BLACKBOARD PROOFS

CSE202 – WEEK 14

1. Approximate SubsetSum

The aim of this proof is to show that the exponentially many elements of $S_i^{(e)}$ are not too far from those of $S_i^{(a)}$, while at the same time, the size of the sets $S_i^{(a)}$ is sufficiently small for the complexity to remain polynomial.

First, the size of the intersection of these sets with intervals is controlled by the interval.

Lemma 1. *For any* $0 < a \leq b$, $|\operatorname{filter}(S,\delta) \cap [a,b]| \leq 1 + \log_{1+\delta}(b/a)$.

Proof. Let $s_1, \dots, s_k$ with $a \leq s_1 < s_2 < \cdots < s_k \leq b$ be the elements of $\operatorname{filter}(S,\delta) \cap [a,b]$. Then by construction of filter,

$$(1+\delta)s_i < s_{i+1} \quad \text{for all } i,$$

so that by induction $s_k \geq (1+\delta)^{k-1}a$ for all k and thus

$$b \geq (1+\delta)^{k-1}a,$$

whence $k \leq 1 + \log_{1+\delta}(b/a)$, as was to be proved. □

Next, each element of $S_i^{(e)}$ is close to an element of $S_i^{(a)}$.

Lemma 2. $\forall s \in S_i^{(e)}, \exists r \in S_i^{(a)},\ r \leq s \leq r(1+\delta)^i$.

Proof. By induction on i. The property is true when $i = 0$ where $S_0^{(e)} = S_0^{(a)} = \{0\}$. Next, assuming that the property holds for $i-1$, let $s \in S_i^{(e)}$. There exists $u \in S_{i-1}^{(e)}$ such that $s = u$ or $s = u + X_i$. We consider the second case, the first one is similar. By assumption, there exists $t \in S_{i-1}^{(a)}$ with $t \leq u \leq t(1+\delta)^{i-1}$ and $t + X_i$ belongs to $S_i^{(a)}$ before filtering, which implies that there exists $r \in S_i^{(a)}$ such that $r \leq t + X_i \leq r(1+\delta)$. Putting all these inequalities together gives

$$\begin{aligned} r \leq t + X_i \leq s = u + X_i \leq t(1+\delta)^{i-1} + X_i &\leq (r(1+\delta) - X_i)(1+\delta)^{i-1} + X_i \\ &= r(1+\delta)^i + (1-(1+\delta)^{i-1})X_i \leq r(1+\delta)^i. \quad \square \end{aligned}$$

The approximation factor follows.

Lemma 3. $s_{-1}^{(e)} \leq s_{-1}^{(a)}(1+\delta)^n$ *and* $(1+\varepsilon/n)^{-n} \geq 1 - \varepsilon$.

Proof. The first part is a consequence of the previous lemma. The second part is obtained as usual from $\log(1+u) \leq u$ for $u \in (-1,\infty)$ and its corollary $\exp(-\epsilon) \geq 1 - \epsilon$ (obtained by setting $u = e^{-\epsilon} - 1$), leading to

$$(1+\varepsilon/n)^{-n} = \exp(-n\log(1+\varepsilon/n)) \geq \exp(-n\varepsilon/n) \geq 1 - \varepsilon. \quad \square$$

Next, we control the sizes of each of the $S_i^{(a)}$.

Lemma 4. $|S_i^{(a)}| \leq |S_{i-1}^{(a)}| + \log_{1+\delta}((1+\delta)i)$.

Proof. Let $T = \text{filter}(S_{i-1}^{(a)} \cup (X_i + S_{i-1}^{(a)}))$ so that $S_i^{(a)} \subset T$. Then Lemma 1 yields

$$|T| \leq |T \cap [0, X_i)| + |T \cap [X_i, \infty)| \leq |S_{i-1}^{(a)}| + 1 + \log_{1+\delta}(\max T/X_i).$$

By induction, $\max T \leq X_1 + \cdots + X_i \leq iX_i$, which concludes the proof. $\square$

Proposition 1. $\sum |S_i^{(a)}| = O(\epsilon^{-1} n^3 \log n)$.

Proof. An induction from the previous lemma gives

$$|S_i^{(a)}| \leq \sum_{j=1}^{i} \log_{1+\delta}((1+\delta)j) \leq i \log_{1+\delta}((1+\delta)i) \leq n \log_{1+\delta}((1+\delta)n),$$

thus by summation,

$$\sum_{i=1}^{n} |S_i^{(a)}| \leq n^2 \log_{1+\delta}((1+\delta)n) = n^2 \left(\frac{\log n}{\log(1+\epsilon/n)} + 1 \right)$$

and the conclusion follows from $\log(1+\epsilon/n) \sim \epsilon/n$ as $\epsilon/n \to 0$. $\square$

2. Max-2-Sat is NP-complete

A clause $a_i \vee b_i \vee c_i$ is satisfied if and only if one of its variables evaluates to 1 (=true).

Consider to the formula

$$(a_i \vee e_i) \wedge (a_i \vee \bar{e}_i) \wedge (b_i \vee f_i) \wedge (b_i \vee \bar{f}_i) \wedge (c_i \vee g_i) \wedge (c_i \vee \bar{g}_i) \wedge (\bar{a}_i \vee \bar{b}_i) \wedge (\bar{a}_i \vee \bar{c}_i) \wedge (\bar{b}_i \vee \bar{c}_i)$$
$$\wedge (d_i \vee h_i) \wedge (d_i \vee \bar{h}_i) \wedge (a_i \vee \bar{d}_i) \wedge (b_i \vee \bar{d}_i) \wedge (c_i \vee \bar{d}_i).$$

Then, if $a_i = b_i = c_i = 0$ up to 10 clauses can be satisfied (setting $d_i = 0$); if $a_i = b_i = c_i = 1$, the highest number of clauses satisfied is 11 (with $d_i = 1$); if only one variable, say a_i, is 1, the maximum is again 11 with $d_i = 0$; if two of them are 1, then again the maximum is 11, with $d_i = 1$.