COMPSCI 632: Approximation Algorithms	October 2, 2017

Lecture 10

Lecturer: Debmalya Panigrahi *Scribe: Aaron Lowe*

1 Overview

Last class we saw the congestion minimization problem. Given a graph $G = (V, E)$ and set of n pairs of vertices (s_i, t_i) we want to choose paths connecting the pairs, so that the number of times any edge appears along a path is minimized.

We had the following LP

$$\begin{aligned} &\min \lambda && \text{such that} \\ &\sum_{p \in P_i} X_p \geq 1 && \text{and} \\ \sum_i &\sum_{p \in P_i, e \in p} x_p \leq \lambda && \text{for each edge } e \\ &x_p \geq 0 \end{aligned}$$

We saw a randomized rounding algorithm in which for each set of paths P_i we choose one path, each with probability proportional to x_p. We finish analyzing the algorithm in this lecture, and show a randomized rounding algorithm for approximating MAX-SAT as well.

2 Congestion Minimization Continued

2.1 Chernoff Bounds

In this subsection, we will see two chernoff bounds (without proof) which we will use to analyze our randomized algorithm for congestion minimization.

Theorem 1. *If $X_1, \ldots, X_n$ are independant random boolean variables (each takes on value 0 or 1), with*

$$\mu = \sum_i \mathbb{E}[X_i]$$

then,

$$Pr\left[\sum_i X_i \geq (1+\varepsilon)\mu\right] \leq \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{(1+\varepsilon)}}\right)^{\mu} \tag{1}$$

$$Pr\left[\sum_i X_i \geq (1+\varepsilon)\mu\right] \leq e^{-\varepsilon^2\mu/c} \qquad \textit{if } 0 < \varepsilon < 1, \textit{ for some constant } c \tag{2}$$

2.2 Application to Minimum Congestion

Let

$$X_i^e \begin{cases} 1 & \text{if } (s_i,t_i) \text{ path chosen contains } e \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\mathbb{E}\left[X_i^e\right] = \sum_{p \in P_i \mid e \in p} X_p.$$

Since $\lambda \geq \mu$

$$\Pr\left[\sum_i X_i^e \geq (1+\varepsilon)\lambda\right] \leq \left(\frac{e^\varepsilon}{(1+\varepsilon)^{(1+\varepsilon)}}\right)^\lambda$$
$$\leq \left(\frac{1}{(\frac{1+\varepsilon}{e})^{1+\varepsilon}}\right)^\lambda$$

Since there are $O(n^2)$ possible edges, we want this bound to be $1/n^c$ for $c \geq 3$ so we can use a union bound over all edges. If we don't know anything about λ we will choose $(1+\varepsilon) = O(\frac{\log n}{\log\log n})$, giving us an $O(\frac{\log n}{\log\log n})$-approximation.

However suppose $\lambda \geq \log n$, then to get the desired bound we only need $(1+\varepsilon)$ to be some constant, giving a constant approximation.

3 MAX-SAT

For the max-sat problem, we are given a collection of m clauses $\{C_j\}$ containing n literals. Each clause is a disjunction of literals: for example we might have $C_1 = (X_1 \vee \overline{X_2} \vee X_3)$. The goal is to give an assignment (true or false) to each literal which maximizes the number of clauses which are satisfied, that is evaluate to true.

A naive randomized algorithm would be to assign each literal X_i true with probability 1/2. If clause C_j has ℓ_j literals,

$$\Pr\left[C_j \text{ is satisfied}\right] = 1 - 1/2^{\ell_j}.$$

In expectation at least $m/2$ clauses will be satisfied, giving a 2-approximation. We observe that as clauses become longer, this bound becomes better.

Remark 1. *In fact for 3-SAT it can be shown that one cannot do better than the 7/8-approximation given by this algorithm, unless $P = NP$.*

However if every clause contained only 1 literal, we could get the optimal solution by assigning each literal to true if it appears positively in more clauses, otherwise false if it appears as a negation in more clauses. This suggests that we need to utilize the bound on opt to get a better bound on our approximation.

We can write a LP for this problem, letting z_j be a variable denoting if C_j is satisfied, y_i denote if literal X_i is true, P_j be the set of positive literals in C_j and N_j the negative.

$$\max \sum z_i$$
$$\sum_{i \in P_j} y_i + \sum_{i \in N_j} (1 - y_i) \geq z_j \text{ for each } j$$
$$0 \leq y_i \leq 1$$
$$0 \leq z_j \leq 1$$

Given the fractional optimal solution to the LP we can round this by setting x_i to true with probability y_i^*. Then we have

$$\Pr[C_j \text{ is sat.}] \geq 1 - \prod_{i \in P_j} (1 - y_i^*) \prod_{i \in N_j} y_i^*$$

The arithmetic-mean geometric-mean inequality says that

$$\frac{\sum_{i=1}^n a_i}{n} \geq \left(\prod_{i=1}^n a_i \right)^{(1/n)}.$$

Using this we have

$$\begin{aligned}
\Pr[C_j \text{ is sat.}] &\geq 1 - \left[\frac{1}{\ell_j} \left(\sum_{i \in P_j} (1 - y_i^*) + \sum_{i \in N_j} y_i^* \right) \right]^{1/\ell_j} \\
&= 1 + \left[\frac{1}{\ell_j} \left(\sum_{i \in P_j} y_i^* + \sum_{i \in N_j} (1 - y_i^*) \right) - \frac{|P_i|}{\ell_j} - \frac{|N_j|}{\ell_j} \right]^{\ell_j} \\
&\geq 1 - \left[1 - \frac{1}{\ell_j} \left(\sum_{i \in P_j} y_i^* + \sum_{i \in N_j} (1 - y_i^*) \right) \right]^{\ell_j} \\
&\geq 1 - \left(1 - \frac{z_j}{\ell_j} \right)^{\ell_j}
\end{aligned}$$

Letting $f(z_j^*) = 1 - \left(1 - \frac{z_j}{\ell_j}\right)^{\ell_j}$ we can see that f is a concave function, with $f(0) = 0$ $f(1) = (1 - 1/\ell_j)^{\ell_j}$ so we have that

$$f(z_j^* \geq \left(1 - (1 - \frac{1}{\ell_j})^{\ell_j} \right) z_j^* \geq (1 - 1/e) z_j^*.$$

We note that the LP-algorithm does better when the clauses are short. So we will combine our two algorithms: simply run both and take the solution which has more clauses maximized.

To analyze this hybrid algorithm, we see that choosing the better of the two is at least as good as picking each algorithm with probability 1/2.

Then we have

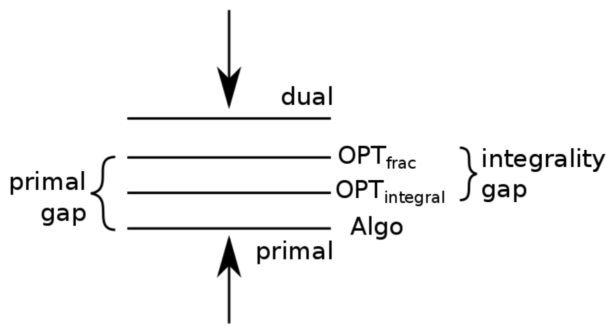
$$\Pr[C_j \text{ is sat.}] \geq 1 - \frac{1}{2} \left(\left(1 - \frac{1}{\ell_j} \right)^{\ell_j} + 2^{-\ell_j} \right)$$

So looking by cases for values of ℓ_j we see

$$\begin{aligned}
\ell_j = 1: &\quad \Pr[C_j \text{ is sat.}] = 3/4 \\
\ell_j = 1: &\quad \Pr[C_j \text{ is sat.}] = 3/4 \\
\ell_j \geq 3: &\quad \Pr[C_j \text{ is sat.}] \geq 1 - 1/2((1-1/e)+1/8) \geq 3/4
\end{aligned}$$

3.1 Integrality Gap

For problems where we construct an integer solution from rounding a fractional solution to a feasible integral solution, and use this ratio as our approximation bound, we can do no better than the ratio between the optimal fraction and optimal integral solution. This is ratio is called the *integrality gap* and bounds how well we can analyze such algorithms (or primal dual arguments as well.)



We will see that for MAX-SAT there is an integrality gap of 4/3 meaning that our algorithm is tight with the integrality-gap.

Consider a MAX-SAT problem instance with 2 literals, x_1, x_2 and 4 clauses

$$\begin{gathered}
(x_1 \vee x_2) \\
(\overline{x_1} \vee x_2) \\
(x_1 \vee \overline{x_2}) \\
(\overline{x_1} \vee \overline{x_2})
\end{gathered}$$

The integral optimal solution is 3 (indeed every assignment satisfies 3 of the 4 clauses.) However setting each variable to 1/2 gives a fractional optimal solution of 4 to the LP. Therefore the integrality gap is 4/3.

4 Summary

In this lecture we saw how Chernoff bounds can be used to give approximation factors for randomized rounding algorithms, and saw another application of randomized rounding for the MAX-SAT problem. Finally we saw how integrality gaps can be used to give bounds on how well LP based approaches can perform.