THE PERRON-FROBENIUS THEOREM

VAIBHAV MISHRA

Abstract. This paper will explore the Perron-Frobenius Theorem and its applications to Markov Chains and stationary distributions. It starts with some background and then proves the theorem. The paper concludes with some other applications of the theorem in other fields.

1. Introduction and History

The Perron-Frobenius Theorem is a very important theorem in linear algebra and helps show that irreducible aperiodic Markov chains have unique stationary distributions among many other applications. It was proved by Oskar Perron in 1908 for positive matrices and Georg Frobenius generalized the theorem for nonegatve matrices in 1912.

2. Background

Definition 2.1 (Positive Matrix)**.** A positive matrix is a matrix whose entries are all positive real numbers.

Definition 2.2 (Spectral Radius)**.** The spectral radius, r of a $n \times n$ square matrix, A, is the maximum of the absolute values of the eigenvalues (λ) of the matrix ($|\lambda| \leq r$).

The concept of the spectral radius is central to our proof of the Perron-Frobenius Theorem as the Perron-Frobenius eigenvalue of a matrix is also the spectral radius of the matrix.

Definition 2.3 (Primitive Matrix)**.** A primitive matrix is a $n \times n$ square nonegative matrix ($A = (a_{ij})$) which to some power is positive. It is said to be primitive if there exists k such that $A^k \geq 0$.

Primitive matrices are important for our proof of the Perron-Frobenius Theorem for nonnegative matrices.

Definition 2.4 (Irreducible Markov Chain)**.** A Markov chain on a state space Ω is irreducible, if for any states $i, j \in \Omega$ there is some positive integer t such that $p^{(t)_{ij}>0}$. An irreducible Markov Chain is where every state can be reached from every other state.

3. Proof of the Perron-Frobenius Theorem

Theorem 3.1. *Let A be an $n \times n$ irreducible square matrix.*

- *There is a positive number r, called the* Perron–Frobenius eigenvalue *of A, such that r is an eigenvalue of A, and is the spectral radius of A.*
- *The eigenspace corresponding to r is 1-dimensional.*
- *There exists an eigenvector v, called the* Perron–Frobenius eigenvector*, corresponding to r such that all the entries of v are positive.*
- *The only nonnegative eigenvectors of A are multiples of v.*

Lemma 3.2. *Let A be a positive $n \times n$ matrix. Then A has a positive eigenvalue, r with a positive corresponding eigenvector, v.*

Proof. Let A be positive matrix in $M_n(\mathbb{R})$. Let S denote the set of vectors x in $\mathbb{R}^n$ that have all entries non-negative and satisfy $\|x\| = 1$. If $x \in S$, then all entries of Ax are positive. Let us define a function $L : S \to \mathbb{R}_{>0}$ as follows. For $x \in S$,

$$L(\mathbf{x}) = \min_{1 \leq i \leq n} \left\{ \frac{(Ax)_i}{x_i} : x_i \neq 0 \right\}.$$

To better understand what the function L does, we can start with $x \in S$. Compare the vectors x and Ax entry by entry. Look at these positions i in which x has a positive entry x_i. For each of these i, the ith entry of Ax is also positive, so it is x_i multiplied by some positive scaling factor α_i. The least of these α_i is what we are calling $L(x)$. It is a positive real number.

That is how $L(x)$ is defined for a particular $x \in S$, and L is a continuous function from S to the set of positive real numbers. Since S is compact, this means that L has a maximum value on S. Call this r, and let $v \in S$ be a vector for which $L(v) = r$.

We will show that r is an eigenvalue of A with v being the corresponding eigenvector, and that v is positive. There are two steps.

1. We show that $Av = rv$. We know that $Av \geq rv$ since $L(v) = r$,this means that $(Av)_i \geq rv_i$ for all i. Thus $Av - rv \geq 0$. this means that $A(Av - rv)$ is a positive vector and so we can choose $\epsilon > 0$ small enough that $A(Av - rv) > \epsilon Av$. The vector Av may not belongs to S, but there is a positive real number c for which $cAv \in S$.

$$A(Av) > (r + \epsilon)Av \implies A(cAv) \geq (r + \epsilon)cAv \implies L(cAv) \geq (r + \epsilon).$$

This contradict the choice of r as the maximum value of L on S, and we conclude that $Av = rv$.

2. We also know that $v \geq 0$ since $v \in S$. It follows that $Av > 0$ - no entry of Av can be equal to zero since v is a non-negative non-zero vector and A is positive. Hence rv and v are both strictly positive. ∎

Lemma 3.3. *The spectral radius of A is r.*

Proof. If we have λ be any eigenvalue of A, and if we have y be the corresponsing eigenvector with $\|y\| = 1$. For every entry, i , of Ay and λ, we have the following:

$$\begin{aligned} y &= Ay \\ \implies \lambda y_i &\leq \sum_{j=1}^{n} A_{ij} y_j \\ \implies |\lambda y_i| &\leq \sum_{j=1}^{n} A_{ij} y_j \\ \implies |\lambda||y_i| &\leq \sum_{j=1}^{n} A_{ij} |y_j|. \end{aligned}$$

∎

Lemma 3.4. *r has geometric multiplicity 1.*

Proof. We know that v is a positive eigenvector of A corresponding to r. Now suppose that u is an eigenvector of A corresponding to r, and that u is independent of v over $\mathbb{C}$.
We assume that the u as entries that are real as r is real. If u has entries that are non-real complex numbers, then the real and imaginary part of u would separately be eigenvectors of A and at least one of them would be independent of v.
As per our assumption, ever element of the 2-dimensional space spanned by u and v (over $\mathbb{C}$ or $\mathbb{R}$) is an eigenvector of A corresponding to r. Since $v > 0$, there is a real number ϵ with the property that $u' = v + \epsilon u$ is a non-negative vector with at least one entry equal to zero. However $u' \neq 0$ since u and v are independent.
This is the required contradiction, since Au' would be positive in this case and would not be a scalar multiple of u'. ∎

Lemma 3.5. *The algebraic multiplicity of r is 1.*

Proof. We must show that A is similar to a (real) matrix A' that has the entry r in the $(1, 1)$ position and zeros throughout the rest of Row 1 and Column 1.

Since A and its transpose have the same characteristic polynomial and hence the same spectrum, the spectral radius of A^T is r. Now, there is a positive column vector w that is an eigenvector of A^T corresponding to r (by the Lemma 3.2). Thus $A^T w = rw$ and the row vector w^T satisfies

$$w^T A = rw^T.$$

Now let U be the $(n - 1)$- dimensional orthogonal complement of w with respect to the ordinary scalar product on $\mathbb{R}^n$:

$$U = \{u \in \mathbb{R}^n : w^T u = 0\}.$$

Let $u \in U$, and consider the vector $Au \in \mathbb{R}^n$. Note that

$$w^T Au = rw^T u = 0,$$

so $Au \in U$ whenever $u \in U$. this means that the subspace U of $\mathbb{R}^n$ is A-invariant. This is because U is the orthogonal complement in $\mathbb{R}^n$ of a left eigenvector of A, it has nothing to do with the positivity of A or the special properties of r and w. However these special properties give us an important extra piece of information.

As v is the corresponding eigenvector of r. Then $v \notin U$ since $w \cdot v = w^T v > 0$, because w and v are both positive. Let $\{b_1, ..., b_{n-1}\}$ be a basis of U. Then $B = \{v, b_1, ..., b_{n-1}\}$ is a basis of $\mathbb{R}^n$.

Now the matrix A' that describes the linear transformation of $\mathbb{R}^n$ determined by left multiplication by A, with respect to the basis B, has the following form:

$$A' = r0...00..B_{(n-1)\times(n-1)}.0$$

Here B is $n \times n$ matrix with real entries. Since A and A' are similar , r occurs as an eigenvalue of both, with the same algebraic multiplicity and with geometric multiplicity 1 in each case. The characteristics polynomial of A' is $(x - r)r_B(x)$, where $r_B(x)$ is the characteristic polynomial of B. If the algebraic multiplicity of r as an eigenvalue of A' exceeds 1, then r is an eigenvalue of B with a corresponding eigenvector $v_B \in \mathbb{R}^{n-1}$. This means that the vector in $\mathbb{R}^n$ obtained by preceding v with a zero entry is an eigenvector of A' corresponding to r. Since e_1 is also an eigenvector of A' corresponding to r, this means that r has geometric multiplicity at least 2 as an eigenvector of A', and hence also as an eigenvector of A. The contradiction leads us to conclude that r occurs once as a root of the characteristic polynomial of A.

■

Lemma 3.6. *Let u be a positive eigenvector of A. then u is a real positive scalar multiple of v.*

Proof. Let μ be the eigenvalue of A to which u corresponds. Then, μ is real and $\mu > 0$, since A and u are positive and $Au = \mu u$. Thus $0 < \mu \leq r$. Choose ϵ small enough that $u' = v - \epsilon u$ is positive. For which i we have

$$(Au')_i = rv_i - \mu\epsilon u_i \geq r(v_i - \epsilon u_i) = ru'_i.$$

Thus $Au' \geq ru'$, which means that $Au' = ru'$ by the maximality of r as a value of the function L. This means that u' is a r-eigenvector of A, which means that u', hence u, is a scalar multiple of v and $\mu = r$.

■

Lemma 3.7. *Suppose that μ is an eigenvalue of A, $\mu \neq r$. Then $|\mu| < p$.*

Proof. Suppose, anticipating contradiction, that $|\mu| = r$, and let y be an eigenvector of A corresponding to μ, with $||y|| = 1$. Let $|y|$ denote the vector in $\mathbb{C}^n$ whose entries are the moduli of the entries of y. Then $|y| \in S$ and for each i we have

$$(A|y|)_i = \sum_j A_{ij}|y_j| = \sum |A_{ij}y_j| \geq |\sum_j A_{ij}y_j| = |\mu y_i| = r|y_i|.$$

Thus $A|y| \geq r|y|$, which means that $|y|$ is a r-eigenvector of A and $|y| = v$. Then equality holds in the triangle inequality above and we have for each i that

$$\sum_j |A_{ij}y_j| = |\sum_j A_{ij}y_j|.$$

So $A_{i1}y_1, A_{i2}y_2, ..., A_{in}y_n$ are complex numbers with the property that the sum of their moduli is the modulus of their sum. This means that they lie on the same ray in the complex plane (a ray is a half-line with its endpoint at 0). Since the numbers A_{ij} are all real and positive, this means that $y_1, ..., y_n$ all lie on the same ray. Hence there is some θ for which $e^{i\theta}y$ is a positive vector. Thus y is a (complex) scalar multiple of a positive vector, and since r is the only eigenvalue of A to have a positive corresponding eigenvector, it follows that $\mu = r$. Thus the only eigenvalue of A to have modulus r is r itself, and every other eigenvalue has modulus strictly less than the spectral radius.

■

4. Application of Perron-Frobenius Theorem to Markov Chains

One of the most useful applications of Perron-Frobenius theorem is its use for Markov chains.

If we have a Markov Chain $X_0, X_1, ...,$ with states in $1, ..., n$, with transition matrix P and let p_t be the distribution of X_t, then $p_{t+1} = Pp_t = P^t P_0$. We know that $\lambda^T P =^T$, where λ is a left eigenvector with eigenvalue 1 (Perron-Frobenius eigenvalue of P). For a irreducible and aperiodic Markov chain, the right Perron-Frobenius eigenvector is the stationary distribution, $P\pi = \pi$. There exists an integer k such that P^k has strictly positive enries iff the Markov chain is aperiodic and irreducible. The stationary distribution of the Markov chain is the unique Perron-Frobenius eigenvector of P^k.

Theorem 4.1. *If we have an irreducible and aperiodic Markov Chain, then it has a unique stationary distribution, π.*

Proof. See [6] for a complete and rigorous proof. ■

Lemma 4.2. *Let A be an $n \times n$ matrix, where A is diagonizable iff A has n linearly independent eigenvectors.*

Proof. Let P be a $n \times n$ matrix with xolumns $z_1, ..., z_n$ and let D be a $n \times n$ diagonal matrix with diagonal entries $\lambda_1, ...,_n$. Then,

$$AP = A[z_1 z_2 ... z_n] = [Az_1 Az_2 ... Az_n],$$

and

$$PD = P = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} = [\lambda_1 z_1 \lambda_2 z_2 ... \lambda_n z_n].$$

Now if we assume that A is diagonizable and $A = PDP^{-1}$, then $AP = PD$, which gives,

$$[Az_1 Az_2 ... Az_n] = [\lambda_1 z_1 \lambda_2 z_2 ... \lambda_n z_n].$$

or

$$Az_1 = \lambda_1 z_1, Az_2 = \lambda_2 z_2, ..., Az_n = \lambda_n z_n.$$

Since P is invertible its columns have to be linearly independent, and also be nonzero. The above equation shows that $(_i, z_i)$ are eigenpairs for $i = 1, ..., n$. Therefore a diagonizable matrix has n linearly independent eigenvectors, where the columns of P are the eigenvectors and the diagonal of D are the eigenvalues.

Now if we have A have n linearly independent eigenvectors $z_i, ...z_n$ with eigenvalues $\lambda_i, ..., \lambda_n$, and we also have a matrix $P = [z_i z_2 ... z_n]$ and a diagonal matrix $D = diag(\lambda_1, ..., \lambda_n)$. Then from above we have $AP = PD$, which can still be trues without having linearly independent eigenvectors. Since the eigenvectors are linearly independent P is invertible and so $A = PDP^{-1}$. ■

Lemma 4.3. *If A has n linearly independent eigenvectors, then so does A^T.*

Proof. Since A hs n linearly independent eigenvectors, A may be diagonalized as $A = PDP^{-1}$, where the columns of P are the eigenvectors that are linearly independent of A and the eigenvalues of A are the diagonal entries of D. Then, we also have $A^T = (P^{-1})^T DP^T$. By the previous lemma, we have the columns of $(P^{-1})^T$ be the n linearly independent eigenvectors of A^T. ∎

5. Google PageRank Example

Non-negative matrixes and graphs are very connected to each other. For example, if we have a non-negative $n \times n$ matrix, B, and the graph $\Gamma(B)$ that is associated with this matrix have the vertices $v_1, ..., v_n$. The graph has an arc from v_i to v_j, iff the entry A_{ij} of A is positive .

Definition 5.1. If we let Γ be a directed graph, with vertices u and v, then a walk from u to v is a sequence of arcs, where the terminal vertex of each arc is the initial vertex of the next one. The length of a walk is the number of arcs that it contains. The zero-nonzero pattern of B is said to be symmetric when B_{ji} is positive whenever B_{ij} is positive. In a symmetric pattern, the graph association with A may be undirected as it has an the arc from v_i to v_j iff it has an arc from v_j to v_i. The adjacency matrix, $A(\Gamma)$ of a directed graph with Γ with vertices $v_1, ..., v_n$ is the matrix whose (i, j) entry is 1 if there is an arc from $v_i to v_j$ in Γ, and 0 otherwise.

Lemma 5.2. *If we let Γ be a directed graph with adjacency matrix A, and let k be a positive integer, then the entry in the (i, j) position of A^k is the number of walks from $v_i to v_j$ in Γ.*

Proof. If $k = 1$, then the (i, j) entry is 1 or 0 as there either is an arc from v_i to v_j or there is not. Now if this lemma is try for A^{k-1}:

$$(A^k)_{ij} = \sum_{m=1}^{n} (A_{im}^{k-1} A_{mj}.$$

Then, A_{im}^{k-1} is the number of walks of length $k - 1$ from v_i to v_m. Therefore, $A_{im}^{k-1} A^{mj}$ is the number of walks of length k in Γ from v_i to v_j that have v_m as their second last vertex. The sum of these numbers over m is the total number of walks of length k from v_i to $v - j$. ∎

If we have B be any non-negative matrix and let A be the (0,1)-matrix with the same zero-nonzero pattern as B, then $A = A(\Gamma(B))$. For every positive integer k, B^k and A^k have the same zero-nonzero pattern. The entry in the (i,j) entry of B^k is positive iff there is a walk of length k from v_i to v_j in $\Gamma(B)$. The matrix B is primitive iff B^k is positive form k, in other words there must be a walk of length k in $\Gamma(B)$ from every vertex u to every vertex v.

The Perron-Frobenius Theorem uses this definition and interpretation of matrices in terms of graphs in the application of the PageRank algorithm which Google uses to assign rankings for webpage searches.We assign vertices to webpages where there is an arc from vertex v_i to vertex v_j if there is a link from Page i to Page j and let n be the number of webpages involved. Now, let A be the transpose of the adjacency matrix of this graph, so that $A_{ij} = 1$ if there is a link from Page j to Page i, 0 otherwise. Also, suppose that the number of links from Page j is d_j. Now if we assign the probability of a surfer on Page j leaving it through a randomly chosen link as 0.85, and the probability that a surfer will move to a randomly chosen page as 0.15. Here the probability that a surfer at Page j will go to Page i is given by: $\frac{0.85}{d_j} + \frac{0.15}{n}$, if there is a link from Page j to Page i and $\frac{0.15}{n}$, if there is not.

Theorem 5.3. *If we let A be a positive $n \times n$ matrix where the entries in every column sum to the same number k, then k is the Perron eigenvalue of A.*

Proof. Let v be the vector that has length of n and whose entries are all 1. Then, $v^T A = kv^T$. Thus v^T is a left eigenvector of A and since v is a positive vector, it follows that k is the Perron eigenvalue of A by the Perron-Frobenius Theorem. ∎

Theorem 5.4. *Let A be a positive $n \times n$ matrix with spectral radius 1 and let v be any positive vector is $\mathbb{R}^n$. Then the sequence $v, Av, A^2v, ...$ converges to the Perron eigenvector of A or to the zero vector.*

Proof. Let $1, \lambda_2, ..., \lambda_n$ be the eigenvalues of, so $|\lambda_i| < 1$ by the Perron-Frobenius Theorem. Assume that A is diagonalizable and let $v_1, ..., v_n$ be a basis of $\mathbb{R}^n$ consisting of eigenvectors of A, where $Av_1 = v_1$ and $Av_i = \lambda_i v_i$. Now, $v = a_1v_1 + a_2v_2 + ... + a_nv_n$, and for a positive integer k,

$$A^k v = a_1 A^K v_1 + a_2 A^K v_2 + ... + a_n A^K v_n = a_1 v_1 + a_2 A^K v_2 + ... + a_n A^K v_n.$$

Since $|\lambda_i| < 1, \lambda_i^k \to 0$ as $k \to \infty$.

∎

For the PageRank matrix, if we let x be the vector whose jth entry x_j is the proportion of surfers who are on Page j. The ith entry of Bx is

$$(Ax)_i = \sum_{j=1}^{n} B_{ij} x_j = \sum_{j=1}^{n} P(j \to i) x_j,$$

Thus $\sum_{j=1}^{n} P(j \to i)x_j$ is the proportion of the whole population that will be at Page i one step after the step whose population distribution is described by x, so the vector Bx converges to y for which $\sum y_i = 1$. Thus y describes the steady state of the system. The pages are ranked in terms of importance according to the entries of this Perron eigenvector.

6. Other Applications of the theorem

The Perron-Frobenius Theorem has many other applications in mathematics like on compact operators as well as non-negative matrices. It also has applications in other fields like modeling population growth, modeling price changes in economics, and power control.

7. References

[1] http://www.maths.nuigalway.ie/ rquinlan/linearalgebra/lecture2526.pdf
[2] https://personalpages.manchester.ac.uk/staff/stefan.guettel/ma/protected/nonneg.pdf
[3] https://www.webpages.uidaho.edu/ stevel/565/lectures/5a
[4] https://www.cc.gatech.edu/ rpeng/CS7540$_S$17/$Feb2PerronFrob.pdf$
[5] http://www.maths.nuigalway.ie/ rquinlan/linearalgebra/lecture3233.pdf
[6] http://faculty.washington.edu/harin/markovchain$_proofs.pdf$