CSC 373 Lecture 29

Announcements:

As posted, weekly TA office hour Fridays 1-2 in Pratt 378.

Term Test 2 question 3 regrading final call

Course evaluations Friday or Monday?

Next assignment/test date; decision now.

Today

Randomized rounding

- ¾ approximation for IP/LP for Max-Sat
- O(log m) approximation for Set cover

The Max-Sat problem as an IP

- In the (general) weighted Max-Sat problem, we are given a CNF formula $F = C_1 \wedge C_2 \ldots \wedge C_m$ over a set of variables $x_1, \ldots, x_n$ with clause C_i having weight w_i. In contrast to Max-*k*-Sat and Exact Max-*k*-Sat, each clause can have any number of literals. Let C_j^+ (resp C_j^-) be the set of all variables occurring positively (resp. negatively) in C_j. For example, if $C_j = x_1 \vee \bar{x}_2 \vee x_3$, we have $C_j^+ = \{x_1, x_3\}$; $C_j^- = \{x_2\}$. An IP formulation of weighted Max-Sat is:
- maximize $\sum_j w_j \cdot z_j$ subject to

  $$\sum_{x_i \in C_j^+} y_i + \sum_{x_i \in C_j^-} (1-y_i) \geq z_j$$

  $$y_i \in \{0,1\} \;;\quad z_j \in \{0,1\}$$

- Here the intended meaning of z_j is that clause C_j will be satisfied and the intended meaning of y_i is that the propositional variable x_i is set true (resp. false) if $y_i = 1$ (resp 0).
- The LP relaxation is $0 \leq y_i \leq 1$, $0 \leq z_j \leq 1$; here we do want the $y_i \leq 1$ and $z_j \leq 1$ constraints. Why?

Randomized rounding the LP

- Since we have forced our fractional solutions to be in [0,1], we can think of each fractional variable as a probability. Then we can do randomized rounding.
- Let $\{y^*_i, z^*_j\}$ be an optimal LP solution so that the $LP\text{-}OPT = \mathrm{sum}\ w_j z^*_j$. We set $y'_i = 1$ with probability y^*_i to obtain an integral solution. We do not need to round the $\{z^*_j\}$ variables since the desired solution is a truth assignment (which will in turn determine which clauses are satisfied). Note that every rounded solution is a solution (i.e. truth assignment) but we will need to use properties of the LP solution to derive an approximation ratio.

The analysis

- Let *C_j* be a clause with *k* literals and by renaming we will assume that *C_j = (x_1 v x_2 …v x_k)*. We are focusing on this one clause so say say x_1 occurred negatively I C_j, we introduce a new variable v_1 to represent {\bar x_1} and then change all occurences of x_1 to be the appropriate occurrence of v_1.
- Let *b_k = 1- (1-1/k)^k*. We will show the *Prob[C_j satisfied (in the rounded solution)]* is at least *b_k z*_j*. By linearity of expectations, the contribution (in expectation) to the rounded solution of a clause *C_j* having *k* literals is then at least *b_k w_j*. (Recall that the LP-OPT is sum_j w_j z*_j) Since *(1-1/k)^k < 1/e* (and converges to *1/*e with *k*), the approx ratio is *>= 1-1/e > .632*. (We will need one further idea to obtain a (3/4) ratio.)

Arithmetic-Geometric mean

- In the analysis, we will need to make use of the arithmetic geometric mean inequality which states that for non negative real values:
- $(1/k)\ \{a_1 + a_2 + \ldots a_k\} \geq k$ th root of the *product* $(a_1 * a_2 * \ldots a_k)$ or equivalently $[(1/k)(a_1 + \ldots + a_k)]^k \geq (a_1 * a_2 \ldots * a_k)$

Analysis continued

- Let C_j be a k literal clause and by renaming assume $C_j = x_1 \vee x_2 \ldots \vee x_k$. C_j is satisfied if not all of the y_i are set to 0 (when we set $y_i = 1$ with probability y^*_i).
- The probability that C_j is satisfied is then $[1 - \prod_i (1-y^*_i)]$.
- By the arithmetic-geometric mean inequality this probability is then at least
  $1 - [(1/k) \{(1-y^*_1) + \ldots + (1-y^*_k)\}]^k$
  $= 1 - [(1-(1/k) (y^*_1 + \ldots y^*_k)]^k \geq 1 - (1-(z^*_j/k)^k$
  where the inequality is by the LP constraint:

  $\sum_{y_i \in C_j^+} y_i + \sum_{y_i \in C_j^-} (1-y_i) \geq z_j$
  (keeping in mind the renaming making literals positive) so that we just have $\sum_{y_i \in C_j^+} y_i$. Hence $y^*_1 + \ldots + y^*_k \geq z^*_j$.

End of analysis for Max-Sat

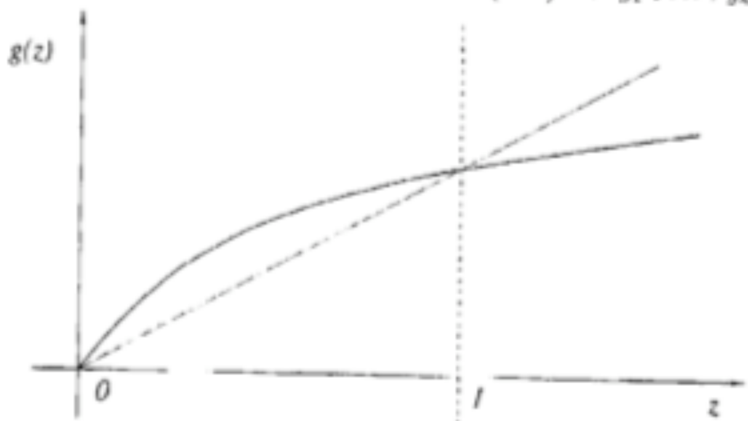
- Define $g(z) = 1 - (1 - z/k)^k$; then $g(z)$ is a concave function with $g(0) = 0$ and $g(1) = b_k$.
- By concavity, $g(z) >= (b_k) z$ for all $0 <= z <= 1$. In particular, $g(z^*) >= b_k z^*$
- Hence if C_j is a clause with k literals, then the $Prob[C_j \text{ satisfied}] >= (b_k) z^*_j$
- Like the more naive randomized alg (used for exact Max-*k*-Sat), this algorithm can also be de-randomized (by solving 2n LPs) to obtain a *(1-1/e)* approximation.
- Since the naïve alg is good for big *k* clauses and the *(1-1/e)* alg is good for small *k* clauses, it turns out (with a little more work) that by taking the best of these two deterministic algorithms, we get a (3/4) approximation.

$$1-\prod_{i=1}^{k}(1-y_i) \geq 1-\left(\frac{\sum_{i=1}^{k}(1-y_i)}{k}\right)^{k} = 1-\left(1-\frac{\sum_{i=1}^{k} y_i}{k}\right)^{k}$$
$$\geq 1-\left(1-\frac{z_c^*}{k}\right)^{k},$$

where the first inequality follows from the arithmetic-geometric mean inequality which states that for nonnegative numbers $a_1, \ldots, a_k$,

$$\frac{a_1+\ldots+a_k}{k} \geq \sqrt[k]{a_1 \times \ldots \times a_k}.$$

The second inequality uses the constraint in LP (16.2) that $y_1+\ldots+y_k \geq z_c$.



Define function g by:

$$g(z) = 1-\left(1-\frac{z}{k}\right)^{k}.$$

This is a concave function with $g(0) = 0$ and $g(1) = \beta_k$. Therefore, for $z \in [0,1]$, $g(z) \geq \beta_k z$. Hence, $\mathbf{Pr}[c$ is satisfied$] \geq \beta_k z_c^*$. The lemma follows. □

Notice that β_k is a decreasing function of k. Thus, if all clauses are of size at most k,

$$\mathbf{E}[W] = \sum_{c \in C} \mathbf{E}[W_c] \geq \beta_k \sum_{c \in C} w_c z_c^* = \beta_k \mathrm{OPT}_f \geq \beta_k \mathrm{OPT}.$$

Set cover IP/LP randomized rounding

There is a very natural and efficient greedy algorithm for solving the weighted set cover problem with approximation h_d where $d = max\_i |S\_i|$. But we want to use this problem to give a final example of IP and randomized rounding. The following randomized algorithm will with high probability produce a cover that is within a factor $O(H\_d) = O(log\ m)$ of the optimum where m is the size of the universe. This is also an opportunity to (re)introduce a little more probability.

There is also a connection between a primal dual approach solving the LP relaxation and the natural deterministic greedy algorithm that achieves approximation ratio *H_d* but we will not have time to discuss primal dual algorithms.

The IP/LP randomized rounding

- The IP is to $\min \sum_i w_i x_i$

  subj to $\sum_{\{i: u_j \in S_i\}} x_i >= 1$

  $x_i \in \{0,1\}$ for IP; $x_i >= 0$ for LP
- We solve this LP

  and find an optimal solution $\{x^*_1, \ldots, x^*_n\}$.

  We know that $x^*_i <= 1$ since in an optimal solution, each x^*_i is at most 1.

We treat the x^*_i values as probabilities and choose S_i (to be in our set cover) with probability x^*_i. This is a covering problem and the chosen sets will most likely not be a cover. So we will have to repeat this process enough times to have a good probability that all elements are covered.

The analysis

- It is easy to calculate the expected cost of the "partial cover" *C'* of sets selected by the LP optimum. Namely,

E*[cost(C')] = sum w_i Prob[S_i is chosen]*
= sum w_i x*_i = OPT-LP

- Now we need to calculate the probability that a given *u_j = u* is not covered. Lets say that *u* occurs in sets *S_1, …, S_k*. The LP solution must satisfy the constraint : *sum_{i: u in S_i} x*_i >=1*.