Notes on Perron-Frobenius Theory

John R. Stembridge

Department of Mathematics
University of Michigan
Ann Arbor, Michigan 48109–1043 USA
Email: jrs@umich.edu

15 June 2011

1. Main results

The Perron-Frobenius Theorem is a collection of facts about the eigenvalues and eigenvectors of real nonnegative matrices. In these notes we provide complete proofs of the main results; the one non-trivial thing we take for granted is the existence of Jordan Canonical Form over the complex field.

We make no claims that anything stated here is new or original.

For a square complex matrix A, let $\rho(A)$ denote its spectral radius; i.e., the maximum of $|\lambda|$ as λ varies over the eigenvalues of A. If $\rho(A) = |\lambda|$, we say that λ is *extremal.*

Lemma 1. *Assume $\rho = \rho(A) > 0$, and let r denote the size of the largest Jordan block associated to an extremal eigenvalue of A.*

(a) *The rate of growth of the entries of A^n is limited; namely,*

$$(A^n)_{i,j} = O(n^{r-1}\rho^n) \quad \text{as } n \to \infty, \text{ for all } i, j.$$

(b) *If λ is the only extremal eigenvalue of A, then for all vectors v, the limit*

$$L(v) := \lim_{n\to\infty} \frac{A^n v}{n^{r-1}\lambda^n}$$

converges, and the range of L is a nonzero subspace of the λ-eigenspace of A.

Proof. Both claims are stable under changes of basis, so we may assume that A is in Jordan Canonical Form. In addition, if the claims hold for A_1 and A_2, then they hold for their direct sum. Thus we may assume that A consists of a single $r \times r$ Jordan block; i.e.,

$$A = \lambda + E, \text{ where } E = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \text{ (if } r = 3\text{, say).}$$

Since $E^r = 0$, one sees that

$$A^n = (\lambda + E)^n = \lambda^n + \binom{n}{1}\lambda^{n-1}E + \cdots + \binom{n}{r-1}\lambda^{n-r+1}E^{r-1},$$

and now (a) follows easily. For (b), we have

$$\lim_{n\to\infty} \frac{A^n}{n^{r-1}\lambda^n} = \frac{1}{(r-1)!\,\lambda^{r-1}}E^{r-1},$$

and the range of this operator is the coordinate space $\mathbb{C}e_1$, the λ-eigenspace of A. $\square$

THEOREM 2. *If all sufficiently high powers of A are real and positive, then the extremal eigenvalues of A are real and positive (i.e., if λ is extremal, then $\lambda = \rho(A) > 0$).*

For a stronger version of this result, see Theorem 5 below.

We note that the above hypothesis does not require A to be nonnegative. For example, if $A = \begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix}$, then A^n is positive for $n = 2$ and $n \geqslant 4$. (See also Corollary 10.)

Proof. (Shamelessly adapted from Wikipedia.) Let $\rho = \rho(A)$. We cannot have $\rho = 0$, otherwise A would be nilpotent and all sufficiently high powers of A would be 0.

Replacing A with A/ρ, we may assume $\rho = 1$.

Now let λ be an extremal eigenvalue of A. If $\lambda \neq 1$, then we can choose m so that A^m is positive and λ^m has negative real part. If a is the smallest diagonal entry of A^m, then $B = A^m - a/2$ is positive and has an eigenvalue $\mu = \lambda^m - a/2$ such that $|\mu| > 1$. However, $B \leqslant A^m$ entry-wise, so the matrix entries of B^n are bounded by the matrix entries of A^{mn}. Since $\rho(A) = 1$, the latter are of polynomial growth by Lemma 1. On the other hand, there is an eigenvector v for B such that $B^n v = \mu^n v$ has exponential growth, a contradiction. $\square$

LEMMA 3. *If A is (real and) nonnegative, then A has a nonnegative eigenvector with eigenvalue $\rho(A)$.*

Proof. We know that $\rho(A)$ is an eigenvalue of A if A is positive (Theorem 2), so the same is true for nonnegative A by continuity.

For the eigenvectors, we may replace A with $1 + A$. Indeed, this has no effect on the eigenspaces, but shifts the spectrum so that $\rho = \rho(A)$ is the unique extremal eigenvalue, and is positive. By Lemma 1(b), it follows that there is an integer $r \geqslant 1$ so that

$$v \mapsto L(v) = \lim_{n\to\infty} \frac{A^n v}{n^{r-1}\rho^n}$$

is a nonzero linear map into the ρ-eigenspace of A. Thus there must be a coordinate vector e_i such that $L(e_i)$ is a ρ-eigenvector (i.e., nonzero), and it is nonnegative, since A is nonnegative. $\square$

Recall that a directed graph is *strongly connected* if there is a directed path between every pair of vertices, or equivalently, there is a closed (directed) path that passes through every vertex. In particular, a graph with one vertex and no edges is strongly connected.

The *support graph* of a square matrix $A = [a_{ij}]$ is a directed graph in which there is an edge from i to j if $a_{ij} \neq 0$. We will say that A is *strongly connected* if this associated graph is strongly connected. It is not hard to show that this is equivalent to the non-existence of a simultaneous permutation of the rows and columns of A having the block triangular form $\left[\begin{smallmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{smallmatrix}\right]$, where A_{11} and A_{22} are square submatrices.

Note that the 1×1 matrix $[\,0\,]$ is strongly connected; it is the only strongly connected matrix with spectral radius 0.

THEOREM 4. *If A is nonnegative and strongly connected, then*

(a) *A has a positive eigenvector v with eigenvalue $\rho(A)$,*
(b) *the $\rho(A)$-eigenspace is one-dimensional,*
(c) *v is the unique nonnegative eigenvector of A (up to scalar multiples), and*
(d) *$\rho(A)$ is a simple root of the characteristic polynomial of A.*

Proof. We may replace A by $A + 1$ so that by Lemma 3, $\rho = \rho(A)$ is the only extremal eigenvalue of A, and $\rho > 0$.

(a) Let v be a nonnegative eigenvector of A with eigenvalue ρ, as provided by Lemma 3. If the first a coordinates of v are positive and the remaining b coordinates of v are zero, then the condition $Av = \rho v$ forces the southwest $b \times a$ submatrix of A to be 0, and thus A could not be strongly connected. Hence v must be positive.

(b) Now suppose that u is another eigenvector in the ρ-eigenspace. Replace u with $-u$ if necessary so that at least one coordinate of u is positive. It follows that if c is the largest scalar such that $v - cu$ is nonnegative, then $v - cu$ has at least one zero coordinate, and hence cannot be an eigenvector, by the argument in the previous paragraph. Since $v - cu$ belongs to the ρ-eigenspace, the only other possibility is that $v = cu$.

(c) Let w^T be a positive left eigenvector[1] for A with eigenvalue ρ. (Apply (a) to A^T.) If u is a nonnegative right eigenvector with eigenvalue λ, we can evaluate $w^T Av$ in two ways, obtaining $\rho w^T u = w^T Au = \lambda w^T u$. However, $w^T u$ is necessarily positive, so $\lambda = \rho$ and u is a multiple of v by (b).

(d) Let r be the multiplicity of ρ as a root of the characteristic polynomial of A. Since the ρ-eigenspace is one-dimensional, it must be the case that A has only one Jordan block associated to ρ, and it has order r. Thus by Lemma 1(b), some matrix entries of A^n grow at the asymptotic rate of $n^{r-1}\rho^n$, so the same must be true for the coordinates of $A^n u$ for any positive vector u. However, for the positive eigenvector v we have $A^n v = \rho^n v$, a contradiction unless $r = 1$. □

[1] It would also be sensible to call w^T a right eigenvector, since $w^T \mapsto w^T A$ is a right action for A. This is one of those situations where standard terminology is in conflict, or at least dyslexic.

The following result strengthens Theorem 2.

THEOREM 5. *If all sufficiently high powers of A are real and positive, then all of the conclusions of Theorem 4 hold, and we have*

$$A^n = \frac{1}{w^T v} \rho^n v w^T + o(\rho^n) \quad \text{as } n \to \infty,$$

where w^T and v denote left and right eigenvectors for A with eigenvalue $\rho = \rho(A)$.

Proof. Choose m so that A^m is positive. We know that ρ is the only extremal eigenvalue of A (Theorem 2), and any extremal eigenvalue/vector for A is also extremal for A^m, so Theorem 4 (applied to A^m) implies that ρ has multiplicity 1 as an eigenvalue of A, and the corresponding eigenvector is positive (if suitably normalized). Similarly, there is no other nonnegative eigenvector for A since the same is true for A^m.

To prove the asymptotic formula, note that both sides are invariant under changes of basis and choices of eigenvectors, so we may assume that A is in Jordan Canonical Form. Since ρ has multiplicity 1, there is a 1×1 Jordan block with eigenvalue ρ, and all other blocks have eigenvalues μ with $|\mu| < \rho$. Powers of these blocks therefore grow at rates asymptotically slower than ρ^n (Lemma 1), so if the blocks are ordered so that e_1^T and e_1 are left and right ρ-eigenvectors, then $A^n = \mathrm{diag}(\rho^n, 0, \ldots, 0) + o(\rho^n) = \rho^n e_1 e_1^T + o(\rho^n)$. □

If A is merely nonnegative and strongly connected, then it may happen that every power of A has entries that vanish, and there may be extremal eigenvalues in addition to $\rho(A)$. In fact, these two (related) misfeatures are the subject of the next section.

2. Periodicity

Define a directed graph to be *m-cyclic* if the vertices may be partitioned into disjoint (nonempty) blocks V_k ($0 \leqslant k < m$) so that every edge is directed from a vertex in V_k to a vertex in V_{k+1} (subscripts taken mod m). In the case $m = 2$, this is the same as the graph being bipartite, but being m-cyclic and m-partite are not the same for $m > 2$.

Similarly we will say that a (square) matrix A is *m-cyclic* if the support graph of A is m-cyclic. Illustrating this in the case $m = 3$, this amounts to saying that there is a suitable simultaneous permutation of the rows and columns of A that has the block form

$$\begin{bmatrix} 0 & A_{12} & 0 \\ 0 & 0 & A_{23} \\ A_{31} & 0 & 0 \end{bmatrix}.$$

Note that the diagonal blocks of zeroes must be square.

LEMMA 6. *If $\mu_1, \ldots, \mu_n$ are the nonzero eigenvalues (with multiplicity) of the product $A_{12}A_{23} \cdots A_{m1}$ of the blocks of an m-cyclic matrix A, then A has exactly mn nonzero eigenvalues (counted by multiplicity), and they are precisely the m-th roots of $\mu_1, \ldots, \mu_n$.*

Proof. Let $B = A_{12}A_{23}\cdots A_{m1}$. We are given that $\det(1-tB) = (1-\mu_1 t)\cdots(1-\mu_n t)$. Taking the determinant of the identity

$$\begin{bmatrix} 1 & -tA_{12} & 0 \\ 0 & 1 & -tA_{23} \\ -tA_{31} & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ t^2 A_{23}A_{31} & 1 & 0 \\ tA_{31} & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1-t^3A_{12}A_{23}A_{31} & -tA_{12} & 0 \\ 0 & 1 & -tA_{23} \\ 0 & 0 & 1 \end{bmatrix}$$

reveals that (in the case $m = 3$)

$$\det(1-tA) = \det(1-t^3B) = (1-\mu_1 t^3)\cdots(1-\mu_n t^3),$$

and the argument for general m is essentially the same. $\square$

By symmetry, the matrix factors appearing in $B = A_{12}A_{23}\cdots A_{m1}$ can be permuted cyclically without affecting the result, yielding the following amusement.

Corollary 7. *If $A_1, \ldots, A_m$ are rectangular matrices, then the nonzero eigenvalues of $A_1\cdots A_m$ (with multiplicity) are unchanged by cyclic permutations of the factors, given that all m permuted products are defined (and therefore are square).*

Define the *periodicity* of a square matrix A to be the greatest common divisor of the lengths of all closed directed paths in its support graph. Note that the periodicity of the 1×1 matrix $[\,0\,]$ is ∞; for all other strongly connected matrices it is an integer $p \geqslant 1$.

Theorem 8. *Let ω be a primitive p-th root of unity. If A is nonnegative and strongly connected with periodicity $p < \infty$, then*

(a) *A is p-cyclic,*
(b) *the eigenvalues of A (with multiplicity) are stable under multiplication by ω,*
(c) *the extremal eigenvalues of A are $\omega^k\rho$ $(0 \leqslant k < p)$, and*
(d) *each extremal eigenvalue is a simple root of the characteristic polynomial of A.*

Proof. (a) Suppose that there are directed paths of lengths k and l from vertex 1 to vertex i in the support graph of A. There must also be a directed path, say of length h, that returns from vertex i to vertex 1. Thus there are closed paths of lengths $k+h$ and $l+h$, so both lengths must be divisible by p and $k = l \bmod p$. In other words, all paths from vertex 1 to vertex i have the same length mod p.

We may therefore partition the vertices of the support graph into p blocks so that block k consists of all vertices reachable from vertex 1 by a directed path of length $k \bmod p$ $(0 \leqslant k < p)$. If there were an edge directed from block k to block l, then there would be a directed path of length $k+1 \bmod p$ from vertex 1 to a vertex in block l, so this could happen only if $l = k+1 \bmod p$. Thus A is p-cyclic.

Having proved that A is p-cyclic, (b) follows immediately from Lemma 6.

It also follows that A^p is block-diagonal, say $A^p = \operatorname{diag}(A_0, \ldots, A_{p-1})$.

Given a set of positive integers with greatest common divisor p, one knows by a theorem of Schur[2] that every sufficiently large multiple of p is a nonnegative integer combination of those integers. Thus by extending a single closed path that passes through every vertex, one can find such paths having a length equal to any sufficiently large multiple of p. Since there must be a directed path of length divisible by p between any two vertices in the same block, it follows that there must also be directed paths whose lengths are any sufficiently large multiple of p. That is, all sufficiently high powers of $A_0, \ldots, A_{p-1}$ are positive.

By Theorem 2, we may deduce that each of $A_0, \ldots, A_{p-1}$ has a unique extremal eigenvalue, and (Theorem 4) the extremal eigenvalue of block A_i has multiplicity 1. On the other hand, each extremal eigenvalue λ of A contributes an extremal eigenvalue λ^p to A^p, so by pigeon-holing, there can be at most p such eigenvalues (with multiplicity) since at most one occurs in each block A_i. By part (b) and a second application of Theorem 4, we know that A has at least p extremal eigenvalues; namely, $\omega^i \rho$ for $0 \leqslant i < p$. Therefore these must be all of the extremal eigenvalues (proving (c)), and they each must occur with multiplicity 1 (proving (d)). □

By decomposing the support graph into strongly connected components, any nonnegative matrix is permutation-equivalent to a block-triangular matrix whose diagonal blocks are strongly connected. Therefore,

Corollary 9. *Every extremal eigenvalue of a nonnegative real matrix is an m-th root of a real number for some m.*

In the proof of Theorem 8, we saw that if A has periodicity p, then the diagonal blocks of all sufficiently high powers of A^p are positive.

Corollary 10. *All sufficiently high powers of a nonnegative matrix A are positive if and only if A is strongly connected and has periodicity 1.*

A further consequence of Theorem 8 is that we can detect the periodicity of a nonnegative strongly connected matrix from its spectrum.

Corollary 11. *A nonnegative strongly connected matrix A has periodicity $p < \infty$ if and only if it has p distinct (nonzero) extremal eigenvalues.*

3. Monotonicity

Recall if A is not strongly connected, then A can be permuted into the block triangular form $\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$. In particular, the eigenvalues of A are those of A_{11} and A_{22}, so we may perturb the entries of A_{12} arbitrarily without affecting the spectral radius of A.

[2] OK, I lied when I said that the only non-trivial thing we would take for granted would be Jordan Canonical Form. Still, you could argue that Schur's theorem is actually pretty well-known and easy. In any case, there is a nice elementary proof that can be found in Section 3.15 of [**W**].

On the other hand, the following result shows that if we increase or decrease any single entry of a nonnegative strongly connected matrix (but keep the matrix nonnegative), then the spectral radius necessarily changes, and in the same direction as the perturbation.

THEOREM 12. *If* $|B| \leqslant A$ *entry-wise, then* $\rho(B) \leqslant \rho(A)$. *Moreover, if equality occurs and* A *is strongly connected, then* $|B| = A$.

Proof. We have $|B^n| \leqslant |B|^n \leqslant A^n$, so the entries of B^n are asymptotically dominated by those of A^n. In particular, if $\rho(A) = 0$, then A is nilpotent, and therefore B must be as well. Otherwise, we may assume $\rho = \rho(A) > 0$, in which case the entries of B^n are $O(n^{r-1}\rho^n)$ for some $r \geqslant 1$ by Lemma 1(a). If λ is an extremal eigenvalue of B and v is an associated eigenvector, then some coordinates of $B^n v = \lambda^n v$ must grow at the asymptotic rate of $\rho(B)^n$, so this is possible only if $\rho(B) \leqslant \rho$.

In the case of equality, we may assume B is nonnegative (i.e., $B = |B|$). Given that A is strongly connected, Theorem 4 implies that A has a positive left eigenvector w^T with eigenvalue ρ, and Lemma 3 implies that B has a nonnegative right eigenvector v with eigenvalue ρ. It follows that

$$w^T(A - B)v = (w^T A)v - w^T(Av) = \rho w^T v - \rho w^T v = 0.$$

However, $A \geqslant B$ and w^T is positive. So if v is positive, this forces $A = B$ and we are done.

Permuting coordinates if necessary, the remaining possibility is that the first a coordinates of v are positive, and the remaining b coordinates are 0. In that case, as noted previously in the proof of Theorem 4, the condition $Bv = \rho v$ forces the southwest $b \times a$ submatrix of B to vanish. Furthermore, the corresponding submatrix of A cannot be zero, otherwise A would not be strongly connected. Therefore, one or more of the first a columns of $w^T(A - B)$ is positive, and hence $w^T(A - B)v > 0$, a contradiction. □

References

[**W**] H. S. Wilf, *Generatingfunctionology* (2nd ed.), Academic Press., New York, 1994. URL: ⟨`http://www.math.upenn.edu/~wilf/DownldGF.html`⟩.