# Homework I, Advanced Algorithms 2020

Due on Friday March 27 at 17:00 (upload one solution per group on moodle). Solutions to many homework problems, including problems on this set, are available on the Internet, either in exactly the same formulation or with some minor perturbation. It is *not acceptable* to copy such solutions. It is hard to make strict rules on what information from the Internet you may use and hence whenever in doubt contact Michael Kapralov. You are, however, allowed to discuss problems in groups of up to three students; it is sufficient to hand in one solution per group.

**1** *(20 pts)* Let $G = (V, E)$ be an undirected graph on $n$ vertices with $nd/2$ edges. Consider the following probabilistic experiment for finding an independent set in $G$ (recall that an independent set is a set of vertices of $G$ with no edges going between them). Delete each vertex of $G$ (together with all its incident edges) independently with probability $1 - 1/d$.

**1a** *(5 pts)* Compute the expected number of vertices and edges that remain after the deletion process.

**1b** *(7 pts)* From this infer that there is an independent set with at least $n/(2d)$ vertices in any graph on $n$ vertices and $nd/2$ edges.

**1c** *(8 pts)* Let $G$ be a 3-regular graph. Suppose that we would like to turn our randomized experiment into an algorithm as follows. We delete every vertex of $G$ independently with probability $2/3$. Then for every edge that remains delete one of its endpoints. Derive an upper bound on the probability that this algorithm finds an independent set smaller than $(1 - \epsilon)n/6$ for a fixed $\epsilon > 0$.

**2** *(20 pts)* Let $G = (V, E)$ be an undirected graph, and let $\mathcal{M} = (E, \mathcal{I})$ denote the graphic matroid of $G$.

**2a** (10 pts) Define

$$\mathcal{I}' = \{S \subset E : E \setminus S \text{ contains a base of } \mathcal{I}\}.$$

Prove that $\mathcal{M}' = (E, \mathcal{I}')$ is a matroid.

**2b** (10 pts) Give a polynomial time algorithm that, given an undirected graph $G = (V, E)$, determines if $G$ contains two edge disjoint spanning trees.

**3** *(20 pts)* **Primal-dual Analysis of Greedy for Set Cover.** In this problem you will use LP duality to analyze the performance of the greedy algorithm for the Set Cover problem:

**Input:** A collection $\mathcal{F} = \{S_1, \ldots, S_m\}$ of subsets of a universe $U$ such that $\bigcup_{i=1}^m S_i = U$.

**Output:** A collection of sets from $\mathcal{F}$ of smallest possible size that covers the universe $U$.

The greedy algorithm for set cover starts with an empty solution and repeatedly picks a set from $\mathcal{F}$ that covers the largest number of not yet covered elements:

```
GREEDY(U, F)
C ← ∅                                  ▷ Elements covered so far
while C ≠ U do
    A ← argmax_{S∈F} |S \ C|           ▷ Pick a set that covers the most yet uncovered elements
    Include A in the solution
    Set price(e) ← 1/|A \ C| for all e ∈ A \ C
    C ← C ∪ A
Output all sets included in the solution
```

Note that Greedy presented above, besides constructing the solution, assigns *prices* to elements of the universe once they are covered – you will use these prices to analyze the approximation ratio achieved by Greedy. To do that, you will apply the weak duality theorem to the LP relaxation of Set Cover and its dual, given below:

(Primal) LP Relaxation

$$\begin{aligned} \textbf{minimize} \quad & \sum_{S \in \mathcal{F}} x_S \\ \textbf{subject to} \quad & \sum_{S : e \in S} x_S \geq 1 \quad \text{for } e \in U \\ & x_S \geq 0 \quad \text{for } S \in \mathcal{F} \end{aligned}$$

(Dual)

$$\begin{aligned} \textbf{maximize} \quad & \sum_{e \in U} y_e \\ \textbf{subject to} \quad & \sum_{e \in U : e \in S} y_e \leq 1 \text{ for } S \in \mathcal{F} \\ & y_e \geq 0 \quad \text{for } e \in U \end{aligned}$$

Let $e_1, \ldots, e_n$ denote the elements of the universe in the order they are covered by Greedy (breaking ties arbitrarily). Let OPT denote the number of sets in an optimal solution.

**3a** (7 pts) Show that for every $k = 1, \ldots, n$ one has $\text{price}(e_k) \leq OPT/(n-k+1)$.

**3b** ($\star$, 5 pts) Let $y_e = \frac{\text{price}(e)}{H_n}$ for every $e \in U$, where $H_n = 1 + 1/2 + 1/3 + \ldots + 1/n$ is the $n$-th Harmonic number. Show that $y$ is a feasible solution to the dual of the LP relaxation of Set Cover given above.

**3c** (8 pts) For every $S \in \mathcal{F}$ let $x_S = 1$ if Greedy picks $S$ and $x_S = 0$ otherwise. Apply the weak duality theorem to $x$ and $y$ to conclude that the solution produced by Greedy contains at most $H_n \cdot OPT$ sets.

**4** *(20 pts)* **Integrality gap of the Set Cover LP.** Consider the following instance of the Set Cover problem. For an even integer $d \geq 1$ let

$$U = \left\{ x \in \{0,1\}^d : \sum_{i=1}^{d} x_i = d/2 \right\},$$

i.e., the universe consists of all binary vectors of length $d$ that have $d/2$ nonzeros. Let the collection $\mathcal{F}$ contain $m = d$ sets $S_1, \ldots, S_m$, defined by

$$S_i = \{x \in U : x_i = 1\}$$

for every $i = 1, \ldots, m$.

**4a** (6 pts) Give a feasible solution to the LP relaxation of Set Cover on the instance above with value bounded by 2.

**4b** (7 pts) Prove that at least $d/2 + 1$ sets are needed to cover the universe $U$.

**4c** (7 pts) Conclude that the integrality gap of the LP relaxation of Set Cover on a universe $U$ of size $n$ is $\Omega(\log n)$.

**5** *(20 pts)* **Implementation.** The objective of this problem is to successfully solve the problem *Hiking Trails* on our online judge. You will find detailed instructions on how to do this on Moodle.