# CSC 373: Algorithm Design and Analysis
# Lecture 25

Allan Borodin

March 20, 2013

# Announcements and Outline

## Announcements

- Lecture this Friday; slower pace for rest of term.
- Next week in tutorial, we will go over all the basic probability concepts that are needed for this part of the course.
- There are many texts which will have a short section on basic probability concepts; for example, see section 13.12 of the Kleinber and Tardos text.

## Today's outline

- Max-Sat as an IP
- Randomized rounding
- Set cover and randomized rounding

# Randomized rounding – The weighted Max-Sat problem

**The weighted Max-Sat problem**

- Given a CNF formula $F = C_1 \wedge C_2 \wedge \ldots \wedge C_m$ 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.
- **Goal:** is to find a truth assignment that maximizes that the total weight of the satisfied clauses.

# The weighted Max-Sat problem as an IP

- Let $C_j^+$ (resp $C_j^-$) be the set of all variables occurring positively (resp. negatively) in $C_j$.
  - For example, if $C_j = x_1 \vee \bar{x}_2 \vee x_3$, then $C_j^+ = \{x_1, x_3\}$ and $C_j^- = \{x_2\}$.

**An IP formulation of weighted Max-Sat**

$$\begin{aligned} \text{maximize} \quad & \sum_{j=1}^{m} W_j z_j \\ \text{s.t.} \quad & \sum_{x_i \in C_j^+} y_i + \sum_{x_i \in C_j^-} (1 - y_i) \geq z_j \qquad \text{for each } j = 1, \ldots, m \\ & z_j, y_j \in \{0, 1\} \qquad \text{for each } j = 1, \ldots, m \end{aligned}$$

- **Intended meaning:** $z_j = 1$ indicates that clause $C_j$ is satisfied; the propositional variable $x_i$ is set true iff $y_i = 1$
- The LP relaxation is $0 \leq y_i \leq 1$ and $0 \leq z_j \leq 1$. 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
$$\{y_1^*, \ldots, y_m^*, z_1^*, \ldots, , z_m^*\}$$
be an optimal LP solution so that the LP-OPT $= \sum_j W_j z_j^*$.

- We set $\hat{y}_i = 1$ 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 can assume that $C_j = (x_1 \vee x_2 \vee \ldots \vee x_k)$.
- Let $b_k = 1 - \left(1 - \frac{1}{k}\right)^k$. We will show (next 3 slides) that

$$\mathbb{P}\big[C_j \text{ is satisfied in the rounded solution}\big] \geq z_j^* b_k$$

- By linearity of expectations, the contribution (in expectation) to the rounded solution of a clause $C_j$ having $k$ literals is then at least $W_j z_j^* b_k$. (Recall that the LP-OPT $= \sum_j W_j z_j^*$.)
- Since $(1 - 1/k)^k < 1/e$ (and converges to $1/e$ when $k \to \infty$), the approx. ratio is at least $1 - 1/e > 0.632$.
- **Note:** We will need one further idea to obtain a $3/4$ approx. ratio.

# Arithmetic-geometric mean inequality

- In the analysis, we will need to make use of the arithmetic geometric mean inequality which states that for non negative real values $a_i$:

$$\frac{a_1 + a_2 + \ldots + a_k}{k} \geq (a_1 \cdot a_2 \cdot \ldots \cdot a_k)^{1/k}$$

- Or equivalently

$$\left(\frac{a_1 + a_2 + \ldots + a_k}{k}\right)^k \geq (a_1 \cdot a_2 \cdot \ldots \cdot a_k)$$

# Analysis continued: check each setp carefully

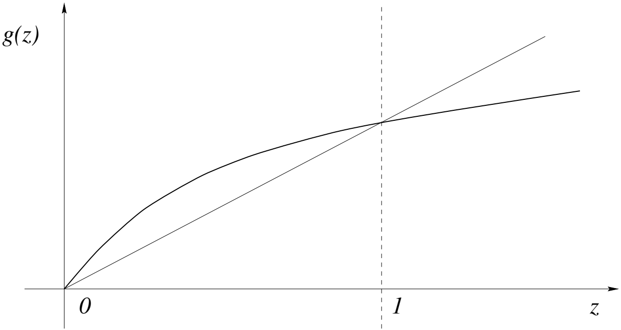
- Let $C_j$ be a clause with $k$ literals and by renaming assume $C_j = (x_1 \vee x_2 \vee \ldots \vee x_k)$.
- $C_j$ is satisfied if not all of the $y_i$ are set to 0 (after setting $y_i = 1$ with probability $y_i^*$).
- Thus, the probability that $C_j$ is satisfied is $1 - \prod_i (1 - y_i^*)$. By the arithmetic-geometric mean inequality, this probability is at least

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

- The inequality is by the LP constraint: $\displaystyle\sum_{y_i \in C_j^+} y_i + \sum_{y_i \in C_j^-} (1 - y_i) \geq z_j$
- **Note:** Keep in mind the renaming making literals positive, so that we just have $\sum_{y_i \in C_j^+} y_i$ .
- Hence it follows that $y_1^* + \ldots + y_k^* \geq z_j^*$.

# End of analysis for Max-Sat

- Define $g(z) = 1 - \left(1 - \frac{z}{k}\right)^k$, then $g(z)$ is a concave function with $g(0) = 0$ and $g(1) = b_k$.




- By concavity, $g(z) \geq b_k z$ for all $z \in [0, 1]$. In particular, $g(z^*) \geq b_k z_j^*$.
- Hence if $C_j$ is a clause with $k$ literals, then the

$$\mathbb{P}\left[C_j \text{ is satisfied in the rounded solution}\right] \geq z_j^* b_k$$

# Some concluding remarks on this Max-Sat algorithm

- Like the more naive randomized algorithm used for exact Max-$k$-Sat, this algorithm can also be derandomized (by solving at most $2n$ LPs) to obtain a $(1 - 1/e)$ approximation.

- Since the naive algorithm is good for big $k$ clauses and the $(1 - 1/e)$ approximation algorithm is good for small $k$ clauses, it turns out that by taking the best solution of these two deterministic algorithms, we get a $\frac{3}{4}$ approximation.

- This is close to the best known approximation ratio for Max-Sat.

# The Set Cover problem

## The Set Cover problem

- Given a set of elements $U = \{1, 2, \ldots, m\}$ (called the universe) and $n$ sets $S = \{S_1, S_2, \ldots, S_n\}$, whose union comprises the universe $U$.
- A subset $S' \subseteq S$ is a cover of $U$ if its union contains all elements of $U$.
- **Goal:** Find a smallest cover $C \subseteq S$.

## Example

- Assume we are given the universe $U = \{1, 2, 3, 4, 5\}$ and sets

$$S = \Big\{\{1, 2, 3\}, \{2, 4\}, \{3, 4\}, \{2, 4, 5\}\Big\}.$$

- Clearly the union of all the sets in $S$ contains all elements in $U$.
- However, we can cover all of the elements with the following, smaller number of sets $\big\{\{1, 2, 3\}, \{2, 4, 5\}\big\}$.

# The weighted Set Cover problem

### The weighted Set Cover problem

- Given a set of elements $U = \{1, 2, \ldots, m\}$ (called the universe) and $n$ sets $S = \{S_1, S_2, \ldots, S_n\}$, whose union comprises the universe $U$.
- Each set $S_i$ now has a weight $w_i$
- **Goal:** Find a cover $C \subseteq S$ with minimum total weight.

# 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 $d = max_i |S_i|$.
- Recall that $O(H_m) = O(\log m)$, where $m$ is the size of the universe.
- But we want to use this problem to give a final example of IP and randomized rounding.
- Note that in the randomized Max-Sat algorithms, we never had to worry about whether or not a solution was feasible since every truth assignment is feasible. The only issue was the approximation ratio.
- 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.
- This is also an opportunity to (re)introduce a little more probability.

# The IP/LP randomized rounding

## An IP formulation of weighted Set Cover

$$\text{minimize} \sum_{i=1}^{n} w_i x_i$$

$$\text{s.t.} \sum_{i\,:\,j \in S_i} x_i \geq 1 \qquad \text{for each } j = 1, \ldots, m$$

$$x_i \in \{0, 1\} \qquad \text{for each } i = 1, \ldots, n$$

- We relax this 0/1 IP by replacing the integrality constraints $x_i \in \{0, 1\}$ by the following constraints:

$$0 \leq x_i \qquad \text{for each } i = 1, \ldots, n$$

- We solve this LP and find an optimal solution $\{x_1^*, \ldots, x_n^*\}$.
- We know $x_i^* \leq 1$ since in an optimal solution, each $x_i^*$ is at most 1.
- Thus, we can treat the $x_i^*$ values as probabilities and choose $S_i$ (to be in our set cover) with probability $x_i^*$.

# Some comments on this randomized rounding

- This is a covering problem and the sets produced by randomized rounding 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.
- We will next show:

This randomized rounding algorithm with high probability produces a cover whose cost is within a factor $O(\log m)$ of the optimum.

# The analysis

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

- Namely,

$$\mathbb{E}[cost(C')] = \sum w_i \cdot \mathbb{P}[S_i \text{ is chosen}] = \sum w_i x_i^* = \text{OPT-LP}$$

- Now we need to calculate the probability that a given $u \in U$ is not covered.

- Let's say that $u$ occurs in sets $S_1, \ldots, S_k$. The LP solution must satisfy the constraint:

$$\sum_{i \,:\, u \in S_i} x_i^* \geq 1$$

# The analysis continued

- Under this constraint, we can minimize the probability that u is covered by $x_i^* = 1/k$ for $1 \leq i \leq k$. Thus,

$$\mathbb{P}[\ u \text{ is not covered }] \leq \left(1 - \frac{1}{k}\right)^k \leq \frac{1}{e}$$

- Suppose now that we run the same randomized rounding algorithm $c \ln m$ times, where $m = |U|$, for some constant $c$. On each iteration, add sets (given by the rounded LP) to the set cover.

- While we may be adding the same set many times (and hence overcounting),

$$\text{the cost of the ``cover''} \leq (c \ln m) \cdot \text{OPT-LP}.$$

- Thus,

$$\mathbb{P}[\ u \text{ is not covered }] \leq \left(\frac{1}{e}\right)^{c \ln m} = \left(\frac{1}{m}\right)^c.$$

# Finishing the analysis

**The union bound**

Let $R_1, \ldots, R_m$ be a set of random events with $\mathbb{P}[R_j] \leq p_j$. Then

$$\mathbb{P}[\text{ at least one } R_j \text{ occurs }] \leq p_1 + \ldots + p_m$$

- Let $R_j$ be the event that element $j$ is not covered. Then by the union bound,

$$\mathbb{P}[\text{ some } u \in U \text{ is not covered }] \leq |U| \left(\frac{1}{m}\right)^c = \left(\frac{1}{m}\right)^{c-1}$$

- Using the Markov inequality we can show that the expected cost is within $O(\log m) \cdot \text{OPT-LP}$ with good probability. Hence, with good probability we get a cover with cost $O(\log m) \cdot \text{OPT-LP}$.

- This certainly shows that with good probability we get a cover with cost $O(\log m) \cdot \text{OPT}$ since $\text{OPT-LP} \leq \text{OPT}$.