University of Freiburg
Dept. of Computer Science
Prof. Dr. F. Kuhn
S. Faour, M. Fuchs, A. Malyusz

Algorithm Theory
Sample Solution Exercise Sheet 13

Due: Friday, 2nd of February 2024, 10:00 am

Remark: For the maximization problems, we are considering the approximation ratio $\alpha := \max_{\text{input instance } I} \frac{OPT(I)}{ALG(I)}$

Exercise 1: Miscellaneous Approximations *(8 Points)*

Let $G = (V, E)$ be an undirected connected graph .

(a) The *minimum dominating set* problem asks to find a dominating set $D \subseteq V$ of minimum size. Show that for $c \geq 2$, the `domset` algorithm from the previous sheet (c.f. sheet 12, exercise 2) computes an $\mathcal{O}(\ln n)$-approximation of a minimum dominating set with probability at least $1 - \frac{2}{n}$. *(3 Points)*

(b) 1. An *independent set* is a set $I \subseteq V$ such that no two nodes in I share an edge in E. The *maximum independent set* problem asks to find an independent set of maximum size. Recall that the *minimum vertex cover* problem asks to find a vertex cover of minimum size. Now, show that both optimization problems are equivalent i.e. finding the minimum-size vertex cover is equivalent to finding the maximum-size independent set . *(1 Point)*

2. Show that the two problems are not equivalent in an approximation-preserving way, i.e it is not true that for all positive integer α, finding an α-approximate minimum vertex cover is equivalent to finding a α-approximate maximum independent set.

Hint: Give a counterexample by finding a family of graphs where one can easily obtain a 2-approximate minimum vertex cover, but this will equivalently find a very bad approximate maximum independent set. *(4 Points)*

Sample Solution

(a) Notice that each node in a minimum dominating set covers at most $d + 1$ nodes, one can then deduce that OPT $\cdot(d + 1) \geq n$, where OPT is the size of a minimum dominating set, and that is enough to show what the question asks for.

(b) 1. One can prove that the statement is true by taking the complement of the result, i.e. the set of vertices is a vertex cover if and only if its complement is an independent set.

2. Consider a complete bipartite graph $K_{n,n}$. One can show that on one hand all the nodes make up a 2-aprroximation to the minimum vertex cover problem, but the complement graph which is the empty graph of size 0 is far away from being a 2-approx of the maximum independent set problem (one can show that a maximum independet set is of size n).

Exercise 2: A Set Cover Variant *(12 Points)*

We consider the following variant of the set cover problem discussed in the lecture. We are given a set of elements X and a collection $\mathcal{S} \subseteq 2^X$ of subsets of X such that $\bigcup_{S \in \mathcal{S}} S = X$. In addition, we are given an integer parameter $k \geq 2$.

Instead of finding a collection $\mathcal{C} \subseteq \mathcal{S}$ of the sets which covers all elements, the goal is to find a set of **at most** k sets $S_1, \ldots, S_k \in \mathcal{S}$ such that the number of covered elements $|S_1 \cup \cdots \cup S_k|$ is maximized. We consider the greedy set cover algorithm from the lecture, but we stop the algorithm after adding k sets.

(a) Show that for $k = 2$, the described greedy algorithm has approximation ratio at most 4/3. *(5 Points)*

(b) Let us now consider a general parameter $k \geq 2$. Show that if an optimal choice of k sets $S_1, \ldots, S_k$ covers ℓ elements, after adding t sets, the greedy algorithm covers at least $\frac{\ell}{k} \cdot \sum_{i=1}^{t} \left(1 - \frac{1}{k}\right)^{i-1}$ elements. *(5 Points)*

(c) Prove that the approximation ratio of the greedy algorithm is at most $\frac{e}{e-1}$. You can use that $(1 - 1/k)^k < e^{-1}$. *(2 Points)*

Sample Solution

(a) Let S denote the optimal solution which covers x *distinct* elements. For the case $k = 2$, the optimal solution consists of one or two non empty sets in $\mathcal{S}$. Let us consider the two cases as follows:

- Assume that the optimal solution consists of only one non empty set, that is, $S = \{S_1\}$. Then S_1 is a set in $\mathcal{S}$ with cardinality x, and all other sets in $\mathcal{S}$ are subset of S_1. Since the greedy approach picks S_1 as the set with maximum cardinality in $\mathcal{S}$, the greedy solution is optimal.
- Assume the optimal solution consists of more than one set, that is $S = \{S_1, S_2\}$. Moreover, let $F = \{F_1, F_2\}$ denote the greedy solution. Since $|S_1 \cup S_2| = x$, then either $|S_1| \geq x/2$ or $|S_2| \geq x/2$. Hence, we can claim that the set with maximum cardinality in $\mathcal{S}$ has cardinality $\geq x/2$, which is picked by the greedy algorithm for the first choice (i.e., $|F_1| \geq x/2$).

  After the first choice of the greedy solution, $\{S_1 \cup S_2\} \setminus F_1$ is the set of uncovered elements in the optimal solution. These elements are covered by S_1 and S_2. Hence, either S_1 or S_2 covers at least half of these elements. Therefore, we can claim that the set in $\mathcal{S}$ which covers the maximum number of uncovered elements covers at least $|\{S_1 \cup S_2\} \setminus F_1|/2$ elements.

  Therefore the greedy solution covers at least $|F_1| + \frac{x - |F_1|}{2} = \frac{x + |F_1|}{2}$ elements *(true since* $|\{S_1 \cup S_2\} \setminus F_1| \geq |\{S_1 \cup S_2\}| - |F_1|$*)*.

  Now, considering $|F_1| \geq x/2$, the greedy solution covers at least $\frac{x + x/2}{2} = \frac{3x}{4}$ elements, which proves the claim on greedy solution's approximation factor.

(b) The core part of the proof is the following claim.

Claim 1. Considering the optimal solution $S = \{S_1, S_2, \ldots, S_k\}$, for any $A \subseteq \bigcup_{i=1}^{k} S_i$, there exists at least one set U in S such that $|U \cap A| \geq \frac{|A|}{k}$. That is, U covers at least $\frac{|A|}{k}$ elements in A.

Proof. Let us assume that there does not exist such a set in S. Then we can say that all of the sets in S covers less than $\frac{|A|}{k}$ elements in A. Since there are at most k sets in S, the union of all the sets in S cannot cover all the elements in A. This contradicts the fact that the the sets in S cover all elements in $X \supseteq A$. □

Let $E = \{e_1, e_2, \ldots, e_\ell\}$ be the set of all elements that are covered by the optimal solution. Moreover, let y_t denote the number of elements that the greedy solution covers after choosing t sets. Hence, the number of uncovered elements by greedy approach in E before it chooses the t^{th} set is $\ell - y_{t-1}$. Based on Claim 1, for the set of uncovered elements in E there exist at least one set in the optimal solution that covers at least $\frac{\ell - y_{t-1}}{k}$ new elements in the optimal solution. Since in step t the greedy algorithm chooses a set in $\mathcal{S}$ which covers the maximum number of uncovered elements, $y_t - y_{t-1} \geq \frac{\ell - y_{t-1}}{k}$. Therefore, we have the following recurrence relation for the number of elements that the greedy solution covers.

$$y_t \geq \frac{\ell}{k} + y_{t-1}\left(1 - \frac{1}{k}\right)$$

Considering the fact that $y_0 = 0$ (before greedy chooses any set, the number of covered elements is zero), by repeated replacement, the above recurrence relation leads to the claim stated in the question.

NB: Alternatively one can prove the statement by arguing by induction on $t \geq 1$.

(c) Considering the recurrence relation achieved in question (b), the number of elements that are covered by greedy solution with k chosen sets is at least

$$\begin{aligned}\frac{\ell}{k} \cdot \sum_{i=1}^{k}\left(1 - \frac{1}{k}\right)^{i-1} &= \frac{\ell}{k} \cdot \frac{1 - \left(1 - \frac{1}{k}\right)^k}{1 - \left(1 - \frac{1}{k}\right)} \\ &\overset{**}{\geq} \frac{\ell}{k} \cdot \frac{1 - \frac{1}{e}}{1 - \left(1 - \frac{1}{k}\right)} \\ &= \ell \cdot \frac{e-1}{e}.\end{aligned}$$

** $\left(1 - \frac{1}{k}\right)^k < \frac{1}{e}$

As a result on any input instance, we can calculate the approximation factor as follows.

$$\frac{Optimal}{Greedy} \leq \frac{\ell}{\ell \cdot \frac{e-1}{e}} = \frac{e}{e-1}$$