Advanced Algorithms

Lecture 9: Average Case Analysis

Lecturer: Nathan Klein

1 Recap

So far, we have focused on *worst-case* analysis, because we wanted to give algorithms that work on all possible inputs.

While worst-case analysis is an important paradigm, often there is a divide between what we can prove in the worst case and how algorithms perform in the real world. There are many examples of massive instances of NP-Hard problems being solved in the real world, even to optimality. To give a famous example, back in 2006, an instance of the traveling salesperson problem (which is, remember, NP-Hard) with 85,900 cities was solved to optimality. Notably, this instance was not somehow hand-crafted to ensure it could be solved: it arose from an actual problem in chip design.

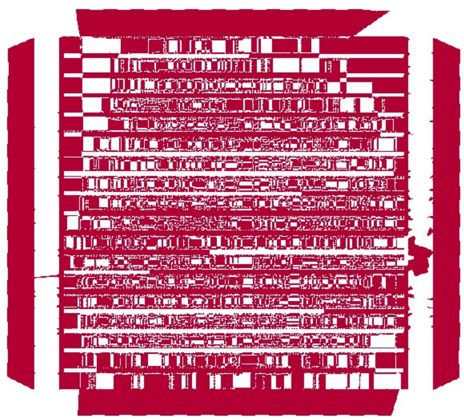

Figure 1: The 85,900 city instance solved in 2006.

This is not really something you would expect for an NP-Hard problem, right? We usually hear things like: to solve a 100-city instance of the TSP, it would take a supercomputer thousands of years. And that's true if the supercomputer needs to do 2^{100} operations, which is roughly what would be needed using the fastest exact algorithms today. And yet, here we are, having solved an instance with 85,900 cities. So it's reasonable to hope that for an *average* instance, things are much better.

2 Average Case Analysis

Since we have focused quite a bit on Max Cut, we're going to analyze Max Cut on random graphs. Remember that the best known approximation algorithm for Max Cut is about 0.878, and it is APX-Hard, so there is no PTAS unless P=NP.

2.1 Erdös-Réyni Model

First, we need to define what a "random graph" actually is in our application. This is one of the drawbacks of average case: often, it's not clear what distribution real-world examples come from. But we can still come up with reasonable average case models and prove things about them.

Definition 2.1 ($G(n,p)$). *In an Erdös-Réyni random graph with n vertices and parameter $p \in [0,1]$, denoted $G(n,p)$, a graph with n vertices $v_1, \ldots, v_n$ is constructed adding each possible edge $\{v_i, v_j\}$ independently with probability p.*

2.2 Average Case for Max Cut: Algorithm

For simplicity, we will focus on the case where $p = \frac{1}{2}$. However, similar results can be proved for any value of p. For the next few lemmas, let $G = (V, E)$ be drawn from $G(n, \frac{1}{2})$.

First, let's recall Hoeffding's inequality from last class for binary random variables.

Theorem 2.2 (Hoeffding's Inequality for Binary Random Variables). *Let $X_1, \ldots, X_n$ be mutually independent binary random variables and $X = \sum_{i=1}^{n} X_i$, $\mathbb{E}[X] = \mu$. Then,*

$$\mathbb{P}[X - \mu \geq t] \leq e^{\frac{-2t^2}{n}}$$

$$\mathbb{P}[|X - \mu| \geq t] \leq 2e^{\frac{-2t^2}{n}}$$

We can use this to show that the graph has about $\frac{n^2}{4}$ edges with high probability.

Fact 2.3. *The number of edges in G is at least $\frac{n(n-1)}{4} - n^{1.5}$ with probability at least $2e^{-4n}$.*

Proof. The expected number of edges is $\frac{1}{2}\binom{n}{2} = \frac{n(n-1)}{4}$. So, by Hoeffding,

$$\mathbb{P}\left[|E| - \mu \leq -n^{1.5}\right] \leq \mathbb{P}\left[||E| - \mu| \geq n^{1.5}\right] \leq 2e^{-2n^3/(n^2/2)} = 2e^{-4n} \qquad \square$$

We can now run our randomized $\frac{1}{2}$-approximation, which recall returned a cut with $|E|/2$ edges in expectation, or we can run a deterministic algorithm (from your homework) which will always return a cut with at least $|E|/2$ edges.

Upshot: In polynomial time, we can obtain a cut with at least $\frac{1}{2}(\frac{n(n-1)}{4} - n^{1.5}) \geq \frac{n^2}{8} - n^{1.5} = (\frac{1}{8} - \frac{1}{\sqrt{n}})n^2$ edges with probability $1 - 2e^{-4n}$.

2.3 Average Case for Max Cut: Upper Bound

Now we will show that in fact the graph has *no* cut with more than $\frac{n^2}{8} + n^{1.5}$ edges with high probability.

Fact 2.4. *For any set $S \subseteq V$,*

$$\mathbb{E}[|\delta(S)|] \leq \frac{n^2}{8}$$

Proof.

$$\mathbb{E}\left[|\delta(S)|\right] = \sum_{e=\{v_i,v_j\}: i\in S, j\notin S} \mathbb{P}\left[e \in E\right] = |S||V \setminus S|\frac{1}{2} \leq \frac{n^2}{8}$$

where in the inequality we use that $\max_{0\leq k\leq n} k(n-k) = \frac{n^2}{2}$. □

Fact 2.5. *For any set $S \subseteq V$,*

$$\mathbb{P}\left[|\delta(S)| \geq \frac{n^2}{8} + \epsilon n^2\right] \leq e^{-8\epsilon^2 n^2}$$

Proof. $|\delta(S)|$ is the sum of independent Bernoulli random variables, so Hoeffding applies. Note that the number of Bernoullis is $|S||V \setminus S| \leq \frac{n^2}{4}$.

$$\begin{aligned}\mathbb{P}\left[|\delta(S)| \geq \frac{n^2}{8} + \epsilon n^2\right] &\leq \mathbb{P}\left[|\delta(S)| - \mu \geq \epsilon n^2\right] && \text{Since } \mu \leq \frac{n^2}{8} \\ &\leq e^{-2\epsilon^2 n^4/(n^2/4)} \leq e^{-8\epsilon^2 n^2}\end{aligned}$$

as claimed. □

We can now employ the union bound to demonstrate that in fact *all* cuts do not exceed $(\frac{1}{8} + o(1))n^2$ edges with high probability. Remember the union bound:

Lemma 2.6 (Union Bound)**.** *Let $A_1, \ldots, A_n$ be a collection of events. Then,*

$$\mathbb{P}\left[\bigcup_{i=1}^{n} A_i\right] \leq \sum_{i=1}^{n} \mathbb{P}\left[A_i\right]$$

Lemma 2.7. *The probability there exists a cut with more than $(\frac{1}{8} + \frac{1}{\sqrt{n}})n^2$ edges is at most e^{-6n}.*

Proof. There are 2^n sets $S \subseteq V$. Let A_S be the event that $|\delta(S)| \geq (\frac{1}{8} + \frac{1}{\sqrt{n}})n^2$ edges. By Fact 2.5, where we set $\epsilon = \frac{1}{\sqrt{n}}$, $\mathbb{P}\left[A_i\right] \leq e^{-8\epsilon^2 n^2} = e^{-8n}$. So, by the union bound,

$$\mathbb{P}\left[\text{exists a cut with more than } (\frac{1}{8} + \frac{1}{\sqrt{n}})n^2 \text{ edges}\right] = \mathbb{P}\left[\bigcup_{S\subseteq V} A_i\right] \leq \sum_{S\subseteq V} \mathbb{P}\left[A_i\right] \leq 2^n e^{-8n}$$ □

Therefore, we are overwhelmingly likely to see no cut with more than this many edges.

2.4 Approximation Guarantee

To obtain our result, we just need to put these two pieces together.

Theorem 2.8. *The deterministic algorithm for Max Cut from the homework which returns a cut with at least $|E|/2$ edges is a $1 - o(1)$ approximation with probability at least $1 - 2^{-n}$ on $G(n, \frac{1}{2})$. The randomized algorithm for Max Cut from class is with probability $1 - 2^{-n}$ a randomized $1 - o(1)$ approximation.*

Proof. Apply the union bound over the event that $\frac{|E|}{2} \le (\frac{1}{8} - \frac{1}{\sqrt{n}})n^2$ and the event that no cut T has $|\delta(T)| \ge (\frac{1}{8} + \frac{1}{\sqrt{n}})n^2$. This implies with probability at least $1 - 2^{-n}$, neither event occurs. In this case, the approximation factor is at least:

$$\frac{(\frac{1}{8} - \frac{1}{\sqrt{n}})n^2}{(\frac{1}{8} + \frac{1}{\sqrt{n}})n^2} = 1 - o(1)$$

in particular, this is a $1 - O(\frac{1}{\sqrt{n}})$ approximation with probability at least $1 - 2^{-n}$. Put another way, for all but an exponentially small fraction of graphs, this is a $1 - o(1)$ approximation. □

2.5 Finding a Certificate

One slightly awkward thing about this algorithm is that *we don't know whether it worked.* In the TSP example, the solvers actually had to produce a proof (sometimes called certificate) that there was no better solution.

So, one would hope that in this case you can also find a proof that there is no cut with value more than $(\frac{1}{2} + o(1))|E|$ with high probability. That turns out to be true as well.

Definition 2.9 (Spectral Norm of a Matrix)**.** *Given a matrix $A \in \mathbb{R}^{n \times n}$, the spectral norm $\|A\|_2$ is its largest singular value. If A is symmetric, this is equivalent to its largest eigenvalue in absolute value.*

On the homework, you will prove the first part of the following fact. The second part follows similarly.

Fact 2.10. *For a symmetric matrix A,*

$$\lambda_1 = \min_{x \neq \mathbf{0}, x \in \mathbb{R}^n} \frac{x^T A x}{x^T x}$$

$$\|A\|_2 = \max\{|\lambda_i|\} = \max\{|\lambda_1|, |\lambda_n|\} = \max_{x \neq \mathbf{0}, x \in \mathbb{R}^n} \frac{|x^T A x|}{x^T x}$$

which implies that for all $x \in \mathbb{R}^n$, we have

$$|x^T A x| \le \|A\|_2 \cdot \|x\|^2$$

Recall for a graph with vertices $v_1, \ldots, v_n$ its adjacency matrix $A \in \{0,1\}^{n \times n}$ is a matrix where $A_{ij} = 1$ if there is an edge between v_i and v_j and 0 otherwise.

For a cut $S \subseteq V$, let $\mathbf{1}_S \in \{0,1\}^n$ be the vector for which $\mathbf{1}_i = 1$ if $v_i \in S$ and 0 otherwise.

Fact 2.11.

$$\mathbf{1}_S^T A \mathbf{1}_{V \setminus S} = |\delta(S)|$$

Proof.

$$\mathbf{1}_S^T A \mathbf{1}_{V \setminus S} = \sum_{i,j \in [n] \times [n]} A_{ij} \mathbb{I}\{v_i \in S\} \mathbb{I}\{v_j \notin S\} = |\delta(S)|$$

□

Lemma 2.12. *Let $G = (V, E)$ with adjacency matrix A. Let J be the matrix of all 1s. Then, where S is the max cut of G:*

$$|\delta(S)| \leq \frac{n^2}{8} + \frac{n}{2}\|A - J/2\|_2$$

Proof.

$$\begin{aligned} |\delta(S)| - \frac{n^2}{8} &= \mathbf{1}_S^T A \mathbf{1}_{V \setminus S} - \frac{n^2}{8} && \text{By Fact 2.11} \\ &\leq \mathbf{1}_S^T (A - J/2) \mathbf{1}_{V \setminus S} && \text{Since } \mathbf{1}_S^T J \mathbf{1}_{V \setminus S} \leq n^2/4 \end{aligned}$$

Now, we can almost relate this back to the Rayleigh quotient, but we have a slight hiccup in that we don't have the same vector being multiplied on the left and right. We'll prove the Rayleigh quotient still applies in Corollary 2.15, so we get a bound on this of:

$$\begin{aligned} &\leq \|A - J/2\|_2 \cdot \|\mathbf{1}_S\|_2 \cdot \|\mathbf{1}_{V \setminus S}\|_2 \\ &\leq \|A - J/2\|_2 \cdot \sqrt{|S|}\sqrt{|V \setminus S|} \\ &\leq \|A - J/2\|_2 \cdot \frac{n}{2} \end{aligned}$$

□

Before we prove Corollary 2.15, let's see how this enables us to finish the proof. First, we will use the following matrix concentration bound. You can prove it by realizing that the sum of a bunch of independent matrices concentrates similarly to the sum of independent random variables. Unfortunately, we don't quite have time to prove it in this class, although I encourage you to check out Luca Trevisan's course for more on this theorem[1] and the field of average case analysis.

Theorem 2.13 ([FK81])**.** *With probability $1 - o(1)$, for $G \sim G(n, \frac{1}{2})$, if A is the adjacency matrix of G,*

$$\|A - J/2\|_2 \leq O(\sqrt{n})$$

But the point is that we know we can *compute* the spectral norm in polynomial time, and with high probability this spectral norm will be at most $O(\sqrt{n})$, meaning that we get a proven upper bound of $\frac{n^2}{8} + n^{1.5}$ with high probability.

Average Case for Max Cut

We have shown that almost all graphs are easy to solve Max Cut on. If you pick one at random from the Erdös-Réyni model for $p = \frac{1}{2}$, you are going to be able to get essentially an optimal solution in polynomial time as well as a proof that it's near optimal. The situation is similar for all values of p.

So, yes, the problem is APX-Hard. But the instances that you can't get within a $1 - o(1)$ factor of optimality are very rare. In the real world, you are probably unlikely to encounter them. However, be careful! Not all problems are like this, and again, it can be difficult to even define a "random" instance. What is, for example, a random image found on the internet? It is definitely not the result of randomly picking the color of each pixel, which would be the natural analog of $G(n, p)$.

[1] See Lecture 6.

2.6 Extra Fact

Here we'll show that it's not a big deal that we are not multiplying by the same vector on the lefthand side and righthand side in Lemma 2.12.

Lemma 2.14. *Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Then for all $x, y \in \mathbb{R}^n$ with $\|x\| = \|y\| = 1$, we have*

$$|x^T Ay| \leq \|A\|_2$$

Proof. Notice that since A is symmetric, we have the identity:

$$x^T Ay = \frac{1}{4}((x+y)^T A(x+y) - (x-y)^T A(x-y))$$

Now, we can apply Fact 2.10 to obtain

$$\begin{aligned} |x^T Ay| &= |\frac{1}{4}((x+y)^T A(x+y) - (x-y)^T A(x-y))| \leq \frac{1}{4}\|A\|_2(\|x+y\|^2 + \|x-y\|^2) \\ &= \frac{1}{2}\|A\|_2(\|x\|^2 + \|y\|^2) = \|A\|_2 \end{aligned}$$

□

Corollary 2.15. *Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Then for all $x, y \in \mathbb{R}^n$,*

$$|x^T Ay| \leq \|x\| \cdot \|y\| \cdot \|A\|_2$$

Proof. Apply the above lemma to the vectors $x/\|x\|$ and $y/\|y\|$ (if either has norm 0, the lemma follows immediately) to obtain:

$$|x^T Ay| = \|x\|\|y\| \cdot |\frac{x}{\|x\|}^T A \frac{y}{\|y\|}| \leq \|x\| \cdot \|y\| \cdot \|A\|_2$$

□

References

[FK81] Z. Füredi and J. Komlós. "The eigenvalues of random symmetric matrices". In: *Combinatorica* 1.3 (1981), pp. 233–241. ISSN: 1439-6912. DOI: 10.1007/BF02579329 (cit. on p. 5).