THE GOOGLE SPECTRUM

VARUN IYER

Abstract. Google's PageRank algorithm is vastly superior to its rival search algorithms because of the mathematics behind it. As one of the many results of Perron's Theorem, the algorithm is a discrete-time finite-state Markov chain that relies upon its transition probability matrix's spectrum for its convergence. This paper introduces the spectrum, Perron's theorem and proof, some of its applications, and basic Markov Chain theory, including, finally, a description of the PageRank algorithm.

Contents

1. Introduction

Larry Page and Sergey Brin developed Google's PageRank algorithm in 1998, still during the early stages of the Internet. By 2002, 2 years before Google's IPO, PageRank was operating on a matrix at the order of 2.7 billion, cited by Cleve Moler as "The World's Largest Matrix Computation" [**4**]. Today Google searches over 30 trillion webpages, 100 billion times a month, and those numbers are still growing as new sites are used and new users Google on a daily basis. The algorithm, that was far more efficient than its rival '90s search engines at ranking web pages based on a given search, is still exceedingly relevant today, with some minor modifications over the years. To understand Google's PageRank is to understand the development and structure of the Internet.

The goal of this paper is to highlight what makes the Google PageRank algorithm so superior to its rival search engines by explaining its core mathematics. To get to this discussion, however, we need to cover several different topics. First, we begin with some spectral graph theory to provide a graphical understanding of the spectrum, a tool that will be used numerous times. Describing some properties of graphs that can be found from the spectrum will provide some intuition for

when we later discuss an example of a large directed graph, the Internet. We move on to stating and proving Perron's Theorem, an elegant foundation to many practical mathematics. A related theorem shows the convergence properties of iterative processes, which we will use to discuss some of Perron's other applications in ecology and economics. We then define the power method, a central tool to the PageRank algorithm and other Perron applications, and then introduce some Markov Chain theory which is crucial to understanding the intuition behind creating PageRank. Finally explaining in detail the rationale behind the Google matrix and the PageRank algorithm, we end with a proof of the Google matrix's spectrum, which is fundamentally responsible for the algorithm's efficiency. Hopefully this paper explains the first revolutionary search algorithm in depth, and will perhaps inspire future algorithms to be modeled from a similar mathematical foundation.

2. The Spectrum of a Graph

Spectral graph theory is the study of the properties of graphs by analyzing the eigenvalues and eigenvectors of the graph's associated matrices, including the adjacency matrix. We begin by defining a graph, and introduce the graph's spectrum. In later sections, we will prove a useful theorem about the spectrum of a positive matrix, and then use the spectrum to study a particular graph, the Internet. For now, this introduction to spectral graph theory should open some ideas to how the manipulation of the matrix of a graph can provide useful information about the graph's structure.

Definition 2.1. A *graph* is an ordered pair $G = (V, E)$, where V is the set of *vertices* and E is the set of *edges*, such that:

$$E \subset \{(x, y) \mid x, y \in V, x \neq y\}.$$

The graph we defined above is an undirected graph. A directed graph contains edges with directions, where an edge that joins v_i to v_j may not necessarily join v_j to v_i. In this section, we limit our discussion to undirected graphs, but the concept of directed graphs will become important in a later section.

If v_i and v_j are vertices of G and $e = \{v_i, v_j\}$ is an edge that joins the two vertices, we call v_i and v_j adjacent to one another. In other words, two vertices are adjacent in a graph if there is a single connecting edge. The degree $d(v)$ of a vertex v is the number of vertices that are adjacent to v. We will now define the adjacency matrix, which will be crucial to understanding the graph's properties.

Definition 2.2. Suppose G is a graph with vertex set $V = \{v_1, v_2, \ldots, v_n\}$. The *adjacency matrix* of graph G is the $n \times n$ matrix $\mathbf{A}$ whose entries a_{ij} are given by:

$$a_{ij} = \begin{cases} 1, & \text{if } v_i \text{ and } v_j \text{ are adjacent;} \\ 0, & \text{otherwise.} \end{cases}$$

The adjacency matrix is both real and symmetric, by definition. Since the labeling of a graph's vertices is arbitrary, properties of the matrix that are invariant under permutations of rows and columns are of particular interest.

Suppose that λ is an eigenvalue of $\mathbf{A}$. Since $\mathbf{A}$ is real and symmetric, λ is also real. This λ is a root of the characteristic polynomial $\det(\lambda\mathbf{I} - \mathbf{A})$. We now have the tools to define the spectrum of a graph.

Definition 2.3. The *spectrum* of graph G is the set of eigenvalues of $\mathbf{A}$ combined with their respective algebraic multiplicities. If the k distinct eigenvalues of $\mathbf{A}$ are $\lambda_0 > \lambda_1 > \ldots > \lambda_k$, then we write

$$\mathbf{Spec} = \begin{pmatrix} \lambda_0 & \lambda_1 & \ldots & \lambda_k \end{pmatrix}$$

Applications of the spectrum will come into play in later sections. For now, we will show some properties of graphs that can be derived from $\mathbf{A}$.

Suppose the characteristic polynomial $\det(\lambda\mathbf{I} - \mathbf{A})$ of G is

$$\chi(G, \lambda) = \lambda^k + c_1\lambda^{k-1} + c_2\lambda^{k-2} + c_3\lambda^{k-3} + \cdots + c_k$$

The coefficients of χ can be expressed in terms of the principal minors of $\mathbf{A}$, where a principal minor is the determinant of a submatrix obtained by taking a subset of the rows and the same subset of the columns.

For the following proposition, we define triangles T as subgraphs containing 3 mutually adjacent vertices.

Proposition 2.4. *The coefficients of the characteristic polynomial χ of graph G have the following properties:*

(1) $c_1 = 0$;
(2) $-c_2$ is the number of edges in G
(3) $-c_3$ is twice the number of triangles in G

Proof. For $i \in \{1, 2, \ldots, k\}$, the value of $(-1)^i c_i$ is the sum of the principal minors of $\mathbf{A}$ with i rows and i columns.

(1) The diagonal elements of $\mathbf{A}$ are all 0, so $c_1 = 0$.
(2) A nonzero principal minor of $\mathbf{A}$ with 2 rows and 2 columns must be

$$\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1.$$

For each pair of adjacent vertices in G, there is one such minor. Hence $(-1)^2 c_2 = -1 \times E$.
(3) The nontrivial possibilities for principal minors of three rows and columns include:

$$\begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix} = 0, \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{vmatrix} = 0, \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{vmatrix} = 2.$$

The only nonzero minor corresponds to 3 mutually adjacent vertices in G, or a triangle T. Thus $(-1)^3 c_3 = 2 \times T$.

□

We end our introduction to spectral graph theory by defining the walk and what it means for a graph to be connected. A *walk* of length l in G, from v_i to v_j, is a finite sequence of vertices of G,

$$v_i = u_0, u_1, \ldots, u_l = v_j,$$

such that consecutive elements of the sequence are adjacent. The powers of the adjacency matrix of a graph can be used to find the number of walks of length l.

Theorem 2.5. *The number of walks of length l in graph G, from v_i to v_j, is element (i, j) of the matrix $\boldsymbol{A}^l$.*

Proof. We will prove this theorem by induction. For walks of lengths 0 and 1, the result is true, as $\mathbf{A}^0 = \mathbf{I}$ and $\mathbf{A}^1 = \mathbf{A}$. Suppose the result is true for walks of length L. For each walk of length $L+1$ from v_i to v_j, there is a walk of length L from v_i to vertices v_k adjacent to v_j. This bijection implies that the number of such $L+1$ walks is

$$\sum_{\{v_k,v_j\}\in E} (\mathbf{A}^L)_{ik} = \sum_{k=1}^{n} (\mathbf{A}^L)_{ik} a_{kj} = (\mathbf{A}^{L+1})_{ij}$$

Hence, the result is true by induction. □

Definition 2.6. A graph is *connected* if each pair of vertices is joined by a walk.

A directed graph is said to be strongly connected if there is a walk of directed edges between any pair of vertices.

For more on spectral and algebraic graph theory, see Bigg [**1**]. In the next section, we move away from graphs to discuss a property of positive square matrices quintessential to many practical applications, stated in Perron's Theorem.

3. Perron's Theorem

In 1907, Oskar Perron proved a theorem that would play a role in numerous fields of practical mathematics. Many measured interactions that take place in the real world can be described by nonnegative or positive matrices, where each element is nonnegative or positive, and a large number of important models take the form of simple linear iterative processes. Such processes begin with an initial state x^0 and evolve recursively as $x^k = \mathbf{A}^k x^0$; Perron tells us when that iterative process will converge using, interestingly enough, the spectrum of matrix $\mathbf{A}$. In this section, we state Perron's Theorem and lay the foundation for a proof shown in 1959 by Samuel Karlin [**2**].

Theorem 3.1. ***(Perron's Theorem)*** *The eigenvalue of largest absolute value of a positive square matrix* $\boldsymbol{A}$ *(a* $n \times n$ *matrix with strictly positive entries) is positive, and belongs to a positive eigenvector, where all of the vector's elements are positive. All other eigenvectors are smaller in absolute value.*

We call the largest absolute value ρ of the eigenvalues of $\mathbf{A}$ the *spectral radius*. If $\mathbf{A}$ is at least nonnegative, ρ is an eigenvalue belonging to a nonnegative eigenvector, which we will show in the proof of Perron's Theorem. The convergence results of an iterative process follow using the *Jordan canonical form* of the matrix $\mathbf{A}$.

Definition 3.2. The *Jordan canonical form* of the matrix $\mathbf{A}$ is defined as follows: For $\mathbf{A} \in \mathbf{C}^n$, there exists a nonsingular $\mathbf{X} \in \mathbf{C}^n$ such that

$$\mathbf{X}^{-1}\mathbf{A}\mathbf{X} = \mathbf{J} = \operatorname{diag}(\mathbf{J}_1, \mathbf{J}_2, \ldots, \mathbf{J}_m),$$

where the index $m \leq n$ is the number of linearly independent eigenvectors of $\mathbf{A}$ with associated eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$. Each $\mathbf{J}_j$ is either the 1×1 matrix (λ_j) or a matrix

$$\mathbf{J}_j = \begin{pmatrix} \lambda_j & 1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \\ 0 & 0 & \ldots & \lambda_j \end{pmatrix}.$$

Lemma 3.3. *For $\boldsymbol{A} \in \boldsymbol{C}^n$ and any $\epsilon > 0$ there exists a norm on $\boldsymbol{C}^n$ such that, under the induced matrix norm,*

$$||\boldsymbol{A}|| \leq \rho(\boldsymbol{A}) + \epsilon.$$

Proof. We can define a modified Jordan form by

$$\hat{\mathbf{J}} = \mathbf{Y}^{-1}\mathbf{A}\mathbf{Y}, \quad \mathbf{Y} = \mathbf{X}\text{diag}(1, \epsilon, \epsilon^2, \ldots, \epsilon^{n-1}),$$

which has the exact same block-diagonal form as $\mathbf{J}$, except every off-diagonal 1 is replaced by ϵ. Any norm on $\mathbf{C}^n$ induces a matrix norm

$$||\mathbf{A}|| = \sup_{\mathbf{x}\neq 0} \frac{||\mathbf{A}\mathbf{x}||}{||\mathbf{x}||}, \text{ for all } \mathbf{A} \in \mathbf{C}^n.$$

It can straightforwardly be shown that

$$||\mathbf{A}||_\infty = \sup_{\mathbf{x}\neq 0} \frac{||\mathbf{A}\mathbf{x}||_\infty}{||\mathbf{x}||_\infty} = \max_{i=1,\ldots,n} \sum_{j=1}^{n} |a_{ij}|, \text{ for all } \mathbf{A} \in \mathbf{C}^n.$$

Because $||\mathbf{x}|| = ||\mathbf{Y}^{-1}\mathbf{x}||_\infty$ is a norm on $\mathbf{C}^n$, we can obtain the result

$$||\mathbf{A}|| = \sup_{\mathbf{x}\neq 0} \frac{||\mathbf{A}\mathbf{x}||}{||\mathbf{x}||} = \sup_{\mathbf{x}\neq 0} \frac{||\mathbf{Y}^{-1}\mathbf{A}\mathbf{x}||_\infty}{||\mathbf{Y}^{-1}\mathbf{x}||_\infty} = \sup_{\mathbf{y}\neq 0} \frac{||\mathbf{Y}^{-1}\mathbf{A}\mathbf{Y}\mathbf{y}||_\infty}{||\mathbf{y}||_\infty} = ||\hat{\mathbf{J}}||_\infty \leq \rho(\mathbf{A}) + \epsilon.$$

□

For any induced matrix norm, it follows that

$$||\mathbf{A}^k|| \leq ||\mathbf{A}||^k, k \geq 0.$$

Theorem 3.4. *For every initial state $\boldsymbol{x}^0$, an iteration $\boldsymbol{x}^k = \boldsymbol{A}^k x^0$ converges if and only if the spectral radius of $\boldsymbol{A}$ has absolute value $|\rho(\boldsymbol{A})| < 1$.*

Proof. Let $|\rho(\mathbf{A})| < 1$ and choose $\epsilon > 0$ such that $\rho(\mathbf{A}) + \epsilon < 1$. The previous lemma implies that $||\mathbf{A}|| < 1$ under a norm. Hence, because $||\mathbf{A}^k|| \leq |\mathbf{A}||^k, k \geq 0$, it is implied that $\lim_{k\to\infty} \mathbf{A}^k = 0$, and an iteration $\mathbf{x}^k = \mathbf{A}^k\mathbf{x}^0$ converges.

Conversely, suppose that $\rho(\mathbf{A}) + \epsilon \geq 1$ and let λ be some eigenvalue such that $\lambda \geq 1$. If $\mathbf{x} \in \mathbf{C}^n, ||\mathbf{x}|| = 1$ is the corresponding eigenvector to λ, then $\mathbf{A}^k\mathbf{x} = \lambda^k\mathbf{x}, k \geq 0$ and thus

$$1 \leq |\lambda|^k = ||\lambda^k\mathbf{x}|| = ||\mathbf{A}^k\mathbf{x}|| \leq ||\mathbf{A}^k|| \leq ||\mathbf{A}||^k$$

shows that the sequence $\{\mathbf{A}^k\}$ does not converge to 0. Hence by contrapositive, an iteration $x^k = \mathbf{A}^k x^0$ will converge only if $|\rho(\mathbf{A})| < 1$. □

If $\lambda = 1$ is an eigenvalue, its eigenspace has full rank, equal to the multiplicity of the root $\lambda = 1$ in the characteristic equation $\chi(\lambda) = det(\lambda\mathbf{I} - \mathbf{A})$. In other words, when $\mathbf{A} > 0$, the iterative process will converge only when the spectral radius is at most 1. The process converges to 0 when $\rho < 1$. When $\rho = 1$, the process converges to the component of the initial state in the eigenspace spanned by the corresponding eigenvector $\mathbf{v} > 0$.

We now define the resolvent function, and prove a lemma for that function that will be used in the proof of Perron's Theorem.

Definition 3.5. The *resolvent* $\mathbf{R}(\lambda)$ is the function

$$\mathbf{R}(\lambda) = (\lambda - \mathbf{A})^{-1},$$

where λ is the set of eigenvalues of the nonnegative matrix $\mathbf{A}$.

The first resolvent identity **(*)** states $\mathbf{R}(\lambda) - \mathbf{R}(\mu) = -(\lambda - \mu)\mathbf{R}(\lambda)\mathbf{R}(\mu)$, and will be used in the proof of the following lemma.

Lemma 3.6. *The coefficients of the resolvent's Laurent expansion*

$$\boldsymbol{R}(\lambda) = \sum_{k=-h}^{\infty} \boldsymbol{A}_k(\lambda - \lambda_0)^k$$

about $\lambda = \lambda_0$ (where λ_0 is not an eigenvalue of $\boldsymbol{A}$) satisfy

$$\boldsymbol{A}_k \cdot \boldsymbol{A}_m = \begin{cases} -\boldsymbol{A}_{k+m+1} & \text{if } k,\ m \geq 0, \\ 0 & \text{if } k < 0, m \geq 0, \\ \boldsymbol{A}_{k+m+1} & \text{if } k,\ m < 0. \end{cases}$$

Proof. Assume $\lambda_0 = 0$, without loss of generality. By integrating both sides of the Laurent expansion counterclockwise about circular contour C centered at the origin, we get the Cauchy integral formula

$$\mathbf{A}_k = \frac{1}{2\pi i}\int_C \frac{\mathbf{R}(\lambda)}{\lambda^{k+1}} d\lambda$$

Using contours C_1 and C_2 of radius $r_1 < r_2$,

$$\begin{aligned} \mathbf{A}_k \cdot \mathbf{A}_m &= \frac{1}{2\pi i}\int_{C_1} \frac{\mathbf{R}(\lambda)}{\lambda^{k+1}} d\lambda \cdot \frac{1}{2\pi i}\int_{C_2} \frac{\mathbf{R}(\mu)}{\mu^{m+1}} d\mu \\ &= \frac{1}{4\pi^2 i^2}\int_{C_1}\int_{C_2} \frac{\mathbf{R}(\lambda)\mathbf{R}(\mu)}{\lambda^{k+1}\mu^{m+1}} \\ &= -\frac{1}{4\pi^2 i^2}\int_{C_1}\int_{C_2} \frac{\mathbf{R}(\lambda) - \mathbf{R}(\mu)}{\lambda^{k+1}\mu^{m+1}(\lambda - \mu)} \ \textbf{(*)} \\ &= \frac{1}{2\pi i}\int_{C_1} \frac{\mathbf{R}(\lambda)}{\lambda^{k+1}} \left(\frac{1}{2\pi i}\int_{C_2} \frac{d\mu}{\mu^{m+1}(\mu - \lambda)} \right) d\lambda \\ &+ \frac{1}{2\pi i}\int_{C_2} \frac{\mathbf{R}(\mu)}{\mu^{m+1}} \left(\frac{1}{2\pi i}\int_{C_1} \frac{d\lambda}{\lambda^{k+1}(\lambda - \mu)} \right) d\mu \end{aligned}$$

Counterclockwise around counter C: $|z| = r$,

$$\frac{1}{2\pi i}\int_C \frac{dz}{z^n(z - w)} = \begin{cases} 0 & \text{if } n \leq 0 \text{ and } |w| > r, \\ -\frac{1}{w^n} & \text{if } n > 0 \text{ and } |w| > r, \\ \frac{1}{w^n} & \text{if } n \leq 0 \text{ and } |w| < r, \\ 0 & \text{if } n > 0 \text{ and } |w| < r, \end{cases}$$

which can be seen be integrating

$$\frac{1}{z^n(z - w)} = -\frac{1}{wz^n}\left[1 + \frac{z}{w} + \left(\frac{z}{w}\right)^2 + \dots\right].$$

When $k < 0$ and $m \geq 0$, plugging in the the Laurent expansions shows both inner integrals of the result of $\mathbf{A}_k \cdot \mathbf{A}_m$ are 0, so $\mathbf{A}_k\mathbf{A}_m = 0$. If both k and m are negative, the inner integral of the first addend is

$$\frac{1}{2\pi i}\int_{C_2} \frac{d\mu}{\mu^{m+1}(\mu - \lambda)} d\lambda = \frac{1}{\lambda^{m+1}},$$

while the inner integral of the second addend is 0. Hence, by the first addend

$$\mathbf{A}_k \cdot \mathbf{A}_m = \frac{1}{2\pi i} \int_{C_1} \frac{\mathbf{R}(\lambda)}{\lambda^{k+m+2}} d\lambda = \mathbf{A}_{k+m+1}.$$

Similarly, for k and m both positive, it can be shown that $\mathbf{A}_k \cdot \mathbf{A}_m = -\mathbf{A}_{k+m+1}$. □

We can now prove Perron's Theorem.

Proof. First, we will show a nonnegative matrix $\mathbf{A}$ has its spectral radius ρ as an eigenvalue, and the eigenspace of $\lambda = \rho$ contains nonnegative eigenvectors.

The resolvent

$$\mathbf{R}(\lambda) = (\lambda - \mathbf{A})^{-1}$$

is analytic everywhere except at its spectrum (λ). Therefore it must be analytic within the radius $|\lambda| > \rho$, and has at least one singularity on the spectral radius $|\lambda| = \rho$.

The Neumann series is $\sum_{k=0}^{\infty} T^k$ of some operator T. The resolvent's Neumann series

$$\mathbf{R}(\lambda) = \frac{1}{\lambda} + \frac{\mathbf{A}}{\lambda^2} + \frac{\mathbf{A}^2}{\lambda^3} + \dots$$

will converge to the first singularity of $\mathbf{R}(\lambda)$, on the region $|\lambda| > \rho$. If $\lambda = \rho$ is not a singularity, then $\lim_{\lambda \to \rho^+} \mathbf{R}(\lambda)$ is finite. Hence

$$\mathbf{R}(\rho) = \frac{1}{\rho} + \frac{\mathbf{A}}{\rho^2} + \frac{\mathbf{A}^2}{\rho^3} + \dots$$

is finite since each coefficient of $\mathbf{A}^n \geq 0$. However, then the Neumann series converges absolutely and uniformly on the closed region $|\lambda| \geq \rho$, which does not allow the series to be centered about the spectral radius circle.

To show the eigenspace of $\lambda = \rho$ contains nonnegative eigenvectors, let us look at the resolvent's Laurent series expansion about $\lambda = \rho$:

$$\mathbf{R}(\lambda) = \sum_{k=-h}^{\infty} \mathbf{A}_k (\lambda - \rho)^k,$$

where $\mathbf{A}_{-h} \neq 0$. We can see $\mathbf{A}_{-h} \geq 0$ since

$$\mathbf{A}_{-h} = \lim_{\lambda \to \rho^+} (\lambda - \rho)^h \mathbf{R}(\lambda) \geq 0.$$

By Lemma 3.6:

$$\mathbf{A}_k \cdot \mathbf{A}_m = \begin{cases} -\mathbf{A}_{k+m+1} & \text{if k, m} \geq 0, \\ 0 & \text{if k} < 0, m \geq 0, \\ \mathbf{A}_{k+m+1} & \text{if k, m} < 0. \end{cases}$$

and thus

$$\mathbf{R}(\lambda)\mathbf{A}_{-h} = \sum_{k=-h}^{\infty} (\lambda - \rho)^k \mathbf{A}_k \mathbf{A}_{-h} = \frac{\mathbf{A}_{-h}}{\lambda - \rho},$$

in other terms,

$$\mathbf{A}\mathbf{A}_{-h} = \rho \mathbf{A}_{-h}.$$

By this, each nonzero column of $\mathbf{A}_{-h}$ is a nonnegative eigenvector of $\mathbf{A}$, associated with the eigenvalue $\lambda = \rho$.

Now assume $\mathbf{A}$ is positive and $\mathbf{v}$ is a nonnegative eigenvector associated with the spectral radius $\lambda = \rho$. Then $0 < \mathbf{A}\mathbf{v} = \rho \mathbf{v}$, and it follows that both ρ and $\mathbf{v}$ are

positive. In other words, each nonnegative column of $\mathbf{A}_{-h}$ is a positive eigenvector of $\lambda = \rho$ for a positive matrix $\mathbf{A}$. However if $h > 1$, by Lemma 3.6, $\mathbf{A}_{-h}$ is nilpotent such that

$$\mathbf{A}_{-h}^2 = 0;$$

on the other hand, every nonzero column of $\mathbf{A}_{-h} > 0$, a contradiction under these conditions. Hence h must equal 1; and so $\lambda = \rho$ is of order 1.

The nonzero elements of the range of $\mathbf{A}_{-1}$ are eigenvectors of $\mathbf{A}$ belonging to $\lambda = \rho$. Conversely, for any eigenvector $\mathbf{v}$ of $\mathbf{A}$ belonging to $\lambda = \rho$, $\mathbf{A}\mathbf{v} = \rho\mathbf{v}$ and so

$$R(\lambda)\mathbf{v} = \frac{\mathbf{v}}{\lambda - \rho}$$

for all λ near ρ, giving by the uniqueness of the Laurent expansion at $\mathbf{v}$ that

$$\mathbf{A}_{-1}\mathbf{v} = \mathbf{v}.$$

Thus the column space of $\mathbf{A}_{-1}$ is the eigenspace of $\lambda = \rho$.

If $\mathbf{v} = (\rho - \mathbf{A})\mathbf{x} \neq 0$ is in the eigenspace belonging to $\lambda = \rho$, by the previous equation and $\mathbf{A}\mathbf{A}_{-h} = \rho\mathbf{A}_{-h}$, $\mathbf{A}_{-1}\mathbf{v} = \mathbf{v} = \mathbf{A}_{-1}(\rho - \mathbf{A})\mathbf{x} = (\rho - \mathbf{A})\mathbf{A}_{-1}\mathbf{x} = 0$. However, $\mathbf{v} \neq 0$, so the operator $\mathbf{A}$ cannot be cyclic at $\lambda = \rho$. Thus $\lambda = \rho$ is a simple eigenvalue of $\mathbf{A}$.

Finally, we will show there are no other eigenvalues on the spectral radius circle. For an arbitrarily small $\epsilon > 0$, consider $\mathbf{A} - \epsilon\mathbf{I} > 0$. Its largest positive eigenvalue is $\rho - \epsilon$, and because $\mathbf{A}$ is nonnegative, that eigenvalue is its spectral radius. Now translate this small circle back to the right by ϵ, and we see that all the other eigenvalues of $\mathbf{A}$ are within the open radius $|\lambda| < \rho$.

In summary, the dominant eigenvalue of a positive matrix is positive and belongs to a positive eigenvector. All other eigenvalues are smaller in modulus. And we are done. □

A more detailed version of the proof, with a discussion of Perron's previous proofs can be found in MacCluer [**2**]. With the proof of Perron's theorem complete, we can now enter a discussion of some of its many applications, in the next section.

4. Some Applications of Perron's Theorem

In this section, we look at two real-world applications of Perron's Theorem, an example in ecology and an example in economics. This will provide an idea of how the spectral radius of nonnegative matrices can be used as a parameter in different linear iterative processes. First, let's look at an application to a simple population model.

Ecological Population Model. First, divide a population into n groups of increasing age, by decades. Let x_i^j be the number of people in the ith age group recorded at the jth time interval. The $(j+1)$th interval will reveal that a fraction of each age group advances into the next, as illustrated by

$$\begin{vmatrix} x_1^{j+1} \\ . \\ . \\ . \\ x_n^{j+1} \end{vmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & . & . & . & a_{1N} \\ a_{21} & 0 & 0 & 0 & . & . & 0 \\ 0 & a_{32} & 0 & . & . & . & . \\ & & & & . & . & 0 \\ 0 & . & . & 0 & . & a_{n,n-1} & 0 \end{pmatrix} \begin{vmatrix} x_1^j \\ . \\ . \\ . \\ x_n^j \end{vmatrix},$$

where the first row of $\mathbf{A} = (a_{ij})$ shows the contribution of each age group to reproduction.

Perron's Theorem reveals that if the dominant nonnegative eigenvalue $\rho > 1$, initial populations can grow without bound. If $\rho < 1$, every population becomes extinct. If $\rho = 1$, depending on the Jordan form of $\mathbf{A}$, a population may oscillate (when there are complex eigenvalues $|\lambda| = 1$), may extinguish or explode, or may converge to a steady state distribution in the eigenspace of $\rho = 1$.

Walrasian Stability of Competitive Markets. Suppose n similar commodities are competing for the consumer's dollar. We approximate the excess demand f_i for commodity i (in the case of high demand, less supply) as a linear function of prices p minus equilibrium prices p^0:

$$f_i \approx \sum_{j=1}^{n} a_{ij}(p_j - p_j^0).$$

Higher prices for one commodity will increase excess demand for the others, so $a_{ij} \geq 0$ for $i \neq j$, and $a_{ii} < 0$.

Once disturbed from equilibrium, the rate of price readjustment must be proportional to excess demand:

$$\dot{p} = \mathbf{KA}(p - p^0),$$

where $\mathbf{K} = \text{diag}(k_1, k_2, \ldots, k_n)$, the diagonal matrix of positive adjustment speeds. Hence, future prices are given by

$$p = p^0 + e^{\mathbf{KA}t}c,$$

where $c = p(0) = p^0$.

The behavior of the price readjustment system is dependent on the eigenvalues of $\mathbf{M} = \mathbf{KA}$. If all eigenvalues lie in the open left half-plane $\text{Re}(s) < 0$, prices return to equilibrium. If some eigenvalues are on the imaginary axis, prices can be exploding or oscillating. If eigenvalues lie in the open right half-plane, some prices may skyrocket.

We define a Metzler matrix as one with nonnegative entries off the diagonal. The matrix $\mathbf{M} = \mathbf{KA}$ is Metzler, and is governed by the following property:

Theorem 4.1. *The rightmost eigenvalue of $\boldsymbol{M}$ is real; all others are strictly to its left.*

Proof. Translate the spectrum of $\mathbf{M}$ to the right by adding an arbitrarily small multiple of the identity, such that $\mathbf{M}+\epsilon\mathbf{I} \geq 0$. By Perron's Theorem, this translated matrix has real dominant eigenvalue $\rho \geq 0$. By translating back, we find that $\mathbf{M}$ must have real eigenvalue $\beta = \rho - \epsilon$, and all other eigenvalues must lie within a radius ρ to the left, tangent to the vertical line $\text{Im}(s) = \beta$ at β. □

Hence, price stability is governed by the real eigenvalue β of $\mathbf{M}$.

These two examples should provide an idea for how the spectrum can dictate properties in real-world applications. In the next section, we return to analyzing matrices of a graph in a particularly revolutionary application of Perron's Theorem to the world's largest directed graph, the Internet.

5. The Power Method and Markov Chains

A common theme in the Perron applications is that the dominant eigenvalue of matrix $\mathbf{A}$ and respective dominant eigenvector must actually be calculated. For this we generally use the power method to approximate the dominant eigenvector.

Definition 5.1. The *power method* is an iterative process of approximating the dominant eigenvalue and respective eigenvector of matrix $\mathbf{A}$. First, choose an initial nonzero vector approximation $\mathbf{v}_0$ in R^n as one of the dominant eigenvectors of $\mathbf{A}$. Form the sequence given by

$$\mathbf{v}_1 = \mathbf{A}\mathbf{v}_0$$
$$\mathbf{v}_2 = \mathbf{A}\mathbf{v}_1 = \mathbf{A}(\mathbf{A}\mathbf{v}_0) = \mathbf{A}^2\mathbf{v}_0$$
$$\mathbf{v}_3 = \mathbf{A}\mathbf{v}_2 = \mathbf{A}(\mathbf{A}^2\mathbf{v}_0) = \mathbf{A}^3\mathbf{v}_0$$
$$\dots$$
$$\mathbf{v}_k = \mathbf{A}\mathbf{v}_{k-1} = \mathbf{A}(\mathbf{A}^{k-1}\mathbf{v}_0) = \mathbf{A}^k\mathbf{v}_0$$

For large powers of k, and proper scaling, $\mathbf{v}_k$ becomes a good approximation of the dominant eigenvector of $\mathbf{A}$. From this eigenvector we can calculate the dominant eigenvalue.

Theorem 5.2. *If v is an eigenvector of matrix A, then its corresponding eigenvalue is given by*

$$\lambda = \frac{\boldsymbol{Av} \cdot \boldsymbol{v}}{\boldsymbol{v} \cdot \boldsymbol{v}}.$$

Proof. Since $\mathbf{v}$ is an eigenvector of $\mathbf{A}$, we know $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ and thus

$$\frac{\mathbf{A}\mathbf{v} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}} = \frac{\lambda\mathbf{v} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}} = \frac{\lambda(\mathbf{v} \cdot \mathbf{v})}{\mathbf{v} \cdot \mathbf{v}} = \lambda.$$

□

Calculating a result by means of repeated operations comes up in many mathematical and computing applications. A particular field of mathematics modeling probabilistic behavior across an iterative process called Markov Chain Theory helps describe many applications of Perron's theorem. To discuss the Google PageRank algorithm, we need some definitions related to Markov chain processes, iterative processes where future states only depend on the current state and a probability distribution. The PageRank vector is a stationary probability distribution vector of a discrete-time, finite-state Markov chain, so it is crucial to first understand some properties of Markov chains. We begin by defining the types of matrices relevant to Markov chains.

Definition 5.3. A square matrix $\mathbf{A}$ with nonnegative entries is *irreducible* if and only if its directed graph is strongly connected. In other words, $\mathbf{A}$ is irreducible if there is a walk of directed edges between any pair of vertices in the directed graph.

Definition 5.4. A matrix $\mathbf{A}$ is *primitive* when $\mathbf{A}$ is a nonnegative irreducible matrix that only has one eigenvalue $\lambda = \rho$ on its spectral radius.

Definition 5.5. A matrix $\mathbf{A}$ is *stochastic* if each row sum is equal to 1. In some cases, we describe these matrices as *row-stochastic*, and matrices whose column sum is equal to 1 to be *column-stochastic*.

Theorem 5.6. *Every stochastic matrix* $\boldsymbol{A}$ *has* $\lambda = 1$ *as an eigenvalue, and this eigenvalue is its dominant eigenvalue.*

Proof. Let $\mathbf{A}$ be a $n \times n$ column-stochastic matrix and $\mathbf{e}$ be an n-dimensional column vector with entries of all 1's. Because $\mathbf{A}$ is column-stochastic, $\mathbf{A}^T\mathbf{e} = \mathbf{e}$, so 1 is an eigenvalue for $\mathbf{A}^T$. Since $\mathbf{A}$ and $\mathbf{A}^T$ have the same eigenvalues, 1 is also an eigenvalue for $\mathbf{A}$.

Suppose $\mathbf{Ax} = \lambda\mathbf{x}$ for some $\lambda > 1$. Since the columns of $\mathbf{A}$ are nonnegative and sum to 1, each element of $\mathbf{A}x$ is a convex combination of the components of $\mathbf{x}$, which can be no greater than the largest component of $\mathbf{x}$. However, because $\lambda > 1$, at least one element of $\lambda\mathbf{x}$ is greater than the largest component of $\mathbf{x}$. Hence, $\lambda > 1$ is impossible, and $\lambda = 1$ is the largest eigenvalue for $\mathbf{A}$.

A similar proof can show that row-stochastic matrices also have 1 has a dominant eigenvalue. □

This property of stochastic matrices will become useful when we apply Perron's Theorem. Now we can begin our discussion of Markov chain theory.

Definition 5.7. A *stochastic process* is a set of random variables $\{X_t\}_{t=0}^{\infty}$ with a common range $\{S_1, S_2, \ldots, S_n\}$, known as the *state space* for the process. The t parameter is generally thought of as time, so X_t can be thought of as the state of the process at time t.

We use the phrase discrete-time when the process considers time as discrete rather than continuous. The phrase finite-state describes the state space as finite rather than infinite. In this discussion, we only use discrete-time finite-state processes.

In the next definition, the notation $P(E|F)$ refers to the conditional probability that event E occurs given event F occurs.

Definition 5.8. A *Markov chain* is a stochastic process that satisfies the property

$$P(X_{t+1} = S_j | X_t = S_{i_t}, X_{t-1} = S_{i_{t-1}}, \ldots, X_0 = S_{i_0}) = P(X_{t+1} = S_j | X_t = S_{i_t})$$

for each $t = 0, 1, 2, \ldots$

Definition 5.9. The *transition probability* $p_{ij}(t) = P(X_t = S_j | X_{t-1} = S_i)$ is the probability of being in state S_j at time t, given that the chain is in state S_i at time $t-1$. The *transition probability matrix* $\mathbf{P}_{n\times n}(t) = (p_{ij}(t))$ is a nonnegative stochastic matrix for each t.

A stationary Markov chain is a chain in which the transition probabilities do not vary with time, such that $p_{ij}(t) = p_{ij}$ for all t. In a stationary Markov chain, the transition probability matrix is a constant stochastic matrix $\mathbf{P} = (p_{ij})$. An irreducible Markov Chain is a chain for which the transition probability matrix $\mathbf{P}$ is an irreducible matrix. An aperiodic Markov chain is an irreducible chain whose transition probability matrix $\mathbf{P}$ is a primitive matrix.

Definition 5.10. A *probability distribution vector* is a nonnegative row vector $\mathbf{p}^T = (p_1, p_2, \ldots, p_n)$ such that $\sum_k p_k = 1$.

Hence, every row in a stochastic matrix is a probability vector. For a Markov chain with a transition probability matrix $\mathbf{P}$, a stationary probability distribution vector is a probability vector π^T such that $\pi^T\mathbf{P} = \pi^T$. The initial distribution vector for an n-state chain is $\mathbf{p}^T(0) = (p_1(0), p_2(0), \ldots, p_n(0))$, where $p_j(0) =$

$P(X_0 = S_j)$. In other words, the initial distribution contains the probabilities that the chain starts in each state. The kth step probability distribution vector is $\mathbf{p}^T(k) = (p_1(k), p_2(k), \ldots, p_n(k))$, where $p_j(k) = P(X_k = S_j)$. $p_j(k)$ is the probability of being in the jth state after the kth step.

How can we know if a Markov chain provides us with a stationary distribution vector? Let $\mathbf{P}$ be the transition probability matrix for an irreducible Markov Chain on states $\{S_1, S_2, \ldots, S_n\}$, and let there be vector π^T such that $\pi^T\mathbf{P} = \pi^T, |\pi|_1 = 1$. For every initial distribution $\mathbf{p}^T(0)$, we know the kth step transition matrix is $\mathbf{P}^k$ and the kth step probability distribution vector is $\mathbf{p}^T(k) = \mathbf{p}^T(0)\mathbf{P}^k$. If $\mathbf{P}$ is primitive such that the chain is aperiodic, and $\mathbf{e}$ is a column vector of all 1's, then

$$\lim_{k\to\infty} \mathbf{P}^k = \mathbf{e}\pi^T \text{ and } \lim_{k\to\infty} \mathbf{p}^T(k) = \pi^T.$$

Thus, π^T is the unique stationary probability distribution vector because it satisfies $\pi^T\mathbf{P} = \pi^T$.

This concludes the necessary definitions for describing the matrices for the calculation of the PageRank vector. For a more thorough introduction to Markov chains, see Tolver [**3**].

6. The Mathematics of Google's PageRank

This section describes one of the most interesting and influential applications of Perron's Theorem to our modern lives: Googling. We describe how the ranking system works and the rationale behind each of the algorithm's matrices. We conclude with a proof of the Google matrix's spectrum, which tells us how many iterations of the algorithm are necessary to accurately rank the web pages. The majority of this section uses Langville and Meyer's book [**4**], so for more information on the algorithm, its speed modifications, its sensitivity, and related ranking algorithms please see their book.

Search engines like Google do three basic things. First, they crawl through the Web and locate all web pages available publicly. Next, they index the data so it can be searched for efficiently using keywords and phrases. Lastly, they rank the importance of each page on the Web, so that when a user searches and a subset of the Web with desired information is found, it is displayed in order of importance. The model for the World Wide Web can hence be described as a directed graph, where the nodes are web pages and the edges represent links that connect one page to another. Ranking the "importance" of a node can be thought of as manipulating the adjacency matrix of the Web such that desired nodes are placed first in a ranking vector.

Page and Brin first used a simple summation equation to rank pages on the Web, found from bibliometrics research (the analysis of citations among academic papers). The PageRank $r(P_i)$ of a page P_i is the sum of the PageRanks of all pages directed to P_i:

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

where B_{P_i} is the set of pages directed to P_i and $|P_j|$ is the number of directed links pointing outwards from page P_j. The problem is, however, in this equation the PageRanks of pages directed to P_i, $r(P_j)$, is unknown. Instead, we define an iterative process, where we initially give each page an equal PageRank of $\frac{1}{n}$, where

n is the number of pages on the Web. With an initial state of $r_0(P_i) = \frac{1}{n}$, the PageRank of page P_i at iteration $k+1$, $r_{k+1}(P_i)$, becomes

$$r_{k+1}(P_i) = \sum_{P_j \in B_{P_i}} \frac{r_k(P_j)}{|P_j|}.$$

This process of ranking pages only computes one rank at a time, so we would rather compute a single $1 \times n$ PageRank vector π^T that holds PageRanks for all pages. To hold data for the Web's directed graph, we describe a hyperlink matrix $\mathbf{H}$.

Definition 6.1. The *hyperlink matrix* $\mathbf{H}$ is an $n \times n$ row-normalized matrix such that

$$H_{ij} = \begin{cases} \frac{1}{|P_i|} & \text{if there is a directed edge from page } i \text{ to page } j \\ 0 & \text{otherwise} \end{cases}$$

Notice that the hyperlink matrix is similar in form to the previously discussed adjacency matrix for a graph, though the entries of the hyperlink matrix are probabilities. $\mathbf{H}$ appears to be more like the stochastic transition probability matrix for a Markov chain; the primary difference is that $\mathbf{H}$ contains dangling nodes, or pages with no directed edges pointing outwards, which appear as rows of 0 in $\mathbf{H}$. All other rows are stochastic, so we call $\mathbf{H}$ substochastic. In later modifications of the hyperlink matrix, probabilities are not uniform for all webpages, but rather dependent on the individual user's search logs and previous page visits to maximize the search's accuracy.

With these new matrices, we can modify our iterative process:

$$\pi^{(k+1)T} = \pi^{(k)T}\mathbf{H}.$$

The original iterative process began with initial distribution vector $\pi^{(0)T} = 1/n\mathbf{e}^T$, where $\mathbf{e}^T$ is a row vector of all 1s. This initial vector was later replaced by $\mathbf{v}^T > 0$, a probability distribution vector called the personalization vector. Using this vector in place of $1/n\mathbf{e}^T$ means probabilities are no longer uniformly distributed among pages, and are dependent on the user's preferences. Thus, a random Web surfer can jump to different pages based on his/her preferences, rather than being equally likely to jump to any page.

Because $\mathbf{H}$ is substochastic, the user/Markov chain is stuck when it reaches a dangling node. To fix this we make a stochasticity adjustment by replacing the $\mathbf{0}^T$ rows of $\mathbf{H}$ with $\mathbf{v}^T$, defining a new stochastic matrix $\mathbf{S}$ and a dangling node vector $\mathbf{a}$.

Definition 6.2. The *stochastic matrix* $\mathbf{S}$ is defined as $\mathbf{S} = \mathbf{H} + \mathbf{a}(\mathbf{v}^T)$, where $a_i = 1$ if page i is a dangling node and 0 otherwise.

Now that we have a stochastic matrix to describe the probability adjacency of pages in the Web, we can use $\mathbf{S}$ as a transition probability matrix for a Markov chain. However, this matrix alone does not guarantee that the PageRank vector π^T will converge. To guarantee a stationary distribution vector of the chain exists, we need to make a primitivity adjustment. With a primitive matrix, we make the Markov chain both irreducible and aperiodic, which fulfills the necessary conditions to find a stationary distribution vector. We now define the Google matrix $\mathbf{G}$.

Definition 6.3. The *Google matrix* $\mathbf{G}$ is defined as $\mathbf{G} = \alpha\mathbf{S} + (1-\alpha)\mathbf{e}\mathbf{v}^T$, where α is a scalar from 0 to 1.

The parameter α reflects the likelihood that a random Web surfer follows hyperlinks in the order of directed edges, rather than jumping, which is reflected by $\mathbf{e}\mathbf{v}^T$. This α greatly affects the number of iterations it takes to calculate the PageRank vector, as a value close to 1 is much more computationally intensive than 0.5. It also affects the sensitivity of the ranking itself, so for a generally stable ranking, α is set to 0.85. We will show the rationale for this in a later theorem.

With this primitivity adjustment, our Google matrix $\mathbf{G}$ is now stochastic, irreducible, aperiodic, and primitive, and can be used by our iterative process to find a stationary probability PageRank vector π^T. Now we see how Perron's Theorem is applicable; following from the fact our matrix $\mathbf{G}$ is stochastic, the dominant eigenvalue is $\lambda_1 = 1$, ensuring that the iterative process converges. Our adjusted iterative process is now

$$\pi^{(k+1)T} = \pi^{(k)T}\mathbf{G}$$

To calculate our PageRank vector π^T, we must solve the following problem:

$$\pi^T = \pi^T\mathbf{G},$$

$$\pi^T\mathbf{e} = 1.$$

The goal is to find the normalized dominant eigenvector of $\mathbf{G}$ corresponding to the dominant eigenvalue $\lambda_1 = 1$. The second equation ensures that π^T is a probability vector. Hence, we can now calculate π^T using the power method, made simpler when we express $\mathbf{G}$ in terms of $\mathbf{H}$:

$$\pi^{(k+1)^T} = \pi^{(k)T}\mathbf{G}$$

$$= \alpha\pi^{(k)T}\mathbf{S} + \frac{1-\alpha}{n}\pi^{(k)T}\mathbf{e}\mathbf{v}^T$$

$$= \alpha\pi^{(k)T}\mathbf{H} + (\alpha\pi^{(k)T}\mathbf{a} + 1 - \alpha)\mathbf{v}^T$$

Each vector-matrix multiplication is $O(n)$ since $\mathbf{H}$ has only about 10 nonzeros per row. The power method is matrix-free and storage-friendly, and is also beneficial because only 50 - 100 iterations are necessary before the process converges, and a satisfactory approximation of the exact PageRank vector emerges. How do we know when the power method converges?

Definition 6.4. The *asymptotic rate of convergence* of the power method applied to a matrix is the rate at which $|\lambda_2/\lambda_1|^k \to 0$, for k iterations of the power method.

Since $\mathbf{G}$ is stochastic and primitive, $\lambda_1 = 1$ and $|\lambda_2| < 1$. Hence, although the largest eigenvalue tells us the iteration converges by Perron's Theorem, it is the second largest eigenvalue that governs the rate of convergence. We will finally prove the Google spectrum theorem, which will tell us after how many iterations the power method converges.

Theorem 6.5. ***(The Google Spectrum Theorem)*** *If the spectrum of the stochastic matrix* $\boldsymbol{S}$ *is* $\{1, \lambda_2, \lambda_3, \ldots, \lambda_n\}$*, then the spectrum of the Google matrix* $\boldsymbol{G} = \alpha\boldsymbol{S} + (1-\alpha)\boldsymbol{e}\boldsymbol{v}^T$ *is* $\{1, \alpha\lambda_2, \alpha\lambda_3, \ldots, \alpha\lambda_n\}$.

Proof. Since $\mathbf{S}$ is stochastic, $(1, \mathbf{e})$ is an eigenpair of $\mathbf{S}$. Let $\mathbf{Q} = (\mathbf{e}\ \mathbf{X})$ be a nonsingular matrix that has the eigenvector $\mathbf{e}$ as its first column. Let $\mathbf{Q}^{-1} = \begin{pmatrix} \mathbf{y}^T \\ \mathbf{Y}^T \end{pmatrix}$. Hence, $\mathbf{Q}^{-1}\mathbf{Q} = \begin{pmatrix} \mathbf{y}^T\mathbf{e} & \mathbf{y}^T\mathbf{X} \\ \mathbf{Y}^T\mathbf{e} & \mathbf{Y}^T\mathbf{X} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$. This gives us two identities: $\mathbf{y}^T\mathbf{e} = 1$ and $\mathbf{Y}^T\mathbf{e} = \mathbf{0}$. Hence, when the similarity transformation is applied to $\mathbf{S}$, we get

$$\mathbf{Q}^{-1}\mathbf{S}\mathbf{Q} = \begin{pmatrix} \mathbf{y}^T\mathbf{e} & \mathbf{y}^T\mathbf{S}\mathbf{X} \\ \mathbf{Y}^T\mathbf{e} & \mathbf{Y}^T\mathbf{S}\mathbf{X} \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{y}^T\mathbf{S}\mathbf{X} \\ 0 & \mathbf{Y}^T\mathbf{S}\mathbf{X} \end{pmatrix}.$$

This tells us that $\mathbf{Y}^T\mathbf{S}\mathbf{X}$ contains the other eigenvalues $\mathbf{S}$, $\{\lambda_2, \ldots, \lambda_n\}$. Thus, when we apply the similarity transformation to $\mathbf{G} = \alpha\mathbf{S} + (1-\alpha)\mathbf{e}\mathbf{v}^T$, we get

$$\begin{aligned} \mathbf{Q}^{-1}(\alpha\mathbf{S} + (1-\alpha)\mathbf{e}\mathbf{v}^T)\mathbf{Q} &= \alpha\mathbf{Q}^{-1}\mathbf{S}\mathbf{Q} + (1-\alpha)\mathbf{Q}^{-1}\mathbf{e}\mathbf{v}^T\mathbf{Q} \\ &= \begin{pmatrix} \alpha & \alpha\mathbf{y}^T\mathbf{S}\mathbf{X} \\ 0 & \alpha\mathbf{Y}^T\mathbf{S}\mathbf{X} \end{pmatrix} + (1-\alpha)\begin{pmatrix} \mathbf{y}^T\mathbf{e} \\ \mathbf{Y}^T\mathbf{e} \end{pmatrix}\begin{pmatrix} \mathbf{v}^T\mathbf{e} & \mathbf{v}^T\mathbf{X} \end{pmatrix} \\ &= \begin{pmatrix} \alpha & \alpha\mathbf{y}^T\mathbf{S}\mathbf{X} \\ 0 & \alpha\mathbf{Y}^T\mathbf{S}\mathbf{X} \end{pmatrix} + \begin{pmatrix} (1-\alpha) & (1-\alpha)\mathbf{v}^T\mathbf{X} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \\ &= \begin{pmatrix} 1 & \alpha\mathbf{y}^T\mathbf{S}\mathbf{X} + (1-\alpha)\mathbf{v}^T\mathbf{X} \\ \mathbf{0} & \alpha\mathbf{Y}^T\mathbf{S}\mathbf{X} \end{pmatrix} \end{aligned}$$

Thus, the spectrum of $\mathbf{G} = \alpha\mathbf{S} + (1-\alpha)\mathbf{e}\mathbf{v}^T$ is $\{1, \alpha\lambda_2, \lambda_3, \ldots, \alpha\lambda_n\}$. □

This theorem tells us that the second eigenvalue of the Google matrix $\mathbf{G}$ is $\alpha\mu_2$, where μ_2 is the second eigenvalue of the stochastic matrix $\mathbf{S}$. The structure of the Web makes it highly likely that $|\mu_2| = 1$ or $|\mu_2| \approx 1$, which means that $|\lambda_2(\mathbf{G})| = \alpha$ or $|\lambda_2(\mathbf{G})| \approx \alpha$. Therefore, the asymptotic rate of convergence of the power method when finding the PageRank vector is (or approximately) the rate at which $\alpha^k \to 0$. The value used by Google is $\alpha = 0.85$ at the time of this paper, so at 50 iterations $\alpha^{50} = 0.85^{50} \approx .0003$, which implies 2-3 places of accuracy for the approximate PageRank vector. Thus, the second element of the Google spectrum dictates the number of iterations of the power method for finding the PageRank vector, for a given accuracy.

Acknowledgments. It is a pleasure to thank my mentor, Kevin Casto, for introducing me to spectral graph theory and describing its relevance to Google's PageRank. Without his guidance and support, I would not have had the idea for the paper, or the resources to write it. I would also like to thank Peter May for hosting the Mathematics Research Experiences for Undergraduates (REU) at the University of Chicago this year, László Babai for introducing me to the wonders of Linear Algebra, and the entire University staff for making accommodations for the program.

References

[1] Norman Biggs. Algebraic Graph Theory, 2nd Edition. Cambridge University Press. 1993.
[2] C. R. MacCluer. The Many Proofs and Applications of Perron's Theorem. SIAM Review, Vol. 42, No. 3, p. 487-498. 2000.
[3] Anders Tolver. An Introduction to Markov Chains. University of Copenhagen. 2016.
[4] Amy N. Langville and Carl D. Meyer. Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press. 2006.