Lecture 10: Fundamental Theorem of Markov Chains, Page Rank

Rafael Oliveira

University of Waterloo
Cheriton School of Computer Science

rafael.oliveira.teaching@gmail.com

October 7, 2021

Overview

- Main Tools
  - Linear Algebra Background
  - Perron-Frobenius
- Main Applications
  - Fundamental Theorem of Markov Chains
  - Page Rank
- Conclusion
- Acknowledgements

Eigenvalues, Eigenvectors and Spectral Radius

- Given a square matrix $A \in \mathbb{R}^{n \times n}$, we say that $\lambda \in \mathbb{C}$ is an *eigenvalue* of A if there is a vector $v \in \mathbb{C}^n$ such that $Av = \lambda v$.

v is an eigenvector corresponding to λ

Eigenvalues, Eigenvectors and Spectral Radius

- Given a square matrix $A \in \mathbb{R}^{n \times n}$, we say that $\lambda \in \mathbb{C}$ is an *eigenvalue* of A if there is a vector $v \in \mathbb{C}^n$ such that $Av = \lambda v$.
- The *spectral radius* of a matrix A, denoted $\rho(A)$, is the maximum absolute value of the eigenvalues of A

$$\rho(A) = \max_{\lambda \text{ eigenvalue of } A} |\lambda|$$

Eigenvalues, Eigenvectors and Spectral Radius

- Given a square matrix $A \in \mathbb{R}^{n \times n}$, we say that $\lambda \in \mathbb{C}$ is an *eigenvalue* of A if there is a vector $v \in \mathbb{C}^n$ such that $Av = \lambda v$.
- The *spectral radius* of a matrix A, denoted $\rho(A)$, is the maximum absolute value of the eigenvalues of A
- *Gelfand's formula*

$$\rho(A) = \lim_{n \to \infty} \|A^n\|_F^{1/n}$$

$$\|A\|_F = \mathrm{tr}[A^T A]^{1/2}$$

Eigenvalues, Eigenvectors and Spectral Radius

- Given a square matrix $A \in \mathbb{R}^{n\times n}$, we say that $\lambda \in \mathbb{C}$ is an *eigenvalue* of A if there is a vector $v \in \mathbb{C}^n$ such that $Av = \lambda v$.
- The *spectral radius* of a matrix A, denoted $\rho(A)$, is the maximum absolute value of the eigenvalues of A
- *Gelfand's formula*

$$\rho(A) = \lim_{n\to\infty} \|A^n\|_F^{1/n}$$

- *Geometric multiplicity*: an eigenvalue λ of A has geometric multiplicity k if the space of eigenvectors of A with eigenvalue λ has dimension k. That is, if dimension of null space of $A - \lambda I$ is k.

$$A = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 3 \end{pmatrix}$$

geometric multiplicity of 1 is 2

$\alpha e_1 + \beta e_2$ is eigenvector of 1

Eigenvalues, Eigenvectors and Spectral Radius

- Given a square matrix $A \in \mathbb{R}^{n\times n}$, we say that $\lambda \in \mathbb{C}$ is an *eigenvalue* of A if there is a vector $v \in \mathbb{C}^n$ such that $Av = \lambda v$.
- The *spectral radius* of a matrix A, denoted $\rho(A)$, is the maximum absolute value of the eigenvalues of A
- *Gelfand's formula*

$$\rho(A) = \lim_{n\to\infty} \|A^n\|_F^{1/n}$$

- *Geometric multiplicity*: an eigenvalue λ of A has geometric multiplicity k if the space of eigenvectors of A with eigenvalue λ has dimension k. That is, if dimension of null space of $A - \lambda I$ is k.
- *Algebraic multiplicity*: an eigenvalue λ of A has algebraic multiplicity k if $(t-\lambda)^k$ is the highest power of $t-\lambda$ dividing $\det(tI - A)$ $= p(t)$

characteristic polynomial of A

$A = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 3 \end{pmatrix}$ $\det(tI - A) = (t-1)^2(t-3)$

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

$$\det(tI - A) = (t-1)^2 \nearrow \text{algebraic mult. of } 1 \text{ is } 2$$

A only has 1 eigenvector : e_1 (eigenvalue is 1)

$\therefore$ geometric multiplicity is 1.

$$\begin{pmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}$$

Eigenvalues, Eigenvectors and Spectral Radius

- Given a square matrix $A \in \mathbb{R}^{n\times n}$, we say that $\lambda \in \mathbb{C}$ is an *eigenvalue* of A if there is a vector $v \in \mathbb{C}^n$ such that $Av = \lambda v$.
- The *spectral radius* of a matrix A, denoted $\rho(A)$, is the maximum absolute value of the eigenvalues of A
- *Gelfand's formula*
$$\rho(A) = \lim_{n\to\infty} \|A^n\|_F^{1/n}$$
- *Geometric multiplicity*: an eigenvalue λ of A has geometric multiplicity k if the space of eigenvectors of A with eigenvalue λ has dimension k. That is, if dimension of null space of $A - \lambda I$ is k.
- *Algebraic multiplicity*: an eigenvalue λ of A has algebraic multiplicity k if $(t-\lambda)^k$ is the highest power of $t - \lambda$ dividing $\det(tI - A)$
- Example:

Positivity Lemma

Lemma (Positivity Lemma)

If $A \in \mathbb{R}^{n \times n}$ is a positive matrix and $u, v \in \mathbb{R}^n$ are distinct vectors such that $u \geq v$, then $Au > Av$. Moreover, there exists $\varepsilon > 0$ such that $Au > (1 + \varepsilon)Av$.

A is positive matrix if every entry

$A_{ij} > 0$.

$u \geq v \iff$ each coordinate $u_i \geq v_i$

Positivity Lemma

Lemma (Positivity Lemma)

If $A \in \mathbb{R}^{n \times n}$ is a positive matrix and $u, v \in \mathbb{R}^n$ are distinct vectors such that $u \geq v$, then $Au > Av$. Moreover, there exists $\varepsilon > 0$ such that $Au > (1 + \varepsilon)Av$.

- Note that

$$A(u - v) = \sum_{j} A_{ij}(u_j - v_j) \geq (\min_{i,j} A_{ij}) \cdot \sum_{j} (u_j - v_j)$$

Positivity Lemma

Lemma (Positivity Lemma)

If $A \in \mathbb{R}^{n\times n}$ is a positive matrix and $u, v \in \mathbb{R}^n$ are distinct vectors such that $u \geq v$, then $Au > Av$. Moreover, there exists $\varepsilon > 0$ such that $Au > (1+\varepsilon)Av$.

- Note that

$$\to \left[A(u-v)\right]_i = \sum_j A_{ij}(u_j - v_j) \geq (\min_{i,j} A_{ij}) \cdot \sum_j (u_j - v_j)$$

$> \min A_{ij} \cdot \mu > 0$

- Since $u_j \geq v_j$ for all j and u, v distinct implies that there is one index k such that $u_k > v_k$, we have

$$\sum_j (u_j - v_j) \geq u_k - v_k > 0$$

$$\left[A(u-v)\right]_i > \mu \cdot \min_{i,j} A_{ij} > 0$$

Positivity Lemma

Lemma (Positivity Lemma)

If $A \in \mathbb{R}^{n\times n}$ is a positive matrix and $u, v \in \mathbb{R}^n$ are distinct vectors such that $u \geq v$, then $Au > Av$. Moreover, there exists $\varepsilon > 0$ such that $Au > (1+\varepsilon)Av$.

- Note that

$$A(u-v) = \sum_{i,j} A_{ij}(u_j - v_j) \geq n \cdot (\min_{i,j} A_{ij}) \cdot \sum_j (u_j - v_j)$$

- Since $u_j \geq v_j$ for all j and u, v distinct implies that there is one index k such that $u_k > v_k$, we have

$$\sum_j (u_j - v_j) \geq u_k - v_k > 0$$

- the moreover part just follows from taking small enough ε.

Perron's Theorem

Theorem (Perron's Theorem)

Let $A \in \mathbb{R}^{n \times n}$ be a positive matrix (i.e., all its entries are positive). Then, the following hold:

1. $\rho(A)$ *is an eigenvalue, and it has a positive eigenvector*
2. $\rho(A)$ *is the only eigenvalue in the complex circle* $|\lambda| = \rho(A)$
3. $\rho(A)$ *has geometric multiplicity 1*
4. $\rho(A)$ *has algebraic multiplicity 1*

(2) says that $\rho(A)$ is <u>only</u> eigenvalue of <u>max</u> absolute value

Perron's Theorem

Theorem (Perron's Theorem)

Let $A \in \mathbb{R}^{n \times n}$ be a positive matrix (i.e., all its entries are positive). Then, the following hold:

1. *$\rho(A)$ is an eigenvalue, and it has a positive eigenvector*
2. *$\rho(A)$ is the only eigenvalue in the complex circle $|\lambda| = \rho(A)$*
3. *$\rho(A)$ has geometric multiplicity 1*
4. *$\rho(A)$ has algebraic multiplicity 1*

- By the definition of $\rho(A)$, there is an eigenvalue $\lambda \in \mathbb{C}$ such that $|\lambda| = \rho(A)$. Let v the a corresponding eigenvector.

$$\rho(A) = \max_{\lambda \text{ eigenvalue of } A} |\lambda|$$

Perron's Theorem

Theorem (Perron's Theorem)

Let $A \in \mathbb{R}^{n \times n}$ be a positive matrix (i.e., all its entries are positive). Then, the following hold:

1. $\rho(A)$ *is an eigenvalue, and it has a positive eigenvector*
2. $\rho(A)$ *is the only eigenvalue in the complex circle* $|\lambda| = \rho(A)$
3. $\rho(A)$ *has geometric multiplicity 1*
4. $\rho(A)$ *has algebraic multiplicity 1*

- By the definition of $\rho(A)$, there is an eigenvalue $\lambda \in \mathbb{C}$ such that $|\lambda| = \rho(A)$. Let v the a corresponding eigenvector.
- Let u be the vector defined by $u_i = |v_i|$. Then, we have

triangle inequality

$$(Au)_i = \sum_j A_{ij} u_j \geq \left| \sum_j A_{ij} v_j \right| = |\lambda v_i| = \rho(A) \cdot u_i$$

$\to (Av)_i$

so $Au \geq \rho(A) u$.

$\langle A_i, u \rangle$ $\quad \sum A_{ij} \cdot |v_j| = \sum |A_{ij} v_j|$

Perron's Theorem - item 1

- We proved $Au \geq \rho(A)u$.
- If inequality strict, then we have

$$A^2u > \rho(A) \cdot Au$$

and there is some positive $\varepsilon > 0$ such that

$$A^2u \geq (1+\varepsilon)\rho(A)Au$$

$Au > \rho(A)u$

$A \cdot (Au)$

$Au > \rho(A)u$

$A \cdot (Au) \geq A \cdot (\rho(A)u) \cdot (1+\epsilon)$

Perron's Theorem - item 1

- We proved $Au \geq \rho(A)u$.
- If inequality strict, then we have

$$A^2 u > \rho(A) \cdot Au$$

and there is some positive $\varepsilon > 0$ such that

$$A^2 u \geq (1+\varepsilon)\rho(A)Au$$

- By induction, we would have

$$A^{n+1} u \geq (1+\varepsilon)^n \cdot \rho(A)^n \cdot Au$$

Perron's Theorem - item 1

- We proved $Au \geq \rho(A)u$.
- If inequality strict, then we have

$$A^2 u > \rho(A) \cdot Au$$

and there is some positive $\varepsilon > 0$ such that

$$A^2 u \geq (1 + \varepsilon)\rho(A)Au$$

- By induction, we would have

$$A^n u \geq (1 + \varepsilon)^{n-1} \cdot \rho(A)^{n-1} \cdot Au$$

- By Gelfand's formula we would have

$$\rho(A) = \lim_{n \to \infty} \|A^n\|_F^{1/n} \geq (1 + \varepsilon)\rho(A)$$

$$\geq [(1+\varepsilon)^{n-1} \cdot \rho(A)^{n-1}]^{1/n} \quad \therefore Au = \rho(A)u$$

which is a contradiction. So equality must hold.

Perron's theorem - items 1 and 2

- We just proved that $\rho(A)$ is an eigenvalue, with eigenvector $u \geq 0$.

Perron's theorem - items 1 and 2

- We just proved that $\rho(A)$ is an eigenvalue, with eigenvector $u \geq 0$.
- Note that $u > 0$ since $\rho(A)u_i = (Au)_i > 0$

$$Au = \rho(A)u \Rightarrow (Au)_i = \sum_{j=1}^{n} A_{ij} u_j \geq \underbrace{A_{ik}}_{>0} \underbrace{u_k}_{>0} > 0$$

$$(Au)_i = \rho(A) \cdot u_i$$

Perron's theorem - items 1 and 2

- We just proved that $\rho(A)$ is an eigenvalue, with eigenvector $u \geq 0$.
- Note that $u > 0$ since $\rho(A)u_i = (Au)_i > 0$
- Now we are ready for item 2: the only eigenvalue on the complex circle $|\mu| = \rho(A)$ is $\rho(A)$

Perron's theorem - items 1 and 2

- We just proved that $\rho(A)$ is an eigenvalue, with eigenvector $u \geq 0$.
- Note that $u > 0$ since $\rho(A)u_i = (Au)_i > 0$
- Now we are ready for item 2: the only eigenvalue on the complex circle $|\mu| = \rho(A)$ is $\rho(A)$
- If we had another eigenvalue $\lambda \neq \rho(A)$ in the circle $|\mu| = \rho(A)$, where z is the eigenvalue (vector) corresponding to λ, by the previous slide, we know that w defined as $w_i = |z_i|$ satisfies

$Az = \lambda z$

$$Aw = \rho(A)w \quad \Leftrightarrow \quad \sum_j A_{ij} w_j = \rho(A) \cdot |z_i| = |\lambda z_i| = |\sum_j A_{ij} z_j|$$

$(Aw)_i = \rho(A) \cdot w_i$ $\quad$ $|\lambda| = \rho(A)$ $\quad$ λ eigenvalue, z eigenvector of λ

for every $1 \leq i \leq n$

Perron's theorem - items 1 and 2

- We just proved that $\rho(A)$ is an eigenvalue, with eigenvector $u \geq 0$.
- Note that $u > 0$ since $\rho(A)u_i = (Au)_i > 0$
- Now we are ready for item 2: the only eigenvalue on the complex circle $|\mu| = \rho(A)$ is $\rho(A)$
- If we had another eigenvalue $\lambda \neq \rho(A)$ in the circle $|\mu| = \rho(A)$, where z is the eigenvalue corresponding to λ, by the previous slide, we know that w defined as $w_i = |z_i|$ satisfies

$$Aw = \rho(A)w \quad \Leftrightarrow \quad \sum_j A_{ij}w_j = \rho(A) \cdot |z_i| = |\lambda z_i| = |\sum_j A_{ij}z_j|$$

for every $1 \leq i \leq n$

- *Lemma*: if the conditions above hold, then there is $\alpha \in \mathbb{C}$ nonzero such that $\alpha z \geq 0$

Proof by squaring both sides and using complex conjugates.

Perron's theorem - items 2 and 3

- But if $\alpha z \geq 0$ and a nonzero vector, we have

$$\lambda(\alpha z) = \alpha \cdot (\lambda z) = \alpha(Az) = A(\alpha z) \geq 0$$

≥ 0

positive

Perron's theorem - items 2 and 3

- But if $\alpha z \geq 0$ and a nonzero vector, we have

$$\lambda(\alpha z) = \alpha \cdot (\lambda z) = \alpha(Az) = A(\alpha z) \geq 0$$

- Thus we know that λ is a non-negative number. However, $\rho(A)$ is the only non-negative number in the circle $|\mu| = \rho(A)$. This concludes item 2.

$$\underbrace{\lambda}_{\geq 0} \cdot \underbrace{(\alpha z_1)}_{> 0} \geq 0$$

Perron's theorem - items 2 and 3

- But if $\alpha z \geq 0$ and a nonzero vector, we have

$$\lambda(\alpha z) = \alpha \cdot (\lambda z) = \alpha(Az) = A(\alpha z) \geq 0$$

- Thus we know that λ is a non-negative number. However, $\rho(A)$ is the only non-negative number in the circle $|\mu| = \rho(A)$. This concludes item 2.
- Now we are ready to prove item 3: the geometric multiplicity of $\rho(A)$ is 1.
- Suppose not, and let u, v be two linearly independent eigenvectors for $\rho(A)$. We can assume that both u, v are real vectors (why?).

$$A(\underline{\phi} + i\underline{\psi}) = \rho(A) \cdot \underline{\phi} + i\rho(A)\underline{\psi}$$

Perron's theorem - items 2 and 3

- But if $\alpha z \geq 0$ and a nonzero vector, we have

$$\lambda(\alpha z) = \alpha \cdot (\lambda z) = \alpha(Az) = A(\alpha z) \geq 0$$

- Thus we know that λ is a non-negative number. However, $\rho(A)$ is the only non-negative number in the circle $|\mu| = \rho(A)$. This concludes item 2.
- Now we are ready to prove item 3: the geometric multiplicity of $\rho(A)$ is 1.
- Suppose not, and let u, v be two linearly independent eigenvectors for $\rho(A)$. We can assume that both u, v are real vectors (why?).
- Let $\beta > 0$ be such that $u - \beta v \geq 0$ and at least one entry is zero.

Perron's theorem - items 2 and 3

- But if $\alpha z \geq 0$ and a nonzero vector, we have

$$\lambda(\alpha z) = \alpha \cdot (\lambda z) = \alpha(Az) = A(\alpha z) \geq 0$$

- Thus we know that λ is a non-negative number. However, $\rho(A)$ is the only non-negative number in the circle $|\mu| = \rho(A)$. This concludes item 2.
- Now we are ready to prove item 3: the geometric multiplicity of $\rho(A)$ is 1.
- Suppose not, and let u, v be two linearly independent eigenvectors for $\rho(A)$. We can assume that both u, v are real vectors (why?).
- Let $\beta > 0$ be such that $u - \beta v \geq 0$ and at least one entry is zero.
- $u - \beta v \neq 0$ since the vectors are linearly independent

Perron's theorem - items 2 and 3

- But if $\alpha z \geq 0$ and a nonzero vector, we have

$$\lambda(\alpha z) = \alpha \cdot (\lambda z) = \alpha(Az) = A(\alpha z) \geq 0$$

- Thus we know that λ is a non-negative number. However, $\rho(A)$ is the only non-negative number in the circle $|\mu| = \rho(A)$. This concludes item 2.
- Now we are ready to prove item 3: the geometric multiplicity of $\rho(A)$ is 1.
- Suppose not, and let u, v be two linearly independent eigenvectors for $\rho(A)$. We can assume that both u, v are real vectors (why?).
- Let $\beta > 0$ be such that $u - \beta v \geq 0$ and at least one entry is zero.
- $u - \beta v \neq 0$ since the vectors are linearly independent
- But for each $1 \leq i \leq n$

$$\rho(A) \cdot (u - \beta v)_i = (A(u - \beta v))_i > 0$$

which contradicts our choice of β. Thus, there cannot be two linearly independent eigenvectors.

Perron-Frobenius

Theorem (Perron-Frbenius)

If a non-negative matrix $A \in \mathbb{R}^{n \times n}$ is aperiodic and irreducible, then the following hold:

1. $\rho(A)$ *is an eigenvalue, and it has a positive eigenvector*
2. $\rho(A)$ *is the only eigenvalue in the complex circle* $|\lambda| = \rho(A)$
3. $\rho(A)$ *has geometric multiplicity 1*
4. $\rho(A)$ *has algebraic multiplicity 1*

Perron-Frobenius

Theorem (Perron-Frbenius)

If a non-negative matrix $A \in \mathbb{R}^{n \times n}$ is aperiodic and irreducible, then the following hold:

1. $\rho(A)$ *is an eigenvalue, and it has a positive eigenvector*
2. $\rho(A)$ *is the only eigenvalue in the complex circle* $|\lambda| = \rho(A)$
3. $\rho(A)$ *has geometric multiplicity 1*
4. $\rho(A)$ *has algebraic multiplicity 1*

- By previous lecture, we saw that A being aperiodic and irreducible implies that there is $m > 0$ such that A^m has all positive entries.

Perron-Frobenius

Theorem (Perron-Frbenius)

If a non-negative matrix $A \in \mathbb{R}^{n \times n}$ is aperiodic and irreducible, then the following hold:

1. $\rho(A)$ *is an eigenvalue, and it has a positive eigenvector*
2. $\rho(A)$ *is the only eigenvalue in the complex circle* $|\lambda| = \rho(A)$
3. $\rho(A)$ *has geometric multiplicity 1*
4. $\rho(A)$ *has algebraic multiplicity 1*

- By previous lecture, we saw that A being aperiodic and irreducible implies that there is $m > 0$ such that A^m has all positive entries.
- Apply Perron's theorem to A^m and note that the eigenvalues of A^m are λ_i^m, where λ_i are the eigenvalues of A

Fundamental Theorem of Markov Chains

- The *return time* from state i to itself is defined as

$$H_{i,i} := \min\{t \geq 1 \mid X_t = i, X_0 = i\}$$

Fundamental Theorem of Markov Chains

- The *return time* from state i to itself is defined as

$$H_{i,i} := \min\{t \geq 1 \mid X_t = i, X_0 = i\}$$

- *Expected return time:* defined as $h_{i,i} := \mathbb{E}[H_{i,i}]$.

Fundamental Theorem of Markov Chains

- The *return time* from state i to itself is defined as

$$H_{i,i} := \min\{t \geq 1 \mid X_t = i, X_0 = i\}$$

- *Expected return time:* defined as $h_{i,i} := \mathbb{E}[H_{i,i}]$.

Theorem (Fundamental Theorem of Markov Chains)

Any finite, irreducible and aperiodic Markov Chain has the following properties:

1. *There exists a unique stationary distribution π, where $\pi_i > 0$ for all $i \in [n]$*
2. *The sequence of distributions $\{p_t\}_{t \geq 0}$ will converge to π, no matter what the initial distribution is*
3. $$\pi_i = \lim_{t \to \infty} P^t_{i,i} = \frac{1}{h_{i,i}}$$

Fundamental Theorem of Markov Chains

Theorem (Fundamental Theorem of Markov Chains)

Any finite, irreducible and aperiodic Markov Chain has the following properties:

1. *There is unique stationary distribution* π*, where* $\pi_i > 0$ *for all* $i \in [n]$
2. *For every distribution* $p_0 \in \mathbb{R}^n_{\geq 0}$,
$$\lim_{t \to \infty} p_0 \cdot P^t = \pi$$
3. $$\pi_i = \lim_{t \to \infty} P^t_{i,i} = \frac{1}{h_{i,i}}$$

Fundamental Theorem of Markov Chains

Theorem (Fundamental Theorem of Markov Chains)

Any finite, irreducible and aperiodic Markov Chain has the following properties:

1. *There is unique stationary distribution* π*, where* $\pi_i > 0$ *for all* $i \in [n]$
2. *For every distribution* $p_0 \in \mathbb{R}^n_{\geq 0}$,
$$\lim_{t\to\infty} p_0 \cdot P^t = \pi$$
3. $$\pi_i = \lim_{t\to\infty} P^t_{i,i} = \frac{1}{h_{i,i}}$$

- The transition matrix P is non-negative, irreducible and aperiodic. So we can apply Perron-Frobenius and prove items 1 and 2.

Fundamental Theorem of Markov Chains

Theorem (Fundamental Theorem of Markov Chains)

Any finite, irreducible and aperiodic Markov Chain has the following properties:

1. *There is unique stationary distribution* π*, where* $\pi_i > 0$ *for all* $i \in [n]$
2. *For every distribution* $p_0 \in \mathbb{R}^n_{\geq 0}$, $\lim_{t\to\infty} p_0 \cdot P^t = \pi$
3. $$\pi_i = \lim_{t\to\infty} P^t_{i,i} = \frac{1}{h_{i,i}}$$

If our underlying graph is undirected:

Fundamental Theorem of Markov Chains

Theorem (Fundamental Theorem of Markov Chains)

Any finite, irreducible and aperiodic Markov Chain has the following properties:

1. *There is unique stationary distribution* π, *where* $\pi_i > 0$ *for all* $i \in [n]$
2. *For every distribution* $p_0 \in \mathbb{R}^n_{\geq 0}$, $\lim_{t\to\infty} p_0 \cdot P^t = \pi$
3. $$\pi_i = \lim_{t\to\infty} P^t_{i,i} = \frac{1}{h_{i,i}}$$

If our underlying graph is undirected:

- If A_G adjacency matrix of $G(V,E)$ and $D = diag(d_1, d_2, \ldots, d_n)$, transition matrix:

$$P = D^{-1} \cdot A_G$$

Fundamental Theorem of Markov Chains

Theorem (Fundamental Theorem of Markov Chains)

Any finite, irreducible and aperiodic Markov Chain has the following properties:

1. *There is unique stationary distribution π, where $\pi_i > 0$ for all $i \in [n]$*
2. *For every distribution $p_0 \in \mathbb{R}^n_{\geq 0}$,* $\lim_{t \to \infty} p_0 \cdot P^t = \pi$
3. $$\pi_i = \lim_{t \to \infty} P^t_{i,i} = \frac{1}{h_{i,i}}$$

If our underlying graph is undirected:

- If A_G adjacency matrix of $G(V, E)$ and $D = diag(d_1, d_2, \ldots, d_n)$, transition matrix:

$$P = D^{-1} \cdot A_G$$

- Note that in this case, easy to guess stationary distribution:

$$\pi_i = \frac{d_i}{2m}, \quad m = |E|$$

Fundamental Theorem of Markov Chains

If our underlying graph is undirected:

Fundamental Theorem of Markov Chains

If our underlying graph is undirected:

- In this case, easy to guess stationary distribution:

$$\pi_i = \frac{d_i}{2m}, \quad m = |E|$$

Fundamental Theorem of Markov Chains

If our underlying graph is undirected:

- In this case, easy to guess stationary distribution:

$$\pi_i = \frac{d_i}{2m}, \quad m = |E|$$

- If A_G adjacency matrix of $G(V, E)$ and $D = diag(d_1, d_2, \ldots, d_n)$, transition matrix:

$$P = D^{-1} \cdot A_G$$

Fundamental Theorem of Markov Chains

If our underlying graph is undirected:

- In this case, easy to guess stationary distribution:

$$\pi_i = \frac{d_i}{2m}, \quad m = |E|$$

- If A_G adjacency matrix of $G(V, E)$ and $D = diag(d_1, d_2, \ldots, d_n)$, transition matrix:

$$P = D^{-1} \cdot A_G$$

  - P not symmetric, but *similar* to a symmetric matrix:

$$D^{1/2} P D^{-1/2} = D^{1/2} D^{-1} A_G D^{-1/2} = D^{-1/2} A_G D^{-1/2} = P'$$

symmetric

Fundamental Theorem of Markov Chains

If our underlying graph is undirected:

- If A_G adjacency matrix of $G(V, E)$ and $D = diag(d_1, d_2, \ldots, d_n)$, transition matrix:

$$P = D^{-1} \cdot A_G$$

  - P not symmetric, but *similar* to a symmetric matrix:

$$D^{1/2} P D^{-1/2} = D^{1/2} D^{-1} A_G D^{-1/2} = D^{-1/2} A_G D^{-1/2} = P'$$

  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!

Fundamental Theorem of Markov Chains

If our underlying graph is undirected:

- If A_G adjacency matrix of $G(V, E)$ and $D = diag(d_1, d_2, \ldots, d_n)$, transition matrix:

$$P = D^{-1} \cdot A_G$$

  - P not symmetric, but *similar* to a symmetric matrix:

$$D^{1/2} P D^{-1/2} = D^{1/2} D^{-1} A_G D^{-1/2} = D^{-1/2} A_G D^{-1/2} = P'$$

  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!
  - Eigenvectors of P are $D^{-1/2} v_i$ where v_i are eigenvectors of P'. And v_i can be taken to form *orthonormal basis*.

Fundamental Theorem of Markov Chains

- Stationary distribution: $\pi_i = \dfrac{d_i}{2m}, \quad m = |E|$
- Transition matrix: $P = D^{-1} \cdot A_G$
  - P *similar* to a symmetric matrix: $P' = D^{-1/2} A_G D^{-1/2}$
  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!
  - Eigenvectors of P are $D^{-1/2} v_i$ where v_i are eigenvectors of P'. And v_i can be taken to form *orthonormal basis*.

Fundamental Theorem of Markov Chains

- Stationary distribution: $\pi_i = \dfrac{d_i}{2m}, \quad m = |E|$
- Transition matrix: $P = D^{-1} \cdot A_G$
  - P *similar* to a symmetric matrix: $P' = D^{-1/2} A_G D^{-1/2}$
  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!
  - Eigenvectors of P are $D^{-1/2} v_i$ where v_i are eigenvectors of P'. And v_i can be taken to form *orthonormal basis*.
  - Graph strongly connected $\Rightarrow$ *Perron-Frobenius* for irreducible non-negative matrices

Fundamental Theorem of Markov Chains

- Stationary distribution: $\pi_i = \dfrac{d_i}{2m}, \quad m = |E|$
- Transition matrix: $P = D^{-1} \cdot A_G$
  - P *similar* to a symmetric matrix: $P' = D^{-1/2} A_G D^{-1/2}$
  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!
  - Eigenvectors of P are $D^{-1/2} v_i$ where v_i are eigenvectors of P'. And v_i can be taken to form *orthonormal basis*.
  - Graph strongly connected $\Rightarrow$ *Perron-Frobenius* for irreducible non-negative matrices
    - *unique* eigenvector whose eigenvalue has *max absolute value*
    - eigenvector has *all positive coordinates*
    - eigenvalue is *positive*

v_i's orthonormal

Fundamental Theorem of Markov Chains

- Stationary distribution: $\pi_i = \dfrac{d_i}{2m}, \quad m = |E|$
- Transition matrix: $P = D^{-1} \cdot A_G$
  - P *similar* to a symmetric matrix: $P' = D^{-1/2} A_G D^{-1/2}$
  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!
  - Eigenvectors of P are $D^{-1/2} v_i$ where v_i are eigenvectors of P'. And v_i can be taken to form *orthonormal basis*.
  - Graph strongly connected $\Rightarrow$ *Perron-Frobenius* for irreducible non-negative matrices
    - *unique* eigenvector whose eigenvalue has *max absolute value*
    - eigenvector has *all positive coordinates*
    - eigenvalue is *positive*
  - This eigenvector is π!

Fundamental Theorem of Markov Chains

- Stationary distribution: $\pi_i = \dfrac{d_i}{2m}, \quad m = |E|$
- Transition matrix: $P = D^{-1} \cdot A_G$
  - P *similar* to a symmetric matrix: $P' = D^{-1/2} A_G D^{-1/2}$
  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!
  - Eigenvectors of P are $D^{-1/2} v_i$ where v_i are eigenvectors of P'. And v_i can be taken to form *orthonormal basis*.
  - Graph strongly connected $\Rightarrow$ *Perron-Frobenius* for irreducible non-negative matrices
    - *unique* eigenvector whose eigenvalue has *max absolute value*
    - eigenvector has *all positive coordinates*
    - eigenvalue is *positive*
  - This eigenvector is π!
  - All random walks converge to π, as we wanted to show.

- Main Tools
  - Linear Algebra Background
  - Perron-Frobenius
- Main Applications
  - Fundamental Theorem of Markov Chains
  - Page Rank
- Conclusion
- Acknowledgements

Page Rank

- **Setting:** we have a directed graph describing relationships between set of webpages.

Page Rank

- **Setting:** we have a directed graph describing relationships between set of webpages.

  There is a directed edge (i, j) if there is a link from page i to page j.

Page Rank

- **Setting:** we have a directed graph describing relationships between set of webpages.

  There is a directed edge (i, j) if there is a link from page i to page j.
- **Goal:** want algorithm to "rank" how important a page is.

Page Rank

- **Setting:** we have a directed graph describing relationships between set of webpages.

  There is a directed edge (i, j) if there is a link from page i to page j.
- **Goal:** want algorithm to "rank" how important a page is.
- *Intuition:* if many other pages link to a particular page, then the linked page must be important!

Other intuition: if important page links to another page this linked page must also be important

Page Rank

- **Setting:** we have a directed graph describing relationships between set of webpages.

  There is a directed edge (i, j) if there is a link from page i to page j.
- **Goal:** want algorithm to "rank" how important a page is.
- *Intuition:* if many other pages link to a particular page, then the linked page must be important!

Algorithm (Page Rank Algorithm)

1. *Initially, each page has pagerank value* $\frac{1}{n}$

Page Rank

- **Setting:** we have a directed graph describing relationships between set of webpages.

  There is a directed edge (i, j) if there is a link from page i to page j.
- **Goal:** want algorithm to "rank" how important a page is.
- *Intuition:* if many other pages link to a particular page, then the linked page must be important!

Algorithm (Page Rank Algorithm)

1. *Initially, each page has pagerank value* $\frac{1}{n}$
2. *In each step, each page:*

Page Rank

- **Setting:** we have a directed graph describing relationships between set of webpages.

  There is a directed edge (i, j) if there is a link from page i to page j.
- **Goal:** want algorithm to "rank" how important a page is.
- *Intuition:* if many other pages link to a particular page, then the linked page must be important!

Algorithm (Page Rank Algorithm)

1. *Initially, each page has pagerank value* $\frac{1}{n}$
2. *In each step, each page:*
   1. *divides its pagerank value equally to its outgoing link,*

Page Rank

- **Setting:** we have a directed graph describing relationships between set of webpages.

  There is a directed edge (i, j) if there is a link from page i to page j.
- **Goal:** want algorithm to "rank" how important a page is.
- *Intuition:* if many other pages link to a particular page, then the linked page must be important!

Algorithm (Page Rank Algorithm)

1. *Initially, each page has pagerank value* $\frac{1}{n}$
2. *In each step, each page:*
   1. *divides its pagerank value equally to its outgoing link,*
   2. *sends these equal shares to the pages it points to,*

Page Rank

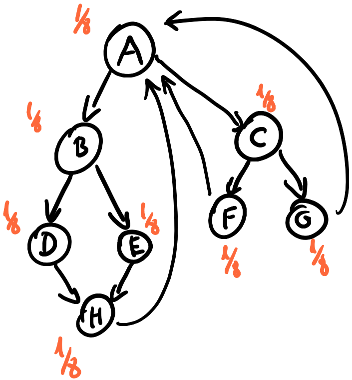
- **Setting:** we have a directed graph describing relationships between set of webpages.

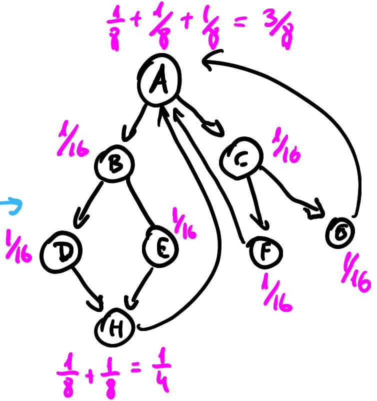
  There is a directed edge (i, j) if there is a link from page i to page j.
- **Goal:** want algorithm to "rank" how important a page is.
- *Intuition:* if many other pages link to a particular page, then the linked page must be important!

Algorithm (Page Rank Algorithm)

1. *Initially, each page has pagerank value* $\frac{1}{n}$
2. *In each step, each page:*
    1. *divides its pagerank value equally to its outgoing link,*
    2. *sends these equal shares to the pages it points to,*
    3. *updates its new pagerank value to be the sum of shares it receives.*

Page Rank - Example





$w_0 \cdot P = w_1$

↳ depends only on graph structure

Page Rank

Algorithm (Page Rank Algorithm)

1. *Initially, each page has pagerank value* $\frac{1}{n}$
2. *In each step, each page:*
   1. *divides its pagerank value equally to its outgoing link,*
   2. *sends these equal shares to the pages it points to,*
   3. *updates its new pagerank value to be the sum of shares it receives.*

Page rank is a Markov chain!

Page Rank

Algorithm (Page Rank Algorithm)

1. *Initially, each page has pagerank value* $\frac{1}{n}$
2. *In each step, each page:*
   1. *divides its pagerank value equally to its outgoing link,*
   2. *sends these equal shares to the pages it points to,*
   3. *updates its new pagerank value to be the sum of shares it receives.*

- *Equilibrium* of pagerank values equal to probabilities of *stationary distribution* of random walk

$$P \in \mathbb{R}^{n \times n}, \quad P_{i,j} = \frac{1}{\delta^{out}(i)}$$

Page Rank

Algorithm (Page Rank Algorithm)

1. *Initially, each page has pagerank value* $\frac{1}{n}$
2. *In each step, each page:*
   1. *divides its pagerank value equally to its outgoing link,*
   2. *sends these equal shares to the pages it points to,*
   3. *updates its new pagerank value to be the sum of shares it receives.*

- *Equilibrium* of pagerank values equal to probabilities of *stationary distribution* of random walk

$$P \in \mathbb{R}^{n\times n}, \quad P_{i,j} = \frac{1}{\delta^{out}(i)}$$

- Pagerank values and transition probabilities satisfy same equations:

$$p_{t+1}(j) = \sum_{i:(i,j)\in E} \frac{p_t(i)}{\delta^{out}(i)} \quad \Rightarrow \quad p_{t+1} = p_t \cdot P$$

Page Rank

Algorithm (Page Rank Algorithm)

1. *Initially, each page has pagerank value $\frac{1}{n}$*
2. *In each step, each page:*
   1. *divides its pagerank value equally to its outgoing link,*
   2. *sends these equal shares to the pages it points to,*
   3. *updates its new pagerank value to be the sum of shares it receives.*

- *Equilibrium* of pagerank values equal to probabilities of *stationary distribution* of random walk

$$P \in \mathbb{R}^{n \times n}, \quad P_{i,j} = \frac{1}{\delta^{out}(i)}$$

- Pagerank values and transition probabilities satisfy same equations:

$$p_{t+1}(j) = \sum_{i:(i,j)\in E} \frac{p_t(i)}{\delta^{out}(i)} \quad \Rightarrow \quad p_{t+1} = p_t \cdot P$$

- If graph finite, irreducible and aperiodic, fundamental theorem guarantees stationary distribnution.

Page Rank

- In practice, directed graph may not satisfy fundamental theorem's conditions

Page Rank

- In practice, directed graph may not satisfy fundamental theorem's conditions
- Modify original graph as follows:

Page Rank

- In practice, directed graph may not satisfy fundamental theorem's conditions
- Modify original graph as follows:
  - Fix number $0 < s < 1$
  - Divide s fraction of its pagerank value to its neighbors,
  - $1 - s$ fraction of its pagerank value to all nodes evenly

Page Rank

- In practice, directed graph may not satisfy fundamental theorem's conditions
- Modify original graph as follows:
  - Fix number $0 < s < 1$
  - Divide s fraction of its pagerank value to its neighbors,
  - $1 - s$ fraction of its pagerank value to all nodes evenly
- Equivalent to the random walk:

Page Rank

- In practice, directed graph may not satisfy fundamental theorem's conditions
- Modify original graph as follows:
  - Fix number $0 < s < 1$
  - Divide s fraction of its pagerank value to its neighbors,
  - $1 - s$ fraction of its pagerank value to all nodes evenly
- Equivalent to the random walk:
  - With probability s go to one of its neighbors (uniformly at random),
  - With probability $1 - s$ go to random page (uniformly at random)

Page Rank

- In practice, directed graph may not satisfy fundamental theorem's conditions
- Modify original graph as follows:
  - Fix number $0 < s < 1$
  - Divide s fraction of its pagerank value to its neighbors,
  - $1 - s$ fraction of its pagerank value to all nodes evenly
- Equivalent to the random walk:
  - With probability s go to one of its neighbors (uniformly at random),
  - With probability $1 - s$ go to random page (uniformly at random)
- Now resulting graph is *strongly connected* and *aperiodic* $\Rightarrow$ unique stationary distribution

Page Rank

- In practice, directed graph may not satisfy fundamental theorem's conditions
- Modify original graph as follows:
  - Fix number $0 < s < 1$
  - Divide s fraction of its pagerank value to its neighbors,
  - $1 - s$ fraction of its pagerank value to all nodes evenly
- Equivalent to the random walk:
  - With probability s go to one of its neighbors (uniformly at random),
  - With probability $1 - s$ go to random page (uniformly at random)
- Now resulting graph is *strongly connected* and *aperiodic* $\Rightarrow$ unique stationary distribution
- This modification does not change "relative importance" of vertices

Conclusion

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

- Page Rank algorithm (today's lecture)
- Approximation algorithms for counting problems [Karp, Luby & Madras]
  - Permanent of non-negative matrices [Jerrum, Vigoda & Sinclair]
- Sampling Problems
  - Gibbs sampling in statistical physics
  - many more places
- Probability amplification without too much randomness (efficient)
  - Random walks on expander graphs
- many more

Acknowledgement

- Lecture based largely on:
  - Hannah Cairns notes on Perron-Frobenius (see link in course webpage)
  - Lap Chi's notes
  - [Motwani & Raghavan 2007, Chapter 6]
- See Lap Chi's notes at `https://cs.uwaterloo.ca/~lapchi/cs466/notes/L11.pdf`
- Also see Lap Chi's notes `https://cs.uwaterloo.ca/~lapchi/cs466/notes/L14.pdf` for a proof of fundamental theorem of Markov chains for undirected graphs.

References I

Motwani, Rajeev and Raghavan, Prabhakar (2007)
Randomized Algorithms

Karp, R.M. and Luby, M. and Madras, N. (1989)
Monte-Carlo approximation algorithms for enumeration problems.
Journal of algorithms, 10(3), pp.429-448.

Jerrum, M. and Sinclair, A. and Vigoda, E. (2004)
A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries.
Journal of the ACM (JACM), 51(4), pp.671-697.