible nonnegative matrix, and suppose that $r > \rho(A)$. Then $rI - A$ is invertible, and in fact $(rI - A)^{-1} = \sum_{j=0}^{\infty} \frac{1}{r^{j+1}} A^j$. In particular, $(rI - A)^{-1}$ has all positive entries.*

3.6 Reducible nonnegative matrices

Recall that a square matrix A is reducible if there is a permutation matrix P such that PAP^T has a 2×2 block upper triangular form. Iterating that definition, it follows that for any reducible nonnegative matrix A, there is a permutation matrix Q such that QAQ^T has the form

$$QAQ^T = \begin{bmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,m} \\ 0 & A_{2,2} & \dots & A_{2,m} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & A_{m,m} \end{bmatrix}, \tag{8}$$

where each $A_{j,j}$ is either square and irreducible, or is a 1×1 zero matrix. The block triangular form in (8) is known as the *Frobenius normal form for A*. It is straightforward to see that for such a matrix A, we have $\rho(A) = \max\{\rho(A_{j,j}) | j = 1, \dots, m\}$, so that for reducible nonnegative matrices, the spectral radius is an eigenvalue. Observe that a reducible nonnegative matrix may or may not have a positive right eigenvector corresponding to the spectral radius; for instance consider the examples

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}.$$

Suppose that we have a reducible nonnegative matrix A, and consider the directed graph $D(A)$. We say that two vertices i, j of $D(A)$ *communicate* if either $i = j$, or there are paths from i to j and from j to i in $D(A)$. Evidently communication is an equivalence relation, and the corresponding equivalence classes (which are sets of indices) are called the *classes* for A. It is straightforward to determine that in fact the classes for A are the same as the subsets of indices that correspond to diagonal blocks in the Frobenius normal form for A. Suppose that we have two classes S and T for A. We say that S has *access* to T if there is an $i \in S$ and a $j \in T$ such that i has access to j. A class is called *final* if it does not have access to any other class. Observe that any class is either a final class, or it has access to a final class. Lastly, we say that a class S is *basic* if $\rho(A[S]) = \rho(A)$, where $A[S]$ is the principal submatrix of A on the rows and columns indexed by S.

The following result addresses the existence of positive eigenvectors for reducible nonnegative matrices.

Theorem 3.22 *Suppose that A is a nonnegative matrix. There is a positive right eigenvector for A corresponding to $\rho(A)$ if and only if the final classes for A coincide with the basic classes for A.*

4 Stochastic matrices and Markov chains

A square nonnegative matrix T of order n is *stochastic* if $T\mathbf{1} = \mathbf{1}$, that is, if every row sum of T is equal to 1. Suppose that we have a nonnegative vector $v \in \mathbb{R}^n$ whose entries sum to 1, and observe that v can be thought of as a probability vector, in the sense that the entries of v constitute a probability distribution on the numbers $1, \ldots, n$. Note also that for any $k \in \mathbb{N}$, the row vector $v^T T^k$ is nonnegative (since both v and T are nonnegative), and that $v^T T^k \mathbf{1} = \mathbf{1}$. Thus each row vector $v^T T^k$ is a probability vector in $\mathbb{R}^n$. A sequence of row vectors of the form $v^T, v^T T, v^T T^2, \ldots$ is an example of what's known as a Markov chain; we now give a more formal development of that notion.

Suppose that $n \in \mathbb{N}$, and consider a sequence of random variables defined on the set $\{1, \ldots, n\}$, say $u_k, k \in \mathbb{N}$. The indices $1, \ldots, n$ are known as the *states* of the Markov chain, and the set $\{1, \ldots, n\}$ is the *state space*. Let $Pr\{A\}$ and $Pr\{A|B\}$ denote the probability of the event A, and the conditional probability of the event A

given event B, respectively. We say that the sequence u_k has the *Markov property* if, for each $k \in \mathbb{N}$

$$Pr\{u_{k+1}|u_1, u_2, \ldots, u_k\} = Pr\{u_{k+1}|u_k\}.$$

A sequence of random variables that enjoys the Markov property is known as a *Markov chain*. A Markov chain is said to be *time homogeneous* if there is a collection of fixed probabilities $t_{i,j}, i, j = 1, \ldots, n$ such that for each $k \in \mathbb{N}$ and $i, j = 1, \ldots, n$ we have

$$Pr\{u_{k+1} = j|u_k = i\} = t_{i,j}.$$

In this setting we refer to the quantity $t_{i,j}$ as the *transition probability from state i to state j* for the Markov chain. For a time homogeneous Markov chain, we may construct the corresponding *transition matrix* T as $T = [t_{i,j}]_{i,j=1,\ldots,n}$. We can also represent the iterates u_k of the Markov chain as vectors

$$v_k \equiv \begin{bmatrix} Pr\{u_k = 1\} \\ Pr\{u_k = 2\} \\ \vdots \\ Pr\{u_k = n\} \end{bmatrix}.$$

Observe that $v_k^T \mathbf{1} = 1$ for each $k \in \mathbb{N}$. It is straightforward to verify that for each $k \in \mathbb{N}$,

$$v_{k+1}^T = v_k^T T.$$

This last relation can be reformulated equivalently as

$$v_{k+1}^T = v_1^T T^k, k \in \mathbb{N}. \tag{9}$$

Evidently we may view the iterates of a time homogeneous Markov chain on the state space $\{1, \ldots, n\}$ as realisation of the power method, whereby powers of the stochastic matrix T are applied to an initial vector v_1^T.

Observe that for any stochastic matrix T, we have $\rho(T) = 1$, with $\mathbf{1}$ as a corresponding right eigenvector.

Theorem 4.1 *(Perron–Frobenius theorem for stochastic matrices) Suppose that T is an irreducible stochastic matrix of order n. Then we have following:*
a) $\rho(T) = 1$ is an algebraically and geometrically simple eigenvalue of T;
b) the right eigenspace corresponding to the eigenvalue 1 is spanned by $\mathbf{1}$;
c) there is a positive left eigenvector w corresponding to the eigenvalue 1, normalised so that $w^T \mathbf{1} = 1$;
d) if T is primitive, then for any eigenvalue $\lambda \neq 1$ we have $|\lambda| < 1$;
e) if T is periodic with period d, then each of $e^{\frac{2\pi i j}{d}}, j = 0, \ldots, d-1$ is an algebraically simple eigenvalue of T, while all remaining eigenvalues have modulus strictly less than 1.

For an irreducible stochastic matrix T, the left eigenvector w of Theorem 4.1 is known as the *stationary distribution vector* for the corresponding Markov chain, and is a central quantity of interest.

Corollary 4.2 *Suppose that T is a primitive stochastic matrix of order n with stationary distribution vector w. Then for any initial vector $v \in \mathbb{R}^n$ with $v \geq 0, \mathbf{1}^T v = 1$, we have*

$$\lim_{k\to\infty} v^T T^k = w^T.$$

Thus, a Markov chain with transition matrix T converges to the stationary distribution vector w, independently of the initial distribution.

In view of Corollary 4.2, we have the interpretation that each entry in the stationary vector represents the long–term probability that the Markov chain is in the corresponding state. So, a large entry in the stationary distribution vector corresponds to a state that is visited frequently by the Markov chain, while a small entry in the stationary distribution vector corresponds to an infrequently visited state.

Example 4.3 Let G be a connected undirected graph on vertices $1, \ldots, n$. Recall that for each $i = 1, \ldots, n$, the *degree* of vertex i, which we denote d_i, is the number of edges incident with vertex i. The *simple random walk on* G is the Markov chain whose states are the vertices $1, \ldots, n$, and where for each pair of vertices $i, j = 1, \ldots, n$, we define the transition probability

$$t_{i,j} = \begin{cases} \frac{1}{d_i}, & \text{if } i \text{ is adjacent to } j \\ 0, & \text{otherwise.} \end{cases}$$

Letting A be the $(0,1)$ adjacency matrix of G and $D = diag(d_1, \ldots, d_n)$, we find readily that $T = D^{-1}A$. Since G is connected, the matrix T is irreducible. Further, letting w be the vector

$$w = \frac{1}{\sum_{j=1}^n d_j} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix},$$

we sse that $w^T T = \frac{1}{\sum_{j=1}^n d_j} \mathbf{1}^T D D^{-1} A = \frac{1}{\sum_{j=1}^n d_j} \mathbf{1}^T A = \frac{1}{\sum_{j=1}^n d_j} \mathbf{1}^T D = w^T$. Hence w is the stationary vector for T. Since w is a scalar multiple of the degree sequence for G, we see that vertices of large degree correspond to frequently visited states in the simple random walk on G.

Theorem 4.4 *Suppose that we have an irreducible stochastic matrix T of order n with stationary vector w. Fix an index k between 1 and $n-1$, and partition T and w conformally as*

$$T = \left[\begin{array}{c|c} T_{1,1} & T_{1,2} \\ \hline T_{2,1} & T_{2,2} \end{array}\right],$$

where $T_{1,1}$ and $T_{2,2}$ are of orders k and $n-k$, respectively. Then we have the following conclusions.
a) The matrices $S_1 \equiv T_{1,1} + T_{1,2}(I - T_{2,2})^{-1}T_{2,1}$ and $S_2 \equiv T_{2,2} + T_{2,1}(I - T_{1,1})^{-1}T_{1,2}$ are irreducible and stochastic, of orders k and $n-k$, respectively.
b) Denoting the stationary distribution vectors of S_1 and S_2 by u_1 and u_2, respectively, the vector w is given by

$$w = \left[\begin{array}{c} a_1 u_1 \\ \hline a_2 u_2 \end{array}\right],$$

where the vector

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

is the stationary vector of the 2×2 matrix

$$C = \begin{bmatrix} u_1^T T_{1,1}\mathbf{1} & u_1^T T_{1,2}\mathbf{1} \\ u_2^T T_{2,1}\mathbf{1} & u_2^T T_{2,2}\mathbf{1} \end{bmatrix}.$$

The matrices S_1, S_2 in Theorem 4.4 are called *stochastic complements*, while the matrix C is known as the *coupling matrix*. Observe that Theorem 4.4 allows for the following 'divide and conquer' strategy for computing the stationary vector.
i) partition T;
ii) compute the stochastic complements S_1, S_2;
iii) find the stationary vectors u_1, u_2 for S_1 and S_2;
iv) find the coupling matrix C and its stationary distribution vector;
v) assemble the stationary vector w from u_1, u_2, a_1, a_2.
Note also that at step iii) we could iterate the divide and conquer strategy if we like.

Suppose that we have a reducible stochastic T matrix written in Frobenius normal form as

$$\begin{bmatrix} T_{1,1} & T_{1,2} & \dots & T_{1,m} \\ 0 & T_{2,2} & \dots & T_{2,m} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & T_{m,m} \end{bmatrix}.$$

Suppose that we have a basic class for T, say corresponding to the irreducible diagonal block $T_{j,j}$. Since T is a stochastic matrix, we see that $T_{j,j}\mathbf{1} \le \mathbf{1}$, and letting u be the stationary vector for $T_{j,j}$, we find that on the one hand $u^T T_{j,j}\mathbf{1} = u^T\mathbf{1} = 1$ (since

$\rho(T_{j,j}) = 1$), while on the other hand $u^T T_{j,j}\mathbf{1} \leq u^T\mathbf{1} = 1$. We deduce that in fact we must have $T_{j,j}\mathbf{1} = \mathbf{1}$, from which it follows that the blocks $T_{j,k}, k = j+1, \ldots, m$, must all be zero blocks. Noting also that necessarily $T_{m,m}$ is a basic class, we deduce that T has the number 1 as an algebraically simple eigenvalue if and only if the only basic class for T corresponds to the block $T_{m,m}$. When that holds, there is again a uniques left eigenvector w of T corresponding to the eigenvalue 1, and is normalised so that $w^T\mathbf{1} = 1$. We refer to this w also as the stationary distribution vector, though observe that it may not be a positive vector.

The irreducible stochastic matrices are, in some sense, fundamental to the study of spectral properties of all irreducible nonnegative matrices, as the following result makes clear.

Theorem 4.5 *Suppose that A is an irreducible nonnegative matrix of order n. Then there is a diagonal matrix D whose diagonal entries are positive, and an irreducible stochastic matrix T such that $A = \rho(A)DTD^{-1}$.*

The argument to establish Theorem 4.5 is straightforward. Let x be a right Perron vector of A, and set $D = diag(x)$; it's then easy to check that $\frac{1}{\rho(A)}D^{-1}AD$ is a stochastic matrix that has the same directed graph as A, and so is irreducible. In view of Theorem 4.5, we see that questions about eigenvalues/eigenvectors of irreducible nonnegative matrices can be referred to the corresponding questions for the irreducible stochastic case.

4.1 An application to web search: the *PageRank* algorithm

We return to the model of the world wide web as a directed graph, where vertices represent web pages, and directed arcs represent links between pages. This time, we adopt the perspective of the theory of Markov chains in order to produce a ranking of the importance of web pages. What we will describe is called the *PageRank* algorithm, which is used by the search engine *Google*.

We begin our description of the algorithm in a simplified setting. Suppose that we have a strongly connected directed graph Δ which models the world wide web, and let A be its adjacency matrix. Let $D = diag(A\mathbf{1})$ – that is, D is the diagonal matrix whose i–th diagonal entry is the number of arcs the leave vertex i; graph theorists call this the *outdegree* of vertex i. Now let T be the stochastic matrix $D^{-1}A$, which is the transition matrix for the simple random walk on the directed graph Δ. (Note the similarity with Example 4.3.) We imagine a 'random surfer' on the web who begins on some initial web page, selects a link at random from the outlinks listed on that page, clicks the selected outlink to arrive on a new page, then repeats the process on each subsequent page. Thus the random surfer is performing a simple random walk on Δ and so the sequence of pages visited is a realisation of the Markov chain with transition matrix T. If T is primitive, then each entry of its stationary distribution vector describes the long–term probability that the surfer is in the corresponding state. The interpretation then is that large entries in the stationary distribution vector correspond to important web pages, in the sense that the internal structure of the web directs the random surfer towards those web pages. The *PageRank* algorithm then proposes to use the stationary distribution vector to rank the importance of the web pages.

There are, however, a couple of implementational features that need to be addressed: i) many web pages, such as pictures, or pdf documents, do not contain any outgoing links; ii) the directed graph for the world wide web is not even close to being strongly connected (let alone primitive). To deal with these issues, two amendments to the approach above are adopted.

i) For each vertex in the directed graph that having outdegree 0, we let the corresponding row of T be $\frac{1}{n}\mathbf{1}^T$, where n is the number of vertices in the directed graph. For vertices having outdegree 1 or more, the corresponding row of T is constructed as above. It is readily verified that the matrix T is stochastic. The random surfer interpretation is now modified slightly – on a page having outgoing links, the surfer clicks one of them at random, while if the surfer lands on a page with no outgoing links, he/she jumps to a page on the web chosen uniformly at random.

ii) As noted above, T is stochastic, but because of the (very) disconnected nature of the world wide web, T likely to be reducible. To address that issue, we take a positive vector $u \in \mathbb{R}^n$ such that $\mathbf{1}^T u = 1$, and a scalar $c \in (0, 1)$, and form the matrix $G = cT + (1-c)\mathbf{1}u^T$. This matrix G is usually known as the *Google matrix* and note that G is positive and stochastic. The Markov chain associated with G still admits an interpretation in terms of a random surfer: at each time step, with probability c the surfer follows links at random as described by T, while with probability $1-c$, the surfer selects a web page with probability determined by the corresponding entry in u, and jumps to that page. The vector u is often known as the *teleportation vector*.

With these modifications in place, the stationary distribution of G is used to to rank the importance of web pages. This stationary distribution is known as the *PageRank* vector. Google reports using the power method to approximate the stationary distribution vector of G – that is, they take an initial nonnegative vector v whose entries sum to 1, and compute a number of terms in the sequence $v^T G^k, k \in \mathbb{N}$. Since G is primitive, we know that this sequence will converge to the stationary distribution vector, but a key practical question is how fast the rate of convergence is. Referring to the Jordan form for G, it is evident that the rate of convergence is governed, asymptotically, by the eigenvalues of G. Fortunately, the following results gives good information regarding the eigenvalues of G.

Theorem 4.6 *Let M be a real square matrix of order n with eigenvalues $\lambda_1, \ldots, \lambda_n$, and assume that λ_1 is an algebraically simple eigenvalue, with corresponding right eigenvector x. Let c be a real scalar and let v be a vector in $\mathbb{R}^n$. Then the eigenvalues of the matrix $cM + (1-c)xv^T$ are: $c\lambda_1 + (1-c)v^T x, c\lambda_2, \ldots, c\lambda_n$.*

Corollary 4.7 *Let T be an $n \times n$ stochastic matrix with eigenvalues $1, \lambda_2, \ldots, \lambda_n$, and suppose that $v \in \mathbb{R}^n$ with $v \geq 0$ and $v^T \mathbf{1} = 1$. Then the eigenvalues of the matrix $G = cT + (1-c)\mathbf{1}v^T$ are $1, c\lambda_2, \ldots, c\lambda_n$.*

From Corollary 4.7, all of the non–Perron eigenvalues of the Google matrix have moduli bounded above by c, and in practice it turns out that G has several eigenvalues with modulus equal to c. Consequently, the asymptotic rate of convergence for the power method applied G is given by c. Interestingly enough for the Google matrix G, it has been shown that a version of the divide and conquer technique inspired by Theorem 4.4, used in conjunction with the power method (applied to the stochastic complements) has an asymptotic rate of convergence that is at least as good as that of the power method applied directly to G.

4.2 Mean first passage times

Suppose that we have an irreducible stochastic matrix T of order n. As noted above, the entries in the stationary distribution vector give us long–term information about the behaviour of the corresponding Markov chain. However, we may be interested in the behaviour of the Markov chain in the short or medium term. To help understand the nature of the Markov chain in this shorter time frame, we consider the so–called first passage times, which we now define.

Fix a pair of indices i and j between 1 and n, and consider a Markov chain with transition matrix T. The *first passage time* from state i to state j, $f_{i,j}$, say, is the random variable taking on the value given by the smallest $k \geq 1$ such that the chain is in state j after k steps, given that the chain started in state i. Let $m_{i,j}$ denote the *mean first passage time* from state i to state j – i.e. the expected value of $f_{i,j}$. The following conditional expectation argument establishes a connection between the $m_{i,j}$s and the entries in T.

Suppose that the Markov chain is initially in state i. We may compute $m_{i,j}$ by conditioning on the state of the Markov chain after one step has been taken. After one step, the chain is in state j (with probability $t_{i,j}$) or it is in some state $k \neq j$ (with probability $t_{i,k}$). We thus find that

$$m_{i,j} = t_{i,j} + \sum_{k=1,\ldots,n,k\neq j} t_{i,k}(m_{k,j} + 1) = 1 + \sum_{k=1,\ldots,n,k\neq j} t_{i,k} m_{k,j}. \tag{10}$$

Let $M = [m_{i,j}]_{i,j=1,\ldots,n}$ denote the *mean first passage matrix* for the Markov chain. Then (10) can be rewritten as

$$M = T(M - M_{dg}) + J, \tag{11}$$

where J is the $n \times n$ all ones matrix, and where for any $n \times n$ matrix A, we define A_{dg} as

$$A_{dg} = \mathrm{diag}\left(\left[\begin{array}{ccc} a_{1,1} & \ldots & a_{n,n} \end{array}\right]\right).$$

Let w denote the stationary distribution vector for T, and note from (11) that $w^T M = w^T T(M - M_{dg}) + w^T J$, which yields $w^T M_{dg} = \mathbf{1}^T$. Hence we find that for each $i = 1, \ldots, n, m_{i,i} = \frac{1}{w_i}$. Observe that this gives us another interpretation of the entries in the stationary distribution vector in terms of the *mean first return times* for the Markov chain associated with T.

We would like to derive a formula for M in terms of T. To do this, we introduce the notation $W = diag(w)$. From (11), we find that $(I - T)M = -TM_{dg} + J = -TW^{-1} + J$. Next, we consider the matrix $I - T + \mathbf{1}w^T$, and note that it's invertible. Since $(I - T + \mathbf{1}w^T)M = -TW^{-1} + \mathbf{1}(\mathbf{1} + w^T M)$, it follows that $M = -(I - T + \mathbf{1}w^T)^{-1}TW^{-1} + \mathbf{1}(\mathbf{1} + w^T M) = (I - (I - T + \mathbf{1}w^T)^{-1})W^{-1} + \mathbf{1}w^T M$. We already know that the diagonal of M is given by W^{-1}, so we find that in fact

$$M = (I - (I - T + \mathbf{1}w^T)^{-1} + J((I - T + \mathbf{1}w^T)^{-1})_{dg})W^{-1}. \tag{12}$$

The matrix $(I - T + \mathbf{1}w^T)^{-1}$ is known as the *fundamental matrix* for the Markov chain.

Suppose that we have an irreducible stochastic matrix T of order n with stationary distribution vector w and mean first passage matrix M. Referring to (12), we see that

$$\begin{aligned} Mw = (I - (I - T + \mathbf{1}w^T)^{-1} + J((I - T + \mathbf{1}w^T)^{-1})_{dg})W^{-1}w = \\ (I - (I - T + \mathbf{1}w^T)^{-1} + J((I - T + \mathbf{1}w^T)^{-1})_{dg})\mathbf{1} = \\ (\mathbf{1}^T((I - T + \mathbf{1}w^T)^{-1})_{dg}\mathbf{1})\mathbf{1}. \end{aligned}$$

Consequently, we see that for each index i between 1 and n, the expression $\sum_{j=1}^n m_{i,j}w_j$ is independent of i and given by the quantity $\mathbf{1}^T((I - T + \mathbf{1}w^T)^{-1})_{dg}\mathbf{1}$. This quantity is known as the *Kemeny constant* for the Markov chain, and can be thought of as the expected first passage time from a given state i to a randomly chosen destination state (where the probability of selecting a particular destination state is given by the corresponding entry in the stationary distribution vector).

Example 4.8 Suppose that $a \in (0,1)$, and let

$$T = \begin{bmatrix} 0 & 0 & 0 & a & 1-a \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

The stationary distribution vector for T is given by

$$w = \frac{1}{5-a}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1-a \end{bmatrix}.$$

The mean first passage matrix for the Markov chain is then given by

$$M = \begin{bmatrix} 5-a & 4-a & 3-a & 2-a & \frac{1+3a}{1-a} \\ 1 & 5-a & 4-a & 3-a & \frac{2+2a}{1-a} \\ 2 & 1 & 5-a & 4-a & \frac{3+a}{1-a} \\ 3 & 2 & 1 & 5-a & \frac{4}{1-a} \\ 4 & 3 & 2 & 1 & \frac{5-a}{1-a} \end{bmatrix}.$$

The Kemeny constant here is equal to $\frac{15-a}{5-a}$.

5 Sample exercises

1. Suppose that A is an $n \times n$ matrix that is singular. The *index of* A is defined as being the smallest $k \in \mathbb{N}$ such that $rank(A^{k+1}) = rank(A^k)$. Prove that the index of A is the same as the order of the largest Jordan block associated with the eigenvalue 0 for A.

2. Let A be an irreducible nonnegative matrix of order n. Let r be the vector of row sums of A – i.e. $r = A\mathbf{1}$. Prove that

$$\min\{r_j | j = 1, \ldots, n\} \leq \rho(A) \leq \max\{r_j | j = 1, \ldots, n\}.$$

Prove that the following are equivalent: i) $\rho(A) = \max\{r_j | j = 1, \ldots, n\}$; ii) $\rho(A) = \min\{r_j | j = 1, \ldots, n\}$; iii) r is a scalar multiple of $\mathbf{1}$.

3. Suppose that A is a $k \times k$ $(0,1)$ matrix. Prove that the matrix AA^T to be primitive if only if, for any pair of indices i, j between 1 and k, there are indices $i \equiv i_0, i_1, i_2, \ldots, i_{r-1}, i_r \equiv j$ and indices $l_0, l_1, \ldots, l_{r-1}$ such that $a_{i_p, l_p} = 1$ for each $p = 0, 1, \ldots r-1$, and $a_{i_{p+1}, l_p} = 1$ for each $p = 0, 1, \ldots r-1$.

4. Suppose that $k \in \mathbb{N}$. What is the smallest $M \in \mathbb{N}$ such that every integer greater than or equal to M can be written in the form $ak + b(k+1)$ for some pair of nonnegative integers a and b?

5. Suppose that A is a square reducible nonnegative matrix written in Frobenius normal form. Prove that every class of A is either a final class, or has access to a final class.

6. Let A be an $n \times n$ nonnegative matrix such that $\rho(A) = 0$. Prove that there is a $k \in \mathbb{N}$ such that $A^k = 0$. Under what circumstances does A have a positive right eigenvector corresponding to the spectral radius?

7. Suppose that T is an irreducible stochastic matrix of order n, partitioned as in Theorem 4.4. Prove that the stochastic complements S_1 and S_2 are also irreducible and stochastic. Give an example to show that even if T is primitive, one of the stochastic complements may not be primitive.

8. Let T be an irreducible stochastic matrix of order n, with eigenvalues $1 \equiv \lambda_1, \lambda_2, \ldots, \lambda_n$. Prove that the Kemeny constant for the Markov chain associated with T is equal to $\sum_{j=2}^{n} \frac{1}{1-\lambda_j}$.

References

[1] A. Berman and R. Plemmons. *Nonnegative Matrices in the Mathematical Sciences.* SIAM, Philadelphia, 1994.

[2] R. Brualdi and H. Ryser. *Combinatorial Matrix Theory.* Cambridge University Press, Cambridge, 1991.

[3] H. Caswell. *Matrix Population Models*, 2nd edition. Sinauer, Sunderland, 2001.

[4] I. Herstein. *Topics in Algebra*, 2nd edition. Wiley, New York, 1975.

[5] R. Horn and C. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, 1985.

[6] A. Langville and C. Meyer. *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, Princeton, 2006.

[7] J. Marsden. *Elementary Classical Analysis*. W.H. Freeman, San Francisco, 1974.

[8] E. Seneta *Non-negative Matrices and Markov Chains*. Springer-Verlag, New York, 1981.