Satisfiability threshold for $2 - SAT$ in new model - work in progress

Mikołaj Pudo

Jagiellonian University
Mathematics and Computer Science Faculty
Foundations of Computer Science Department

CLA, 2009

Table of contents

Table of contents

$k - SAT$ problem

- $V_m = \{x_1, x_2, \ldots, x_m\}$ - set of boolean variables,
- $C = \pm x_{i_1} \vee \ldots \vee \pm x_{i_k}$ - k-clause,
- $F = (\pm x_{1_1} \vee \ldots \vee \pm x_{1_k}) \wedge \ldots \wedge (\pm x_{n_1} \vee \ldots \vee \pm x_{n_k})$ - $k - CNF$ formula,
- $\Omega_k(n, m)$ - set of $k - CNF$ formulas with n clauses over set of m variables,

$k - SAT$ problem

- $V_m = \{x_1, x_2, \dots, x_m\}$ - set of boolean variables,
- $C = \pm x_{i_1} \vee \dots \vee \pm x_{i_k}$ - k-clause,
- $F = (\pm x_{1_1} \vee \dots \vee \pm x_{1_k}) \wedge \dots \wedge (\pm x_{n_1} \vee \dots \vee \pm x_{n_k})$ - $k - CNF$ formula,
- $\Omega_k(n, m)$ - set of $k - CNF$ formulas with n clauses over set of m variables,

Satisfiability probability

- $\mathcal{A}\colon V_m \to \{0, 1\}$ - valuation,
- $\mathcal{A}(\phi)$ - set of all solutions for formula ϕ,
- $c = \frac{n}{m}$ - clause to variables density,
- $S_k(m, c) = Pr[\phi \in \Omega_k(mc, m) \, \text{satisfiable}]$,

Satisfiability probability

- $\mathcal{A}\colon V_m \to \{0, 1\}$ - valuation,
- $\mathcal{A}(\phi)$ - set of all solutions for formula ϕ,
- $c = \frac{n}{m}$ - clause to variables density,
- $S_k(m, c) = Pr[\phi \in \Omega_k(mc, m)\,\text{satisfiable}]$,

Satisfiability threshold conjecture

Conjecture

For each $k \geq 2$ there is c_k such that for all $\varepsilon > 0$

$$\lim_{m \to \infty} S_k(m, c_k - \varepsilon) = 1$$

and

$$\lim_{m \to \infty} S_k(m, c_k + \varepsilon) = 0.$$

The state of the art

- For $k = 2$ it is known that $c_2 = 1$,
- For $k \geq 3$ we don't know...
- Experiments show that: $c_3 \approx 4.25 \pm 0.05$,
- Lower and upper bounds for interval containig c_k (if it exists), eg. $3.42 < c_3 < 4.506$.

The state of the art

- For $k = 2$ it is known that $c_2 = 1$,
- For $k \geq 3$ we don't know...
- Experiments show that: $c_3 \approx 4.25 \pm 0.05$,
- Lower and upper bounds for interval containig c_k (if it exists), eg. $3.42 < c_3 < 4.506$.

Table of contents

Variables permutation

Consider two formulas over $V = \{x, y, z\}$:

$$(x \vee \overline{y}) \wedge (y \vee z)$$

and

$$(z \vee \overline{x}) \wedge (x \vee y)$$

So far they were considered different, but actually they don't differ that much.
Take permutation $\sigma : V \to V$ such that:
$\sigma(x) = z, \sigma(y) = x, \sigma(z) = y$

Variables permutation

Consider two formulas over $V = \{x, y, z\}$:

$$(x \vee \overline{y}) \wedge (y \vee z)$$

and

$$(z \vee \overline{x}) \wedge (x \vee y)$$

So far they were considered different, but actually they don't differ that much.
Take permutation $\sigma : V \rightarrow V$ such that:
$\sigma(x) = z, \sigma(y) = x, \sigma(z) = y$

- $\Phi_k(n, m)$ - set of $k - CNF$ formulas with n clauses over set of m variables,
- $S_k(m, c) = Pr[\phi \in \Phi_k(mc, m) \, \text{satisfiable}]$,
- For each $k \geq 2$ find c_k such that for all $\varepsilon > 0$

$$\lim_{m \to \infty} S_k(m, c_k - \varepsilon) = 1$$

and

$$\lim_{m \to \infty} S_k(m, c_k + \varepsilon) = 0$$

- Since $2 - SAT$ is always the easiest part of the game…

- $\Phi_k(n, m)$ - set of $k - CNF$ formulas with n clauses over set of m variables,
- $S_k(m, c) = Pr[\phi \in \Phi_k(mc, m)\,\text{satisfiable}]$,
- For each $k \geq 2$ find c_k such that for all $\varepsilon > 0$

$$\lim_{m \to \infty} S_k(m, c_k - \varepsilon) = 1$$

and

$$\lim_{m \to \infty} S_k(m, c_k + \varepsilon) = 0$$

- Since $2 - SAT$ is always the easiest part of the game...

Sketch of proof

- Construct a graph from each $2 - CNF$ formula,
- Describe all graphs constructed from unsatisfiable formulas,
- Count those graphs and show that the number tends to zero.

Table of contents

Graph construction

Every formula $\phi \in \Phi(n, m)$ can be transformed into a directed graph with coloring function $G_\phi = (V, E, \alpha)$, where:

- V - vertex set, $|V| = 4 \cdot n$
- E - edge set,
- $\alpha : V \to \{1 \dots m\}$ - coloring function.

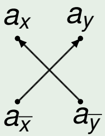
G_ϕ is made of 4-element gadgets.

Gadgets

Consider formula ϕ. Every clause $a = (x \vee y) \in \phi$ gives one 4-element gadget with vertices: $a_x, a_{\overline{x}}, a_y, a_{\overline{y}}$, and two gadget edges: $a_{\overline{x}} \to a_y, a_{\overline{y}} \to a_x$.

Example

$(x \vee y)$



Mid-gadget edges

For all $a, b \in \phi$, if $y \in a \wedge \overline{y} \in b$, then take gadgets constructed from a and b and join them with mid-gadget edges: $a_y \rightarrow b_y$ and $b_{\overline{y}} \rightarrow a_{\overline{y}}$

Example

$$\phi = a \wedge b$$
$$a = (x \vee y)$$
$$b = (\overline{y} \vee z)$$

Graph coloring

Construct $\alpha : V \rightarrow \{1 \ldots m\}$ such that for all $a, b \in \phi$ if $(y \in a \vee \overline{y} \in a)$ and $(y \in b \vee \overline{y} \in b)$, then $\alpha(a_y) = \alpha(a_{\overline{y}}) = \alpha(b_y) = \alpha(b_{\overline{y}})$.

Example

$$(\overline{x_1} \vee \overline{x_2}) \wedge$$
$$(x_2 \vee \overline{x_3}) \wedge$$
$$(x_3 \vee x_1) \wedge$$
$$(\overline{x_4} \vee x_5)$$

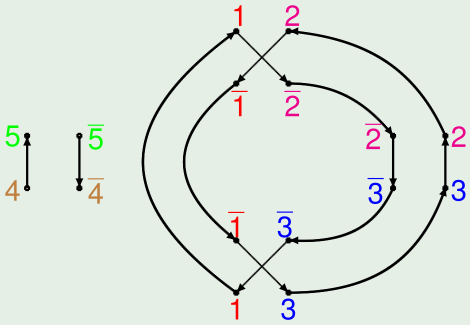
$x_1 \longrightarrow 1$

$x_2 \longrightarrow 2$

$x_3 \longrightarrow 3$

$x_4 \longrightarrow 4$

$x_5 \longrightarrow 5$



Contradictory cycle

Definition

A cycle in G_ϕ is a *contradictory cycle* if it contains at least two vertices marked with the same color, but with opposit signs. Such vertices are called *contradictory vertices*. Mid-gadget edges connecting contradictory vertices are called *contradictory edges*.

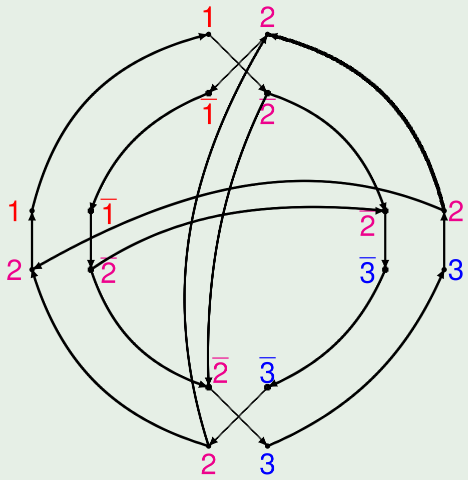
Lemma

Formula $\phi \in \Phi(n, m)$ is unsatisfiable if and only if G_ϕ contains contradictory cycle.

Example

$(\overline{x_1} \vee \overline{x_2})\wedge$
$(x_2 \vee \overline{x_3})\wedge$
$(x_3 \vee \overline{x_2})\wedge$
$(\overline{x_2} \vee \overline{x_1})\wedge$

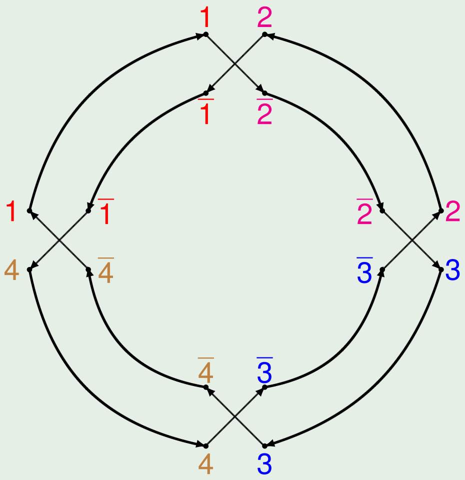
$x_1 \longrightarrow 1$
$x_2 \longrightarrow 2$
$x_3 \longrightarrow 3$



Example

$(\overline{x_1} \vee \overline{x_2})\wedge$
$(x_2 \vee x_3)\wedge$
$(\overline{x_3} \vee \overline{x_4})\wedge$
$(x_4 \vee x_1)\wedge$

$x_1 \longrightarrow 1$
$x_2 \longrightarrow 2$
$x_3 \longrightarrow 3$
$x_4 \longrightarrow 4$



Random variable

Definition

Let $X_{n,m} : \Phi(n, m) \to \mathbb{N}$ be a random variable such that for a formula $\phi \in \Phi(n, m)$, $X_{n,m}(\phi)$ is equal to the number of contradictory cycles in graph G_ϕ.

Markov's inequality

Since the random variable $X_{n,m}(\phi)$ gives the number of contradictory cycles in random formula ϕ, by Lemma only those formulas for which $X_{n,m}(\phi) \geq 1$ holds are unsatisfiable.

We can use Markov's inequality:

$$Pr\left(\phi \in \Phi(n,m) | \phi \text{ - unsatisfiable}\right) =$$
$$= Pr\left(\phi \in \Phi(n,m) | X_{n,m}(\phi) \geq 1\right) \leq E[X_{n,m}],$$

where:

$E[X_{n,m}]$ - expected value of random variable $X_{n,m}$.

Markov's inequality

Since the random variable $X_{n,m}(\phi)$ gives the number of contradictory cycles in random formula ϕ, by Lemma only those formulas for which $X_{n,m}(\phi) \geq 1$ holds are unsatisfiable.
We can use Markov's inequality:

$$Pr\left(\phi \in \Phi(n,m) | \phi \text{ - unsatisfiable}\right) =$$
$$= Pr\left(\phi \in \Phi(n,m) | X_{n,m}(\phi) \geq 1\right) \leq E[X_{n,m}],$$

where:
$E[X_{n,m}]$ - expected value of random variable $X_{n,m}$.

All we have to do is:

- Find $E[X_{n,m}]$,
- Show that $E[X_{n,m}] = o(1)$, assuming $n = c \cdot m$, and $c < 1$.

$$E[X_{n,m}] = \sum_{\phi\in\Phi(n,m)} X_{n,m}(\phi)\cdot Pr\left(\phi\in\Phi\left(n,m\right)\right) =$$

$$= \frac{\sum_{\phi\in\Phi(n,m)} X_{n,m}(\phi)}{|\Phi(n,m)|}$$

The denominator of $E[X_{n,m}]$ is easy:

$$|\Phi(n,m)| = \left\{ {2n \atop m} \right\} 2^{2n},$$

where:
$\left\{ {2n \atop m} \right\}$ - Stirling number of the second kind.
The numerator requires more perspiration…

$$E[X_{n,m}] = \sum_{\phi \in \Phi(n,m)} X_{n,m}(\phi) \cdot Pr\left(\phi \in \Phi\left(n,m\right)\right) =$$

$$= \frac{\sum_{\phi \in \Phi(n,m)} X_{n,m}(\phi)}{|\Phi(n,m)|}$$

The denominator of $E[X_{n,m}]$ is easy:

$$|\Phi(n,m)| = \left\{ {2n \atop m} \right\} 2^{2n},$$

where:
$\left\{ {2n \atop m} \right\}$ - Stirling number of the second kind.
The numerator requires more perspiration...

$$E[X_{n,m}] = \sum_{\phi \in \Phi(n,m)} X_{n,m}(\phi) \cdot Pr\left(\phi \in \Phi\left(n,m\right)\right) =$$

$$= \frac{\sum_{\phi \in \Phi(n,m)} X_{n,m}(\phi)}{|\Phi(n,m)|}$$

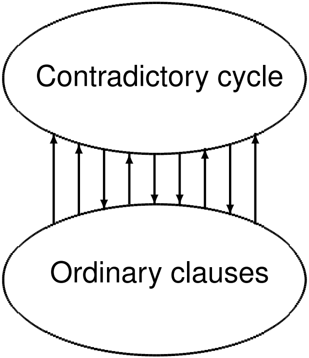
The denominator of $E[X_{n,m}]$ is easy:

$$|\Phi(n,m)| = \left\{ {2n \atop m} \right\} 2^{2n},$$

where:
$\left\{ {2n \atop m} \right\}$ - Stirling number of the second kind.
The numerator requires more perspiration. . .

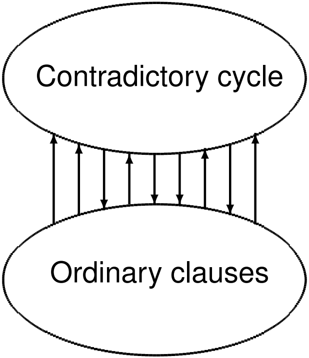
Numerator of $E[X_{n,m}]$

Choose clauses for the contradictory cycle.
Count all contradictory cycles which can be built on those clauses



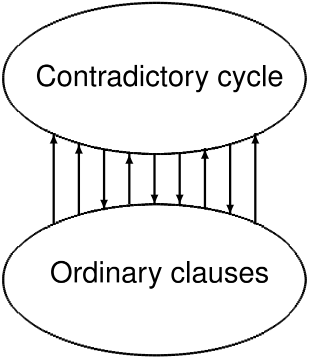
Numerator of $E[X_{n,m}]$

Color ordinary clauses with some variables and signs.



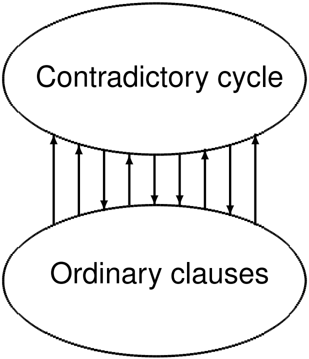
Numerator of $E[X_{n,m}]$

Some variables from ordinary clauses can be joined with variables from contradictory cycle.



Numerator of $E[X_{n,m}]$

Remember to count all possible choices of clauses for contradictory cycle, and all lengths of those cycles.



A scary formula

$$\sum_{\phi \in \Phi(n,m)} X_{n,m}(\phi) =$$

$$\sum_{a=2}^{n} \sum_{b=1}^{\lfloor \frac{a}{2} \rfloor} \sum_{r=0}^{a-b} n^{\underline{a}} \cdot a^{\underline{(2b)}} \cdot \frac{(m-r)!}{(m-a+b)!} \binom{a-b}{r} \left\{ {2n-2a \atop m-r} \right\} 2^{2n-2b},$$

where:

$n^{\underline{a}} = \frac{n!}{(n-a)!}$ - falling factorial power

Even more scary formula

$$E[X_{n,m}] = \frac{\sum_{a=2}^{n} \sum_{b=1}^{\lfloor \frac{a}{2} \rfloor} \sum_{r=0}^{a-b} \frac{n^{\underline{a}} \cdot a^{\underline{(2b)}} \cdot (m-r)!}{a \cdot 4^{b} \cdot (m-a+b)!} \binom{a-b}{r} \left\{ {2n-2a \atop m-r} \right\}}{\left\{ {2n \atop m} \right\}},$$

where:

$\left\{ {n \atop m} \right\}$ - Stirling number of the second kind,

$n^{\underline{a}} = \frac{n!}{(n-a)!}$

Since $n = mc$

$$E[X_{mc,m}] =$$

$$= \sum_{a=2}^{mc} \sum_{b=1}^{\lfloor \frac{a}{2} \rfloor} \sum_{r=0}^{a-b} \frac{(mc)^{\underline{a}} \cdot a^{\underline{(2b)}} \cdot (m-r)!}{a \cdot 4^b \cdot (m-a+b)!} \binom{a-b}{r} \frac{\left\{ {2mc-2a \atop m-r} \right\}}{\left\{ {2mc \atop m} \right\}}$$

What next?

What next?

1. Find upper bound for c_2
2. Find upper and lower bound for $c_k, k \geq 3$
3. Attack conjecture

Thank you.