ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Final Exam, Advanced Algorithms 2017-2018

- You are only allowed to have a handwritten A4 page written on both sides.
- Communication, calculators, cell phones, computers, etc... are not allowed.
- Your explanations should be clear enough and in sufficient detail that a fellow student can understand them. In particular, do not only give pseudo-code without explanations. A good guideline is that a description of an algorithm should be such that a fellow student can easily implement the algorithm following the description.
- **You are allowed to refer to material covered in the lecture notes** including theorems without reproving them.
- **Problems are not necessarily ordered by difficulty.**
- **Do not touch until the start of the exam.**

Good luck!

Name: _____ **N° Sciper:** _____

Problem 1 / 10 points	Problem 2 / 15 points	Problem 3 / 15 points	Problem 4 / 15 points	Problem 5 / 22 points	Problem 6 / 23 points

Total / 100

1 *(10 pts)* **Submodular functions.** Let $N = \{1, 2, 3, 4, 5, 6, 7, 8\}$ and consider the submodular function $f : 2^N \to \mathbb{R}$ defined by

$$f(S) = \min(|S|, 5) \qquad \text{for every } S \subseteq N.$$

What is the value of $\hat{f}(0, 1/4, 1/2, 1/4, 3/4, 1/2, 3/4, 1)$ where $\hat{f}$ denotes the Lovász extension of f?

(In this problem you are asked to calculate $\hat{f}(0, 1/4, 1/2, 1/4, 3/4, 1/2, 3/4, 1)$. Recall that you are allowed to refer to material covered in the lecture notes.)

Solution:

2 *(15 pts)* **LSH for Manhattan distance.** Recall the Manhattan distance function that we saw in class: for any d-dimensional Boolean vectors $p, q \in \{0,1\}^d$, the Manhattan distance is defined by

$$\text{dist}(p,q) = \|p-q\|_1 = |\{i : p_i \neq q_i\}|\,.$$

Design a locality sensitive hash (LSH) family $\mathcal{H}$ of functions $h : \{0,1\}^d \to \{0,1,2,3\}$ such that for any $p, q \in \{0,1\}^d$,

$$\Pr_{h\sim\mathcal{H}}[h(p) = h(q)] = \left(1 - \frac{\text{dist}(p,q)}{d}\right)^2 .$$

(In this problem you are asked to explain the hash family and show that it satisfies the above property. Recall that you are allowed to refer to material covered in the lecture notes.)

Solution:

3 *(15 pts)* **Estimating cheese consumption of fondue lovers.** Professor Ueli von Gruyères worked hard last year to calculate the yearly cheese consumption of each individual in Switzerland. Specifically, let U be the set of all persons in Switzerland. For each person $i \in U$, Ueli calculated the amount $w_i \in \mathbb{R}_{\geq 0}$ (in grams) of the yearly cheese consumption of person i. However, to help Coop and Migros in their supply-chain management, he needs to calculate the total cheese consumption of those persons that prefer fondue over raclette. That is, if we let $F \subseteq U$ be those that prefer fondue over raclette, then Ueli wants to calculate

$$W_F = \sum_{i \in F} w_i \,.$$

The issue is that Ueli does not know the set F and he does not have the time or energy to ask the preferences of all persons. He therefore designs two estimators that only ask a single person:

Estimator $\mathcal{A}_1$: Let $W = \sum_{i \in U} w_i$. Sample person i with probability $\frac{w_i}{W}$ and output W if i prefers fondue and 0 otherwise.

Estimator $\mathcal{A}_2$: Sample person i with probability $\frac{1}{|U|}$ and output $|U| \cdot w_i$ if i prefers fondue and 0 otherwise.

Let X_1 and X_2 be the random outputs of $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively. Ueli has shown that $\mathcal{A}_1$ and $\mathcal{A}_2$ are unbiased estimators and he has also bounded their variances:

$$\mathbb{E}[X_1] = \mathbb{E}[X_2] = W_F, \qquad \mathrm{Var}[X_1] \leq W^2 \qquad \text{and} \qquad \mathrm{Var}[X_2] \leq |U| \sum_{i \in U} w_i^2 \,.$$

However, Ueli is now stuck because the variances are too high to give any good guarantees for the two estimators. We are therefore going to help Ueli by designing a new estimator with good guarantees while still asking the preferences of relatively few persons.

For a fixed small parameter $\epsilon > 0$, your task is to design and analyze an estimator that outputs a random value Y with the following guarantee:

$$\Pr[|Y - W_F| \geq \epsilon W] \leq 1/3 \,. \tag{1}$$

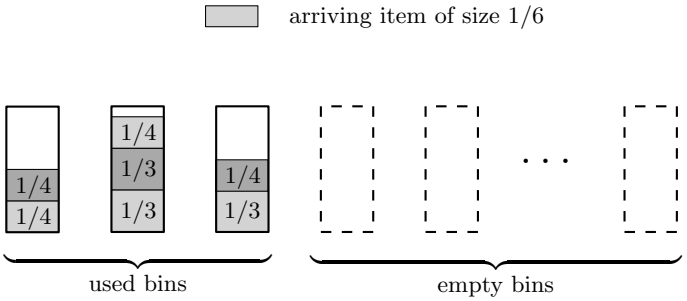
Your estimator should ask at most $3/\epsilon^2$ persons about their preferences.

(In this problem you are asked to (i) design an estimator that asks at most $3/\epsilon^2$ persons about their preferences and (ii) prove that it satisfies the guarantee (1). Recall that you are allowed to refer to material covered in the lecture notes.)

Solution:

(This page is intentionally left blank.)

4 *(15 pts)* **Online bin-packing.** Recall the online bin-packing problem that we saw in Exercise Set 10: We are given an unlimited number of bins, each of capacity 1. We get a sequence of items one by one each having a size of at most 1, and are required to place them into bins as we receive them. Our goal is to minimize the number of bins we use, subject to the constraint that no bin should be filled to more than its capacity. An example is as follows:



Here, seven items have already arrived that we have packed in three bins. The newly arriving item of size 1/6 can either be packed in the first bin, third bin, or in a new (previously unused) bin. It cannot be packed in the second bin since $1/3 + 1/3 + 1/4 + 1/6 > 1$. If it is packed in the first or third bin, then we still use three bins, whereas if we pack it in a new bin, then we use four bins.

In this problem you should, assuming that all items have size at most $0 < \epsilon \leq 1$, design and analyze an online algorithm for the online bin-packing problem that uses at most

$$\frac{1}{1-\epsilon}\text{OPT} + 1 \text{ bins}, \tag{2}$$

where OPT denotes the minimum number of bins an optimal packing uses. In the above example, $\epsilon = 1/3$.

(In this problem you are asked to (i) design the online algorithm and (ii) prove that it satisfies the guarantee (2)*. Recall that you are allowed to refer to material covered in the lecture notes.)*

Solution:

(This page is intentionally left blank.)

5 *(consisting of subproblems **a**-**b**, 22 pts)* **Finding large cuts.**

5a *(9 pts)* Consider the following algorithm that takes as input an undirected graph $G = (V, E)$:

> SimpleCut($G = (V, E)$):
>
> 1. Let $\mathcal{H}$ be a 2-universal family of hash functions $h : V \to \{0, 1\}$.
> 2. Select $h \in \mathcal{H}$ at random.
> 3. **return** the vertex set $S = \{v \in V : h(v) = 0\}$.

Prove the following:

In expectation, the set S returned by SimpleCut cuts at least $|E|/2$ edges.

(In this problem you are asked to prove the above statement. Recall that you are allowed to refer to material covered in the lecture notes.)

Solution:

5b *(13 pts)* Design and analyze a polynomial-time algorithm for the following problem:

> **Input:** a vertex set V.
>
> **Output:** vertex subsets $S_1, S_2, \ldots, S_\ell \subseteq V$ with the following property:
>
> For every set of edges $E \subseteq \binom{V}{2}$, there is an $i \in \{1, 2, \ldots, \ell\}$ such that
>
> $$|\{e \in E : |e \cap S_i| = 1\}| \geq |E|/2\,,$$
>
> i.e., S_i cuts at least half the edges in $G = (V, E)$.

We remark that, since your algorithm should run in time polynomial in $n = |V|$, it can output at most polynomially (in n) many vertex sets. We also emphasize that the algorithm does **not** take the edge set E as input.

(In this problem you are asked to (i) design the algorithm, (ii) show that it runs in time polynomial in n, and (iii) prove that the output satisfies the property given in the problem statement. Recall that you are allowed to refer to material covered in the lecture notes.)

Solution:

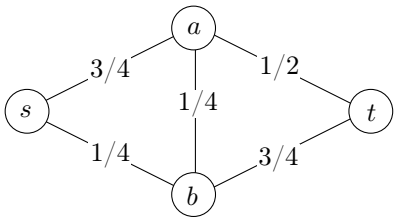
(This page is intentionally left blank.)

6 *(consisting of subproblems **a**-**b**, 23 pts)* **Finding a min s,t-cut using linear programming.** Consider an undirected graph $G = (V, E)$ and let $s \neq t \in V$. In the minimum (unweighted) s,t-cut problem, we wish to find a set $S \subseteq V$ such that $s \in S$, $t \notin S$ and the number of edges crossing the cut is minimized.

We shall use a linear program to solve this problem. Let P be the set of all paths between s and t in the graph G. The linear program has a variable y_e for each edge $e \in E$ and is defined as follows:

$$\begin{aligned} \text{minimize} \quad & \sum_{e \in E} y_e \\ \text{subject to} \quad & \sum_{e \in p} y_e \geq 1 \quad \forall p \in P, \\ & y_e \geq 0 \qquad \forall e \in E. \end{aligned}$$

For example, consider the following graph where the numbers on the edges depict the y_e-values of a feasible solution to the linear program:



The values on the edges depict a feasible but not optimal solution to the linear program. That it is feasible follows because each y_e is non-negative and $\sum_{e \in p} y_e \geq 1$ for all $p \in P$. Indeed, for the path s, b, a, t we have $y_{\{s,b\}} + y_{\{b,a\}} + y_{\{a,t\}} = 1/4 + 1/4 + 1/2 = 1$, and similar calculations for each path p between s and t show that $\sum_{e \in p} y_e \geq 1$. That the solution is not optimal follows because its value is 2.5 whereas an optimal solution has value 2.

6a *(10 pts)* Let OPT denote the number of edges crossing a minimum s,t-cut and let OPT_{LP} denote the value of an optimal solution the linear program. Prove that $\text{OPT}_{\text{LP}} \leq \text{OPT}$.

(In this problem you are asked to prove $\text{OPT}_{\text{LP}} \leq \text{OPT}$. Recall that you are allowed to refer to material covered in the lecture notes.)

Solution:

(This page is intentionally left blank.)

6b *(13 pts)* Prove that $\mathrm{OPT} \leq \mathrm{OPT}_{\mathrm{LP}}$, where OPT and $\mathrm{OPT}_{\mathrm{LP}}$ are defined as in **6a**.

Hint: Round a feasible linear programming solution y. In the (randomized) rounding it may be helpful to consider, for each vertex $v \in V$, the length of the shortest path from s to v in the graph where edge $e \in E$ has length y_e. For example, in the graph and linear programming solution depicted in the problem statement, we have that the length of the shortest path from s to a equals $1/2$.

(In this problem you are asked to prove $\mathrm{OPT} \leq \mathrm{OPT}_{\mathrm{LP}}$*. Recall that you are allowed to refer to material covered in the lecture notes.)*

Solution:

(This page is intentionally left blank.)