Max SAT — Given a set of \( m \) clauses (CNF)

in \( n \) Boolean variables \( x_1, \ldots, x_n \), find True/False assignment
to variables to make a max. number of clauses true.

e.g. \((\overline{a_1} \lor a_2) (a_1 \lor \overline{a_3}) (a_2 \lor a_3) (\overline{a_3})\)

\( a_3 = F \quad a_2 = T \) but can't get all clauses

Max is 3.

[Decision version] NP-complete — usual SAT.

NP-complete even for clauses of size \( \leq 2 \) (though deciding

if all clauses can be satisfied is poly. time)

called Max 2-SAT.

to prove NP-complete: reduce 3-SAT to Max 2-SAT

turn \((x \lor y \lor z)\) into \( \sim 10 \) 2-SAT clauses

s.t. \((x \lor y \lor z)\) is satisfied iff \( \sim 7 \) of the 2-SAT clauses

are satisfied.

Randomized Algorithm for Max-SAT.

Pick Truth Value Assignment at random (\( \frac{1}{2} \))

Analysis For any clause \( C \)

\[ \Pr[\text{C is satisfied}] = 1 - \Pr[\text{C not satisfied}] \]

\[ = 1 - \frac{1}{2^t} \quad t = \# \text{variables in } C \]

\[ \geq (1 - \frac{1}{2}) = \frac{1}{2} \]

Expected \# clauses satisfied = \[ \sum_{\text{clause } C} E(\sigma_C) \],

\[ \sigma_C = \begin{cases} 1 & \text{if } C \text{ is satisfied} \\ 0 & \text{else} \end{cases} \]

\[ = \sum_C E(\sigma_C) \geq \frac{m}{2} \]

using linearity of expectation.

Note: cannot use analogy

with coin tosses because not independent.

Expected approx. factor is \( \frac{1}{2} \)

This alg. is better for large clauses. Eg. 2-SAT: \( \frac{3}{4} \). 3-SAT: \( \frac{7}{8} \)
one consequence:

**Theorem.** There always exists a truth value assignment that satisfies at least half the clauses.

**Proof.** By the "probabilistic method." *

If the expected value of a random variable is $\alpha$,

then there exists at least one value for the variable $\geq \alpha$.

* Many powerful consequences.

We can de-randomize the above alg. — get a deterministic algorithm to satisfy at least half the clauses.

**Example**

<table>
<thead>
<tr>
<th>clauses</th>
<th>$\bar{x}_1 \bar{x}_2 x_1 \lor x_2 \lor \bar{x}_3 \bar{x}_1 x_2 \lor x_3 \lor x_2 \lor x_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$ clauses satisfied</td>
<td>$\frac{1}{2} + \frac{3}{4} + \frac{7}{8} + \frac{7}{8} + \frac{7}{8} = 3\frac{7}{8}$</td>
</tr>
<tr>
<td>$\alpha$ Esat</td>
<td>$(1-\frac{1}{2^t}) t = # variables$</td>
</tr>
</tbody>
</table>

If we set $x_1 = False$ we get expectation

$\frac{1}{2} + \frac{1}{2} + \frac{3}{4} + 1 + \frac{3}{4} = 4$

$\frac{1}{2}$ is $\bar{x}_1$ in clause, else $(1-\frac{1}{2^t}) t = \# vars$, not counting $x_1$.

If we set $x_1 = True$ we get expectation

$0 + 1 + 1 + \frac{3}{4} + 1 = 3\frac{3}{4}$

**Note:** $Esat = \frac{1}{2} Esat (x_1 = F) + \frac{1}{2} Esat (x_1 = T)$

so one choice gives expectation at least $Esat$

$3\frac{7}{8} = \frac{1}{2} \cdot 4 + \frac{1}{2} \cdot 3\frac{3}{4}$

We pick $x_1 = False$ because it gives higher expectation.
Continue to $x_2$

Setting $x_2$ to False gives expectation

$$1 + 1 + \frac{1}{2} + 1 + \frac{1}{2} = 4$$

Setting $x_2$ to True gives

$$1 + 0 + 1 + 1 + 1 = 4$$

Here all clauses are settled and we have 4 satisfied clauses, this example can be generalized.

Improved Approx. Alg. for Max SAT.

Formulate as Integer Linear Program (ILP)

Solve LP relaxation. Use "randomized rounding".

Make variables

$$x_i \quad i=1 \cdots n \quad \text{for each Boolean variable } a_i$$

$$y_i \quad i=1 \cdots m \quad \text{for each clause}$$

Max $\sum y_i$

One constraint per clause.

E.g. $c_1 = (\overline{a_1} \lor a_2)$

$$y_1 \leq (1-x_1) + x_2 \quad \text{i.e. in order to set } y_1 = 1 \text{ we need } x_2 = 1 \quad \text{or } x_1 = 0 \quad \text{(or both)}$$

$$0 \leq x_i \leq 1 \quad i=1 \cdots n$$

$$0 \leq y_i \leq 1 \quad i=1 \cdots m \quad \text{this is LP relaxation}$$

If $x_i, y_i \in \{0,1\}$ then this ILP is exactly Max SAT.

Use poly. time LP alg. to solve.

Set $a_i = \frac{1}{2}$ with Prob $x_i$: "randomized rounding"

This gives a truth value assignment.
Analysis for Max 2-SAT.

\[ \text{e.g. } C_2 = (a_1 \lor a_2) \]

in LP \[ y_2 \leq x_1 + x_2 \]

\[ \text{Prob}\{C_2\} \geq \frac{3}{4} \text{Prob}\{a_1 \lor a_2\} \]

\[ \geq x_1 + x_2 - \left(\frac{x_1 + x_2}{2}\right)^2 \quad \text{since geometric mean \leq arithmetic mean} \]

\[ \geq y_2 - \frac{y_2^2}{4} \]

\[ = \frac{3}{4} y_2 \]

Since \[ y_2 \leq x_1 + x_2 \leq 2 \]

and the \( y \)-\( x \)-axis is increasing for \( x \in [0, 2] \)

\[ \text{Prob}\{C_2\} \text{ satisfied} \geq \frac{3}{4} y_2 \]

Expected # clauses satisfied

\[ \geq \frac{3}{4} \sum y_i = \frac{3}{4} \text{OPT}_{\text{LP}} \geq \frac{3}{4} \text{OPT}_{\text{ILP}} = \frac{3}{4} \text{OPT}_{\text{Max SAT}} \]

Analysis of general case (which we won't do)

shows this method is better for small clauses.

Previous alg. better for large clauses.

Can combine & de-randomize to get approx factor \( \frac{9}{14} \)

for general Max SAT. We won't do details.

Best known: Goemans & Williamson '94 \( \approx 0.878 \) approx factor.

Lower bound: no approx factor \( \geq 0.942 \) unless \( P = NP \).
We've seen:
\[
\begin{align*}
&\text{metric TSP} & \quad & 1.5 \text{ approx} \quad (\text{we saw 2-approx}) \\
&\text{vertex cover} & \quad & 2 \\
&\text{set cover} & \quad & O(\log n)
\end{align*}
\]
\[
\begin{align*}
&\text{max cut} & \quad & \frac{1}{2} \\
&\text{max SAT} & \quad & \frac{3}{4}
\end{align*}
\]

Questions
- Which problems have constant factor approx alg
- How close to 1 can constant be?

Note: These questions are only relevant assuming P ≠ NP.

Next day: Problems where approx factor can get arbitrarily close to 1

PTAS poly-time approximation scheme (of course run-time increases)

Hardness of Approximation

results of the form:

If we could approx. problem $Q$ in poly-time with constant factor / PTAS, then P = NP.

Some results like this are straightforward.

Example. Recall from 1st lecture: 2-approx for Travelling Salesman Problem for points in plane.

Lemma. If we could 2-approx. general TSP in poly-time then P = NP.

Proof. Prove we could solve Ham. cycle in poly time

Let $G = (V, E)$ be instance of Ham. cycle
Construct $G' = (V, E')$ a complete graph

$w(e) = 1$ for $e \in E$ original edges

$w(e) = n+2$ for $e \in E' - E$

By assumption, we can find a TSP tour $T$ in $G'$

$w(T) \leq 2 \cdot \text{OPT}_{TSP}$

Observe: $w(T) = n$ (if $T \subseteq E$) or $w(T) \geq (n-1)+(n+2) > 2n$

Claim. $G$ has Ham. cycle iff $w(T) = n$

i.e. $T \subseteq E$

If $\Rightarrow$ $G$ has Ham. cycle. Then $\text{OPT}_{TSP} = n$.

So $w(T) \leq 2n$. Then $w(T) = n$

$\Leftarrow$ obvious.