HONORS THESIS: AN EXPLORATION IN PERRON'S THEOREM

N. BURKE

ABSTRACT. In this paper, we will explain Perron's Theorem and Perron-Frobenius theory from linear algebra. We considered the historical perspective when completing research and have included our findings accordingly. Applications and examples will also be investigated, including applications to Markov chains, Leslie's population model, and Google's Page Rank algorithm.

CONTENTS

Date: December 18, 2020.

1. Introduction

1.1. **A Personal Reflection.** My interest in mathematics began during 5th grade. My teacher showed the class some upper level concepts, like bases, and I was captivated. For the rest of the year, I devoted my recess time to nailing the math concepts. After elementary school, my interest deepened. I had an enrichment tutor in high school who introduced me to problems like the four color theorem which she correctly guessed I would enjoy. Later, I learned that she was teaching me combinatorics. In my freshman year of college, I was part of my first group project with other women in the math department. Since then, many of them have changed majors. I went through a period of questioning the path I chose, as well. However, I realized I love what I am learning and I want to continue my pursuit of mathematics. Last year, I completed a research project for the University Honors Program where I studied unfoldings of polyhedra. It was the best experience I've had in college and it made me realize I was interested in applied mathematics.

Now that I am a university student, small talk always includes the questions "What are you studying?" and "What will you do with your degree?" After I answer that I am studying math, what follows are the questions "Are you going to be a teacher? Or a statistician?" It is important to be able to explain what I can actually do as an applied mathematician. In order to illustrate to the general population what a mathematician can do, as well as deepen my knowledge in an interesting subject, I am presenting a study of a specific theorem in applied mathematics: Perron's theorem.

1.2. **Background on Perron's Theorem.**

I will be explaining select versions of Perron's theorem in order of their creation. Different versions of the theorems will be used for applications as they are appropriate.

Perron's theorem was first stated and proved in 1907. Perron's theorem allows us to determine quite a few conclusions about positive matrices. It tells us that positive square matrices behave how one might intuitively expect in that they have at least one positive eigenvalue and positive eigenvector. Perron-Frobenius theory is an extension of this theorem on nonnegative matrices [4].

1.3. **Related Work.**

See Baldin [3] for a relevant algebra background. To learn linear algebra from the basics, we recommend *Elementary Linear Algebra* by Howard Anton [1]. We refer the reader to "Why didn't Charasoff and Remak use Perron-Frobenius mathematics?" by Parys Wilfried [15] for an in-depth look at the historical perspective during Perron's and Frobenius' work in the 1900s. Hawkins [7] explains what led to the first version of Perron's theorem. For more applications and proofs of Perron's theorem, we recommend MacCluer's work [11]. Meyer [14], Sternberg [17], Baldin [3], and Carnia [5] give information on how Perron-Frobenius theory relates to graph theory.

We refer the reader to Appendix 8 for background mathematics definitions and concepts, which we have put in an appendix in order to not interrupt from our historical narrative.

2. Perron

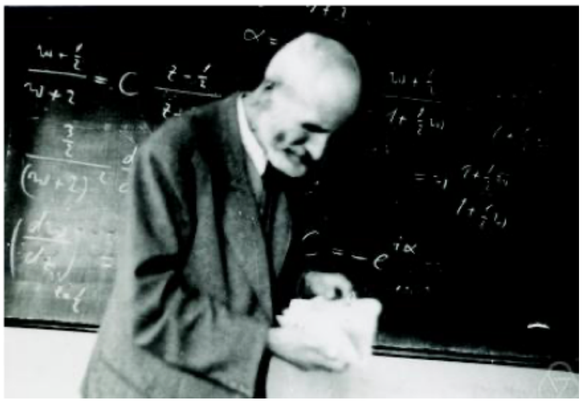

Figure 1. Oskar Perron at Munich in 1948 from Pierce [16]

Perron's theorem came about from his *Habilitationsschrift*, a rework and an extension of Stolz's theorem, which Oskar Perron published in 1907 [7]. The topic of the paper was the convergence criteria of partial sums which corresponded to the coefficients of Jacobi's algorithm. Perron first introduced a 2×2 matrix to the problem and considered the characteristic roots. To continue towards his conclusions on convergence, Perron used limits of ratios of the elements in the matrix to find properties of the eigenvalue ρ. It followed that, in some cases, an eigenvalue ρ_0 must be the largest absolute value root or have the largest multiplicity. Then, he found that the necessary case relates to strict nonnegative requirements. After considering this problem, Perron wrote a lemma and a corollary that was very close to what we call Perron's theorem. He then published a new paper, *Towards the Theory of Matrices*, which included Perron's theorem and his proof [7].

See the appendix for a definition of characteristic root and eigenvalue 8.5 and multiplicity 8.9.

Theorem 2.1. Perron's theorem: Let A be any square matrix such that $A > 0$. Then A has a real-valued characteristic root $\rho_0 > 0$ of multiplicity one such that $\rho_0 > |\rho|$ for all other characteristic roots ρ of A. Moreover, a positive eigenvector associated with ρ_0 exists [7].

See the appendix for a definition of eigenvector 8.6.

Perron's first proof was "not explained well" according to Hawkins. In his second paper, *Zur Theorie der Matrices*, Perron made several improvements [7]. One such improvement was his limit lemma.

Lemma 2.2. Perron's limit lemma: Let $A = (a_{ij})$ be an $n \times n$ matrix with $a_{ij} > 0$ for all i, j. If the entries of the matrix power A^m is denoted by $(a_{ij}^{(m)})$, then;

(1) $\lim_{m\to\infty} \frac{a_{ij}^{(m)}}{a_{nj}^{(m)}}$ exists as a finite number that is independent of j. The number is denoted by $\frac{x_i}{x_n}$.

(2) $\rho' = \lim_{m\to\infty} \frac{a_{ij}^{(m+1)}}{a_{ij}^{(m)}}$ exists as a finite positive number that is independent of i and j.

(3) If $\mathbf{x} = (x_1, \ldots, x_n)^T$ with the x_i as in (1), then $A\mathbf{x} = \rho'\mathbf{x}$.
Hence, ρ' is a positive characteristic root of A [7].

Proof. To prove theorem 2.1, Perron used induction on the size of $A_{ij} > 0$ with entries $a_{ij} > 0$ for all $i, j = 0, 1, \ldots, n$. We know the claim is true for a 1×1 matrix. So, we assume the claim to be true for all $A = (a_{ij}) > 0 \quad n \times n$ matrices with $i, j = 0, 1, \ldots, n$. Then, let $\phi_{n+1}(\rho) = \det(\rho I - A)$. Note, we define $B_{n\times n}$ to be the submatrix of A obtained by deleting the first row and column. Then $B = (a_{ij}) > 0$ with $i, j = 1, 2, \ldots, n$ and we let $\phi_n(\rho) = \det(\rho I - B)$. Then, by the induction hypothesis, ϕ_n has a maximal positive root σ_0. Also, $\phi_n(\rho)$ has at least one term ρ^n so $\lim_{\rho\to\infty} \phi_n(\rho) = \infty$. Thus $\phi(\rho) > 0$ for all $\rho > \sigma$. We think Perron used Laplace expansions of determinants by cofactors next [7].

See the appendix for a definition of det 8.4.

The Laplace expansion by cofactors of the 0-th row is

$$\phi_{n+1}(\rho) = (\rho - a_{00})\phi_n(\rho) - \sum_{i=1}^{n} a_{0i}[\,\mathrm{Adj}(\rho I - A)]_{\,i0} \tag{1}$$

Note, the $(0, i)$ cofactor is the $(i, 0)$ cofactor. Then

See the appendix for a definition of Adj 8.1 and cofactor 8.3.

$$[\,\mathrm{Adj}(\rho I - A)]_{\,i0} = \sum_{i=1}^{n} a_{j0}[\,\mathrm{Adj}(\rho I - B)]_{\,ij} \tag{2}$$

By substituting equation 2 into equation 1,

$$\phi_{n+1}(\rho) = (\rho - a_{00})\phi_n(\rho) - (a_{01} \ldots a_{0n})\mathrm{Adj}(\rho I - B)(a_{10} \ldots a_{n0})^T \tag{3}$$

Note, $(a_{01} \ldots a_{0n})$ and $(a_{10} \ldots a_{n0})^T$ are strictly positive. Rewriting equation 3 gives

$$\frac{\phi_{n+1}(\rho)}{\phi_n(\rho)} = \rho - a_{00} - (a_{01} \ldots a_{0n})\mathbf{d}(\rho), \quad \mathbf{d}(\rho) = \frac{\mathrm{Adj}(\rho I - B)(a_{10} \ldots a_{n0})^T}{\phi_n(\rho)} \tag{4}$$

Then Perron makes use of lemma 2.2 to show that $\phi_{n+1} = 0$ for at least one $\rho \in (\sigma_0, \infty)$. Thus A has a maximal positive root $\rho_0 > \sigma_0$ [7, 11].

To show that ρ_0 has multiplicity one, we consider

$$\frac{d}{d\rho}\left[\frac{\phi_{n+1}(\rho)}{\phi_n(\rho)}\right]_{\rho_0} = \frac{\phi'_{n+1}(\rho)}{\phi_n(\rho)}$$

If this is true, then $\phi'_{n+1}(\rho) > 0$ since we already know $\phi_n(\rho) > 0$ which would make ρ_0 simple. Taking the derivative of equation 4 gives

See the appendix for a definition of simple eigenvalue 8.15.

$$\frac{d}{d\rho}\left[\frac{\phi_{n+1}(\rho)}{\phi_n(\rho)}\right]_{\rho_0} = 1 - (a_{01} \ldots a_{0n})\mathbf{d}'(\rho).$$

Then to show that $\mathbf{d}'(\rho) < 0$, Perron noted that $\mathbf{d}(\rho)$ has an expansion of the form $\mathbf{d}(\rho) = \frac{1}{\rho}((a_{10} \ldots a_{n0})^T) + (\frac{1}{\rho})^k$ such that $k > 1$. The derivative of this expression is $\mathbf{d}'(\rho) = \frac{-1}{\rho^2}((a_{10} \ldots a_{n0})^T) + \frac{-k}{\rho^{k+1}}$. So we do indeed have $\mathbf{d}'(\rho) < 0$ for a sufficiently large ρ. In fact, $\mathbf{d}'(\rho_0) < 0$. Therefore ρ_0 has multiplicity one [7].

Next, Perron shows the positivity of $\mathrm{Adj}(\rho_0 I - B)$. The positivity of the cofactors of $\rho_0 I - A$ means that $\mathrm{Adj}(\rho_0 I - B) > 0$ and so its columns are positive characteristic vectors for ρ_0. In other words, the eigenvector $\mathbf{x}$

associated with ρ_0 is positive. Note, $\rho_0 > \sigma_0$ implies that $\mathbf{d}(\rho_0) > 0$ and $\phi_n(\rho) > 0$ so $\mathrm{Adj}(\rho I - A)(a_{10} \dots a_{n0})^T > 0$. Then the cofactor expansion from equation 2 shows that $[\,\mathrm{Adj}(\rho_0 I - A)]_{i0} = [\,\mathrm{Adj}(\rho_0 I - B)(a_{10} \dots a_{n0})^T]_i$ for $i = 1, \dots, n$ and also $[\,\mathrm{Adj}(\rho_0 I - A)]_{00} = \phi_n(\rho_0) > 0$. So, the cofactors of $\rho_0 I - A$ are all positive and the eigenvector $\mathbf{x}$ associated with ρ_0 is positive [7, 11]. □

Example 2.3. Say matrix M is

$$M = \begin{bmatrix} 1 & 2 \\ 8 & 1 \end{bmatrix}.$$

Note,

$$M^2 = \begin{bmatrix} 17 & 4 \\ 16 & 17 \end{bmatrix}, M^3 = \begin{bmatrix} 49 & 38 \\ 152 & 49 \end{bmatrix}, M^4 = \begin{bmatrix} 353 & 136 \\ 544 & 353 \end{bmatrix}, \text{ and } M^5 = \begin{bmatrix} 1441 & 842 \\ 3368 & 1441 \end{bmatrix}.$$

Then, we consider $\frac{1}{5^n} M^n$. We compute

$$\frac{1}{5^3} M^3 = \begin{bmatrix} 0.392 & 0.304 \\ 1.216 & 0.392 \end{bmatrix}, \frac{1}{5^4} M^4 \approx \begin{bmatrix} 0.565 & 0.218 \\ 0.870 & 0.565 \end{bmatrix}, \text{ and } \frac{1}{5^5} M^5 \approx \begin{bmatrix} 0.461 & 0.269 \\ 1.078 & 0.461 \end{bmatrix}.$$

And,

$$\frac{1}{5^{20}} M^{20} \approx \begin{bmatrix} 0.500 & 0.250 \\ 1.000 & 0.500 \end{bmatrix}.$$

One might guess that

$$\lim_{n \to \infty} \frac{1}{5^n} M^n = \begin{bmatrix} 0.5 & 0.25 \\ 1 & 0.5 \end{bmatrix}.$$

In fact, $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is a right eigenvector corresponding to the eigenvalue equal to 5 and also $\begin{bmatrix} 2 & 1 \end{bmatrix}$ is a left eigenvector of the same eigenvalue. So, it makes sense that 5 is the exponential growth rate. If we pick $R = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $L = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \end{bmatrix}$, then

$$RL = \begin{bmatrix} 0.5 & 0.25 \\ 1 & 0.5 \end{bmatrix};$$

see [4]. As you can see, the matrix M is a positive $n \times n$ matrix, and so are its powers. If we consider the characteristic equation $\lambda^2 - 2\lambda - 15 = (\lambda - 5)(\lambda + 3) = 0$, we can see that the eigenvalues for M are 5 and -3, each with multiplicity one. As guaranteed by Perron's theorem, the eigenvalue 5 is positive, has multiplicity one, and is the maximal eigenvalue.

This work clearly has some interesting results, and yet it was still lacking in that it required strict positivity and some techniques outside of linear algebra. Perron stated "Although this is a purely algebraic theorem, nevertheless I have not succeeded in proving it with the customary tools of algebra. The theorem remains valid, by the way, when the a_{ik} are only partly positive but the rest are zero provided only that a certain power of the matrix A exists for which none of the entries are zero" [7].

3. Frobenius

Frobenius was an expert on matrix theory, so he likely came to know of Perron's theorem when Perron published his paper *Towards the Theory of Matrices* [7]. Frobenius published a proof for Perron's theorem that avoided the limit lemma in 1908. He also considered what could be said if the matrix was nonnegative, instead of strictly positive. In Frobenius' next paper, published in 1909, he gave an even stronger proof of the theorem which used inner product notation. With this technique, he proved other propositions of a similar vein. In 1912, Frobenius published another paper which started with the exploration of the possible characteristic roots for a nonnegative matrix where nonnegative eigenvectors exist. This led to the separation between primitive and non-primitive irreducible matrices in his theorem conditions. Frobenius included one application in his paper: determinants. However, Perron-Frobenius theory was not yet seen as clearly applicable [7].

See the appendix for a definition of primitive matrices 8.12.

Theorem 3.1 ([7, 8])**.** The Perron-Frobenius theorem: Let A be an $n \times n$ matrix with positive real entries. Then the following are true.

(1) A has a positive characteristic root and hence a maximal positive root ρ_0. Furthermore, ρ_0 has multiplicity one, and $\mathrm{Adj}(\rho_0 I - A) > 0$ for all $\rho_0 \geq \rho$.
(2) If ρ is any other characteristic root of A, then $|\rho| < \rho_0$.

Proof. The first part is proved again using induction, with the improvement that $\mathrm{Adj}(\rho_0 I - A) > 0$ for all $\rho_0 \geq \rho$ is in the hypothesis. Then, cofactor expansions quickly show that most of statement (1) is correct. Unlike Perron, Frobenius used a different approach to prove that ρ_0 has multiplicity one. Let $\phi(\rho_0) = \det(\rho_0 I - A)$ denote the characteristic polynomial of A. Then, the sum of the cofactors is equal to

See the appendix for a definition of characteristic polynomial 8.2

$$\Phi(\rho_0) = \sum_{\alpha=1}^{n} (-1)^{\alpha+\alpha} \phi_{\alpha\alpha}(\rho_0), \tag{5}$$

where $\phi_{\alpha\alpha}(\rho_0)$ denotes the α-th principal minor determinant of $(\rho_0 I - A)$. Equation 5 shows that $\phi_{\alpha\alpha}(\rho_0) = [\,\mathrm{Adj}(\rho_0 I - A)]_{\alpha\alpha} > 0$ because $\alpha + \alpha$ must be even and so the cofactors are even and thus $\Phi > 0$. Since $\phi_{\alpha\alpha}(\rho_0)$ is the (α, α) entry of $\mathrm{Adj}(\rho_0 I - A)$ and all other entries are also positive, we have $\mathrm{Adj}(\rho_0 I - A) > 0$. Equivalently, ρ_0 has multiplicity 1 [8, 11]. □

See the appendix for a definition of minor 8.8.

The proof of part (2) was much briefer than Perron's in his 1908 paper, furthermore Frobenius was able to write a much improved version in 1909. Frobenius had by then discovered Brioschi's work and inner product considerations, leading him to a proof using inner product notation [8].

Example 3.2. We will consider an example of The Perron-Frobenius Theorem 3.1 first on a 2 by 2 matrix. Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

where $a, b, c, d > 0$. The characteristic polynomial is

$$\phi(\lambda) = \det(\lambda I - A) = \lambda^2 - (a+d)\lambda + (ad - bc).$$

The discriminant is $(a-d)^2 + 4bc > 0$ so the polynomial has two distinct real roots;

$$\lambda_0(A) = \frac{(a+d) + \sqrt{(a-d)^2 + 4bc}}{2}, \ \lambda_1(A) = \frac{(a+d) - \sqrt{(a-d)^2 + 4bc}}{2}.$$

Since $a, b, c, d > 0$, the determinant can only be positive. So, there exists a real eigenvalue for A. Also as $a, b, c, d > 0, \lambda_0(A) > 0$, and so A has a positive real eigenvalue. Note,

$$\frac{(a+d) + \sqrt{(a-d)^2 + 4bc}}{2} > \left| \frac{(a+d) - \sqrt{(a-d)^2 + 4bc}}{2} \right|.$$

Then, $\lambda_0(A) > |\lambda_1(A)|$. Hence, $\lambda_0(A)$ is a simple positive eigenvalue and also $\mathrm{Adj}(\lambda_0 I - A) > 0$. As you can see, the theorem holds true for a positive 2×2 matrix [10].

In 1912, Frobenius' published further theorems.

Theorem 3.3. Irreducible matrix theorem: If $A \neq 0$ is an irreducible matrix, then ρ_0 is simple and positive and $\mathrm{Adj}(\rho_0 I - A) > 0$. Hence there is a $\mathbf{v} > 0$ such that $A\mathbf{v} = \rho_0 \mathbf{v}$. All other characteristic roots ρ satisfy $|\rho| \leq \rho_0$ [7].

Theorem 3.4. Primitive matrix theorem: An irreducible matrix A is primitive if and only if $A^v > 0$ for some power v [7].

Theorem 3.5 ([7])**.** Imprimitive matrix theorem: Let $A \neq 0$ be an $n \times n$ imprimitive matrix, and let k denote the number of characteristic roots of A with absolute value equal to ρ_0. Then,

(1) $A^k \sim_\sigma \begin{pmatrix} R_{11} & 0 & \dots & 0 \\ 0 & R_{22} & \dots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \dots & R_{kk} \end{pmatrix}$, meaning A is permutationally similar to the matrix and where each square block R_{ii} is primitive;
(2) If the characteristic polynomial of A is expressed in the notation $\phi(\rho) = \det(\rho I - A) = \rho^n + a_1\rho^{n_1} + a_2\rho^{n_2} + \dots, \quad a_i \neq 0$, then k is the greatest common divisor of the differences $n - n_1, n - n_2, \dots$;
(3) If ρ is any characteristic root of A, then so is $\epsilon\rho$ where $\epsilon = e^{2\pi i/k}$;
(4) The k roots with absolute value ρ_0 are given by $\rho_i = \epsilon^i \rho_0$, $i = 0, \dots, k-1$, and they are all simple;
(5) If the characteristic polynomial of A is expressed in the notation $\phi(\rho) = \rho^n + b_1\rho^{n-k} + b_2\rho^{n-2k} + \dots + b_m\rho^{n-mk}$, where $b_m \neq 0$ but $b_i = 0$, for some $i < m$, is possible, and if $\psi(\rho) = \rho^m + b_1\rho^{m-1} + b_2\rho^{m-2} + \dots + b_m$, then $\psi(\rho)$ has a simple positive root that is larger than the absolute value of any other root.

4. Markov

Markov published his work on the theory of probability at the same time as Perron published his work on matrices. His paper was translated into German by the time Frobenius' had published his last work, although Frobenius was apparently unaware of Markov's work. The two of them independently came to many of the same conclusions. Clearly, Markov chains are one of the earliest applications of Perron-Frobenius theory. Even so, it took until the 1930s for interest in Markov chains to become widespread. Consequently, Markov chains were developed thoroughly and in more generality, including non-negative stochastic matrices. However, Markov chains and Perron-Frobenius theory still were not linked. Richard von Mises made a close connection in 1931 in his book *The Calculus of Probabilities and its Application to Statistics and Theoretical Physics* [7]. He used Frobenius' conclusions in his application and was aware of the connection to Markov Chains. It was V.I. Romanovsky who explicitly cited Frobenius in relation to Markov chains in his "Investigations on Markoff chains" in 1936 [7].

Definition 4.1. A Markov chain is a system that transitions from one state to another based solely on the current state [13, 1].

Example 4.2. Consider the vector $\mathbf{v} = (x, 1-x)$ multiplied by powers of the transition matrix

$$M = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}.$$

Then define $\mathbf{v}_1 = M\mathbf{v}, \mathbf{v}_2 = M^2\mathbf{v}, \ldots, \mathbf{v}_n = M^n\mathbf{v}$. As you can see, this chain only depends on the previous step since we can write $\mathbf{v}_1 = M\mathbf{v}, \mathbf{v}_2 = M\mathbf{v}_1, \mathbf{v}_3 = M\mathbf{v}_2, \ldots, \mathbf{v}_n = M\mathbf{v}_{n-1}$. In this case, $\mathbf{v}_\infty = (0.6, 0.4)^T$. Note,

$$M\mathbf{v}_\infty = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix} \begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix} = \mathbf{v}_\infty.$$

We will explain this phenomenon later. Note, M is a strictly positive 2×2 matrix. When we calculate eigenvectors and eigenvalues, we see that the eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = 1/2$, each with multiplicity one. So, $\lambda_1 > |\lambda_2|$. Their eigenvectors are $(0.6, 0.4)^T$ for λ_1 which is positive as we expect, and $(-1, 1)^T$ for λ_2. Also, $\mathbf{v}_\infty$ is the eigenvector with eigenvalue 1. This sort of vector convergence only happens with special matrices, not all Markov chains [10].

This leads us to the following definition.

Definition 4.3. A stochastic matrix is a square matrix which describes the transitions of a Markov chain. In the matrix, each entry represents a probability, so all entries satisfy $0 \leq P_{ij} \leq 1$ where the probability of going from i to j is P_{ij}, and where $\sum_{j=1} P_{i,j} = 1$ [18, 1].

Theorem 4.4. Markov's theorem: A stochastic $n \times n$ matrix P which has a positive matrix power P^m for some m has a final state matrix

$$P^\infty = \lim_{k\to\infty} P^k = \begin{pmatrix} p_1 & p_2 & \dots & p_n \\ p_1 & p_2 & \dots & p_n \\ \vdots & & & \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

where the row vector $\mathbf{p}$ is the unique positive vector with $p_1+p_2+\ldots+p_n = 1$ that satisfies $\mathbf{p}P = \mathbf{p}$ [12, 17, 1].

In fact, Markov wrote all of his results without the use of matrix notation. He used the language of determinants and systems of linear equations [7]. We will be using matrix notation and Perron-Frobenius theory, which Markov was lacking.

Proof. First we require that a stochastic matrix $M = (p_{ij})$ is positive for some power which lets us apply the Perron-Frobenius theorem 3.1. This gives us an eigenvalue λ which is positive, non-repeated, and has one positive eigenvector $\mathbf{v}$ associated. Since M's row sums equal 1, then we have the column vector $\mathbf{1}$ which is full of 1s and is an eigenvector. By definition, $\mathbf{v}P = \mathbf{v}\lambda = (\mathbf{1})P = (\mathbf{1})1 = \mathbf{1}$. So, we have $\lambda = 1$ for some eigenvector $\mathbf{v} = (\mathbf{1})$ [17, 12].

Let's assume we have an eigenvalue $\lambda_i \neq 1$. So, $P\mathbf{x} = \lambda_i\mathbf{x}$ and $\mathbf{x} \neq \mathbf{0}$. Also, as $P\mathbf{v} = \mathbf{v}$ we know $\mathbf{x} \neq a\mathbf{v}$ where a is some constant. Let $\mathbf{x} = (x_i, \ldots, x_n)^t$ and set $m = \max_i |x_i|$. We then assume that for at least one value of i, we'll call it i_0, it holds that $|x_{i_0}| < m$. If it was the case that all $|x_i| = m$, then all entries would be equal as stochastic matrices are nonnegative. Then, we could normalize the vector to be $\mathbf{v}$ which would then imply that $\lambda_i = \lambda = 1$. So, $|\lambda_i|m < 1m$ and thus $|\lambda_i| < 1$; see [7].

Then, M has a unique positive left eigenvector $\mathbf{p}$ which can be normalized to give $\mathbf{p} * P = \mathbf{p}$ [17, 12].

Next, we consider the convergence of P^k. By theorem 3.1, we can say that all other eigenvalues λ_i are such that $|\lambda_i| < \lambda = 1$. So, P^k converges to P^∞ based on Jordan canonical form. Note, $P^\infty = \lim_{k\to\infty} P^{k+1} = (\lim_{k\to\infty} P^k) * P = P^\infty * P$. Thus, each row of P^∞ is $\mathbf{p}$. [12] □

See the appendix for a definition of Jordan form 8.7.

Markov chains in combination with Perron-Frobenius theory have many applications. Markov's theorem 4.4 can be applied to several economic models. It can also be used in the case of an unfair coin [11]. Biophysicists use the theorem for modeling the decay of "the fraction of mutant proteins that fold stably to their native structure as a function of the number of amino acid substitutions" as stated by [10]. Markov's theorem can be employed for almost any repeated experiment where one outcome depends on the previous outcome [4].

We will now explain several applications in detail.

Example 4.5. Andrey Markov wanted to study the proportions of vowels and consonants in *Eugene Onegin* [4]. In 1913, he published his results. Markov defined four variables: the proportion of vowels that followed a vowel, consonants that followed a vowel, vowels that followed a consonant, and consonants that followed a consonant. We will represent his results in the stochastic matrix

$$P = \begin{array}{c} \\ \text{Vow} \\ \text{Con} \end{array} \begin{array}{c} \begin{array}{cc} \text{Vow} & \text{Con} \end{array} \\ \begin{bmatrix} 0.128 & 0.872 \\ 0.663 & 0.337 \end{bmatrix} \end{array}.$$

Looking at the powers of P, we see

$$P^2 \approx \begin{array}{c} \\ \text{Vow} \\ \text{Con} \end{array} \begin{array}{c} \begin{array}{cc} \text{Vow} & \text{Con} \end{array} \\ \begin{bmatrix} 0.595 & 0.405 \\ 0.308 & 0.692 \end{bmatrix} \end{array}, P^3 \approx \begin{array}{c} \\ \text{Vow} \\ \text{Con} \end{array} \begin{array}{c} \begin{array}{cc} \text{Vow} & \text{Con} \end{array} \\ \begin{bmatrix} 0.345 & 0.655 \\ 0.498 & 0.502 \end{bmatrix} \end{array}, \dots,$$

$$P^{11} \approx \begin{array}{c} \\ \text{Vow} \\ \text{Con} \end{array} \begin{array}{c} \begin{array}{cc} \text{Vow} & \text{Con} \end{array} \\ \begin{bmatrix} 0.431 & 0.569 \\ 0.432 & 0.568 \end{bmatrix} \end{array}, P^{12} \approx \begin{array}{c} \\ \text{Vow} \\ \text{Con} \end{array} \begin{array}{c} \begin{array}{cc} \text{Vow} & \text{Con} \end{array} \\ \begin{bmatrix} 0.432 & 0.568 \\ 0.432 & 0.568 \end{bmatrix} \end{array}.$$

In fact,

$$\lim_{k \to \infty} P^k = \begin{array}{c} \\ \text{Vow} \\ \text{Con} \end{array} \begin{array}{c} \begin{array}{cc} \text{Vow} & \text{Con} \end{array} \\ \begin{bmatrix} 0.432 & 0.568 \\ 0.432 & 0.568 \end{bmatrix} \end{array}.$$

As you can see, the row vectors are the same and they both still add to 1. This is in keeping with the calculation for the total sample proportion of vowels at 0.432 and the proportion of consonants at 0.568 [4].

Example 4.6. Take for example the the population dynamics of Colorado's dragons in the mountains. Let's say the ratio of the dragons in the mountains compared to dragons outside of the mountains starts at $\frac{30}{120} = 25\%$, leaving 75% not in the mountains. We will call this $\mathbf{v}_0 = (0.25, 0.75)^T$. Every year, 80% of the dragons in the mountains move away and 12% of dragons move into the mountains. Thus, we have the stochastic matrix

$$Q = \begin{bmatrix} 0.80 & 0.12 \\ 0.20 & 0.88 \end{bmatrix}.$$

So we can multiply Q by our starting vector to get

$$\mathbf{v}_1 = Q\mathbf{v}_0 = Q \begin{bmatrix} 0.25 \\ 0.75 \end{bmatrix} = \begin{bmatrix} 0.29 \\ 0.71 \end{bmatrix}$$

which represents the fact that, after one year, the ratio of dragons in the mountains increases to 29% while the population outside of the mountains is 71%. Note, the columns of Q add to 1 and Q is positive. This does not change for Q^k. We find that the eigenvalues of Q are $\lambda_1 = 1$ and $\lambda_2 = 0.68$ with eigenvectors $\mathbf{u}_1 = (0.375, 0.625)^T$ and $\mathbf{u}_2 = (-1, 1)^T$ respectively. Based on the Markov chain,

$$\mathbf{v}_0 = \begin{bmatrix} 0.25 \\ 0.75 \end{bmatrix} = \begin{bmatrix} 0.375 \\ 0.625 \end{bmatrix} + 0.125 \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Then,

$$\mathbf{v}_1 = 1 \begin{bmatrix} 0.375 \\ 0.625 \end{bmatrix} + 0.68(0.125) \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

and

$$\mathbf{v}_k = Q^k\mathbf{v} = \begin{bmatrix} 0.375 \\ 0.625 \end{bmatrix} + (0.68)^k(0.125) \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

The eigenvector $\mathbf{u}_1$ with $\lambda_1 = 1$ is the steady state at $\mathbf{v}_\infty$. The effect of the other eigenvector tends towards zero as k increases since $\lambda_2 < 1$. Thus, the dragon population in Colorado will eventually be 45/120 in the mountains and 75/120 outside of the mountains [10]. So, Q^T fits the conditions for theorem 4.4 and indeed satisfies its conclusions.

Example 4.7. The Ehrenfest Experiment

This thought experiment was proposed by P. and T. Ehrenfest to model quantum systems.

Consider a box with n molecules trapped inside where n is even. The box is divided (by an imaginary wall) into two congruent halves we will call 1 and 2. We consider the state of the system $S = s_1, \ldots, s_n$ to be the number of molecules in half 1 which will be between 0 and n. Suppose that initially there are some molecules in each half and that at time k, half 1 has i molecules. We assign all molecules in the box with an integer $z = \{1, \ldots, n\}$.

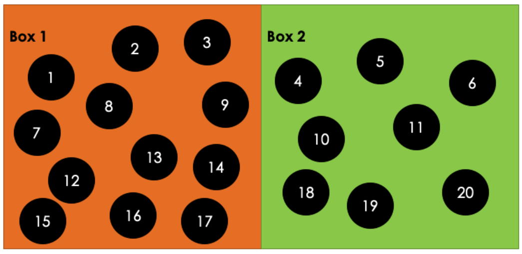

FIGURE 2. Image demonstrating the Ehrenfest Experiment with 20 molecules

Then, we pick a random integer from z and move its associated molecule to the other half of the box. So, we have a sequence of independent events where x_k is the number of molecules in half 1 after the kth trial. We want to know the probability $P(x_k = n+i \mid x_{k-1} = n+m; k)$, that is, the probability that there are $n+i$ molecules in half 1 after k trials given that there were $n+m$ in half 1 in the previous trial. Then, we have the conditional probabilities $P(x_{k+1} = i+1 \mid x_k = i) = 1 - i/n$ and $P(x_{k+1} = i-1 \mid x_k = i) = i/n$. Let transition matrix Q be $Q = (p_{ij})$, where $i, j \in \{1, \ldots, n-1\}$. So, each time we choose a molecule, the transition probabilities are

$$p_{0,1} = 1,\ p_{i,i-1} = \frac{i}{n},\ p_{i,i+1} = 1 - \frac{i}{n},\ p_{n,n-1} = 1$$

and all other entries are 0. Note, $Q^{n+2} > 0$, so the Markov chain must converge to the final state $Q^\infty = \mathbf{v}p$ where $p = (p_1, p_2, \ldots, p_n)$ such that $p_i = (1 - \frac{i-1}{n})p_{i-1} + \frac{i+1}{n}p_{i+1}$ for $1 \leq i \leq n-1$ and $p_0 = \frac{p_1}{n}$, $p_n = \frac{p_{n-1}}{n}$ by theorem 4.4. The recursion has the unique solution

$$p_i = \binom{n}{i} 2^{-n} = \binom{n}{i} (\frac{1}{2})^i (1 - \frac{1}{2})^{n-i}.$$

The solution is the binomial distribution with mean $\mu = \frac{n}{2}$. Thus, as a molecule is picked many times, on average, half of the molecules will be in half 1, which is why it is nice when n is even. This experiment gives a

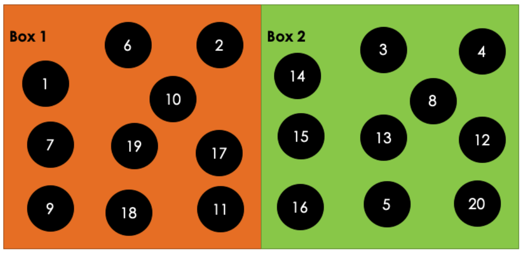

FIGURE 3. Image demonstrating the result of the Ehrenfest Experiment with 20 molecules

more complete model of heat exchange than what classical physics describes [9, 12, 11].

5. Post - Markov

We now survey advances in Perron-Frobenius theory that were developed after Markov's time.

5.1. Leslie.

In 1945, Patrick Leslie introduced a model for growth of a stratified population, a population that can be divided into subgroups. Leslie defined variables b_i, the expected number of daughters produced by a female in age group i, and s_i, the proportion of females in the ith age group who survive to the next age group. We are interested only in the case where all the $s_i > 0$ and $b_n > 0$ [17, 14].

Definition 5.1. Leslie Model: $\mathbf{f}(t+1) = L\mathbf{f}(t)$ where $\mathbf{f}(t)$ and $\mathbf{f}(t+1)$ are population vectors and L is the Leslie Matrix.

Leslie Matrix:

$$L = \begin{pmatrix} b_1 & b_2 & \dots & & b_n \\ s_1 & 0 & \dots & & 0 \\ 0 & s_2 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & 0 & s_{n-1} & 0 \end{pmatrix}$$

The element b_i denotes the birth rate, that is, the number of off-spring produced by an individual of age i at time t. The element s_i denotes the survival rate, that is, the probability that an individual in age class i will survive to age class $i+1$. So, $0 < s_i \leq 1$ and $b_i \geq 0$. All other entries are zeros. Each component is the number of individuals of a particular age class i such that $\mathbf{f}(t) = (f_1(t), f_2(t), \dots, f_n(t))^T$, where $f_i(t)$ is the number of individuals at time t of age class i [3, 5].

Example 5.2. The Rabbit Problem

We will consider a hypothetical rabbit population Let a_t represent the number of adult pairs of rabbits at the end of month t and let y_t be the number of youth pairs of rabbits at the end of month t. Also, let c_t represent the total count of pairs of rabbits at the end of month t. Each youth pair takes 2 months to mature to adulthood. In this model, both adults and youth give birth to a pair at the end of every month, but after the adult pair reproduces, the pair dies. Let the initial $a_0 = 0$ and $y_0 = 1$ as we start with one pair of youth rabbits. Then we have the following:

(1) At the end of month 0, $a_t = 0$ and $y_t = 1$ and $c_t = 1$.
(2) At the end of month 1, $a_t = 1$ and $y_t = 1$ and $c_t = 2$.
(3) At the end of month 2, $a_t = 1$ and $y_t = 2$ and $c_t = 3$.
(4) At the end of month 3, $a_t = 2$ and $y_t = 3$ and $c_t = 5$.
(5) At the end of month 4, $a_t = 3$ and $y_t = 5$ and $c_t = 8$.

And so on.

Note, the total number of rabbit pairs at the end of month n is equal to the sum of the number of pairs at the end of the two previous months $c_t = c_{t-1} + c_{t-2}$. You may recognise this as the Fibonacci sequence. Then we have

$$\begin{bmatrix} y_{t+1} \\ a_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_t \\ a_t \end{bmatrix} \text{ where } \begin{bmatrix} y_0 \\ a_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Which can be written as

$$\mathbf{f}(t+1) = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{f}(t),$$

where $\mathbf{f}(0) = (1, 0)^T$ [3]. Let

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

Although the Leslie model was created in 1945, the Fibonacci sequence was recognised much, much earlier. So we were able to identify this particular Leslie model as something that likely predates all others. In fact, the eigenvalues of L are the golden ratio $\frac{1 \pm \sqrt{5}}{2}$ [17].

Note, in this model there are some unrealistic traits which we want to account for. These include, but are not limited to; all pairs surviving to adulthood, each pair giving birth to exactly one pair, and adult rabbits and youth having the same reproductive rate. So, we generalize the model using survival rates and fertility rates. Let $b_1, b_2 \geq 0$ where these numbers replace the number of offspring with the fertility rate b_i. So, b_1 is the fertility rate of the youth age class and b_2 is the fertility rate of the adult age class. Then, replace the 1 with the survival rate s where $0 < s \leq 1$. Thus,

$$\mathbf{f}(t+1) = \begin{bmatrix} b_1 & b_2 \\ s & 0 \end{bmatrix} \mathbf{f}(t).$$

Here we have the Leslie model for two age classes [3].

5.1.1. *The New Perron-Frobenius Theorem.* In 1950 the Collatz-Wielandt formula was added to the Perron Frobenius theorem, among others [14].

Theorem 5.3 ([17, 3, 14, 5])**.** Perron-Frobenius theorem: Let $A \geq 0$ be an irreducible $n \times n$ matrix.

(1) $r \in \rho(A)$ is the Perron root or Perron eigenvalue with $r > 0$ and if λ is any other eigenvalue, then $r \geq |\lambda|$.
(2) If A is primitive, all other eigenvalues of A satisfy $r > |\lambda|$.
(3) r has algebraic multiplicity and geometric multiplicity of 1.

(4) The Perron eigenvector $\mathbf{v}$ is the unique vector defined by $A\mathbf{v} = r\mathbf{v}$, $\mathbf{v} > 0$, and $\|\mathbf{v}\|_1 = 1$. Any nonnegative eigenvector of A is a multiple of $\mathbf{v}$.
(5) If $0 \leq B \leq A$ and $B \neq A$, then every eigenvector σ of B satisfies $|\sigma| < r$.
(6) The Collatz-Wielandt Formula: $r = \max_{\mathbf{x} \in \mathcal{N}} f(\mathbf{x})$, where

$$f(\mathbf{x}) = \min_{1 \leq i \leq n, x_i \neq 0} \frac{[A\mathbf{x}]_i}{x_i}$$

and $\mathcal{N} = \{\mathbf{x} \mid \mathbf{x} \geq 0 \text{ with } \mathbf{x} \neq 0\}$

Proof.

(1) We know this is true from theorem 3.3. Now let's prove it. Let $A_k = A + \left(\frac{1}{k}\right) E$ where E is a matrix full of 1's and $k = 1, 2, \ldots, n$. Clearly, A_k are all positive since $A \geq 0, \quad k > 0, \quad E > 0$. Let $r_k > 0$ be the Perron root and $\mathbf{v}_k$ the corresponding eigenvector for A_k. Then $\{\mathbf{v}_k\}_{k=1}^{\infty} \subseteq \{\mathbf{x} \in \mathbb{R} : ||\mathbf{x}||_1 = 1\}$ so $\lim_{s\to\infty} A_s\}_{k=1}^{\infty}$ is bounded. The Bolzano-Weierstrass theorem gives that the bounded real sequence has a convergent subsequence. So, $\{\mathbf{v}_s\}_{s=1}^{\infty} \to \mathbf{z}$ for some $\mathbf{z} \in \mathbb{R}^n$ and increasing sequence s. $\mathbf{z} > 0$ since $\mathbf{v}_s > 0$. Note, k is increasing so $A_1 > A_2 > \ldots > A$. By the triangle inequality, $|A^s| \leq |A|^s$ so $|A_1|_{\infty}^s \geq |A_2|_{\infty}^s \geq \ldots \geq |A|_{\infty}^s$ so $|A_1^s|_{\infty}^{1/s} \geq |A_2^s|_{\infty}^{1/s} \geq \ldots \geq |A^s|_{\infty}^{1/s}$ so

$$\lim_{s\to\infty} |A_1^s|_{\infty}^{1/s} \geq \lim_{s\to\infty} |A_2^s|_{\infty}^{1/s} \geq \ldots \geq \lim_{s\to\infty} |A^s|_{\infty}^{1/s}.$$

See the appendix for a definition of spectral radius 8.16

We know the spectral radius $\rho(A) = \lim_{k\to\infty} ||A^k||^{1/k}$. Thus, $r_1 \geq r_2 \geq \ldots \geq r$. So we have $\lim_{s\to\infty} r_s = r^* \geq r$. However, $\lim_{s\to\infty} A_s = A$. Then,

$$A\mathbf{z} = \lim_{s\to\infty} A_s \lim_{s\to\infty} \mathbf{v}_s = \lim_{s\to\infty} r_s \mathbf{v}_s = r^*\mathbf{z}$$

so we have $A\mathbf{z} = r^*\mathbf{z}$. Hence r^* is an eigenvalue so $r^* \in \sigma(A) \implies r^* = r$. So, statement (1) is true [3].
(2) Note, proving that $|\lambda| < r$ for the positive matrix implies the same for the primitive matrix since the eigenvalues of A^k are the k-th powers of the eigenvalues of A.

Lemma 5.4. Let $B > 0$ be a positive matrix with $r = 1$. Then all other eigenvalues of B satisfy $|\lambda| < 1$.

Proof. Suppose that $\mathbf{z}$ is an eigenvector of B with eigenvalue λ and $|\lambda| = 1$. Then $|\mathbf{z}| = |\lambda\mathbf{z}| = |B\mathbf{z}| \leq |B||\mathbf{z}| = B|\mathbf{z}|$ using the notation $(|B\mathbf{z}|)_i = |b_{i1}z_1 + b_{i2}z_2 + \ldots + b_{in}z_n|$. Then, $|\mathbf{z}| \leq B|\mathbf{z}|$. Let $\mathbf{y} := B|\mathbf{z}| - |\mathbf{z}|$. So, $\mathbf{y} \geq 0$. Suppose for a contradiction that $\mathbf{y} \neq 0$. Then, $B\mathbf{y} > 0$ and $B|\mathbf{z}| > 0$, so there exists $\epsilon > 0$ such that $B\mathbf{y} > \epsilon B|\mathbf{z}|$ and thus $B(B|\mathbf{z}| - |\mathbf{z}|) > \epsilon B|\mathbf{z}|$. Equivalently,

$$E(B|\mathbf{z}|) > B|\mathbf{z}|, \text{ where } E := \frac{1}{1+\epsilon} B.$$

This would imply that $E^k B|\mathbf{z}| > B|\mathbf{z}|$ for all k. However, the eigenvalues of E are all less than one in absolute value. So, $E^k \to 0$. Then, all entries of $B|\mathbf{z}|$ are less than or equal to zero, which is a contradiction to the fact that $B|\mathbf{z}| > 0$. Therefore, $|\mathbf{z}|$ must be an eigenvector of B with eigenvalue 1. Note, $|B\mathbf{z}| = |\mathbf{z}|$ so $|B\mathbf{z}| = B|\mathbf{z}|$ which can only occur when all entries of $\mathbf{z}$ are of the same sign. Then, $\mathbf{z}$ must be a multiple of the eigenvector $\mathbf{x}$ since there are no other eigenvectors with all entries of the same sign other than the multiples of $\mathbf{x}$ by part (4). So, $\lambda = 1 = r$. $\square$

Note that the matrix B is found by dividing any positive matrix by its r. Thus, we have proved part (2) [17].

(3) Let Λ be a diagonal $n \times n$ matrix, with entries $\lambda_1, \dots, \lambda_n$ along the diagonal. Expanding $\det(\Lambda I - A)$ along the i-th row shows that

$$\frac{\partial}{\partial \lambda_i} \det(\Lambda I - A) = \det(\Lambda_{i,i} - A_{i,i})$$

where the matrix subscript i, i indicates the minor. Substituting $r = \lambda_i$ and the chain rule gives

$$\frac{d}{d\lambda} \det(rI - A) = \sum_i \det(rI - A_{(i,i)}).$$

Then, each of the matrices $rI - A_{(i,i)}$ has a strictly positive determinant and thus the algebraic and geometric multiplicity of $rI - A_{(i,i)}$ are each one [17].

(4) Suppose A has an eigenvector $\mathbf{v}$ with strictly positive entries and let λ_v denote the corresponding eigenvalue, so that $A\mathbf{v} = \lambda_v \mathbf{v}$. Define B so that

$$\mathbf{v} = \begin{pmatrix} v_1 & 0 & \dots & 0 \\ 0 & v_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & v_n \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = B \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Then

$$A \cdot B \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \lambda_v B \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \implies B^{-1} A B \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \lambda_v \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

As

$$B^{-1} = \begin{pmatrix} v_1^{-1} & 0 & \dots & 0 \\ 0 & v_2^{-1} & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & v_n^{-1} \end{pmatrix},$$

the matrix $BAB^{-1} \geq 0$. We assume that $\mathbf{v} = (1, \dots, 1)^T$ which we can claim without loss of generality since similar matrices have the same eigenvalues. Then $\lambda_v = \sum_{j=1}^n a_{ij}$ for each $1 \leq i \leq n$. This is clearly a positive real number, unless $A = 0$. Let $\mathbf{z} = (z_1, \dots, z_n)^T$

and $\|\mathbf{z}\| = \max_{i=1,\ldots,n} |z_i|$. So for any $\mathbf{z} \in B^n$ we have

$$|a_{i1}z_1 + \ldots + b_{in}z_n| \leq |b_{i1}||z_1| + \ldots + |b_{in}||z_n| \leq \sum_{j=1}^{n} b_{ij} \max_{i=1,\ldots,n} |z_i| = \lambda_v \|\mathbf{z}\|. \tag{6}$$

by the triangle inequality. Then, $\|A\mathbf{z}\| \leq \lambda_v \|\mathbf{z}\|$ and, for some eigenvector λ_1 and its eigenvalue $\mathbf{v}_1$, $\|A\mathbf{z}\| = |\lambda_1|\|\mathbf{v}_1\| \leq \lambda_v\|\mathbf{z}\|$. Thus, $\lambda_v \geq |\lambda_1|$ and so $\lambda_v = r$. So the Perron eigenvector is unique [10].

(5) Define

$$L(\mathbf{z}) := \max\{s : s\mathbf{z} \leq A\mathbf{z}\} = \min_{1 \leq i \leq n, z_i \neq 0} \frac{(A\mathbf{z})_i}{z_i}$$

and λ_L its Perron root and Q denotes the positive orthant. Suppose that $0 \leq B \leq A, B \neq A$. If $\mathbf{z} \in Q$ is a vector such that $s\mathbf{z} \leq B\mathbf{z}$, then since $B\mathbf{z} \leq A\mathbf{z}$, we get $s\mathbf{z} \leq A\mathbf{z}$. So, $\lambda_{L_B}(\mathbf{z}) \leq \lambda_{L_A}(\mathbf{z})$ for all $\mathbf{z}$ and hence $0 \leq B \leq A \implies \lambda_L(B) \leq \lambda_L(A)$. Then,

$$0 \leq B \leq A, B \neq A \implies r(B) \leq r(A).$$

We have followed the proof by Sternberg [17].

(6) Let $\beta = f(\mathbf{x})$ for some arbitrary $\mathbf{x} \in \mathcal{N}$. Since $f(\mathbf{x}) \leq [A\mathbf{x}]_i/x_i$ for each i, we have that $\beta \leq [A\mathbf{x}]_i/x_i$ for each i. So, $\beta x_i \leq [A\mathbf{x}]_i$, which implies that $0 \leq \beta\mathbf{x} \leq A\mathbf{x}$. Next, let $\mathbf{q}_k^T$ be the left-hand Perron root for A_k. Also let $\mathbf{p}$ and $\mathbf{q}^T$ be the respective right-hand and left-hand Perron roots for A associated with the Perron root ρ. For all $x \in \mathcal{N}$, $\quad k > 0$ we have $\mathbf{q}^T\mathbf{x} > 0$. Then, $\beta\mathbf{x} \leq A\mathbf{x} \leq A_k\mathbf{x} \implies \beta\mathbf{q}_k^T\mathbf{x} \leq (\mathbf{q}_k^T A_k)\mathbf{x} = r_k\mathbf{q}_k^T\mathbf{x} \implies \beta \leq \rho_k \implies \beta = f(\mathbf{x}) \leq r$ since $r_k \to r$. Therefore, $f(\mathbf{x}) \leq r$ for all $\mathbf{x} \in \mathcal{N}$ [3, 14].

Now we want to show that $r = \max_{\mathbf{x} \in \mathcal{N}} f(\mathbf{x})$. Since $f(\mathbf{x}) \leq r$, $r \geq \max_{\mathbf{x}}$ Note,

$$f(\mathbf{p}) = \min_{1 \leq i \leq n} \frac{[A\mathbf{p}]_i}{p_i} = \min_{1 \leq i \leq n} \frac{[r\mathbf{p}]_i}{p_i} = r \min_{1 \leq i \leq n} \frac{p_i}{p_i} = r.$$

Since $f(\mathbf{p}) = r$ for $\mathbf{p} \in \mathcal{N}$, then $r = \max_{\mathbf{x} \in \mathcal{N}} f(\mathbf{x})$ as r cannot be larger than $f(\mathbf{x})$ and still equal $f(\mathbf{p})$ [3].

$\square$

Theorem 5.5. Let A be a nonnegative irreducible matrix and $r = \rho(A)$. Then A is primitive if and only if

$$\lim_{k \to \infty} \left(\frac{A}{r}\right) = G = \frac{\mathbf{p}\mathbf{q}^T}{\mathbf{q}^T\mathbf{p}} > 0$$

where $\mathbf{p}$ and $\mathbf{q}$ are the respective Perron vectors for A and A^T. Also, G is the spectral projector of r onto $N(A - rI)$ along $R(A - rI)$ (where $R(P) = N(I - P)$, $R(I - P) = N(P)$) [3, 14, 5].

Proof. By the Perron-Frobenius theorem 5.3, $1 = \rho(A/r)$ is an eigenvalue for A/r with multiplicity 1. We know matrix A is primitive if and only if A/r is primitive. Also, A is primitive if and only if there exists a $k > 0$ such that $A^k > 0$ by theorem 3.4. We see that $(A/r)^k = (1/r^k)(A^k)$ so when $A^k > 0$, then $(A/r)^k > 0$. However, by definition A/r can only be primitive if $1 = \rho(A/r)$ is the only eigenvalue on its spectral radius. We take it as a fact that $\lim_{k\to\infty}(A/r)^k$ must exist since $1 = \rho(A/r)$ is a simple eigenvalue and also the only eigenvalue in the spectral radius. Also,

$$\lim_{k\to\infty}\left(\frac{A}{r}\right)^k$$

is a projector (8.13) onto $N\left(I - \frac{A}{r}\right)$ along $R\left(I - \frac{A}{r}\right)$. Then $\lim_{k\to\infty} A^k$ exists and is equal to the projector onto $N(I - A)$ along $R(I - A)$. Then,

$$G = \lim_{k\to\infty}\left(\frac{A}{r}\right)^k = \frac{\mathbf{p}\mathbf{q}^T}{\mathbf{q}^T\mathbf{p}} > 0$$

is a projector since a linear projector on some V is a projector if and only if it is equal to its square [3]. □

See the appendix for a definition of projector 8.13 and reducible 8.14.

Example 5.6. The matrix

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

is primitive. A has eigenvalues $\rho = 2$ and -1, so the associated Perron vector $\mathbf{p}$ is $(1,1)^T$ and A^T has the same eigenvalues with $\mathbf{q} = (1,2)^T$. Then we check

$$\frac{\mathbf{p}\mathbf{q}^T}{\mathbf{q}^T\mathbf{p}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \end{bmatrix} \div \begin{bmatrix} 1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}\frac{1}{3} > 0.$$

So, theorem 5.5 confirms that A is primitive [3].

Now, we will consider the Leslie model again with these theorems in our tool belt. Recall: $\mathbf{f}(t+1) = L\mathbf{f}(t)$ from definition 5.1. First, we divide the population into age classes $G_1, G_2, \ldots, G_n$ such that each class covers the same amount of time. Let $f_k(t)$ represent the number of individuals in G_k at time t where $k = 2, 3, \ldots, n$. Then we see $f_1(t+1) = f_1(t)b_1 + f_2(t)b_2 + \ldots + f_n(t)b_n$, $f_k(t+1) = f_{k-1}(t)s_{k-1}$. Note, $\mathbf{F}(t) = (F_1(t), F_2(t), \ldots, F_n(t))^T$ with

$$F_k(t) = \frac{f_k(t)}{f_1(t) + f_2(t) + \ldots + f_n(t)},$$

finds the percentage of the population in G_k at time t [3, 14]. We know L is primitive when $b_i > 0, b_{i+1} > 0$ for some i. Also, for any Leslie matrix where $a_n > 0$ and $b_i > 0$, L is irreducible [17, 5]. Then, $r = \rho(L) > 0$ is an eigenvalue of the matrix by theorem 5.3. As L is primitive,

$$\lim_{k\to\infty}\left(\frac{L}{r}\right)^t = G = \frac{\mathbf{p}\mathbf{q}^T}{\mathbf{q}^T\mathbf{p}} > 0$$

by theorem 5.5. So, $\mathbf{f}(t) = L^T\mathbf{f}(0)$ follows from induction. If $\mathbf{f}(0) = 0$, then

$$\lim_{t\to\infty}\frac{\mathbf{f}(t)}{r^t} = \lim_{t\to\infty}\frac{L^t\mathbf{f}(0)}{r^t} = G\mathbf{f}(0) = \frac{\mathbf{p}\mathbf{q}^T\mathbf{f}(0)}{\mathbf{q}^T\mathbf{p}}.$$

In fact,

$$\mathbf{F}^* = \lim_{t\to\infty}\mathbf{F}(t) = \lim_{t\to\infty}\frac{\mathbf{f}(t)}{\|\mathbf{f}(t)\|_1} = \lim_{t\to\infty}\frac{\mathbf{f}(t)/r^t}{\|\mathbf{f}(t)\|_1/r^t} = \mathbf{p}.$$

We have found $\mathbf{F}^*$, the steady-state age distribution for the Leslie matrix L [3, 14].

Example 5.7. Let us consider the rabbit problem again. Recall; the total number of rabbit pairs at the end of month n is equal to the sum of the number of pairs at the end of the two previous months $c_t = c_{t-1} + c_{t-2}$. Then we have

$$\begin{bmatrix} y_{t+1} \\ a_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_t \\ a_t \end{bmatrix} \text{ where } \begin{bmatrix} y_0 \\ a_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Which can be written as

$$\mathbf{f}(t+1) = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{f}(t),$$

where $\mathbf{f}(0) = (1, 0)^T$ [3]. Let

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

The eigenvalues for L are $\frac{1 \pm \sqrt{5}}{2}$. Then, the Perron root is $\frac{1+\sqrt{5}}{2}$ with the corresponding eigenvector

$$\mathbf{v}_1 = \begin{bmatrix} 1 + \sqrt{5} \\ 2 \end{bmatrix}.$$

Then, the steady state of the rabbit population will be $1 + \sqrt{5}$:2 youth to adults by the above.

5.2. **Google Page Ranking Algorithm.** In 1998, Sergey Brin and Larry Page developed Google's PageRank. It is very well know now, even acknowledging that it was not the first page ranking algorithm. The algorithm is designed to place value on a webpage based on the links between pages [6]. Generally, the more links to a page, the higher its page rank [17]. Also, if an "important" page links to a page, that page rank should increase. So we see, importance of links is not equal. Even before the internet, a page rank algorithm could be used to rank journals based on the citations from other journals [6].

Ideally, the ranking for a Google result satisfies an equation similar to

$$r(P_i) = \sum_{P_j \in B_{P_i}} \frac{r(P_j)}{|P_j|}$$

where $r(P)$ is the desired ranking, B_{P_i} is the set of "pages" pointing to P_i, and $|P_j|$ is the number of links out of P_j. If $\mathbf{r}$ denotes the row vector whose i-th entry is $r(P_i)$ and H denotes the matrix whose i, j entry is $\frac{1}{|P_i|}$ if there is a link from P_i to P_j, then the equation becomes $\mathbf{r} = \mathbf{r}H$. The matrix H is of size $n \times n$ where n is the number of pages at the current time.

Example 5.8. Consider an imaginary internet where only four pages exist. We will call them page 1, page 2, page 3, and page 4. Page 1 links to pages 2, 3, and 4. Page 2 links to pages 3 and 4. Page 3 links to page 1. Page 4 links to pages 1 and 3. We want to calculate their rankings and assume they start out with equal rank at 1/4 each. Since page 1 links out three times, its weight is 1/3. Page 3 only links out once, so its weight is 1. The other two pages are assigned weights similarly. Then, we can write this as the transition matrix A where

$$A = \begin{array}{c c} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{bmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix} \end{array} \text{ and } \mathbf{v}_0 = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}.$$

To update the ranking, we multiply the starting vector $\mathbf{v}_0$ by A until we reach a steady state. Then,

$$\mathbf{v}^* = \mathbf{v}_0 \cdot \lim_{k\to\infty} A^k \approx \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{bmatrix} 0.387 \\ 0.129 \\ 0.290 \\ 0.193 \end{bmatrix}$$

which tells us that page 1 has the highest page rank at about 0.387. Now, if we find the Perron root for the transition matrix A and the associated eigenvector, we get eigenvalue 1 with eigenvector

$$\mathbf{v}_1 = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{bmatrix} 12 \\ 4 \\ 9 \\ 6 \end{bmatrix} \approx 31 \cdot \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{bmatrix} 0.387 \\ 0.129 \\ 0.290 \\ 0.193 \end{bmatrix}.$$

Notice how page 1 ended up with the highest rank even though 3 is linked to more. As we intended, the page rank was affected by weight of a page's link [19].

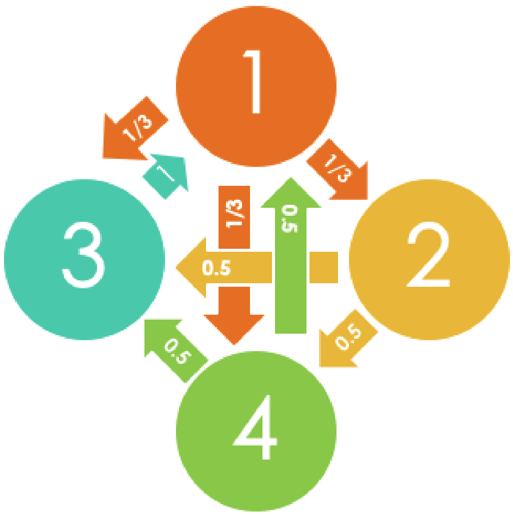

FIGURE 4. A depiction of Example 5.8

In reality, n is a very large number, so the page rank problem is best solved by iteration. While H is very large, it is also sparse (mostly zeros) so the calculation is quite doable [17]. Some rows in this vector will be all zeros, so Brin and Page cleverly replaced these rows with entries $\frac{1}{n}$. Let $S := H + \frac{1}{n}\mathbf{a} \otimes \mathbf{e}$ where $\mathbf{a}$ is the column vector whose i-th entry is one and $a_i = 0$ otherwise and $\mathbf{e}$ is the row vector full of ones. Now, S is a Markov chain matrix whose row sums all equal one. Then, to make S primitive, they set

$$G := \alpha S + (1-\alpha)\frac{1}{n}J$$

where J is full of ones and $\alpha \in \mathbb{R}$, $0 < \alpha < 1$. Thus, G is positive and G^k converges to a matrix whose rows are all equal to $\mathbf{s}$ where $\mathbf{s} = \mathbf{s}G$ by theorem 4.4. However, G may not be sparse, so computing powers is a nuisance. Consider the iteration

$$\mathbf{s}_{k+1} = \mathbf{s}_k G = \alpha \mathbf{s}_k S + \frac{1-\alpha}{n}\mathbf{s}_k J = \alpha \mathbf{s}_k H + \frac{1}{n}(\alpha \mathbf{s}_k \mathbf{a} + 1 - \alpha)\mathbf{e}.$$

Now, we have only sparse multiplication [17].

To clarify why G converges, let $1, \lambda_2, \lambda_3, \ldots$ be the spectrum of S and let $1, \mu_2, \mu_3, \ldots$ be the spectrum of G arranged in decreasing order such that $\lambda_2 < 1, \mu_2 < 1$. Then, if $\lambda_i = \alpha\mu_i$, $i = 2, 3, \ldots, n$, we see that $\lambda_2 < \alpha$ since $\mu_2 < 1$ [17].

Now, we can solve the problem using eigenvectors guaranteed by the Perron-Frobenius theorem 3.1 instead of a lot of matrix multiplication. However this may not be very fast either. That is why Brin and Page used the previously mentioned iteration until convergence was reached. This convergence is guaranteed by Markov's theorem 4.4. In fact, a good approximation is reached fairly quickly [19].

Before PageRank, Google returned searches based on the occurrence of the search words. However, this returned pages that used the searched words most, not the ones that were most relevant. For example, a page reviewing Colorado State University that used the phrase "Colorado State University" most might be the first result, over Colorado State University's official website. Adding in the PageRank algorithm helped fix this problem [19].

6. Conclusion

Perron's theorem is continuously developing. The theorem is over one hundred years old. It started off as a small conclusion in a paper about convergence criteria of partial sums. Mathematicians have continued to add conclusions ever since it was first stated. Now, Perron-Frobenius theory is an entire topic in mathematics. Seeing all the work that's been done on the theorem over the years, it is clear that others agree, Perron-Frobenius Theory is worth studying.

The applications of Perron's theorem are wide, varied and evolving. His theorem is working in the background of things we interact with daily. The

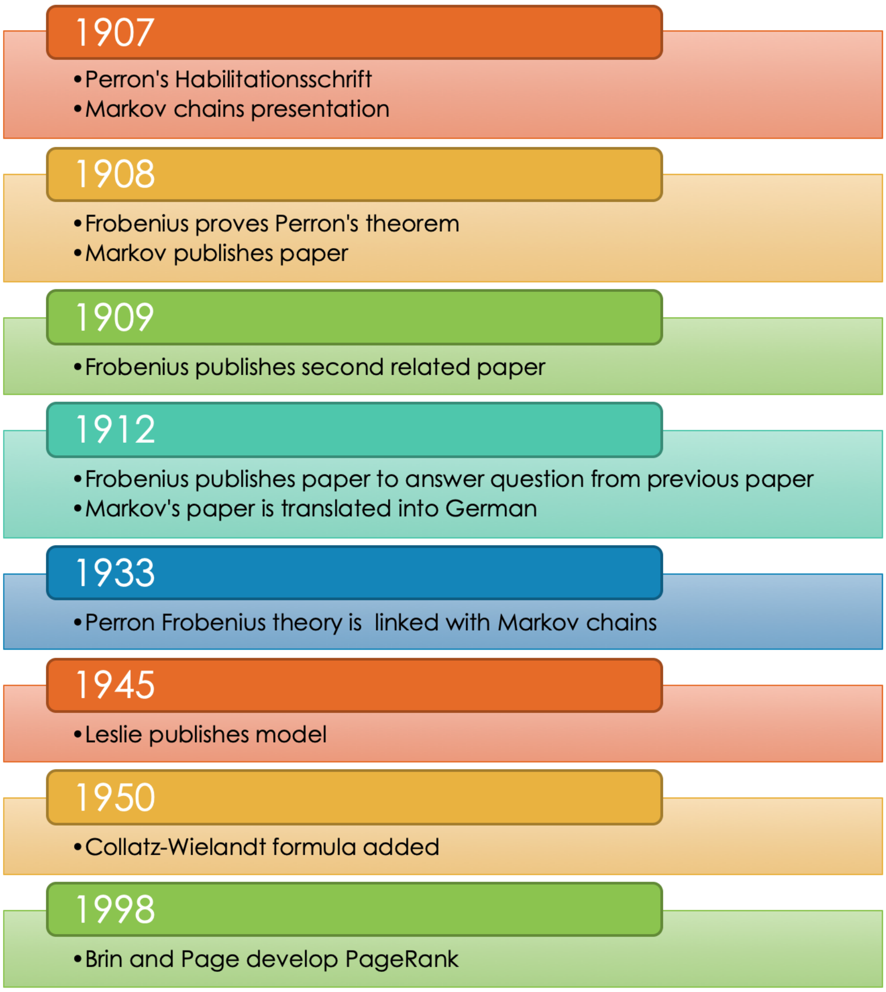

FIGURE 5. Timeline of Perron's Theorem

Perron-Frobenius theorem has applications that range from simple to complex, such as a simple population model as well as a molecular thermodynamic equilibrium model. Future applications may be found in fields you would least expect.

As the history of Perron's theorem reveals, mathematicians come up with new math using building blocks of the past, and then make constant improvements. Mathematicians not only improve upon theorems, they also enhance proofs and applications. They even find surprising new connections, such as the link between matrix theory and the convergence of partial sums.

In answer to the question; "What will you do with your degree?" I would say; "I will answer interesting questions like Perron and Frobenius did. I will solve real problems using math like Markov did. And, I will think of new ways to use old math like the mathematicians who linked Perron-Frobenius theory to new problems like the Leslie model and Google's PageRank algorithm." My exploration of Perron's theorem has reinforced that something that isn't well known can still have a large impact. There are many brilliant mathematicians who we owe modern math to. I may not know their names but I use their contributions. I intend continue their legacy.

7. Bibliography

References

[1] Anton, Howard. *Elementary Linear Algebra.* 11th ed., Wiley, 2014.

[2] Axler, Sheldon. *Linear Algebra Done Right.* 3rd ed., *Springer*, 2015.

[3] Baldin, Jerika. "The Perron-Frobenius Theorem and its Application to Population Dynamics." 2017. https://www.lakeheadu.ca/sites/default/files/uploads/77/Baldin.pdf

[4] Bernhardt, Chris. "Powers of Positive Matrices." *Mathematics Magazine.* 2018. pp. 218-227. https://doi.org/10.1080/0025570X.2018.1446615.

[5] Carnia, E, et al. "Mathematical Properties of nxn Nonnegative Matrix: Case of Irreducible Leslie Matrix." *Journal of Physics: Conference Series*, vol. 1280, 2019. https://iopscience.iop.org/article/10.1088/1742-6596/1280/2/022048.

[6] Emerging Technology from the arXiv. "Scientist Finds PageRank-Type Algorithm from the 1940s." MIT Technology Review, 17 Feb. 2010. www.technologyreview.com/2010/02/17/264032/scientist-finds-pagerank-type-algorithm-from-the-1940s/.

[7] Hawkins, Thomas. "Continued fractions and the origins of the Perron–Frobenius theorem." *Springer*, 2008. https://link.springer.com/article/10.1007/s00407-008-0026-x.

[8] Hawkins, Tomas "Nonnegative Matrices." *The Mathematics of Frobenius in Context: a Journey Through 18th to 20th Century Mathematics*, Springer, 2013, pp. 607–650.

[9] Kac, Mark. "Random Walk and the Theory of Brownian Motion." *The American Mathematical Monthly*, vol. 54, no. 7, Sept. 1947, pp. 369–391, doi:10.2307/2304386.

[10] Khim, Suyeon. "The Frobenius-Perron Theorem." 2007. https://www.scribd.com/document/71078313/Khim.

[11] MacCluer, C. R. "The Many Proofs and Applications of Perron's Theorem." 42, *SIAM Review.* 2000. pp. 487-498. https://epubs.siam.org/doi/pdf/10.1137/S0036144599359449.

[12] MacCluer, C. R. "Perron's Theorem and Markov Chains." 2005. https://users.math.msu.edu/users/maccluer/Class/415/Perron.pdf.

[13] "Markov Chain." *Wikipedia*, Wikimedia Foundation, 2020. en.wikipedia.org/wiki/Markov_chain.

[14] Meyer, Carl D. "Chapter 8: Perron-Frobenius Theory of Nonnegative Matrices." *Matrix Analysis and Applied Linear Algebra*, SIAM, Society for Industrial and Applied Mathematics, 2000, pp. 661–703.

[15] Parys, Wilfried."Why Didn't Charasoff and Remak Use Perron-Frobenius Mathematics?" *The European Journal of the History of Economic Thought* 21 (2014): 991–1014. https://doi.org/10.1080/09672567.2014.951672.

[16] Pierce, Rod. "Eigenvector and Eigenvalue." *Math Is Fun*, 20 Sept. 2020. www.mathsisfun.com/algebra/eigenvalue.html.
[17] Sternberg, Shlomo. "Chapter 9 The Perron-Frobenius Theorem." *Dynamical Systems*, Dover Publications, Inc., 2010, pp. 175–195.
[18] "Stochastic Matrix." *Wikipedia*, Wikimedia Foundation, 2020. en.wikipedia.org/wiki/Stochastic_matrix.
[19] Tanase, Raluca, and Remus Radu. "Lecture #3." *PageRank Algorithm - The Mathematics of Google Search*, 2009. pi.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/lecture3.html.

8. Background Mathematics

Definition 8.1. If A is any $n \times n$ matrix and C_{ij} is the cofactor of A_{ij}, then the matrix

$$\begin{pmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} \end{pmatrix}$$

is called the matrix of cofactors from A and the transpose of this matrix is called the <u>adjoint</u> of A and is denoted by $\text{Adj}(A)$ [1].

Definition 8.2. Let A be an $n \times n$ matrix. The expansion of the determinant $\det(\lambda I - A) = 0$ (the characteristic equation) gives $\lambda^n + c_1\lambda^{n-1} + ... + c_n = 0$ which is called the <u>characteristic polynomial</u> of A [2].

Definition 8.3. The <u>cofactor</u> of a matrix $A_{n\times n}$ is the (i, j)-minor (M_{ij}) times a sign factor: $C_{ij} = (-1)^{i+j}M_{ji}, \quad 1 \leq i, j \leq n$ [1].

Definition 8.4. The <u>determinant</u> of A is the product of the eigenvalues of A, with each eigenvalue repeated according to its multiplicity. The determinant is denoted $\det(A)$ [2]. The cofactor expansions are $\det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \dots + a_{nj}C_{nj}$ and $\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \dots + a_{in}C_{in}$ [1].

Definition 8.5. A number λ is called an <u>eigenvalue</u> of A if there exists $\mathbf{v}$ in the vector space such that $\mathbf{v} \neq 0$ and $A\mathbf{v} = \lambda\mathbf{v}$. This number is also called the characteristic root [2]. Equivalently, eigenvalues are the solutions to the characteristic equation $\det(\lambda I - A) = 0$ [1].

Definition 8.6. If λ is an eigenvalue of A, then a vector $\mathbf{v}$ in the vector space is called an <u>eigenvector</u> of A corresponding to λ when $\mathbf{v} \neq 0$ and $A\mathbf{v} = \lambda\mathbf{v}$ [2].

Definition 8.7. The <u>Jordan canonical form</u> refers to a basis called a Jordan basis with the block diagonal matrix

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_p \end{pmatrix}$$

where each A_j is an upper-triangluar matrix of the form

$$A_j = \begin{pmatrix} \lambda_1 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{pmatrix}$$

[2].

Definition 8.8. If A is an $n \times n$ matrix, the minor of entry a_{ij} is the determinant of the matrix with the i-th row and j-th column deleted. This minor is denoted $M_{i,j}$ [1].

Definition 8.9. Algebraic multiplicity is the number of times λ appears as a root of the characteristic polynomial. Geometric multiplicity is the number of linearly independent eigenvectors associated with λ or in other words, $\text{geomult}_A(\lambda) = \text{dimension}N(A - \lambda I)$ [3].

Definition 8.10. A nonnegative matrix is a matrix whose elements are all zero or positive. $\{\mathbf{x} \mid \mathbf{x} \geq 0 \text{ with } \mathbf{x} \neq 0\}$ means that a vector $\mathbf{x}$ can have entries greater than or equal to zero, but at least one entry must be nonzero.

Definition 8.11. A positive matrix is a matrix whose elements are all strictly positive.

Definition 8.12. A primitive matrix is a nonnegative matrix that is irreducible and has one diagonal element [5]. A non-negative irreducible matrix A having only one eigenvalue, $r = \rho(A)$, on its spectral circle is said to be a primitive matrix. Equivalently, a matrix is called primitive if there exists a positive integer k such that A^k is a positive matrix [3].

Definition 8.13 ([3])**.** Let X and Y be vector spaces over C and let $\mathbf{v}$ be in vector space V. The unique linear operator P defined by $P\mathbf{v} = \mathbf{x}$ is called the projector onto X along Y and P has the following properties:

(1) $P^2 = P$
(2) $R(P) = \{\mathbf{x} \mid P\mathbf{x} = \mathbf{x}\}$
(3) $R(P) = N(I - P) = X$ and $R(I - P) = N(P) = Y$

Definition 8.14. A matrix $A \geq 0$ is said to be a reducible matrix when there exists a permutation matrix P such that

$$P^T A P = \begin{pmatrix} X & Y \\ 0 & Z \end{pmatrix}$$

where X and Z are both square [3]. Note, the matrix must be at least 2×2 [7]. Otherwise, A is said to be an irreducible matrix [3].

Definition 8.15. An eigenvalue λ is called simple if $\text{algmult}A(\lambda) = 1$. If A is a square matrix then λ is a semi-simple eigenvalue if and only if $\text{algmult}A(\lambda) = \text{geomult}A(\lambda)$ [3].

Definition 8.16. For any square matrix A, the spectral radius is the maximum eigenvalue of A in absolute value, that is, the number $\rho(A) = \max_{\lambda \in \omega(A)} |\lambda|$ where $\omega(A)$ is the set of distinct eigenvalues [5, 3].