Phase Transitions and Computational Complexity

Amin Coja-Oghlan

Warwick/Frankfurt

The spectre of NP-hardness

NP-complete problems [Cook, Karp 1970s]

- Literally hundreds of fundamental problems are *NP-hard*:
  - Boolean Satisfiability [model checking, VLSI, ...],
  - Graph Colouring [timetabling/allocation, ...],
  - Travelling Salesman [scheduling problems, ...],
  - ...
- P≠NP conjecture: no efficient algoithm in the *worst case*.
- $\rightsquigarrow$ *"Average Case analysis of algorithms"*

Phase transitions in discrete structures

- Colouring random graphs or hypergraphs
- Independent sets in random graphs
- Random k-SAT

Phase transitions in discrete structures

- Colouring random graphs or hypergraphs
- Independent sets in random graphs
- Random k-SAT

- Conjectured phase transitions. . . *[experiments, physics]*
- . . . appear to affect the combinatorial and computational nature.

Outline

- Random k-SAT: "classical" stuff.
- The physics perspective (non-rigorous).
- Physics-inspired rigorous stuff.
- Open problems.

The k-SAT problem

- Fix $k \geq 3$ and let $x_1, x_2, \ldots, x_n$ be *Boolean variables*.
- Given an expression of the form

$$\underbrace{(x_1 \vee \bar{x}_{17} \vee \cdots \vee x_{29})}_{k\text{-clause}} \wedge \underbrace{(\bar{x}_{11} \vee x_2 \vee \cdots \vee \bar{x}_1)}_{k\text{-clause}} \wedge \cdots$$

- ...find a *Boolean assignment* that makes the entire formula true.

The k-SAT problem

- Fix $k \geq 3$ and let $x_1, x_2, \ldots, x_n$ be *Boolean variables*.
- Given an expression of the form

$$\underbrace{(x_1 \vee \bar{x}_{17} \vee \cdots \vee x_{29})}_{k\text{-clause}} \wedge \underbrace{(\bar{x}_{11} \vee x_2 \vee \cdots \vee \bar{x}_1)}_{k\text{-clause}} \wedge \cdots$$

- ...find a *Boolean assignment* that makes the entire formula true.

The naive algorithm

- Check all 2^n *possible assignments*!
- (Prohibitive even for $n = 100$.)
- Yet no better worst-case algorithm is known.

Random k-SAT

The random k-SAT model

- Fix $k \geq 3$ and let $x_1, \ldots, x_n$ be Boolean variables.
- Generate k-clauses $C_1, \ldots, C_m$ *uniformly and independently*.
- Let $\mathbf{\Phi} = C_1 \wedge \cdots \wedge C_m$.

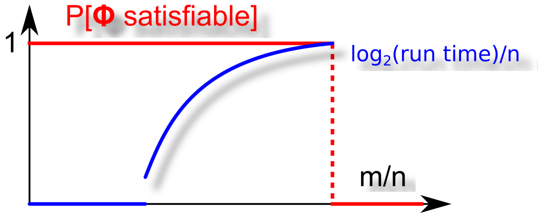
Random k-SAT

The random k-SAT model

- Fix $k \geq 3$ and let $x_1, \ldots, x_n$ be Boolean variables.
- Generate k-clauses $C_1, \ldots, C_m$ *uniformly and independently*.
- Let $\mathbf{\Phi} = C_1 \wedge \cdots \wedge C_m$.

Kirkpatrick, Selman (experimental) [Science 1994]

- There occurs a satisfiability phase transition.
- *Industrial SAT solvers* require exponential time.



Random k-SAT

Theorem [Achlioptas, Peres: JAMS 2004]

The random formula $\mathbf{\Phi}$ is...

- ...satisfiable if $m/n < 2^k \ln 2 - \Theta(k)$.
- ...unsatisfiable if $m/n > 2^k \ln 2$.

Random k-SAT

Theorem [Achlioptas, Peres: JAMS 2004]

The random formula $\mathbf{\Phi}$ is...

- ...satisfiable if $m/n < 2^k \ln 2 - \Theta(k)$.
- ...unsatisfiable if $m/n > 2^k \ln 2$.

Proof

Non-constructive ('2nd moment method').

Random k-SAT

Hunting the k-SAT threshold

- The precise k-SAT threshold is *unknown* for any $k \geq 3$.
- (In fact, it is not known to exist.)

Algorithms for random k-SAT

Question

- The *threshold* is $r_k \sim 2^k \ln 2$.
- For what m/n can we *find* satisfying assignments *efficiently*?

Algorithms for random k-SAT

Question

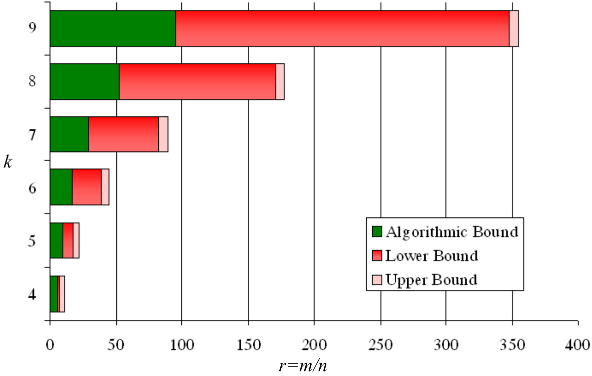
- The *threshold* is $r_k \sim 2^k \ln 2$.
- For what m/n can we *find* satisfying assignments *efficiently*?

Algorithm	*Density* $m/n < \cdots$	
Pure Literal	$\sim 2\ln(k)/k$	BFU 1993
Walksat, proven	$0.04 \cdot 2^k/k$	ACO, Frieze 2012
Walksat, conjectured	$2^k/k$	Monasson, Semerijan 2003
Shortest Clause	$\frac{e^2}{8} \cdot 2^k/k$	Chvatal, Reed 1992
Unit Clause	$\frac{e}{2} \cdot 2^k/k$	Chao, Franco 1990
SC+backtracking	$1.817 \cdot 2^k/k$	Frieze, Suen 1996
`Fix`	$\sim 2^k \ln(k)/k$	ACO 2009

Algorithms for random k-SAT

Question

- The *threshold* is $r_k \sim 2^k \ln 2$.
- For what m/n can we *find* satisfying assignments *efficiently*?



The statistical mechanics perspective

- “Mean-field models of disorered systems” (such as glasses).
- Phase transitions in glasses hypothesized by *Kauzmann* (1948).

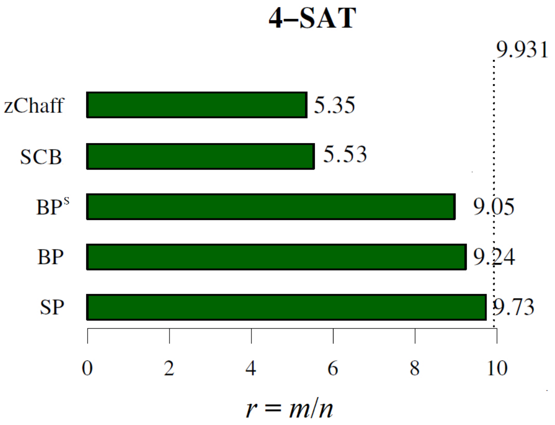
Enter the physicists

Mezard, Parisi, Zecchina [Science 2002]

- The (non-rigorous) cavity method.
- “Replica symmetry breaking”.
- New algorithms: *Belief/Survey Propagation* guided decimation.

Experiments [Kroc, Sabharwal, Selman 2009]

For small k, BP/SP guided decimation fare *extremely* well.



The cavity method

A thought experiment

- Generate a random formula $\mathbf{\Phi}$. Fix it.

The cavity method

A thought experiment

- Generate a random formula $\mathbf{\Phi}$. Fix it.
- Choose a *satisfying assignment* σ at random.

The cavity method

A thought experiment

- Generate a random formula $\mathbf{\Phi}$. Fix it.
- Choose a *satisfying assignment* σ at random.
- For a bunch of variables $y_1, \ldots, y_l$, how can we characterise

$$(\sigma(y_1), \ldots, \sigma(y_l))?$$

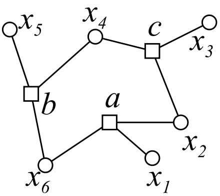
The cavity method

A thought experiment

- Generate a random formula **Φ**. Fix it.
- Choose a *satisfying assignment* σ at random.
- For a bunch of variables $y_1, \ldots, y_l$, how can we characterise
$$(\sigma(y_1), \ldots, \sigma(y_l))?$$

The factor graph $G(\Phi)$

- *Vertices:* clauses and variables.
- *Edges* between clauses and the variables they contain $\rightsquigarrow$ metric.



The cavity method

Replica symmetry/non-reconstruction $[m/n < 2^k \ln(k)/k]$

- Correlations are purely local.

The cavity method

Replica symmetry/non-reconstruction $[m/n < 2^k \ln(k)/k]$

- Correlations are purely local.
- Fix a variable y.
- Let $\mathcal{B}_{\geq\omega} = \{\text{variables at distance } \geq \omega \text{ from } y\}$.

The cavity method

Replica symmetry/non-reconstruction $[m/n < 2^k \ln(k)/k]$

- Correlations are purely local.
- Fix a variable y.
- Let $\mathcal{B}_{\geq\omega} = \{\text{variables at distance} \geq \omega \text{ from } y\}$.
- Then

$$\mathrm{E}\left\|\sigma(y) - \sigma(y|\mathcal{B}_{\geq\omega})\right\|_{tv} \to 0.$$

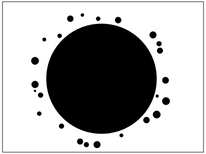
The cavity method

Replica symmetry/non-reconstruction $[m/n < 2^k \ln(k)/k]$

- Correlations are purely local.
- Fix a variable y.
- Let $\mathcal{B}_{\geq\omega} = \{\text{variables at distance} \geq \omega \text{ from } y\}$.
- Then

$$\mathrm{E}\left\|\sigma(y) - \sigma(y|\mathcal{B}_{\geq\omega})\right\|_{tv} \to 0.$$

- Geometrically, $\mathcal{S}(\mathbf{\Phi}) = \{\text{all satisfying assignments}\}$ is a "giant ball".



The cavity method

Replica symmetry breaking $[m/n > 2^k \ln(k)/k]$

- There are long-range correlations.

The cavity method

Replica symmetry breaking $[m/n > 2^k \ln(k)/k]$

- There are long-range correlations.
- Fix a variable y.
- Let $\mathcal{B}_{\geq\omega} = \{\text{variables at distance} \geq \omega \text{ from } y\}$.

The cavity method

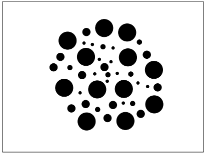
Replica symmetry breaking $[m/n > 2^k \ln(k)/k]$

- There are long-range correlations.
- Fix a variable y.
- Let $\mathcal{B}_{\geq\omega} = \{\text{variables at distance } \geq \omega \text{ from } y\}$.
- Then $\sigma(y)$ is *determined* by $\sigma(y|\mathcal{B}_{\geq\omega})$ w.p. $1-\varepsilon$.

The cavity method

Replica symmetry breaking $[m/n > 2^k \ln(k)/k]$

- There are long-range correlations.
- Fix a variable y.
- Let $\mathcal{B}_{\geq\omega} = \{\text{variables at distance} \geq \omega \text{ from } y\}$.
- Then $\sigma(y)$ is *determined* by $\sigma(y|\mathcal{B}_{\geq\omega})$ w.p. $1-\varepsilon$.
- Geometrically, $\mathcal{S}(\mathbf{\Phi}) = \{\text{all satisfying assignments}\}$ "shatters".



The cavity method

Replica symmetry breaking $[m/n > 2^k \ln(k)/k]$

- There are long-range correlations.
- Fix a variable y.
- Let $\mathcal{B}_{\geq\omega} = \{\text{variables at distance} \geq \omega \text{ from } y\}$.
- Then $\sigma(y)$ is *determined* by $\sigma(y|\mathcal{B}_{\geq\omega})$ w.p. $1-\varepsilon$.

Still, so long as

$$m/n < r_{cond} = 2^k \ln 2 - \frac{3}{2}\ln 2 + o_k(1),$$

for $y_1, \ldots, y_l$ *mutually far apart* we have

$$\mathrm{E}\left\|\sigma(y_1, \ldots, y_l) - \sigma(y_1) \otimes \cdots \otimes \sigma(y_l)\right\|_{tv} \to 0.$$

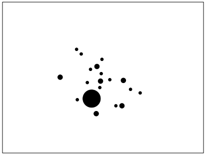
The cavity method

Condensation $[r_{cond} < m/n < r_{sat}]$

- A few large clusters dominate.
- In particular,

$$\mathrm{E}\left\|\sigma(y_1,\ldots,y_l)-\sigma(y_1)\otimes\cdots\otimes\sigma(y_l)\right\|_{tv}\not\to 0.$$

- Still, there is an abundance of small clusters.



Belief Propagation

- Suppose that

$$\mathrm{E}\left\|\sigma(y_1,\ldots,y_l)-\sigma(y_1)\otimes\cdots\otimes\sigma(y_l)\right\|_{tv}\to 0.$$

Belief Propagation

- Suppose that

$$\mathrm{E}\left\|\sigma(y_1,\ldots,y_l)-\sigma(y_1)\otimes\cdots\otimes\sigma(y_l)\right\|_{tv}\to 0.$$

- Locally, the factor graph is a *tree*.

Belief Propagation

- Suppose that

$$\mathrm{E}\left\|\sigma(y_1,\ldots,y_l)-\sigma(y_1)\otimes\cdots\otimes\sigma(y_l)\right\|_{tv}\to 0.$$

- Locally, the factor graph is a *tree*.
- Therefore, for a *fixed* $\mathbf{\Phi}$, there is a *recurrence* for

$$\mu(y)=\mathrm{E}_{\sigma\in\mathcal{S}(\mathbf{\Phi})}\left[\sigma(y)\right].$$

- $\rightsquigarrow$ Belief Propagation equations.

Survey propagation

- The *Survey propagation distribution* on $\mathcal{S}(\mathbf{\Phi})$:
  - Pick a cluster $\mathcal{C}$ uniformly.
  - Pick a solution $\sigma' \in \mathcal{C}$ uniformly.

Survey propagation

- The *Survey propagation distribution* on $\mathcal{S}(\mathbf{\Phi})$:
  - Pick a cluster $\mathcal{C}$ uniformly.
  - Pick a solution $\sigma' \in \mathcal{C}$ uniformly.
- For this modified measure we have

$$\mathrm{E}\left\|\sigma'(y_1,\ldots,y_l) - \sigma'(y_1)\otimes\cdots\otimes\sigma'(y_l)\right\|_{tv} \to 0.$$

- $\rightsquigarrow$ Survey Propagation equations.

Survey propagation

- The *Survey propagation distribution* on $\mathcal{S}(\mathbf{\Phi})$:
  - Pick a cluster $\mathcal{C}$ uniformly.
  - Pick a solution $\sigma' \in \mathcal{C}$ uniformly.
- For this modified measure we have

$$\mathrm{E}\left\|\sigma'(y_1,\ldots,y_l)-\sigma'(y_1)\otimes\cdots\otimes\sigma'(y_l)\right\|_{tv}\to 0.$$

- $\rightsquigarrow$ Survey Propagation equations.

- Yields a prediction on the *k-SAT threshold*...
- ...in the form of an infinite-dimensional variational problem.

Message passing algorithms

Belief Propagation guided decimation

- For $i = 1, \ldots, n$ do
-     use BP to approximate $\mu(x_i)$.
-     set x_i to *true* with the resulting probability.
-     simplify the formula accordingly.

Message passing algorithms

Belief Propagation guided decimation

- For $i = 1, \ldots, n$ do
- use BP to approximate $\mu(x_i)$.
- set x_i to *true* with the resulting probability.
- simplify the formula accordingly.

A similar scheme yields *Survey Propagation guided decimation*.

Proving the physics hypotheses

- Hunting the k-SAT threshold
- Shattering/Condensation
- Message passing algorithms

Hunting the k-SAT threshold

- Best current bounds: 2nd moment method [Achlioptas, Peres 2004]
- Requires *symmetry*.

Hunting the k-SAT threshold

- Best current bounds: 2nd moment method [Achlioptas, Peres 2004]
- Requires *symmetry*.

- Survey Propagation equations $\rightsquigarrow$ *upper bound* [Franz, Leone 2003]
- Generalises to other problems [Panchenko, Talagrand 2004]
- *Sherrington-Kirkpatrick* model [Talagrand 2006]

Hunting the k-SAT threshold

- Best current bounds: 2nd moment method [Achlioptas, Peres 2004]
- Requires *symmetry*.

- Survey Propagation equations $\rightsquigarrow$ *upper bound* [Franz, Leone 2003]
- Generalises to other problems [Panchenko, Talagrand 2004]
- *Sherrington-Kirkpatrick* model [Talagrand 2006]

Theorem [ACO, Panagiotou 2012]

The threshold for 2-coloring k-uniform hypergraphs is

$$r_{2-col} = 2^{k-1}\ln 2 - \frac{\ln 2}{2} - \frac{1}{4} + \tilde{O}(2^{-k}).$$

The solution space geometry

Loose/frozen variables

Let Φ be a k-CNF, σ a *satisfying assignment*, and x a variable.

- x is loose if there is a satisfying assignment τ such that

$$\sigma(x) \neq \tau(x) \text{ and } \mathrm{dist}(\sigma, \tau) \leq \ln(n).$$

The solution space geometry

Loose/frozen variables

Let Φ be a k-CNF, σ a *satisfying assignment*, and x a variable.

- x is loose if there is a satisfying assignment τ such that

$$\sigma(x) \neq \tau(x) \text{ and } \mathrm{dist}(\sigma, \tau) \leq \ln(n).$$

- x is frozen if for any satisfying assignment τ

$$\sigma(x) \neq \tau(x) \Rightarrow \mathrm{dist}(\sigma, \tau) = \Omega(n).$$

The solution space geometry

Theorem [Achlioptas, ACO 2008]

Choose $\sigma \in \mathcal{S}(\mathbf{\Phi})$ uniformly at random.

1. $r < (1 - o_k(1))2^k \ln k/k \Rightarrow$ all but $o_k(1)n$ vars are *loose* w.h.p.
2. $r > (1 + o_k(1))2^k \ln k/k \Rightarrow$ all but $o_k(1)n$ vars are *frozen* w.h.p.

The solution space geometry

Theorem [Achlioptas, ACO 2008]

Choose $\sigma \in \mathcal{S}(\mathbf{\Phi})$ uniformly at random.

1. $r < (1 - o_k(1))2^k \ln k/k \Rightarrow$ all but $o_k(1)n$ vars are *loose* w.h.p.
2. $r > (1 + o_k(1))2^k \ln k/k \Rightarrow$ all but $o_k(1)n$ vars are *frozen* w.h.p.

A geometric perspective

For $r > (1 + o_k(1))2^k \ln k/k \dots$

- …there is *shattering*.
- … $\mathcal{S}(\mathbf{\Phi})$ resembles an *error correcting code*.

The solution space geometry

Theorem [Achlioptas, ACO 2008]

Choose $\sigma \in \mathcal{S}(\mathbf{\Phi})$ uniformly at random.

1. $r < (1 - o_k(1))2^k \ln k/k \Rightarrow$ all but $o_k(1)n$ vars are *loose* w.h.p.
2. $r > (1 + o_k(1))2^k \ln k/k \Rightarrow$ all but $o_k(1)n$ vars are *frozen* w.h.p.

A geometric perspective

For $r > (1 + o_k(1))2^k \ln k/k \ldots$

- ...there is *shattering*.
- ...$\mathcal{S}(\mathbf{\Phi})$ resembles an *error correcting code*.

Molloy 2012: *exact* threshold for freezing in random graph colouring.

The solution space geometry

Theorem [ACO, Zdeborova 2012]

There is a *condensation transition* in random hypergraph 2-coloring at

$$r_{cond} = 2^{k-1} \ln 2 - \ln 2 + o_k(1).$$

- Large clusters dominate.
- Belief Propagation breaks.

Message passing algorithms

- The best "combinatorial" algorithm succeeds for $m/n < 2^k \ln(k)/k$.
- Do BP/SP perform better?

Message passing algorithms

- The best "combinatorial" algorithm succeeds for $m/n < 2^k \ln(k)/k$.
- Do BP/SP perform better?

Theorem [ACO 2011]

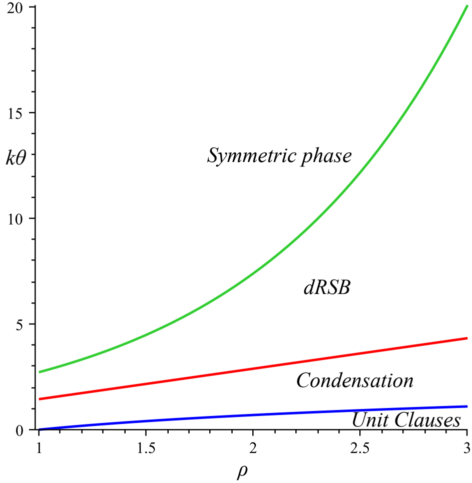
BP guided decimation fails for

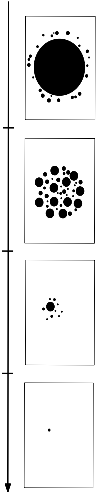
$$m/n > c \cdot 2^k/k$$

for some constant $c > 0$ w.h.p.

The phase diagram [ACO, Pachon-Pinzon 2011]

Let $m/n = \varrho \cdot 2^k/k$, $\theta = 1 - t/n$ =fraction of *unassigned* vars.





Open problems

- The BP/SP equations.
- Simulated annealing and sampling.
- Local computations.
- Hiding solutions.

The BP/SP equations

- Can we use BP/SP to compute *marginals*?

The BP/SP equations

- Can we use BP/SP to compute *marginals*?
- BP equations $\rightsquigarrow$ we can get the *partition function*

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{E}\left[\ln \#\text{solutions}\right].$$

- (*Condensation* should be a barrier for BP.)

The BP/SP equations

- Can we use BP/SP to compute *marginals*?
- BP equations $\rightsquigarrow$ we can get the *partition function*

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{E}\left[\ln \#\text{solutions}\right].$$

- (*Condensation* should be a barrier for BP.)
- SP equations $\rightsquigarrow$ we can find the *complexity*

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{E}\left[\ln \#\text{clusters}\right]$$

The BP/SP equations

- Can we use BP/SP to compute *marginals*?
- BP equations $\rightsquigarrow$ we can get the *partition function*

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{E}\left[\ln \#\text{solutions}\right].$$

- (*Condensation* should be a barrier for BP.)
- SP equations $\rightsquigarrow$ we can find the *complexity*

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{E}\left[\ln \#\text{clusters}\right]$$

- . . . and thus the *exact threshold*.

The BP/SP equations

- Can we use BP/SP to compute *marginals*?
- BP equations $\rightsquigarrow$ we can get the *partition function*

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{E}\left[\ln \#\text{solutions}\right].$$

- (*Condensation* should be a barrier for BP.)
- SP equations $\rightsquigarrow$ we can find the *complexity*

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{E}\left[\ln \#\text{clusters}\right]$$

- ...and thus the *exact threshold*.

Montanari, Shah 2007: BP works for $m/n < \ln(k)/k$.

Simulated annealing

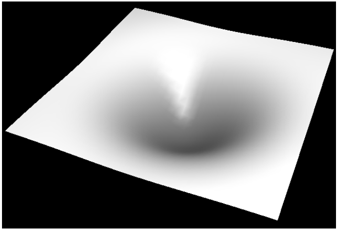
Question

Is *freezing/shattering* an algorithmic barrier?

Simulated annealing

Question

Is *freezing/shattering* an algorithmic barrier?

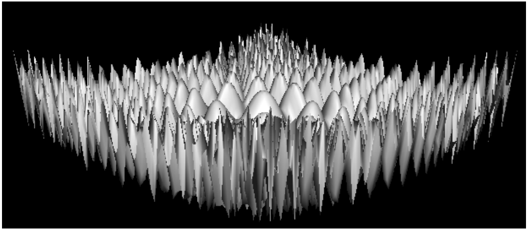


- *Symmetric regime:* $\mathcal{S}(\mathbf{\Phi})$ is the bottom of a smooth "valley".
- Conjecture: simulated annealing succeeds in *polynomial time*.
- In fact, simulated annealing should sample.

Simulated annealing

Question

Is *freezing/shattering* an algorithmic barrier?



- *Shattered regime:* $\mathcal{S}(\mathbf{\Phi})$ is the bottom of a ragged landscape.
- Conjecture: simulated annealing requires *exponential time*.
- (Very) partial results: Jerrum 1992; ACO, Efthymiou 2011.

Local computations

Question

Is *freezing/shattering* an algorithmic barrier?

Local algorithms

- f : {formulas with *tree factor graphs*} $\to [0,1]$.

Local computations

Question

Is *freezing/shattering* an algorithmic barrier?

Local algorithms

- $f : \{$formulas with *tree factor graphs*$\} \to [0, 1]$.
- For $i = 1, \ldots, n$ do
- let $p_i = f($depth ω neighborhood of $x_i)$.
- set x_i to true w.p. p_i.
- simplify the formula.

Local computations

Question

Is *freezing/shattering* an algorithmic barrier?

Local algorithms

- f : {formulas with *tree factor graphs*} $\to [0, 1]$.
- For $i = 1, \dots, n$ do
- let $p_i = f$(depth ω neighborhood of x_i).
- set x_i to true w.p. p_i.
- simplify the formula.

Can such *local algorithms* succeed beyond $c \cdot 2^k/k$?

Hiding solutions

Observation [Achlioptas, ACO 2008]

- Suppose it's difficult to find *frozen* solutions.

Hiding solutions

Observation [Achlioptas, ACO 2008]

- Suppose it's difficult to find *frozen* solutions.
- We can easily *generate* a pair $(\boldsymbol{\Phi}, \sigma)$ such that
  - $\boldsymbol{\Phi}$ is uniformly random.
  - $\sigma \in \mathcal{S}(\boldsymbol{\Phi})$.
  - σ is frozen.

Hiding solutions

Observation [Achlioptas, ACO 2008]

- Suppose it's difficult to find *frozen* solutions.
- We can easily *generate* a pair $(\mathbf{\Phi}, \sigma)$ such that
  - $\mathbf{\Phi}$ is uniformly random.
  - $\sigma \in \mathcal{S}(\mathbf{\Phi})$.
  - σ is frozen.
- $\rightsquigarrow$ identification scheme based on SAT.

Conclusion

- Physics hypotheses $\rightsquigarrow$ *blueprint* for a very nice theory.
- Precise thresholds, message passing algorithms, ...
- Open problems:
  - a rigorous *cavity method*,
  - understanding *message passing algorithms*,
  - analysing "recurrent" algorithms.
- Are random problems *easy* or *hard*?