

## Cook's theorem

- Recall the language SAT:

$$\text{SAT} = \{\langle \varphi \rangle \mid \varphi \text{ is a satisfiable boolean formula}\}$$

- Theorem:** SAT is **NP**-complete
- Last class we proved that  $\text{SAT} \in \text{NP}$
- It remains to prove that SAT is **NP**-hard
- Let  $L$  be an arbitrary language in **NP**. We must show that  $L \leq_p \text{SAT}$ .
- Let  $M$  be an NTM and  $q$  a polynomial such that  $L(M) = L$  and  $T_M(n) \leq q(n)$ ,  $\forall n \in \mathbb{N}$ . Assume that  $M$  has input alphabet  $\{0, 1\}$ , tape alphabet  $\{0, 1, \sqcup\}$ , and add new transitions  $\delta(q_{\text{accept}}, a) = (q_{\text{accept}}, a, R)$  and  $\delta(q_{\text{reject}}, a) = (q_{\text{reject}}, a, R)$  for each  $a \in \{0, 1, \sqcup\}$ . These extra transitions allow us to assume that all computations of  $M$  on an input of size  $n$  have length exactly  $q(n)$ . Assume the states of  $M$  are  $\{q_0, \dots, q_m\}$ .
- For an input  $w$ ,  $\varphi_w$  will be a formula such that:

$$M \text{ has an accepting computation on input } w \iff \varphi \text{ is satisfiable}$$

- Let  $n = |w|$ , and let  $N = q(n)$ . Since  $M$  takes at most  $N$  steps on input  $w$ , it can only reach tape squares at distance  $N$  or less from its initial position. The idea is to have variables representing the contents of these  $2N + 1$  cells, the state of  $M$ , and the position of  $M$ 's tape head, at each of the  $N$  steps that  $M$  may take. The formula  $\varphi_w$  will ensure that the settings of these variables represent a valid sequence of configurations of  $M$  on input  $w$ , and that the final configuration in the sequence is an accepting configuration.
- The variables of  $\varphi_w$  are as follows:

$$\begin{array}{ll} B_{i,j} & - N \leq i \leq N, 0 \leq j \leq N \\ T_{i,j} & - N \leq i \leq N, 0 \leq j \leq N \\ H_{i,j} & - N \leq i \leq N, 0 \leq j \leq N \\ Q_{i,j} & 0 \leq i \leq m, 0 \leq j \leq N \end{array}$$

The idea is as follows:  $B_{i,j}$  is true iff the  $i$ -th cell is blank at the end of step  $j$  (the end of step 0 is considered to be the initial configuration). If  $B_{i,j}$  is false, then  $T_{i,j}$  is true iff the  $i$ -th cell contains a 1 at the end of step  $j$ .  $H_{i,j}$  is true iff the tape-head is scanning cell  $i$  at the end of step  $j$ . Finally,  $Q_{i,j}$  is true iff  $M$  is in state  $q_i$  at the end of step  $j$ .

- It is clear that these variables can be used to represent a sequence of  $N+1$  configurations of  $M$ . The formula  $\varphi_w$  will be constructed so that every satisfying assignment must set the variables so that they correspond to a valid computation of  $M$  on input  $w$ .
- The first thing is to ensure that for every  $j$ , at most one of the variables  $Q_{i,j}$  is true, since  $M$  can only be in one state. Define the subformula ONE-STATE $_j$ :

$$\text{ONE-STATE}_j = \bigwedge_{i \neq i'} \neg(Q_{i,j} \wedge Q_{i',j})$$

Define ONE-STATE =  $\bigwedge_{j=0}^N \text{ONE-STATE}_j$ .

- Similarly, we need to ensure that for every  $j$ , at most one of the variables  $H_{i,j}$  is true. Define ONE-POS $_j$ :

$$\text{ONE-POS}_j = \bigwedge_{i \neq i'} \neg(H_{i,j} \wedge H_{i',j})$$

and let ONE-POS =  $\bigwedge_{j=0}^N \text{ONE-POS}_j$ .

- Next, we have a subformula INIT that ensure that the variables with  $j = 0$  represent  $I_M(w)$  (the initial configuration of  $M$  on input  $w$ ).

$$\text{INIT} = Q_{0,0} \wedge H_{0,0} \wedge \left( \bigwedge_{i=-N}^{-1} B_{i,0} \right) \wedge \left( \bigwedge_{i=n}^N B_{i,0} \right) \wedge \left( \bigwedge_{i=0}^{n-1} (\neg B_{i,0} \wedge W_i) \right)$$

where  $W_i = T_{i,0}$  if the  $i$ -th bit of  $w$  is 1, and  $W_i = \neg T_{i,0}$  otherwise.

- Finally, we need to ensure that the configuration after  $j+1$  steps follows from the configuration after  $j$  steps via a transition of  $M$ . For each  $a \in \{0, \dots, m\}$ ,  $d \in \{0, 1, \sqcup\}$ ,  $t \in \{-N, \dots, N\}$ , and  $j \in \{0, \dots, N\}$  we have a subformula TRANS $_{a,d,t,j}$  which says “if, after  $j$  steps,  $M$  is scanning cell  $t$  and is in state  $q_a$ , and cell  $t$  contains the symbol  $d$  at the end of  $j$  steps, then the state, tape position, and the contents of cell  $t$  after  $j+1$  steps follow via a transition of  $M$  and the other cells are unchanged.”
- The subformula TRANS $_{a,d,t,j}$  is best illustrated by example. Suppose that  $M$  has the following transition:

$$\delta(q_a, 0) = \left\{ (q_b, 1, R), (q_c, 0, L) \right\}$$

Then  $\text{TRANS}_{a,0,t,j}$  is as follows.

$$(Q_{a,j} \wedge H_{t,j} \wedge \neg B_{t,j} \wedge \neg T_{t,j}) \rightarrow \\ \left( Q_{b,j+1} \wedge H_{t+1,j+1} \wedge \neg B_{t,j+1} \wedge T_{t,j+1} \wedge \left( \bigwedge_{i \neq t} (B_{i,j+1} \leftrightarrow B_{i,j} \wedge T_{i,j+1} \leftrightarrow T_{i,j}) \right) \right) \\ \vee \left( Q_{c,j+1} \wedge H_{t-1,j+1} \wedge \neg B_{t,j+1} \wedge \neg T_{t,j+1} \wedge \left( \bigwedge_{i \neq t} (B_{i,j+1} \leftrightarrow B_{i,j} \wedge T_{i,j+1} \leftrightarrow T_{i,j}) \right) \right)$$

- Then TRANS is defined as:

$$\text{TRANS} = \bigwedge_{a,d,t,j} \text{TRANS}_{a,d,t,j}$$

- Now we are ready to define  $\varphi_w$ . It is:

$$\varphi_w = \text{ONE-STATE} \wedge \text{ONE-POS} \wedge \text{INIT} \wedge \text{TRANS} \wedge Q_{\text{accept},N}$$

- If  $w \in L = L(M)$  then  $M$  has an accepting computation on input  $w$ , and setting the variables of  $\varphi_w$  to correspond to this computation results in a satisfying truth assignment for  $\varphi_w$ . On the other hand, if  $\varphi_w$  is satisfiable then there is a truth assignment  $\tau$  that satisfies  $\varphi_w$ , and by construction the variables must correspond to an accepting computation of  $M$  on input  $w$ , implying  $w \in L(M) = L$ . So  $\varphi_w$  is satisfiable iff  $w \in L$ .
- We also need to argue that the function  $f(w) = \varphi_w$  is computable in time polynomial in  $|w|$ . Let's first consider the size of  $\varphi_w$ . The number of states of  $M$  is constant, so the length of ONE-STATE and ONE-POS is  $O(N)$ . Similarly the length of INIT is  $O(N)$ . The length of  $\text{TRANS}_{a,d,t,j}$  is  $O(N)$ , and so the length of TRANS is  $O(N^3)$ . Thus the length of  $\varphi_w$  is  $O(q(n)^3)$ , where  $n = |w|$ . Since  $q$  is a polynomial, so is  $q(n)^3$ . It is not too hard to convince yourself that  $\varphi_w$  can be produced in time polynomial in its length.