The NP-completeness of Subset Sum

Pilu Crescenzi and Viggo Kann

University of Florence and KTH

October 2011
Basic definitions

- **Class NP**
  - Set of decision problems that admit “short” and efficiently verifiable solutions
  - Formally, \( L \in \text{NP} \) if and only if there exist
    - polynomial \( p \)
    - polynomial-time machine \( V \)
    - such that, for any \( x \),
      \[
      x \in L \iff \exists y (|y| \leq p(|x|) \land V(x, y) = 1)
      \]

- **Polynomial-time reducibility**
  - \( L_1 \leq L_2 \) if there exists polynomial-time computable function \( f \) such that, for any \( x \),
    \[
    x \in L_1 \iff f(x) \in L_2
    \]

- **NP-complete problem**
  - \( L \in \text{NP} \) is NP-complete if any language in NP is polynomial-time reducible to \( L \)
    - Hardest problem in NP
Basic results

- Cook-Levin theorem
  - Sat problem
    - Given a boolean formula in conjunctive normal form (disjunction of conjunctions), is the formula satisfiable?
  - Sat is NP-complete

- 3-Sat
  - Each clause contains exactly three literals
  - 3-Sat is NP-complete

- Simple proof by local substitution
  - \( l_1 \Rightarrow (l_1 \lor y \lor z) \land (l_1 \lor y \lor \overline{z}) \land (l_1 \lor \overline{y} \lor z) \land (l_1 \lor \overline{y} \lor \overline{z}) \)
  - \( l_1 \lor l_2 \Rightarrow (l_1 \lor l_2 \lor y) \land (l_1 \lor l_2 \lor \overline{y}) \)
  - \( l_1 \lor l_2 \lor l_3 \Rightarrow l_1 \lor l_2 \lor l_3 \)
  - \( l_1 \lor l_2 \lor \cdots \lor l_k \Rightarrow \)

  \[(l_1 \lor l_2 \lor y_1) \land (\overline{y_1} \lor l_3 \lor y_2) \land (\overline{y_2} \lor l_4 \lor y_3) \land \cdots \land (\overline{y_{k-3}} \lor l_{k-1} \lor l_k)\]
Problem definition: Subset Sum

Given a (multi)set $A$ of integer numbers and an integer number $s$, does there exist a subset of $A$ such that the sum of its elements is equal to $s$?

- No polynomial-time algorithm is known
- Is in NP (short and verifiable certificates):
  - If a set is “good”, there exists subset $B \subseteq A$ such that the sum of the elements in $B$ is equal to $s$
  - Length of $B$ encoding is polynomial in length of $A$ encoding
  - There exists a polynomial-time algorithm that verifies whether $B$ is a set of numbers whose sum is $s$:
    - Verify that $\sum_{a \in B} a = s$
**Reduction of 3-Sat to Subset Sum:**

- **n variables** $x_i$ and **m clauses** $c_j$
- For each variable $x_i$, construct numbers $t_i$ and $f_i$ of $n + m$ digits:
  - The $i$-th digit of $t_i$ and $f_i$ is equal to 1
  - For $n + 1 \leq j \leq n + m$, the $j$-th digit of $t_i$ is equal to 1 if $x_i$ is in clause $c_{j-n}$
  - For $n + 1 \leq j \leq n + m$, the $j$-th digit of $f_i$ is equal to 1 if $\overline{x_i}$ is in clause $c_{j-n}$
  - All other digits of $t_i$ and $f_i$ are 0

**Example:**

$$(x_1 \lor x_2 \lor x_3) \land (\overline{x_1} \lor \overline{x_2} \lor x_3) \land (\overline{x_1} \lor x_2 \lor \overline{x_3}) \land (x_1 \lor \overline{x_2} \lor x_3)$$

<table>
<thead>
<tr>
<th>Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$f_1$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$t_2$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$f_2$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$t_3$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$f_3$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
For each clause \( c_j \), construct numbers \( x_j \) and \( y_j \) of \( n + m \) digits:
- The \((n + j)\)-th digit of \( x_j \) and \( y_j \) is equal to 1
- All other digits of \( x_i \) and \( y_j \) are 0

**Example:**

\[
(x_1 \lor x_2 \lor x_3) \land (\overline{x_1} \lor \overline{x_2} \lor x_3) \land (\overline{x_1} \lor x_2 \lor \overline{x_3}) \land (x_1 \lor \overline{x_2} \lor x_3)
\]

<table>
<thead>
<tr>
<th>Number</th>
<th>( i )</th>
<th>( j )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>( x_1 )</td>
<td>0 0 0</td>
<td>1 0 0 0</td>
</tr>
<tr>
<td>( y_1 )</td>
<td>0 0 0</td>
<td>1 0 0 0</td>
</tr>
<tr>
<td>( x_2 )</td>
<td>0 0 0</td>
<td>0 1 0 0</td>
</tr>
<tr>
<td>( y_2 )</td>
<td>0 0 0</td>
<td>0 1 0 0</td>
</tr>
<tr>
<td>( x_3 )</td>
<td>0 0 0</td>
<td>0 0 1 0</td>
</tr>
<tr>
<td>( y_3 )</td>
<td>0 0 0</td>
<td>0 0 1 0</td>
</tr>
<tr>
<td>( x_4 )</td>
<td>0 0 0</td>
<td>0 0 0 1</td>
</tr>
<tr>
<td>( y_4 )</td>
<td>0 0 0</td>
<td>0 0 0 1</td>
</tr>
</tbody>
</table>

Finally, construct a sum number \( s \) of \( n + m \) digits:
- For \( 1 \leq j \leq n \), the \( j \)-th digit of \( s \) is equal to 1
- For \( n + 1 \leq j \leq n + m \), the \( j \)-th digit of \( s \) is equal to 3
Proof of correctness

- Show that Formula satisfiable $\implies$ Subset exists:
  - Take $t_i$ if $x_i$ is true
  - Take $f_i$ if $x_i$ is false
  - Take $x_j$ if number of true literals in $c_j$ is at most 2
  - Take $y_j$ if number of true literals in $c_j$ is 1

Example

- $(x_1 \lor x_2 \lor x_3) \land (\overline{x_1} \lor x_2 \lor x_3) \land (\overline{x_1} \lor x_2 \lor \overline{x_3}) \land (x_1 \lor \overline{x_2} \lor x_3)$
- All variables true

<table>
<thead>
<tr>
<th></th>
<th>$i$</th>
<th></th>
<th>$j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>$t_1$</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$t_2$</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$t_3$</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$x_2$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$y_2$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$x_3$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$y_3$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$x_4$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$s$</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Show that Subset exists ⇒ Formula satisfiable:

- Assign value true to $x_i$ if $t_i$ is in subset
- Assign value false to $x_i$ if $f_i$ is in subset
- Exactly one number per variable must be in the subset
  - Otherwise one of first $n$ digits of the sum is greater than 1
- Assignment is consistent
- At least one variable number corresponding to a literal in a clause must be in the subset
  - Otherwise one of next $m$ digits of the sum is smaller than 3
- Each clause is satisfied