Lecture 9: Random Walks, Markov Chains, Mixing Time

Rafael Oliveira

University of Waterloo
Cheriton School of Computer Science

rafael.oliveira.teaching@gmail.com

May 24, 2024

Overview

What is a Random Walk?

Given a graph $G(V, E)$

1. random walk starts from a vertex v_0
2. at each time step it moves *uniformly* to a *random neighbor* of the current vertex in the graph

$$v_{t+1} \leftarrow_R N_G(v_t)$$

What is a Random Walk?

Given a graph $G(V, E)$

1. random walk starts from a vertex v_0
2. at each time step it moves *uniformly* to a *random neighbor* of the current vertex in the graph

$$v_{t+1} \leftarrow_R N_G(v_t)$$

Basic questions involving random walks:

What is a Random Walk?

Given a graph $G(V, E)$

1. random walk starts from a vertex v_0
2. at each time step it moves *uniformly* to a *random neighbor* of the current vertex in the graph

$$v_{t+1} \leftarrow_R N_G(v_t)$$

Basic questions involving random walks:

- *Stationary distribution:* does the random walk converge to a "stable" distribution? If it does, what is this distribution?

What is a Random Walk?

Given a graph $G(V, E)$

1. random walk starts from a vertex v_0
2. at each time step it moves *uniformly* to a *random neighbor* of the current vertex in the graph

$$v_{t+1} \leftarrow_R N_G(v_t)$$

Basic questions involving random walks:

- *Stationary distribution:* does the random walk converge to a “stable” distribution? If it does, what is this distribution?
- *Mixing time:* how long does it take for the walk to converge to the stationary distribution?

What is a Random Walk?

Given a graph $G(V, E)$

1. random walk starts from a vertex v_0
2. at each time step it moves *uniformly* to a *random neighbor* of the current vertex in the graph

$$v_{t+1} \leftarrow_R N_G(v_t)$$

Basic questions involving random walks:

- *Stationary distribution:* does the random walk converge to a "stable" distribution? If it does, what is this distribution?
- *Mixing time:* how long does it take for the walk to converge to the stationary distribution?
- *Hitting time:* starting from a vertex v_0, what is expected number of steps until it reaches a vertex v_f?

What is a Random Walk?

Given a graph $G(V, E)$

1. random walk starts from a vertex v_0
2. at each time step it moves *uniformly* to a *random neighbor* of the current vertex in the graph

$$v_{t+1} \leftarrow_R N_G(v_t)$$

Basic questions involving random walks:

- *Stationary distribution:* does the random walk converge to a "stable" distribution? If it does, what is this distribution?
- *Mixing time:* how long does it take for the walk to converge to the stationary distribution?
- *Hitting time:* starting from a vertex v_0, what is expected number of steps until it reaches a vertex v_f?
- *Cover time:* how long does it take to reach every vertex of the graph at least once?

Random Walk: Example

- Suppose $G(V, E) = K_n$, the complete graph, $a, b \in V$ two vertices

Random Walk: Example

- Suppose $G(V, E) = K_n$, the complete graph, $a, b \in V$ two vertices
  1. What is expected number of steps to reach b in simple random walk starting at a? (i.e., hitting time)

Random Walk: Example

- Suppose $G(V, E) = K_n$, the complete graph, $a, b \in V$ two vertices
  1. What is expected number of steps to reach b in simple random walk starting at a? (i.e., hitting time)
  2. Starting from a, what is expected number of steps to visit all vertices? (i.e, cover time)

Random Walk: Example

- Suppose $G(V, E) = K_n$, the complete graph, $a, b \in V$ two vertices
  1. What is expected number of steps to reach b in simple random walk starting at a? (i.e., hitting time)
  2. Starting from a, what is expected number of steps to visit all vertices? (i.e, cover time)
  3. Stationary Distribution?

Random Walk: Example

- Suppose $G(V, E) = K_n$, the complete graph, $a, b \in V$ two vertices
  1. What is expected number of steps to reach b in simple random walk starting at a? (i.e., hitting time)
  2. Starting from a, what is expected number of steps to visit all vertices? (i.e, cover time)
  3. Stationary Distribution?
  4. Mixing time? (we'll do it later)

Random Walk: Example

- Suppose $G(V, E) = K_n$, the complete graph, $a, b \in V$ two vertices
  1. What is expected number of steps to reach b in simple random walk starting at a? (i.e., hitting time)
  2. Starting from a, what is expected number of steps to visit all vertices? (i.e, cover time)
  3. Stationary Distribution?
  4. Mixing time? (we'll do it later)
- **Practice question:** Compare question 2 to coupon collector problem!

What is a Markov Chain?

Random walk is a special kind of *stochastic process*:

$$\Pr[X_t = v_t \mid X_0 = v_0, \ldots, X_{t-1} = v_{t-1}] = \Pr[X_t = v_t \mid X_{t-1} = v_{t-1}]$$

What is a Markov Chain?

Random walk is a special kind of *stochastic process*:

$$\Pr[X_t = v_t \mid X_0 = v_0, \ldots, X_{t-1} = v_{t-1}] = \Pr[X_t = v_t \mid X_{t-1} = v_{t-1}]$$

Probability that we are at vertex v_t at time t only depends on the state of our process at time $t-1$.

What is a Markov Chain?

Random walk is a special kind of *stochastic process*:

$$\Pr[X_t = v_t \mid X_0 = v_0, \ldots, X_{t-1} = v_{t-1}] = \Pr[X_t = v_t \mid X_{t-1} = v_{t-1}]$$

Probability that we are at vertex v_t at time t only depends on the state of our process at time $t-1$.

Process is "*forgetful*/*memoryless*"

Markov chain is characterized by this property.

Why study Markov Chains and Random Walks?

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

Why study Markov Chains and Random Walks?

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

- Page Rank algorithm (next lecture)

Why study Markov Chains and Random Walks?

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

- Page Rank algorithm (next lecture)
- Approximation algorithms for counting problems [Karp, Luby & Madras]
    - Permanent of non-negative matrices [Jerrum, Vigoda & Sinclair] (*great final project topic!*)

Why study Markov Chains and Random Walks?

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

- Page Rank algorithm (next lecture)
- Approximation algorithms for counting problems [Karp, Luby & Madras]
  - Permanent of non-negative matrices [Jerrum, Vigoda & Sinclair] (*great final project topic!*)
- Sampling Problems
  - Gibbs sampling in statistical physics (*great final project topic!*)
  - many more places

Why study Markov Chains and Random Walks?

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

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

Why study Markov Chains and Random Walks?

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

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

Representing Finite Markov Chains

Markov chain can be seen as weighted directed graph.

Representing Finite Markov Chains

Markov chain can be seen as weighted directed graph.

- Vertex is a state of Markov chain
- edge (i, j) corresponds to transition probability from i to j

Representing Finite Markov Chains

Markov chain can be seen as weighted directed graph.

- Vertex is a state of Markov chain
- edge (i, j) corresponds to transition probability from i to j

- Markov Chain *irreducible* if underlying directed graph is *strongly connected* (i.e. there is directed path from i to j for any pair $i, j \in V$)

Representing Finite Markov Chains

Markov chain can be seen in weighted adjacency matrix format.

Representing Finite Markov Chains

Markov chain can be seen in weighted adjacency matrix format.

- $P \in \mathbb{R}^{n \times n}$ transition matrix

Representing Finite Markov Chains

Markov chain can be seen in weighted adjacency matrix format.

- $P \in \mathbb{R}^{n \times n}$ transition matrix
- entry $P_{i,j}$ corresponds to transition probability to i from j

Representing Finite Markov Chains

Markov chain can be seen in weighted adjacency matrix format.

- $P \in \mathbb{R}^{n \times n}$ transition matrix
- entry $P_{i,j}$ corresponds to transition probability to i from j
- $p_t \in \mathbb{R}^n$ probability vector: $p_t(i) :=$ Pr[being at *state i* at *time t*]

Representing Finite Markov Chains

Markov chain can be seen in weighted adjacency matrix format.

- $P \in \mathbb{R}^{n \times n}$ transition matrix
- entry $P_{i,j}$ corresponds to transition probability to i from j
- $p_t \in \mathbb{R}^n$ probability vector: $p_t(i) := \Pr[\text{being at } \textit{state } i \text{ at } \textit{time } t]$
- Transition given by

$$p_{t+1} = P \cdot p_t$$

Properties of Markov Chains

- *Period* of a state i is:

$$\gcd\{t \in \mathbb{N} \mid P_{i,i}^t > 0\}$$

That is, gcd of all times t such that the probability of starting at state i and being back at i at time t is positive

Properties of Markov Chains

- *Period* of a state i is:

$$\gcd\{t \in \mathbb{N} \mid P_{i,i}^{t} > 0\}$$

  That is, gcd of all times t such that the probability of starting at state i and being back at i at time t is positive
- State i is *aperiodic* if its period is 1.

Properties of Markov Chains

- *Period* of a state i is:

$$\gcd\{t \in \mathbb{N} \mid P^t_{i,i} > 0\}$$

  That is, gcd of all times t such that the probability of starting at state i and being back at i at time t is positive
- State i is *aperiodic* if its period is 1.
- Markov Chain aperiodic if *all states* are aperiodic (otherwise periodic)

Properties of Markov Chains

- *Period* of a state i is:

$$\gcd\{t \in \mathbb{N} \mid P^t_{i,i} > 0\}$$

  That is, gcd of all times t such that the probability of starting at state i and being back at i at time t is positive
- State i is *aperiodic* if its period is 1.
- Markov Chain aperiodic if *all states* are aperiodic (otherwise periodic)
  - Bipartite graphs yield periodic Markov Chains

Properties of Markov Chains

- *Period* of a state i is:

$$\gcd\{t \in \mathbb{N} \mid P^t_{i,i} > 0\}$$

  That is, gcd of all times t such that the probability of starting at state i and being back at i at time t is positive
- State i is *aperiodic* if its period is 1.
- Markov Chain aperiodic if *all states* are aperiodic (otherwise periodic)
  - Bipartite graphs yield periodic Markov Chains

Lemma

For any finite, irreducible and aperiodic Markov Chain, there exists $T < \infty$ such that

$$P^t_{i,j} > 0 \quad \textit{for any } i, j \in V \textit{ and } t \geq T.$$

See proof in reference of [Häggström, Chapter 4].

Stationary Distributions

Definition (Stationary Distribution)

A stationary distribution of a Markov Chain is a probability distribution $\pi \in \mathbb{R}^n$ such that

$$P\pi = \pi.$$

Stationary Distributions

Definition (Stationary Distribution)

A stationary distribution of a Markov Chain is a probability distribution $\pi \in \mathbb{R}^n$ such that

$$P\pi = \pi.$$

- Informally, π is an "equilibrium/fixed point" state, as we have $\pi = P^t\pi$ for any $t \geq 0$.

Stationary Distributions

Definition (Stationary Distribution)

A stationary distribution of a Markov Chain is a probability distribution $\pi \in \mathbb{R}^n$ such that

$$P\pi = \pi.$$

- Informally, π is an "equilibrium/fixed point" state, as we have $\pi = P^t \pi$ for any $t \geq 0$.
- **Intuition:** If we run finite, irreducible and aperiodic Markov Chain long enough, we will converge to a stationary distribution.

Stationary Distributions

Definition (Stationary Distribution)

A stationary distribution of a Markov Chain is a probability distribution $\pi \in \mathbb{R}^n$ such that

$$P\pi = \pi.$$

- Informally, π is an "equilibrium/fixed point" state, as we have $\pi = P^t \pi$ for any $t \geq 0$.
- **Intuition:** If we run finite, irreducible and aperiodic Markov Chain long enough, we will converge to a stationary distribution.
  - what do you mean by converge here?

Stationary Distributions

Definition (Stationary Distribution)

A stationary distribution of a Markov Chain is a probability distribution $\pi \in \mathbb{R}^n$ such that

$$P\pi = \pi.$$

- Informally, π is an "equilibrium/fixed point" state, as we have $\pi = P^t \pi$ for any $t \geq 0$.
- **Intuition:** If we run finite, irreducible and aperiodic Markov Chain long enough, we will converge to a stationary distribution.
  - what do you mean by converge here?
- Given two distributions $p, q \in \mathbb{R}^n$, their *total variational distance* is

$$\Delta_{TV}(p, q) = \frac{1}{2} \sum_{i=1}^{n} |p_i - q_i| = \frac{1}{2} \cdot \|p - q\|_1$$

Stationary Distributions

Definition (Stationary Distribution)

A stationary distribution of a Markov Chain is a probability distribution $\pi \in \mathbb{R}^n$ such that

$$P\pi = \pi.$$

- Informally, π is an "equilibrium/fixed point" state, as we have $\pi = P^t \pi$ for any $t \geq 0$.
- **Intuition:** If we run finite, irreducible and aperiodic Markov Chain long enough, we will converge to a stationary distribution.
  - what do you mean by converge here?
- Given two distributions $p, q \in \mathbb{R}^n$, their *total variational distance* is

$$\Delta_{TV}(p, q) = \frac{1}{2} \sum_{i=1}^{n} |p_i - q_i| = \frac{1}{2} \cdot \|p - q\|_1$$

- p_t *converges* to q iff $\lim_{t \to \infty} \Delta_{TV}(p_t, q) = 0$

Mixing Time of Markov Chains

Definition (Mixing Time)

The ε-mixing time of a Markov Chain is the smallest t such that

$$\Delta_{TV}(p_t, \pi) \leq \varepsilon$$

regardless of the initial starting distribution p_0.

Mixing Time of Markov Chains

Definition (Mixing Time)

The ε-mixing time of a Markov Chain is the smallest t such that

$$\Delta_{TV}(p_t, \pi) \leq \varepsilon$$

regardless of the initial starting distribution p_0.

- For complete graph, eigenvalues $\lambda_1 = 1, \lambda_2 = \cdots = \lambda_n = -1/(n-1)$, corresponding eigenvectors $v_1, \ldots, v_n$ (orthonormal)

Hitting Time

- Given states i, j in a Markov chain, the *hitting time* from state i to state j is defined as

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

We say $T_{i,j} = \infty$ if the Markov chain never visits j starting from i.

Hitting Time

- Given states i, j in a Markov chain, the *hitting time* from state i to state j is defined as

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

  We say $T_{i,j} = \infty$ if the Markov chain never visits j starting from i.
- The *mean hitting time* $\tau_{i,j} := \mathbb{E}[T_{i,j}]$

Hitting Time

- Given states i, j in a Markov chain, the *hitting time* from state i to state j is defined as

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

  We say $T_{i,j} = \infty$ if the Markov chain never visits j starting from i.
- The *mean hitting time* $\tau_{i,j} := \mathbb{E}[T_{i,j}]$
- *Hitting time lemma*: For any *finite*, *irreducible*, *aperiodic* Markov chain, and for any two states i, j (not necessarily distinct) we have that:

$$\Pr[T_{i,j} < \infty] = 1 \quad \text{and} \quad \mathbb{E}[T_{i,j}] < \infty$$

Proof of Hitting Time Lemma

- We know that we can find $M < \infty$ such that $(P^M)_{i,j} > 0$ for all i, j, since our Markov chain is finite, irreducible and aperiodic.

Proof of Hitting Time Lemma

- We know that we can find $M < \infty$ such that $(P^M)_{i,j} > 0$ for all i, j, since our Markov chain is finite, irreducible and aperiodic.
- set $\alpha := \min_{i,j}(P^M)_{i,j}$ and for every $t \geq 0$ let X_t be the state of the Markov chain at step t

Proof of Hitting Time Lemma

- We know that we can find $M < \infty$ such that $(P^M)_{i,j} > 0$ for all i, j, since our Markov chain is finite, irreducible and aperiodic.
- set $\alpha := \min_{i,j}(P^M)_{i,j}$ and for every $t \geq 0$ let X_t be the state of the Markov chain at step t
- Note that

$$\Pr[T_{i,j} > M] \leq \Pr[X_M \neq j] \leq 1 - \alpha$$

Proof of Hitting Time Lemma

- We know that we can find $M < \infty$ such that $(P^M)_{i,j} > 0$ for all i, j, since our Markov chain is finite, irreducible and aperiodic.
- set $\alpha := \min_{i,j}(P^M)_{i,j}$ and for every $t \geq 0$ let X_t be the state of the Markov chain at step t
- Note that

$$\Pr[T_{i,j} > M] \leq \Pr[X_M \neq j] \leq 1 - \alpha$$

- Moreover, we can prove:

$$\begin{aligned} \Pr[T_{i,j} > 2M] &= \Pr[T_{i,j} > M] \cdot \Pr[T_{i,j} > 2M \mid T_{i,j} > M] \\ &\leq (1-\alpha) \cdot \Pr[X_{2M} \neq j \mid T_{i,j} > M] \\ &\leq (1-\alpha)^2 \end{aligned}$$

Proof of Hitting Time Lemma

- We know that we can find $M < \infty$ such that $(P^M)_{i,j} > 0$ for all i, j, since our Markov chain is finite, irreducible and aperiodic.
- set $\alpha := \min_{i,j}(P^M)_{i,j}$ and for every $t \geq 0$ let X_t be the state of the Markov chain at step t
- Note that

$$\Pr[T_{i,j} > M] \leq \Pr[X_M \neq j] \leq 1 - \alpha$$

- Moreover, we can prove:

$$\begin{aligned} \Pr[T_{i,j} > 2M] &= \Pr[T_{i,j} > M] \cdot \Pr[T_{i,j} > 2M \mid T_{i,j} > M] \\ &\leq (1 - \alpha) \cdot \Pr[X_{2M} \neq j \mid T_{i,j} > M] \\ &\leq (1 - \alpha)^2 \end{aligned}$$

- Iterating, we have $\Pr[T_{i,j} > \ell M] \leq (1 - \alpha)^\ell$

Proof of Hitting Time Lemma

- Iterating, we have $\Pr[T_{i,j} > \ell M] \le (1-\alpha)^{\ell}$
- Thus, we have

$$\mathbb{E}[T_{i,j}] = \sum_{n\ge 1} \Pr[T_{i,j} \ge n] = \sum_{n \ge 0} \Pr[T_{i,j} > n] \le M/\alpha < \infty$$

Fundamental Theorem of Markov Chains

- The *return time* from state i to itself is $T_{i,i}$
- *Expected return time:* defined as $\tau_{i,i} := \mathbb{E}[T_{i,i}]$.

Fundamental Theorem of Markov Chains

- The *return time* from state i to itself is $T_{i,i}$
- *Expected return time:* defined as $\tau_{i,i} := \mathbb{E}[T_{i,i}]$.

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. $$\pi_i = \lim_{t \to \infty} P^t_{i,i} = \frac{1}{\tau_{i,i}}$$

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. $$\pi_i = \lim_{t \to \infty} P^t_{i,i} = \frac{1}{\tau_{i,i}}$$

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. $$\pi_i = \lim_{t \to \infty} P^t_{i,i} = \frac{1}{\tau_{i,i}}$$

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:

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

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. $$\pi_i = \lim_{t\to\infty} P^t_{i,i} = \frac{1}{\tau_{i,i}}$$

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:

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

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

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

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 $Av = \lambda v$.

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

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*

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

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*

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

- *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.

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*
$$\rho(A) = \lim_{t \to \infty} \|A^t\|_F^{1/t}$$
- *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 $\det(tI - A)$

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*

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

- *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 $Au > (1+\varepsilon)Av$.

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

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

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 $Av > Av$. Moreover, there exists $\varepsilon > 0$ such that $Au > (1+\varepsilon)Av$.

- Note that

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

- 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

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

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

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

- 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 circumference* $|\lambda| = \rho(A)$
3. $\rho(A)$ *has geometric multiplicity 1*
4. $\rho(A)$ *has algebraic multiplicity 1*

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 circumference* $|\lambda| = \rho(A)$
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.

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 circumference $|\lambda| = \rho(A)$*
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

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

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

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

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

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

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

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

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

- By Gelfand's formula we would have

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

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 $\rho(A)u_i = (Au)_i > 0$

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 circumference $|\mu| = \rho(A)$ 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 circumference $|\mu| = \rho(A)$ is $\rho(A)$
- If we had another eigenvalue $\lambda \neq \rho(A)$ in the circumference $|\mu| = \rho(A)$, where z is the eigenvector 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 $1 \leq i \leq n$

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 circumference $|\mu| = \rho(A)$ is $\rho(A)$
- If we had another eigenvalue $\lambda \neq \rho(A)$ in the circumference $|\mu| = \rho(A)$, where z is the eigenvector 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

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

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.

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?).

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-Frobenius)

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-Frobenius)

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-Frobenius)

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

Acknowledgement

- Lecture based largely on:
  - Lap Chi's notes
  - [Motwani & Raghavan 2007, Chapter 6]
  - [Häggström]
- 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.

Häggström, Olle (2002)
Finite Markov Chains and Algorithmic Applications
Cambridge University Press