COMPSCI 614: Randomized Algorithms with Applications to Data Science

Prof. Cameron Musco

University of Massachusetts Amherst. Spring 2024.

Lecture 24 (Final Lecture!)

Logistics

- Optional Problem Set 5 due 5/13 at 11:59pm.
- Final exam will be **Tuesday 5/14, 10:30-12:30pm in the classroom**. See Piazza post for info on study materials.
- I will hold additional final review office hours **Monday 5/13 from 3-4:30pm**.
- Final project due the last day of finals: Friday 5/17 – if you have questions as you come into the last couple of weeks of the project feel free to reach out.
- Please fill our SRTIs when you get a chance!

Summary

Last Time: Convex relaxation and randomized rounding.

- High level idea of convex relaxation for approximating NP-hard problems.
- Deterministic rounding for vertex cover. Randomized rounding for set cover.
- SDP relaxation and hyperplane rounding for max-cut (Goemans-Williamson algorithm)

Today: The Probabilistic Method (not on the exam)

- From probabilistic proofs to algorithms via the method of conditional expectations.
- The Lovasz local lemma for events with 'bounded' correlation.

The Probabilistic Method

The Basic Idea: Suppose we want to prove that a combinatorial object satisfying a certain property exists. Then it suffices to exhibit a random process that produces such an object with probability > 0.

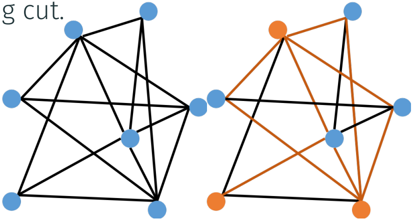
We have already seen examples of this – e.g. the JL Lemma and Newman's Theorem reducing private coin communication complexity to public coin communication complexity (Problem Set 2).

A common tool: For a random variable with $\mathbb{E}[X = \mu]$, $\Pr[X \geq \mu] > 0$ and $\Pr[X \leq \mu] > 0$.

Example 1: Max-Cut

Prove that for any graph with m edges, there exists a cut containing at least $m/2$ edges.

Consider a random partition of the nodes (each node is included independently in each half with probability 1/2). Let X be the size of the corresponding cut.



We have $\mathbb{E}[X] =$

Therefore, $\Pr[X \geq m/2] > 0$. So every graph with m edges has a cut containing at least $m/2$ edges.

Example 2: 3-SAT

Prove that for any 3-SAT formula, there is some assignment of the variables such that at least 7/8 of the clauses are true.

Consider a random assignment of the variables. And let X be the number of satisfied clauses.

$$(x_1 \vee \bar{x}_2 \vee x_4) \wedge (x_2 \vee \bar{x}_4 \vee x_3) \wedge (x_2 \vee x_3 \vee x_4) \wedge \ldots$$

What is $\mathbb{E}[\mathsf{X}]$?

So, $\Pr[\mathsf{X} \geq 7/8m] > 0$. So there is an assignment satisfying at least 7/8 of the clauses in every 3-SAT formula.

From Existence to Efficient Algorithms

Simple Max-Cut Approximation: A randomly sampled partition cuts $m/2$ edges in expectation. But how many partitions do we need to sample before finding a cut of size at least $m/2$ with good probability?

Let p be the probability of finding a cut of size $\geq m/2$. Then:

$$\mathbb{E}[X] = \frac{m}{2} \leq (1-p) \cdot \left(\frac{m}{2} - 1\right) + p \cdot m$$
$$\implies \frac{1}{\frac{m}{2}+1} \leq p.$$

How many attempts do we need to take to find a large cut with probability at least $1-\delta$? $O(m \cdot \log(1/\delta))$

We can also derandomize this algorithm in a very simple way.

Let $x_1, x_2, \ldots \in \{0,1\}$ indicate if the vertices are included on one side of the random partition.

Consider determining these randsom variables sequentially.

$$\frac{m}{2} = \mathbb{E}[X] = \frac{1}{2}\mathbb{E}[X|x_1 = 1] + \frac{1}{2}\mathbb{E}[X|x_1 = 0].$$

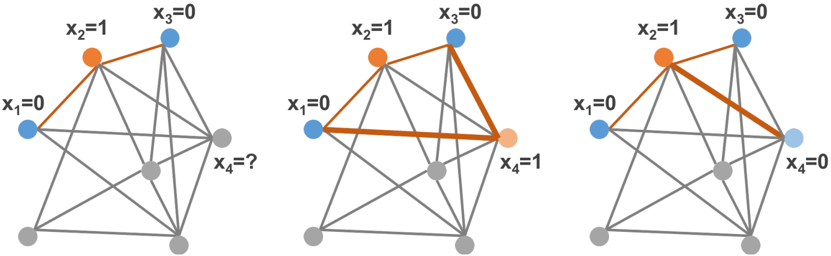
Set $x_1 = v_1$ such that $\mathbb{E}[X|x_1 = v_1] \geq \frac{m}{2}$ Then we have:

$$\frac{m}{2} \leq \mathbb{E}[X|x_1 = v_1] = \frac{1}{2}\mathbb{E}[X|x_1 = v_1, x_2 = 1] + \frac{1}{2}\mathbb{E}[X|x_1 = v_1, x_2 = 0]$$

Set $x_2 = v_2$ such that $\mathbb{E}[X|x_1 = v_1, x_2 = v_2] \geq \frac{m}{2}$. And so on...

Conditional Expectations for Cuts

How can we pick v_i such that $\mathbb{E}[X | x_1 = v_1, \ldots, x_{i-1} = v_{i-1}] \geq \frac{m}{2}$?



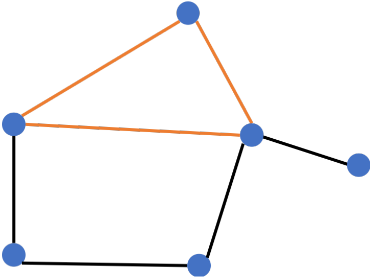
$\mathbb{E}[X | x_1 = 0, \ldots, x_4 = 1] = \frac{1}{2} \cdot 10 + 2 = 7 \mathbb{E}[X | x_1 = 0, \ldots, x_4 = 0] = \frac{1}{2} \cdot 10 + 1 = 6$

Natural greedy approach: add vertex i to the side of the cut to which it has fewest edges.

Yields a 1/2 approximation algorithm for max-cut. Recall that 16/17 is the best possible assuming $P \neq NP$, and .878 is the best known

Large Girth Graphs

The girth of a graph is the length of its shortest cycle.

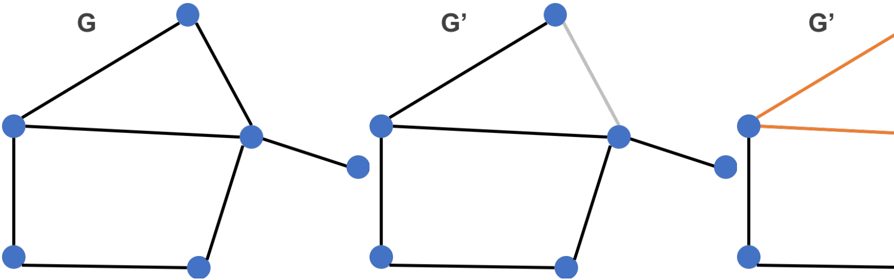


Natural Question: How large can the girth be for a graph with m edges?

Erdös Girth Conjecture: For any $k \geq 1$, there exists a graph with $m = \Omega(n^{1+1/k})$ edges and girth $2k + 1$.

Relevance to Spanners

A spanner is a subgraph that approximately preserves shortest path distances. We say G' is a spanner for G with stretch t if for all u, v $d_{G'}(u,v) \leq t \cdot d_G(u,v)$.



Even when G' excludes a single edge, $t \geq girth(G) - 1$.

Erdös Girth Conjecture $\implies$ there are no generic spanner constructions with $o(n^{1+1/k})$ edges and stretch $\leq 2k - 1$.

Large Girth Graphs via Probabilistic Method

Theorem (Weaker Version of Girth Conjecture)

For any fixed $k \geq 3$, there exists a graph with n nodes, $\Omega(n^{1+1/k})$ edges, and girth $k+1$.

Sample and Modify Approach: Let G be an Erdös-Renyi random graph, where each edge is included independently with probability $p = n^{1/k-1}$. Remove one edge from every cycle in G with length $\leq k$, to get a graph with girth $k+1$.

Let X be the number of edges in the graph and Y be the number of cycles of length $\leq k$. Suffices to show $\mathbb{E}[X - Y] = \Omega(n^{1+1/k})$.

$$\mathbb{E}[X] = \frac{n(n-1)}{2} \cdot p = \frac{1}{2} \cdot \left(1 - \frac{1}{n}\right) \cdot n^{1+1/k}.$$

$$\mathbb{E}[Y] = \sum_{i=3}^{k} \binom{n}{i} \cdot \frac{(i-1)!}{2} \cdot p^i \leq \sum_{i=3}^{k} n^i p^i = \sum_{i=3}^{k} n^{i/k} < k \cdot n.$$

Large Girth Graphs via Probabilistic Method

So far: An Erdös-Renyi random graph with $p = n^{1/k-1}$ has expected number of edges (X) and cycles of length $\leq k-1$ (Y) bounded by:

$$\mathbb{E}[X] = \frac{1}{2} \cdot \left(1 - \frac{1}{n}\right) \cdot n^{1+1/k}$$
$$\mathbb{E}[Y] < k \cdot n.$$

When k is fixed and n is sufficiently large, $k \cdot n \ll n^{1+1/k}$. Thus,

$$\mathbb{E}[X - Y] = \Omega(\mathbb{E}[X]) = \Omega(n^{1+1/k}),$$

proving the theorem.

Lovasz Local Lemma

Probabilities of Correlated Events

Suppose we want to sample a random object that avoids n 'bad events' $E_1, \ldots, E_n$.

E.g., we want to sample a random assignment for variables that satisfies a a k-SAT formula with n clauses. E_i is the event that clause i is not satisfied.

If the E_i are independent, and $\Pr[E_i] < 1$ for all i then:

$$\Pr\left[\neg \bigcup_{i=1}^{n} E_i\right] = \prod_{i=1}^{n}(1 - E_i) > 0.$$

What if the events are not independent?

If $\sum_{i=1}^{n} \Pr[E_i] < 1$ then by a union bound,

$$\Pr\left[\neg \bigcup_{i=1}^{n} E_i\right] \geq 1 - \sum_{i=1}^{n} E_i > 0.$$

As n gets large, the union bound gets very weak – each event has to occur with probability $< 1/n$ on average.

Bounded Correlation

Consider events $E_1, \ldots, E_n$ where E_i is independent of any $j \notin \Gamma(i)$ (the neighborhood of i in the dependency graph)

E.g., consider randomly assigning variables in a k-SAT formula with n clauses, and let E_i be the event that clause i is unsatisfied.

$$(x_1 \vee \bar{x}_2 \vee x_3) \wedge (x_2 \vee \bar{x}_4 \vee x_3) \wedge (x_4 \vee x_5 \vee x_6) \wedge (\neg x_4 \vee x_6 \vee x_7) \ldots$$

Theorem (Lovasz Local Lemma)

Suppose for a set of events $E_1, E_2, \ldots, E_n$, $\Pr[E_i] \leq p$ for all i, and that each E_i is dependent on at most d other events E_j (i.e., $|\Gamma(i)| \leq d$, then if $4dp \leq 1$:

$$\Pr\left[\neg \bigcup_{i=1}^{n} E_i\right] > (1-2p)^n > 0.$$

In the worse case, $d = n - 1$ and this is similar to the union bound. But it can be much stronger.

LLL Application: k-SAT

Theorem

If no variable in a k-SAT formula appears in more than $\frac{2^k}{4k}$ clauses, then the formula is satisfiable.

Let E_i be the event that clause i is unsatisfied by a random assignment. $\Pr[E_i] \leq \frac{1}{2^k} = p$.

$|\Gamma(i)| \leq k \cdot \frac{2^k}{4k} = \frac{2^k}{4} = d$

So $4dp = 4 \cdot \frac{1}{2^k} \cdot \frac{2^k}{4} \leq 1$, and thus $\Pr\left[\neg \bigcup_{i=1}^{n} E_i\right] > 0$. I.e., a random assignment satisfies the formula with non-zero probability.

Algorithmic LLL

Important Question: Given an Lovasz Local Lemma based proof of the existence, can we convert it into an efficient algorithm?

Moser and Tardos [2010] prove that a very natural algorithm can be used to do this.

Let $E_1, \ldots, E_n$ be events determined by a set of independent random variables $V = \{v_1, \ldots, v_m\}$. Let $v(E_i)$ be the set of variables that E_i depends on.

Resampling Algorithm:

1. Assign $v_1, \ldots, v_m$ random values.
2. While there is some E_i that occurs, reassign random values to all varables in $v(E_i)$.
3. Halt when an assignment is found such that no E_i occurs.

Algorithmic LLL

Theorem (Algorithmic Lovasz Local Lemma)

Consider a set of events $E_1, E_2, \ldots, E_n$ determined by a finite set of random variables V. If for all i, $\Pr[E_i] \leq p$ and $|\Gamma(i)| \leq d$, and if $ep(d+1) \leq 1$, then Resampling *finds an assignment of the variables in V such that no event E_i occurs. Further, the algorithm makes $O(\frac{n}{d})$ iterations in expectation.*

Application to k-SAT: Consider a k-SAT formula where no variable appears in more than $\frac{2^k}{5k}$ clauses. Let E_i be the event that clause i is unsatisfied by a random assignment

$$\Pr[E_i] \leq \frac{1}{2^k} = p \quad \text{and} \quad |\Gamma(i) \leq k \cdot \frac{2^k}{5k} = \frac{2^k}{5} = d.$$

Have $ep(d+1) \leq \frac{e}{5} + \frac{e}{2^k} \leq 1$ as long as $k \geq 3$, so the theorem applies, giving a polynomial time algorithm for this variant of k-SAT.

Thanks for a great semester!