Computability and Complexity

Lecture 4, Thursday August 10, 2023

Ari Feiglin

Let us define the following decision problem,

$$\mathsf{3SAT} = \left\{\varphi \,\middle|\, \begin{array}{l}\varphi \text{ is a satisfiable boolean formula in CNF, and each part of the conjunction is the disjunction of at} \\ \text{most three literals}\end{array}\right\}$$

This is a restriction of the decision problem SAT. We've actually already proved that $\mathsf{3SAT}$ is **NP**-complete since our proof of the Cook-Levin theorem actually gave us a reduction from CSAT to $\mathsf{3SAT}$, as our reduction to SAT defined a formula where each part of the conjunction is the disjunction of three literals. This shows that SAT is not **NP**-hard since we can't restrict the number of literals in each disjunction, as $\mathsf{3SAT}$ is also **NP**-hard.

But it turns out that

$$\mathsf{2SAT} = \left\{\varphi \,\middle|\, \begin{array}{l}\varphi \text{ is a satisfiable boolean formula in CNF, and each part of the conjunction is the disjunction of at} \\ \text{most two literals}\end{array}\right\}$$

is in **P**.

Example 4.1:

Let us define the following decision problem,

$$\mathsf{Clique} = \{(G,k) \mid G \text{ is an unordered graph with a clique whose size is at least } k\}$$

then Clique is **NP**-complete.

Obviously Clique is in **NP**, as we can define the verifier $V\big((G,k),C\big)$ and verify that C is a clique of G of size $\geq k$. Since $|C| \leq |G|$ for a clique, and this can be done in polynomial time on $|C|$, this is a polynomial proof system as required.

Let us define a reduction from IS to Clique, meaning $\mathsf{IS} \leq \mathsf{Clique}$ and so Clique is **NP**-hard. Given an input $\big(G = (V,E),k\big)$ we define the graph $G' = (V,E^c)$. Then $(G,k) \in \mathsf{IS}$ if and only if $(G',k) \in \mathsf{Clique}$ (ie. $(G,k) \mapsto (G',k)$ is a Karp reduction from IS to Clique). If $(G',k) \in \mathsf{Clique}$ then suppose C is a clique of G of size $\geq k$, then for every $u,v \in C$ then $(u,v) \in E^c$ and so $(u,v) \notin E$. So C is an independent set of size $\geq k$ in G', and so $(G,k) \in \mathsf{IS}$ as required. The proof for the converse is similar.

Thus Clique is indeed **NP**-complete as required.

Definition 4.2:

If $G = (V,E)$ is a graph, a **vertex cover** is a set of vertices S which touches every edge in G. In other words, for every $(u,v) \in E$, either u or v is in S.

Example 4.3:

We define the following decision problem,

$$\mathsf{VertexCover} = \{(G,k) \mid G \text{ has a vertex cover whose size is at most } k\}$$

We will show that $\mathsf{VertexCover}$ is **NP**-complete.

It is easy to see that $\mathsf{VertexCover}$ is in **NP**. Notice that C is a vertex covering if and only if $V \setminus C$ is an independent set: if $u,v \in V \setminus C$ then $(u,v) \notin E$ (as then either u or v would be in C). And if $V \setminus C$ is an independent set, then for every $(u,v) \in E$ then u or v cannot be in $V \setminus C$ (ie. one is in C) as $V \setminus C$ is independent.

So G has a vertex covering of size k if and only if it has an independent set of size $|V| - k$, and therefore the mapping $(G,k) \mapsto (G, |V| - k)$ is a Karp reduction from IS to $\mathsf{VertexCover}$, and therefore $\mathsf{VertexCover}$ is **NP**-complete as required.

Definition 4.4:

A **dominating set** of a graph $G = (V, E)$ is a set of vertices S such that for every $u \in V$, either u is in S or u has a neighbor which is in S.

Example 4.5:

We define the following decision problem,

$$\mathsf{DominatingSet} = \{(G, k) \mid G \text{ has a dominating set whose size is at most } k\}$$

We will show that $\mathsf{DominatingSet}$ is **NP**-complete.

Again, it is easy to see that $\mathsf{DominatingSet}$ is in **NP**. We will prove this by defining a reduction from $\mathsf{VertexCover}$ to $\mathsf{DominatingSet}$. Notice that if C is a vertex cover, and there are no isolated vertices, then C is a dominating set: for $u \in V$ there exists a $(u, v) \in E$ and thus $u \in C$ or $v \in C$ as C is a vertex cover.

Suppose $G = (V, E)$ is a graph (without isolated vertices), we define a new graph $G' = (V', E')$ where

$$V' = V \cup \{uv \mid (u, v) \in E\}, \quad E' = E \cup \{(u, uv), (uv, v) \mid (uv) \in E\}$$

So for every edge in G, we insert a vertex which is also connected to both ends of the edge. We claim that $(G, k) \mapsto (G', k)$ is a Karp reduction from $\mathsf{VertexCover}$ to $\mathsf{DominatingSet}$.

If $G = (V, E)$ has isolated vertices, then we remove the isolated vertices and then construct G'. Since a vertex cover need not contain isolated vertices, we can assume that they don't (we are minimizing the size of the vertex covers and dominating sets).

If $(G, k) \in \mathsf{VertexCover}$ then the vertex cover of size $\leq k$ in G is also a dominating set in G'. Suppose that C is a vertex cover in G, then for every $x \in V'$, if

(1) $x = u \in V$ then since C is a vertex cover, and u is not isolated, there exists a $(u, v) \in E$ and so $u \in C$ or $v \in C$ as required.

(2) $x = uv$ then since $(u, v) \in E$, either u or v is in C and so x has a neighbor in C, as required.

and so C is a dominating set in G'. And therefore $(G', k) \in \mathsf{DominatingSet}$ as required.

And if $(G', k) \in \mathsf{DominatingSet}$ then let S be a dominating set of size $\leq k$ in G', then let us define a new set S', where for every $x \in S$ if

(1) $x = u \in V$, add u to S'.

(2) $x = uv$ then add either u or v to S'.

Then $|S'| \leq |S| \leq k$, and S' is a vertex cover of G: if $(u, v) \in E$ then since $uv \in V'$ and S is a dominating set, either $uv \in S$, or $u \in S$, or $v \in S$. This means that either $u \in S'$ or $v \in S'$, and thus S' is indeed a vertex cover of G. Therefore $(G, k) \in \mathsf{VertexCover}$ as required.

Definition 4.6:

A **Hamiltonian path** in a graph is a path which visits every vertex exactly once.

Example 4.7:

We define the decision problem

$$\mathsf{DHP} = \{G \mid G \text{ is a directed graph which has a Hamiltonian path}\}$$

We claim that this is **NP**-complete.

It is easy to see that this is in **NP**. We will define a reduction from SAT to DHP. Suppose we are given a boolean formula in CNF,

$$\varphi = \bigwedge_{i=1}^{m} \bigvee_{j=1}^{n} \varepsilon_{ij} x_j$$

We define a graph $G = (V, E)$ where we define the following types of vertices:

(1) For each variable x_i we define $3m+3$ copies of it as vertices, which we will denote $x_{i,1}, \dots, x_{i,3m+3}$.

(2) For $i = 1, \dots, m$ we add a vertex b_i which corresponds to the ith disjunction in φ.

(3) We add start and end nodes, s and t.

We also define the following types of edges

(1) For each variable x_i, we define the edges $(x_{i,j}, x_{i,j+1})$ and $(x_{i,j+1}, x_{i,j})$.

(2) For each variable x_i, we define the edges $(x_{i,1}, x_{i+1,1})$, $(x_{i,1}, x_{i+1,3m+3})$, $(x_{i,3m+3}, x_{i+1,1})$, and $(x_{i,3m+3}, x_{i+1,3m+3})$.

(3) For each disjunction D_i and each variable x_j which appears in D_i, then

  (i) if x_j appears in D_i as-is, then we add edges $(x_{j,3i}, b_i)$ and $(b_i, x_{j,3i+1})$.

  (ii) if $\neg x_j$ appears in D_i, then we add edges $(x_{j,3i+1}, b_i)$ and $(b_i, x_{j,3i})$.

(4) We add edges $(s, x_{1,1})$ and $(s, x_{1,3m+3})$, and $(x_{n,1}, t)$ and $(x_{n,3m+3,t})$.

Now, if $\varphi \in \mathsf{SAT}$ then suppose τ satisfies it. Then we define a Hamiltonian path in G as follows:

(1) We start at S.

(2) For every $i = 1, \dots, n$ if τ_i is true then we move on the row $x_{i,1}, \dots, x_{i,3m+3}$ from left to right. Otherwise we move from right to left.

   At each $x_{i,j}$ we check if we can visit some b_k and continue (ie. if we are going from left to right, we must check if we can go from b_k to $x_{i,j+1}$). If we can then we go to that b_k and then $x_{i,j\pm1}$ (if we are going from left to right, then it is $+1$, and right to left is -1).

   If we can't go to some b_k, then we go to the next $x_{i,j\pm1}$ (again, the sign depends on the direction of movement). If $j \pm 1 = 0$ or $3m+4$ (ie we've reached the end of the row), then we go to $x_{i+1,1}$ or $x_{i+1,3m+3}$ depending on whether on the row $x_{i+1,1}, \dots, x_{i+1,3m+3}$ depending on if we are moving left or right on the row for $x_{i+1,j}$.

(3) Once we get to $x_{n,1}$ or $x_{n,3m+3}$, and this is the final vertex in $x_{n,j}$, then we go to t.

This is a well-defined path. We claim it is Hamiltonian. Since we necessarily traverse every vertex of the form $x_{i,j}$ or s or t, we must confirm that we also visit every vertex of the form b_i. For every b_i, some $\varepsilon_{ij} x_j$ must be satisfied by τ, and so if we let j the minimum such value, we will visit b_i on the row of x_j.

Now suppose G has a Hamiltonian path. Suppose that from $x_{i,j}$ we visit b_k, then from b_k we go to $x_{a,b}$. Suppose for the sake of a contradiction that $a \neq i$. Further suppose that on the x_ith row, we are going from left to right. So let j be the minimum such j where this occurs on the x_ith row, so by this point we must have visited all $x_{i,j'}$ for $j' \leq j$. Then at some other point we must go back to the x_i row, and since j is the minimum where this anomaly occurred, we must go to $x_{i,j'}$ for some $j' > j$ and visit $x_{i,j}$ from its right. But then from $x_{i,j}$ we will not have a place to go (since it can only go to $x_{i\pm1,j}$, which have been visited), and thus we cannot have reached t (this must be the final vertex in the Hamiltonian cycle).

So this means that if we go from $x_{i,j}$ to b_k then we return to $x_{i\pm1,j}$, depending on the direction of movement in the x_ith row. So if we go from left to right in x_i, let τ_i be true, and otherwise let it be false. This satisfies φ as each disjunction (b_i) is satisfied.

Thus $\varphi \mapsto G$ is a reduction from SAT to DHP, and so DHP is $\mathbf{NP}$-complete.