THE PERRON-FROBENIUS THEOREM IN RELATION TO POSITIVE, STOCHASTIC MATRICES

[redacted]

1. Introduction

A primary interest when studying Markov chains is the existence of a stationary distribution, and whether this state is convergent from any beginning state. The primary theorem applied to this question is the Perron-Frobenius theorem, which makes significantly stronger statements regarding non-negative square matrices. This paper will seek to prove the limited case of the Perron-Frobenius theorem when applied to the transition matrix of a finite-state Markov chain. Although the general theorem can be proven via a variety of approaches, this limited proof will seek to primarily use elementary linear algebra results. This paper will subsequently discuss approaches to extend this result as well as current applications of this result in the use of finite-state Markov chains.

2. Definitions and prerequisite results

Definition 2.1. A square matrix P is (column) stochastic if it satisfies

$$p_{ij} \geq 0 \quad \forall i, j, \qquad \sum_i p_{ij} = 1 \quad \forall j$$

Definition 2.2. A matrix (or vector) A is positive, and denoted $A > 0$ if it satisfies

$$a_{ij} > 0$$

Theorem 2.3. *Let A be a square matrix, with $R_j = \sum_{j \neq i} |a_{ij}|$, the sum of absolute values of the off-diagonal entries of a column. Every eigenvalue lies in one of the disks on the complex plane centered at a_{jj} with radius R_j.*

This theorem is known as the Gershgorin Circle Theorem.

Theorem 2.4. *Let $\rho(A)$ be the spectral radius of A. $\forall v, \quad \|Av\| \leq \rho(A)v$.*

This lemma follows directly from the construction of the spectral radius as a limit of vector norms.

3. Proof of Theorem

In this section we present the proof of the Perron-Frobenius theorem in relation to positive, stochastic matrices in order to make the following statement regarding the convergence of finite-state Markov chains.

Theorem 3.1. *Let P be a positive, stochastic matrix. $\exists!\pi$ positive, stochastic such that $\|P^n x_0 - \pi\| \leq \alpha^n$, $\alpha < 1$ for any starting vector x_0 stochastic.*

This statement is directly equivalent to the existence of a unique stationary distribution for any finite-state Markov chain, provided that the transition probability between any two states is strictly positive. In addition, the Markov chain will converge to this steady state regardless of its initial state geometrically. The subsequent lemmas will develop the structure of the eigenvalues of a positive, stochastic matrix. The theorem proof itself will be focused on demonstrating its convergent behavior.

Lemma 3.2. *A positive stochastic matrix P has an eigenvalue $\lambda = 1$.*

Proof. Consider the characteristic polynomials of P and P^T. Since the eigenvalues of P are the solution of the polynomial $det(A - \lambda I) = det(A - \lambda I)^T = det(A^T - \lambda I)$, it follows P and P^T have the same eigenvalues. Since the columns of P sum to 1, the rows of P^T similarly sum to 1. By stochasticity $P^T 1 = 1$, where 1 is the column vector of ones. So $\lambda = 1$ is an eigenvalue of P^T and subsequently P. □

Lemma 3.3. *A positive stochastic matrix P has a spectral radius $\rho(P) = 1$, and $\forall \lambda_i \neq 1$, $|\lambda_i| < 1$, where λ_i are the eigenvalues of the matrix.*

Proof. The proof of this lemma follows geometrically from 2.3. Since every diagonal entry of a positive stochastic matrix lies in the interval $(0, 1)$, every Gershgorin disk is centered on the real axis on the interval $(0, 1)$. For each disk centered at p_{jj}, stochasticity ensures it has radius $1 - p_{jj}$. The combination of these conditions require each Gershgorin disk to be internally tangent to the unit disk at 1 on the real axis. As such, the union of the Gershgorin disks lie completely within the unit disk, so $\rho(P) \leq 1$. Since $\lambda = 1$ is an eigenvalue of P, it follows $\rho(P) = 1$. Additionally since every Gershgorin disk is internally tangent at one, and has a strictly smaller radius than one, the union of the disks does not intersect with the unit disk at any other point, implying that $\lambda = 1$ is the only eigenvalue with $|\lambda| = 1$. □

Lemma 3.4. *$\nexists v$ non-negative such that $Pv > v$ if P is a positive stochastic matrix.*

Proof. Suppose for contradiction such a v exists. If v is non-negative, Pv is positive. However, if $Pv > v$ entry-wise, $\|Pv\| > \|v\|$. Since $\rho(P) = 1$ by Lemma 3.3, this contradicts Theorem 2.4. □

Lemma 3.5. *A positive stochastic matrix P has an eigenvector $z > 0$ corresponding to $\lambda = 1$.*

Proof. By Lemma 3.2, P has eigenvalue $\lambda = 1$. Let x be a corresponding eigenvector, so it follows $Px = x, x \neq 0$. Consider $z = P|x|$. Since $|x|$ has all non-negative elements, and at least one non-zero element, and P has all positive elements, it follows $z > 0$. $z = P|x| \geq |Px| = |x|$ by triangle inequality. So, $y = z - |x| \geq 0$. Suppose $y > 0$. Positivity implies

$$0 < Py = Pz - P|x| = Pz - z \rightarrow Pz > z$$

This contradicts 3.4 so $y = 0$. So, $z = |x|$, which implies $Pz = z$, and by construction $z > 0$. □

Lemma 3.6. *The geometric multiplicity of eigenvalue $\lambda = 1$ of a stochastic positive matrix P is* 1.

Proof. By Lemma 3.5, P has eigenvalue pair $(x, 1)$ where $x > 0$. Suppose $Py = y$, with x, y linearly independent. Find ϵ such that $z = x + \epsilon y \geq 0$, but z is not strictly greater than zero, which can clearly be done since x is strictly positive, and y is linearly independent (they span a two dimensional space). $Pz > 0$, since z is non-negative, and P is strictly positive (as reasoned before). However, $z = Pz$, so $z > 0$, which contradicts the assumption that z is not strictly greater than zero. So the assumption x, y linearly independent is false, and $\lambda = 1$ has geometric multiplicity one. □

Lemma 3.7. *If P is a positive stochastic matrix, there exists a unique π such that $P\pi = \pi$, and π is stochastic*

Proof. By Lemma 3.5 $\exists x$ such that $Px = x$, with x positive. By Lemma 3.6 x is unique up to scaling, so $\pi = \frac{x}{\sum_i x_i}$ is stochastic, and is the only stochastic vector which satisfies $P\pi = \pi$. □

Lemma 3.8. *The algebraic multiplicity of eigenvalue $\lambda = 1$ of a stochastic positive matrix P is* 1.

Proof. Suppose $\lambda = 1$ is not a simple eigenvalue. The Jordan representation of P implies $(A - I)^k y = 0, \quad (A - I)^{k-1} y \neq 0$. So, $(A - I)^{k-1} y$ is an eigenvector corresponding to $\lambda = 1$, and by Lemma 3.6, y can be chosen so $x = (A - I)^{k-1} y$, with $x > 0$. Let

$$z = (A - I)^{k-2} y$$

$$(A - I)z = (A - I)^{k-1} y = x \rightarrow Az - z = x$$

Since x is strictly positive, $Az > z$, and by triangle inequality $A|z| > |z|$. However, this contradicts 3.4, so the eigenvalue is simple. □

A collection of these results, forming a restricted version of the Perron-Frobenius theorem is as follows.

Theorem 3.9. *Let P be a positive, stochastic matrix. The following holds.*

(1) *P has eigenvalue $1 = \rho(P)$. (Lemma 3.2)*
(2) *For any eigenvalue $\lambda \neq 1$ of P $|\lambda| < 1$. (Lemma 3.3).*
(3) *The corresponding eigenspace of $\lambda = 1$ has multiplicity one, it is simple. (Lemma 3.8).*
(4) *There exists a unique stochastic vector π corresponding to $\lambda = 1$, satisfying $P\pi = \pi$. (Lemma 3.7)*

Based on these results, the statement regarding convergence, Theorem 3.1 will be proved below.

Proof. Let x_0 be a stochastic vector. Let $v_1, ...v_n$ be an ordered normed (by $L1$ norm) basis satisfying the Jordan canonical form, $P = TJT^{-1}$. Let $x_0 = a_1v_1 + ...a_nv_n$. Since π has geometric multiplicity one, $v_1 = \pi$.

$$P^k x_0 = (TJT^{-1})^k x_0 = TJ^kT^{-1}x_0 = TJ^k(a_1e_1 + ...a_ne_n) = TJ^ka_1e_1 + TJ^k(a_2e_2 + ...a_ne_n)$$

Since the eigenvalue corresponding to the first Jordan block is 1 by theorem, and the diagonals of the Jordan blocks after the first are less than one by theorem $J_i^k \to 0$, with a convergence of λ_i^k, leaving the second term to vanish. So $P^k x_0 \to a_1v_1$, and since a stochastic matrix times a stochastic vector is stochastic, $P^k x_0 \to v_1 = \pi$ with a geometric convergence rate with ratio λ_i^k, where λ_i is the second largest eigenvalue. □

4. Discussion

The results obtained in Theorem 3.9 can be generally applied to any positive matrix, regardless of its spectral radius. The majority of the proof proceeds similarly by considering the matrix normed by its spectral radius. The theorem is slightly weaker for non-negative matrices, and the convergence result is conditional on the periodicity of the matrix. Convergence results for Markov chains in particular can also be obtained through various more modern probabilistic methods, most notably through work by Diaconis, and tighter bounds on the convergence ratio can be obtained. However, when examining this particular treatment of the Perron-Frobenius theorem, the primary necessary building blocks of the theorem are classical results from linear algebra and the spectral behavior of matrices.

5. Application

One of the most notable modern examples of the application of the convergent behavior of positive, stochastic matrices is the Google algorithm, PageRank. Specifically, the underlying ranking of pages in a Google search is determined by their relative importance. This importance is determined by the steady-state of the Markov chain that represents a hypothetical user randomly selecting a (certain type) of link on each web page they visit. This leads to the fundamental assumption of the PageRank algorithm: more important sites will have greater amounts of web traffic flowing to them. This series of assumptions converts the question of ranking websites to the determination of a stationary distribution of a (particularly large) stochastic matrix. In order to leverage the results previously described, PageRank makes two important modifications to the transition matrix. The first is assuming that if a page has no outgoing links, it corresponds to a uniform column vector, a user will randomly select any web page when on a web page with no current links. The second modification is to ensure the positivity required of the transition matrix, by a slight modification, $P' = cP + (1 - c)E$, where E is a stochastic matrix of uniform columns, each positive. Suitable choice of E fulfills the requirement of the theorem. The choice of the column vector introduces a preference or bias controllable by Google, however, a uniform vector can be used to remain completely neutral. In addition, the constant, known as the damping factor, can be controlled to preference computational stability and performance of the algorithm. In every case, the eventual ranking factor is determined by the stationary distribution of P', which as it is guaranteed positive, converges relatively quickly by the result obtained.

References

[1] Richard Bellman, *Introduction to Matrix Analysis, 2nd ed*, SIAM, 1997.
[2] T. Haveliwala, S. Kamvar, The Second Eigenvalue of the Google Matrix, *Stanford*.
[3] R. Horn, C. Johnson, *Matrix Analysis, 2nd ed*, Cambridge, 1985.
[4] C.R. MacCluer, The Many Proofs and Applications of Perron's Theorem, *SIAM Review*, 42, 487-498 (2000).