# SUBSET-SUM

$$\text{SUBSET-SUM} = \{\langle S,t\rangle \mid S = \{x_1, x_2, \dots x_k\},$$
$$\exists T \subseteq \{1,2,\dots k\}, \text{ s.t } \sum_{i\in T} x_i = t\}$$

Given a set $S$ and a target sum $t$, is there a subset of $S$, whose elements add up to $t$?

Let $S = \{6, 20, 32, 16, 5\}$.

Then $\langle S, 54\rangle \in$ SUBSET-SUM

$\langle S, 42\rangle \in$ SUBSET-SUM

$\langle S, 34\rangle \notin$ SUBSET-SUM

$\langle S, 1\rangle \notin$ SUBSET-SUM.

**Theorem 7.56**: SUBSET-SUM is NP-complete.

**Proof**: We need to show two things

1) SUBSET-SUM $\in$ NP and (2) 3-SAT $\leq_p$ SUB-SUM.

SUBSET-SUM $\in$ NP. Already shown in lecture 46.

On input $\langle S, t\rangle$ :

1. Non deterministically select | reject each of $x_1, x_2, \dots x_k$. $\longrightarrow O(k)$ time
2. Add all the selected $x_i$ and verify if they add up to $t$. $\longrightarrow O(k)$
3. If sum $= t$, then accept. Else, reject.

Now we focus on 3-SAT $\leq_p$ SUBSET-SUM.

Given a 3-CNF formula $\phi$, we need to construct $S, t$ such that

$$\langle \phi \rangle \in \text{3-SAT} \iff \langle S, t\rangle \in \text{SUBSET-SUM}.$$

Let $\phi$ be a formula with $\underbrace{n \text{ variables}}_{x_1, x_2, \dots x_n}$ and $m$ clauses. We build the SUBSET-SUM instance as follows:

We construct $S$ with $2(n+m)$ numbers.

Each variable $x_i$ in $\phi$ corresponds to two numbers in $S$ — $y_i$ and $z_i$.

Each clause $C_j$ corresponds to two numbers — $g_j$ and $h_j$.

$$
\begin{array}{c|cccccc|ccccc|}
 & 1 & 2 & 3 & 4 & \cdots & n & \overset{C_1}{1} & \overset{C_2}{2} & \overset{C_3}{3} & \cdots & \overset{C_m}{m} \\
\hline
y_1 & 1 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 \\
z_1 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
y_2 & 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 1 \\
z_2 & 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 \\
y_3 & & & 1 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 \\
z_3 & & & 1 & 0 & \cdots & 0 & & & & & \\
\vdots & & & & & \ddots & \vdots & & \vdots & \vdots & & \vdots \\
y_n & & & & & & 1 & & & & & \\
z_n & & & & & & 1 & & & & & \\
\hline
g_1 & & & & & & & 1 & 0 & 0 & \cdots & 0 \\
h_1 & & & & & & & 1 & 0 & 0 & \cdots & 0 \\
g_2 & & & & & & & 0 & 1 & 0 & \cdots & 0 \\
h_2 & & & & & & & 0 & 1 & 0 & \cdots & 0 \\
g_3 & & & & 0 & & & 0 & 0 & 1 & \cdots & 0 \\
h_3 & & & & & & & 0 & 0 & 1 & \cdots & 0 \\
\vdots & & & & & & & & \ddots & & & \vdots \\
g_m & & & & & & & 0 & 0 & 0 & \cdots & 1 \\
h_m & & & & & & & 0 & 0 & 0 & \cdots & 1 \\
\hline
\end{array}
$$

All nos in decimal

$$t = \underbrace{1\ 1\ 1\ \cdots\cdots\ 1}_{n}\ \underbrace{3\ 3\ 3\ \cdots\ 3}_{m}$$

In the top right quadrant, we choose the entries in the following manner.

- → If $x_i$ is in $C_j$, put 1 against $y_i$ and $C_j$
- → If $\overline{x_i}$ is in $C_j$, put 1 against $z_i$ and $C_j$
- → 0's in all the other cells of the top right quadrant.

Important: All the numbers are to be read as decimal numbers.

$$S = \{ y_i, z_i, g_j, h_j \mid 1 \le i \le n, 1 \le j \le m \}.$$

The table has $2(n+m)^2$ entries. So the construction of $S$ is in polynomial time.

Now we need to show that:

$\phi$ is satisfiable $\Longleftrightarrow$ $S$ has a subset that sums to $t$.

($\Longrightarrow$) Suppose $\phi \in$ 3-SAT. This means that there is a way to assign True / False to the variables $x_i$ such that each clause is satisfied.

We pick a subset of $S$ as follows.

$x_i$ is TRUE in sat. assignment
$\Rightarrow$ Choose $y_i$ in $S$, and not $z_i$

$x_i$ is FALSE $\Rightarrow$ Choose $z_i$ in $S$, and not $y_i$.

Since the assignment is a satisfying assignment, each column marked $C_j$ has at least one 1 under it from the first $2n$ rows. And at most three 1's.

We include $g_j$ alone, or both $g_j$ and $h_j$ to make up the sum under $C_j$ to 3.

$\langle S, t \rangle \in$ SUBSET-SUM.

($\Leftarrow$) Suppose $\langle S, t \rangle \in$ SUBSET-SUM.

Let us note the following:

→ All the digits of the numbers in S are 0/1.

→ Each column has at most five 1's.
So no carry is possible while adding the numbers.

→ We have a sum 1 in the first n columns.
This means for each $i \leq n$, exactly one of $y_i$ or $z_i$ is in the subset.

We claim that the following is a satisfying assignment:

→ If $y_i$ is in the subset, set $x_i$ to TRUE.

→ If $z_i$ is in the subset, set $x_i$ to FALSE.

In the last m columns, we need the sum to be 3 each. The contribution from the last 2m rows can be at most 2. So for each $C_j$, there must be a contribution of at least 1 from $y_i$ or $z_i$.

→ If contribution from $y_i$, then $x_i$ appears in that clause, and is set to TRUE.

→ If contribution from $z_i$, then $\overline{x_i}$ appears in that clause, and $x_i$ is set to FALSE.

In either case, the clause is satisfied.

Hence $\langle \phi \rangle \in$ 3-SAT.