

Advanced Algorithms - Homework 3

Kishlaya Jaiswal

December 2, 2019

Exercise 1

Proof. i) Given an instance of hitting set (E, \mathcal{S}) where E is the ground set and \mathcal{S} is a collection of subsets of E . We can reduce it to an instance of set cover as follows:

Let universe $\mathcal{U} = \mathcal{S}$ and subsets $\mathcal{S}' = E$ such that $S \in \mathcal{S}'$ contains all those elements $u \in \mathcal{U}$ where $S \in u$ in the hitting set formulation.

Conversely, given $(\mathcal{U}, \mathcal{S}')$ an instance of set cover, reduce it to hitting set as follows:

Let ground set $E = \mathcal{S}'$ and subsets $\mathcal{S} = \mathcal{U}$ such that $S \in \mathcal{S}$ contains all those elements of $e \in E$ where $S \in e$ in set cover formulation.

Observe that these reductions are inverses of each other and hence there is a bijection between all instances of set cover and all instances of hitting set.

Now we show that the cost of a feasible solution remains preserved under these reductions.

Consider (E, \mathcal{S}) an instance of hitting set and let F be feasible solution, so $F \cap S_i \neq \phi \forall S_i \in \mathcal{S}$, then the set cover instance will be (\mathcal{S}, E) and now F corresponds to sets picked as $F \subseteq E$. F covers the entire universe \mathcal{S} as $F \cap S_i \neq \phi$.

Similarly, if $(\mathcal{U}, \mathcal{S}')$ is an instance of set cover with a feasible solution $\mathcal{T} \subseteq \mathcal{S}'$ such that $\bigcup_{T \in \mathcal{T}} T = \mathcal{U}$ then consider the equivalent hitting set instance $(\mathcal{S}', \mathcal{U})$. Clearly, $T_i \cap \mathcal{U} \neq \phi$ as \mathcal{T} covers the universe \mathcal{U} . Hence \mathcal{T} is a feasible solution for hitting set.

Thus we have shown that there is a bijection between feasible solutions of same cost of any equivalent instances of set cover and hitting set. Hence they are essentially the same problem.

Furthermore, from the cost preserving bijection we also get that $OPT(SC) = OPT(HS)$ for any two equivalent instances.

Since we have $\log n$ approximation for set cover we also get a $\log n$ approximation for hitting set.

Also, if know that size of each set $S \leq f$ in the hitting set instance then we also have a f -approximation.

ii) Let U be the universe with n elements. We pick an element from the universe with probability p (to be decided later), that is we sample F at random from the set of all possible subsets.

Let

$$Y_i = \begin{cases} 1 & \text{if } S_i \text{ is satisfied} \\ 0 & \text{otherwise} \end{cases}$$

Note that,

$$Pr[Y_i = 1] = Pr[S_i \text{ is satisfied}] = Pr[|F \cup S_i| = 1] = |S_i|p(1-p)^{|S_i|-1}$$

Denote by $N = \sum Y_i$ number of satisfied sets. Then we have,

$$E(N) = \sum_{i=1}^m E(Y_i) = \sum_{i=1}^m Pr[Y_i = 1] = \sum_{i=1}^m |S_i|p(1-p)^{|S_i|-1}$$

Suppose the size of all sets is the same, say k . Set $p = \frac{1}{k}$ then

$$E(N) = \sum_{i=1}^m k \left(\frac{1}{k}\right) \left(1 - \frac{1}{k}\right)^{k-1} = m \left(1 - \frac{1}{k}\right)^{k-1}$$

Case 1: $k = 1$. We set $p = 1$ that is we pick all the elements. In this case, assuming the sets were all distinct, we have indeed achieved optimum as $E(N) = m = OPT$

Case 2: Since $\left(\frac{k}{k-1}\right)^{k-1} = \left(1 + \frac{1}{k-1}\right)^{k-1} \leq e$ and Since, $OPT \leq m$,

we get $E(N) \geq \frac{1}{e} OPT$

iii) Say the universe \mathcal{U} has n elements. We consider the partition $(0, n] = (\frac{n}{2}, n] \cup (\frac{n}{4}, \frac{n}{2}] \cup (\frac{n}{8}, \frac{n}{4}] \cup \dots$. Now, size of each set S_i lies in one of this partition. Since there are $\lg n$ many partitions, there exists one partition $(\frac{k}{2}, k]$ such that atleast $\frac{m}{\lg n}$ many sets have size in $(\frac{k}{2}, k]$. We shall consider only these sets. Then repeating the experiment same as above, we get

$$E(N) \geq \sum_{i: |S_i| \in (k/2, k]} \frac{k}{2} \left(\frac{1}{k}\right) \left(1 - \frac{1}{k}\right)^{k-1} \geq \frac{m}{2 \lg n} \left(1 - \frac{1}{k}\right)^{k-1} \geq \frac{1}{2e \lg n} OPT$$

iv) We have the following LP formulation for the unique hitting set instance:

$$\begin{aligned} & \text{Maximize } \sum_{j=1}^m y_j \\ & \sum_{i \in S_j} x_i = y_j & \forall j \leq m \\ & x_i \in \{0, 1\} & \forall i \leq n \\ & y_j \in \{0, 1\} & \forall j \leq m \end{aligned}$$

The above LP has a optimal solution $OPT = m$ because it is given that given instance has the perfect hitting set property.

Relaxing the above LP, we get:

$$\begin{aligned}
& \text{Maximize } \sum_{i=1}^m y_j \\
& \sum_{i \in S_j} x_i = y_j \quad \forall j \leq m \\
& 0 \leq x_i \leq 1 \quad \forall i \leq n \\
& 0 \leq y_j \leq 1 \quad \forall j \leq m
\end{aligned}$$

Let (x^*, y^*) be the optimal solution. Then we choose the i^{th} element with probability x_i^* . Let $q_k = \prod_{j \in S_k} (1 - x_j^*)$, then probability that a set S_k is satisfied is

$$\sum_{i \in S_k} x_i^* \prod_{\substack{j \in S_k \\ j \neq i}} (1 - x_j^*) = \sum_{i \in S_k} \frac{x_i^*}{1 - x_i^*} q_k$$

Hence,

$$\begin{aligned}
E(\# \text{ satisfied sets}) &= \sum_{k=1}^m \sum_{i \in S_k} \frac{x_i^*}{1 - x_i^*} q_k \\
&= \sum_{k=1}^m \sum_{i \in S_k} \frac{q_k}{1 - x_i^*} x_i^* \\
&\geq \sum_{k=1}^m \sum_{i \in S_k} \left(1 - \frac{1}{e}\right) x_i^* \\
&= \left(1 - \frac{1}{e}\right) \sum_{k=1}^m \sum_{i \in S_k} x_i^* \\
&= \left(1 - \frac{1}{e}\right) \sum_{k=1}^m y_k^* \\
&\geq \left(1 - \frac{1}{e}\right) m \\
&= \left(1 - \frac{1}{e}\right) OPT
\end{aligned}$$

v) Suppose all sets have size 2. Let us call this problem as 2–hitting set.

Claim: 2–hitting set is as hard to approximate as vertex cover.

Proof: Given an instance (V, E) of vertex cover. Consider it as an instance of 2–hitting set with ground set V and each (u, v) edge as a subset.

Now clearly, if we have an α -approximation for 2-hitting set we get an α -approximation for vertex cover. Furthermore, the above reduction is a poly-time reduction. Hence 2-hitting set is also NP-hard. ■

Suppose the sets satisfy $|S_i| = 2 \forall S_i$ and it has the perfect hitting set property. Say the ground set is V and subsets S . Then consider the graph (V, S) where the vertices are the ground set and edges are $(u, v) \in S_i \forall S_i \in S$.

Since it has the perfect hitting property, $\exists F \subseteq V$ such that $|F \cap S_i| = 1$

We claim that $(F, V \setminus F)$ is a bi-partition.

Proof: Suppose there is an edge S_i between two elements of $F \implies |F \cap S_i| = 2$ contradiction to perfect hitting property .

Suppose there is an edge S_i between two elements of $V \setminus F \implies |F \cap S_i| = \phi$ which is a contradiction. ■

Hence the graph so obtained is a bipartite graph.

For bipartite graphs, we can find minimum vertex cover in poly-time (Konig-egervary theorem) hence we can find $F =$ minimum vertex cover, in poly-time.

Therefore, we can solve hitting set $(E, \{S_i\})$ instance with

- $\forall S_i |S_i| = 2$
- it has hitting property

in polynomial time.

□

Exercise 2

In class, we have seen a $\frac{1}{2}(\sqrt{5} - 1)$ -approximation algorithm for MAX SAT using biased coins and a randomized $(1 - \frac{1}{e})$ -approximation using randomized LP rounding. We show that choosing the better of the two solutions gives a $\frac{3}{4}$ -approximation for MAX SAT.

Proof. Let W_0 denote the cost of the solution using biased coins randomized algorithm, W_1 denote the cost of the solution using unbiased coins randomized algorithm and W_2 denote the cost of the solution using randomized LP-rounding.

We recall that

$$E(W_0) \geq E(W_1)$$

$$E(W_1) \geq \sum_{j=1}^m w_j \left(1 - \left(\frac{1}{2}\right)^{l_j}\right)$$

and

$$E(W_2) \geq \sum_{j=1}^m w_j y_j^* \left(1 - \left(1 - \frac{1}{l_j}\right)^{l_j}\right)$$

where w_j is the weight of j-th clause and y_j^* is optimal solution of relaxed LP.

Choosing better of the two solutions corresponds to $W = \max(W_0, W_2)$, then we have $\max(W_0, W_2) \geq \frac{1}{2}W_0 + \frac{1}{2}W_2$. Thus,

$$\begin{aligned}
E(W) &\geq E\left(\frac{1}{2}W_0 + \frac{1}{2}W_2\right) \\
&= \frac{1}{2}E(W_0) + \frac{1}{2}E(W_2) \\
&\geq \frac{1}{2}E(W_1) + \frac{1}{2}E(W_2) \\
&\geq \frac{1}{2}\left(\sum_{j=1}^m w_j \left(1 - \left(\frac{1}{2}\right)^{l_j}\right)\right) + \frac{1}{2}\left(\sum_{j=1}^m w_j y_j^* \left(1 - \left(1 - \frac{1}{l_j}\right)^{l_j}\right)\right) \\
&\geq \frac{1}{2}\left(\sum_{j=1}^m w_j y_j^* \left(1 - \left(\frac{1}{2}\right)^{l_j}\right)\right) + \frac{1}{2}\left(\sum_{j=1}^m w_j y_j^* \left(1 - \left(1 - \frac{1}{l_j}\right)^{l_j}\right)\right) \\
&\quad (\text{because } 0 \leq y_j^* \leq 1) \\
&= \frac{1}{2}\sum_{j=1}^m w_j y_j^* \left(\left(1 - \left(\frac{1}{2}\right)^{l_j}\right) + \left(1 - \left(1 - \frac{1}{l_j}\right)^{l_j}\right)\right)
\end{aligned}$$

Remains to show that $\forall n > 0$

$$\left(1 - \left(\frac{1}{2}\right)^n\right) + \left(1 - \left(1 - \frac{1}{n}\right)^n\right) \geq \frac{3}{2}$$

Let

$$f(x) = 2 - \left(\frac{1}{2}\right)^x - \left(1 - \frac{1}{x}\right)^x$$

Then $f'(x) = -\left(\frac{1}{2}\right)^x \ln \frac{1}{2} - \left(1 - \frac{1}{x}\right)^x \left(\ln\left(1 - \frac{1}{x}\right) + \frac{1}{x-1}\right) > 0$. Hence f is increasing for $x > 1$ and we get our desired result. \square

Exercise 2

Proof. Given a graph G , we perform the following experiment E , n times on it:

Remove each edge of G with probability p and retain it with probability $1-p$. If the resulting subgraph is disconnected, then set $X_i = 1$ otherwise, set $X_i = 0$.

Now, set $X = \frac{X_1 + X_2 + \dots + X_t}{t}$

We have already seen that $E(X) = p_{fail} = \mu$ and $V(X) = \frac{\mu - \mu^2}{t} \leq \frac{\mu}{t}$ and so

$$Pr[|X - \mu| > \mu\epsilon] \leq \frac{1}{t\mu\epsilon^2} \leq \frac{poly(n)}{t\epsilon^2}$$

as $\mu = p_{fail} = \Omega\left(\frac{1}{poly(n)}\right)$ is given.

If we set this failure probability to atmost $\frac{1}{4}$ then we get $t \geq \frac{4}{\epsilon^2} poly(n)$.

Now, we repeat the above entire experiment $s = 2r + 1$ times, to get i.i.d. random values Y_1, Y_2, \dots, Y_s . Note that, we have $E(Y_i) = E(X)$ and $V(Y_i) = V(X)$ and so

$$Pr[|Y_i - \mu| > \mu\epsilon] \leq \frac{1}{t\mu\epsilon^2}$$

Let $Y = \text{median}(Y_1, Y_2, \dots, Y_{2r+1})$.

We want to have failure probability atmost δ for Y that is $Pr[|Y - \mu| > \mu\epsilon] \leq \delta$.

For that, we note that if $|Y - \mu| > \mu\epsilon$ then two of the following can happen:

- $Y < (1 - \epsilon)\mu$. Since Y was the median, we must have $r + 1$ many experiments to fail from below, that is there are $r + 1$ many values for i such that $Y_i < (1 - \epsilon)\mu$.
- $Y > (1 + \epsilon)\mu$. Since Y was the median, we must have $r + 1$ many experiments to fail from above, that is there are $r + 1$ many values for i such that $Y_i > (1 + \epsilon)\mu$.

Now, we set $p = \frac{1}{t\mu\epsilon^2}$ and $q = 1 - p$. Note that $p \leq \frac{1}{4} \implies q \geq \frac{3}{4}$.

We look at some inequalities first:

$$\forall k \leq r, \frac{q^k p^{2r+1-k}}{q^r p^r} = p \left(\frac{p}{q} \right)^{r-k} \leq 1 \quad (1)$$

$$\sum_{k=0}^r \binom{2r+1}{k} \leq \sum_{k=0}^{2r+1} \binom{2r+1}{k} = 2^{2r+1} \quad (2)$$

$$4q - 4q^2 = 1 - (2q - 1)^2 \leq 1 - \left(\frac{1}{2} \right)^2 \leq \frac{3}{4} \quad (3)$$

Hence, we get

$$\begin{aligned}
Pr[|Y - \mu| > \mu\epsilon] &= Pr[|Y_i - \mu| > \mu\epsilon \text{ for atleast } r + 1 \text{ many values of } i] \\
&= Pr[|Y_i - \mu| > \mu\epsilon \text{ for exactly } r + 1 \text{ many values of } i] \\
&\quad + Pr[|Y_i - \mu| > \mu\epsilon \text{ for exactly } r + 2 \text{ many values of } i] + \dots \\
&\quad \dots + Pr[|Y_i - \mu| > \mu\epsilon \text{ for exactly } 2r + 1 \text{ many values of } i] \\
&= \sum_{k=r+1}^{2r+1} \binom{2r+1}{k} p^k q^{2r+1-k} \\
&= \sum_{k=0}^r \binom{2r+1}{k} q^k p^{2r+1-k} \\
&\leq \sum_{k=0}^r \binom{2r+1}{k} q^r p^r \quad \text{from (1)} \\
&= (q - q^2)^r \sum_{k=0}^r \binom{2r+1}{k} \\
&\leq (q - q^2)^r 2^{2r} \quad \text{from (2)} \\
&= (4q - 4q^2)^r \\
&\leq \left(\frac{3}{4}\right)^r \quad \text{from (3)} \\
&\leq \delta
\end{aligned}$$

Now, we get

$$r \geq \frac{\log \frac{1}{\delta}}{\log \frac{4}{3}} \implies s \geq \frac{2 \log \frac{1}{\delta}}{\log \frac{4}{3}} + 1$$

Hence, $s = \Omega(\log \frac{1}{\delta})$.

So total time required is $t' = st \geq \frac{4}{\epsilon^2} \text{poly}(n) \log \frac{1}{\delta}$. Hence $t' = \Omega(\log \frac{1}{\delta})$. \square

Exercise 3

Proof. For an assignment a , denote by $p(a)$ the probability of choosing that assignment where each variable is set to 1 (or 0) with probability p (or $1 - p$) independently. Further, for any given set S of assignments, denote by $p(S) = \sum_{a \in S} p(a)$.

Let S_i be the set of all satisfying assignments for the clause C_i and set $\mathcal{S} = \bigcup_{i=1}^m S_i$.

Then $p(\mathcal{S}) = \sum_{a \in \mathcal{S}} p(a)$ is the number we need to approximate.

We first make a $(m \times |\mathcal{S}|)$ table T where the rows are indexed by the clauses and columns are indexed by the assignments \mathcal{S} . For each entry (C_i, a) in the T , set it to $p(a)$ if a satisfies C_i otherwise set it to 0. Mark the first non-zero entry in each column as special.

Now we perform the following experiment:

- Choose a clause C_i with probability $\frac{p(S_i)}{p(U)}$ where $p(U) = \sum p(S_i)$

- Having chosen a clause C_i , choose an assignment a from set S_i at random
- Set $X_n = 1$ iff the chosen assignment a is special, otherwise $X_n = 0$

$$E[X_n] = Pr[X_n = 1] = \frac{p(\mathcal{S})}{p(U)}$$

Now let $X = \frac{\sum_{n \leq t} X_n}{t}$ and $Y = p(U)X$.

Then Y approximates $p(\mathcal{S})$ as $E[Y] = p(\mathcal{S})$ and by law of large numbers, for large enough t , $Y \sim p(\mathcal{S})$.

[Note: We can compute $p(U) = \sum p(S_i)$ in polynomial time as $p(S_i) = \sum p^{p_i}(1-p)^{n_i}$ where the sum runs over all possibilities for the variables not occurring in C_i] \square

Exercise 4

Proof. .

- Probability two balls B_i and B_j fall into the same bin = $\sum_{k=1}^n$ Probability B_i falls in bin k and B_j fall in bin k = $\sum_{k=1}^n (1/n^2) = \boxed{\frac{1}{n}}$
- Let $X_{ij} = 1$ iff ball B_i and ball B_j collide otherwise $X_{ij} = 0$. Then the total number of collisions $X = \sum_{i \neq j} X_{ij}$.

$$\text{Hence, expected number of collisions} = E[X] = \sum_{i \neq j} E[X_{ij}] = \sum_{i \neq j} \frac{1}{n} = \boxed{\binom{m}{2} \frac{1}{n}}$$

- Probability that a particular bin is empty = $\boxed{\left(1 - \frac{1}{n}\right)^m}$.

Let $Y_k = 1$ iff bin k is empty. Then total number of empty bins $Y = \sum_{k=1}^n Y_k$.

Hence, expected number of empty bins = $E[Y] = \sum_{k=1}^n E[Y_k] = \sum_{k=1}^n (1 - \frac{1}{n})^m =$

$$\boxed{n \left(1 - \frac{1}{n}\right)^m}.$$

- Probability that a particular bin has atleast k balls = $\boxed{\binom{m}{k} \left(\frac{1}{n}\right)^k}$
- Probability that all bins have atmost k balls = $1 - \text{Probability that some bin has atleast } k+1 \text{ balls} \geq 1 - n(\text{Probability that a particular bin has atleast } k \text{ balls})$

$$\text{Probability that a particular bin has atleast } k \text{ balls} \leq \binom{m}{k} \left(\frac{1}{n}\right)^k \leq \left(\frac{me}{k}\right)^k \left(\frac{1}{n}\right)^k = \left(\frac{me}{nk}\right)^k$$

Assume that $\boxed{m \leq \frac{3}{e}n}$, then

$$\text{Probability that a particular bin has atleast } k \text{ balls} \leq \left(\frac{3}{k}\right)^k$$

Setting $k = \frac{3 \ln n}{\ln \ln n}$, we get that, Probability that a particular bin has atleast k balls $\leq \frac{1}{n^2}$

Hence, probability that all bins have atmost k balls $\geq 1 - \frac{1}{n}$

- Let Z_i = number of balls in bin i and $Z = \max\{Z_i | 1 \leq i \leq n\}$. Let $k = \frac{3 \ln n}{\ln \ln n}$. We know that $P(Z < k) \geq 1 - \frac{1}{n}$ and so $P(Z \geq k) \leq \frac{1}{n}$. Now,

$$\begin{aligned} E[Z] &= \sum_i i P(Z = i) \\ &= \sum_{i < k} i P(Z = i) + \sum_{i \geq k} i P(Z = i) \\ &\leq (k-1) P(Z < k) + n P(Z \geq k) \leq (k-1) P(Z < k) + 1 \\ &\leq (k-1) + 1 = k = \frac{3 \ln n}{\ln \ln n} \end{aligned}$$

Thus, $E[Z] = \frac{\ln n}{\ln \ln n} (1 + O(1))$

Analysis:

1. Stirling's approximation:

$$e\sqrt{n} \left(\frac{n}{e}\right)^n \leq n! \leq \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

2. $\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$

$$\begin{aligned} \binom{n}{k} &\leq \frac{n!}{k!(n-k)!} \\ &\leq \frac{e\sqrt{n} \left(\frac{n}{e}\right)^n}{\sqrt{2\pi k} \left(\frac{k}{e}\right)^k \sqrt{2\pi(n-k)} \left(\frac{n-k}{e}\right)^{n-k}} \\ &\leq \frac{e}{2\pi} \sqrt{\frac{n}{k(n-k)}} \left(\frac{n}{k}\right)^k \left(\frac{n}{n-k}\right)^{n-k} \\ &\leq \frac{e}{\pi} \left(\frac{n}{k}\right)^k \left(1 + \frac{k}{n-k}\right)^{n-k} \\ &\leq \frac{e}{\pi} \left(\frac{n}{k}\right)^k e^k \\ &\leq \left(\frac{ne}{k}\right)^k \end{aligned}$$

3. $(\ln n)/n$ is a decreasing function $\forall n \geq 1$ (decreases to 0) and hence $(\ln n)/n \leq \frac{1}{3}$ eventually.

4. If $k = \frac{3 \ln n}{\ln \ln n}$, then $\left(\frac{3}{k}\right)^k \leq \frac{1}{n^2}$ for large enough n

$$\begin{aligned}
 \left(\frac{3}{k}\right)^k &= \exp\left(k \ln \frac{3}{k}\right) \\
 &= \exp\left(\frac{3 \ln n}{\ln \ln n} \ln\left(\frac{\ln \ln n}{\ln n}\right)\right) \\
 &= \exp\left(\frac{3 \ln n}{\ln \ln n} (\ln \ln \ln n - \ln \ln n)\right) \\
 &= \exp\left(-3 \ln n + \ln n \frac{3 \ln \ln \ln n}{\ln \ln n}\right) \\
 &\leq \exp(-3 \ln n + \ln n) \\
 &= \exp(-2 \ln n) \\
 &= \frac{1}{n^2}
 \end{aligned}$$

□