CS 351: Solutions to Problem Set 3

(1) Show that the language $\overline{\text{2SAT}} = \{\varphi \mid \varphi$ is a Boolean formula in CNF that has at most two literals per clause, and φ is *not* satisfiable$\}$ is in NL by designing an appropriate proof system.

In 2SAT, each clause has exactly two literals, and by suitably using negations, we may write each clause in the form $\ell_1 \to \ell_2$, where ℓ_1 and ℓ_2 are literals (e.g., $x_3 \vee \neg x_7$ becomes $x_7 \to x_3$ (or $\neg x_3 \to \neg x_7$)).

In our proof system, the proof for the unsatisfiability of a $\overline{\text{2SAT}}$ instance φ is of the form $\langle x, C_1, \ldots, C_s, C'_1, \ldots, C'_t \rangle$, where $s, t \geq 1$, and $C_1, \ldots, C_s, C'_1, \ldots, C'_t$ are clauses of φ such that:

(a) $C_1 = (\ell_{11} \to \ell_{12}), C_2 = (\ell_{21} \to \ell_{22}), \ldots, C_s = (\ell_{s1} \to \ell_{s2})$, and for $1 \leq i < s$, we have $\ell_{i,2} = \ell_{i+1,1}$, and $\ell_{11} = x$ and $\ell_{s2} = \neg x$.

(b) $C'_1 = (\ell'_{11} \to \ell'_{12}), C'_2 = (\ell'_{21} \to \ell'_{22}), \ldots, C'_t = (\ell'_{t1} \to \ell'_{t2})$, and for $1 \leq i < t$, we have $\ell'_{i,2} = \ell'_{i+1,1}$, and $\ell'_{1,1} = \neg x$ and $\ell'_{t2} = x$.

That is, our proof of unsatisfiability consists of a variable x and two sets of clauses; the conjunction of the clauses in the first set logically implies $x \to \neg x$ and the conjunction of the clauses in the second set logically implies $\neg x \to x$.

Note that checking whether a given proof satisfies the above conditions can be easily done in logspace with one-way access to the proof.

For the soundness (which is trivial) and completeness, see Papadimitriou's book, proof of Theorem 9.1. Notice that the algorithm that constructs a truth assignment from the assumption that there is no valid proof of unsatisfiability does *not* have to be a logspace algorithm.

(2a) *VC Dimension.* Given an $m \times n$ table A of 0's and 1's, A is said to *shatter* a set C of d columns if the restriction of A to the columns in C contains every possible string in $\{0,1\}^d$. The *Vapnik–Cervonenkis dimension (VC dimension)* of A is the largest d such that A shatters some set of d columns. The decision problem VCD is defined by $\text{VCD} = \{(A, d) \mid$ the VC dimension of A is $d\}$.

This is a slightly tricky problem because the input size has two parameters m and n. Note that the VC dimension of an $m \times n$ table can never exceed $\log m$ (so exponential in $\log m$ is still polynomial in m). VCD is in $\Sigma_2^p \cap \Pi_2^p$. For $\text{VCD} \in \Sigma_2^p$, note that VC dimension of A is d ($\leq \log m$) iff

$$(\exists C \subseteq [n])\ (\forall C' \subseteq [n])\ \big[|C| = d, A|_C \supseteq \{0,1\}^d\ \wedge$$
$$|C'| > d \Rightarrow (|C'| > \log_2 m,\ \text{or for some } z \in \{0,1\}^{|C'|}, z \notin A|_{C'})\big].$$

The fact that the condition in the "[...]" is polynomial-time verifiable uses the fact that d could be at most $\log m$ so 2^d is polynomial in m. Note also that since the conditions on C and C' can be verified independent of each other, we have

$$(A,d) \in \textsc{vcd} \text{ iff } (\forall C' \subseteq [n])\ (\exists C \subseteq [n])\ \big[|C| = d, A|_C \supseteq \{0,1\}^d\ \wedge$$
$$|C'| > d \Rightarrow (|C'| > \log_2 m, \text{ or for some } z \in \{0,1\}^{|C'|}, z \notin A|_{C'})\big],$$

so $\textsc{vcd} \in \Pi_2^p$.

Incidentally, as formulated, $\textsc{vcd} \in \text{DTIME}[mn^{O(\log m)}] = \text{DTIME}[N^{O(\log N)}]$, where $N = nm$, the size of the input.

(2b) *Clique Number.* Given a graph G, define $\alpha(G)$, the *clique number* of G to be the size of the largest clique in G, that is, the number of vertices in the largest complete (vertex-induced) subgraph of G. The decision problem CLIQUE-NUMBER is defined by CLIQUE-NUMBER $= \{(G,k) \mid \alpha(G) = k\}$.

CLIQUE-NUMBER $\in \Sigma_2^p \cap \Pi_2^p$.

CLIQUE-NUMBER $\in \Sigma_2^p$ since

$$\alpha(G) = k \text{ iff } (\exists S \subseteq V(G))\ (\forall T \subseteq V(G))\ [|S| = k \wedge \text{for every } u, v \in S,$$
$$(u,v) \in E(G) \wedge |T| > |S| \Rightarrow \text{there exist } u, v \in T,\ (u,v) \notin E(G)].$$

CLIQUE-NUMBER $\in \Pi_2^p$ since

$$\alpha(G) = k \text{ iff } (\forall T \subseteq V(G))\ (\exists S \subseteq V(G))\ [|S| = k \wedge \text{for every } u, v \in S,$$
$$(u,v) \in E(G) \wedge |T| > |S| \Rightarrow \text{there exist } u, v \in T,\ (u,v) \notin E(G)].$$

(2c) Problem 17.3.11 in Papadimitriou's book.

Notation: $\tau = \langle u_1, \ldots, u_k \rangle$ denotes a sequence of k distinct vertices, and $cost(\tau)$ denotes the cost of a tour of the k cities in the order given by τ, and for a subset A of $\tau = \{u_1, \ldots, u_k\}$, $\tau|_A$ denotes the restriction of the tour τ to the vertices in A. Note that given a tour τ, $cost(\tau)$ can be computed in polynomial time.

An instance T (an $n \times n$ matrix of distances) of TSP has the master tour property iff $(\exists \sigma = \langle v_1, \ldots, v_n \rangle)(\forall S \subseteq [n])(\forall \tau = \langle u_1, \ldots, u_{|S|} \rangle)[cost(\tau) \geq cost(\sigma|_S)]$.

(3) Reading assignment.