Will prove the following are NP-complete

- Exact Cover
- Solving most linear equations
- Sparse solutions to linear equations

Then discuss

- Low-rank matrix completion
- Non-negative matrix factorization
- Subset Sum
- 3-colorability

Exact Cover

Algebraic statement:

input is a m-by-n matrix A with $\{0,1\}$ entries.

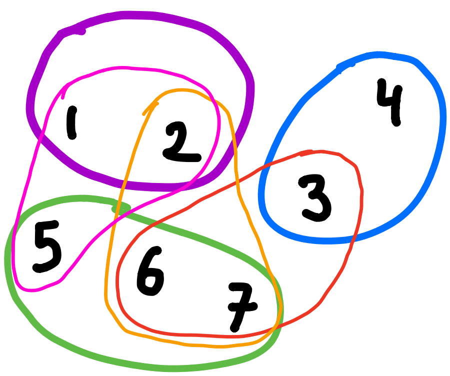
answer is "yes" if $\exists\, x \in \{0,1\}^n$ s.t. $Ax = \mathbb{1}$

Combinatorial statement:

Given sets $A_1, \dots, A_n$, each a subset of $\{1, \dots, k\}$

does there exist $S \subseteq \{1, \dots, n\}$ s.t.

$\forall i \in \{1..k\}$ there is exactly one $j \in S$ s.t. $i \in A_j$



That is, a collection of sets that covers each element exactly once

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \stackrel{?}{=} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Is NP-complete. Clearly in NP.

Will prove hardness later.

k-Lin: given $a_1, \dots, a_m \in \mathbb{R}^n$, $b_1, \dots, b_m \in \mathbb{R}$, $k \in \mathbb{N}$
does there exist an $x \in \mathbb{R}^n$ s.t.
$\# \{ i : a_i^T x = b_i \} \geq k$.

Is there an x that satisfies at least
k of the equations?

Note: is easy when $k = m$.

Exact Cover $\leq_p$ k-Lin
We will reduce Exact Cover to k-Lin,
thereby proving that k-Lin is NP-hard.

Let A be the input to Exact Cover, and
let its dimensions be m-by-n.
Let a_i be the i^{th} row of A.
The list of equations will include $a_i^T x = 1$

But, we also need equations to force the entries
of x to be in $\{0, 1\}$.
So, add in equations $e_j^T x = 1$ <u>and</u> $e_j^T x = 0$,
for all $1 \leq j \leq n$, i.e. $x(j) = 1$ and $x(j) = 0$

There are now $m + 2n$ equations.
Set $k = m + n$.

$A \in$ Exact Cover $\Rightarrow$ equations, $k \in k$-Lin

if $Ax = 1$ and $x \in \{0,1\}^n$,

then the m equations $a_i^T x = 1$ are satisfied,

as are n of the equations $e_j^T x = 1$, $e_j^T x = 0$

equations, $k \in k$-Lin $\Rightarrow$ $A \in$ Exact Cover.

It is only possible to satisfy one of the

equations $e_j^T x = 1$ and $e_j^T x = 0$

So, at least n equations must be unsatisfied.

As there are $m+2n$ equations and at least

$k = m+n$ are satisfied,

it must be the case that for every $1 \le i \le n$,

$a_i^T x = 1$

and for every j, one of the equations

$e_j^T x = 1$ and $e_j^T x = 0$ is satisfied.

$\Rightarrow x \in \{0,1\}^n$

Sparse-Lin: Sparse solutions to linear equations.

Given $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $k \in \mathbb{N}$

Does there exist $x \in \mathbb{R}^n$ s.t. $Ax = b$

and $\#\{j : x(j) \neq 0\} = k$?

That is, x has exactly k non-zeros.

Sparse-Lin is NP-complete.

Is in NP because can check if $Ax=b$ in polynomial time, and x can not be too big.

Is NP-hard because Exact Cover $\leq_p$ Sparse-Lin

Let $\hat{A}$ be the matrix that is input to Exact Cover.

We will, of course, require $\hat{A}x = \mathbb{1}$

But, we need to force $x \in \{0,1\}^n$.

To do so, we add a vector of n variables, y, and equations $x(j) + y(j) = 1$, for $1 \leq j \leq n$.

We then set $k = n$.

$$A = \begin{pmatrix} \hat{A} & 0 \\ I & I \end{pmatrix} \quad b = \begin{pmatrix} \mathbb{1} \\ \mathbb{1} \end{pmatrix}$$

$\hat{A} \in$ Sparse-Lin $\Rightarrow$ equations have k-sparse solution

proof: let $x \in \{0,1\}^n$ satisfy $\hat{A}x = \mathbb{1}$, and set $y(j) = 1 - x(j)$.

Exactly n variables are 1 and n are 0.

$(k = n)$

equations have a k-sparse solution $\Rightarrow \hat{A} \in$ Sparse-Lin.

As $k=n$, this means that n variables must be zero.

As we can not have $x(j)=0$ and $y(j)=0$, it must be the case that

for every j $x(j)=0$ and $y(j)=1$

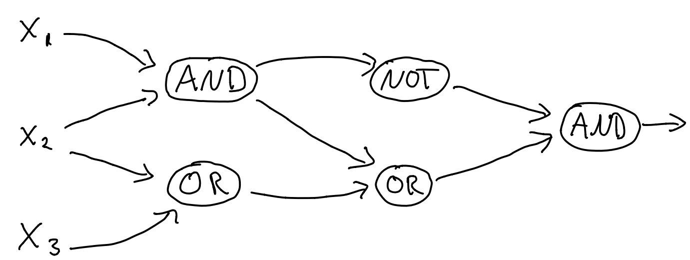
or $x(j)=1$ and $y(j)=0$

So, $x \in \{0,1\}^n$, and $\hat{A}x = \mathbb{1}$

Exact Cover is NP-hard.

because C-SAT $\le_p$ Exact Cover.

Recall C-SAT: given a circuit, is there an input that makes the output 1?



To ease description, we will give better names to elements of the set than $\{1..k\}$.

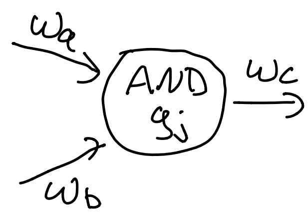
Let $g_1, \dots, g_k$ be the AND, OR, NOT gates.
Create an element for each, called $G_1, \dots, G_k$
Call the arrows connecting gates (and inputs) wires $w_1, \dots, w_m$, and make an element for each, $W_1, \dots, W_m$

Finally, for each wire we create 4 more elements I_j^0, I_j^1, O_j^0 and O_j^1, which stand for in and out on wire j.

The idea is that if a wire is transmitting a bit b, then that corresponds to elements I_j^b and O_j^b

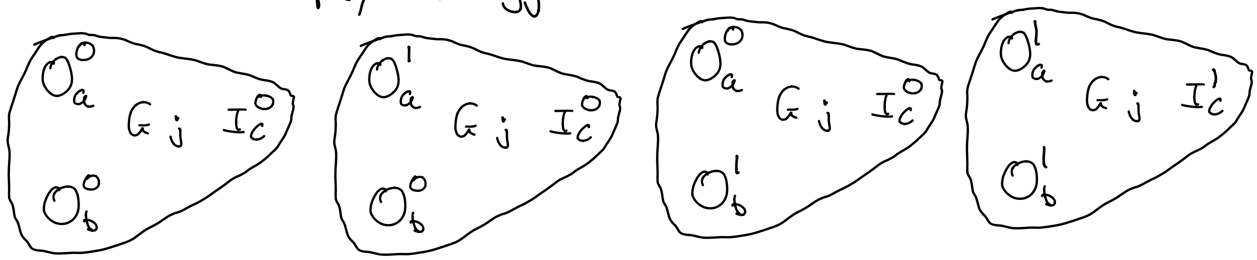
We now need to describe the sets.

These will correspond to allowable input/output relations at the gates and the wires.



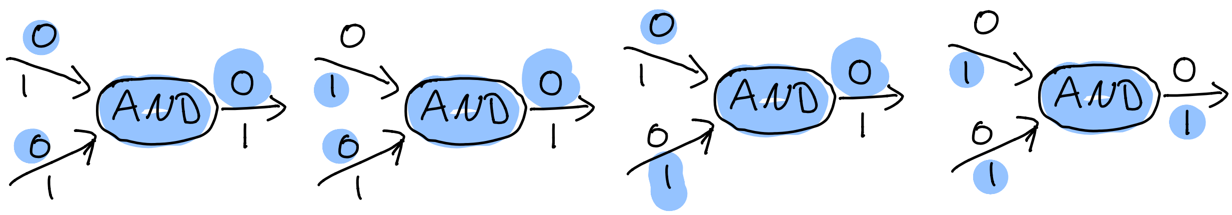
If w_a and w_b are the wires going in to gate g_j and w_c is the wire going out,
and $\alpha \in \{0,1\}$, $\beta \in \{0,1\}$, $\gamma \in \{0,1\}$ are values
s.t. $g_j(\alpha, \beta) = \gamma$
then we include set $\{O_a^\alpha, O_b^\beta, I_c^\gamma, G_j\}$

For example, if g_j is an AND we create sets



These are the only 4 sets containing G_j

This may be clearer if I draw the circuit, skip names, and just highlight elements in the set



If g_j is a NOT, it gets 2 sets like

$$O_a^0 \quad G_j \quad I_c^1 \qquad \text{and} \qquad O_a^1 \quad G_j \quad I_c^0$$

Similarly, an OR gate gets 4 sets like

$$\{O_a^0, O_b^0, O_c^0, G_j\}$$
$$\{O_a^1, O_b^0, O_c^1, G_j\}$$
$$\{O_a^0, O_b^1, O_c^1, G_j\}$$
$$\{O_a^1, O_b^1, O_c^1, G_j\}$$

If a gate has many outputs, we include all of them. For the output gate we only include the sets in which the output is true.

For each gate, we should choose the set that corresponds to the values on the wires attached to it.

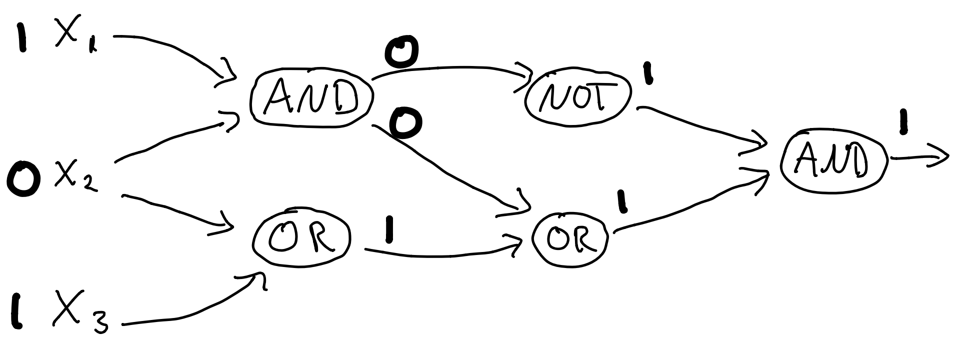
For wire w_i, we create two sets $\{I_i^0, w_i, O_i^0\}$ and $\{I_i^1, w_i, O_i^1\}$ we include in the exact cover the one of these that does NOT correspond to the values on the wire.

For input X_i, with outputs on wires $w_a, \dots, w_c$, we create two sets:

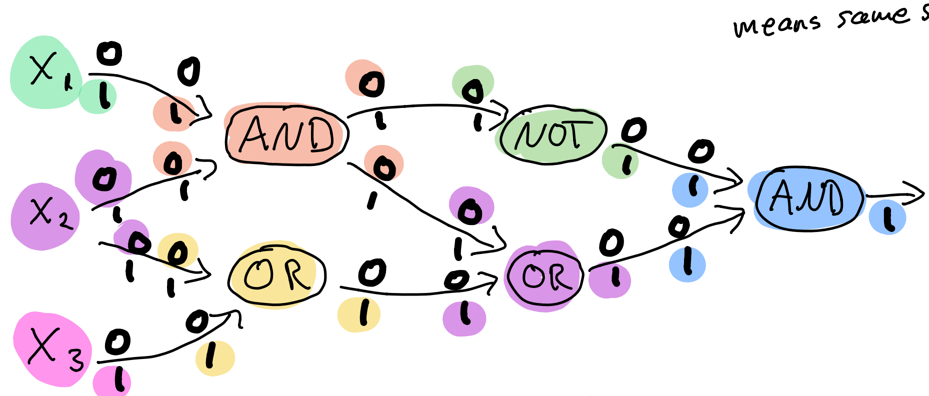
$$\{X_i, O^0_a, \dots, O^0_c\} \quad \text{and} \quad \{X_i, O^1_a, \dots, O^1_c\}$$

I now claim, and sketch, that this system has an exact set cover iff the circuit is satisfiable.

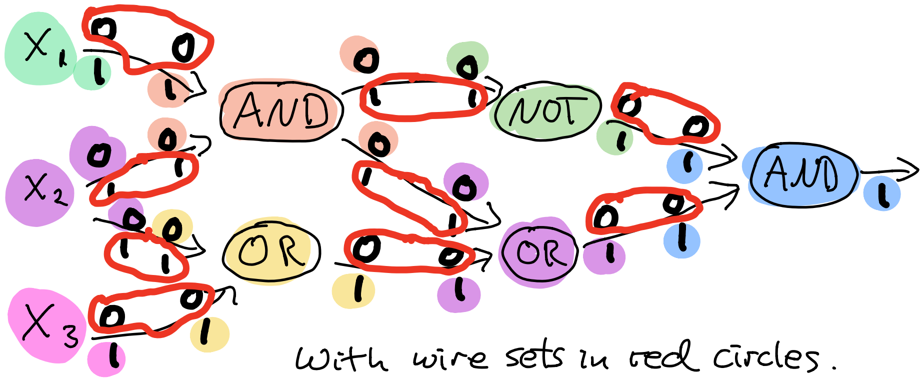
I'll just sketch the correspondance between a satisfying assignment and an exact cover.



same color means same set.



Just the gate sets



With wire sets in red circles.

If each element is in exactly one set:
the wire sets guarantee exactly one value, 0/1, is unused on each wire.

The gate sets guarantee that the used values on wires obey the rules of the gates.

More NP-complete problems (until time runs out)

Subset Sum.

Input integers $a_1, \ldots, a_n$, t, where t is "target"

Answer "yes" if $\exists S \subseteq \{1, \ldots, n\}$ s.t. $\sum_{i \in S} a_i = t$

Algebraic version: Given $a \in \mathbb{Z}^n$ and t is there an $x \in \{0,1\}^n$ s.t. $a^T x = t$?

Exact Cover $\leq_p$ Subset Sum

Exact Cover asks if $\exists x \in \{0,1\}^n$ s.t. $Ax = \mathbb{1}$

This holds iff for all y, $y^T A x = y^T \mathbb{1}$

Think of $a = y^T A$ and $t = y^T \mathbb{1}$.

But we want one y for all x.

Note $Ax \in \{0, 1, \ldots, m\}^n$

So, set $y = (1, m+1, (m+1)^2, \ldots, (m+1)^{n-1})$

Claim $Ax = \mathbb{1}$ iff $y^T A x = y^T \mathbb{1}$

<u>proof</u>: for $z \in \{0, 1, \ldots, m\}^n$,

$y^T z$ uniquely determines z

NMF: (non-negative matrix factorization)

Given a matrix $A \in \mathbb{R}_+^{m \times n}$ of rank k

"yes" if there exist $U \in \mathbb{R}_+^{m \times k}$ and $V \in \mathbb{R}_+^{k \times n}$

s.t. $UV = A$

(Vavasis '09, maybe Shitov '16)

Low-rank matrix completion:

Given $A \in \mathbb{R}^{m \times n}$, integer k, and

$S \subseteq \{1..m\} \times \{1..n\}$

"yes" if $\exists\, B \in \mathbb{R}^{m \times n}$ of rank $\le k$ s.t.

$A(i,j) = B(i,j)$ for all $(i,j) \in S$.

Consider entries not in S unknown.

Can we fill in the unknown entries to get rank $\le k$.

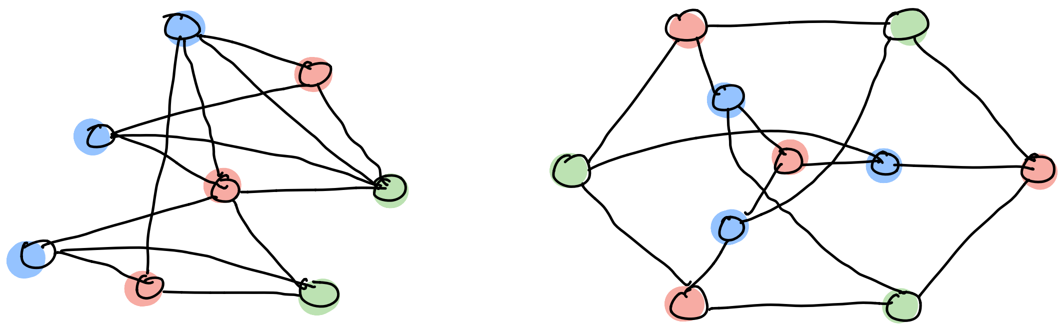
(Peeters '96)

Both can be proved NP-hard by reductions from k-color.

Given $G = (V, E)$, a k-coloring of G is a

$c: V \to \{1, .., k\}$ (colors)

s.t. for all $(a,b) \in E$ $c(a) \neq c(b)$



k-color: Given G and k,
does G have a k-coloring?

Is hard even for $k=3$.

For every $\delta > 0$, if can distinguish
n^{δ} colorable from
$n^{1-\delta}$ colorable,
then $NP \subseteq ZPP$ (zero-error randomized polynomial time)