Agenda

What have we done?

- Probabilistic thinking!
- Balls and Bins
- Probabilistic Method
- Foundations of DTMC

Next

- Applications of DTMC

Example 1: a Randomized 2-SAT Algorithm

2-SAT Problem: given a 2-CNF formula φ, find a satisfying truth assignment

A Randomized Algorithm for 2-SAT

1. Repeat the following at most m times
2. Pick a truth assignment t at random
3. If $t(\varphi) \neq \text{TRUE}$, then choose a non-satisfying clause, flip a literal
4. Else ($t(\varphi) = \text{TRUE}$), return t
5. Return not satisfiable after m steps

Analysis of the Algorithm

- If there's no satisfying truth assignment, we're OK.
- Suppose there's one truth assignment $\bar{t}$
- At step i, let Y_i be the Hamming distance from t to $\bar{t}$
- Then, we want $Y_i = 0$ for some i. We know

$$\begin{aligned} \text{Prob}[Y_{i+1} = k-1 \mid Y_i = k] &\geq 1/2 \\ \text{Prob}[Y_{i+1} = k+1 \mid Y_i = k] &\leq 1/2 \end{aligned}$$

- $\{Y_i\}_{i\geq 0}$ is generally not a Markov chain
- Define a Markov chain $\{X_i\}_{i\geq 0}$, state space $I = \{0, 1, \cdots, n\}$

$$\begin{aligned} \text{Prob}[X_{i+1} = k-1 \mid X_i = k] &= 1/2 \\ \text{Prob}[X_{i+1} = k+1 \mid X_i = k] &= 1/2 \end{aligned}$$

Analysis of the Algorithm

- $\{Y_i\}_{i\geq 0}$ "leans left" heavier than $\{X_i\}_{i\geq 0}$
- Expected number of steps until some $Y_i = 0$ is at most expected number of steps until some $X_i = 0$
- Key: we do know how to compute the expected number of steps until $\{X_i\}_{i\geq 0}$ "hits" $\{0\}$.
- The mean hitting times $\mu_i = \mu_i^{\{0\}}$ are minimal non-negative solutions to the following

$$\begin{aligned}
\mu_0 &= 0 \\
\mu_i &= 1 + \frac{1}{2}(\mu_{i-1} + \mu_{i+1}),\ 1 \leq i \leq n-1 \\
\mu_n &= 1 + \mu_{n-1}
\end{aligned}$$

- Induction gives $\mu_{n-i} = n^2 - i^2$

Analysis of the Algorithm

- Let X be the number of steps until 0 is reached
- Conditioning on the initial state, $\mathsf{E}[X] \leq n^2$
- By Markov

$$\mathsf{Prob}[X \geq m] \leq \frac{n^2}{m} = \frac{1}{2}$$

  for $m = 2n^2$
- Run independently k times, error probability is reduced to $\frac{1}{2^k}$

Example 2: Undirected s, t-Connectivity (UNSTCONN)

UNSTCONN

Given a graph G, two vertices s and t. Is there an s, t-path?

The Low-Space Solution

Start from s, keep walking on the graph randomly for some time. If t is found: return YES, otherwise return NO.

Main question

How many steps must be taken so that the false negative probability is at most ϵ?

Need some basic results on random walks on graphs

Random Walks on Graphs – Basic Observations

Definition

$G = (V, E)$ a finite and connected undirected graph. A random walk on G is a Markov chain on V where $p_{uv} = 1/\deg(u)$

Basic observations

- The walk is irreducible and recurrent, has an invariant distribution

$$\pi_v = \frac{\deg(v)}{2m}, \quad m = |E|.$$

- Thus, it is positive recurrent
- The walk is aperiodic iff G is not bipartite, in which case π is the limit distribution

We will assume G is non-bipartite henceforth

Random Walks on Graphs – Basic Parameters

- $\mu_{u,v}$: hitting time (or *mean hitting time*) is the expected number of steps to hit v starting from u
- $\kappa(u, v) = \mu_{u,v} + \mu_{v,u}$: commute time
- Cover time $\mathcal{C}(G)$ is the expected number of steps until all nodes are visited. If no starting node is specified, we mean the worst case, i.e. starting from the node with worst cover time.
- Mixing rate measures how fast the walk converges (defined precisely later)

Question

Determine the hitting times and cover time for a random walk on K_n

Random Walks on Graphs – Basic Results

Lemma

$\mu_{v,v} = \frac{1}{\pi_v} = \frac{2m}{\deg(v)}$

Lemma

For any edge $uv \in E(G)$, $\kappa(u, v) \leq 2m$

Theorem (Cover time bound)

$\mathcal{C}(G) \leq 2m(n - 1)$

Back to the UNSTCONN Problem

- If there's no s, t-path, the algorithm returns NO correctly
- If there's a path, the walk hits t in expected time at most $\mathcal{C}(G) \leq 2mn < 2n^3$
- Applying Markov as usual, a walk of length $4n^3$ is sufficient to make false positive probability $\leq 1/2$
- Important to note: the algorithm uses only $O(\log n)$-space, so
  $$\text{UNSTCONN} \in \mathbf{RL} \text{ (randomized log-space)}$$
- In 2004, Omer Reingold showed a beautiful result that the problem can be solved *deterministically* in log-space, implying $\mathbf{L} = \mathbf{SL}$
- His result makes use of *expanders* which we'll discuss next

Another Proof of Convergence to Equilibrium

- Let us first assume G is d-regular, $\mathbf{A}$ is the adjacency matrix
- Transition probability matrix for the random walk is the normalized adjacency matrix $\hat{\mathbf{A}} = \frac{1}{d}\mathbf{A}$
- Both $\hat{\mathbf{A}}$ and $\mathbf{A}$ are real and symmetric
- λ is an eigenvalue of $\mathbf{A}$ iff $\hat{\lambda} = \lambda/d$ is an eigenvalue of $\hat{\mathbf{A}}$ with the same eigenvector
- The set of eigenvalues of $\mathbf{A}$ is called the spectrum of the graph G

Spectral Theorem for Real and Symmetric Matrices

Theorem (Spectral Theorem for Real and Symmetric Matrices)

Let $\mathbf{A}$ be any real and symmetric $n \times n$ matrix, then there is an orthogonal matrix $\mathbf{Q}$ (columns are orthogonal) such that

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$$

where $\mathbf{\Lambda}$ is a real diagonal matrix with entries $(\lambda_1, \ldots, \lambda_n)$.
In particular, the columns $\mathbf{q}_1, \cdots, \mathbf{q}_n$ of $\mathbf{Q}$ are orthogonal eigenvectors of $\mathbf{A}$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$

Properties of Spectrum of a Graph

Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the spectrum of a d-regular graph G, then

- $\mathbf{1}$ is a d-eigenvector, i.e. $\mathbf{A1} = d\mathbf{1}$
- $\lambda_1 = d$
- $|\lambda_i| \leq d$ for all i
- G is connected if and only if $\lambda_i < d$ for all $i \geq 2$
- G is not bipartite if and only if $\lambda_n \neq -d$

Thus, if G is connected and not bipartite, its spectrum is

$$d = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n > -d$$

In particular,

$$\lambda(G) = \max\{|\lambda_2|, |\lambda_n|\} < \lambda_1 = d$$

Another Proof of Convergence for Random Walks on G

- The uniform distribution $\mathbf{u} = \mathbf{1}/n$ is a 1-eigenvector of $\hat{\mathbf{A}}$, thus the uniform distribution is an invariant distribution of the random walk.

$$\hat{\lambda}(G) = \max\{|\hat{\lambda}_2|, |\hat{\lambda}_n|\} < \hat{\lambda}_1 = 1$$

- Let $\mathbf{u}_2, \dots, \mathbf{u}_n$ be the other orthogonal eigenvectors of $\hat{\mathbf{A}}$, then for any initial distribution π,

$$\pi = c_1\mathbf{u} + c_2\mathbf{u}_2 + \cdots c_n\mathbf{u}_n$$

- Since π is a distribution, $\langle \pi, \mathbf{u}\rangle = 1/n$, implying $c_1 = 1$. Thus,

$$\pi = \mathbf{u} + c_2\mathbf{u}_2 + \cdots c_n\mathbf{u}_n$$

- Consequently, we obtain another proof of the convergence theorem:

$$\hat{\mathbf{A}}^k\pi = \mathbf{u} + c_2\lambda_2^k\mathbf{u}_2 + \cdots c_n\lambda_n^k\mathbf{u}_n$$

Summary: Large Spectral Gap ⇒ Fast Convergence

- G: finite, connected and non-bipartite
- Then, $|\hat{\lambda}_i| \leq \hat{\lambda}(G) < 1, \forall i \geq 2$.
- Moreover, we proved

$$\hat{\mathbf{A}}^k \pi = \mathbf{u} + c_2 \lambda_2^k \mathbf{u}_2 + \cdots c_n \lambda_n^k \mathbf{u}_n$$

which implies

$$\lim_{k \to \infty} \hat{\mathbf{A}}^k \pi = \mathbf{u},$$

for any initial distribution π of the random walk.

- $1 - \hat{\lambda}(G)$ is called the spectral gap (equivalently, $d - \lambda(G)$)
- Large spectral gap ⇒ fast convergence

How to Measure Convergence Speed?

- Want to know how far $\pi^{(k)} = \hat{\mathbf{A}}^k \pi$ is from $\mathbf{u}$
- Could try l_1, l_2, l_∞ norms, but what do they mean probabilistically?

Definition (Total Variation Distance)

Given two distributions P and Q on a countable sample space Ω, the total variation distance between P and Q is the largest possible difference in probabilities that the two distributions can assign to the same event. Namely,

$$\|P - Q\| = \sup_{A \subset \Omega} |P(A) - Q(A)|,$$

where $P(A) = \sum_{\omega \in A} P(\omega)$, $Q(A) = \sum_{\omega \in A} Q(\omega)$.

Variation Distance on Countable State Space

Lemma

When Ω is countable,

$$\|P - Q\| = \frac{1}{2} \sum_{\omega \in \Omega} |P(\omega) - Q(\omega)|$$

In other words, the distance is half the l_1-norm of the vector $P - Q$.

Convergence Speed (Mixing Rate) of Random Walks

Theorem

Let G be finite, connected, non-bipartite, and d-regular with $\hat{\lambda}(G) < 1$. Then, for any initial distribution π we have

$$\|\hat{\mathbf{A}}^k \pi - \mathbf{u}\|_1 \leq \sqrt{n}\hat{\lambda}^k.$$

Proof.

$$\|\hat{\mathbf{A}}\pi - \mathbf{u}\|_2 = \|\hat{\mathbf{A}}(\pi - \mathbf{u})\|_2 \leq \hat{\lambda}\|\pi - \mathbf{u}\|_2 \leq \hat{\lambda}$$

Inductively,

$$\|\hat{\mathbf{A}}^k \pi - \mathbf{u}\|_2 \leq \hat{\lambda}^k \|\pi - \mathbf{u}\|_2 \leq \hat{\lambda}^k$$

Cauchy-Schwartz completes the proof. □

(Note: there's also a notion of convergence in entropy.)

Mixing Lemma

Lemma (Expander Mixing Lemma)

Let $G = (V, E)$ be a d-regular graph on n vertices, $\lambda = \lambda(G)$, $\hat{\lambda} = \hat{\lambda}(G)$. Let $(S, T) = \{(u, v) \mid u \in S, v \in T\}$ (set of ordered pairs). Then, $\forall S, T \subseteq V$,

$$\left| |(S,T)| - \frac{d|S||T|}{n} \right| \leq \lambda \sqrt{|S||T|}$$

Simple Consequences, when $\hat{\lambda} < 1$

- Maximum independent set size of G is at most $\hat{\lambda} n$
- Chromatic number is at least $1/\hat{\lambda}$
- Diameter is $O(\log n)$

Spectral Expansion and Edge Expansion

- Edge boundary $\partial(S)$ of $S \subset V$:

$$\partial(S) = |(S, \bar{S})|$$

- Edge expansion ratio $h(G)$ of G:

$$h(G) = \min_{S \subset V, |S| \leq n/2} \frac{|\partial S|}{|S|}$$

also called Cheeger constant or Cheeger number of G

Theorem (Connection between edge expansion and spectral gap)

Let G be d-regular with spectrum $\lambda_1 \geq \cdots \geq \lambda_n$. Then,

$$\frac{d - \lambda_2}{2} \leq h(G) \leq \sqrt{2d(d - \lambda_2)}$$

Expanders: Finally!

Three definitions which are more or less equivalent. (We only consider regular expanders for simplicity.)

Definition (Spectral Expander)

A d-regular graph G is called an α-spectral expander if $\hat{\lambda}(G) \leq \alpha$.

Definition (Edge Expander)

A d-regular graph G is called an β-edge expander if $h(G) \geq \beta$.

Definition (Vertex Expander)

A d-regular graph G is called an γ-vertex expander if

$$\min_{S \subset V, |S| \leq n/2} \frac{|\Gamma(S)|}{|S|} \geq \gamma.$$

They are More or Less Equivalent

We have seen a relationship between spectral expansion and edge expansion before. We connect spectral vs. vertex expansion below.

Lemma

An α-spectral expander is also a $\frac{2}{\alpha^2+1}$-vertex expander

Lemma

A β-vertex expander is also an α-spectral expander with

$$\alpha = \sqrt{1 - \frac{(\beta-1)^2}{d^2(8+4(\beta-1)^2)}}$$

It is straightforward to connect vertex vs. edge expansions.

Family of Expanders

Definition

A sequence of d-regular graphs $\{G_i\}_{i=1}^{\infty}$ is a family of spectral expanders if there exists $\epsilon > 0$ such that $\hat{\lambda}(G_i) \leq 1 - \epsilon$ for all i.
$n_i = |V(G_i)|$ are required to be strictly increasing.

(Families of vertex- and edge-expanders are defined similarly.)
Intuitively, good families of expanders (i.e. usable for most applications) satisfy the following

- $\{n_i\}$ is not increasing too fast (e.g., $n_{i+1} \leq n_i^2$ is good)
- the G_i can be generated in polynomial time

Good Families of Expanders

Let $\{G_i\}$ be a family of d-regular expanders, where $\{n_i\}$ is increasing but not too fast

- the family is called mildly explicit if there's an algorithm generating the ith graph G_i in time polynomial in i
- the family is called very explicit if there's an algorithm which, on inputs i and $v \in [n_i]$ and $k \in [d]$, computes the kth neighbor of v in G_i in time polynomial in (the binary representation of) the input (i, v, k)

Example: Margulis Construction

Margulis (1973) constructed the following expander family

For every integer m,

- $V(G_m) = \mathbb{Z}_m \times \mathbb{Z}_m$
- neighbors of (x, y) are $(x + y, y), (x - y, y), (x, y + x), (x, y - x), (x + y + 1, y), (x - y + 1, y), (x, y + x + 1), (x, y - x + 1)$ (all operations done in the ring $\mathbb{Z}_m$)
- this is a family of very explicit 8-regular expanders

Applications of Expanders and Random Walks on Expanders

Numerous

- Constructing good topologies for P2P networks
- Taking double cover of expanders gives bipartite expanders (remember magical graph from the first weak): can be used to construct good error-correcting codes, superconcentrators, concentrators, good interconnection networks, etc.
- Construct parallel sorting networks of size $O(n \lg n)$ (a huge result!)
- Efficient error reduction in probabilistic algorithms
- Metric embedding
- PCP Theorem and many other results in complexity theory (remember Reingold's result)
- ...

(See Bulletin of the AMS Survey by Hoory, Linial, and Wigderson)

Application: Efficient Error Reduction

Some problems whose most efficient solutions are randomized algorithms:

- Primality testing: is the input a prime?
- Polynomial identity checking: is $P(x)Q(x) \equiv R(x)$ for given polynomials P, Q, R under some finite field
- Matrix indentity checking: is $\mathbf{AB} = \mathbf{C}$ for matrices on finite fields

These are examples of decision problems Π which have a randomized poly-time (Monte Carlo) algorithm A satisfying the following:

- On input x of size n, A uses $r = r(n)$ random bits

$$
\begin{aligned}
x \in \Pi_{\text{YES}} &\implies \operatorname*{Prob}_{s \in \{0,1\}^r}[A(x,s) = \text{YES}] \geq 1/2 \\
x \in \Pi_{\text{NO}} &\implies \operatorname*{Prob}_{s \in \{0,1\}^r}[A(x,s) = \text{NO}] = 1
\end{aligned}
$$

$\mathbf{RP}$ is the complexity class consisting of these kinds of problems

The Straightforward Way to Reduce Error Probability

- Pick k random strings $s_1, \dots, s_k \in \{0,1\}^r$
- Return NO only if all $A(x, s_i)$ say NO
- False negative probability is $(1/2)^k$
- Number of random bits used is kr

Another view of this method

- For any input x, let B_x be the set of strings $s \in \{0,1\}^r = \Omega$ for which $A(x,s)$ gives the wrong answer for x
- If $x \in \Pi_{\text{YES}}$, $|B_x| \leq |\Omega|/2$
- The algorithm sample k independent points from Ω
- Probability that all k points are bad (i.e. in B_x) is at most $(1/2)^k$

Sampling by Random Walk on Expanders

- Let $G = (V, E)$ be a very explicit d-regular α-spectral expander where $V = \Omega$, where $\alpha < 1/2$
- Instead of sampling k independent points, start from a uniformly random vertex s_0 of G and take a random walk of length k: $s_0, s_1, \dots, s_k$
- Run $A(x, s_i)$ and return NO only if all $A(x, s_i)$ says NO
- Number of random bits used is $r + k \log_2 d = r + O(k)$
- Error probability equal the probability that all $s_0, \dots, s_k$ stay inside B_x

Theorem (Ajtai-Komlós-Szemerédi, 1987)

Let $G = (V, E)$ be an α-spectral expander. Consider $B \subset V$, $|B| \leq \beta|V|$. The probability that a random walk of length k (with uniformly chosen initial vertex) stays inside B the entire time is at most $(\alpha + \beta)^k$.

Proof of AKS Theorem

- Let P be the orthogonal projection onto the coordinates B, i.e. $P = (p_{ij})$ where $p_{ij} = 1$ iff $i = j \in B$, $p_{ij} = 0$ otherwise.
- The probability that the length-k walk stays in B is

$$\|(P\hat{\mathbf{A}})^k P\mathbf{u})\|_1$$

  (conditioned on $s_0 \in B$, apply Chapman-Kolmogorov equation)
- Next, for any vector $\mathbf{v}$,

$$\|P\hat{\mathbf{A}}P\mathbf{v})\|_2 \leq (\alpha + \beta)\|\mathbf{v}\|_2$$

- Finally,

$$\begin{aligned}
\|(P\hat{\mathbf{A}})^k P\mathbf{u})\|_1 &\leq \sqrt{n}\|(P\hat{\mathbf{A}})^k P\mathbf{u})\|_2 \\
&= \sqrt{n}\|(P\hat{\mathbf{A}}P)^k \mathbf{u})\|_2 \\
&\leq \sqrt{n}(\alpha + \beta)^k \|u\|_2 \\
&= (\alpha + \beta)^k
\end{aligned}$$

How Big can the Spectral Gap be?

- The best expander is K_n, whose spectrum is

$$[n-1, -1, -1, \dots, -1]$$

- However, we are interested in cases where $n \gg d$. In this case,

Theorem (Alon-Boppana)

If G is d-regular with n vertices, then,

$$\lambda(G) \geq 2\sqrt{d-1} - o_n(1)$$

- A d-regular graph G is called a Ramanujan Graph if $\lambda(G) \leq 2\sqrt{d-1}$
- Amazingly: Ramanujan graphs can be constructed explicitly when $d-1$ is any prime power. (Using Cayley graphs of projective linear groups.)

Many Other Known Constructions

- Most random graphs are expanders
- Most random graphs **are** Ramanujan graphs!!!
- Many explicit constructions based on Cayley graphs
- Zig-Zag Product! (Lead to Reingold's result)

TBD