Question #1: 2-out-of-4 Set Cover

In the 2-out-of-4 Set Cover problem you are given n, m and m sets $A_1, A_2, \ldots, A_m$ all subsets of $\{1, \ldots, n\}$ of size 4 (i.e. $A_1, \ldots, A_m \subseteq \{1, \ldots, n\}$, $|A_1| = \cdots = |A_m| = 4$) and asked to find the smallest possible subset of $\{1, \ldots, n\}$ that has at least two elements from each A_j, i.e. find the smallest $S \subseteq \{1, \ldots, n\}$ such that $|S \cap A_j| \geq 2$ for all $1 \leq j \leq m$.
As it turns out this problem is NP-hard which means that it is firmly believed by most researchers that no efficient algorithm can solve it exactly. Instead in this problem our goal is to find an α-approximation algorithm for a reasonably small constant α. In other words we want to find an efficient algorithm that always outputs a set S that has at least two elements from each A_j and whose size is no more than the size of the smallest possible such set *times* α for some $\alpha > 1$ which is reasonably small.

Solution We will use a linear program in our algorithm. To do this lets pretend for now that we want to solve the problem exactly using a linear program. What would be a natural choice of variables for a linear program formulating the 2-out-of-4 Set Cover problem? The most natural choice is to have one variable x_i for each element $i \in \{1, \ldots, n\}$ which signifies if that element is chosen to be in our solution set n. Given these variables it is natural for the objective function of our linear program to be $x_1 + x_2 + \cdots + x_n$. It is also natural to put two constraints for each $i \in \{1, \ldots, n\}$, one $x_i \geq 0$ and the other $x_i \leq 1$ because our hope is that x_i will be 1 to signify that the element i is selected to be in our solution set and 0 otherwise. There is one more type of constraint that we should add to our linear program. Remember that in the problem we are trying to solve for each j the solution set has to select at least two element from the set A_j it is then natural to add one constraint for each $1 \leq j \leq m$ of the form $\sum_{i \in A_j} x_i \geq 2$. The resulting linear program can be seen in Figure 1.

$$\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{n} x_i && && (1)\\
\text{variables:} \quad & x_1, x_2, \ldots, x_n && && (2)\\
& x_i \geq 0 && \forall 1 \leq i \leq n && (3)\\
& x_i \leq 1 && \forall 1 \leq i \leq n && (4)\\
& \sum_{i \in A_j} x_i \geq 2 && \forall 1 \leq j \leq m && (5)
\end{aligned}$$

Figure 1: Our linear program.

It is tempting to claim that the answer to our original problem (2-out-of-4 Set Cover) is just the answer to this linear program. Remember that to prove such a thing we should show that the answer to the LP is no more than the answer to the original problem and the answer to the original problem is no more than the answer to the LP. The first of these two statements are true and will be helpful to us.

Lemma 1. *The optimal solution to the LP has objective value less than or equal to the size of the best solution to the original (2-out-of-4 Set Cover) problem.*

Proof. Consider the optimal solution to the original problem and call it S. We will construct a valid solution to the LP with objective value $|S|$. This would imply that the optimal solution to the LP has objective value no more than $|S|$ (it can have even smaller objective value though.) Define the solution

$$x_i = \begin{cases} 1 & \text{if } i \in S, \\ 0 & \text{if } i \notin S. \end{cases}$$

Clearly this solution satisfies the constraints (3) and (4). It also satisfies constraints (5) because S is a valid solution of the 2-out-of-4 Set Cover problem we are trying to solve so for every $1 \leq j \leq m$, $|S \cap A_j| \geq 2$ so $\sum_{i \in A_j} x_i \geq 2$. It is also clear that the objective value of this solution is $\sum_i x_i = |S|$. □

The second statement one would want to prove, i.e. the from any answer to the LP one can construct an answer to the original problem S where $|S|$ is equal to the objective value of the LP is just not true! The reason is similar to what we saw in class for the Vertex Cover problem in general graphs; an LP solution can assigned $x_i = 1/4$ in which case it is not clear at all how to construct a solution to 2-out-of-4 Set Cover from this LP solution. Instead we will show to construct a solution to 2-out-of-4 Set Cover from an LP solution with $|S|$ *comparable* (but not necessarily equal) to the obective value of the LP.

Lemma 2. *Given a solution $x_1, \ldots, x_n$ to the LP in Figure 1 we can construct a solution S to the original 2-out-of-4 Set Cover problem such that $|S|$ is at most three times the objective value of the LP solution.*[1]

Proof. It is natural to define S as the set of i such that x_i is large. To be more precise we would like to define S as

$$S = \{i : x_i \geq \beta\},$$

for some number β. What is the correct choice for β? Remember that the property we require from S is that for each set A_j at least two element from A_j are in S. But the only tool we have for making sure S has such a property is that $x_1, \ldots, x_n$ is an LP solution so $\sum_{i \in A_j} x_i \geq 2$. Imagine that the set A_j is $A_j = \{1, 2, 3, 4\}$ so what we really need to do is to choose β in such a way such that if $x_1 + x_2 + x_3 + x_4 \geq 2$ then at least two of $x_1, \ldots, x_4$ at at least β. It is not hard to check that correct choice is to have $\beta = 1/3$. This way if at most one of $x_1, \ldots, x_4$ is at least β then that variable can contribute at most 1 to the sum $x_1 + x_2 + x_3 + x_4$ and the rest of the variable can constribute less than $1/3$ each so in total the sum has to be less than 2 which means that $x_1, \ldots, x_n$ is not an LP solution!
So to wrap things up we set

$$S = \{i : x_i \geq 1/3\},$$

and argue that for any $1 \leq j \leq m$ we have $\sum_{i \in A_j} x_i \geq 2$ so from the variables $\{x_i : i \in A_j\}$ at least 2 are more than or equal to $1/3$ so the set S selects at least two elemets from A_j for every j. On the other hand the set S contain all the i such that $x_i \geq 1/3$ so,

$$|S| = \sum_{i: x_i \geq 1/3} 1 \leq \sum_{i: x_i \geq 1/3} 3x_i \leq \sum_{i=1}^{n} 3x_i = 3 \times \text{"objective value of the LP solution"}. \qquad \square$$

Combining Lemmas 1 and 2 it is clear that the following algorithm for 2-out-of-4 Set Cover always outputs a solution no bigger than three times the size of the optimal solution.

[1] Do not worry about the number 3 that appeared magically here for now. The choice will become clear as we try to construct the set S and prove the lemma.

Input: n, m and sets $A_1, A_2, \ldots, A_m \subseteq \{1, \ldots, n\}$, such that $|A_1| = \cdots = |A_m| = 4$.
Output: $S \subseteq \{1, \ldots, n\}$ such that $|A_j \cap S| \geq 2$ for all $1 \leq j \leq m$.

Construct the LP of Figure 1.
$LP1 \leftarrow$ The LP of Figure 1
$x_1, \ldots, x_n \leftarrow$`LPSolver(`$LP1$`)`
$S \leftarrow \emptyset$
for $i \leftarrow 1$ **to** n **do**
  if $x_i \geq 1/3$ **then**
    $S \leftarrow S \cup \{i\}$
  end
end
return S