CMPT 710 - Complexity Theory: Lecture 9

Valentine Kabanets

October 2, 2007

1 NP-completeness of 3-COL and SubsetSum

3-COL $= \{G \mid G$ is a 3-colorable graph$\}$ (recall that a 3-colorable graph is a graph whose vertices may be colored with colors 0,1, and 2 in such a way that the endpoints of every edge receive different colors).

Theorem 1. *3-COL is* NP*-complete.*

Proof. We need to prove that

1. 3-COL is in NP, and
2. 3-COL is NP-hard (i.e., every language $L \in \mathsf{NP}$ reduces to 3-COL).

We prove (1) by giving the following NP algorithm for 3-COL: Given a graph G, nondeterministically guess an assignment of colors 0,1,2 to the vertices of G; check (in deterministic polytime) that the guessed coloring is proper, i.e., that no edge has both of its endpoints colored with the same color.

To prove (2), we reduce NAE-SAT to 3-COL. Given a 3-CNF formula $\phi(x_1, \dots, x_n)$, we construct a graph G_ϕ such that $\phi \in$NAE-SAT iff $G_\phi \in$3-COL. Our graph G_ϕ will have
vertices:

- $x_1, \bar{x}_1, x_2, \bar{x}_2, \dots, x_n, \bar{x}_n$ (i.e., one vertex for each literal),
- a vertex u, and
- a triple of vertices for each clause $C_j = (l_{j_1} \vee l_{j_2} \vee l_{j_3})$ labeled by $v(j, l_{j_1}), v(j, l_{j_2}), v(j, l_{j_3})$, respectively.

Our graph will have
edges:

- $(x_i, \bar{x}_i)$ for each $1 \le i \le n$,
- (u, x_i) and $(u, \bar{x}_i)$ for each i,

- the triple of vertices corresponding to a clause will be connected to each other (i.e., every clause $C_j = (l_{j_1} \vee l_{j_2} \vee l_{j_3})$ corresponds to a triangle on the vertices $v(j, l_{j_1}), v(j, l_{j_2}), v(j, l_{j_3})$, and
- for every clause $C_j = (l_{j_1} \vee l_{j_2} \vee l_{j_3})$, there are three edges $(v(j, l_{j_1}), l_{j_1})$, $(v(j, l_{j_2}), l_{j_2})$, and $(v(j, l_{j_3}), l_{j_3})$ (i.e., each vertex in a clause-triangle is connected to the corresponding literal-vertex).

We now prove the correctness of our reduction. First, assume that G_ϕ is 3-colorable. Without loss of generality, the vertex u is colored with color 2. So, each of the literal-vertices connected to u will get colors 0 or 1. Our truth assignment will set variable x_i to True, if vertex x_i is colored with color 1; and to False, if vertex x_i is colored with 0. Now we argue that this assignment is satisfying for ϕ in the NAE-SAT sense, i.e., that every clause of ϕ has at least one true literal and at least one false literal.

Consider any clause $C_j = (l_{j_1} \vee l_{j_2} \vee l_{j_3})$ corresponding to the triangle on vertices $v(j, l_{j_1}), v(j, l_{j_2}), v(j, l_{j_3})$ of G_ϕ. Suppose that all literals in C_j are assigned True by our truth assignment. Then it means that the vertices $l_{j_1}, l_{j_2}, l_{j_3}$ are all colored with color 1. So color 1 cannot be used to color the vertices of the triangle on $v(j, l_{j_1}), v(j, l_{j_2}), v(j, l_{j_3})$. But we cannot color a triangle with just two colors! A contradiction. So, at least one literal in clause C_j is assigned False. A similar argument shows that at least one literal in C_j is assigned True. So ϕ is satisfied in the NAE-SAT sense.

Now we prove the other direction. Given an assignment a to ϕ which satisfies ϕ in the NAE-SAT sense, we define a coloring for G_ϕ as follows. The vertex u gets color 2. A vertex x_i gets color 1, if x_i is set to True by the assignment a, and color 0 otherwise. Consider the triangle corresponding to each clause $C_j = (l_{j_1} \vee l_{j_2} \vee l_{j_3})$. Since assignment a is satisfying in the NAE-SAT sense, there is at least one true literal and at least one false literal in C_j. W.l.o.g., assume that l_{j_1} is True, and l_{j_2} is False. Then we color the vertex $v(j, l_{j_1})$ with color 0, the vertex $v(j, l_{j_2})$ with color 1, and the vertex $v(j, l_{j_3})$ with color 2. It is not hard to verify that this coloring is indeed a proper 3-coloring of our graph. □

Another NP-complete problem is the Subset Sum problem: Given numbers $a_1, \ldots, a_n, T$ in binary, decide if there is a subset $S \subseteq \{1, \ldots, n\}$ such that $\sum_{i \in S} a_i = T$.

Theorem 2. *Subset Sum problem is* NP*-complete.*

Proof. The fact that it is in NP is obvious. To show NP-hardness, we do a reduction from 3-SAT. Given a 3cnf on n variables and m clauses, we define the following matrix of decimal digits. The rows are labeled by literals (i.e., x and $\bar{x}$ for each variable x), the first n columns are labeled by variables, and another m columns by clauses.

For each of the first n columns, say the one labeled by x, we put 1's in the two rows labeled by x and $\bar{x}$. For each of the last m columns, say the one corresponding to the clause $\{x, \bar{y}, z\}$, we put 1's in the three rows corresponding to the literals occurring in that clause, i.e., rows x, $\bar{y}$, and z. We also add $2m$ new rows to our table, and for each clause put two 1's in the corresponding column so that each new row has exactly one 1. Finally, we create the last row to contain 1's in the first n columns and 3 in the last m columns.

The $2n + 2m$ rows of the constructed table are interpreted as decimal representations of $k = 2n + 2m$ numbers $a_1, \dots, a_k$, and the last row as the decimal representation of the number T. The output of the reduction is $a_1, \dots, a_k, T$.

Now we prove the correctness of the described reduction. Suppose we start with a satisfying assignment to the formula. We specify the subset S as follows: For every literal assigned the value True (by the given satisfying assignment), put into S the corresponding row. That is, if x_i is set to True, add to S the number corresponding to the row labeled with x_i; otherwise, put into S the number corresponding to the row labeled with $\bar{x}_i$. Next, for every clause, if that clause has 3 satisfied literals (under our satisfying assignment), don't put anything in S. If the clause has 1 or 2 satisfied literals, then add to S 2 or 1 of the dummy rows corresponding to that clause. It is easy to check that the described subset S is such that the sum of the numbers yields exactly the target T.

For the other direction, suppose we have a subset S that makes the subset sum equal to T. Since the first n digits in T are 1, we conclude that the subset S contains exactly one of the two rows corresponding to variable x_i, for each $i = 1, \dots, n$. We make a truth assignment by setting to True those x_i which were picked by S, and to False those x_i such that the row $\bar{x}_i$ was picked by S. We need to argue that this assignment is satisfying. For every clause, the corresponding digit in T is 3. Even if S contains 1 or 2 dummy rows corresponding to that clause, S must contain at least one row corresponding to the variables, thereby ensuring that the clause has at least one true literal. □