18.409: An Algorithmist’s Toolkit

Lecture 4

Jonathan Kelner

Today

- Random walks on graphs
- Start approximate counting

Administrivia

- Reminder: sign up for scribing
- Scribe notes for lectures 2 and 3 posted
- TA: Mayank Varia (mhv@mit.edu)
- He’ll coordinate the scribe notes
  - Please e-mail him for old LaTeX and send him your notes when they’re done

Random Walks

- Let $G=(V,E)$ be an undirected graph
- Start at some vertex $v$
- Repeat:
  - Move to a uniformly random neighbor of the current vertex
- See Matlab
Probability Vectors
- Let $p_t(v)$ denote the probability you’re at vertex $v$ at time $t$
- Can think of $p_t$ as a vector in $\mathbb{R}^n$
- $\sum_{i \in V} p_t(i) = 1$

How Probability Vectors Change
- If you’re at vertex $u$ at time $t$, you’re at each neighbor $v$ of $u$ at time $t+1$ with probability $1/d(u)$
- So
  \[ p_{t+1}(v) = \sum_{(u,v) \in E} \Pr[\text{at } u \text{ at time } t] \cdot \Pr[\text{go from } u \text{ to } v \text{ at time } t+1] \]
  \[ = \sum_{(u,v) \in E} p_t(u) \cdot \frac{1}{d(u)} \]

Walk Matrices
- Can write this as a matrix $W = W_G$:
  \[ [W_G]_{i,j} = \begin{cases} \frac{1}{d(j)} & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases} \]
- Note unfortunate ordering: $[W_G]_{i,j}$ is prob of going from $j$ to $i$
- $W_G = A D^{-1}$

Stationary Distribution
- Let $\pi(u) = \frac{d(u)}{\sum_{v \in V} d(v)}$
- Claim 1: $\pi$ is a probability vector
  Proof: blackboard
- Claim 2: $W \pi = \pi$
  Proof: blackboard
- So if walk has prob distribution given by $\pi$ at step $t$, it will have same distribution at step $t+1$
- Means $\pi$ is an eigenvector of eigenvalue 1
Convergence

- See Matlab examples
- A natural guess is that, if we walk around long enough, we eventually end up arbitrarily close to the stationary distribution
- Not true!
  - See Matlab examples

What’s the Problem

- We have a bipartite graph
- If we are on the left at time 0, we’ll always be on the left at even times and right at odd times
  - So can’t ever converge to stationary distribution
- Easy fix: a lazy random walk
  - At time t:
    - Take a step of the random walk with prob 0.5
    - Stay where you are with prob 0.5
  Can show this breaks periodicity
  Gives new matrix \( W' = (W + \text{Id})/2 \)

A Better Matrix

\[
W = AD^{-1}, \quad W' = (\text{Id} + AD^{-1})/2
\]

- W and W’ aren’t symmetric, which will make things complicated
- Let’s define new matrices that are:
  - The *normalized walk matrix* is
    \[
    N = D^{-1/2} W D^{1/2} = D^{-1/2} A D^{-1/2}
    \]
  - The *normalized lazy walk matrix is*
    \[
    N' = D^{-1/2} W' D^{1/2} = (\text{Id} + D^{-1/2} A D^{-1/2})/2
    \]

More on the Normalized Walk Matrix

- **Claim:** N and W have the same eigenvalues and related eigenvectors
- **Proof and precise statement:** Blackboard
- So normalized walk matrix has eigenvector \( D^{1/2} \mathbf{1} \) with eigenvalue 1
- Likewise with \( N' \) and \( W' \)
Connections to Laplacians

- We’ve used Laplacian L.
- I’ve mentioned the existence of the normalized Laplacian \( L' = D^{-1/2} L D^{-1/2} \)
  - Used in Chung’s book
- Claim: \( N = \text{Id} - L' \)
- Proof: Blackboard
- So eigenvalues of \( N \) are given by
  - 1- (eigenvalues of \( L' \))
  - Makes sense to order them the opposite way:
    \[ 1 = \mu_1 \geq \mu_2 \geq \ldots \geq \mu_n \]
  - Can now translate over theorems about Laplacians to get theorems about these \( \mu \)s.

\( \ell_2 \) Convergence

- Let the **spectral gap** \( \lambda := 1 - \mu_2' \)
- Define \( \ell_2 \) distance between prob dists \( p \) and \( q \) by \( ||p - q||_2 = \sum (p_i - q_i)^2 \)
- **Theorem**: Let \( p_0 \) be an initial probability distribution, and let \( p_t \) be the distribution after \( t \) steps of the lazy random walk.
  \[ ||p_t - \pi||_2 \leq (1 - \lambda)^t \sqrt{\frac{\max_x d(x)}{\min_y d(y)}} \]
  - Will do regular case on blackboard
  - Irregular case is similar

\( \ell_\infty \) Convergence

- **Theorem**: For any \( v \),
  \[ |p_t(v) - \pi(v)| \leq (1 - \lambda)^t \sqrt{\frac{d(v)}{\min_y d(y)}} \]
- Pretty much the identical proof
  - I’ll leave it as an exercise

Restating Our Old Theorems

- We now get the following statements:
  - All \( \mu \) are in [-1,1]
  - If G is connected then \( \mu_2 < 1 \)
    - Implies unique stable distribution
  - The -1 eigenvalues only occur for bipartite graphs (exercise)
- Let \( \mu_i' \) be the \( i \)th eigenval of \( N' \)
  - All \( \mu_i' \) are in [0,1]
  - If G is connected, \( \mu_i' < 1 \)
Cheeger’s inequality carries over too. The isoperimetric number is replaced with a new quantity that we’ll call the conductance $\Phi$.

Note that literature isn’t too consistent on conductance vs. isoperimetric number.

**Definition:** For a subset $S \subseteq V$, the conductance of the corresponding cut is

$$\Phi(S) = \frac{e(S)}{\min \left( \sum_{v \in S} d(v), \sum_{v \notin S} d(v) \right)}$$

$$\Phi(G) = \min_{S \subseteq V} \Phi(S)$$

Cheeger’s Inequality

Cheeger’s inequality now becomes

$$\Theta(1) \Phi^2(G) \leq 1 - \mu^2 \leq \Theta(1) \Phi(G)$$

So $\Phi(G)$ related to rate of convergence.

Proving bounds on $\Phi$ lets us prove a walk mixes quickly.

Some Pictures

Why should cuts be related to mixing?
Everyone’s Favorite Example

- Compute $\pi$ by throwing darts at a dartboard

- If you randomly pick a point in the square, probability get one in the circle = area(circle)/area(square) = $\pi/4$
- Suppose you pick $n$ points in $[-1,1]^2$
- $E[number in unit circle] = n \pi/4$
- So return that $\pi = (\text{#points in circle})*4/n$
- **Question:** How close will this be to the right answer?

Chernoff Bounds

- Suppose we have a random variable $r \in \{0,1\}$ s.t. $\Pr[r=0] = p$, $\Pr[r=1] = (1-p)$
- Draw $n$ independent samples $r_1, \ldots, r_n$
- Let $R = \sum r_i$
- $E[R] = E[\sum r_i] = \sum E[r_i] = np$
- We’ll say “$R \epsilon$-approximates $E[R]$” if $(1-\epsilon)E[R] \leq R \leq (1+\epsilon)E[R]$
- So it’s a multiplicative error measure

Chernoff Bounds (cont.)

- **Theorem [one version of Chernoff bound]:** The probability $R$ fails to $\epsilon$-approximate $E[R]$ is
  $$\Pr[|R - E[R]| \geq \epsilon R] \leq 2e^{-n\epsilon^2/12}$$
  $$= 2e^{-E[R]\epsilon^2/12}$$
- Some notes about this:
  - Pretty near tight
  - Need independent trials, or else no longer true
  - Multiplicative error, not additive
  - For fixed $\epsilon$, falls off exponentially in $n$
  - So if you have failure probability 0.5, can get failure probability 1/2^4 by performing $m = nk$ trials

More on Chernoff Bounds

- $\Pr[|R - E[R]| \geq \epsilon R] \leq 2e^{-n\epsilon^2/12}$
  $$= 2e^{-E[R]\epsilon^2/12}$$
- Smaller $n$ requires more trials
- If you want an $\epsilon$-approximation with probability 1-\delta, need
  $$N \geq \Theta\left(\frac{\log \delta}{\epsilon^2}\right)$$
- In words: need enough trials to get $\Theta(\log \delta / \epsilon^2)$ successes
Back to the Dartboard

- So, if we want to estimate \( \pi \) within, say, 5%, with probability at least 0.99:
  - \( \varepsilon = 0.05 \)
  - \( \delta = 1/100 \)
  - Need
    \[
    N \geq \Theta \left( \frac{\log(1/100)}{(\pi/4)(0.05)^2} \right)
    \]
  - Easy to make \( \delta \) smaller, harder to make \( \varepsilon \) smaller

When Do We Run Into Trouble?

- If we’re bad at darts
  - Big dartboard, small circle
  - If \( p = 1/\text{exponential} \), need exponentially many trials to expect constant number of successes
- If we have trouble throwing darts at all
  - That is, if it is hard to draw samples uniformly from ambient space
- We’ll run into both problems and develop some techniques for (sometimes) fixing them

A Little Terminology

- An \((\varepsilon, \delta)\) \textbf{approximation scheme} for some quantity is an algorithm for computing an \(\varepsilon\)-approximation with probability at least \(1-\delta\)
- \textbf{A fully polynomial randomized approximation scheme (FPRAS)} is an \((\varepsilon,\delta)\) approximation scheme that runs in time \(\text{poly}(n,1/\varepsilon, \log 1/\delta)\)

A Nontrivial Example: Counting DNF Solutions

- Have boolean variables \(x_1,\ldots,x_n\)
  - “Literal” means an \(x_i\) or its negation
- Recall that a DNF formula \(F\) is a disjunction (OR) of conjunctive (AND) clauses:
  \[F = C_1 \lor \ldots \lor C_m\]
  where each \(C_i\) is an AND of literals, e.g.,
  \[F = (x_1 \land \overline{x}_3) \lor (x_2) \lor (x_2 \land \overline{x}_1 \land x_3 \land x_4)\]
Counting DNF Solutions

\[ F = (x_1 \land \overline{x_3}) \lor (x_2) \lor (x_2 \land \overline{x_1} \land x_3 \land x_4) \]

- Easy to check if there exists satisfying assignment
- Really hard to exactly count them
  - \#P hard (so at least as hard as NP-complete problems)
  - We’ll see how to, in poly time, get an \( \varepsilon \)-approximation with probability > 1-1/2^n
    - I.e., an FPRAS

How Not to Do It

- Naive Monte Carlo algorithm
  - One trial:
    - Randomly assign 0 or 1 to each \( x_i \)
    - Check if this is a satisfying assignment
  - Repeat \( m \) times, and say get \( k \) successes
  - Estimate
    - \# solutions \( \approx (k/m) \cdot 2^n \)
  - Why not?
    - \#solutions/ \( 2^n \) can be exponentially small
    - So need exponentially many trials...

How to Fix the Problem
• Want to count yellow boxes
• We’ll:
  1. Sample \( X = \{ \text{blue} \} \cup \{ \text{yellow} \} \)
  2. Count number yellow we get
  3. Multiply by \(|X|\)
• Now success prob \( \geq 1/m \)
• Sample by col instead of by row
• If clause \( i \) has \( k_i \) vars in it, its col has \( 2^{n-k_i} \) satisfying assigs (and we know what they are)
• So sample each col proportional to \( 2^{n-k_i} \)
• Then pick a random sat. assig. of this clause and see if it’s first in row
• \(|X| = \sum 2^{n-k_i}\)