MATH36001 Perron–Frobenius Theory 2015

"In addition to saying something useful, the Perron–Frobenius theory is elegant. It is a testament to the fact that beautiful mathematics eventually tends to be useful, and useful mathematics eventually tends to be beautiful."
Carl D. Meyer, Matrix Analysis and Applied Linear Algebra (2000)

1 Positive and Nonnegative Matrices

If A and B are $m \times n$ real matrices we write $A \geq B$ if $a_{ij} \geq b_{ij}$ for all i, j and $A > B$ if $a_{ij} > b_{ij}$ for all i, j. If $A \geq 0$ we say that A is **nonnegative**. A is **positive** if $A > 0$.

Thus $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ are positive, while $\begin{bmatrix} 1 & 0 \\ 3 & 4 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$ are nonnegative (but not positive).
Although a nonnegative real number which is not positive must be zero, the same is not true for vectors and matrices.

We denote by $|x| = (|x_i|)$ or $|A| = (|a_{ij}|)$ the vector or matrix of absolute values of the elements of x or A; clearly $|x| \geq 0$ and $|A| \geq 0$. The following facts will be needed.

Fact 1 Let $A \in \mathbb{R}^{n \times n}$ and, $x, y \in \mathbb{R}^n$. Then

- $|Ax| \leq |A|\,|x|,$ (1)
- $A > 0$ and $x \geq 0$, $x \neq 0 \;\Rightarrow\; Ax > 0,$ (2)
- $A \geq 0 \;\Rightarrow\; A^k \geq 0$ for all $k \geq 1$, $A > 0 \;\Rightarrow\; A^k > 0$ for all $k \geq 1,$ (3)
- $A \geq 0$ and $x > y > 0 \;\Rightarrow\; Ax > Ay.$ (4)

We will need $\rho(A) = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}$, the spectral radius of A.

Nonnegativity is a natural property of many measured quantities. As a consequence, nonnegative matrices arise in many branches of science and engineering. These include probability theory (Markov chains), population models, iterative methods in numerical analysis, economics (input–output models), epidemiology, stability analysis and physics.

Example 1 (Leontief's input-output matrix) This is one of the first successes of mathematical economics. Consider the consumption matrix

$$A = \begin{bmatrix} 0.4 & 0 & 0.1 \\ 0 & 0.1 & 0.8 \\ 0.5 & 0.7 & 0.1 \end{bmatrix} \qquad \begin{array}{l} \text{(steel)} \\ \text{(food)} \\ \text{(labour)} \end{array}$$

in which a_{ij} gives the amount of product j that is needed to create one unit of product i. The matrix A is nonnegative. Let $p_1 \geq 0$, $p_2 \geq 0$ and $p_3 \geq 0$ be the amount of steel, food and labour we start with, respectively. The amount consumed is Ap, and it leaves a net production of $y = p - Ap$.

Can we produce $y_1 \geq 0$ units of steel, $y_2 \geq 0$ units of food and $y_3 \geq 0$ units of labour?
In other words, given $y \geq 0$, the problem is to find a vector $p \geq 0$ such that $p - Ap = y$ or $p = (I - A)^{-1}y$. If $(I - A)^{-1}$ exists and is nonnegative then clearly $p = (I - A)^{-1}y \geq 0$ since $y \geq 0$. So the real question is: *Given $A \geq 0$, when is $(I - A)^{-1}$ a nonnegative matrix?* This question is easily answered once we know the main fact about nonnegative matrices.

Theorem 1 (Nonnegative eigenpairs) *If $A \geq 0$ then $\rho(A)$ is an eigenvalue of A and there exists an associated eigenvector $x \geq 0$ such that $Ax = \rho(A)x$.*

It follows from Theorem 1 that if $A \geq 0$, $\lambda_{\max}(A) = \rho(A)$ is real and nonnegative. Three cases need to be considered.

1) If $\lambda_{\max}(A) = \rho(A) > 1$ with corresponding eigenvector $x \geq 0$ then x is also an eigenvector of $(I - A)^{-1}$ since $(I - A)^{-1}x = x/(1 - \rho(A))$. But $1/(1 - \rho(A)) < 0$ so $(I - A)^{-1}$ takes a positive vector x and sends it to a nonpositive vector $x/(1-\rho(A))$, which means that $(I-A)^{-1}$ cannot be nonnegative.

2) If $\lambda_{\max}(A) = \rho(A) = 1$ then $(I - A)^{-1}$ fails to exist.

3) If $\lambda_{\max}(A) = \rho(A) < 1$, the nonnegativity of $(I - A)^{-1}$ is addressed with the following lemma.

Lemma 2 *If $\rho(A) < 1$ then $I - A$ is nonsingular and $(I - A)^{-1} = \sum_{k=0}^{\infty} A^k$.*

[With a stronger assumption on A, a similar result is proved in the "Norms" handout, Ex. 9.]

It follows from Lemma 2 that $(I - A)^{-1} = \sum_{k=0}^{\infty} A^k$ is a sum of nonnegative matrices, since $A \geq 0 \Rightarrow A^k \geq 0$ for all $k \geq 1$ by (3) in Fact 1. Hence $(I - A)^{-1}$ is nonnegative.

For the 3×3 matrix A of Example 1, $\rho(A) = 0.9$. Hence $(I - A)^{-1}$ is nonnegative and $p = (I - A)^{-1}y \geq 0$.

2 Positive Matrices

We first concentrate on matrices whose entries are positive.

Theorem 3 (Perron's theorem) *If $A \in \mathbb{R}^{n\times n}$ and $A > 0$ then*

(i) *$\rho(A) > 0$ (positive spectral radius).*

(ii) *$\rho(A)$ is an eigenvalue of A.*

(iii) *There is an eigenvector x with $x > 0$ and $Ax = \rho(A)\,x$.*

(iv) *The eigenvalue $\rho(A)$ has algebraic multiplicity 1 ($\rho(A)$ is a simple eigenvalue).*

(v) *All the other eigenvalues are less than $\rho(A)$ in absolute value, i.e., $\rho(A)$ is the only eigenvalue of maximum modulus.*

Proof. The proof is not examinable. We prove statements (i)–(iii) only.

(i) By definition $\rho(A) \geq 0$ with $\rho(A) = 0$ only when $\Lambda(A) = \{0\}$, in which case A is nilpotent, i.e, $A^k = O$ for some $k > 0$ (see the "Theory of Eigensystems" handout). But when each $a_{ij} > 0$, A cannot be nilpotent so $\rho(A) > 0$.

(ii)–(iii) If (μ, x) is an eigenpair of A such that $|\mu| = \rho(A) \equiv \rho$, then

$$\rho|x| = |\mu||x| = |\mu x| = |Ax| \leq |A||x| = A|x| \;\Rightarrow\; \big(A - \rho I\big)|x| \geq 0.$$

We have to show that equality holds. We argue by contradiction. Let $y = \big(A - \rho I\big)|x|$ and suppose that $y \neq 0$. Then $A > 0$ and $y \geq 0$ implies $Ay > 0$ using property (2) of Fact 1.

Since $z \equiv A|x| > 0$ there exists $\epsilon > 0$ such that $Ay > \epsilon z$, or equivalently, $\bigl(\rho + \epsilon\bigr)^{-1} Az > z$ since $y = z - \rho|x|$. Writing this inequality as $Bz > z$, where $B = \bigl(\rho + \epsilon\bigr)^{-1} A$ and successively multiplying both sides by B while using property (3) of Fact 1 produces

$$B^2 z > Bz > z, \quad B^3 z > B^2 z > Bz > z, \; \ldots \; \Rightarrow \; B^k z > z \quad \forall\, k = 1, 2, \ldots.$$

But $\lim_{k\to\infty} B^k = 0$ since $\rho(B) = \rho/(\epsilon + \rho) < 1$ so in the limit we have $0 > z$ which contradicts the fact that $z > 0$. Consequently, $0 = y = \bigl(A - \rho I\bigr)|x|$ thus $|x|$ is an eigenvector for A associated with the eigenvalue ρ.

The proof is completed by observing that $|x| = \rho^{-1} A|x| = \rho^{-1} z > 0$. $\square$

Theorem 4 (Powers of positive matrices) *If $A > 0$, x is any positive eigenvector of A corresponding to $\rho(A)$, and y is any positive eigenvector of A^T corresponding to $\rho(A) = \rho(A^T)$ then*

$$\lim_{k\to\infty} \left(\frac{A}{\rho(A)}\right)^k = \frac{xy^T}{y^T x} > 0.$$

Proof. Let $B = A/\rho(A)$. By Perron's theorem, $\rho(B) = 1$ is a simple eigenvalue and all other eigenvalues are less than 1 in absolute value. It follows that the Jordan form of B must be of the form $\begin{bmatrix} 1 & 0 \\ 0 & \widetilde{J} \end{bmatrix}$, where $\rho(\widetilde{J}) < 1$. Then $\widetilde{J}^k \to 0$ as $k \to \infty$ so that $J^k \to \operatorname{diag}(1, 0, \ldots, 0)$ as $k \to \infty$. Therefore if $B = XJX^{-1}$ is the Jordan canonical decomposition of B then

$$\lim_{k\to\infty} (A/\rho(A))^k = \lim_{k\to\infty} B^k = X \begin{bmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} X^{-1} = xp^T,$$

where x is the first column of X and p^T is the first row of X^{-1}. Let $G = xp^T$. Now y is a positive eigenvector of B^T corresponding to the eigenvalue 1. Hence $(B^T)^k y = y$ or equivalently $y^T B^k = y^T$ for all $k \geq 1$ and hence $y^T G = y^T$, that is, $(y^T x)p^T = y^T$. But $x, y > 0$ so $y^T x > 0$ and $G = xp^T = \frac{xy^T}{y^T x} > 0$. $\square$

3 Irreducible Nonnegative Matrices

Theorem 1 is as far as Perron's theorem can be generalized to nonnegative matrices without additional hypotheses. For example $A = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ shows that properties (i)–(iv) in Perron's Theorem are lost; and $A = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ shows that the property (v) is also lost. However, Frobenius showed properties (i)–(iv) of Perron's Theorem still hold for nonnegative matrices that are irreducible.

$A \in \mathbb{R}^{n\times n}$ is a **reducible matrix** when there exists a permutation matrix P such that

$$P^T AP = \begin{bmatrix} X & Y \\ 0 & Z \end{bmatrix},$$

where X and Z are both square. A is said to be **irreducible** if it is not reducible. Clearly positive matrices are irreducible. On the other hand, any matrix that has a zero row or column is reducible. A useful approach to ascertain whether a matrix is irreducible is through the directed graph of a matrix. The **directed graph** of an $n \times n$ matrix is obtained by connecting n points $P_1, \dots, P_n$ (on the real line or in the plane) by a *directed link* from P_i to P_j if $a_{ij} \neq 0$. A directed graph is **strongly connected** if for any two points P_i and P_j, there is a finite sequence of directed links from P_i to P_j.

Example 2 Draw the directed graphs of

$$A_1 = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 2 \\ 1 & 0 & 0 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 2 & 2 & 2 \end{bmatrix}.$$

The graph of A_1 is strongly connected. Neither of the graphs of A_2 or A_3 are strongly connected.

Fact 2 $A \geq 0$ is irreducible if and only if its directed graph is strongly connected.

Theorem 5 (Perron–Frobenius theorem) *If $A \geq 0$ is irreducible then*

(i) $\rho(A) > 0$.

(ii) $\rho(A)$ *is an eigenvalue of* A.

(iii) *There is an eigenvector x with $x > 0$ and $Ax = \rho(A)\, x$.*

(iv) $\rho(A)$ *is an eigenvalue of algebraic multiplicity* 1.

$\lambda_{\max}(A) = \rho(A)$ is called the **Perron root**. The **Perron vector** is the unique vector p defined by

Explain uniqueness

$$Ap = \rho(A)p, \quad p > 0, \quad \|p\|_1 = 1.$$

4 Stochastic Matrices and Markov Chains

One of the most elegant applications of Perron–Frobenius theory is the algebraic development of the theory of finite Markov chains. A **stochastic matrix** is a nonnegative matrix $P \in \mathbb{R}^{n \times n}$ in which each row sum is equal to 1:

see MATH 30017: Markov Processes

$$\sum_{j=1}^{n} p_{ij} = 1, \qquad i = 1, 2, \dots, n.$$

Theorem 6 *$P \geq 0$ is stochastic if and only if P has eigenvalue 1 with associated eigenvector $e = [1 \quad 1 \quad \dots \quad 1]^T$. Furthermore $\rho(P) = 1$ for a stochastic matrix P.*

Proof. If P is stochastic then the condition $\sum_{j=1}^{n} p_{ij} = 1$, $i = 1, 2, \ldots, n$ can be rewritten as $Pe = e$. Hence 1 is an eigenvalue with eigenvector e. Conversely, $Pe = e$ implies $\sum_{j=1}^{n} p_{ij} = 1$, $i = 1, 2, \ldots, n$ and hence P is stochastic.

For the last part of the theorem we use $\rho(P) \leq \|P\|_\infty = 1$ so $\rho(P) = 1$ since 1 is an eigenvalue. $\square$

A **Markov chain** is a probabilistic process in which the future development of the process is completely determined by the present state and not at all in the way it arose. Markov chains serve as models for describing systems that can be in a number of different states $s_1, s_2, s_3, \ldots$. The Markov chain is finite if the number of states is finite. At each time step the system moves from state s_i to state s_j with probability $p_{ij} \geq 0$. The matrix $P = [p_{ij}]$ is called a **transition matrix**. Clearly $P \geq 0$ and since $\sum_{j=1}^{n} p_{ij} = 1$ $(i = 1, 2, \ldots, n)$, P is stochastic. It can be shown that every Markov chain defines a stochastic matrix and conversely.

A **probability distribution vector** is a vector $p \geq 0$ such that $e^T p = 1$. To a Markov chain with n states is associated an initial probability distribution vector $p^{(0)} = \begin{bmatrix} p_1^{(0)} & \ldots & p_n^{(0)} \end{bmatrix}$, where $p_i^{(0)}$ is the probability that the chain starts in state i. Then the jth component of $p^{(1)} = P^T p^{(0)}$ gives the probability of being in state j after one step. Note that $e^T p^{(1)} = e^T P^T p^{(0)} = e^T p^{(0)} = 1$ so $p^{(1)}$ is again a probability distribution vector. The vector $p^{(k)} = P^T p^{(k-1)} = (P^T)^k p^{(0)}$ is called the kth step probability distribution vector.

An important problem is to find the **stationary probability distribution vector** p of a Markov chain defined by

$$P^T p = p, \quad e^T p = 1.$$

Theorem 7 *Assume that a Markov chain has a positive transition matrix P. Then independent of the initial probability distribution vector $p^{(0)}$,*

$$\lim_{k\to\infty} p^{(k)} = p,$$

where $p^{(k)} = (P^T)^k p^{(0)}$ and p is the stationary probability distribution vector.

Proof. Since P is stochastic, $\rho(P) = 1$ and $Pe = e$, where $e = \begin{bmatrix} 1 & \ldots & 1 \end{bmatrix}^T$. The stationary probability distribution vector p satisfies $P^T p = p$, $e^T p = 1$. Theorem 4 applied to $P^T > 0$ (with ($x = p$ and $y = e$) gives

$$\lim_{k\to\infty} p^{(k)} = \lim_{k\to\infty} (P^T)^k p^{(0)} = \frac{pe^T}{e^T p} p^{(0)} = \frac{e^T p^{(0)}}{e^T p} p = p$$

since, by definition, $e^T p^{(0)} = 1$. $\square$

Example 3 An association of four regions R_1, R_2, R_3 and R_4 trade in a certain nonrenewable commodity. Assume that α units of the commodity are shipped between and within the regions according to the following matrix of fractions

$$Q = \begin{bmatrix} 0.1 & 0.2 & 0.1 & 0.1 \\ 0.7 & 0.6 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.7 \end{bmatrix},$$

where q_{ij} is the fraction of the commodity present in R_j which is shipped each day to R_i. Suppose that the α units of the commodity are distributed so that there are x_1, x_2, x_3, x_4 units in regions R_1, R_2, R_3, R_4, respectively, today. This is an example of a Markov chain. Since $q_{ij}x_j$ is the amount shipped today from R_j to R_i, $\sum_{j=1}^{4} q_{ij}x_j$ is the amount present at R_i tomorrow. In general, $Q^k x$ gives the distribution of the commodity in each region k days from today.

What will the distribution of the commodity be like far into the future?

Clearly Q is positive and since $Q^T e = e$ with $e = [1 \quad 1 \quad \dots \quad 1]^T$, Q^T is stochastic. So by Theorem 6, $\lambda = 1$ is the largest eigenvalue of Q^T (and Q). As you may verify $p = [6 \quad 16 \quad 11 \quad 11]^T/44$ is the stationary probability distribution vector ($Qp = p$ and $e^T p = 1$). Now applying Theorem 7 with $p^{(0)} = x/\alpha$ so that $e^T p^{(0)} = 1$ gives

$$\lim_{k\to\infty} \alpha Q^k (x/\alpha) = \alpha p = \frac{\alpha}{44} [6 \quad 16 \quad 11 \quad 11]^T.$$

Notice that this limit depends only on Q and α and not at all on x. This tells us that regardless of how the α units were distributed among the four regions initially, in the long run region R_i will have αp_i of them, $i = 1, \dots, 4$.

Suppose now that the commodity is shipped between and within the regions according to the following matrix of fractions:

$$Q = \begin{bmatrix} X & 0 \\ Y & I \end{bmatrix}, \text{ where } X = \begin{bmatrix} 0.1 & 0.3 \\ 0.7 & 0.6 \end{bmatrix}, \quad Y = \begin{bmatrix} 0.1 & 0 \\ 0.1 & 0.1 \end{bmatrix}, \quad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Note that $Q \geq 0$ is reducible and $Q^T e = e$ so Q^T is stochastic. Moreover,

$$Q^k = \begin{bmatrix} X^k & 0 \\ Y \sum_{j=0}^{k-1} X^j & I \end{bmatrix}, \quad k \geq 1.$$

But $\rho(X) \leq \|X\|_1 = 0.9$ so $\lim_{k\to\infty} X^k = O$. Also by Lemma 2, $I - X$ is nonsingular and

$$\sum_{j=0}^{\infty} X^j = (I - X)^{-1} = \begin{bmatrix} 0.9 & -0.3 \\ -0.7 & 0.4 \end{bmatrix}^{-1} = \begin{bmatrix} 8/3 & 3 \\ 14/3 & 6 \end{bmatrix}.$$

Consequently, $\lim_{k\to\infty} Y \sum_{j=0}^{k-1} X^j = \begin{bmatrix} 4/15 & 1/5 \\ 11/15 & 4/5 \end{bmatrix}$ and

$$\lim_{k\to\infty} Q^k x = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 4/15 & 1/5 & 1 & 0 \\ 11/15 & 4/5 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \frac{1}{15} \begin{bmatrix} 0 \\ 0 \\ 4x_1 + 3x_2 + 15x_3 \\ 11x_1 + 12x_2 + 15x_4 \end{bmatrix}.$$

So eventually R_3 and R_4 will have all the commodity and how many units they have depends on the initial distribution vector x.

Example 4 (Google's PageRank) The web search engine Google uses a so-called PageRank vector to determine the order in which web pages are displayed. The limiting probability that an infinitely dedicated random surfer visits any particular web page is its PageRank. The

PageRank vector depends only on the hyper-link structure of the web graph, but it does not depend on the contents of the web pages. A page has high rank if other pages with high rank link to it. Computing PageRank amounts to computing the stationary distribution of a stochastic matrix, the "Google matrix".

The Google matrix G is a convex combination of two stochastic matrices:

$$G = \alpha S + (1-\alpha)ev^T, \quad 0 < \alpha < 1,$$

where $S \geq 0$, stochastic, represents the hyper-link structure of the web graph and $v > 0$ such that $e^T v = 1$ is the so-called personalization vector.

The Google matrix G is positive since $(1-\alpha)ev^T > 0$ and stochastic since $Ge = e$. The Perron–Frobenius theory tells us that 1 is the largest eigenvalue of G. The PageRank vector p is the Perron vector (or stationary probability distribution vector) of G^T i.e., p satisfies $G^T p = p$, $p > 0$, $e^T p = 1$. It is the only eigenvector with nonnegative components and for any initial probability distribution vector $p^{(0)}$, $\lim_{k\to\infty}(G^T)^k p^{(0)} = p$. The components of p are the Google PageRanks.

Due to the huge dimension of the Google matrix (close to 50 billion), efficiently computing the stationary distribution vector p is a challenge.

Exercises

1. (a) If $0 \neq x \geq 0$ and $A > 0$, show that $Ax > 0$.

(b) Find an example of a nonzero, nonnegative, 2×2 matrix A and a nonzero, nonnegative, 2×1 vector x such that $Ax = 0$.

(c) If $A > 0$ and $z \geq w$, show that $Az \geq Aw$, with equality iff $z = w$.

2. Let A be any 2×2 positive stochastic matrix. Show that A has the form $\begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$ for some α, β such that $0 < \alpha, \beta < 1$. Find the Perron root and Perron vector of A^T.

3. Let $A = \begin{bmatrix} 0 & 1 & 0 \\ 3 & 0 & 3 \\ 0 & 2 & 0 \end{bmatrix}$. Show that $A \geq 0$ is irreducible. Find its Perron root and Perron vector.

4. Let $A \geq 0$ be irreducible. Show that there are no nonnegative eigenvectors for A other than the Perron vector and its positive multiples.

5. Explain why $\operatorname{rank}(I - P) = n - 1$ for every irreducible stochastic matrix $P \in \mathbb{R}^{n\times n}$.

6. Let $A \in \mathbb{R}^{n\times n}$ be nonnegative. Show that if there is $x > 0$ such that $Ax = \rho(A)x$ then $P = \big(1/\rho(A)\big)D^{-1}AD$ is stochastic, where $D = \operatorname{diag}(x_1, x_2, \ldots, x_n)$.