CSE 521 Winter 2006
Notes on Approximation Algorithm for MaxSAT

Instructor: Venkatesan Guruswami

1 Randomized Rounding

Given an instance of MaxSAT consisting of m clauses $C_1, \ldots, C_m$ over n variables $x_1, x_2, \ldots, x_n$, the following is an integer linear programming formulation of the problem. In what follows, we use the notation $\mathsf{pos}(C_j) = \{i \mid x_i \text{ occurs positively in } C_j\}$, and $\mathsf{neg}(C_j) = \{i \mid x_i \text{ occurs negatively in } C_j\}$

The ILP formulation: We use a variable y_i to indicate if x_i is set to True of False, and a variable z_j to indicate whether a clause is satisfied or not.
Maximize $\sum_j z_j$
subject to

$\forall j$, $\sum_{i \in \mathsf{pos}(C_j)} y_i + \sum_{i \in \mathsf{neg}(C_j)} (1 - y_i) \geq z_j$

$y_i \in \{0,1\}$ for $i = 1, 2, \ldots, n$,

$z_j \in \{0,1\}$ for $j = 1, 2, \ldots, m$.

The LP relaxation is obtained by relaxing the integrality constraints on y_i, z_j to $0 \leq y_i \leq 1$ and $0 \leq z_j \leq 1$.

Fact: The optimum value of the above LP, c_{LP}, is at least OPT, where OPT is the maximum number of clauses that can be satisfied in the given MaxSAT instance.

The approximation algorithm proceeds by solving the above LP to find an optimal solution (y^*, z^*). It then sets each x_i independently to True with probability y_i^* and False with probability $1 - y_i^*$. This technique is called **Randomized Rounding**, and is a powerful, widely used one in approximation algorithms.

Let Z denote the random variable that equals the number of clauses satisfied by the above algorithm. We wish to estimate $\mathbf{E}[Z]$, and compare it with OPT. For each $j = 1, 2, \ldots, m$, define the indicator random variable Z_j for the event that C_j is satisfies. Then $Z = \sum_{j=1}^m Z_j$, so we turn to estimating $\mathbf{E}[Z_j]$. For each integer $k \geq 1$, define $\alpha_k = 1 - (1 - 1/k)^k$.

Lemma 1 *If C_j has k distinct literals, then*

$$\mathbf{E}[Z_j] \geq \alpha_k z_j^* = \left(1 - \left(1 - \frac{1}{k}\right)^k\right) z_j^* \ .$$

Proof: We can assume without loss of generality that all literals of C_j are positive. (Indeed, if not and x_i appears negated, we can just replace x_i with $\bar{x}_i$ throughout and change y_i^* to $1 - y_i^*$, without affecting z_j^* and Z_j.) Also by renaming variables, we assume $C_j = (x_1 \vee x_2 \vee \ldots x_k)$.

The probability that C_j is satisfied is

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

where we used the AM-GM inequality and the fact that $\sum_{i=1}^{k} y_i^* \geq z_j^*$.

Now the function $f(z) = 1 - (1 - z/k)^k$ is a concave function with $f(0) = 0$ and $f(1) = 1 - (1 - 1/k)^k$, and so $f(z) \geq (1 - (1 - 1/k)^k)z = \alpha_k z$ for $0 \leq z \leq 1$. It follows that the probability that C_j is satisfied, which is also $\mathbf{E}[Z_j]$, is at least $(1 - (1 - 1/k)^k)z_j^*$, as claimed. □

Since $(1 - 1/k)^k \leq 1/e$ for all $k \geq 1$, it follows that $\alpha_k \geq (1 - 1/e)$ for all $k \geq 1$. Hence $\mathbf{E}[Z_j] \geq (1 - 1/e)z_j^*$ for each j. Therefore

$$\mathbf{E}[Z] = \sum_{j=1}^{m} \mathbf{E}[Z_j] \geq (1 - 1/e)\sum_{j=1}^{m} z_j^* = (1 - 1/e)c_{\mathrm{LP}} \geq (1 - 1/e)\mathsf{OPT} \ .$$

We therefore have a randomized approximation algorithm which delivers a solution with expected value at least $(1 - 1/e)$ times the maximum number of satisfiable clauses.

2 Improving the approximation ratio

We now consider the following algorithm: Pick $b \in \{0, 1\}$ uniformly at random. If $b = 0$ run the above algorithm, and if $b = 1$ pick a random, independent assignment to the x_i's.

Define the notation $\beta_k = (1 - 2^{-k})$ for $k \geq 1$. Using the same notation as above, for this algorithm, we have $\mathbf{E}[Z_j \mid b = 0] \geq \alpha_k z_j^*$ and $\mathbf{E}[Z_j \mid b = 1] = \beta_k z_j^*$. Combining these we get

$$\mathbf{E}[Z_j] = \frac{1}{2}\,\mathbf{E}[Z_j \mid b = 0] + \frac{1}{2}\,\mathbf{E}[Z_j \mid b = 1] \geq \frac{\alpha_k + \beta_k}{2} z_j^* \ .$$

Now $\alpha_1 + \beta_1 = \alpha_2 + \beta_2 = 3/2$ and $\alpha_k + \beta_k \geq 7/8 + (1 - 1/e) \geq 3/2$ for $k \geq 3$. Therefore

$$\mathbf{E}[Z] = \sum_{j=1}^{m} \mathbf{E}[Z_j] \geq \frac{3}{4}\sum_{j=1}^{m} z_j^* = \frac{3}{4}c_{\mathrm{LP}} \geq \frac{3}{4}\mathsf{OPT} \ .$$

We thus have a randomized 4/3-approximation algorithm for MaxSAT.