COSC 341
Theory of Computing
Lecture 19
The Cook-Levin Theorem

Stephen Cranefield
stephen.cranefield@otago.ac.nz

Lecture slides (mostly) by Michael Albert

Keywords: SAT, $\mathbf{NP}$-complete

The theorem and the plan

Theorem (Cook-Levin theorem)

SAT *is NP-complete.*

The plan:

- Given a non-deterministic-by-oracle Turing machine, M, an input word, w of size n, and a time-bound of An^c for M:
  - Film the operation of M on input w for An^c steps
  - Convert that film into an instance of SAT by encoding its configuration in boolean variables as frames and then linking them all together consistently.
  - Do this in such a way that the instance has a satisfying assignment if (and only if) the computation that M would run on w accepts for some choice of oracle tape.

What's a configuration?

- The position of the read-write heads (main tape, oracle tape)
- The current state
- The complete current contents of the tapes

We can use Boolean variables whose intended interpretation (if true) will be something like "the symbol at position 6 of the main tape at time 13 is an a".

And then, the game is to figure out a polynomially-sized set of clauses that ensure that every satisfying assignment represents a legitimate accepting computation.

Variables used

We refer to variables using functions that take parameters and encapsulate some variable-naming scheme.

Name	**Interpretation**	**Number of variables**
STATE(q, t)	M is in state q at time t.	$\lvert Q\rvert \times (An^c + 1)$
M-HEAD(j, t)	The read/write head on the main tape is at position j at time t.	$(An^c + 1)^2$
O-HEAD(j, t)	The read/write head on the oracle tape is at position j at time t.	$(An^c + 1)^2$
M-SYMBOL(s, j, t)	Symbol s is on the main tape at position j at time t.	$\lvert\Gamma\rvert \times (An^c + 1)^2$
O-SYMBOL(s, j, t)	Symbol s is on the oracle tape at position j at time t.	$\lvert\Gamma\rvert \times (An^c + 1)^2$

Example clause

Suppose we want to encode the constraint that:

Position 3 *of the main tape contains exactly one symbol at time* 23

We can easily state that there is *some* symbol in that position at time 23:

$$\bigvee_{s \in \Sigma} \text{M-SYMBOL}(s, 3, 23).$$

But how can we say there is *only one* symbol there?

A key idea used throughout the proof

- Suppose we have a set of variables $\{x_1, x_2, \ldots, x_k\}$ and we want to ensure that exactly one of them is true in a satisfying assignment.
- We can use the following clauses:

$$x_1 \vee x_2 \vee x_3 \vee \cdots \vee x_k$$
$$\neg x_1 \vee \neg x_2$$
$$\neg x_1 \vee \neg x_3$$
$$\cdots$$
$$\neg x_1 \vee \neg x_k$$
$$\neg x_2 \vee \neg x_3$$
$$\cdots$$
$$\neg x_{k-1} \vee \neg x_k$$

- There are $1 + \frac{k(k-1)}{2}$ clauses, which is polynomial in size of the number of variables (k). Also, each clause is of size at most k, so we have polynomially many clauses of polynomial size.
- If we only say that polynomially many times during the proof, then the total size of our problem representation will be polynomial.

Returning to our example clause

We can use the trick in the previous slide to add the constraint that there is *only one* symbol at position 3 of the main tape at time 23. We include in our CNF formula to be satisfied the following clauses:

$$\neg\text{M-SYMBOL}(s, 3, 23) \vee \neg\text{M-SYMBOL}(s', 3, 23)$$

for all pairs $s, s' \in \Sigma$ with $s \neq s'$.

Another example clause

If position 7 *is not under the oracle read/write head at time* 12 *then the symbol at that position is the same at time* 13.

We need to encode this separately for each symbol $s \in \Gamma$:

$$\neg\text{O-HEAD}(7, 12) \wedge \text{O-SYMBOL}(s, 7, 12) \implies \text{O-SYMBOL}(s, 7, 13).$$

Here we use the logical implication symbol ' $\implies$ ' to give us a more readable way of referring to the clause:

$$\text{O-HEAD}(7, 12) \vee \neg\text{O-SYMBOL}(s, 7, 12) \vee \text{O-SYMBOL}(s, 7, 13).$$

This is equivalent to the first version because:

$$x \implies y \text{ is equivalent to } \neg x \vee y$$

and $\neg(\neg\text{O-HEAD}(7, 12) \wedge \text{O-SYMBOL}(s, 7, 12))$ is equivalent to $\text{O-HEAD}(7, 12) \vee \neg\text{O-SYMBOL}(s, 7, 12)$ using one of De Morgan's laws.

The details

Better presented through the notes (Notes 17).

We will see how to express clauses in various ‘groups’ (or themes):

- Constraints that ensure the representation of the tapes at each time step are valid.
- An encoding of the initial configuration of the TM.
- The constraint that the TM is in the accepting state at time An^c.
- Constraints ensuring that the representation of each state and its following state are related by a correct TM operational step.