高等演算法 HW3

許博翔

May 17, 2024

Notation 1. Let n be a positive integer. $[n] := \{1, 2, \dots, n\}$.

Problem 0.

Problem 1, 2, 3, 4: All by myself.

Problem 5, 6: Discuss with: B10401113 張有朋

Problem 1. Consider $(x_1 \lor x_2) \land (x_1 \lor \neg x_2) \land (\neg x_1 \lor x_2) \land (\neg x_1 \lor \neg x_2)$.

Equivalent ILP:

$\max(z_1 + z_2 + z_3 + z_4)$.

$$\text{subject to} : \begin{cases} y_1 + y_2 \geq z_1 \\ y_1 + 1 - y_2 \geq z_2 \\ 1 - y_1 + y_2 \geq z_3 \\ 1 - y_1 + 1 - y_2 \geq z_4 \\ y_i, z_c \in \{0, 1\} \end{cases}.$$

One can see that in LP, we can set $y_1 = y_2 = \frac{1}{2}$, and get $z_1 = z_2 = z_3 = z_4 = 1$, which maximizes $z_1 + z_2 + z_3 + z_4 = 4$.

But since exactly one of the 4 clauses above must be false, $\max(z_1 + z_2 + z_3 + z_4) = 3$.

$\therefore$ the integrality gap is $\frac{3}{4}$ in this case.

Note that it can't be more than $\frac{3}{4}$ since in the following problem, we'll find a solution ALG that satisfies $\frac{3}{4}OPT \leq \frac{3}{4}OPT(LP) \leq ALG \leq OPT$.

$\therefore$ MAX-SAT has integrality gap $\frac{3}{4}$.

Problem 2.

Lemma 2.1. Let $f(x) = 1 - \frac{1}{4^x} - \frac{3}{4}x$. For $0 \le x \le 1$, $f(x) \ge 0$.

Proof. It's obvious that f is continuous and differentiable in $\mathbb{R}$.
$f'(x) = \ln 4 \frac{1}{4^x} - \frac{3}{4}$.
$\Rightarrow f'(x) > 0 \iff \ln 4 \frac{1}{4^x} > \frac{3}{4} \iff 4^x < \frac{4 \ln 4}{3} \iff x < \log_4(\frac{4 \ln 4}{3}) \approx 0.443$.
$\therefore f$ is increasing in $(-\infty, \log_4(\frac{4 \ln 4}{3}))$ and decreasing in $(\log_4(\frac{4 \ln 4}{3}), \infty)$.
$\Rightarrow \forall x \in [0, \log_4(\frac{4 \ln 4}{3})], x \ge f(0) = 0$, and $\forall x \in [\log_4(\frac{4 \ln 4}{3}), 1], x \ge f(1) = 0$.
$\therefore f(x) \ge 0$ for all $x \in [0, 1]$. ■

Let c be a clause.
The probability that c is satisfied $= 1 - \prod_{i \in S_c^+} (1 - 4^{y_i^* - 1}) \prod_{i \in S_c^-} (1 - (1 - 4^{y_i^* - 1})) \overset{\because 1 - 4^{y_i^* - 1} \le 4^{-y_i^*}}{\ge}$
$1 - \prod_{i \in S_c^+} 4^{-y_i^*} \prod_{i \in S_c^-} 4^{y_i^* - 1} = 1 - (\frac{1}{4})^{\sum_{i \in S_c^+} y_i^* + \sum_{i \in S_c^-} (1 - y_i^*)}$.
By the restrictions in LP, there is $\sum_{i \in S_c^+} y_i^* + \sum_{i \in S_c^-} (1 - y_i^*) \ge z_c^*$.
$\therefore$ the probability that c is satisfied $\ge 1 - (\frac{1}{4})^{\sum_{i \in S_c^+} y_i^* + \sum_{i \in S_c^-} (1 - y_i^*)} \ge 1 - (\frac{1}{4})^{z_c^*} \overset{\text{by } \textbf{Lemma (2.1)}}{\ge}$
$\frac{3}{4} z_c^*$.
$\therefore$ the expected number of clauses that are satisfied $\ge \frac{3}{4} \sum_c z_c^*$.

Problem 3. Let V denote the vertex set, and E denote the edge set.

Algorithm:

For every vertex, color it with one of the k colors uniform randomly and independently.

For every edge (u, v), $\Pr[(u, v) \in S] = \Pr[u, v$ have the different colors $] = 1 - \frac{1}{k}$.
$\therefore$ the expected size of S is $(1 - \frac{1}{k})|E| \ge (1 - \frac{1}{k})OPT$, and this is a randomized $(1 - \frac{1}{k})$-approximation algorithm.
Derandomize:

Suppose that $V = [n]$.

Let $[k]$ denote the k colors.

Run the following algorithm with parameter m to obtain the coloring $c_m : [n] \to [k]$.

When $m = n$, the algorithm is deterministic.

- for $i = 1$ to m:
  - Choose j s.t. $|\{1 \le i' \le i-1 : c_m(i') \neq j, (i', i) \in E\}|$ is maximized. – (1)
  - Set $c_m(i) = j$ (i.e. color the vertex i with j).
- for $i = m+1$ to n:
  - Uniformly choose j from k.
  - Set $c_m(i) = j$ (i.e. color the vertex i with j).

Let the resulting S of the algorithm be S_m.
$\mathrm{E}[|S_0|] \ge (1 - \frac{1}{k})OPT$, which has been proved above.
$\mathrm{E}[|S_i|] = |\{(u,v) \in E : u < v < i, c_i(u) \neq c_i(v)\}| + |\{(u,i) \in E : u < i, c_i(u) \neq c_i(i)\}| + (1 - \frac{1}{k})|\{(u,v) \in E : u < v, v > i\}| \overset{\text{By (1)}}{\ge} |\{(u,v) \in E : u < v < i, c_i(u) \neq c_i(v)\}| + (1 - \frac{1}{k})|\{(u,i) \in E : u < i\}| + (1 - \frac{1}{k})|\{(u,v) \in E : u < v, v > i\}| = \mathrm{E}[|S_{i-1}|]$.
$\therefore$ the algorithm with parameter $m = n$, which is a deterministic algorithm, satisfied $|S_n| \ge |S_{n-1}| \ge \cdots \ge |S_0| \ge (1 - \frac{1}{k})OPT$.
Clearly, setting $c_n(i) = j$ above is $O(1)$.
One can first store the neighborhood of each vertex, and then in (1), run though all neighbors of i.
Since each edge will be run for twice in (1), the running time of this algorithm is $O(|V| + |E|)$.

Problem 4.

Lemma 4.1. If a process succeeds with probability at least p, where $p > 0$ is a constant, and each time the process runs, the results are independent, then the expected number of times running the process to get a success is at most $\frac{1}{p}$.

Proof. Let T denote the number of times running the process to get a success.
Suppose that the probability that the process succeeds $= q$.
$\mathrm{E}[T] = 1 + \Pr[\text{ the process fails }]\mathrm{E}[T] \le 1 + (1-p)\mathrm{E}[T]$.

$\Rightarrow p\mathrm{E}[T] \leq 1$.
$\Rightarrow \mathrm{E}[T] \leq \dfrac{1}{p}$. ■

Lemma 4.2. The expected time complexity of step 3 is $O(i)$.

Proof. The probability that inserting j in the rebuild process fails = the probability that collisions happen on both slots $\overset{\text{Both tables have at most } j \text{ non-empty slots}}{\leq} (\frac{j}{4i^{1.5}})^2 \leq (\frac{i}{4i^{1.5}})^2 = \frac{1}{16i}$.
$\Rightarrow$ the probability that at least one of the above fails $\leq i \times \frac{1}{16i} = \frac{1}{16}$.
$\therefore$ the probability that the rebuild process succeeds $= 1-$ the probability that at least one of the above fails $\geq \frac{15}{16}$.
By **Lemma (4.1)**, the expected number of times the rebuild process is run $\leq \frac{16}{15} = O(1)$.
Since the rebuild process needs to insert at most i items, its time complexity is $O(i)$.
$\therefore$ the expected time complexity of step 3 $= O(i)O(1) = O(i)$. ■

Lemma 4.3. For any i, the expected total number of times step 3 is run when the i, $(i+1)$, ..., $(2i-1)$-th insertion arrive is $O(1)$.

Proof. If step 3 is run for 0 times, then **Lemma (4.3)** holds clearly.
Suppose that step 3 is run for the first time when the j-th insertion arrives.
For all $k = j+1, j+2, \ldots, 2i-1$, let p_k denote the probability that step 3 is run when the k-th insertion arrives.
$p_k =$ the probability that collisions happen on both slots $\overset{\text{Both tables have at most } k \text{ non-empty slots}}{\leq} (\frac{k}{\text{table size}})^2 \leq (\frac{k}{4j^{1.5}})^2 \leq (\frac{2i}{4j^{1.5}})^2 \leq (\frac{2i}{4i^{1.5}})^2 = \frac{1}{4i}$.
$\therefore$ the expected number of step 3 is called $= 1 + \sum_{k=j+1}^{2i-1} p_k \leq 1 + \sum_{k=i}^{2i-1} p_k \leq 1 + \sum_{k=i}^{2i-1} \frac{1}{4i} = \frac{3}{2} = O(1)$. ■

Let t_i denote the total expected running time when the i, $i+1$, ..., $2i-1$-th insertion arrive.
By **Lemma (4.3)**, t_i is ($O(i)$ step 1 or 2) + ($O(1)$ step 3) $\overset{\text{By } \textbf{Lemma (4.2)}}{=} O(i)O(1) + O(1)O(i) = O(i)$.

Let k be an integer such that $n \leq 2^k < 2n$.

The total expected time complexity $\leq t_1 + t_2 + t_4 + t_8 + \cdots + t_{2^k} = O(1) + O(2) + O(4) + O(8) + \cdots + O(2^k) = O(2^{k+1}) = O(4n) = O(n)$.

Problem 5.

Lemma 5.1. For all $x \geq 0$, there is $f(x) = e^{-x} + x - 1 \geq 0$.
That is, $1 - x \leq e^{-x}$.

Proof. $f'(x) = -e^{-x} + 1 \geq 0$ for all $x \geq 0$.
$\therefore f$ is increasing in $(0, \infty)$.
$\because f(0) = 1 + 0 - 1 = 0$.
$\therefore f(x) \geq 0$ for all $x \geq 0$. ■

Suppose that there are i points $P_1, P_2, \ldots, P_i$ on the circle, where the intervals between any two points are greater than $\dfrac{1}{n^2}$. Pick the $i+1$-th point P_{i+1} uniformly at random on the circle. Let E_j denote the event that the interval formed by P_j and the P_{i+1} is less than $\dfrac{1}{2n^2}$.
Since E_j, E_k are disjoint for all $j \neq k$.
$\therefore$ The probability that the intervals between any two points are still greater than $\dfrac{1}{n^2} \leq 1 - \Pr[E_1 \cup E_2 \cup \cdots \cup E_i] = 1 - (\Pr[E_1] + \Pr[E_2] + \cdots + \Pr[E_i]) = 1 - \dfrac{i}{2n^2}$.
$\therefore$ after inserting n points, the probability that the intervals between any two points are greater than $\dfrac{1}{n^2} \leq (1 - \dfrac{1}{2n^2})(1 - \dfrac{2}{2n^2}) \cdots (1 - \dfrac{n}{2n^2}) \overset{\text{By } \textbf{Lemma (5.1)}}{\leq} e^{-\frac{1}{2n^2}} e^{-\frac{2}{2n^2}} \cdots e^{-\frac{n}{2n^2}} = e^{-\frac{n(n+1)}{4n^2}} \leq e^{-\frac{n^2}{4n^2}} = e^{-\frac{1}{4}}$.
$\therefore$ the probability that the size of the smallest interval is less than $\dfrac{1}{n^2} = 1-$ the probability that the intervals between any two points are greater than $\dfrac{1}{n^2} \geq 1 - e^{-\frac{1}{4}} = \Omega(1)$.

Problem 6. Suppose that the $a_1, a_2, \ldots, a_{2n}$ are n points chosen uniform randomly in $[0, 1]$.

WLOG suppose that the intervals of the i-th slot are $[a_1, a_1 + x], [a_2, a_2 + y] \pmod 1$, and WLOG suppose that $a_2 - a_1 \pmod 1 \leq \dfrac{1}{2}$.
$\Pr[x + y < \dfrac{1}{4n\sqrt{n}}] < \Pr[x < \dfrac{1}{4n\sqrt{n}} \wedge y < \dfrac{1}{4n\sqrt{n}}]$.

Let $A_1 := [a_1, a_1 + \frac{1}{4n\sqrt{n}}), A_2 := [a_2, a_2 + \frac{1}{4n\sqrt{n}})$.
There are two cases that $x < \frac{1}{4n\sqrt{n}} \wedge y < \frac{1}{4n\sqrt{n}}$.
Case 1: $a_2 \in A_1$.
x is guaranteed to $< \frac{1}{4n\sqrt{n}}$ in this case.
The probability that case 1 happens $= \frac{1}{4n\sqrt{n}}$.
$\Pr[y < \frac{1}{4n\sqrt{n}}] = \Pr[a_3 \in A_2 \vee \cdots \vee a_{2n} \in A_2] \leq \sum_{i=3}^{2n} \Pr[a_i \in A_2] = \sum_{i=3}^{2n} \frac{1}{4n\sqrt{n}} = \frac{2n-3}{4n\sqrt{n}} \leq \frac{2n}{4n\sqrt{n}} = \frac{1}{2\sqrt{n}}$.
Case 2: $a_2 \notin A_1$.
The probability that case 2 happens $= 1 - \frac{1}{4n\sqrt{n}}$.
By the assumption that $a_2 - a_1 \pmod 1 \leq \frac{1}{2}$, there is $A_1 \cap A_2 = \emptyset$.
$\Pr[x < \frac{1}{4n\sqrt{n}} \wedge y < \frac{1}{4n\sqrt{n}}] = \Pr[\bigcup_{i=3}^{2n} \bigcup_{j=i+1}^{2n} (a_i \in A_1 \wedge a_j \in A_2) \vee (a_i \in A_2 \wedge a_j \in A_1)] \leq \sum_{i=3}^{2n} \sum_{j=i+1}^{2n} \Pr[(a_i \in A_1 \wedge a_j \in A_2) \vee (a_i \in A_2 \wedge a_j \in A_1)] = \sum_{i=3}^{2n} \sum_{j=i+1}^{2n} 2 \times \frac{1}{4n\sqrt{n}} \times \frac{1}{4n\sqrt{n}} = \binom{2n-2}{2} \times 2 \times \frac{1}{16n^3} \leq \frac{4n^2}{16n^3} = \frac{1}{4n}$.
$\therefore \Pr[x + y < \frac{1}{4n\sqrt{n}}] \leq \frac{1}{4n\sqrt{n}} \times \frac{1}{2\sqrt{n}} + (1 - \frac{1}{4n\sqrt{n}} \times \frac{1}{4n} \leq \frac{1}{8n^2} + \frac{1}{4n} \leq \frac{1}{2n}$.
$\Pr[$ the size of the smallest interval $< \frac{1}{4n\sqrt{n}}] = \Pr[\bigcup_{i=1}^{n}($ the size of the interval of the i-th slot $< \frac{1}{4n\sqrt{n}} \leq \sum_{i=1}^{n} \Pr[$ the size of the interval of the i-th slot $< \frac{1}{4n\sqrt{n}}] \leq \sum_{i=1}^{n} \frac{1}{2n} = \frac{1}{2}$.
$\therefore$ the size of the smallest interval is at least $\frac{1}{4n\sqrt{n}}$ with probability $\geq 1 - \frac{1}{2} = \frac{1}{2} = \Omega(1)$.