Homework 7

Problem 14.1

Modify algorithm 14.1 so that it picks all sets that are nonzero in the fractional solution (we call the modified algorithm R). Show that the algorithm also achieves a factor of f.

We first give the LP relaxation and its dual for set cover.

Primal:

$$\begin{aligned} \textit{minimize} \quad & \sum_{S \in \mathcal{S}} c(S) x_S & (1) \\ s.t. \quad & \forall e \in U, \sum_{S: e \in S} x_S \geq 1 & (2) \\ & \forall S \in \mathcal{S}, x_S \geq 0 & (3) \end{aligned}$$

Dual:

$$\begin{aligned} \textit{maximize} \quad & \sum_{e \in U} y_e & (4) \\ s.t. \quad & \forall S \in \mathcal{S}, \sum_{e: e \in S} y_e \leq c(S) & (5) \\ & \forall e \in U, y_e \geq 0 & (6) \end{aligned}$$

Suppose we have an optimal solution $y^* = \sum_{e \in U} y_e$ for the dual problem. We know that this objective value is the sum of the y_e values for each element in the universe. We also know that each constraint in the dual corresponds to a set; specifically, each constraint says that all elements within a set must be less than or equal to the cost of taking that set.

If we summed up all the constraints of the dual, we would get the following sum:

$$\sum_{S \in \mathcal{S}} \sum_{e \in S} y_e \quad (7)$$

This is similar to the objective value for the dual, except that it recounts elements that appear in multiple sets. However, by definition of f, each element can be obtained in at most f sets, so the most the objective value of the dual could be is

$$OPT_D^* = \sum_{e \in U} y_e \leq \frac{\sum_{S \in \mathcal{S}} \sum_{e \in S} y_e}{f} \quad (8)$$

We will multiply both sides of this inequality by f to get

$$f * OPT_D^* \leq \sum_{S \in \mathcal{S}} \sum_{e \in S} y_e \tag{9}$$

Given an optimal primal solution x^*, for every nonzero value x_j^*, we know from the primal complementary slackness conditions that $\forall \{S : x_s^* > 0\}, \sum_{e \in S} y_e = c(S)$. If we break down the sum of all dual constraints into two disjoint sets (based on whether or not the constraint's corresponding x_s^* value is equal to 0), we have:

$$\sum_{S \in \mathcal{S}} \sum_{e \in S} y_e = \sum_{S : x_s^* = 0} \sum_{e \in S} y_e + \sum_{S : x_s^* > 0} \sum_{e \in S} y_e \tag{10}$$

Getting rid of the first term on the RHS (which can't be negative), and plugging in the known cost from complementary slackness conditions, we have

$$\sum_{S \in \mathcal{S}} \sum_{e \in S} y_e \geq \sum_{S : x_s^* > 0} \sum_{e \in S} y_e = \sum_{S : x_s^* > 0} c(S) \tag{11}$$

The RHS of this equation is the output value of our modified algorithm R, because for every $x_s^* > 0$, we round x_s^* up to 1 and pay the entire $c(s)$. So we know that

$$output(R) = \sum_{S : x_s^* > 0} c(S) \leq \sum_{S \in \mathcal{S}} \sum_{e \in S} y_e \leq f * OPT_D^* \tag{12}$$

Because of LP duality theorem, and because the optimal value of the relaxed LP cannot be larger than the optimal value of the integer problem,

$$output(R) \leq f * OPT_D^* = f * OPT_P^* \leq f * OPT \tag{13}$$