COMP 360 - Fall 2015 - Assignment 4

Due: 6:00 pm Nov 24th.

General rules: In solving these questions you may collaborate with other students but each student has to write his/her own solution. There are in total 110 points, but your grade will be considered out of 100. You should drop your solutions in the assignment drop-off box located in the Trottier Building.

---

1. (10 Points) Show that the following problem is in PSPACE:

    - Input: A CNF ϕ.
    - Output: The number of the truth assignments that satisfy ϕ.

    Solution: One can generate all the possible 2^n truth assignments, one by one (reusing the memory). Each such truth assignment takes n space. We will also have a variable A that is equal to the number of truth assignments satisfying the formula found so far. Every time that a new truth assignment is generated, we check whether it satisfies the formula or not, and then update the variable A accordingly. Note that A takes only $\log 2^n = n$ bits of memory. Hence in total the required space is going to be $O(n)$.

2. (10 Points) Given a set P of n points on the plane, consider the problem of finding the smallest circle containing all the points in P. Show that the following is a 2-factor approximation algorithm for this problem. Pick a point x in P, and set r to be the distance of the farthest point in P from x. Output the circle centered at x with radius r.

    Solution: First we show that all the points will be in the circle outputted by algorithm: If y is the point farthest from x, then $r = d(x, y)$. If there's a point p outside the circle, then $d(p, x) > r = d(x, y)$ which cannot be because y is the farthest point from x.

    Now we show that this is a 2-factor approximation: Suppose x' and r' are the center and radius of the optimal solution. Since x and y are inside the circle we have $d(x, c) \geq r'$ and $d(y, c) \geq r'$ so $d(x, c) + d(y, c) \geq 2r'$. From the triangle inequality we have $d(x, c) + d(y, c) \geq d(x, y) = r$ from combining these two we have $2r' \geq r$.

3. (10 Points) Consider the following optimization version of the Subset-Sum problem: Given positive integers $\{w_1, \ldots, w_n\}$ and a positive integer m. We want to find a set $S \subseteq \{1, \ldots, n\}$ such that $\sum_{i \in S} w_i \leq m$ and is maximized. Show that the following is a $\frac{1}{2}$-factor approximation algorithm:

    - Set $S := \emptyset$.
    - Sort the numbers such that $w_1 \geq w_2 \geq \ldots \geq w_n$.
    - For $i = 1, \ldots, n$:
    - if it is possible add i to S without violating $\sum_{i \in S} w_i \leq m$, then add i to S.

Solution: This follows just by definition of algorithm. Suppose $W^* \leq m$ is the optimal solution and the algorithm picked $w_1 \geq w_2 \geq \cdots \geq w_k$ and stopped. We show that $\sum_{i=1}^{k} w_i \geq W^*/2$. Suppose otherwise, then $w_{k+1} \leq w_k < W^*/2$ and hence $\sum_{i=1}^{k+1} w_i < W^* \leq m$, which shows that the algorithm should not have stopped after adding w_k, a contradiction.

4. Consider the MAX-SAT problem: Given a CNF formula ϕ on variables $x_1, \ldots, x_n$, find a truth assignment to the variables that maximizes the total number of satisfied clauses.

   (a) (10 Points) Show that the following is a $\frac{1}{2}$-factor approximation algorithm for MAX-SAT (meaning that: the output of the algorithm is always at least half of the optimum): Let σ_{true} be the truth assignment that assigns True to every variable, and σ_{false} be the truth assignment that assigns False to every variable. Compute the number of clauses satisfied by σ_{true} and σ_{false}, and output the better assignment.

   (b) (5 Points) Give a tight example: An input instance where this algorithm performs as bad as the $\frac{1}{2}$ factor.

   Solution: (a) holds because every clause in CNF is satisfied either with σ_{true} or with σ_{false}. Indeed, if a clause is not satisfiable with all the variables being true, it means that all the variables in this clause are with $\neg$, but then the clause would have been satisfiable with σ_{false}. As for (b), suppose $n = 2k$ and we have the following CNF- $(x_1 \& x_2 \& \ldots \& x_k \& \neg x_{k+1} \& \neg x_{k+2} \ldots \& \neg x_{2k}$, then clearly all clauses are satisfiable when $x_1 = x_2 = \ldots x_k = T$ and $x_{k+1} = x_{k+2} = \ldots x_{2k} = F$, however both σ_{true} and σ_{false} assignments will satisfy only half of the clauses.

5. (10 Points) Problem 10 of Chapter 11: Suppose you are given an $n \times n$ grid graph G. Associated with each node v is an integer weight $w(v) \geq 0$. You may assume that all the weights are distinct. Your goal is to choose an independent set S of nodes of the grid, so that the sum of the weights of the nodes in S is as large as possible. (The sum of the weights of the nodes in S will be called its total weight.) Consider the following greedy algorithm for this problem.

   - Start with $S := \emptyset$.
   - While some node remains in G:
     - Pick a node v of maximum weight.
     - Add v to S.
     - Delete v and its neighbors from G
   - Endwhile.

   Show that this algorithm returns an independent set of total weight at least $\frac{1}{4}$ times the maximum total weight of any independent set in the grid graph G.

   Solution: Since for every node v picked we remove the neighbors, the algorithm will not output any connected nodes thus the algorithm gives an independent set. Suppose that we pick a node v at some point in the algorithm. Let $v_1, \ldots, v_4$ be its neighbours. Note that none of $v_1, \ldots, v_4$ have been picked at this point (otherwise v would have been deleted). Since v has the maximum weight among the remaining vertices, we have $\text{weight}(v) \geq \text{weight}(v_i)$ for $i = 1 \ldots 4$. So

   $$4 \times \text{weight}(v) \geq \sum_{i=1}^{4} \text{weight}(v_i).$$

   If the optimal algorithm doesn't choose v and chooses a subset (or all four) of the neighbors instead, then it could be at most 4 times better.

6. Consider a directed bipartite graph $G = (V, E)$. We want to eliminate all the directed cycles of length 4 by removing a smallest possible set of vertices.

   (a) (5 points) Let $\mathcal{C}_4$ denote the set of all cycles of length 4 in the graph. Show that the following integer program models the problem:

$$\begin{array}{lll} \min & \sum_{v\in V} x_v & \\ \text{s.t.} & \sum_{u\in C} x_u \geq 1 & \forall C \in \mathcal{C}_4 \\ & x_u \in \{0,1\} & u \in V \end{array}$$

   Solution: For each vertex v, we have a variable x_v. These variables are 0/1 valued. The meaning of $x_v = 1$ is that we remove vertex v from the graph. The meaning of $x_v = 0$ is that we keep vertex v. Let OPT denote the optimum value for the original problem. Let OPT_{ip} denote the optimum value for the integer program. Let x^* be an optimum solution of the integer program. By the inequality constraint, the integer program will pick at least one vertex from each 4-cycle. Thus removing the vertices corresponding to $x^* = 1$ will remove all the 4-cycles. Therefore we have $\text{OPT} \leq \text{OPT}_{ip}$. On the other hand, take a minimum set of vertices whose removal kills all the 4-cycles. Setting $x_v = 1$ for these vertices clearly produces a feasible solution for the integer program. Therefore $\text{OPT}_{ip} \leq \text{OPT}$.

   (b) (5 points) Why does the optimal solution to the following relaxation provides a lower bound for the optimal answer to the above integer linear program? In other words why it is not necessary to have the constraints $x_u \leq 1$ in the relaxation?

$$\begin{array}{lll} \min & \sum_{v\in V} x_v & \\ \text{s.t.} & \sum_{u\in C} x_u \geq 1 & \forall C \in \mathcal{C}_4 \\ & x_u \geq 0 & \forall u \in V \end{array}$$

   Solution: We claim that in any optimum solution x^*, $x_u^* \leq 1$ for all u. Suppose there exists some u such that $x_u^* > 1$. Round down the value of this variable to 1. Note that all the inequality constraints will still be satisfied. So we still have a feasible solution. On the other hand, the optimum value will go down, which is a contradiction.

   (c) (10 points) Give a simple 4-factor approximation algorithm for the problem based on rounding the solution to the above linear program.

   Solution: As before, let x^* be the optimum solution. The rounding is as follows. If $x_u^* \geq 1/4$, set $x_u^* = 1$, otherwise set $x_u^* = 0$. First let's check that we get a feasible solution to our problem. In each inequality constraint, it must be the case that at least one of the variables has value $\geq 1/4$. Thus in our rounded solution, we pick at least one vertex from each 4-cycle. So we kill all the 4-cycles as required. Let OPT^* be the optimum for the linear program, let OPT be the optimum for the original problem and let A be the value obtained by rounding the optimum of the linear program. Clearly $\text{OPT}^* \leq \text{OPT}$. Also, by our rounding scheme, we have $A \leq 4\text{OPT}^*$. Thus, $A \leq 4\text{OPT}$, i.e. our solution is within a factor 4 of the optimum.

   (d) (15 points) Let L and R denote the set of the vertices in the two parts of the bipartite graph. (Every edge has one endpoint in L and one endpoint in R). Let x^* denote an optimal solution to the linear program in Part (b). We round x^* in the following way:

   For every $u \in V$,

   - if $u \in R$ and $x_u^* \geq 1/2$, set $\widehat{x}_u = 1$.
   - if $u \in L$ and $x_u^* > 0$, set $\widehat{x}_u = 1$.
   - Otherwise set $\widehat{x}_u = 0$.

Show that $\widehat{x}$ is a feasible solution to the integer linear program.

Solution: Observe that each 4-cycle contains two vertices from L and two vertices from R. Consider an inequality constraint of the linear program (so we are considering a fixed 4-cycle). If $x_u^* > 0$ for one of the two vertices in L, $\widehat{x}_u$ will be set to 1 and therefore this inequality will be satisfied. On the other hand, if $x_u^* = 0$ for both vertices in L, then it must be the case that $x_v^* \geq 1/2$ for one of the vertices in R. Thus this vertex will be rounded to 1 and the inequality will be satisfied.

(e) (10 points) Consider the dual of the relaxation:

$$\begin{array}{lll} \max & \sum_{C\in\mathcal{C}_4} y_C & \\ \text{s.t.} & \sum_{C\in\mathcal{C}_4, u\in C} y_C \leq 1 & \forall u \in V \\ & y_C \geq 0 & \forall C \in \mathcal{C}_4 \end{array}$$

and let y^* be an optimal solution to the dual. Use the complementary slackness to prove the following statement: For every $C \in \mathcal{C}_4$ either we have $|\{u : \widehat{x}_u = 1\}| \leq 3$ or $y_C^* = 0$.

Solution: Suppose $|\{u : \widehat{x}_u = 1\}| > 3$. Then all the variables for that cycle must be rounded to 1. For that to happen, it must be that $x_u^* \geq 1/2$ for the vertices in R and $x_u^* > 0$ for the vertices in L. Thus, we must have $\sum_{u\in C} x_u^* > 1$, i.e. the constraint is not tight. By complementary slackness, this means $y_C^* = 0$.

(f) (10 points) Use the complementary slackness and the previous parts to show that our rounding algorithm is a 3-factor approximation algorithm.

Solution: As mentioned before, we have $\sum_{u\in V} x_u^* = \text{OPT}^* \leq \text{OPT}$. Thus, we are done once we show

$$\sum_{u\in V} \widehat{x}_u \leq 3\text{OPT}^*.$$

Note that if $\widehat{x}_u = 1$, $x_u^* > 0$. Therefore, by complementary slackness, $\sum_{C\in\mathcal{C}_4, u\in C} y_C^* = 1$. The variables $\widehat{x}_u$ are 0/1 valued, so we can write

$$\sum_{u\in V} \widehat{x}_u = \sum_{u\in V} \widehat{x}_u \sum_{C\in\mathcal{C}_4, u\in C} y_C^* = \sum_{u\in V} \sum_{C\in\mathcal{C}_4, u\in C} \widehat{x}_u y_C^*.$$

We now change the order of the sums and get

$$\sum_{u\in V} \sum_{C\in\mathcal{C}_4, u\in C} \widehat{x}_u y_C^* = \sum_{C\in\mathcal{C}_4} \sum_{u\in C} \widehat{x}_u y_C^* = \sum_{C\in\mathcal{C}_4} y_C^* \sum_{u\in C} \widehat{x}_u.$$

From part (e) of the question, we know that if $y_C^* \neq 0$, then $\sum_{u\in C} \widehat{x}_u \leq 3$. Therefore the above quantity can be upper bounded by $3\sum_{C\in\mathcal{C}_4} y_C^* = 3\text{OPT}^*$ (the equality follows from duality). Putting things together, we have shown

$$\sum_{u\in V} \widehat{x}_u \leq 3\text{OPT}^*$$

as required.