Sunghyeon Jo
Course: Project-Hardness
Instructor: Sunghyeon Jo

Homework - #2
Term: Fall 2022

Problem 1

Decide each problem in P or NP-Complete. If it is in P, decribe a polynomial time algorithm. You can assume that SAT, 3SAT, Cycle-through-2-vertices are NP-Complete (Note that this is **NOT** the easiest problem of the assignment).

- *SAT*. Given a boolean formula with CNF form. Decide if it is satisfiable.
- *3SAT*. Given a boolean formula with CNF form, and every clause of the formula has at most three literal. Decide if it is satisfiable.
- *Cycle-through-2-vertices*. Given directed graph $G = (V, E)$ and $s, t \in V$. Decide if there is a simple cycle in G which contains two vertices s, t.

1. *Horn-SAT*. Given a boolean formula with CNF form, and every clause of the formula has at most one positive literal. Decide if it is satisfiable.
2. *Set-Cover*. Given $n, m, k \in \mathbb{N}$ and m sets $A_1, A_2, .., A_m \subset \{1, 2, .., n\}$. Decide if it exists a subset $S \subset \{1, 2, .., n\}$ such that $S \cap A_j \neq \phi$ and $|S| \leq k$.
3. *Set-Cover-2*. Given $n, m, k \in \mathbb{N}$ and m sets $A_1, A_2, .., A_m \subset \{1, 2, .., n\}$. Decide if it exists a subset $S \subset \{1, 2, .., m\}$ such that $\cup_{i \in S} A_i = \{1, 2, .., n\}$ and $|S| \leq k$.
4. *Cycle-1-mod-3*. Given directed graph $G = (V, E)$. Decide if there is a simple cycle in G which its length is 1 modulo 3.
5. *Cycle-K*. Given directed graph $G = (V, E)$ and an integer $K \geq |V|^2$. Decide if there is a closed walk in G which its length is equal to K. You can use arbitrary source(research, wiki, etc..) for solve this problem.

Solution:

1. *Horn-SAT*.

   If there is no clause with single positive literal, $(0, 0, .., 0)$ satisfies it. Otherwise, assign variables in those clause to 1 and remove assigned variable in each clauses (if already satisfied, remove the clause). remaining clauses have at most one positive literal, so we can repeat it until no clauses with single positive literal. Then just assign 0 for remaining variables. It is guaranteed that a satisfying assignment. Therefore, it is **always satisfiable** and there is linear algorithm for finding satisfying assignment.

2. *Set-Cover*. Given $n, m, k \in \mathbb{N}$ and m sets $A_1, A_2, .., A_m \subset \{1, 2, .., n\}$. Decide if it exists a subset $S \subset \{1, 2, .., n\}$ such that $S \cap A_j \neq \phi$ and $|S| \leq k$.

   Given boolean formula $\varphi = C_1 \wedge ...C_s$ of 3SAT form with variables $x_1, .., x_t$. Let's make Set-Cover instance like below:

   - $n = 2t, m = t + s, k = t$
   - $A_i = \{i, n + i\}$. $i \in S$ means $x_i = 1$ and $n + i \in S$ means $x_i = 0$.
   - A_{t+i} corresponds to C_i. For example, If $C = x_i \vee \neg x_j \vee x_k$, corresponding set A is $\{i, n + j, k\}$.

   $k = t$ guarantees that exactly one of $i, n+i$ are in S and each clause C_i is equivalent to the condition $S \cap A_{t+i} \neq \phi$. Therefore, existence of the Set-Cover is equivalent to satisfiability of φ and hence 3SAT $\leq_p$ Set-Cover. Therefore, Set-Cover is NP-Complete.

3. *Set-Cover-2*. Given $n, m, k \in \mathbb{N}$ and m sets $A_1, A_2, .., A_m \subset \{1, 2, .., n\}$. Decide if it exists a subset $S \subset \{1, 2, .., m\}$ such that $\cup_{i \in S} A_i = \{1, 2, .., n\}$ and $|S| \leq k$.

   We will show Set-Cover can be reduced to Set-Cover-2. Consider an instance of Set-Cover and define $B_i = \{k | i \in A_k\}$. If $S = \{s_1, ..., s_k\} \subset \{1, 2, .., n\}$ is a solution of Set-Cover problem, then S is a solution of Set-Cover-2 instance given as B_i as subset of $\{1, 2, .., m\}$ and vise versa. Therefore, Set-Cover $\leq_p$ Set-Cover-2. Since Set-Cover is NP-Complete, Set-Cover-2 is also NP-Complete.

4. *Cycle-1-mod-3*. Given directed graph $G = (V, E)$. Decide if there is a simple cycle in G which its length is 1 modulo 3.

   We want to show that Cycles-through-2-vertices $\leq_p$ Cycles-1-mod-3. Moreover, for any fixed integer $B \leq 3$ and $1 \leq A < B$, Cycles-A-mod-B can be reduced in the same way.

   We replace an directed edge by a direct path with constant length (add intermediate vertices on the edge).

   - For any edges outgoing from s, we replace the edge by the path with length 2.
   - For any edges outgoing from t, in the same way replace it by the length 2 path.
   - Otherwise, we replace the edge by the path with length 3.

   Then, there is a simple cycle through s and t in the original graph iff we can find a cycle with length $3k + 1$ in the graph. Therefore, Cycles-through-2-vertices $\leq_p$ Cycles-1-mod-3 and hence Cycles-1-mod-3 is NP-Complete.

   In general, we can build Cycle-A-mod-B instance by giving the different edge length. Replace path length to $(2, 2, 3)$ to (x, y, B) such that $x + y \equiv B$ and x, y are not divided by B. Therefore, Cycles-A-mod-B is NP-Complete.

5. *Cycle-K*. Given directed graph $G = (V, E)$ and an integer $K \geq |V|^2$. Decide if there is a closed walk in G which its length is equal to K. You can use arbitrary source(research, wiki, etc..) for solve this problem.

   Let A as the adjacency matrix of the graph, there is a closed walk with length K if and only if $Tr(A^K) > 0$. Thus O(log K) matrix multiplication does. Since matrix multiplication can be done in $O(N^3)$ time, it is in P. Large numbers can be avoided by use or operation instead of addition.

   I'll introduce a linear-time algorithm for this problem. Define **period** of an strongly connected graph as the gcd(great common divisor) of all closed walks' length. Then period can be calculated as follows:

   - Do DFS traversal starting at arbitrary vertex s.
   - Let d_v by distance between s and v in DFS tree.
   - Calculate g = gcd of $d_v - d_u - 1$ for all edge $u \rightarrow v$.

   Then g is the period. Proof of this is in `http://www.math.clemson.edu/~shierd/Shier/markov.pdf`.

   Since every closed walk can be decomposed into simple cycles, thsere are simple cycles of length $l_1, .., l_k < n$ such that $gcd(l_1, ..., l_k) = g$

   From proposition 2.2 of `https://www.cis.upenn.edu/~cis5110/Frobenius-number.pdf`, we can see that there always be a closed walk of length L if $g|L$ and $L \geq n^2$.

   Threfore, Cycle-K can be solved in linear time by following steps.

   - Find SCC(Strongly Connected Component)s of G.
   - Calculate period of each SCC and check if it divides K.
   - If there is no period divisor of K, there is no length K closed walk. Otherwise, length K closed walk exists in G.

Problem 2
We didn't look GAP-SAT in the lecture. **GAP-3SAT[ρ, 1]** is the gap version of 3SAT such that an instance is given in the form of 3CNF, decide if it is satisfiable or there is no assignment that satisfies ρ of the clauses(It is guaranteed that input is included in those two cases).

- Let GAP-E3SAT is almost same to GAP-3SAT but all clauses has exactly 3 literals and there is **no two same** literals in a clause. Show that GAP-E3SAT$[7/8, 1]$ is in P.

Solution:

Given boolean formula φ for GAP-E3SAT.

Each clause of φ has 3 literals and they are distinct. If a clause have both positive and negative literal for a variable, the clause is constantly true. Otherwise, the clause have three literals of distinct variables. Therefore, for every clause C_i of φ, $E[C_i] \geq 7/8$.

Due to linearity of expectation, expected satisfying clauses of ϕ by random assignment is at least 7/8 of all clauses. Thus, there is an assignment that satisfies at least 7/8 of clauses. Therefore, GAP-3SAT[7/8, 1] is in P since it is always satisfiable.

Moreover, there is an algorithm finding satisfying assignment and terminates in polynomial time expected. Let's calculate maximum probability of failing satisfying φ of N clauses and M variables, when assignment is random. Among the 2^M assignemnts, let S be successful assignments and F be failed assignments cannot satisfy 7/8 of clauses. Then, it holds:

Σ_F (7/8 - (# of satisfying clauses)/N) $\leq \Sigma_S$ ((# of satisfying clauses)/N - 7/8).

Since 7/8 - (# of satisfying clauses)/$N \geq 1/8N$ for failing assignments, $|F|/8N \leq |S|/8$ holds. Therefore, it is almost guaranteed that there are satisfying assignment among $c \cdot N$ random assignments. An algorithm that setting variables randomly and check the satisfiability until finding solution terminates after checking expected $O(N)$ assignments.

Problem 3
Show that 3SAT $\leq_p$ MAX-2SAT.

Solution:

Let's think about a clause $a \vee b \vee c$ (Note that a, b, c are literal, not variable. So they could be either positive or negative).
We want to make a set of 2CNF clauses such that the number of maximum satisfying clauses(#MAXSAT) differs only when $a \vee b \vee c = 0$.
Consider eight 2CNFs : $(a \vee b), (b \vee c), (c \vee a), (\neg a \vee y), (\neg b \vee y), (\neg c \vee y), (\neg y \vee \neg y), (\neg y \vee \neg y)$
Assume that value of a, b, c are fixed and we can choose the value of y.
When two or three of a, b, c are 1, then #MAXSAT $= 6$ can be achieved by setting $y = 1$.
When only one literal is 1, then #MAXSAT $= 6$ can be achieved by setting $y = 0$.
When $a = b = c = 0$, then #MAXSAT $= 5$ since first three clauses cannot be satisfied.
Therefore, #MAXSAT $= 6$ if and only if $a \vee b \vee c = 1$.

Given 3CNF formula φ composed of N clauses, we can build corresponding 2CNF formula ϕ by below rule.

- For each clause $a \vee b \vee c$, replace it to $(a \vee b) \wedge (b \vee c) \wedge (c \vee a) \wedge (\neg a \vee y) \wedge (\neg b \vee y) \wedge (\neg c \vee y) \wedge (\neg y \vee \neg y) \wedge (\neg y \vee \neg y)$

Note that additional variable y should be different variable for different clauses.

Then (#MAXSAT of ϕ) $= 6N$ if and only if φ is satisfiable.
Since the number of variables and clauses of new formula are linear to original formula, there is a polynomial reduction from 3SAT to MAX-2SAT.

Therefore, 3SAT $\leq_p$ MAX-2SAT.

Problem 4
We skipped some procedure in the proof of Schafer's Dichotomy Theorem. Prove the following lemma.

- If relation R is not bijunctive, then $Rep(\{R, [x \neq y], [x \vee y]\})$ contains the relation "exactly one of x,y,z"

Solution:

This problem was way harder than my first thought :(

Consider a relation $R(x_1, .., x_n) \subset \{0,1\}^n$. For an assignment s and a set of variables $Y \subset V = \{x_1, .., x_n\}$, define $s \bigoplus Y$ as an assignment that differs to s by only variables in Y.
Also, for two assignment s_1, s_2, define $s_1 \bigoplus s_2$ as set of variables that differs in two assignments.
In original paper, Schaefer proved a lemma that those two condition is equivalent:

(a) R is bijunctive.

(b) For every $A, B \subset V$, it holds that if $r, r \bigoplus A, r \bigoplus B \in R$ then $r \bigoplus (A \cap B) \in R$.

I'll prove (b) $\Rightarrow$ (a) in somewhat different way to Schaefer by define the concept of closed relation.

Define a relation $R(x_1, .., x_n) \in \{0,1\}^n$ is **closed** when

- If $(a_1, ..., a_n) \in \{0,1\}^n$ satisfies that $\forall_{1 \leq i,j \leq n} \exists_{r \in R} (a_i, a_j) = (r_i, r_j)$, then $(a_1, .., a_n) \in R$. (I'll write assignment r as $(r_1, ..., r_n)$, it means the value of x_i is r_i in assignment r.)

For example, $R_1 = \{(0,0,1), (0,1,0), (0,0,1)\}$ is not closed since $(0,0,0) \notin R_1$.

I claim that (b) $\Rightarrow$ R is closed $\Rightarrow$ (a).

1. R is closed $\Rightarrow$ (a)
   Assume that R is closed.
   For each $1 \leq i, j \leq n$, let $Z_{i,j}$ be a set of distinct value of pair (x_i, x_j) for elements of R.
   If $(a_1, ..., a_n)$ satisfies that $(a_i, a_j) \in Z_{i,j}$ for all i, j, then $(a_1, ..., a_n)$ should be included in R because R is closed. Therefore, R is the set of elements which satisfy all $Z_{i,j}$ conditions. For each $(z_i, z_j) \notin Z_{i,j}$, we can think of an 2CNF clause corresponds to (z_i, z_j). For example, if $(0,0) \notin Z_{i,j}$ then at least one of x_i and x_j should be 1, so $x_i \vee x_j$ is corresponding 2CNF clause. Consider a 2CNF formula φ consists of clauses generated by all missing elements of $Z_{i,j}$ for every i, j. It is easy to see that $s \in \varphi \Leftrightarrow s \in R$. Therefore, R is bijunctive if R is closed.

2. (b) $\Rightarrow$ R is closed
   Assume that R is not closed.
   Then there is an assignment $a = (a_1, ..., a_n) \notin R$ satisfies that $\forall_{1 \leq i,j \leq n} \exists_{r \in R} (a_i, a_j) = (r_i, r_j)$.
   So $p, q, r \in R$ exists such that $(p_1, p_2) = (a_1, a_2), (q_1, q_3) = (a_1, a_3), (r_2, r_3) = (a_2, a_3)$
   If (a_1, a_2, a_3) is not equal to all three of $(p_1, p_2, p_3), (q_1, q_2, q_3), (r_1, r_2, r_3)$, let $A = p \bigoplus q, B = p \bigoplus r$. Since (b) holds and $p, q = p \bigoplus A, r = p \bigoplus B$ are elements of R, $p \bigoplus (A \cap B) \in R$. The value of (x_1, x_2, x_3) in assignment $p \bigoplus (A \cap B)$ is equal to (a_1, a_2, a_3). Therefore, there is an element of R which have same value with a for first 3 variables x_1, x_2, x_3. In the same way, we can show that $(a_1, a_2, a_3, ..., a_n) \in R$. Since it is a contradiction, R is closed.

Now we will start the main proof.

If a relation R is not bijunctive, R does not holds (b). Therefore, there is an assignment s and $A, B \subset V$ such that $s, s \bigoplus A, s \bigoplus B \in R$ but $s \bigoplus (A \cap B) \notin R$.

Form a relation R' as negate all variables x_i which $s_i = 1$. Then $R' \in Rep(\{R, [x \neq y], [x \vee y]\})$ and $(0, 0, ..., 0) \in R'$.

Form a relation $B(y_0, y_1, y_2, y_3)$ by replacing each variable x_i of R' as following 4 cases:

- $x_i \in A \cap B \to \neg y_1$
- $x_i \in A \setminus B \to y_2$
- $x_i \in B \setminus A \to y_3$
- $x_i \in (A \cup B)^C \to y_0$.

$(0,1,0,0), (0,0,1,0), (0,0,0,1) \in B, (0,0,0,0) \notin B$ holds.
Form a relation $B' = B \wedge (\neg y_1 \vee \neg y_2) \wedge (\neg y_2 \vee \neg y_3) \wedge (\neg y_3 \vee \neg y_1) \wedge (\neg y_0 \vee \neg y_0)$.
Then there is solution (y_0, y_1, y_2, y_3) iff exactly one of y_1, y_2, y_3 is 1.

Therefore, $Rep(\{R, [x \neq y], [x \vee y]\})$ contains the relation "exactly one of x,y,z".

Problem 5
Show that NAE-E3SAT is NP-Complete using Schaefer's Dichotomy Theorem.

Solution:
Consitder a relation $R(x, y, z) = (x \vee y \vee z) \wedge (\neg x \vee \neg y \vee \neg z)$. It is obvious that this relation can be written in NAE-E3SAT form. It's enough to show that R is not in 6 of P cases.

- Trivially, R is neither 1-valid nor 0-valid.
- A 2CNF clause is either $x \vee \neg x$ form or not satisfied by at least 2 assignments of (x, y, z). First case is ignoreable and latter case should satisfy except 2 assignments: $(0,0,0), (1,1,1)$. But any 2CNF clauses not holds the condition. Therefore, R is bijunctive.
- Assume that R is equivalent to boolean formula ϕ. There could not be nontrivial 2-literal clauses (see above) and only $x \vee y \vee z, \neg x \vee \neg y \vee \neg z$ are possible clauses. but $x \vee y \vee z$ have 3 positive literal and $\neg x \vee \neg y \vee \neg z$ have 3 negative literal. Since $R(x, y, z) \neq x \vee y \vee z$ and $R(x, y, z) \neq \neg x \vee \neg y \vee \neg z$, R is neither weakly negative nor weakly positive.
- Nontrivial affine formulas are not satisfied by at least half of assignments, but $R(x, y, z)$ is satisfied by 6 out of 8 assignments. Therefore R is not affine.

From above, we can conclude that NAE-E3SAT is NP-Complete.
